Лащевська Н. О., Зінченко О. В., Коротков С. С.

ТЕОРІЯ РОЗПІЗНАВАННЯ ОБРАЗІВ ТА КЛАСИФІКАЦІЇ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ

Частина 1. Подання, перетворення, реставрація та ознаковий опис образів

Навчальний посібник

Лащевська Н. О., Зінченко О. В., Коротков С. С.

ТЕОРІЯ РОЗПІЗНАВАННЯ ОБРАЗІВ ТА КЛАСИФІКАЦІЇ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ

Частина 1. Подання, перетворення, реставрація та ознаковий опис образів

Навчальний посібник

Київ
2026

УДК 004.93:004.8(075.8)
Т33

Рекомендовано до видання рішенням Вченої ради Національного університету біоресурсів і природокористування України (протокол № 4 від «22» жовтня 2026 р.)

Рецензенти:

Сторчак Каміла Павлівна, д.т.н., професор, завідувач кафедри інформаційних систем та технологій, Державний університет інформаційно-комунікаційних технологій, Київ, Україна;

Приходько Ірина Олександрівна, к.т.н., доцент, доцент кафедри прикладної радіоелектроніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;

Шкарупило Вадим Вікторович, д.т.н., професор, професор кафедри комп'ютерних систем, мереж та кібербезпеки, Національний університет біоресурсів і природокористування України, Київ, Україна.

Т33 Теорія розпізнавання образів та класифікації в системах штучного інтелекту. Частина 1. Подання, перетворення, реставрація та ознаковий опис образів : навч. посіб. / Н. О. Лащевська, О. В. Зінченко, С. С. Коротков. Київ : НУБіП України, 2026. 250 с.

При підготовці навчального видання використовувалися інструменти штучного інтелекту виключно з метою стилістичного редагування тексту, перевірки орфографії, граматики та вдосконалення мовного оформлення рукопису. Зміст навчального видання, наукові положення, висновки та авторські матеріали розроблені авторами самостійно. Остаточну перевірку та редагування тексту здійснено авторами.

Посібник забезпечує змістовий модуль 1 «Подання, перетворення, реставрація та ознаковий опис образів» робочої програми навчальної дисципліни «Теорія розпізнавання образів та класифікації в системах штучного інтелекту» та її лабораторний практикум. Розглянуто постановку задачі розпізнавання образів і місце розпізнавання в системах штучного інтелекту; подання сигналів і зображень, дискретизацію і квантування; лінійні моделі формування і деградації образів, моделі шумів і критерії якості; дискретні ортогональні перетворення (Фур'є, Уолша — Адамара, Хаара, косинусне, Карунена — Лоева, RTF); частотно-часовий аналіз і вейвлети; методи попереднього оброблення зображень; методи реставрації образів, зокрема метод умовної деконволюції в базисах перетворень Фур'є, RTF і Адамара; розпізнавання за еталоном на основі кореляційних методів, погодженої фільтрації та нормального ортогонального перетворення; формування, нормування і відбір ознак. Кожен розділ містить обчислювальні приклади мовою Python, контрольні запитання і завдання для

самостійної роботи; у додатках наведено лабораторний практикум, тестові завдання і термінологічний словник.

Для здобувачів вищої освіти першого (бакалаврського) рівня, які навчаються за спеціальностями галузі знань «Інформаційні технології», а також для здобувачів другого (магістерського) рівня, аспірантів і фахівців, які працюють у галузі оброблення сигналів, комп’ютерного зору та штучного інтелекту.

УДК 004.93:004.8(075.8)

© Лащевська Н. О., Зінченко О. В.,
Коротков С. С., 2026
© НУБіП України, 2026

ВІДОМОСТІ ПРО АВТОРІВ



Лащевська Наталія Олександрівна — к.т.н., доцент, доцент кафедри комп’ютерних наук, Національний університет біоресурсів і природокористування України. Дисципліни, що викладає: Теорія розпізнавання образів та класифікації в системах штучного інтелекту, Методологія та організація наукових досліджень.
Наукові інтереси: Інформаційні технології, цифрова обробка зображень, машинне навчання, нейронні мережі.
Основні здобутки: понад 50 наукових публікацій у фахових виданнях та виданнях, які індексуються у базі Scopus та WoS, в тому числі понад 20 праць навчально-методичного характеру. Електронна адреса: lastchevska@nubip.edu.ua



Зінченко Ольга Валеріївна — д.т.н, професор, завідувач кафедри комп'ютерних наук, Національний університет біоресурсів і природокористування України. Дисципліни, що викладає: Методи та системи штучного інтелекту
Технології штучного інтелекту.
Наукові інтереси: Інформаційні технології, штучний інтелект, машинне навчання, нейронні мережі, кіберзахист із використанням машинного навчання, розпізнавання рукописного тексту, глибоке навчання, обробка зображень. Основні здобутки: понад 70 наукових публікацій у фахових виданнях та виданнях, які індексуються у базі Scopus, в тому числі 7 навчальних посібників).
Електронна адреса: olga.zinchenko@nubip.edu.ua



Коротков Сергій Станіславович — PhD, доцент
Кафедра комп’ютерних наук
Національний університет біоресурсів і природокористування України
Дисципліни, що викладає:
Технології розробки інформаційних управляючих систем (ICT)
3D моделювання
Методи комп’ютерного моделювання інтелектуальних систем
Наукові інтереси: Інформаційні технології, машинне навчання, комп’ютерні мережі, програмування
Основні здобутки: понад 10 наукових публікацій у фахових виданнях та виданнях, які індексуються у базі Scopus, в тому числі, 3 монографії та 3 навчальні посібники.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ 17
ПЕРЕДМОВА 19
ВСТУП 21
РОЗДІЛ 1. ЗАДАЧА РОЗПІЗНАВАННЯ ОБРАЗІВ У СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ 23
1.1. Розпізнавання образів як складова штучного інтелекту 23
1.2. Основні поняття: образ, клас, ознака 24
1.3. Типові задачі розпізнавання 26
1.4. Узагальнена структура системи розпізнавання 27
1.5. Підходи до розпізнавання образів 28
1.5.1. Розпізнавання за еталоном (шаблонний підхід) 28
1.5.2. Статистичний підхід 29
1.5.3. Структурний (синтаксичний) підхід 29
1.5.4. Нейромережевий підхід і глибоке навчання 29
1.5.5. Гібридні системи 30
1.6. Міри подібності та відстані в просторі ознак 30
1.7. Навчання, узагальнення та перенавчання 32
1.8. Показники якості розпізнавання 33
1.8.1. Матриця помилок 33
1.8.2. Помилки першого і другого роду 33
1.8.3. ROC-аналіз 34
1.9. Галузі застосування 35
1.10. Програмний інструментарій 35
1.11. Приклади розв’язання задач 36
Висновки до розділу 1 37
РОЗДІЛ 2. СИГНАЛИ ТА ЗОБРАЖЕННЯ ЯК НОСІЇ ІНФОРМАЦІЇ ПРО ОБРАЗ 39
2.1. Сигнал як фізичний носій образу 39
2.2. Зображення як двовимірний сигнал 41
2.2.1. Модель формування зображення 41
2.2.2. Цифрове зображення 41
2.3. Дискретизація сигналів 42
2.3.1. Теорема відліків 42
2.3.2. Спектр дискретизованого сигналу і накладання спектрів 42
2.3.3. Дискретизація зображень 43
2.3.4. Дискретизація і похибки диференціювання 44
2.4. Квантування 45
2.4.1. Рівномірне квантування і шум квантування 45

2.4.2. Візуальні ефекти квантування зображень 45
2.5. Подання кольорових зображень 46
2.5.1. Колірна модель RGB 46
2.5.2. Моделі HSV і HSI 46
2.5.3. Моделі YCbCr і Lab 47
2.5.4. Багатоспектральні зображення 48
2.6. Статистичні характеристики зображень 48
2.6.1. Гістограма 48
2.6.2. Моменти і ентропія 49
2.6.3. Кореляційні характеристики 49
2.7. Просторові відношення між пікселями 50
2.7.1. Сусідство і зв'язність 50
2.7.2. Відстані на растрі 51
2.8. Особливості реєстрації образів у різних фізичних системах 51
2.9. Практична робота з цифровими зображеннями в Python 52
2.10. Приклади розв'язання задач 53
Висновки до розділу 2 54
РОЗДІЛ 3. ЛІНІЙНІ СИСТЕМИ ФОРМУВАННЯ ОБРАЗУ ТА МОДЕЛІ ЙОГО ДЕГРАДАЦІЇ 56
3.1. Причини спотворення образів 56
3.2. Лінійні системи та імпульсна характеристика 57
3.2.1. Принцип суперпозиції 57
3.2.2. Ізопланарні та анізопланарні системи 57
3.2.3. Модель деградації з адитивним шумом 58
3.3. Дискретна згортка та її матричне подання 59
3.3.1. Дискретна згортка 59
3.3.2. Теплицева і циркулянтна матриці 60
3.3.3. Матричне подання двовимірної згортки 61
3.4. Моделі імпульсних характеристик і їх ідентифікація 62
3.4.1. Аналітичні моделі 62
3.4.2. Експериментальна ідентифікація 63
3.5. Моделі шуму та його ідентифікація 64
3.5.1. Типи шумів 64
3.5.2. Статистичний опис шуму 65
3.5.3. Практичні методи оцінювання дисперсії шуму 66
3.6. Критерії якості образів 67
3.7. Некоректність оберненої задачі 68
3.8. Моделювання деградації образу в Python 70
3.9. Приклади розв'язання задач 70

Висновки до розділу 3 71
РОЗДІЛ 4. ОРТОГОНАЛЬНІ ПЕРЕТВОРЕННЯ. ДИСКРЕТНЕ ПЕРЕТВОРЕННЯ ФУР'Є 74
4.1. Образ як вектор і його розклад за базисом 74
4.1.1. Основна ідея ортогональних перетворень 74
4.1.2. Ортогональність, нормованість і повнота 74
4.1.3. Матрична форма перетворення 75
4.1.4. Перетворення матриць операторів 76
4.2. Неперервне перетворення Фур'є 76
4.3. Дискретне перетворення Фур'є 77
4.3.1. Визначення 77
4.3.2. Фізичний зміст номера трансформанти 78
4.3.3. Основні властивості ДПФ 79
4.3.4. Розтікання спектра і вікна 80
4.3.5. Ефект Гіббса 81
4.4. Теорема про циклічну згортку і діагоналізація циркулянтних матриць .. 82
4.4.1. Теорема про циклічну згортку 82
4.4.2. Матрична форма: діагоналізація 82
4.4.3. Обчислення лінійної згортки через ДПФ 83
4.5. Швидке перетворення Фур'є 83
4.6. Двовимірне дискретне перетворення Фур'є 84
4.6.1. Визначення і сепарабельність 84
4.6.2. Подання і тлумачення двовимірного спектра 85
4.6.3. Амплітуда і фаза 85
4.7. Фільтрація зображень у частотній області 86
4.7.1. Загальна схема 86
4.7.2. Фільтри нижніх частот 86
4.7.3. Фільтри верхніх частот і підвищення різкості 87
4.8. Кореляція в частотній області і суміщення образів 88
4.8.1. Взаємна кореляція 88
4.8.2. Фазова кореляція 88
4.9. Спектральні ознаки образів 89
4.10. Практична реалізація 90
4.11. Приклади розв'язання задач 91
Висновки до розділу 4 91
РОЗДІЛ 5. ДИСКРЕТНІ ОРТОГОНАЛЬНІ ПЕРЕТВОРЕННЯ З ДІЙСНИМ ЯДРОМ 94
5.1. Вступні зауваження 94
5.2. Функції та перетворення Уолша 94

5.2.1. Функції Радемахера і Уолша 94
5.2.2. Матриця Адамара і кронекерів добуток 95
5.2.3. Впорядкування функцій Уолша 96
5.3. Швидке перетворення Уолша — Адамара 97
5.4. Згортка і перетворення Уолша — Адамара 99
5.4.1. Діадна згортка 99
5.4.2. Циклічна згортка в базисі Адамара: блочно-діагональна структура 99
5.4.3. Двовимірне перетворення Уолша — Адамара 100
5.5. Перетворення Хаара 101
5.6. Дискретне косинусне перетворення 102
5.6.1. Від косинусного ряду до дискретного перетворення 102
5.6.2. Двовимірне ДКП і стиснення зображень 103
5.7. Енергетична компактність і перетворення Карунена — Лоева 104
5.7.1. Перетворення Карунена — Лоева 104
5.7.2. Чому ДКП близьке до оптимального 104
5.7.3. Порівняльний експеримент 105
5.8. Теорема про диференціювання оригіналу для дискретних ортогональних перетворень 106
5.8.1. Загальне формулювання 106
5.8.2. Розв’язання лінійних диференціальних рівнянь в області трансформант 107
5.8.3. Власні числа різницевого оператора і перетворення RTF 107
5.9. Інші перетворення з дійсним ядром 109
5.10. Порівняльна характеристика перетворень 110
5.11. Програмна реалізація 110
5.12. Приклади розв’язання задач 111
Висновки до розділу 5 112
РОЗДІЛ 6. ЧАСТОТНО-ЧАСОВИЙ АНАЛІЗ: ВІКОННЕ ПЕРЕТВОРЕННЯ ФУР’Є ТА ВЕЙВЛЕТИ 115
6.1. Стаціонарні та нестаціонарні образи 115
6.2. Локальне (віконне) перетворення Фур’є 116
6.2.1. Визначення 116
6.2.2. Роздільна здатність і принцип невизначеності 117
6.2.3. Обмеження віконного перетворення 117
6.3. Неперервне вейвлет-перетворення 118
6.3.1. Визначення 118
6.3.2. Материнські вейвлети 119
6.3.3. Вейвлети і особливості сигналів 120
6.4. Дискретне вейвлет-перетворення і кратномасштабний аналіз 120
6.4.1. Від неперервного до дискретного перетворення 120

6.4.2. Кратномасштабний аналіз ... 120
6.4.3. Алгоритм Малла ... 121
6.4.4. Двовимірне дискретне вейвлет-перетворення ... 123
6.5. Придушення шуму методом вейвлет-порогової обробки ... 123
6.5.1. Принцип ... 123
6.5.2. Вибір порогу ... 124
6.5.3. Приклади ... 124
6.6. Вейвлети і фільтри Габора у формуванні ознак ... 126
6.6.1. Вейвлет-ознаки текстури ... 126
6.6.2. Фільтри Габора ... 126
6.6.3. Частотно-часові ознаки одновимірних сигналів ... 127
6.7. Вейвлет-стиснення зображень ... 127
6.8. Програмна реалізація ... 128
6.9. Приклади розв’язання задач ... 129
Висновки до розділу 6 ... 130
РОЗДІЛ 7. ПОПЕРЕДНЄ ОБРОБЛЕННЯ ЗОБРАЖЕНЬ У ПРОСТОРОВІЙ ОБЛАСТІ ... 132
7.1. Завдання попереднього оброблення ... 132
7.2. Точкові перетворення яскравості ... 132
7.2.1. Основні функції перетворення ... 132
7.2.2. Вирівнювання гістограми ... 133
7.2.3. Адаптивне вирівнювання гістограми ... 134
7.3. Просторова фільтрація ... 135
7.3.1. Лінійні фільтри ... 135
7.3.2. Нелінійні фільтри ... 135
7.3.3. Підвищення різкості ... 136
7.4. Виділення контурів ... 137
7.4.1. Градієнтні оператори ... 137
7.4.2. Оператор Лапласа — Гаусса ... 138
7.4.3. Детектор Канні ... 138
7.5. Порогова сегментація ... 139
7.5.1. Глобальний поріг ... 139
7.5.2. Метод Оцу ... 140
7.5.3. Адаптивні пороги ... 140
7.6. Математична морфологія ... 141
7.6.1. Базові операції над бінарними образами ... 141
7.6.2. Перетворення відстані і скелет ... 142
7.6.3. Морфологія півтонових зображень ... 143
7.7. Геометричні перетворення та нормування образів ... 143

7.7.1. Афінні перетворення 143
7.7.2. Інтерполяція 143
7.8. Програмна реалізація 144
7.9. Приклади розв'язання задач 145
Висновки до розділу 7 145
РОЗДІЛ 8. РЕСТАВРАЦІЯ ОБРАЗІВ: ІНВЕРСНА І ВІНЕРІВСЬКА ФІЛЬТРАЦІЯ, УМОВНА ДЕКОНВОЛЮЦІЯ 148
8.1. Постановка задачі реставрації 148
8.2. Інверсна фільтрація 149
8.2.1. Реставрація в області трансформант Фур'є 149
8.2.2. Псевдоінверсна фільтрація 150
8.3. Фільтр Вінера 150
8.3.1. Виведення 150
8.3.2. Практичні варіанти 151
8.4. Метод умовної деконволюції в натуральних координатах 151
8.4.1. Апріорна інформація 151
8.4.2. Умова на енергію шуму 152
8.4.3. Умова гладкості 152
8.4.4. Метод множників Лагранжа і розв'язок 153
8.4.5. Обчислювальний приклад 154
8.4.6. Обчислювальна складність 155
8.5. Умовна деконволюція в області трансформант Фур'є 155
8.5.1. Перехід в частотну область 155
8.5.2. Рівняння для параметра λ 156
8.5.3. Порівняння з фільтром Вінера 157
8.5.4. Зв'язок з регуляризацією Тихонова 157
8.5.5. Обчислювальний експеримент 158
8.5.6. Крайові ефекти і ефект Гіббса 159
8.6. Узагальнена дискретна фільтрація за методом найменших квадратів 160
8.7. Деконволюція за критерієм форми результуючої імпульсної характеристики 161
8.7.1. Ідея методу 161
8.7.2. Формалізація 161
8.8. Програмна реалізація 163
8.9. Приклади розв'язання задач 164
Висновки до розділу 8 165
РОЗДІЛ 9. РЕСТАВРАЦІЯ ОБРАЗІВ У БАЗИСАХ ПЕРЕТВОРЕНЬ RTF ТА АДАМАРА 168
9.1. Вступні зауваження 168

9.2. Перетворення RTF 168
9.2.1. Теорема про диференціювання оригіналу в базисі RTF 168
9.2.2. Аналіз лінійних систем у базисі RTF 169
9.2.3. Формування частотної характеристики в базисі RTF 169
9.3. Умовна деконволюція в базисі RTF 171
9.4. Умовна деконволюція в області трансформант Адамара 172
9.4.1. Загальний вираз для довільного ортогонального базису 172
9.4.2. Структура блоків матриці G_1Had 173
9.4.3. Символьний метод формування матриць 174
9.4.4. Обернення матриці деградації 175
9.4.5. Алгоритм реставрації одновимірного образу в базисі Адамара 176
9.5. Реставрація двовимірних образів 176
9.6. Порівняння результатів реставрації в різних базисах 177
9.6.1. Результати експериментальних досліджень 177
9.6.2. Методичне зауваження: що залежить від базису, а що — ні 178
9.7. Програмна реалізація 179
9.8. Приклади розв’язання задач 180
Висновки до розділу 9 180
РОЗДІЛ 10. РОЗПІЗНАВАННЯ ЗА ЕТАЛОНОМ: КОРЕЛЯЦІЙНІ МЕТОДИ, ПОГОДЖЕНА ФІЛЬТРАЦІЯ ТА НОРМАЛЬНЕ ПЕРЕТВОРЕННЯ 183
10.1. Задача розпізнавання за еталоном 183
10.2. Кореляційні методи 183
10.2.1. Взаємна кореляція 183
10.2.2. Нормована взаємна кореляція 184
10.3. Лінійний погоджений фільтр 185
10.3.1. Виведення 185
10.3.2. Цифровий погоджений фільтр 185
10.4. Недоліки класичного погодженого фільтра та альтернативні детектори 186
10.4.1. Обмеження оптимальності 186
10.4.2. Детектори подібності 187
10.5. Нормалізація ортогональних перетворень 188
10.5.1. Ідея 188
10.5.2. Нормалізація за кроком 189
10.5.3. Нормалізація за рівнем 189
10.5.4. Коефіцієнт трансформант і гострота 190
10.6. Нормальне ортогональне перетворення 191
10.6.1. Визначення 191
10.6.2. Побудова матричного оператора 191

10.6.3. Властивості нормального спектра 192
10.6.4. Нормальне перетворення для двовимірних образів 193
10.7. Аутентифікація особи за динамічним підписом 194
10.7.1. Постановка задачі 194
10.7.2. Навчання 194
10.7.3. Аутентифікація 195
10.7.4. Експериментальні дослідження і обмеження методу 195
10.8. Порівняння лінійної погодженої та нормальної фільтрації 196
10.9. Класифікація з уточненням 197
10.10. Програмна реалізація 197
10.11. Приклади розв'язання задач 198
Висновки до розділу 10 199
РОЗДІЛ 11. ФОРМУВАННЯ, НОРМУВАННЯ ТА ВІДБІР ОЗНАК 202
11.1. Місце формування ознак у системі розпізнавання 202
11.2. Геометричні ознаки бінарних образів 202
11.3. Моменти та моментні інваріанти 203
11.3.1. Геометричні моменти 203
11.3.2. Інваріанти Ху 204
11.4. Ознаки контуру 205
11.5. Яскравісні та статистичні ознаки 206
11.6. Текстурні ознаки 206
11.6.1. Матриці суміжності рівнів яскравості 206
11.6.2. Локальні бінарні шаблони 207
11.6.3. Спектральні та багатомасштабні текстурні ознаки 207
11.7. Гістограми орієнтованих градієнтів і локальні ознаки 208
11.7.1. Гістограма орієнтованих градієнтів 208
11.7.2. Характерні точки і їх дескриптори 208
11.8. Нормування ознак 209
11.9. Зменшення розмірності 209
11.9.1. Метод головних компонент 209
11.9.2. Лінійний дискримінантний аналіз 210
11.10. Відбір ознак 211
11.10.1. Критерії інформативності 211
11.10.2. Стратегії відбору 211
11.10.3. Порівняльний експеримент 212
11.11. Практичні рекомендації щодо формування ознак 212
11.12. Програмна реалізація 213
11.13. Приклади розв'язання задач 214
Висновки до розділу 11 214

РОЗДІЛ 12. НАСКРІЗНИЙ ПРИКЛАД: РОЗПІЗНАВАННЯ СПОТВОРЕНИХ СИМВОЛІВ 217
12.1. Мета і постановка задачі 217
12.2. Структура досліджуваної системи 218
12.3. Методика експерименту 219
12.4. Результати 219
12.5. Аналіз результатів 221
12.6. Порівняння методів реставрації в складі системи розпізнавання 222
12.7. Обмеження і напрямки розвитку 223
12.8. Програмна реалізація експерименту 223
12.9. Приклад оцінювання значущості відмінностей 224
Висновки до розділу 12 225
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 227
ДОДАТОК А. ПРОГРАМНЕ СЕРЕДОВИЩЕ ТА ЛАБОРАТОРНИЙ ПРАКТИКУМ 231
А.1. Встановлення програмного середовища 231
А.2. Загальні вимоги до виконання і оформлення лабораторних робіт 231
А.3. Перелік лабораторних робіт 233
Лабораторна робота № 0. Вступ до курсу та формування командного AI-проєкту 233
Лабораторна робота № 1. Попередній аналіз та підготовка даних для систем розпізнавання образів 234
Лабораторна робота № 2. Попередня обробка даних та формування простору ознак 234
Лабораторна робота № 3. Класифікатор k найближчих сусідів (k-NN) 235
Лабораторна робота № 4. Наївний байєсівський класифікатор 235
Лабораторна робота № 5. Класифікатор на основі дерева рішень 235
Лабораторна робота № 6. Класифікатор на основі логістичної регресії 236
Лабораторна робота № 7. Комплексна оцінка та порівняння класифікаторів 236
Лабораторна робота № 8. Зниження розмірності методом головних компонент (PCA) 237
Лабораторна робота № 9. Класифікатор випадкового лісу 237
Лабораторна робота № 10. Класифікатори на основі бустингу 237
Лабораторна робота № 11. Класифікатор на основі багатошарового персептрона (MLP) 238
Лабораторна робота № 12. Фінальний захист командного проєкту 238
А.4. Шаблон блокнота лабораторної роботи 238
ДОДАТОК Б. ТЕСТОВІ ЗАВДАННЯ ДЛЯ САМОПЕРЕВІРКИ 240
Б.1. До розділу 1 240
Б.2. До розділу 2 240
Б.3. До розділу 3 241
Б.4. До розділу 4 241
Б.5. До розділу 5 241

Б.6. До розділу 6 242
Б.7. До розділу 7 242
Б.8. До розділу 8 243
Б.9. До розділу 9 243
Б.10. До розділу 10 244
Б.11. До розділу 11 244
Б.12. До розділу 12 244
Б.13. Ключ відповідей 245
ДОДАТОК В. УКРАЇНСЬКО-АНГЛІЙСЬКИЙ СЛОВНИК ОСНОВНИХ ТЕРМІНІВ 246
ДОДАТОК Г. ЗВЕДЕННЯ ОСНОВНИХ ФОРМУЛ 249

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ

Основні позначення

f, $f(n)$, $f(x,y)$ — вихідний (неспотворений) образ: вектор відліків, одновимірний сигнал, зображення

L, $L(x,y)$ — спостережуваний (деградований, спотворений) образ

$\hat{\mathrm{f}}$ — оцінка образу (результат реставрації)

$g(x,y)$, G — імпульсна характеристика системи відображення та матричний оператор згортки

V, $V(x,y)$ — адитивний шум; σ^2 — його дисперсія; ε^2 — енергія шуму

C, $C(k)$ — оператор (і його спектр) умови гладкості

λ — множник Лагранжа (параметр методу умовної деконволюції)

W, $\mathrm{W_н}$ — матричний оператор ортогонального перетворення (нормований)

F, $\mathrm{F_н}$ — матричний оператор ДПФ (нормований)

H_N, $\mathrm{Had_н}$ — матриця Адамара порядку N (нормована)

$X(k)$ — трансформанти (спектральні коефіцієнти) образу

β_k — власні частоти перетворення RTF

x — вектор ознак; ω_k — k-й клас

R — кореляційна (коваріаційна) матриця

E — одинична матриця

$*$ — згортка; $\otimes$ — кронекерів добуток

M{·} — математичне сподівання

$k_{\text{тр}}$ — коефіцієнт трансформант

Скорочення

ВПФ — віконне перетворення Фур’є

ДВП — дискретне вейвлет-перетворення

ДКП — дискретне косинусне перетворення

ДПУА — дискретне перетворення Уолша — Адамара

ДПФ — дискретне перетворення Фур’є

КЛП — перетворення Карунена — Лоева

КМА — кратномасштабний аналіз

НВП — неперервне вейвлет-перетворення

ПУА — перетворення Уолша — Адамара

ФНЧ, ФВЧ — фільтр нижніх, верхніх частот

ФРТ — функція розсіювання точки

ШІ — штучний інтелект

ШПФ — швидке перетворення Фур’є

ШПУА — швидке перетворення Уолша — Адамара

AUC — площа під ROC-кривою (*Area Under Curve*)

CLAHE — адаптивне вирівнювання гістограми з обмеженням контрасту

EER — рівень рівних помилок (*Equal Error Rate*)

FAR, FRR — частки хибних прийнять і хибних відмов

GLCM — матриця суміжності рівнів яскравості

HOG — гістограма орієнтованих градієнтів

LBP — локальний бінарний шаблон
LDA — лінійний дискримінантний аналіз
LoG — оператор Лапласа — Гаусса
MSE — середньоквадратична похибка
NCC — нормована взаємна кореляція
PCA — метод головних компонент
PSNR — пікове відношення сигнал/шум
ROC — робоча характеристика приймача
RTF — дискретне перетворення з власними частотами різницевого оператора
SNR — відношення сигнал/шум
SSIM — індекс структурної подібності
STFT — віконне (короткочасне) перетворення Фур’є

ПЕРЕДМОВА

Навчальний посібник підготовлено для забезпечення змістового модуля 1 «Подання, перетворення, реставрація та ознаковий опис образів» робочої програми навчальної дисципліни «Теорія розпізнавання образів та класифікації в системах штучного інтелекту» та її лабораторного практикуму. Дисципліна поєднує дві традиції, що історично розвивалися паралельно: радіотехнічну традицію цифрового оброблення сигналів і зображень, у якій центральне місце посідають лінійні системи, ортогональні перетворення, фільтрація і реставрація, та традицію теорії розпізнавання образів і машинного навчання, що зосереджується на поданні образів у просторі ознак, статистичних вирішальних правилах і навчанні за прикладами. Сучасні системи штучного інтелекту, які сприймають навколишній світ через камери, мікрофони, радіолокатори і медичні сканери, потребують обох складових: без якісного подання і попереднього оброблення образів найдосконаліший класифікатор неспроможний працювати надійно, а без класифікатора оброблений образ залишається лише «гарним зображенням».

Відповідно до робочої програми дисципліна складається з двох змістових модулів, тому посібник видається у двох частинах. Перша частина, що пропонується увазі читача, забезпечує змістовий модуль 1 і присвячена шляху від фізичного сигналу до вектора ознак: поданню образів, моделям їх формування і спотворення, ортогональним перетворенням, попередньому обробленню і реставрації, найпростішим методам розпізнавання за еталоном і формуванню ознак. Друга частина забезпечуватиме змістовий модуль 2 і буде присвячена статистичним і детермінованим класифікаторам, кластеризації, нейромережевим методам і глибокому навчанню.

Особливістю посібника є детальний виклад методів реставрації образів, зокрема методу умовної деконволюції в базисах перетворень Фур’є, RTF і Адамара, а також методів нормалізації ортогональних перетворень і нормального перетворення для розпізнавання за еталоном. Ці розділи ґрунтуються на результатах наукових досліджень кафедри радіоприймання та оброблення сигналів Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» [1–12], які раніше були викладені в дисертаційній роботі [2], фахових публікаціях і конспекті лекцій [1]. У посібнику цей матеріал перероблено, систематизовано і пов’язано із загальною теорією розпізнавання образів; усі запозичені положення супроводжуються посиланнями на першоджерела. Загальнотеоретичний матеріал викладено з урахуванням класичних підручників і монографій з цифрового оброблення зображень [13–16], теорії розпізнавання образів [17–21], ортогональних перетворень [22, 23] і вейвлет-аналізу [24, 25], а також сучасної літератури з комп’ютерного зору і машинного навчання [21, 26–32].

Посібник має виразну практичну спрямованість. Усі рисунки, що ілюструють роботу методів, отримано в результаті обчислювальних експериментів, проведених спеціально для цього видання; більшість розділів містить фрагменти програм мовою Python, які дозволяють відтворити ці

експерименти. Розділ 12 містить наскрізний приклад дослідження системи розпізнавання спотворених символів, у якому поєднано майже всі розглянуті методи.

Кожний розділ починається з переліку результатів навчання і завершується висновками, контрольними запитаннями та завданнями для самостійної роботи різного рівня складності — від перевірки окремих формул до невеликих дослідницьких проєктів. У додатку А наведено опис програмного середовища і лабораторного практикуму, побудованого як наскрізний командний проєкт (лабораторні роботи № 0–12), у додатку Б — тестові завдання для самоперевірки з ключами відповідей, у додатку В — українсько-англійський словник основних термінів.

Для успішного опанування матеріалу необхідні знання з вищої математики (лінійна алгебра, математичний аналіз, теорія ймовірностей), основ теорії сигналів і програмування. Розділи 1–3 та 7 можуть вивчатися з мінімальною математичною підготовкою; розділи 4–6 і 8–10 потребують упевненого володіння матричним численням.

Автори висловлюють щиру подяку колегам, у співпраці з якими були отримані результати, що лягли в основу розділів 8–10, та рецензентам за цінні зауваження.

Особистий внесок авторів: Н. О. Лащевська — передмова, вступ, розділи 2–5, 8–10, розділ 12 (спільно з С. С. Коротковим), додаток Г, загальне редагування. Це подання і перетворення зображень, моделі спотворень, ортогональні перетворення та реставрація. Розділи 8–10 спираються на її дисертацію, а 2–5 — на цифрову обробку зображень; О. В. Зінченко — розділи 1, 6, 7, 11, додаток В. Це основи розпізнавання та місце в системах ШІ, вейвлети, попереднє оброблення й сегментація, формування ознак. Вони відповідають її інтересам у штучному інтелекті, обробці зображень і розпізнаванні рукописного тексту; С. С. Коротков — програмні реалізації обчислювальних прикладів, розділ 12 (спільно з Н. О. Лащевською), додатки А і Б. Це лабораторний практикум, тести й код на Python, що відповідає його інтересам у програмуванні та комп'ютерному моделюванні.

ВСТУП

Розпізнавання образів — одна з найстаріших і водночас найдинамічніших галузей штучного інтелекту. Її історія налічує близько семи десятиліть: від перших читальних автоматів і перцептрона 1950-х років через статистичну теорію розпізнавання 1960–1970-х, структурні методи і експертні системи 1980-х, методи опорних векторів і ансамблі класифікаторів 1990–2000-х до глибоких нейронних мереж, які з 2010-х років є домінуючим підходом у галузі [17–21, 26, 27]. За цей час змінювалися домінуючі підходи, але незмінною залишалася базова схема: образ реєструється, перетворюється, описується ознаками і класифікується.

Розвиток глибокого навчання іноді створює враження, що класичні методи оброблення сигналів і зображень втратили значення: мережа нібито сама «навчиться» всього необхідного. Досвід показує, що це не так. По-перше, базові операції нейронних мереж — згортка, підвибірка, нормування — це ті самі операції цифрового оброблення сигналів, і їх розуміння необхідне для свідомого проєктування мереж [16, 26, 29]. По-друге, у багатьох прикладних задачах (медична діагностика, радіолокація, промисловий контроль, біометрія) обсяги розмічених даних обмежені, а вимоги до надійності, пояснюваності і передбачуваності поведінки системи високі; тут моделі, що спираються на фізику формування образу, мають суттєві переваги [2, 33]. По-третє, реальні умови експлуатації завжди відрізняються від умов навчання: змінюється освітлення, з’являються розмиття і шум. Методи, що явно моделюють ці чинники і компенсують їх (реставрація, нормування, інваріантні ознаки), підвищують стійкість систем будь-якого типу [13, 28]. Наскрізний приклад розділу 12 демонструє це кількісно.

Матеріал посібника структуровано відповідно до шляху образу через систему розпізнавання.

Розділ 1 вводить основні поняття теорії розпізнавання (образ, клас, ознака, простір ознак), класифікує задачі і підходи, описує структуру системи розпізнавання, міри подібності та показники якості.

Розділ 2 присвячено сигналам і зображенням як носіям інформації про образ: їх класифікації, дискретизації і квантуванню, колірним моделям, статистичним характеристикам, просторовим відношенням між пікселями.

Розділ 3 описує лінійну модель формування і деградації образу, матричне подання згортки, моделі імпульсних характеристик і шумів, методи їх ідентифікації, критерії якості, а також некоректність оберненої задачі реставрації.

Розділ 4 містить загальну теорію ортогональних перетворень і детальний виклад дискретного перетворення Фур’є, його швидких алгоритмів, двовимірного варіанта, фільтрації в частотній області і спектральних ознак.

Розділ 5 розглядає перетворення з дійсним ядром — Уолша — Адамара, Хаара, косинусне, Карунена — Лоева, — а також теорему про диференціювання оригіналу в довільному базисі і перетворення RTF.

Розділ 6 присвячено частотно-часовому аналізу: віконному перетворенню Фур'є, неперервному і дискретному вейвлет-перетворенням, придушенню шуму і формуванню ознак на їх основі.

Розділ 7 описує методи попереднього оброблення зображень у просторовій області: точкові перетворення, фільтрацію, виділення контурів, сегментацію, математичну морфологію, геометричне нормування.

Розділи 8 і 9 присвячено реставрації образів: інверсній і вінерівській фільтрації, методу умовної деконволюції в натуральних координатах і в області трансформант Фур'є, узагальненій фільтрації в довільному базисі, деконволюції за критерієм форми імпульсної характеристики, а також модифікаціям методу умовної деконволюції в базисах перетворень RTF і Адамара.

Розділ 10 розглядає розпізнавання за еталоном: кореляційні методи, погоджену фільтрацію і детектори подібності, нормалізацію ортогональних перетворень, нормальне перетворення і його застосування для аутентифікації особи за динамічним підписом.

Розділ 11 присвячено формуванню ознак (геометричних, моментних, контурних, текстурних, градієнтних), їх нормуванню, зменшенню розмірності і відбору.

Розділ 12 містить наскрізний приклад — дослідження системи розпізнавання спотворених символів з реставрацією.

Розділи посібника пов'язані між собою, але допускають і вибіркове вивчення. Для курсу з орієнтацією на комп'ютерний зір рекомендовано послідовність 1, 2, 3, 4, 7, 6, 11, 10, 12; для курсу з орієнтацією на оброблення сигналів — 1–6, 8–10, 12.

РОЗДІЛ 1. ЗАДАЧА РОЗПІЗНАВАННЯ ОБРАЗІВ У СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ

Навчальні цілі. Після вивчення розділу ви зможете: пояснити зміст понять «образ», «клас», «ознака», «простір ознак»; сформулювати основні типи задач розпізнавання; описати узагальнену структуру системи розпізнавання та етапи її життєвого циклу; порівняти основні підходи до розпізнавання; обирати міри подібності та показники якості класифікації відповідно до умов задачі.

1.1. Розпізнавання образів як складова штучного інтелекту

Здатність розрізняти предмети, звуки, запахи, ситуації та відносити їх до відомих категорій є однією з базових властивостей живих організмів. Людина впізнає обличчя знайомих у натовпі, відрізняє рукописну цифру «1» від «7», за звуком двигуна визначає несправність автомобіля, а лікар за рентгенограмою встановлює наявність патологічних змін. У кожному з цих випадків відбувається перехід від сукупності первинних вимірювань (освітленостей точок сітківки, коливань тиску повітря, концентрацій молекул) до висновку про належність спостережуваного об’єкта до певного класу. Саме цей перехід і становить предмет теорії розпізнавання образів.

Перші технічні системи розпізнавання з’явилися в середині XX століття разом із цифровими обчислювальними машинами. Вже в 1950-х роках створювалися пристрої читання друкованих символів, а модель перцептрона Ф. Розенблатта [29] стала однією з перших спроб побудувати навчувану систему, що імітує механізми зорового сприйняття. У 1960–1970-х роках сформувалися статистична теорія розпізнавання [17, 18, 21] та структурний (синтаксичний) підхід, була розроблена більшість класичних методів попереднього оброблення зображень [13–15]. Сучасний етап пов’язаний з методами машинного навчання та глибокими нейронними мережами [19, 26], які в багатьох прикладних задачах досягли або перевершили точність людини-експерта.

У структурі штучного інтелекту (ШІ) розпізнавання образів посідає особливе місце. Якщо під ШІ розуміти здатність технічної системи сприймати навколишнє середовище, будувати його внутрішнє подання, приймати рішення та діяти для досягнення цілей [27], то розпізнавання забезпечує саме першу ланку цього ланцюга — *сприйняття*. Інтелектуальний агент (мобільний робот, система підтримки лікарських рішень, безпілотний автомобіль) отримує від сенсорів потоки «сирих» даних: зображення, відеопослідовності, акустичні сигнали, показання датчиків. Перш ніж планувати дії, агент мусить перетворити ці дані на символьний опис: «попереду пішохід», «на знімку новоутворення діаметром 8 мм», «голос належить зареєстрованому користувачу». Без надійного розпізнавання всі наступні рівні інтелектуальної обробки втрачають сенс.

Водночас розпізнавання образів не зводиться лише до вибору класифікатора. Досвід розроблення реальних систем свідчить, що якість кінцевого рішення значною мірою визначається етапами, які передують класифікації: умовами реєстрації сигналу, якістю попереднього оброблення, способом формування ознак [13, 18]. Зашумлене, розфокусоване чи неоднорідно

освітлене зображення може зробити безсилим навіть найдосконаліший класифікатор. Саме тому перша частина посібника присвячена поданню образів, їх перетворенням, попередньому обробленню, реставрації та формуванню ознак, а також найпростішим методам розпізнавання за еталоном. Друга частина буде присвячена статистичним класифікаторам, кластеризації та нейромережевим методам.

1.2. Основні поняття: образ, клас, ознака

Термін «образ» (англ. *pattern*) у літературі вживається в кількох значеннях. У теорії розпізнавання образом називають формалізований опис об'єкта (явища, процесу, ситуації), поданий у вигляді, придатному для оброблення технічною системою [17]. Наприклад, образом рукописної літери є матриця яскравостей пікселів її зображення; образом мовного звуку — послідовність відліків акустичного сигналу; образом пацієнта в діагностичній системі — набір результатів аналізів. Важливо розуміти, що образ — це не сам об'єкт, а його *подання*, яке залежить від способу спостереження і може містити спотворення.

У радіотехнічній та медичній літературі, а також у працях з реставрації [2, 33], образом часто називають безпосередньо одновимірний сигнал або двовимірне зображення, отримане системою відображення. У цьому посібнику обидва вживання є рівноправними: коли йдеться про реставрацію чи перетворення, «образ» означає сигнал або зображення; коли йдеться про класифікацію — формалізований опис об'єкта загалом.

Класом (категорією) називають множину образів, що мають спільні суттєві властивості з точки зору поставленої задачі. Класи позначатимемо $\omega_1, \omega_2, \dots, \omega_K$, де K — кількість класів. Поділ на класи завжди визначається метою: ті самі зображення облич можна розбити на класи «чоловік/жінка», «посміхається/не посміхається» або на класи, що відповідають окремим особам. Отже, клас є не властивістю об'єктів самою по собі, а результатом домовленості, закріпленої в постановці задачі.

Ознакою (англ. *feature*) називають кількісну або якісну характеристику образу, яка використовується для прийняття рішення про його належність до класу. Ознаки поділяють на:

- *кількісні (числові)* — площа об'єкта, середня яскравість, частота основного тону, енергія спектральної смуги;
- *порядкові* — ступінь вираженості (низький, середній, високий);
- *номінальні (категоріальні)* — колір із фіксованого переліку, тип текстури;
- *бінарні* — наявність або відсутність певного елемента (отвору в символі, кута на контурі).

Сукупність n числових ознак утворює вектор ознак

$$\mathrm{x} = [x_1, x_2, \dots, x_n]^T, \tag{1.1}$$

де T — знак транспонування. Множину всіх можливих значень вектора ознак називають простором ознак $X \subseteq \mathbb{R}^n$. Кожний образ відображається точкою цього простору, а клас — деякою областю (або хмарою точок). Геометрична інтерпретація дозволяє наочно сформулювати мету розпізнавання: поділити

простір ознак на області $\Omega_1, \ldots, \Omega_K$ так, щоб образи кожного класу потрапляли переважно до «своєї» області.

На рис. 1.1 наведено два характерні випадки розташування класів у двовимірному просторі ознак. У першому з них (рис. 1.1, а) класи можна розділити прямою лінією — такі класи називають *лінійно роздільними*. У другому (рис. 1.1, б) один клас охоплює інший, і жодна пряма не здатна їх розділити, проте задачу розв'язує нелінійна межа (коло). Цей простий приклад ілюструє важливу думку: складність класифікатора, необхідного для розв'язання задачі, визначається не лише природою об'єктів, а й вдалістю обраних ознак. Якщо в другому випадку замість декартових координат (x_1, x_2) використати відстань до початку координат $r = \sqrt{x_1^2 + x_2^2}$, класи стануть роздільними за однією ознакою з простим порогом.

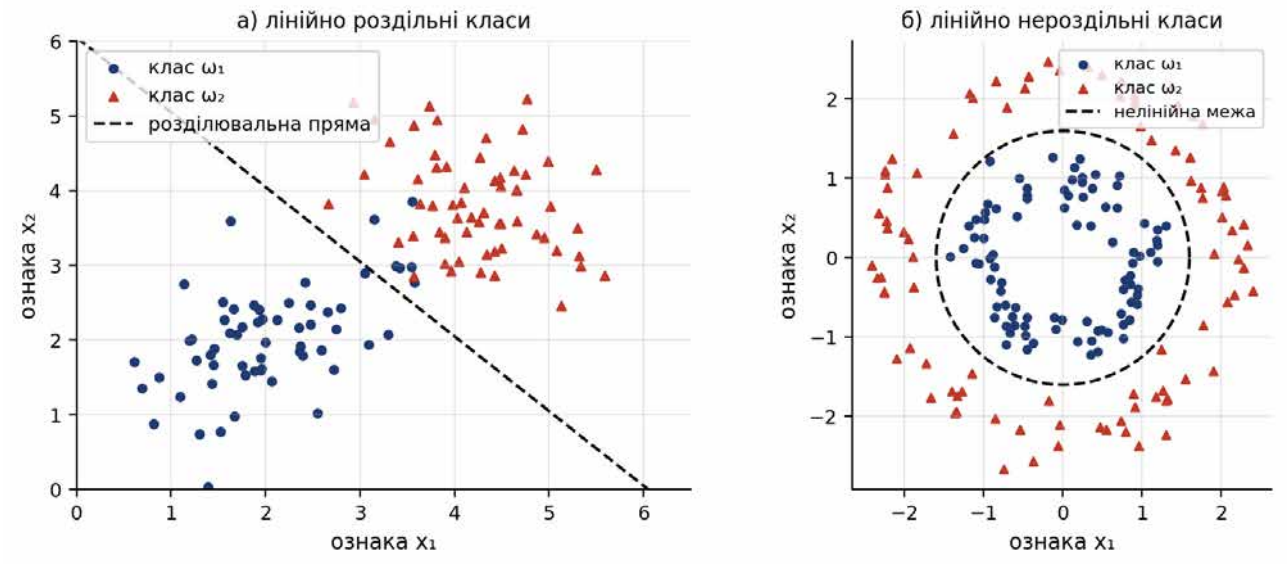


Рис. 1.1. Розташування класів у просторі ознак

Розрізняють також первинне подання образу і його ознаковий опис. Первинне подання — це те, що безпосередньо надходить від сенсора: для цифрового зображення розміром 512×512 пікселів це вектор розмірності 262 144. Безпосередньо класифікувати такі вектори вкрай складно: вони мають величезну розмірність, містять надлишкову інформацію та сильно залежать від умов реєстрації (зсуву, масштабу, освітлення). Тому між первинним поданням і класифікатором розміщують етапи попереднього оброблення та формування ознак, мета яких — отримати компактний, інформативний і стійкий до заважаючих чинників опис.

Вимоги до «хороших» ознак можна сформулювати так [18, 20]:

1. *Інформативність (розрізнювальна здатність)* — значення ознаки для різних класів мають суттєво відрізнятися.
2. *Стійкість (інваріантність)* — значення ознаки мають мало змінюватися під дією чинників, що не стосуються класу: зсуву, повороту, зміни масштабу, яскравості, шуму.
3. *Незалежність* — ознаки не повинні дублювати одна одну; сильно корельовані ознаки збільшують розмірність, не додаючи інформації.

4. *Обчислювальна доступність* — ознака має обчислюватися з прийнятними витратами часу та пам'яті.
5. *Інтерпретованість* — у багатьох застосуваннях (медицина, безпека) бажано, щоб ознака мала зрозумілий фізичний або змістовний сенс.

Ці вимоги часто суперечать одна одній. Наприклад, повна інваріантність до повороту робить неможливим розрізнення цифр «6» і «9». Тому вибір ознак завжди є компромісом, який визначається конкретною задачею.

1.3. Типові задачі розпізнавання

Залежно від того, яку інформацію має система на етапі навчання і яке рішення вона повинна прийняти, виділяють кілька основних типів задач [17, 20].

Класифікація (розпізнавання з учителем). Задано скінченну множину класів $\{\omega_1, \dots, \omega_K\}$ і навчальну вибірку — множину образів із відомими мітками класів:

$$T = \{(\mathrm{x}_i, y_i)\}_{i=1}^{M}, \quad y_i \in \{1, \dots, K\}. \tag{1.2}$$

Потрібно побудувати *вирішальне правило* (класифікатор) $d: X \to \{1, \dots, K\}$, яке з найменшою кількістю помилок відносить нові, не бачені під час навчання образи до класів. Приклади: розпізнавання рукописних цифр, діагностика захворювань за медичними зображеннями, класифікація типів земної поверхні на космічних знімках.

Кластеризація (розпізнавання без учителя, таксономія). Мітки класів невідомі; іноді невідома навіть кількість класів. Потрібно розбити множину образів на групи (кластери) так, щоб образи всередині групи були схожі між собою і відрізнялися від образів інших груп. Кластеризація використовується для сегментації зображень, пошуку типових режимів роботи обладнання, попереднього аналізу даних [20, 21].

Верифікація (аутентифікація). Система має перевірити твердження «цей образ належить класу ω_k», наприклад: «підпис належить особі з логіном X». Відповідь бінарна — «так» або «ні». Аутентифікація є основою біометричних систем доступу; приклад такої системи на основі динамічного підпису розглянуто в розділі 10 [1].

Ідентифікація. Потрібно визначити, якому з багатьох відомих класів належить образ (наприклад, кому з бази даних належить відбиток пальця), або встановити, що образ не належить жодному з них (задача з «відкритою множиною»).

Виявлення (детекція). Потрібно встановити наявність об'єкта заданого класу у сцені та визначити його положення: пошук облич на фотографії, виявлення цілі на радіолокаційному зображенні, локалізація дефектів на поверхні виробу. Класичними методами виявлення сигналу відомої форми є погоджена фільтрація та кореляційні методи (розділ 10).

Сегментація. Потрібно поділити зображення на області, що відповідають окремим об'єктам або їх частинам, тобто фактично класифікувати кожний піксель.

Відновлення (реставрація) образу формально не є задачею класифікації, проте тісно з нею пов'язане. Якщо зареєстрований образ спотворений

неідеальною системою відображення і шумом, то перед розпізнаванням доцільно отримати оцінку неспотвореного образу [2, 33]. Методам реставрації присвячено розділи 3, 8 і 9.

Окремо слід згадати задачі *регресії* (оцінювання неперервної величини за образом, наприклад, віку людини за фотографією) та *структурного опису* (побудови опису сцени як сукупності об’єктів і відношень між ними). Вони виходять за межі цього посібника, але використовують ті самі методи подання та попереднього оброблення.

1.4. Узагальнена структура системи розпізнавання

Незважаючи на різноманіття застосувань, більшість систем розпізнавання мають подібну функціональну структуру (рис. 1.2). Розглянемо призначення кожного блоку.



Рис. 1.2. Узагальнена структура системи розпізнавання образів

Сенсор (система реєстрації) перетворює фізичні властивості об’єкта на електричний сигнал і далі — на цифрові дані. Для зображень це оптична система і матриця фотоприймачів, для медичних зображень — рентгенівський, ультразвуковий чи магнітно-резонансний апарат, для радіолокації — антенна система і приймач. На цьому етапі відбуваються дискретизація та квантування (розділ 2), а також неминучі спотворення: розмиття через неточкову імпульсну характеристику системи, геометричні викривлення, шум (розділ 3) [13, 14].

Попереднє оброблення має на меті поліпшити якість первинних даних і привести їх до стандартного вигляду. Сюди відносять придушення шуму, корекцію контрасту і яскравості, реставрацію розмитих зображень, геометричне нормування (усунення зсуву, повороту, зміни масштабу), сегментацію — виділення об’єкта інтересу з фону. Методам попереднього оброблення присвячено розділи 4–9.

Формування ознак перетворює попередньо оброблений образ на вектор ознак (1.1). Ознаки можуть бути геометричними (площа, периметр, моменти), яскравісними (статистики гістограми), текстурними, спектральними (коефіцієнти ортогональних перетворень) тощо (розділ 11). Часто після обчислення великої кількості «кандидатів» виконують *відбір*

найінформативніших ознак або *зменшення розмірності* (наприклад, методом головних компонент).

Класифікатор реалізує вирішальне правило, що зіставляє вектору ознак мітку класу. Простішими є класифікатори за еталоном (мінімальною відстанню), складнішими — статистичні (байєсівські), класифікатори на основі опорних векторів, дерев рішень, нейронних мереж [18, 19, 26].

Інтерпретація рішення перетворює мітку класу на дію або повідомлення, враховуючи вартість помилок, контекст та додаткову інформацію.

Життєвий цикл системи розпізнавання включає два принципово різні етапи. На етапі навчання (проєктування) за навчальною вибіркою визначають параметри попереднього оброблення, вибирають ознаки та налаштовують класифікатор. На етапі експлуатації система обробляє нові образи з фіксованими параметрами. Між ними обов'язково має бути етап тестування — оцінювання якості на незалежній вибірці, яка не використовувалася під час навчання. У деяких системах (наприклад, в біометричних) навчання триває і під час експлуатації: успішно розпізнані образи поповнюють базу еталонів, що дозволяє враховувати повільні зміни характеристик [1].

Важливо усвідомлювати, що помилка на будь-якому етапі поширюється на всі наступні. Розглянемо простий приклад. Якщо система розпізнавання номерних знаків автомобілів отримує розмите внаслідок руху зображення, сегментація символів виконується неточно, ознаки окремих символів спотворюються, і класифікатор помиляється — хоча сам класифікатор може бути бездоганним. Усунення розмиття методами реставрації (розділи 8, 9) на етапі попереднього оброблення може радикально підвищити точність усієї системи.

1.5. Підходи до розпізнавання образів

Історично сформувалося кілька підходів до побудови систем розпізнавання, які відрізняються способом подання образів і характером вирішальних правил [17, 18, 20].

1.5.1. Розпізнавання за еталоном (шаблонний підхід)

Найпростіший і найстаріший підхід полягає в порівнянні образу з еталонами (шаблонами) класів. Для кожного класу ω_k зберігається еталон s_k — типовий представник або усереднений образ класу. Невідомий образ x відносять до класу, еталон якого найбільш подібний до x:

$$d(\mathrm{x}) = \arg\min_k \rho(\mathrm{x}, \mathrm{s}_k), \tag{1.3}$$

де $\rho(\cdot,\cdot)$ — міра відмінності (відстань). Еквівалентно можна максимізувати міру подібності, наприклад коефіцієнт кореляції. Шаблонний підхід добре працює, коли образи одного класу мало відрізняються один від одного і від еталона, а умови реєстрації контрольовані: розпізнавання друкованих символів фіксованого шрифту, пошук відомого фрагмента на зображенні, аутентифікація за підписом після нормування. Його математичною основою є кореляційні методи, погоджена фільтрація та нормальні ортогональні перетворення, детально розглянуті в розділі 10.

1.5.2. Статистичний підхід

У статистичному підході образ розглядається як реалізація випадкового вектора, а кожний клас описується розподілом імовірностей $p(\mathrm{x} \mid \omega_k)$ та апріорною ймовірністю $P(\omega_k)$. Оптимальне (у сенсі мінімуму ймовірності помилки) вирішальне правило — байєсівське — відносить образ до класу з максимальною апостеріорною ймовірністю:

$$d(\mathrm{x}) = \operatorname{argmax}_k P(\omega_k \mid \mathrm{x}) = \operatorname{argmax}_k p(\mathrm{x} \mid \omega_k)\, P(\omega_k). \qquad (1.4)$$

Якщо розподіли ознак у класах є нормальними, правило (1.4) зводиться до порівняння квадратичних форм — відстаней Махаланобіса до центрів класів [18, 21, 34]. Класифікатор за критерієм максимальної правдоподібності на основі кореляційної матриці відліків сигналу розглянуто в [1]; ця тема детально висвітлюватиметься в другій частині посібника.

1.5.3. Структурний (синтаксичний) підхід

У структурному підході образ подається як композиція простіших елементів (примітивів) і відношень між ними. Наприклад, контур символу описується послідовністю відрізків і дуг, а хромосома на мікроскопічному зображенні — послідовністю сегментів певної форми. Класи задаються граматиками, які породжують допустимі ланцюжки примітивів, а розпізнавання зводиться до синтаксичного аналізу. Структурний підхід ефективний для образів з чітко вираженою ієрархічною будовою (креслення, схеми, рукописні тексти), але чутливий до шуму та помилок виділення примітивів [17, 20].

1.5.4. Нейромережевий підхід і глибоке навчання

Штучні нейронні мережі є універсальними апроксиматорами функцій і можуть навчатися вирішальних правил безпосередньо за прикладами. Сучасні згорткові нейронні мережі об’єднують етапи формування ознак і класифікації: ознаки (фільтри згорткових шарів) не задаються розробником, а «вивчаються» з даних [26]. Це дозволило досягти видатних результатів у розпізнаванні зображень, мовлення, тексту. Проте нейромережевий підхід вимагає великих обсягів розмічених даних, значних обчислювальних ресурсів, а отримані моделі часто важко інтерпретувати.

Слід підкреслити, що навіть у системах глибокого навчання методи, які розглядаються в першій частині посібника, не втрачають значення. Згортка — базова операція згорткової мережі — є тією самою дискретною згорткою, що описує формування та фільтрацію зображень (розділ 3). Фільтри перших шарів навченої мережі зазвичай нагадують детектори контурів і фільтри Габора (розділи 6, 7). Нормування яскравості та контрасту, придушення шуму, аугментація навчальних даних геометричними перетвореннями є стандартними етапами конвеєра навчання. Розуміння класичних методів дозволяє свідомо проєктувати і діагностувати сучасні системи [29, 30].

1.5.5. Гібридні системи

На практиці найчастіше застосовують комбінації підходів: наприклад, попереднє оброблення і сегментацію класичними методами, формування ознак за допомогою ортогональних перетворень, а класифікацію — статистичним або нейромережевим класифікатором. Ще одним прикладом є багатоступеневі системи, в яких швидкий, але менш надійний класифікатор приймає рішення в «очевидних» випадках, а складніший і точніший вмикається лише тоді, коли перший не впевнений. Такий алгоритм класифікації з уточненням, що поєднує класифікатор на основі нормального перетворення з класифікатором за критерієм максимальної правдоподібності, запропоновано в [1] (див. розділ 10).

1.6. Міри подібності та відстані в просторі ознак

Майже всі методи розпізнавання явно чи неявно спираються на поняття «близькості» образів. Функцію $\rho(\mathrm{x},\mathrm{y})$ називають метрикою (відстанню), якщо для будь-яких векторів $\mathrm{x},\mathrm{y},\mathrm{z}$ виконуються аксіоми:

1. $\rho(\mathrm{x},\mathrm{y}) \geq 0$, причому $\rho(\mathrm{x},\mathrm{y}) = 0$ тоді й лише тоді, коли $\mathrm{x} = \mathrm{y}$;
2. $\rho(\mathrm{x},\mathrm{y}) = \rho(\mathrm{y},\mathrm{x})$ (симетрія);
3. $\rho(\mathrm{x},\mathrm{z}) \leq \rho(\mathrm{x},\mathrm{y}) + \rho(\mathrm{y},\mathrm{z})$ (нерівність трикутника).

Найуживанішим є сімейство метрик Мінковського

$$\rho_p(\mathrm{x},\mathrm{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}, \quad p \geq 1. \tag{1.5}$$

При $p = 2$ отримуємо евклідову відстань, при $p = 1$ — манхеттенську (відстань міських кварталів), при $p \to \infty$ — відстань Чебишова $\rho_\infty = \max_i |x_i - y_i|$. Форму «одиничних кіл» цих метрик, тобто множин точок, рівновіддалених від початку координат, показано на рис. 1.3. Від вибору метрики залежить форма областей, які класифікатор за мінімальною відстанню відносить до одного класу. Манхеттенська відстань менш чутлива до великих відхилень окремих ознак (викидів), ніж евклідова; відстань Чебишова визначається найгіршою з ознак [18, 20].

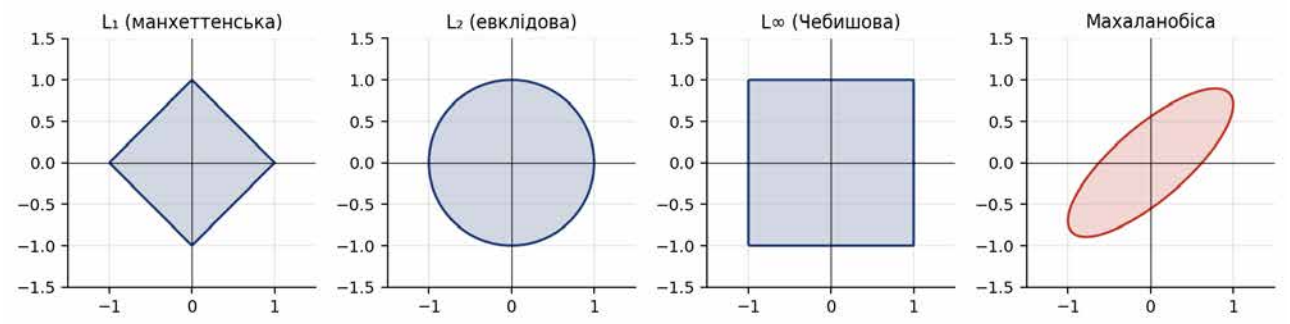


Рис. 1.3. Одиничні «кола» різних метрик у двовимірному просторі ознак

Метрики Мінковського мають суттєвий недолік: вони не враховують ні різних масштабів ознак, ні кореляції між ними. Якщо одна ознака вимірюється в міліметрах і змінюється в межах сотень одиниць, а інша є безрозмірним коефіцієнтом з діапазоном від 0 до 1, то евклідова відстань визначатиметься

практично лише першою ознакою. Найпростішим засобом боротьби з цим є нормування ознак, наприклад, приведення до нульового середнього та одиничного середньоквадратичного відхилення (z-нормування):

$$\tilde{x}_i = \frac{x_i - \mu_i}{\sigma_i}, \tag{1.6}$$

де μ_i і σ_i — середнє значення та середньоквадратичне відхилення i-ї ознаки, оцінені за навчальною вибіркою.

Більш загальним рішенням є відстань Махаланобіса [34]

$$\rho_M(\mathrm{x}, \mu) = \sqrt{(\mathrm{x} - \mu)^T \mathrm{R}^{-1} (\mathrm{x} - \mu)}, \tag{1.7}$$

де μ — вектор математичних сподівань (центр класу), R — коваріаційна (кореляційна) матриця ознак класу. Відстань Махаланобіса враховує і різні дисперсії ознак, і їх взаємну кореляцію: поверхні рівної відстані є еліпсоїдами, орієнтованими вздовж головних осей розсіювання класу (рис. 1.3, праворуч). Для нормального розподілу ознак квадрат відстані Махаланобіса безпосередньо пов'язаний з густиною ймовірності, тому вона відіграє центральну роль у статистичному розпізнаванні. Практична складність полягає в оберненні матриці R, яка для сигналів з великою кількістю відліків зазвичай є погано обумовленою; методи подолання цієї складності за допомогою розкладу на власні вектори та нормального перетворення наведено в [1].

Поряд із відстанями широко використовують міри подібності, які набувають найбільшого значення для ідентичних образів. Найважливішими з них є:

- скалярний добуток (некоригована кореляція) $s(\mathrm{x}, \mathrm{y}) = \mathrm{x}^T \mathrm{y} = \sum_i \mathrm{x}_i y_i$ — основа класичного погодженого фільтра;
- косинусна подібність

$$\cos\theta = \frac{\mathrm{x}^T \mathrm{y}}{\| \mathrm{x} \| \| \mathrm{y} \|}, \tag{1.8}$$

яка залежить лише від кута між векторами і не залежить від їх довжини, тобто інваріантна до зміни масштабу (контрасту) образу;

- коефіцієнт кореляції Пірсона — косинусна подібність центрованих векторів (з відніманням середніх значень), інваріантна і до масштабу, і до зсуву рівня (яскравості):

$$r(\mathrm{x}, \mathrm{y}) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}. \tag{1.9}$$

Для бінарних векторів застосовують відстань Геммінга — кількість позицій, у яких вектори відрізняються, та коефіцієнт Жаккара — відношення кількості спільних одиниць до кількості позицій, де хоча б в одному векторі є одиниця [20].

Між евклідовою відстанню та скалярним добутком існує простий зв'язок:

$$\| \mathrm{x} - \mathrm{y} \|^2 = \| \mathrm{x} \|^2 + \| \mathrm{y} \|^2 - 2\,\mathrm{x}^T \mathrm{y}. \tag{1.10}$$

Якщо норми векторів фіксовані (наприклад, усі образи попередньо нормовані до одиничної енергії), то мінімізація евклідової відстані еквівалентна

максимізації скалярного добутку і косинусної подібності. Якщо ж норми змінюються, ці критерії дають різні результати. Цей, на перший погляд, простий факт має важливі наслідки для порівняння детекторів подібності сигналів, що розглядатиметься в розділі 10.

Приклад 1.1. Нехай еталон s=[1,2,3]^T, а спостережувані образи x_1=[2,4,6]^T (той самий образ з подвоєним контрастом) і x_2=[1,3,2]^T. Евклідові відстані: ρ(x_1,s)=√(1+4+9)≈3,74, ρ(x_2,s)=√(0+1+1)≈1,41 — за евклідовою мірою ближчим є x_2. Косинусна подібність: cosθ_1=28/(√56 √14)=1, cosθ_2=13/(√14 √14)≈0,93 — за косинусною мірою x_1 ідентичний еталону. Яка міра правильна, залежить від того, чи є зміна контрасту суттєвою для задачі.

1.7. Навчання, узагальнення та перенавчання

Мета навчання системи розпізнавання — не запам’ятати навчальну вибірку, а побудувати правило, яке правильно класифікує *нові* образи. Здатність правильно працювати на даних, не використаних під час навчання, називають узагальнювальною здатністю.

Розглянемо класифікатор, який просто запам’ятовує всі навчальні образи і відносить новий образ до класу найближчого з них (правило найближчого сусіда). На навчальній вибірці він не помиляється ніколи, проте на нових даних його точність може виявитися невисокою, особливо якщо навчальні образи містять шум або помилки розмітки. Ситуацію, коли модель надто точно підлаштовується під особливості навчальної вибірки, зокрема під випадковий шум, і внаслідок цього гірше працює на нових даних, називають перенавчанням (англ. *overfitting*). Протилежна ситуація — недонавчання — виникає, коли модель надто проста, щоб відобразити реальні відмінності між класами [19, 35].

Для контролю узагальнювальної здатності наявні дані поділяють на частини:

- навчальна вибірка — для налаштування параметрів класифікатора;
- валідаційна вибірка — для вибору структури моделі, ознак і так званих гіперпараметрів (наприклад, кількості сусідів у методі k найближчих сусідів, параметра регуляризації);
- тестова вибірка — для остаточного, незалежного оцінювання якості; її не можна використовувати для жодних рішень щодо побудови системи.

Якщо обсяг даних малий, застосовують перехресну перевірку (крос-валідацію): дані ділять на Q частин, навчають Q разів, щоразу залишаючи одну частину для контролю, і усереднюють результати [35].

Особливо важливою проблемою для розпізнавання зображень і сигналів є прокляття розмірності [35]. Зі зростанням кількості ознак обсяг простору ознак зростає експоненційно, і навчальна вибірка фіксованого обсягу стає все більш «розрідженою». Емпіричне правило свідчить, що кількість навчальних образів кожного класу має принаймні в 5–10 разів перевищувати кількість ознак; для статистичних класифікаторів, що оцінюють коваріаційні матриці, ця вимога ще жорсткіша. Наприклад, якщо образом є сигнал зі $N = 256$ відліками, то оцінка коваріаційної матриці розміром 256×256 за 20–30 реалізаціями принципово неможлива: матриця буде виродженою. Звідси випливає необхідність зменшення

розмірності (розділ 11) або спеціальних методів, що дозволяють обійти обернення коваріаційної матриці [1].

1.8. Показники якості розпізнавання

1.8.1. Матриця помилок

Результати тестування класифікатора на K класів зручно подавати у вигляді матриці помилок (англ. *confusion matrix*) C розміром $K \times K$, елемент c_{ij} якої дорівнює кількості образів істинного класу i, віднесених класифікатором до класу j. Діагональні елементи відповідають правильним рішенням, позадіагональні — помилкам. Аналіз матриці помилок показує не лише загальну кількість помилок, а й те, які саме класи плутаються між собою, що часто підказує, яких ознак бракує [35, 36].

Для двох класів («позитивний» — наявність об'єкта, «свій» користувач, хворий пацієнт; «негативний» — відсутність) матриця помилок містить чотири величини (табл. 1.1).

Таблиця 1.1

Матриця помилок для бінарної класифікації

Істинний клас	Рішення «позитивний»	Рішення «негативний»
Позитивний	TP (істинно позитивні)	FN (хибно негативні, пропуск)
Негативний	FP (хибно позитивні, хибна тривога)	TN (істинно негативні)

На їх основі обчислюють основні показники:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}; \tag{1.11}$$

$$\text{Precision} = \frac{TP}{TP + FP}; \qquad \text{Recall} = \text{TPR} = \frac{TP}{TP + FN}; \tag{1.12}$$

$$F_1 = \frac{2\,\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}; \qquad \text{FPR} = \frac{FP}{FP + TN}. \tag{1.13}$$

Точність (accuracy) — частка правильних рішень загалом. Вона є зручним, але оманливим показником для незбалансованих класів: якщо лише 1 % пацієнтів хворі, класифікатор, який завжди відповідає «здоровий», має точність 99 %, будучи абсолютно некорисним. *Влучність* (precision) показує, яка частка позитивних рішень правильна; *повнота*, або *чутливість* (recall, TPR), — яку частку справжніх позитивних образів виявлено. *Специфічність* дорівнює $1 -$ FPR. Показник F_1 є гармонічним середнім влучності і повноти.

1.8.2. Помилки першого і другого роду

У задачах верифікації та виявлення прийнято термінологію перевірки статистичних гіпотез. Нехай основна гіпотеза H_1 — «образ належить заявленому класу» (підпис справжній), альтернативна H_0 — «образ чужий». Помилка першого роду полягає у відхиленні правильної гіпотези: справжній користувач

не отримує доступу (англ. *False Rejection Rate*, FRR). Помилка другого роду — прийняття хибної гіпотези: доступ отримує зловмисник (англ. *False Acceptance Rate*, FAR). Зауважимо, що в різних галузях нумерація родів помилок може відрізнятися залежно від того, яку гіпотезу вважають основною; тому завжди слід явно пояснювати, про яку помилку йдеться [1].

Більшість систем верифікації обчислюють деяку міру подібності (або відмінності) образу та еталона і порівнюють її з порогом θ. Зміна порогу змінює співвідношення помилок: жорсткіший поріг зменшує FAR, але збільшує FRR (рис. 1.4, а). Точку, в якій FAR = FRR, називають рівнем рівних помилок (EER, *Equal Error Rate*); вона часто слугує узагальненою характеристикою біометричної системи. Вибір робочого порогу визначається вартістю помилок: у системі доступу до банківського сховища допустимо частіше відмовляти справжнім користувачам, ніж пропустити зловмисника.

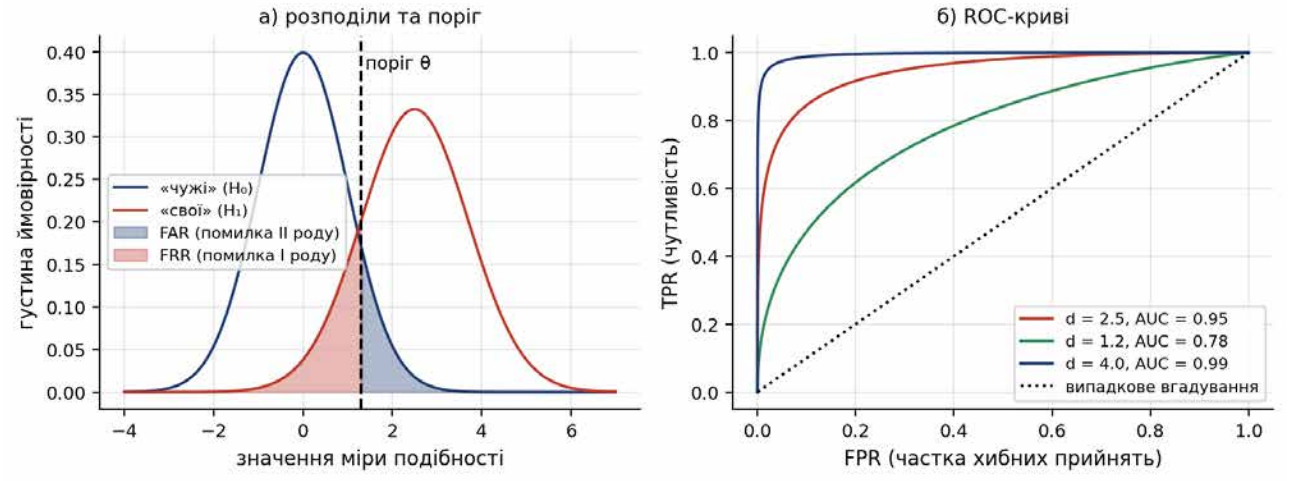


Рис. 1.4. Вибір порогу та ROC-криві для різної роздільності класів

1.8.3. ROC-аналіз

Залежність чутливості TPR від частки хибних тривог FPR при зміні порогу від $-\infty$ до $+\infty$ називають ROC-кривою (*Receiver Operating Characteristic*) [36]. Термін походить з радіолокації, де аналогічні криві («робочі характеристики приймача») використовували для оцінювання якості виявлення сигналів на фоні шуму. ROC-крива ідеального класифікатора проходить через точку $(0,1)$; діагональ відповідає випадковому вгадуванню. Інтегральною характеристикою є площа під ROC-кривою (AUC), що дорівнює ймовірності того, що випадково обраний позитивний образ отримає вищу оцінку, ніж випадково обраний негативний. На рис. 1.4, б показано ROC-криві для трьох значень параметра роздільності d — відстані між центрами розподілів оцінок для «своїх» і «чужих», віднесеної до їх розсіювання. Чим більша роздільність, тим ближче крива до лівого верхнього кута і тим більша AUC. Саме підвищення роздільності — головна мета попереднього оброблення та вдалого вибору ознак.

1.9. Галузі застосування

Методи розпізнавання образів застосовуються практично в усіх галузях, де виникає потреба в автоматичному аналізі сигналів і зображень [28, 31, 32]. Наведемо найважливіші з них.

Медична діагностика. Аналіз рентгенограм, комп'ютерних і магнітно-резонансних томограм, ультразвукових зображень, гістологічних препаратів; аналіз біосигналів (електрокардіограм, електроенцефалограм). Особливістю медичних застосувань є висока ціна помилки та обмежена якість первинних даних: зображення спотворені розмиттям через рух пацієнта, неточкову імпульсну характеристику апаратури, шум [2, 33, 37]. Тому реставрація образів тут має особливе значення.

Біометрія і безпека. Ідентифікація та верифікація особи за обличчям, відбитками пальців, райдужною оболонкою, голосом, динамічним підписом [1]; відеоспостереження; виявлення заборонених предметів на рентгенівських зображеннях багажу.

Дистанційне зондування Землі. Класифікація типів земної поверхні на супутникових знімках, моніторинг сільськогосподарських угідь, виявлення наслідків надзвичайних ситуацій, картографування [13, 28].

Радіолокація і радіонавігація. Виявлення і розпізнавання цілей за радіолокаційними сигналами та зображеннями, класифікація радіовипромінювань.

Промисловість. Автоматичний контроль якості продукції, виявлення дефектів поверхонь, сортування, навігація промислових роботів [31].

Транспорт. Розпізнавання номерних знаків, дорожніх знаків, пішоходів і перешкод у системах допомоги водієві та автономних транспортних засобах.

Обробка документів. Оптичне розпізнавання друкованих і рукописних текстів, аналіз структури документів, розпізнавання формул і таблиць.

1.10. Програмний інструментарій

Практичні приклади в посібнику наведено мовою Python, яка стала фактичним стандартом у галузі машинного навчання та обробки даних. Використовуються такі бібліотеки:

- NumPy [38] — робота з багатовимірними масивами, лінійна алгебра, швидке перетворення Фур'є;
- SciPy [39] — обробка сигналів (scipy.signal), багатовимірна фільтрація (scipy.ndimage), спеціальні функції, оптимізація;
- Matplotlib — побудова графіків і відображення зображень;
- scikit-image [40] — алгоритми обробки зображень і тестові зображення;
- PyWavelets [41] — вейвлет-перетворення;
- scikit-learn [42] — класифікатори, кластеризація, метрики якості, зменшення розмірності.

Встановлення середовища та перелік лабораторних робіт наведено в додатку А. Нижче показано мінімальний приклад: обчислення матриці помилок

і основних показників якості для бінарного класифікатора з пороговим правилом.

```
import numpy as np

rng = np.random.default_rng(0)
# оцінки подібності для "своїх" (1) та "чужих" (0) образів
s_pos = rng.normal(2.5, 1.2, 500)
s_neg = rng.normal(0.0, 1.0, 500)
scores = np.r_[s_pos, s_neg]
labels = np.r_[np.ones(500), np.zeros(500)]

theta = 1.3                    # поріг рішення
pred = (scores >= theta).astype(int)
TP = np.sum((pred == 1) & (labels == 1))
FN = np.sum((pred == 0) & (labels == 1))
FP = np.sum((pred == 1) & (labels == 0))
TN = np.sum((pred == 0) & (labels == 0))
acc = (TP + TN) / len(labels)
prec, rec = TP / (TP + FP), TP / (TP + FN)
FAR, FRR = FP / (FP + TN), FN / (FN + TP)
print(f"Accuracy={acc:.3f} Precision={prec:.3f} Recall={rec:.3f}")
print(f"FAR={FAR:.3f} FRR={FRR:.3f}")
```

Змінюючи поріг theta у циклі та зберігаючи пари (FAR, 1 − FRR), можна побудувати ROC-криву і знайти рівень рівних помилок.

1.11. Приклади розв’язання задач

Задача 1.1. Класифікатор медичних зображень протестовано на 1000 пацієнтах, з яких 50 хворих. Класифікатор правильно виявив 45 хворих, а 30 здорових помилково визнав хворими. Обчисліть показники якості.

Розв’язання. Складемо матрицю помилок: $TP = 45$, $FN = 50 - 45 = 5$, $FP = 30$, $TN = 950 - 30 = 920$. Тоді за (1.11)–(1.13):

- точність Accuracy $= (45 + 920)/1000 = 0{,}965$;
- влучність Precision $= 45/(45 + 30) = 0{,}60$;
- повнота (чутливість) Recall $= 45/(45 + 5) = 0{,}90$;
- $F_1 = 2 \cdot 45/(2 \cdot 45 + 30 + 5) = 0{,}72$;
- частка хибних тривог FPR $= 30/(30 + 920) \approx 0{,}032$.

Висока точність (96,5 %) створює оманливе враження: лише 60 % позитивних рішень правильні. Для медичного скринінгу ці показники можуть бути прийнятними (пропущено лише 10 % хворих), але кожен позитивний результат потребує додаткової перевірки.

Задача 1.2. Два класи мають центри $\mu_1 = [0{,}0]^T$, $\mu_2 = [2{,}0]^T$ і однакову коваріаційну матрицю $R = \begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{bmatrix}$. До якого класу слід віднести образ $x = [1{,}1;\ 1{,}0]^T$ за мінімумом евклідової відстані і за мінімумом відстані Махаланобіса?

Розв'язання. Квадрати евклідових відстаней: $\rho_1^2 = 1,1^2 + 1,0^2 = 2,21$; $\rho_2^2 = 0,9^2 + 1,0^2 = 1,81$. За евклідовою мірою образ належить класу ω_2.

Обернена коваріаційна матриця: $\det R = 1 - 0,64 = 0,36$, $R^{-1} = \frac{1}{0,36}\begin{bmatrix} 1 & -0,8 \\ -0,8 & 1 \end{bmatrix} \approx \begin{bmatrix} 2,778 & -2,222 \\ -2,222 & 2,778 \end{bmatrix}$. Квадрати відстаней Махаланобіса (1.7):

$\rho_{M1}^2 = 2,778 \cdot 1,21 - 2 \cdot 2,222 \cdot 1,1 + 2,778 \cdot 1,0 \approx 1,25$;

$\rho_{M2}^2 = 2,778 \cdot 0,81 - 2 \cdot 2,222 \cdot (-0,9) \cdot 1,0 + 2,778 \cdot 1,0 \approx 9,03$.

За відстанню Махаланобіса образ належить класу ω_1. Пояснення: ознаки сильно корельовані, тож образи класу ω_1 розташовані вздовж діагоналі $x_1 = x_2$; точка $[1,1;\ 1,0]$ лежить майже на цій діагоналі і є «типовою» для ω_1, хоча евклідово ближча до центра ω_2.

Висновки до розділу 1

1. Розпізнавання образів забезпечує ланку сприйняття в системах штучного інтелекту: воно перетворює первинні дані сенсорів на символьний опис, на основі якого приймаються рішення.
2. Образ — це формалізоване подання об'єкта, придатне для оброблення; клас — множина образів, об'єднаних за суттєвою для задачі ознакою; ознака — характеристика образу, що використовується для прийняття рішення. Геометрично задача класифікації зводиться до поділу простору ознак на області класів.
3. Основними типами задач є класифікація, кластеризація, верифікація, ідентифікація, виявлення та сегментація; реставрація образів є важливою допоміжною задачею.
4. Система розпізнавання включає реєстрацію, попереднє оброблення, формування ознак, класифікацію та інтерпретацію рішення. Якість кінцевого результату значною мірою визначається етапами, що передують класифікації.
5. Розрізняють шаблонний, статистичний, структурний та нейромережевий підходи; на практиці переважають гібридні системи.
6. Вибір міри відстані чи подібності визначає, які зміни образу система вважатиме суттєвими. Відстань Махаланобіса враховує масштаби та кореляцію ознак, косинусна міра — інваріантна до масштабу.
7. Якість системи оцінюють на незалежній тестовій вибірці за допомогою матриці помилок, показників точності, влучності, повноти, помилок першого і другого роду та ROC-аналізу.

Питання для обговорення та самоперевірки

1. Що розуміють під образом у теорії розпізнавання? Чим образ відрізняється від самого об'єкта?
2. Чому поділ образів на класи визначається метою задачі, а не лише властивостями об'єктів? Наведіть приклад.
3. Які вимоги висувають до ознак? Чому ці вимоги можуть суперечити одна одній?

4. Що таке простір ознак? Як у ньому інтерпретуються образ і клас?
5. Поясніть різницю між задачами класифікації, кластеризації, верифікації та ідентифікації.
6. Назвіть основні блоки системи розпізнавання та їх призначення.
7. Чим відрізняються етапи навчання, тестування та експлуатації системи розпізнавання?
8. Сформулюйте вирішальне правило класифікатора за мінімальною відстанню до еталона.
9. У чому полягає байєсівське вирішальне правило?
10. Які переваги та недоліки має структурний підхід?
11. Чому методи класичного оброблення сигналів залишаються актуальними для систем глибокого навчання?
12. Запишіть метрику Мінковського. Які окремі випадки вона включає?
13. У яких випадках доцільно використовувати відстань Махаланобіса? Яка практична складність пов'язана з її обчисленням?
14. Чим косинусна подібність відрізняється від коефіцієнта кореляції Пірсона?
15. Що таке перенавчання? Як його виявити?
16. Поясніть зміст прокляття розмірності.
17. Чому точність (accuracy) може бути неадекватним показником якості?
18. Що таке помилки першого і другого роду в задачі верифікації? Що таке EER?
19. Як будується ROC-крива? Що характеризує площа під нею?
20. Наведіть приклади застосування розпізнавання образів у вашій майбутній професійній діяльності.

Завдання для самостійної роботи

1. Для векторів $\mathrm{x} = [3,1,4,1]^T$ і $\mathrm{y} = [2,7,1,8]^T$ обчисліть відстані ρ_1, ρ_2, ρ_∞, косинусну подібність і коефіцієнт кореляції Пірсона.
2. Згенеруйте дві двовимірні нормально розподілені вибірки з різними коваріаційними матрицями. Побудуйте класифікатор за мінімальною евклідовою відстанню до центрів класів і за мінімальною відстанню Махаланобіса. Порівняйте точність на тестовій вибірці.
3. Доведіть співвідношення (1.10). Покажіть, що для векторів одиничної норми $\| \mathrm{x} - \mathrm{y} \|^2 = 2(1 - \cos\theta)$.
4. Користуючись прикладом коду з підрозділу 1.10, побудуйте ROC-криву, обчисліть AUC та рівень рівних помилок.
5. Наведіть приклад задачі, в якій ознака має бути інваріантною до повороту, і приклад, де така інваріантність шкідлива.
6. Для задачі медичної діагностики, в якій хворі становлять 2 % обстежуваних, проаналізуйте, які показники якості слід використовувати і чому.

РОЗДІЛ 2. СИГНАЛИ ТА ЗОБРАЖЕННЯ ЯК НОСІЇ ІНФОРМАЦІЇ ПРО ОБРАЗ

Навчальні цілі. Після вивчення розділу ви зможете: класифікувати сигнали за характером аргументу і значень; пояснити, як дискретизація та квантування впливають на інформацію про образ; обирати крок дискретизації і кількість рівнів квантування; описувати цифрові зображення в різних колірних моделях; обчислювати основні статистичні характеристики зображення; використовувати поняття сусідства, зв'язності та відстані на растрі.

2.1. Сигнал як фізичний носій образу

Будь-яка система розпізнавання отримує інформацію про об'єкт через сигнал — фізичний процес, параметри якого змінюються відповідно до властивостей об'єкта. Для зору таким процесом є розподіл інтенсивності світла у площині зображення, для слуху — коливання тиску повітря, для радіолокатора — відбита від цілі електромагнітна хвиля. Математичною моделлю сигналу слугує функція одного або кількох аргументів: часу t, просторових координат (x, y) або (x, y, z), довжини хвилі тощо [13, 16].

Сигнали класифікують за кількома ознаками.

За кількістю аргументів розрізняють одновимірні сигнали $s(t)$ (мовний сигнал, електрокардіограма, траєкторія пера під час введення підпису), двовимірні $f(x, y)$ (півтонове зображення), тривимірні (томографічний об'єм, відеопослідовність $f(x, y, t)$) та багатовимірні. Кольорове зображення можна розглядати як три двовимірні сигнали або як векторну функцію $\mathrm{f}(x, y) = [f_R, f_G, f_B]^T$. У теорії реставрації та розпізнавання прийнято однаковою мірою називати одновимірний сигнал і двовимірне зображення *образом* [2, 33]: більшість методів формулюється для одновимірного випадку, а потім поширюється на двовимірний, наприклад, шляхом розгортання зображення в вектор рядок за рядком.

За характером аргументу і значень сигнали поділяють на чотири типи (рис. 2.1):

- аналогові — неперервні за аргументом і за значенням;
- дискретні — визначені лише в окремі моменти часу (точки простору), але з довільними значеннями;
- квантовані — неперервні за аргументом, але з обмеженою множиною допустимих значень;
- цифрові — дискретні і за аргументом, і за значенням; саме такі сигнали обробляються обчислювальними засобами.

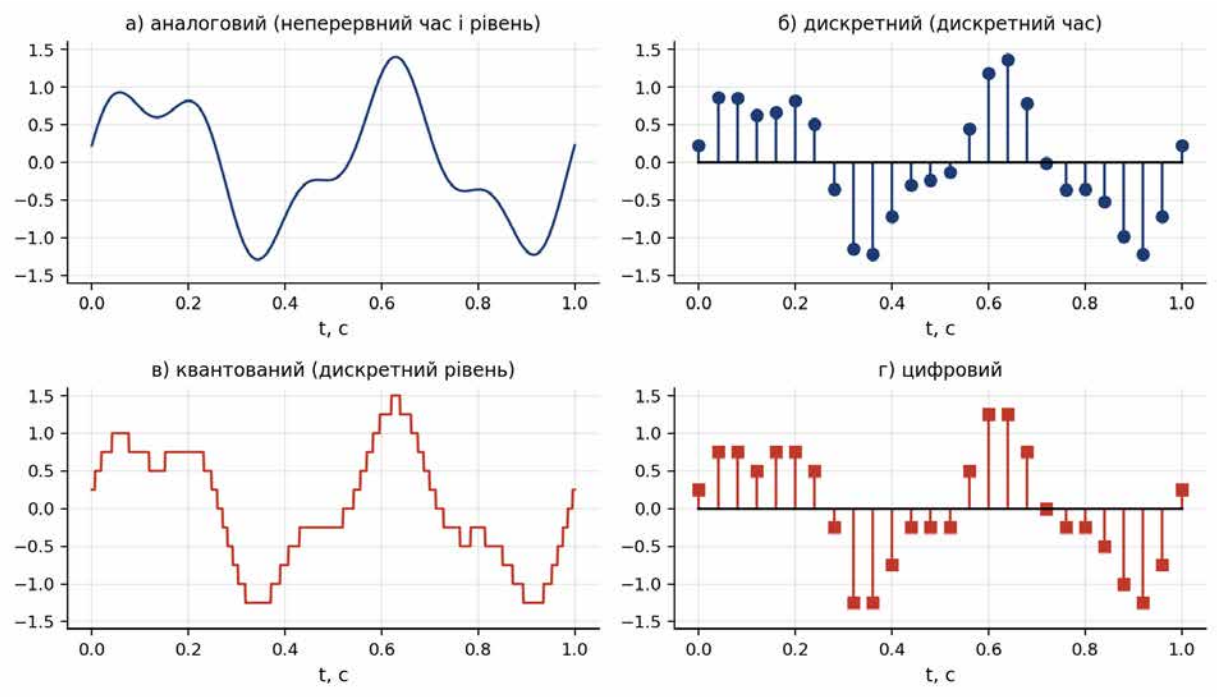


Рис. 2.1. Типи сигналів: аналоговий, дискретний, квантований і цифровий

За передбачуваністю сигнали бувають детермінованими, значення яких у будь-який момент можна точно обчислити, та випадковими (стохастичними), що описуються лише ймовірнісними характеристиками. Реальні сигнали, які надходять до систем розпізнавання, завжди містять випадкову складову — шум, а образи одного класу являють собою різні реалізації деякого випадкового процесу. Саме тому методи розпізнавання значною мірою спираються на статистичний опис [17, 18].

За поведінкою статистичних характеристик у часі розрізняють стаціонарні сигнали, характеристики яких не залежать від положення на осі часу, і нестаціонарні. Для зображень аналогічне поняття — однорідність (гомогенність) поля: статистики однорідного поля не залежать від координат. Строго кажучи, реальні зображення майже ніколи не є однорідними: ділянки фону, контури, текстури мають різні характеристики. Проте гіпотеза однорідності суттєво спрощує математичний апарат і широко використовується в методах фільтрації та реставрації [2, 33]. Методи аналізу нестаціонарних сигналів, які дозволяють досліджувати їх локальні властивості, розглянуто в розділі 6.

Ще одна корисна класифікація пов’язана з поняттям енергії сигналу. Сигнал $s(t)$ називають енергетичним, якщо його енергія скінченна:

$$E = \int_{-\infty}^{\infty} s^2(t)\,dt < \infty, \tag{2.1}$$

і потужностним, якщо скінченною є середня потужність $P = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} s^2(t)\,dt$. Для дискретного сигналу $s(n)$ скінченної довжини N енергія дорівнює сумі квадратів відліків, що в матричній формі записується як скалярний добуток вектора відліків на себе:

$$E = \sum_{n=0}^{N-1} s^2(n) = \mathrm{s}^T\mathrm{s}. \tag{2.2}$$

Цей вираз неодноразово використовуватиметься надалі, зокрема під час опису енергії шуму в методі умовної деконволюції (розділ 8).

2.2. Зображення як двовимірний сигнал

2.2.1. Модель формування зображення

Зображенням називають двовимірну функцію $f(x, y)$, значення якої в точці з просторовими координатами (x, y) пропорційне інтенсивності (яскравості) або іншій фізичній величині, що реєструється системою. Для оптичних зображень значення $f(x, y)$ визначається добутком освітленості сцени $i(x, y)$ та коефіцієнта відбиття поверхні $r(x, y)$:

$$f(x, y) = i(x, y)\, r(x, y), \quad 0 < i(x, y) < \infty, \quad 0 < r(x, y) < 1. \tag{2.3}$$

Освітленість зазвичай змінюється у просторі повільно, тоді як коефіцієнт відбиття містить різкі перепади на межах об’єктів. Ця модель є основою методів компенсації нерівномірного освітлення: логарифмування перетворює добуток (2.3) на суму, після чого повільну складову освітленості можна придушити високочастотною фільтрацією (так звана гомоморфна фільтрація [13]).

Для інших видів зображень значення функції $f(x, y)$ має інший фізичний зміст: у рентгенографії — ослаблення рентгенівського випромінювання вздовж променя, в ультразвуковій діагностиці — інтенсивність відбитого ехосигналу, в радіолокації — ефективна площа розсіювання, у тепловізійних системах — температура. Проте з точки зору оброблення всі вони описуються однаково — як двовимірні масиви чисел.

2.2.2. Цифрове зображення

Для оброблення на ЕОМ неперервне зображення перетворюють у цифрову форму в два етапи: дискретизація (вибірка значень у вузлах прямокутної сітки) та квантування (заміна неперервного діапазону яскравостей скінченною множиною рівнів). Результатом є цифрове зображення — матриця

$$\mathrm{F} = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}, \tag{2.4}$$

елементи якої називають пікселями (від англ. *picture element*). Перший індекс m зазвичай нумерує рядки (вертикальна координата, спрямована донизу), другий n — стовпці (горизонтальна координата). Значення пікселів є цілими числами з діапазону $[0, L-1]$, де $L = 2^b$ — кількість рівнів квантування, b — розрядність. Для більшості півтонових зображень $b = 8$, $L = 256$; для медичних і наукових зображень використовують 12–16 біт.

Зауважимо, що в математичних моделях часто зручно розглядати зображення не як матрицю, а як вектор довжини MN, отриманий розгортанням матриці рядок за рядком (лексикографічне впорядкування):

$$\mathrm{f} = [f(0,0), \dots, f(0, N-1), f(1,0), \dots, f(M-1, N-1)]^T. \quad (2.5)$$

У такому поданні будь-яке лінійне перетворення зображення (фільтрація, розмиття, ортогональне перетворення) записується як множення вектора f на матрицю розміром $MN \times MN$. Це дозволяє використовувати апарат лінійної алгебри, але призводить до матриць величезних розмірів: для зображення 1024×1024 матриця має $10^6 \times 10^6$ елементів [2]. Подолання цієї проблеми за рахунок переходу в область ортогональних перетворень є однією з центральних тем посібника (розділи 4, 5, 8, 9).

2.3. Дискретизація сигналів

2.3.1. Теорема відліків

Дискретизація полягає в заміні неперервного сигналу $s(t)$ послідовністю його значень (відліків) у моменти $t_n = n\Delta t$, де Δt — крок (інтервал) дискретизації; величину $f_д = 1/\Delta t$ називають частотою дискретизації. Природно постає питання: чи можна за відліками точно відновити неперервний сигнал, і якщо так, то яким має бути крок?

Відповідь дає теорема відліків (теорема Котельникова — Шеннона) [16, 43]: сигнал $s(t)$, спектр якого обмежений частотою f_{max} (тобто $S(f) = 0$ при $|f| > f_{max}$), повністю визначається своїми відліками, взятими з кроком

$$\Delta t \le \frac{1}{2f_{max}}, \quad (2.6)$$

і може бути відновлений за формулою

$$s(t) = \sum_{n=-\infty}^{\infty} s\,(n\Delta t)\,\frac{\sin(\pi(t - n\Delta t)/\Delta t)}{\pi(t - n\Delta t)/\Delta t}. \quad (2.7)$$

Мінімальну частоту дискретизації $f_д = 2f_{max}$ називають частотою Найквіста. Ряд (2.7) є розкладом сигналу за зсунутими функціями sinc, які утворюють ортогональний базис у просторі сигналів з обмеженим спектром. Фізично відновлення за формулою (2.7) відповідає пропусканню послідовності відліків через ідеальний фільтр нижніх частот із частотою зрізу $f_д/2$.

2.3.2. Спектр дискретизованого сигналу і накладання спектрів

Щоб зрозуміти, що відбувається при порушенні умови (2.6), розглянемо спектр дискретизованого сигналу. Ідеальну дискретизацію можна подати як множення сигналу на періодичну послідовність дельта-імпульсів $\sum_n \delta\,(t - n\Delta t)$. Спектр такої послідовності також є послідовністю дельта-імпульсів з періодом $f_д$, а множенню в часовій області відповідає згортка спектрів. Тому спектр

дискретизованого сигналу є періодичним повторенням спектра неперервного сигналу:

$$S_д(f) = \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} S\left(f - k f_д\right). \tag{2.8}$$

Якщо $f_д \geq 2f_{max}$, копії спектра не перекриваються, і вихідний спектр можна виділити фільтром. Якщо ж $f_д < 2f_{max}$, сусідні копії накладаються одна на одну (рис. 2.2, б), і складові з частотами $f > f_д/2$ «перекидаються» в смугу $\left[0, f_д/2\right]$, видаючи себе за низькочастотні. Це явище називають накладанням спектрів, або аліасингом (від англ. *alias* — псевдонім). На рис. 2.2, а показано, як гармоніка частотою 9 Гц при дискретизації з частотою 10 Гц дає точно такі самі відліки, як гармоніка 1 Гц: після дискретизації ці два сигнали принципово нерозрізнювані.

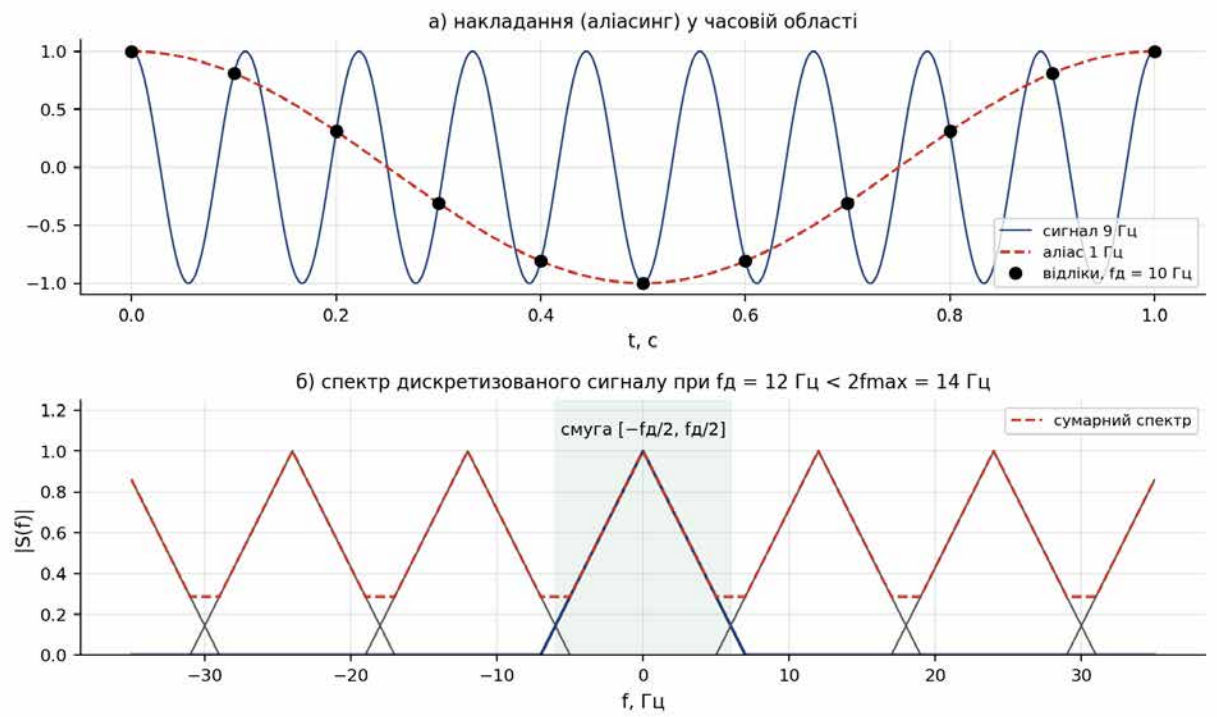


Рис. 2.2. Ефект накладання спектрів при недостатній частоті дискретизації

Аліасинг незворотний: жодна подальша обробка не дозволяє відокремити справжні низькочастотні складові від «псевдонімів». Тому перед дискретизацією сигнал необхідно пропустити через антиаліасинговий фільтр нижніх частот, який придушує складові вище $f_д/2$. У системах реєстрації зображень роль такого фільтра виконують розмиття оптичної системи, скінченні розміри світлочутливих елементів матриці та спеціальні оптичні фільтри.

2.3.3. Дискретизація зображень

Для зображень теорема відліків формулюється окремо для кожної просторової координати. Якщо двовимірний спектр зображення обмежений просторовими частотами u_{max} і v_{max} (вимірюються в періодах на одиницю

довжини), кроки дискретизації мають задовольняти умови $\Delta x \leq 1/(2u_{max})$, $\Delta y \leq 1/(2v_{max})$.

Порушення цих умов призводить до характерних артефактів: на зображеннях з дрібними періодичними структурами (тканини, решітки, черепиця) з'являються хибні низькочастотні візерунки — муар. На рис. 2.3 показано тестове зображення «зонна пластинка», просторова частота якого зростає від центра до країв. При простому проріджуванні (вибір кожного четвертого пікселя) по краях виникають хибні концентричні кільця, яких немає у вихідному зображенні (рис. 2.3, б). Попереднє згладжування гаусівським фільтром перед проріджуванням усуває муар ціною втрати дрібних деталей, які все одно не можуть бути передані з меншою роздільною здатністю (рис. 2.3, в).



Рис. 2.3. Муар при проріджуванні зображення та його усунення попередньою фільтрацією

Кількість пікселів, якою представлено зображення, визначає його просторову роздільну здатність. Її зменшення призводить до втрати дрібних деталей і появи «блочної» структури (рис. 2.4). Для задач розпізнавання важливо, щоб найменші деталі, суттєві для розрізнення класів, займали принаймні 2–3 пікселі. Наприклад, для розпізнавання друкованих символів зазвичай достатньо 20–30 пікселів на висоту символу, тоді як для розпізнавання облич мінімальна прийнятна відстань між зіницями становить близько 30–60 пікселів.



Рис. 2.4. Вплив просторової роздільної здатності

2.3.4. Дискретизація і похибки диференціювання

Дискретизація впливає не лише на можливість відновлення сигналу, а й на точність обчислення його похідних, які широко використовуються для виділення

контурів (розділ 7) та в математичних моделях лінійних систем. Похідна неперервного сигналу в дискретному випадку замінюється скінченною різницею, наприклад, різницею «назад»:

$$s'(n\Delta t) \approx \frac{s(n) - s(n-1)}{\Delta t}. \tag{2.9}$$

Для гармонічного сигналу $e^{j\omega t}$ точна похідна дорівнює $j\omega\, e^{j\omega t}$, тоді як різниця (2.9) дає множник $\left(1 - e^{-j\omega\Delta t}\right)/\Delta t$. При $\omega\Delta t \ll 1$ цей множник наближено дорівнює $j\omega$, проте з наближенням частоти до частоти Найквіста розбіжність зростає. Цей простий факт лежить в основі перетворення RTF [5, 6], у якому замість частот $j\omega_k$ використовуються комплексні числа $\beta_k = 1 - e^{-j2\pi k/N}$, що точно відповідають різницевому оператору. Детально перетворення RTF та його застосування для реставрації образів розглянуто в розділі 9.

2.4. Квантування

2.4.1. Рівномірне квантування і шум квантування

Квантування — це заміна значення сигналу найближчим з дозволених рівнів. При рівномірному квантуванні з кроком q значення s замінюється на

$$s_q = q \cdot \mathrm{round}(s/q). \tag{2.10}$$

Різниця $e = s_q - s$ називається похибкою (шумом) квантування. Якщо крок малий порівняно з діапазоном зміни сигналу, похибку можна вважати випадковою величиною, рівномірно розподіленою на інтервалі $[-q/2, q/2]$, з нульовим середнім і дисперсією

$$\sigma_e^2 = \frac{1}{q}\int_{-q/2}^{q/2} e^2\, de = \frac{q^2}{12}. \tag{2.11}$$

Для b-розрядного квантувача, діапазон якого відповідає повній амплітуді гармонічного сигналу, відношення потужності сигналу до потужності шуму квантування становить

$$\mathrm{SNR}_q \approx 6{,}02\, b + 1{,}76 \text{ дБ}. \tag{2.12}$$

Кожний додатковий розряд покращує відношення сигнал/шум квантування приблизно на 6 дБ [16]. Для 8-розрядного представлення отримуємо близько 50 дБ, що перевищує динамічний діапазон більшості оптичних сенсорів і людського зору при фіксованому рівні адаптації.

2.4.2. Візуальні ефекти квантування зображень

Зменшення кількості рівнів квантування зображення призводить до появи хибних контурів — помітних меж між ділянками з різними рівнями яскравості там, де у вихідному зображенні яскравість змінювалася плавно (рис. 2.5) [13]. Людське око дуже чутливе до таких меж на плавних переходах (небо, обличчя), тоді як на текстурованих ділянках зниження кількості рівнів майже непомітне. Для задач розпізнавання часто достатньо 16–64 рівнів, а для деяких (розпізнавання тексту, аналіз силуетів) — двох (бінарне зображення). Проте для

попереднього оброблення, особливо для реставрації, бажано мати якнайбільшу розрядність: операції фільтрації та деконволюції підсилюють похибки, і шум квантування може стати помітним.

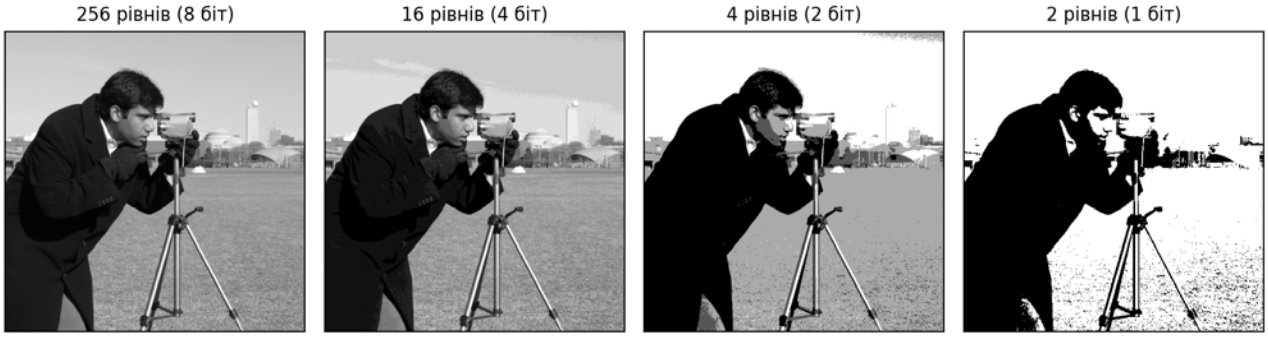


Рис. 2.5. Вплив кількості рівнів квантування на зображення

Крім рівномірного, застосовують нерівномірне квантування, при якому крок менший у тих діапазонах значень, які трапляються частіше або до яких система сприйняття більш чутлива. Приклади — логарифмічне квантування мовних сигналів у телефонії, гамма-кодування яскравості в стандартах відео і кольорового простору sRGB [14, 15].

Кількість даних, необхідних для зберігання цифрового зображення без стиснення, дорівнює $M \cdot N \cdot b$ біт. Наприклад, кольорове зображення 1920×1080 з 8 бітами на кожен із трьох колірних каналів займає приблизно 6,2 Мбайт. Звідси велике значення методів стиснення, які, як буде показано в розділі 5, тісно пов’язані з ортогональними перетвореннями.

2.5. Подання кольорових зображень

2.5.1. Колірна модель RGB

Людське око містить три типи колбочок з максимумами чутливості в довгохвильовій (червоній), середньохвильовій (зеленій) та короткохвильовій (синій) частинах видимого спектра. Тому будь-який колір, що сприймається людиною, можна з достатньою точністю відтворити змішуванням трьох базових кольорів. На цьому ґрунтується адитивна модель RGB, у якій колір пікселя задається трьома компонентами (R, G, B), кожна з яких зазвичай квантується 8 бітами. Кольорове зображення в моделі RGB — це три матриці (канали) однакового розміру (рис. 2.6, верхній ряд) [13, 15].

Модель RGB зручна для пристроїв відображення, але не дуже зручна для аналізу: три її компоненти сильно корельовані (зміна освітленості змінює всі три одночасно), а зміст «колір об’єкта» виражений у ній неявно. Тому для розпізнавання часто переходять до інших колірних моделей.

2.5.2. Моделі HSV і HSI

У моделях HSV (Hue, Saturation, Value) та HSI (Hue, Saturation, Intensity) колір описується величинами, близькими до інтуїтивного сприйняття людини:

- тон H — кут на колірному колі (0° — червоний, 120° — зелений, 240° — синій), що визначає «назву» кольору;

- насиченість S — ступінь «чистоти» кольору, відмінності від сірого;
- яскравість V (або інтенсивність I).

Для моделі HSV з компонентами RGB, нормованими до діапазону $[0,1]$, позначимо $C_{max} = \max(R,G,B)$, $C_{min} = \min(R,G,B)$, $\Delta = C_{max} - C_{min}$. Тоді $V = C_{max}$, $S = \Delta / C_{max}$ (при $C_{max} > 0$), а тон обчислюється залежно від того, яка компонента максимальна:

$$H = 60° \times \begin{cases} \left(\dfrac{G-B}{\Delta}\right) \bmod 6, & C_{max} = R, \\ \dfrac{B-R}{\Delta} + 2, & C_{max} = G, \\ \dfrac{R-G}{\Delta} + 4, & C_{max} = B. \end{cases} \tag{2.13}$$

Головна перевага цих моделей для розпізнавання полягає в тому, що тон майже не залежить від освітленості: тінь на червоному об'єкті змінює V, але мало змінює H. Тому сегментація об'єктів за кольором (наприклад, дорожніх знаків, плодів, шкіри) зазвичай виконується в просторі тону і насиченості (рис. 2.6, нижній ряд) [13].

2.5.3. Моделі YCbCr і Lab

Моделі типу YCbCr відокремлюють яскравісну складову Y від двох кольорорізницевих C_b, C_r. Для 8-розрядних даних за рекомендацією ITU-R BT.601 [15]:

$$Y = 0{,}299\,R + 0{,}587\,G + 0{,}114\,B. \tag{2.14}$$

Вагові коефіцієнти (2.14) відображають різну чутливість ока до основних кольорів: найбільшу — до зеленого, найменшу — до синього. Вираз (2.14) є стандартним способом перетворення кольорового зображення на півтонове. Оскільки зір людини значно менш чутливий до просторових деталей кольору, ніж до деталей яскравості, кольорорізницеві складові можна передавати з меншою роздільною здатністю; на цьому ґрунтуються стандарти стиснення JPEG і MPEG.

Модель CIE Lab побудована так, що евклідова відстань між кольорами приблизно відповідає відмінності, яку сприймає людина (перцептивна рівномірність). Тому вона зручна в задачах, де міра подібності кольорів має узгоджуватися з людським сприйняттям, наприклад, у контролі якості фарбування [14, 15].

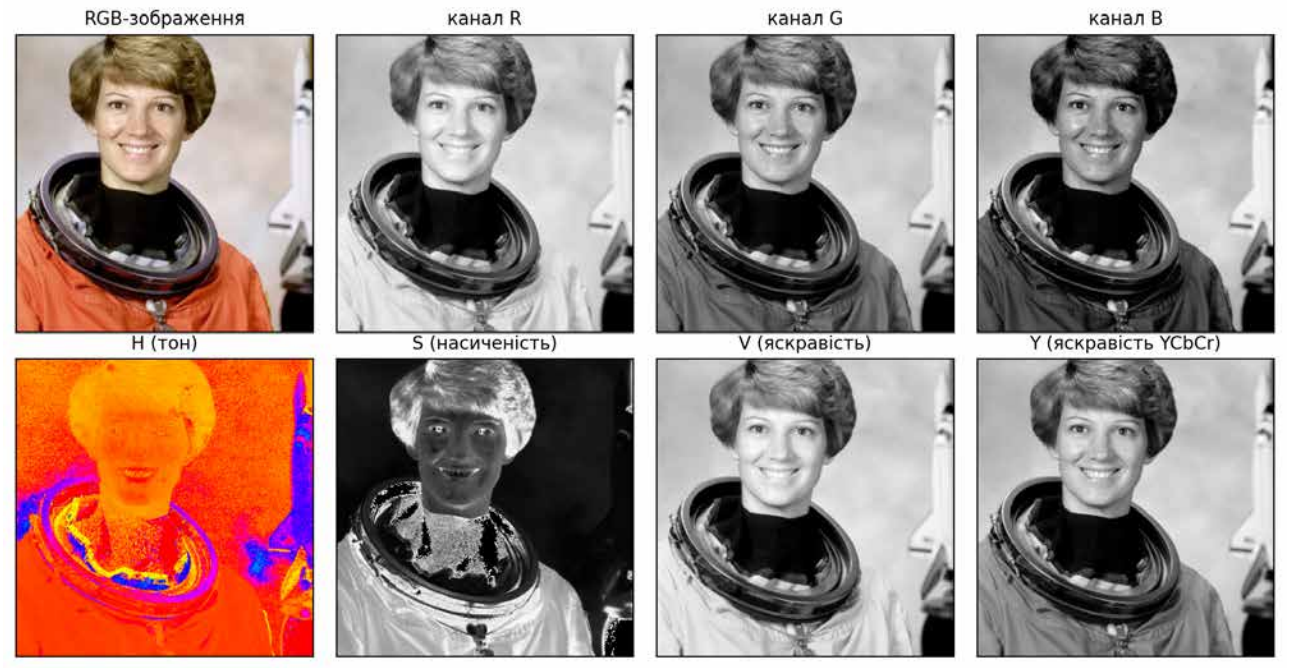


Рис. 2.6. Компоненти кольорового зображення в моделях RGB, HSV та YCbCr

2.5.4. Багатоспектральні зображення

У дистанційному зондуванні, медицині, контролі якості використовують багатоспектральні та гіперспектральні зображення, які містять від кількох до сотень спектральних каналів. Кожний піксель такого зображення є вектором ознак — спектральною сигнатурою матеріалу. Класифікація пікселів за спектральними сигнатурами є класичною задачею статистичного розпізнавання, у якій особливо гостро постають проблеми високої розмірності та кореляції ознак (розділ 11) [17, 35].

2.6. Статистичні характеристики зображень

2.6.1. Гістограма

Гістограмою півтонового зображення називають функцію $h(r_k) = n_k$, де r_k — k-й рівень яскравості, n_k — кількість пікселів з цим рівнем, $k = 0,1, \dots, L-1$. Нормована гістограма

$$p(r_k) = \frac{n_k}{MN} \tag{2.15}$$

є оцінкою ймовірності появи рівня r_k і в сумі дорівнює одиниці. Гістограма не містить інформації про просторове розташування пікселів, проте дає цінне уявлення про загальний характер зображення (рис. 2.7): темне зображення має гістограму, зміщену до малих рівнів; світле — до великих; низькоконтрастне — вузьку гістограму; добре експоноване і контрастне зображення використовує весь діапазон рівнів. Двомодальна гістограма часто свідчить про наявність об’єкта на фоні, що контрастує з ним; на цьому ґрунтуються методи порогової сегментації (розділ 7). Методи перетворення гістограми з метою поліпшення візуальної якості також розглядаються в розділі 7.

2.6.2. Моменти і ентропія

За нормованою гістограмою обчислюють основні статистичні характеристики:

$$\mu = \sum_{k=0}^{L-1} r_k \; p(r_k), \qquad \sigma^2 = \sum_{k=0}^{L-1} (r_k - \mu)^2 \; p(r_k), \tag{2.16}$$

тобто середню яскравість і дисперсію (квадрат середньоквадратичного відхилення, що характеризує контраст). Центральні моменти вищих порядків $\mu_n = \sum_k (r_k - \mu)^n \, p(r_k)$ характеризують асиметрію ($n = 3$) і гостроверхість ($n = 4$) розподілу.

Ентропія гістограми

$$H = -\sum_{k=0}^{L-1} p\,(r_k) \log_2 p(r_k) \tag{2.17}$$

вимірюється в бітах і характеризує середню кількість інформації на піксель за умови, що пікселі незалежні. Ентропія максимальна ($H = \log_2 L$) для рівномірного розподілу яскравостей і мінімальна (нуль) для однорідного зображення. Для зображення, наведеного на рис. 2.7 праворуч, вузька гістограма відповідає меншій ентропії. Ентропія використовується як ознака текстури, як критерій вибору порогу сегментації та як нижня межа для методів стиснення без втрат [13, 14].

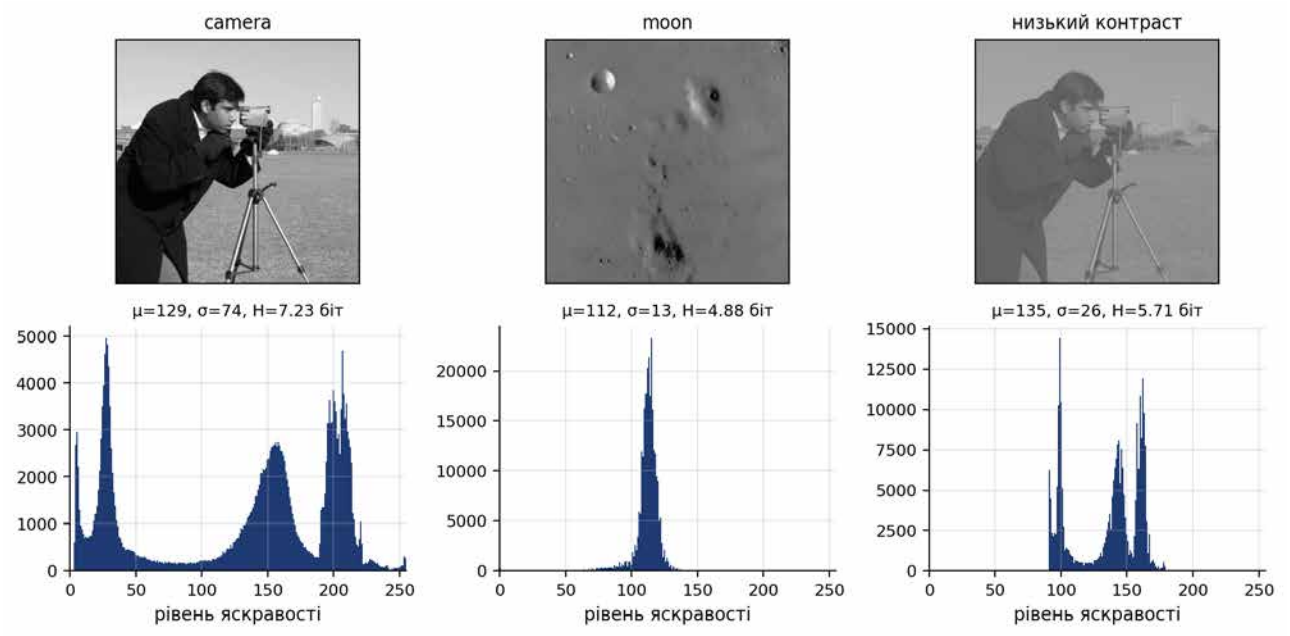


Рис. 2.7. Зображення та їх гістограми із середнім значенням μ, середньоквадратичним відхиленням σ та ентропією H

2.6.3. Кореляційні характеристики

Гістограма описує розподіл значень окремих пікселів, але не їх взаємозв’язок. Сусідні пікселі природних зображень, як правило, сильно корельовані. Кількісно цей зв’язок описується автокореляційною функцією

$$R_{ff}(\Delta m, \Delta n) = \frac{1}{MN}\sum_{m}\sum_{n} f\,(m,n)\, f(m+\Delta m, n+\Delta n), \qquad (2.18)$$

або автоковаріаційною функцією, у якій замість значень f беруться відхилення від середнього. Для однорідних (гомогенних) і ергодичних полів усереднення за ансамблем реалізацій можна замінити усередненням по площі зображення [2], що й відображає формула (2.18). Відповідно до теореми Вінера — Хінчина перетворення Фур'є від автокореляційної функції дає спектральну густину потужності — основну характеристику, що використовується при побудові оптимального фільтра Вінера (розділ 8) [13, 16].

Висока кореляція сусідніх пікселів означає, що зображення в природному поданні є надлишковим: знаючи значення одного пікселя, можна з високою точністю передбачити значення сусіднього. Усунення цієї надлишковості — декореляція — є основною метою ортогональних перетворень, що застосовуються для стиснення і формування ознак (розділи 4, 5, 11).

2.7. Просторові відношення між пікселями

2.7.1. Сусідство і зв'язність

Піксель p з координатами (m,n) має чотирьох горизонтальних і вертикальних сусідів з координатами $(m\pm1,n)$, $(m,n\pm1)$; їх множину називають 4-сусідством $N_4(p)$. Разом із чотирма діагональними сусідами $(m\pm1,n\pm1)$ вони утворюють 8-сусідство $N_8(p)$ (рис. 2.8).

Два пікселі називають 4-зв'язними (8-зв'язними), якщо між ними існує шлях з пікселів, що мають задану властивість (наприклад, належать об'єкту), у якому кожні два послідовні пікселі є 4-сусідами (8-сусідами). Максимальну множину зв'язних пікселів об'єкта називають зв'язною компонентою. Виділення зв'язних компонент (маркування) бінарного зображення є стандартною операцією, що дозволяє розділити сцену на окремі об'єкти і обчислити ознаки кожного з них [13, 15].

Вибір типу зв'язності має важливі наслідки. Діагональна лінія завтовшки в один піксель є зв'язною у 8-сусідстві, але розпадається на окремі точки в 4-сусідстві. З іншого боку, якщо для об'єкта і для фону використати однакову 8-зв'язність, можна отримати парадокс: замкнений діагональний контур не відокремлює внутрішню частину від зовнішньої. Тому зазвичай для об'єкта використовують 8-зв'язність, а для фону — 4-зв'язність, або навпаки.

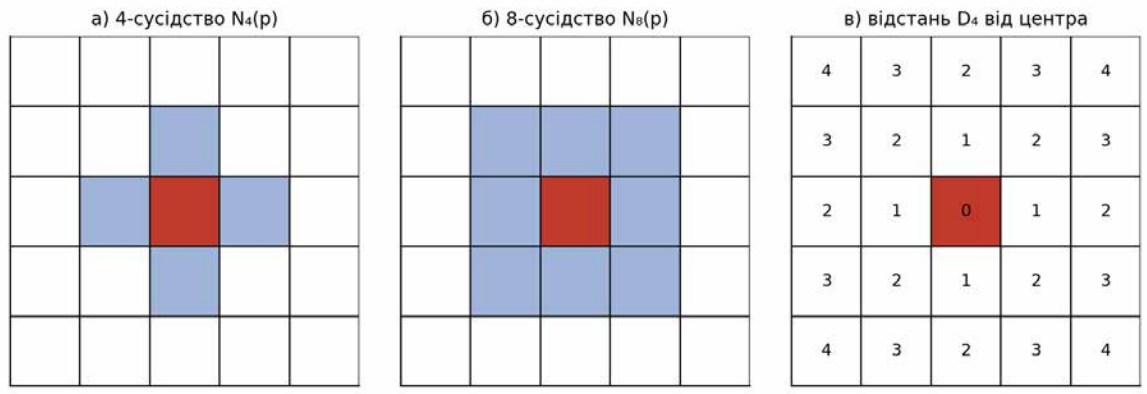


Рис. 2.8. Сусідство пікселів та відстань на растрі

2.7.2. Відстані на растрі

Для пікселів $p = (m_1, n_1)$ і $q = (m_2, n_2)$ використовують такі відстані:

- евклідова: $D_E(p,q) = \sqrt{(m_1 - m_2)^2 + (n_1 - n_2)^2}$;
- «міських кварталів»: $D_4(p,q) = |m_1 - m_2| + |n_1 - n_2|$;
- «шахової дошки»: $D_8(p,q) = \max(|m_1 - m_2|, |n_1 - n_2|)$.

Пікселі, що мають $D_4 = 1$, утворюють 4-сусідство, а пікселі з $D_8 = 1$ — 8-сусідство. Ці відстані є окремими випадками метрик Мінковського (1.5), розглянутих у розділі 1. Вони використовуються в морфологічних операціях, алгоритмах побудови скелета та перетворення відстані (distance transform), що застосовується для обчислення ознак форми [13, 15].

2.8. Особливості реєстрації образів у різних фізичних системах

Хоча після оцифрування всі зображення подаються однаково — матрицями чисел, — фізичний принцип реєстрації визначає характер спотворень, які доведеться враховувати під час попереднього оброблення і реставрації. Розглянемо коротко найважливіші типи систем.

Оптичні системи (фото- і відеокамери, мікроскопи, телескопи) формують зображення об’єктивом на матриці фотоприймачів (ПЗЗ або КМОН). Основні джерела спотворень: дифракція на апертурі і аберації об’єктива (неточкова імпульсна характеристика, що зазвичай залежить від положення в кадрі), розфокусування, рух під час експозиції, фотонний (пуасонівський) шум і тепловий шум електроніки, нелінійність передавальної характеристики (гамма-корекція), нерівномірність чутливості пікселів, інтерполяція кольору в матрицях з кольоровими фільтрами (фільтр Байєра). Для систем дистанційного зондування додаються спотворення, пов’язані з атмосферою (турбулентність, розсіювання) [13, 14].

Рентгенівські системи реєструють ослаблення рентгенівського випромінювання при проходженні крізь об’єкт. Неточковість імпульсної характеристики зумовлена скінченними розмірами фокальної плями рентгенівської трубки, розсіюванням у детекторі (люмінофорі) і мимовільним рухом пацієнта; шум має квантову природу і тому залежить від дози опромінення. Зменшення дози, бажане з медичних міркувань, неминуче підвищує рівень шуму, тому методи реставрації та придушення шуму мають тут

особливе значення [2, 13]. У класичній рентгенівській томографії і скіаграфії розмиття рухом описується саме моделями типу (3.19) [2].

Ультразвукові системи формують зображення за ехосигналами, відбитими від неоднорідностей тканин. Роздільна здатність у напрямку поширення хвилі визначається тривалістю зондувального імпульсу, у поперечному — шириною діаграми спрямованості, яка змінюється з глибиною (анізопланарність). Характерною рисою є спекл-шум — мультиплікативна завада, що виникає внаслідок інтерференції хвиль, розсіяних дрібними неоднорідностями [37].

Радіолокаційні системи, зокрема радіолокатори із синтезованою апертурою, формують зображення за допомогою когерентного оброблення відбитих сигналів. Роздільна здатність визначається шириною спектра зондувального сигналу і довжиною синтезованої апертури, а основним джерелом спотворень є спекл-шум, помилки компенсації траєкторії носія і нелінійності трактів. Погоджена фільтрація (розділ 10) лежить в основі формування таких зображень.

Тепловізійні системи реєструють власне інфрачервоне випромінювання об’єктів. Їм притаманні низька роздільна здатність, нерівномірність чутливості елементів матриці (що потребує калібрування), низький контраст і значний шум.

Отже, вибір моделі деградації (розділ 3) і методів оброблення завжди має спиратися на фізику конкретної системи реєстрації. Методи, що добре працюють для оптичних зображень з адитивним гаусівським шумом, можуть бути малоефективними для ультразвукових чи радіолокаційних зображень зі спекл-шумом, і навпаки.

2.9. Практична робота з цифровими зображеннями в Python

Наведемо приклади базових операцій з цифровими зображеннями: завантаження, перетворення типу даних, обчислення гістограми та статистичних характеристик, перетворення кольорової моделі, зміни роздільної здатності та кількості рівнів квантування [38, 40].

```
import numpy as np
from skimage import data, color, transform, img_as_float

rgb = data.astronaut()                # масив 512x512x3, uint8
gray = color.rgb2gray(rgb)            # Y = 0.2125R + 0.7154G + 0.0721B
print(gray.dtype, gray.min(), gray.max())   # float64 у діапазоні [0, 1]

# гістограма, середнє, СКВ, ентропія
img8 = (gray * 255).astype(np.uint8)
h = np.bincount(img8.ravel(), minlength=256)
p = h / h.sum()
mu = np.sum(np.arange(256) * p)
sigma = np.sqrt(np.sum((np.arange(256) - mu) ** 2 * p))
H = -np.sum(p[p > 0] * np.log2(p[p > 0]))
print(f"mu={mu:.1f}, sigma={sigma:.1f}, H={H:.2f} біт")

# зменшення роздільної здатності з антиаліасинговим фільтром
```

```
small = transform.resize(gray, (128, 128), anti_aliasing=True)

# квантування до L рівнів
L = 8
q = np.floor(gray * L * 0.9999) / (L - 1)

# перехід до моделі HSV
hsv = color.rgb2hsv(rgb)
hue, sat, val = hsv[..., 0], hsv[..., 1], hsv[..., 2]
```

Зверніть увагу на дві поширені помилки. Перша — виконання арифметичних операцій з даними типу uint8, що призводить до переповнення (наприклад, $200 + 100 = 44$ за модулем 256); перед обробкою зображення варто перетворити на тип з плаваючою комою функцією img_as_float. Друга — різні бібліотеки використовують різні вагові коефіцієнти для перетворення на півтонове зображення (стандарти BT.601 і BT.709) [15], тому результати можуть дещо відрізнятися.

2.10. Приклади розв'язання задач

Задача 2.1. Сигнал містить гармоніки з частотами 3 Гц і 7 Гц і дискретизується з частотою 10 Гц. Які частоти будуть спостерігатися в спектрі дискретизованого сигналу в смузі $[0{,}5]$ Гц?

Розв'язання. Частота Найквіста $f_д/2 = 5$ Гц. Гармоніка 3 Гц лежить у смузі і спостерігається без змін. Гармоніка 7 Гц перевищує частоту Найквіста; згідно з (2.8) її копії розташовані на частотах $7 - 10k$ Гц, і в смугу $[-5{,}5]$ потрапляє частота -3 Гц, тобто в амплітудному спектрі вона проявиться на частоті $|7 - 10| = 3$ Гц. Отже, обидві гармоніки дадуть пік на 3 Гц і стануть нерозрізнюваними — типовий приклад незворотного аліасингу.

Задача 2.2. Яке середньоквадратичне відхилення шуму квантування має 8-розрядне зображення з нормованим діапазоном яскравостей $[0{,}1]$? Яке відношення сигнал/шум квантування для гармонічного сигналу повної амплітуди?

Розв'язання. Крок квантування $q = 1/255 \approx 0{,}00392$. За (2.11) $\sigma_e = q/\sqrt{12} \approx 0{,}00113$, тобто близько 0,11 % діапазону. За (2.12) $\mathrm{SNR}_q \approx 6{,}02 \cdot 8 + 1{,}76 \approx 49{,}9$ дБ. Цю оцінку корисно пам'ятати під час реставрації: шум квантування є «нижньою межею» рівня шуму, і, як показано в розділі 8, навіть він робить пряме обернення оператора розмиття неможливим.

Задача 2.3. Зображення має чотири рівні яскравості з імовірностями $0{,}5$; $0{,}25$; $0{,}125$; $0{,}125$. Обчисліть ентропію і порівняйте з максимально можливою.

Розв'язання. За (2.17) $H = -(0{,}5\log_2 0{,}5 + 0{,}25\log_2 0{,}25 + 2 \cdot 0{,}125\log_2 0{,}125) = 0{,}5 + 0{,}5 + 0{,}75 = 1{,}75$ біт. Максимальна ентропія для чотирьох рівнів $\log_2 4 = 2$ біти. Отже, при незалежних пікселях стиснення без втрат теоретично дозволяє зменшити обсяг даних на $12{,}5\,\%$ порівняно з рівномірним 2-бітовим кодуванням.

Висновки до розділу 2

1. Сигнал є фізичним носієм інформації про образ. Сигнали класифікують за кількістю аргументів, характером аргументу і значень (аналогові, дискретні, квантовані, цифрові), передбачуваністю та стаціонарністю. Реальні образи завжди містять випадкову складову.
2. Цифрове зображення є результатом дискретизації та квантування і подається матрицею пікселів або вектором, отриманим розгортанням матриці. Векторне подання дозволяє описувати лінійні перетворення матрицями, проте розміри цих матриць дуже великі.
3. Теорема відліків визначає максимальний крок дискретизації сигналу з обмеженим спектром. Її порушення призводить до незворотного накладання спектрів (аліасингу), що проявляється на зображеннях як муар. Для запобігання цьому перед дискретизацією застосовують фільтрацію нижніх частот.
4. Квантування вносить похибку з дисперсією $q^2/12$; кожний розряд покращує відношення сигнал/шум приблизно на 6 дБ. Недостатня кількість рівнів спричиняє хибні контури.
5. Для аналізу кольорових зображень доцільно використовувати колірні моделі, в яких розділено яскравісну та хроматичні складові (HSV, YCbCr, Lab).
6. Гістограма, моменти, ентропія та автокореляційна функція є основними статистичними характеристиками зображення; вони використовуються і як ознаки, і для обґрунтування методів оброблення.
7. Поняття сусідства, зв’язності та відстаней на растрі лежать в основі виділення окремих об’єктів і обчислення їх геометричних ознак.

Питання для обговорення та самоперевірки

1. Чим відрізняються аналоговий, дискретний, квантований і цифровий сигнали?
2. Що таке стаціонарний сигнал і однорідне (гомогенне) поле зображення? Чому ця гіпотеза важлива для методів оброблення?
3. Як визначається енергія дискретного сигналу? Запишіть її в матричній формі.
4. Поясніть модель формування оптичного зображення як добутку освітленості та коефіцієнта відбиття.
5. Що таке лексикографічне впорядкування пікселів? Яка розмірність матриці лінійного перетворення зображення 256×256?
6. Сформулюйте теорему відліків. Що таке частота Найквіста?
7. Чому спектр дискретизованого сигналу періодичний?
8. У чому полягає ефект накладання спектрів? Чому він незворотний?
9. Як проявляється аліасинг на зображеннях? Як йому запобігти?
10. Чому різницева апроксимація похідної дає похибку, що зростає з частотою?
11. Виведіть формулу для дисперсії шуму рівномірного квантування.

12. Що таке хибні контури і коли вони виникають?
13. Які переваги мають моделі HSV і YCbCr порівняно з RGB для задач розпізнавання?
14. Що характеризує гістограма зображення? Яку інформацію вона не містить?
15. Що таке ентропія зображення? Коли вона максимальна?
16. Як пов’язані автокореляційна функція і спектральна густина потужності?
17. Чим відрізняються 4- і 8-зв’язність? Чому для об’єкта і фону рекомендують використовувати різні типи зв’язності?
18. Які відстані використовують на растрі?

Завдання для самостійної роботи

1. Сигнал $s(t) = \cos(2\pi \cdot 3t) + 0{,}5\cos(2\pi \cdot 7t)$ дискретизують з частотою 10 Гц. Визначте, які частоти будуть присутні в дискретизованому сигналі в смузі $[0{,}5]$ Гц. Перевірте результат чисельно, обчисливши ДПФ.
2. Побудуйте зображення «зонна пластинка» розміром 512×512 і дослідіть муар при проріджуванні в 2, 3, 4 рази з попередньою гаусівською фільтрацією та без неї. Підберіть мінімальне значення параметра σ фільтра, при якому муар стає непомітним.
3. Квантуйте тестове зображення до $L = 2{,}4{,}8{,}16{,}32$ рівнів. Для кожного випадку обчисліть середньоквадратичну похибку і порівняйте її з теоретичним значенням $q^2/12$.
4. Напишіть функцію перетворення RGB → HSV за формулами (2.13) і порівняйте результат з бібліотечною функцією.
5. Обчисліть гістограми і ентропії для п’яти зображень різного характеру (портрет, текст, текстура, медичне зображення, однорідний фон). Поясніть отримані значення.
6. Обчисліть автокореляційну функцію (2.18) рядка зображення і визначте, на якій відстані коефіцієнт кореляції спадає до 0,5.

РОЗДІЛ 3. ЛІНІЙНІ СИСТЕМИ ФОРМУВАННЯ ОБРАЗУ ТА МОДЕЛІ ЙОГО ДЕГРАДАЦІЇ

Навчальні цілі. Після вивчення розділу ви зможете: описувати систему відображення як лінійну систему з імпульсною характеристикою; записувати модель деградації образу в інтегральній, дискретній і матричній формах; розрізняти лінійну та циклічну згортку і відповідні теплицеві та циркулянтні матриці; моделювати типові розмиття і шуми; оцінювати характеристики шуму та імпульсну характеристику за спостережуваними даними; кількісно оцінювати якість образу; пояснювати некоректність оберненої задачі.

3.1. Причини спотворення образів

Жодна реальна система реєстрації не передає образ ідеально. Між об’єктом та цифровим зображенням, яке надходить до системи розпізнавання, лежить ланцюг фізичних перетворень, кожне з яких вносить свої спотворення. Знання природи цих спотворень і вміння описати їх математично є передумовою як для усунення спотворень (реставрації), так і для побудови ознак, стійких до них.

За джерелом та характером спотворення образів можна згрупувати так [2, 13, 33]:

1. Геометричні спотворення — зсув, поворот, зміна масштабу, перспективні та дисторсійні викривлення, нерівномірність растру, зміна координат під час зчитування.
2. Неоднорідність поля образу — нерівномірність освітлення, різна чутливість окремих елементів сенсора, нелінійність передавальної характеристики зчитувальних пристроїв.
3. Локальні спотворення через неточкову імпульсну характеристику — втрата чіткості внаслідок аберацій оптики, розфокусування, дифракції, руху об’єкта чи камери під час експозиції, турбулентності атмосфери, скінченних розмірів джерела випромінювання.
4. Шуми — як корельовані з образом (зернистість фотоматеріалу, структура растру, фотонний шум), так і некорельовані (адитивний тепловий шум електроніки, мультиплікативний, модуляційний).
5. Спотворення через обмежену кількість проєкцій при реконструкції томографічних зображень.
6. Втрата або спотворення контурів реконструйованих об’єктів.

Кожній групі спотворень відповідає своя сукупність методів корекції: геометрична реставрація (глобальна і локальна), точкова корекція (гомогенізація і лінеаризація), узагальнена обернена фільтрація (лінійна — в області натуральних координат або в області ортогональних перетворень, і нелінійна — байєсівська), реконструкція з проєкцій, відновлення контурів (перетворення Хафа, морфологічні методи) [2, 33]. Об’єктом детального розгляду в цьому та наступних розділах є спотворення третьої і четвертої груп, що описуються лінійною моделлю деградації з адитивним шумом. Саме ця модель лежить в основі більшості методів фільтрації та реставрації.

Слід розрізняти два підходи до поліпшення образу. Суб'єктивне поліпшення (англ. *enhancement*) має на меті зробити зображення приємнішим або зручнішим для сприйняття людиною і спирається переважно на досвід та евристики (розділ 7). Об'єктивна реставрація (англ. *restoration*) спирається на математичну модель спотворення та апріорну інформацію про його параметри і має на меті отримати оцінку неспотвореного образу, найближчу до істинної за деяким математичним критерієм [2, 33]. Об'єктивні методи складніші і потребують більше обчислювальних ресурсів, проте дозволяють отримати суттєво кращі результати і, що особливо важливо для систем розпізнавання, є відтворюваними. Зауважимо, що сам об'єктивний критерій близькості щоразу обирається дослідником, тобто суб'єктивно; тому коректне порівняння методів реставрації можливе лише за одним і тим самим критерієм.

3.2. Лінійні системи та імпульсна характеристика

3.2.1. Принцип суперпозиції

Систему відображення можна розглядати як оператор $\mathcal{H}$, який перетворює вхідний (ідеальний) образ f на вихідний $L = \mathcal{H}\{f\}$. Оператор називають лінійним, якщо для будь-яких образів f_1, f_2 і чисел a_1, a_2 виконується принцип суперпозиції [13, 16]:

$$\mathcal{H}\{a_1 f_1 + a_2 f_2\} = a_1\,\mathcal{H}\{f_1\} + a_2\,\mathcal{H}\{f_2\}. \tag{3.1}$$

Будь-який двовимірний образ можна подати як суму (інтеграл) зсунутих і зважених дельта-функцій Дірака:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\,(x',y')\,\delta(x-x',y-y')\,dx'\,dy'. \tag{3.2}$$

Застосувавши до (3.2) лінійний оператор і скориставшись принципом суперпозиції, отримаємо

$$L(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g\,(x,y,x',y')\,f(x',y')\,dx'\,dy', \tag{3.3}$$

де $g(x,y,x',y') = \mathcal{H}\{\delta(x-x',y-y')\}$ — відгук системи на точкове джерело, розташоване в точці (x',y'). Цю функцію називають імпульсною характеристикою системи, а в оптиці — функцією розсіювання точки (ФРТ, англ. *point spread function*, PSF) [13, 14]. Отже, лінійна система повністю описується своєю імпульсною характеристикою: знаючи, як система відображає точку, можна обчислити відображення будь-якого образу.

3.2.2. Ізопланарні та анізопланарні системи

У загальному випадку (3.3) імпульсна характеристика залежить від положення точкового джерела: краї кадру об'єктива зазвичай розмиті сильніше, ніж центр. Такі системи називають анізопланарними (просторово-неінваріантними). Якщо ж відгук на точкове джерело в усіх точках поля зображення має однакову форму і лише зсувається разом із джерелом, тобто

$$g(x,y,x',y') = g(x-x',y-y'), \tag{3.4}$$

систему називають ізопланарною (просторово-інваріантною). Для одновимірних сигналів, що змінюються в часі, аналогічною властивістю є стаціонарність (інваріантність до зсуву в часі), яка для фізичних систем зі сталими параметрами виконується завжди. Для ізопланарної системи вираз (3.3) набуває вигляду двовимірного інтеграла згортки (інтеграла Дюамеля):

$$L(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g\,(x-x', y-y')\, f(x',y')\, dx'\, dy' = g(x,y) * f(x,y), \qquad (3.5)$$

де символ $*$ позначає операцію згортки (конволюції). Для одновимірного випадку

$$L(t) = \int_{-\infty}^{\infty} g\,(t-\tau)\, f(\tau)\, d\tau. \qquad (3.6)$$

Згортка комутативна ($g * f = f * g$), асоціативна і дистрибутивна відносно додавання. Її фізичний зміст наочно ілюструє рис. 3.1: кожний відлік вхідного образу «розмивається» за формою імпульсної характеристики, а вихідний образ є сумою всіх таких розмитих копій. Різкі перепади (межі об'єктів) згладжуються, вузькі деталі розширюються і зменшуються за амплітудою, близько розташовані деталі зливаються. Саме ці ефекти визначають роздільну здатність системи — мінімальну відстань між двома точковими об'єктами, при якій їх ще можна розрізнити [14].

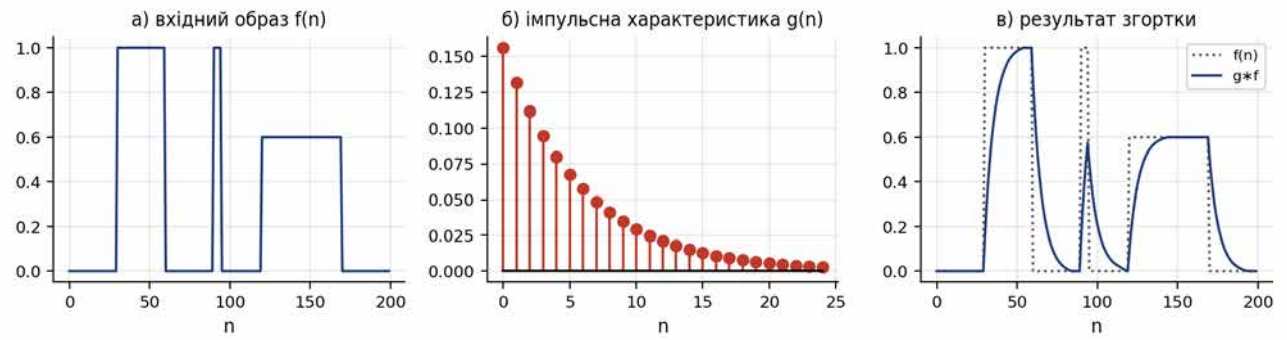


Рис. 3.1. Дискретна згортка одновимірного образу з експоненційною імпульсною характеристикою

Гіпотеза ізопланарності, як і гіпотеза лінійності, на практиці виконується лише наближено. Багато сенсорів мають нелінійну передавальну характеристику (наприклад, фотоплівка має логарифмічну характеристичну криву, а цифрові камери застосовують гамма-корекцію). Тому перед застосуванням лінійних моделей систему слід *лінеаризувати* — застосувати до даних функцію, обернену до передавальної характеристики. Анізопланарну систему часто розглядають як кусково-ізопланарну: поле зображення ділять на фрагменти, у межах кожного з яких імпульсна характеристика вважається сталою [2, 14].

3.2.3. Модель деградації з адитивним шумом

Додавши до (3.3) складову, що описує шум, отримуємо загальну модель деградації образу [2, 13, 33]:

$$L(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g\,(x,y,x',y')\,f(x',y')\,dx'\,dy' + V(x,y), \tag{3.7}$$

де $f(x,y)$ — ідеальний (неспотворений) образ, g — результуюча імпульсна характеристика каналу відображення, $V(x,y)$ — адитивний шум, $L(x,y)$ — спостережуваний (деградований) образ. Для ізопланарної системи

$$L(x,y) = g(x,y) * f(x,y) + V(x,y). \tag{3.8}$$

Модель (3.8) описує процес спотворення ідеального образу. Якщо імпульсна характеристика не є дельта-функцією, то навіть за відсутності шуму спостережуваний образ суттєво відрізняється від ідеального. Задача реставрації полягає в тому, щоб за спостережуваним образом L, відомою (або оціненою) імпульсною характеристикою g та певною інформацією про шум V знайти оцінку $\hat{f}$, яка в обраному сенсі найближча до f. Ця задача розглядатиметься в розділах 8 і 9; тут же зупинимося на тому, як отримати потрібну апріорну інформацію — імпульсну характеристику та характеристики шуму. Варто наголосити, що успіх реставрації значною мірою визначається саме наявністю і точністю апріорної інформації про причини деградації [2].

3.3. Дискретна згортка та її матричне подання

3.3.1. Дискретна згортка

Перейдемо до дискретної форми моделі. Підставимо в (3.6) $t = k\Delta t$, $\tau = n\Delta t$ і замінимо інтеграл сумою, а диференціал — кроком дискретизації:

$$L(k\Delta t) \approx \sum_{n} g\,(k\Delta t - n\Delta t)\,f(n\Delta t)\,\Delta t. \tag{3.9}$$

Нормуючи крок ($\Delta t = 1$), отримаємо дискретну згортку

$$L(k) = \sum_{n} g\,(k-n)\,f(n). \tag{3.10}$$

Якщо імпульсна характеристика має M ненульових відліків $g(0), \ldots, g(M-1)$, а образ — N відліків, то результат лінійної згортки містить $N + M - 1$ відліків: образ «розширюється» на довжину імпульсної характеристики мінус один.

Для двовимірних образів дискретна згортка має вигляд

$$L(m,n) = \sum_{i}\sum_{j} g\,(m-i, n-j)\,f(i,j). \tag{3.11}$$

Двовимірну імпульсну характеристику невеликого розміру (наприклад, 3×3 або 5×5) в обробці зображень називають маскою, ядром або апертурою фільтра. Зауважимо, що операція, яку в більшості бібліотек машинного навчання називають «згорткою» (зокрема, в згорткових нейронних мережах), насправді є взаємною кореляцією — без дзеркального відображення ядра. Для симетричних ядер ця відмінність несуттєва [26].

Якщо двовимірна імпульсна характеристика сепарабельна, тобто $g(m,n) = g_1(m)\,g_2(n)$, то двовимірну згортку можна виконати як послідовність

одновимірних згорток за рядками і стовпцями. Це зменшує кількість операцій на піксель з K^2 до $2K$ для ядра розміру $K \times K$. Сепарабельною є, зокрема, гаусівська імпульсна характеристика.

3.3.2. Теплицева і циркулянтна матриці

Дискретну згортку (3.10) можна записати в матричній формі

$$\mathrm{L} = \mathrm{G}\,\mathrm{f}, \tag{3.12}$$

де f і L — вектори-стовпці відліків вхідного і вихідного образів, а G — дискретний оператор згортки. Структура матриці G залежить від того, як враховуються значення образу за межами інтервалу спостереження.

При лінійній (аперіодичній) згортці вважається, що за межами інтервалу образ дорівнює нулю. Тоді елементи матриці залежать лише від різниці індексів $g_{kn} = g(k-n)$, а сама матриця є теплицевою — сталою вздовж кожної діагоналі (рис. 3.2, а) [2, 14]. Для $N = 6$ і імпульсної характеристики з трьома відліками маємо

$$\mathrm{G}_T = \begin{bmatrix} g_0 & 0 & 0 & 0 & 0 & 0 \\ g_1 & g_0 & 0 & 0 & 0 & 0 \\ g_2 & g_1 & g_0 & 0 & 0 & 0 \\ 0 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & g_2 & g_1 & g_0 & 0 \\ 0 & 0 & 0 & g_2 & g_1 & g_0 \end{bmatrix}. \tag{3.13}$$

При циклічній (круговій) згортці образ вважається періодичним з періодом N, тобто відліки, що «виходять» за правий край, повертаються з лівого. Тоді кожний рядок матриці є циклічним зсувом попереднього на одну позицію праворуч, а матриця називається циркулянтною (рис. 3.2, б):

$$\mathrm{G}_C = \begin{bmatrix} g_0 & 0 & 0 & 0 & g_2 & g_1 \\ g_1 & g_0 & 0 & 0 & 0 & g_2 \\ g_2 & g_1 & g_0 & 0 & 0 & 0 \\ 0 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & g_2 & g_1 & g_0 & 0 \\ 0 & 0 & 0 & g_2 & g_1 & g_0 \end{bmatrix}. \tag{3.14}$$

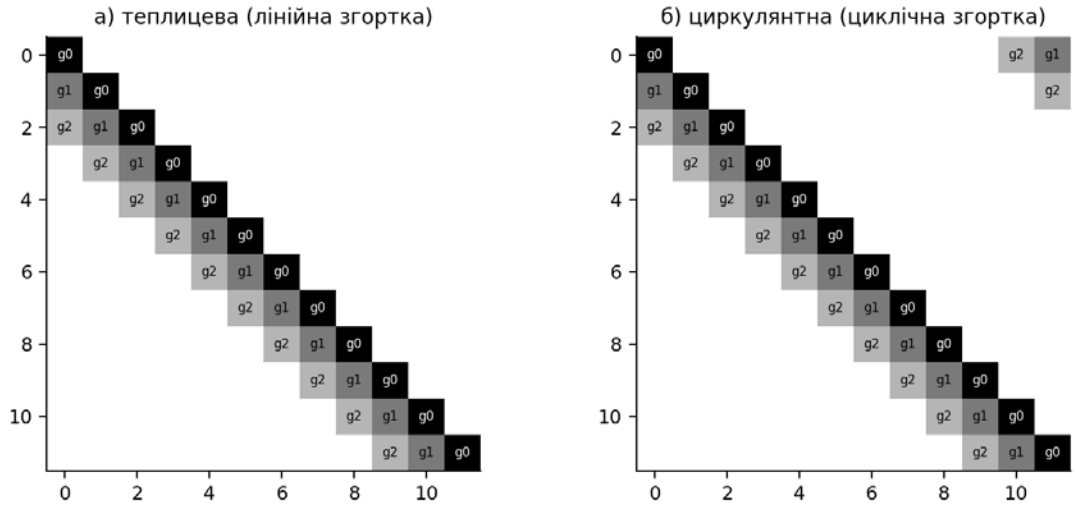


Рис. 3.2. Структура матриць дискретного оператора згортки

Різниця між (3.13) і (3.14) зосереджена лише у верхньому правому куті, проте вона має принципове значення. Як буде показано в розділі 4, *будь-яка циркулянтна матриця діагоналізується дискретним перетворенням Фур'є* — це матрична форма теореми про згортку [2, 16]. Саме ця властивість дозволяє замінити громіздке обернення матриці G поелементним діленням спектрів. Для перетворень з дійсним ядром (наприклад, Адамара) циркулянтна матриця приводиться не до діагонального, а до блочно-діагонального вигляду (розділи 5, 9). Платою за використання циклічної моделі є так звані крайові ефекти: відліки біля правого краю образу впливають на відліки біля лівого. На практиці їх послаблюють доповненням образу нулями, дзеркальним продовженням або плавним «зважуванням» країв.

Розв'язання задачі реставрації в натуральних координатах за відсутності шуму формально зводиться до обернення матриці:

$$\mathrm{f} = \mathrm{G}^{-1}\mathrm{L}. \tag{3.15}$$

Цей приклад демонструє можливості матричного опису дискретизованих моделей. Проте, як показано в підрозділі 3.7, на практиці формула (3.15) майже ніколи не може бути застосована безпосередньо.

3.3.3. Матричне подання двовимірної згортки

Для двовимірного образу розміром $N \times N$, розгорнутого у вектор довжини N^2 згідно з (2.5), двовимірна згортка також записується у формі (3.12), проте матриця G має розмір $N^2 \times N^2$ і блочно-теплицеву (або блочно-циркулянтну) структуру: вона складається з $N \times N$ блоків розміром $N \times N$, кожний з яких є теплицевою (циркулянтною) матрицею, утвореною одним рядком двовимірної імпульсної характеристики, а самі блоки розташовані за теплицевим (циркулянтним) законом. Для зображення 256×256 ця матриця має $65\,536 \times 65\,536 \approx 4{,}3 \cdot 10^9$ елементів; для 1024×1024 — близько 10^{12} елементів. Пряме обернення такої матриці методом Гаусса потребує порядку $(N^2)^3$ операцій, тобто 10^{18} для зображення 1024×1024 [2], що робить розв'язання задачі в

натуральних координатах практично неможливим. Це основна причина переходу до області ортогональних перетворень.

3.4. Моделі імпульсних характеристик і їх ідентифікація

3.4.1. Аналітичні моделі

Найкращим способом визначення імпульсної характеристики є її аналітичне виведення з фізичної моделі системи відображення [2]. Розглянемо найбільш поширені моделі.

Дифракція на апертурі. Для оптичної системи з круглим вхідним отвором діаметром D функція зіниці $P(x, y)$ дорівнює одиниці всередині отвору і нулю поза ним. Імпульсна характеристика некогерентної дифракційно-обмеженої системи визначається автокореляцією функції зіниці в частотній області або, у просторовій області, квадратом модуля її перетворення Фур'є (так званий диск Ейрі). Ширина центрального максимуму пропорційна $\lambda d/D$, де λ — довжина хвилі, d — відстань від апертури до площини зображення. Отже, збільшення апертури зменшує дифракційне розмиття.

Гаусівська модель. У багатьох випадках ізотропну втрату чіткості, спричинену сукупністю незалежних чинників (аберацій, розсіювання в середовищі, скінченних розмірів пікселів), з прийнятною точністю апроксимують гаусівською функцією

$$g(x, y) = k_1 \exp\{-k_r\, r^2\}, \qquad r^2 = x^2 + y^2, \tag{3.16}$$

де константи k_1, k_r знаходять апроксимацією вимірювань; зазвичай записують $k_r = 1/(2\sigma^2)$, де σ — параметр ширини. Гаусівська модель приваблива тим, що її перетворення Фур'є також є гаусівською функцією, вона сепарабельна і не має нулів у частотній області.

Атмосферна турбулентність. Для довгоекспозиційних знімків крізь турбулентну атмосферу використовують модель

$$g(x, y) = k_1 \exp\left\{-(k_2 x^2 + k_3 y^2)^{5/6}\right\}, \tag{3.17}$$

яка для ізотропного поля ($k_2 = k_3$) має «важчі хвости», ніж гаусівська функція [2]. Показник $5/6$ випливає з колмогоровської теорії турбулентності.

Розфокусування. У наближенні геометричної оптики точка, що лежить поза площиною фокусування, відображається рівномірно освітленим диском радіусом R, пропорційним величині розфокусування:

$$g(x, y) = \begin{cases} \dfrac{1}{\pi R^2}, & x^2 + y^2 \le R^2, \\ 0, & \text{інакше.} \end{cases} \tag{3.18}$$

На відміну від гаусівської, ця імпульсна характеристика має нулі у частотній області (на концентричних колах), що, як ми побачимо, ускладнює реставрацію.

Рух під час експозиції. Нехай під час експозиції тривалістю T зображення зсувається у площині сенсора за законом $x_0 = a(t)$, $y_0 = b(t)$. Тоді зареєстрований образ

$$L(x,y) = \int_{-T/2}^{T/2} f\left(x - a(t),\, y - b(t)\right) dt. \tag{3.19}$$

Перейдемо до частотної області. Змінивши порядок інтегрування і виконавши заміну змінних $\xi = x - a(t)$, $\eta = y - b(t)$, отримаємо

$$L(u,v) = F(u,v) \int_{-T/2}^{T/2} e^{-j\left(u\,a(t) + v\,b(t)\right)}\, dt = F(u,v)\, G(u,v), \tag{3.20}$$

тобто спотворення має характер згортки, а частотна характеристика системи $G(u,v)$ виражається через траєкторію руху [2]. Для рівномірного прямолінійного руху на відстань ℓ під кутом φ імпульсна характеристика є відрізком прямої довжиною ℓ, а її частотна характеристика має вигляд функції sinc вздовж напрямку руху — також з нулями. Подібні моделі використовуються, наприклад, у медичній рентгенографії для врахування мимовільних рухів пацієнта.

На рис. 3.3 наведено імпульсні характеристики чотирьох розглянутих типів (верхній ряд) і результати розмиття ними тестового зображення (нижній ряд). Зверніть увагу, що розмиття рухом має чітко виражений напрямок, а розфокусування зберігає «пласкі» ділянки і відносно різкі межі великих об’єктів, але знищує дрібні деталі.

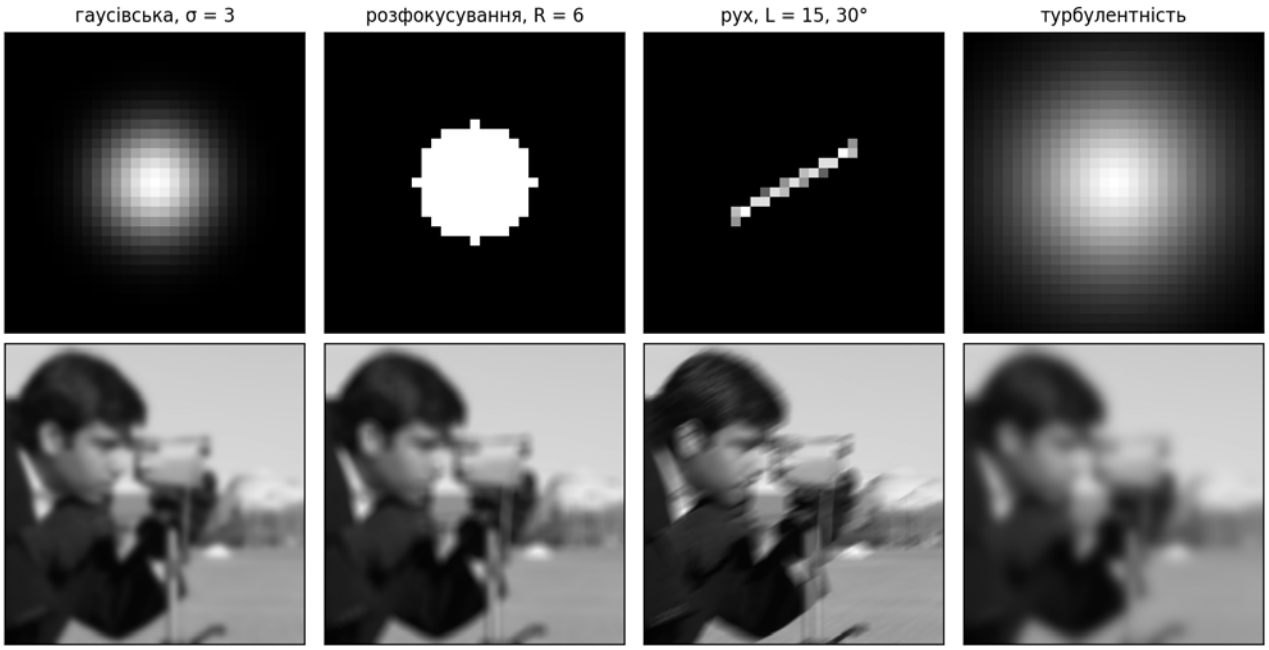


Рис. 3.3. Імпульсні характеристики (вгорі) та відповідні розмиті зображення (внизу)

3.4.2. Експериментальна ідентифікація

У багатьох важливих випадках механізми втрати чіткості настільки складні, що отримати задовільну аналітичну модель не вдається. Тоді імпульсну характеристику доводиться вимірювати [2].

Відображення точкового джерела. Найпряміший спосіб — зареєструвати зображення спеціального тест-об’єкта (фантома) з «точковими» джерелами, розміри яких значно менші за розміри деталей зображення. Іноді точкові джерела

наявні в самій сцені: зірки на астрономічних знімках, «точкові цілі» на ультразвукових або радіолокаційних зображеннях. Тоді імпульсну характеристику можна оцінити апостеріорно за спотвореним образом. Недоліком методу є низьке відношення сигнал/шум: енергія точкового джерела мала.

Метод краю. Значно стійкішим є відображення різкого прямолінійного краю (межі між темною і світлою ділянками). Профіль зображення краю впоперек нього називають функцією розсіювання краю (ESF, *edge spread function*). Похідна від ESF дає функцію розсіювання лінії (LSF) — одновимірний переріз імпульсної характеристики, а модуль її перетворення Фур'є — модуляційну передавальну функцію (МПФ, англ. MTF), яка є стандартною характеристикою якості оптичних систем (рис. 3.4). Оскільки диференціювання підсилює шум, ESF перед диференціюванням зазвичай згладжують або апроксимують параметричною моделлю. Для ізотропних систем однієї LSF достатньо для відновлення двовимірної імпульсної характеристики; для анізотропних вимірювання повторюють для країв різної орієнтації.

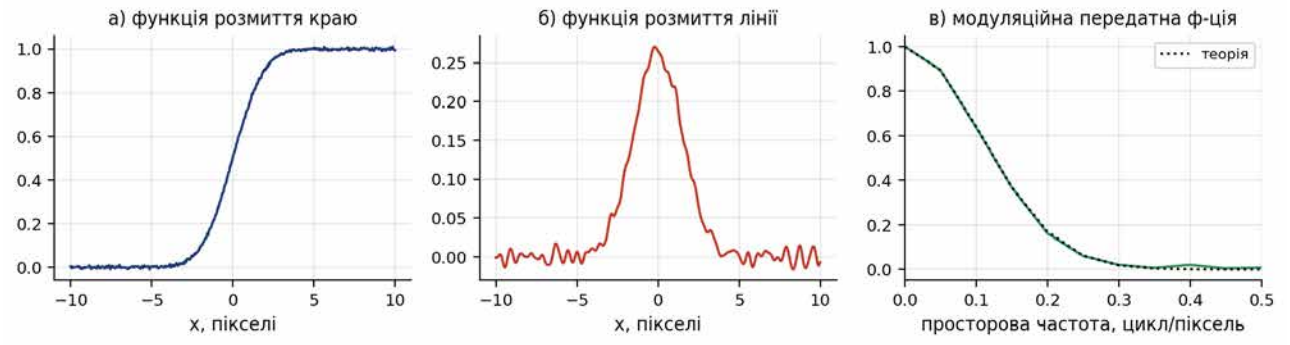


Рис. 3.4. Оцінювання характеристик системи методом краю

Сліпа ідентифікація. Якщо ні тест-об'єкт, ні точкові джерела недоступні, імпульсну характеристику намагаються оцінити безпосередньо за спотвореним зображенням, використовуючи припущення про статистику неспотворених зображень (наприклад, про характерний спад спектра) або положення нулів у частотній області для розмиття рухом і розфокусуванням. Такі методи *сліпої деконволюції* є предметом активних сучасних досліджень, проте виходять за межі посібника.

3.5. Моделі шуму та його ідентифікація

3.5.1. Типи шумів

Шумом називають випадкову складову спостережуваного образу, яка не несе інформації про об'єкт. Найважливішими моделями шуму в системах реєстрації зображень є такі (рис. 3.5).

Адитивний гаусівський шум $V(x, y)$ з нульовим середнім і дисперсією σ^2 не залежить від образу і має нормальний розподіл миттєвих значень. Він виникає головним чином через тепловий шум електронних кіл сенсора і підсилювачів. Якщо значення шуму в різних точках некорельовані, його спектральна густина потужності стала, і такий шум називають білим. Біла гаусівська модель

математично найзручніша, тому на ній ґрунтується більшість класичних методів фільтрації та реставрації [2].

Імпульсний шум («сіль і перець») проявляється як окремі пікселі, значення яких замінено на мінімальне або максимальне (чорні і білі точки). Його причинами є збої передавання, дефектні елементи сенсора, помилки аналого-цифрового перетворення. Лінійні фільтри погано справляються з імпульсним шумом; для нього ефективні нелінійні (медіанні) фільтри (розділ 7) [13].

Пуасонівський (фотонний, дробовий) шум зумовлений квантовою природою світла: кількість фотонів, зареєстрованих елементом сенсора за час експозиції, є випадковою величиною з розподілом Пуассона, дисперсія якого дорівнює середньому. Отже, цей шум залежить від сигналу: у світлих ділянках його абсолютна величина більша, але відносна — менша. Пуасонівський шум домінує при низькій освітленості, в рентгенографії, астрономії, флуоресцентній мікроскопії [13, 37].

Мультиплікативний шум (спекл) описується моделлю $L = f\,(1 + V)$ і характерний для систем з когерентним опроміненням: ультразвукових, радіолокаційних (зокрема радіолокаторів із синтезованою апертурою), лазерних. Логарифмування перетворює мультиплікативний шум на адитивний, що дозволяє застосовувати до нього методи, розроблені для адитивної моделі [13].

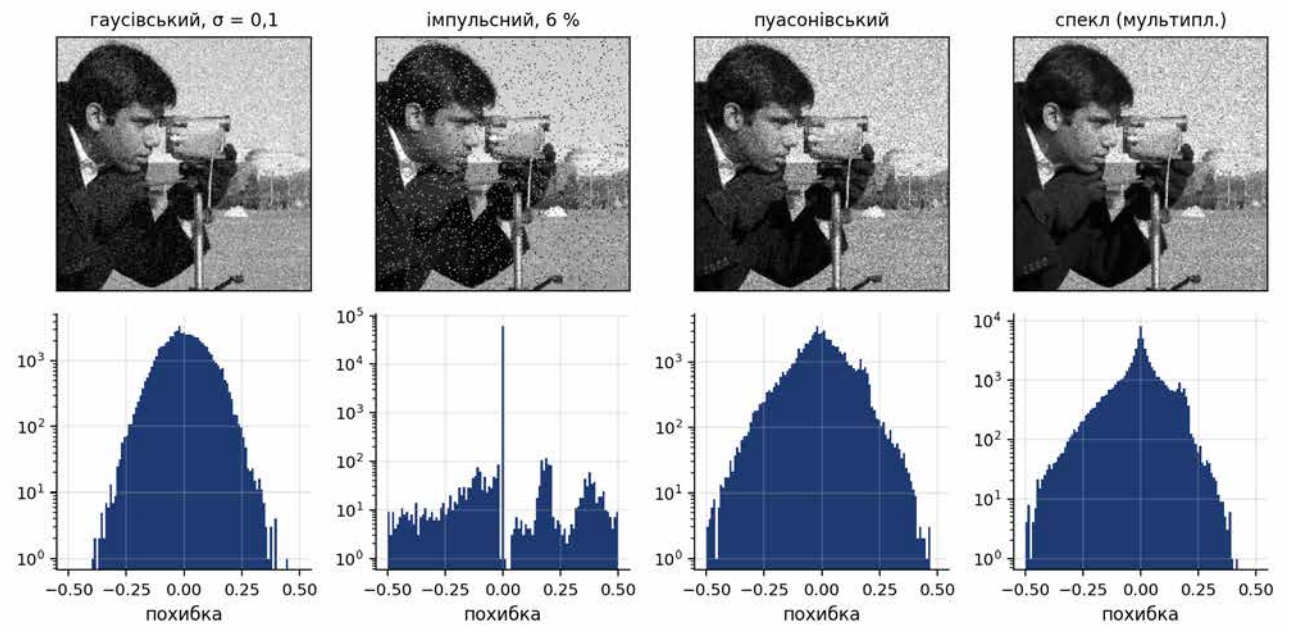


Рис. 3.5. Різні типи шуму (вгорі) та гістограми похибки в логарифмічному масштабі (внизу)

3.5.2. Статистичний опис шуму

Шумову функцію $V(x, y)$ у конкретному образі розглядають як реалізацію двовимірного випадкового поля [2]. Повним описом такого поля є багатовимірна густина ймовірності, розмірність якої дорівнює кількості пікселів. Такий опис надзвичайно громіздкий, і отримати його на практиці неможливо. Тому обмежуються моментами першого і другого порядку, які можна оцінити за наявними даними:

математичним сподіванням

$$\mu(x,y) = \mathrm{M}\{V(x,y)\}, \tag{3.21}$$

дисперсією

$$\sigma_V^2(x,y) = \mathrm{M}\{\big(V(x,y) - \mu(x,y)\big)^2\}, \tag{3.22}$$

та автокореляційною функцією

$$R_{VV}(x_1,y_1,x_2,y_2) = \mathrm{M}\{V(x_1,y_1)\,V(x_2,y_2)\}, \tag{3.23}$$

де $\mathrm{M}\{\cdot\}$ — оператор математичного сподівання (усереднення за ансамблем реалізацій).

Якщо доступна велика кількість реалізацій шуму, незалежних від образу (наприклад, кадри, зареєстровані із закритим об'єктивом), ці характеристики можна обчислити безпосередньо як середні за ансамблем. Аналіз їх залежності від координат дозволяє зробити висновок про однорідність шумового поля. Однак оцінювання за ансамблем є трудомістким, особливо для автокореляційної функції, а в багатьох випадках репрезентативна вибірка реалізацій просто недоступна.

Тому зазвичай приймають певні гіпотези про характер шуму [2]:

- гіпотеза однорідності (гомогенності): середнє μ і дисперсія σ^2 не залежать від координат, а автокореляційна функція залежить лише від зсуву: $R_{VV}(\Delta x, \Delta y)$. Альтернативною характеристикою є спектральна густина потужності $S_{VV} = \mathcal{F}\{R_{VV}(\Delta x, \Delta y)\}$. З однорідності шуму зазвичай випливає і припущення про його незалежність від образу, що, строго кажучи, виконується не завжди (наприклад, не виконується для пуасонівського шуму);
- гіпотеза ергодичності: якщо в розпорядженні є лише одна реалізація (одне спотворене зображення), усереднення за ансамблем замінюють усередненням за просторовими координатами в межах зображення;
- гіпотеза білого гаусівського шуму: спектр шуму рівномірний, а миттєві значення розподілені нормально.

Кожна гіпотеза спрощує методи реставрації, але звужує клас образів, для яких ці методи коректні. Перевірка гіпотез сама по собі є непростою задачею, і її результати слугують підставою для висновку про довіру до результатів реставрації.

3.5.3. Практичні методи оцінювання дисперсії шуму

Часто не існує жодної «чистої» реалізації шуму. Тоді аналізують ділянки спотвореного образу, на яких, на думку дослідника, відсутні деталі об'єкта, тобто шум спостерігається на фоні сталої інтенсивності [2]. Зрозуміло, що таке оцінювання можливе лише за гіпотези ергодичності. Якщо на однорідній ділянці Ω, що містить K пікселів, середня яскравість $\bar{L}_\Omega$, то незміщена оцінка дисперсії

$$\hat{\sigma}^2 = \frac{1}{K-1} \sum_{(x,y)\in\Omega} (L(x,y) - \bar{L}_\Omega)^2. \tag{3.24}$$

Енергію шуму в образі з M відліками (для дискретного поля — сума квадратів відліків шуму) при нульовому середньому можна оцінити як

$$\varepsilon^2 = \mathrm{V}^T\mathrm{V} \approx M\,\sigma^2. \tag{3.25}$$

Саме енергія шуму ε^2 є тією мінімальною апріорною інформацією про шум, якої потребує метод умовної деконволюції (розділ 8). Оскільки формат образу зазвичай великий ($M \sim 10^3$ для одновимірних і 10^5–10^6 для двовимірних образів), вибірка відліків шуму є репрезентативною, і оцінка (3.25) досить точна.

Для автоматичного оцінювання без ручного вибору однорідної ділянки широко застосовують робастну оцінку за вейвлет-коефіцієнтами: у високочастотній діагональній субсмузі вейвлет-розкладу (розділ 6) переважна більшість коефіцієнтів зумовлена шумом, тому

$$\hat{\sigma} = \frac{\mathrm{median}(|d_{HH}|)}{0{,}6745}, \tag{3.26}$$

де d_{HH} — коефіцієнти діагональної деталізуючої субсмуги першого рівня, а константа $0{,}6745$ — медіана модуля стандартної нормальної величини [24]. Медіана, на відміну від середньоквадратичного значення, практично не реагує на поодинокі великі коефіцієнти, зумовлені контурами об’єктів.

3.6. Критерії якості образів

Для порівняння методів оброблення та оцінювання впливу спотворень потрібні кількісні критерії. Якщо доступний еталонний (неспотворений) образ f, якість образу $\hat{f}$ оцінюють повнореференсними критеріями.

Середньоквадратична похибка (MSE)

$$\mathrm{MSE} = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left(f(m,n) - \hat{f}(m,n)\right)^2 \tag{3.27}$$

і пов’язане з нею пікове відношення сигнал/шум (PSNR)

$$\mathrm{PSNR} = 10\log_{10}\frac{f_{max}^2}{\mathrm{MSE}} = 20\log_{10}\frac{f_{max}}{\sqrt{\mathrm{MSE}}}\ \text{дБ}, \tag{3.28}$$

де f_{max} — максимально можливе значення пікселя (255 для 8-розрядних зображень або 1 для нормованих). Типові значення PSNR для зображень прийнятної якості становлять 30–40 дБ [13, 44].

Використовують також відношення сигнал/шум (SNR) як відношення енергії (дисперсії) образу до енергії похибки, та покращення відношення сигнал/шум (ISNR) — різницю між SNR відновленого та спотвореного образів, яке безпосередньо характеризує ефективність методу реставрації.

MSE і PSNR прості для обчислення та аналітичного аналізу, тому саме вони найчастіше використовуються як критерії оптимізації. Проте вони погано узгоджуються з візуальним сприйняттям. На рис. 3.6 наведено чотири спотворення еталонного зображення з *однаковим* значенням PSNR = 24 дБ: зсув яскравості, зниження контрасту, адитивний шум і розмиття. Візуально вони

сприймаються зовсім по-різному: перші два майже не впливають на розпізнаваність деталей, тоді як шум і розмиття значно погіршують її.



Рис. 3.6. Спотворення з однаковим PSNR і різним індексом структурної подібності

Краще узгоджується зі сприйняттям індекс структурної подібності (SSIM) [44], що порівнює локальні середні значення, дисперсії та коваріацію двох зображень у ковзному вікні:

$$\mathrm{SSIM}(\mathrm{x},\mathrm{y}) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}, \tag{3.29}$$

де μ_x, μ_y — локальні середні, σ_x^2, σ_y^2 — локальні дисперсії, σ_{xy} — локальна коваріація, c_1, c_2 — малі сталі, що запобігають діленню на нуль. Добуток у (3.29) поєднує порівняння яскравості, контрасту і структури; індекс набуває значення 1 для ідентичних зображень. Підсумкове значення є середнім за всіма положеннями вікна. Як видно з рис. 3.6, SSIM чітко розрізняє «нешкідливі» та «руйнівні» спотворення з однаковим PSNR.

Якщо еталон недоступний, застосовують безреференсні критерії, що оцінюють, наприклад, різкість (енергію градієнта, ширину перепадів), рівень шуму (3.26) або наявність характерних артефактів (блочності, звону) [44]. Для систем розпізнавання найбільш змістовним «критерієм якості» оброблення є, зрештою, точність розпізнавання на тестовій вибірці.

3.7. Некоректність оберненої задачі

Повернімося до формули (3.15). Здавалося б, якщо імпульсна характеристика відома, а матриця G невироджена, задачу реставрації розв’язано. Насправді ж модель (3.12) описує лише ідеальний випадок. Реальний спостережуваний образ містить шум:

$$\mathrm{L} = \mathrm{G}\,\mathrm{f} + \mathrm{V}, \tag{3.30}$$

і пряме обернення дає

$$\mathrm{G}^{-1}\mathrm{L} = \mathrm{f} + \mathrm{G}^{-1}\mathrm{V}. \tag{3.31}$$

Друга складова (3.31) — шум, «пропущений» через обернену матрицю. Щоб зрозуміти її величину, скористаємося сингулярним розкладом матриці $\mathrm{G} = \mathrm{U}\Sigma\mathrm{W}^T$, де U, W — ортогональні матриці, $\Sigma = \mathrm{diag}(s_1, \ldots, s_N)$, $s_1 \geq s_2 \geq \cdots \geq s_N > 0$ — сингулярні числа. Тоді

$$\mathrm{G}^{-1}\mathrm{V} = \sum_{i=1}^{N} \frac{\mathrm{u}_i^T \mathrm{V}}{s_i} \mathrm{w}_i. \tag{3.32}$$

Для операторів розмиття сингулярні числа швидко спадають (рис. 3.7, а): великим сингулярним числам відповідають «гладкі» сингулярні вектори (низькі просторові частоти), малим — швидко осцилюючі (високі частоти). Проєкції шуму $\mathrm{u}_i^T \mathrm{V}$ мають приблизно однаковий порядок для всіх i (білий шум рівномірно розподілений за всіма компонентами), тому ділення на малі s_i катастрофічно підсилює високочастотні складові шуму [45]. Як видно з рис. 3.7, б, навіть шум із середньоквадратичним відхиленням 10^{-3} (тобто 0,1 % від амплітуди сигналу) робить результат прямого обернення повністю непридатним.

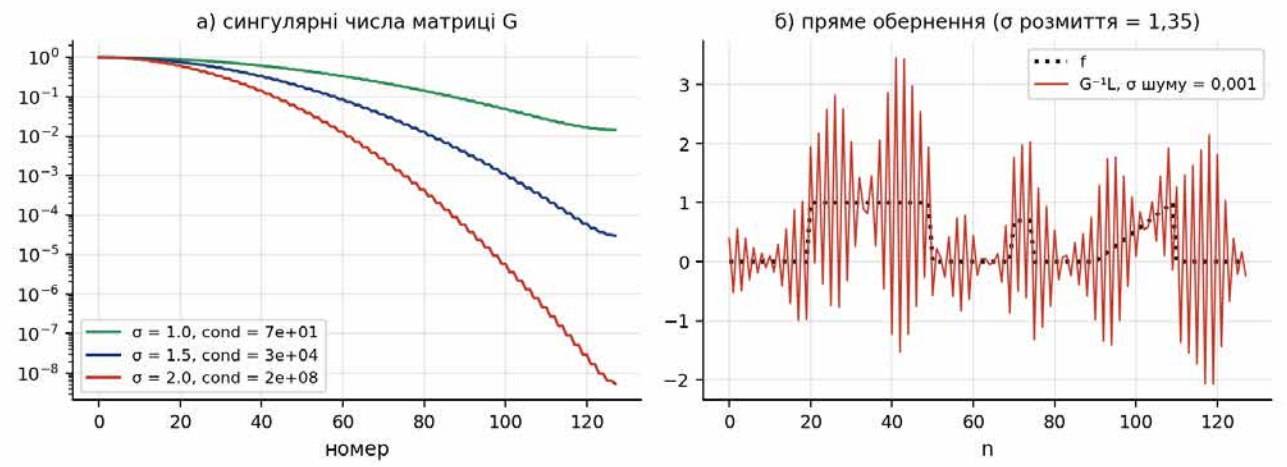


Рис. 3.7. Сингулярні числа оператора гаусівського розмиття та результат прямого обернення зашумленого образу

Кількісною мірою чутливості розв’язку системи лінійних рівнянь до збурень є число обумовленості [45]

$$\mathrm{cond}(\mathrm{G}) = \frac{s_{max}}{s_{min}}, \qquad \frac{\| \delta \mathrm{f} \|}{\| \mathrm{f} \|} \le \mathrm{cond}(\mathrm{G}) \frac{\| \delta \mathrm{L} \|}{\| \mathrm{L} \|}. \tag{3.33}$$

Відносна похибка розв’язку може в $\mathrm{cond}(\mathrm{G})$ разів перевищувати відносну похибку даних. Для гаусівського розмиття з $\sigma = 2$ пікселі число обумовленості вже перевищує 10^8, а для розфокусування і розмиття рухом матриця може бути взагалі виродженою (її частотна характеристика має нулі).

Задачі, розв’язок яких не існує, не єдиний або нестійкий до малих збурень даних, називають некоректно поставленими (за Адамаром). Задача реставрації образів належить саме до цього класу [2, 33, 45]: формально вона може мати нескінченну кількість розв’язків, що узгоджуються з даними в межах похибки, і серед них трапляються як близькі до істинного образу, так і такі, що не мають з ним нічого спільного. Звідси випливає принциповий висновок: розв’язання задачі реставрації неможливе без залучення додаткової апріорної інформації про шукану оцінку — її гладкість, обмеженість енергії шуму, статистичні властивості образів певного класу. Методи, що реалізують цю ідею —

регуляризація Тихонова, фільтрація Вінера, умовна деконволюція, — розглядаються в розділах 8 і 9.

3.8. Моделювання деградації образу в Python

Наведений нижче код моделює деградацію зображення гаусівським розмиттям та адитивним білим шумом, оцінює дисперсію шуму за однорідною ділянкою та за вейвлет-коефіцієнтами, а також обчислює показники якості [40, 41, 44].

```
import numpy as np
from scipy import ndimage as ndi
from skimage import data, img_as_float, metrics
import pywt

f = img_as_float(data.camera())
rng = np.random.default_rng(0)

# імпульсна характеристика гаусівського розмиття
r = 10
y, x = np.mgrid[-r:r+1, -r:r+1]
g = np.exp(-(x**2 + y**2) / (2 * 2.0**2)); g /= g.sum()

sigma = 0.02
L = ndi.convolve(f, g, mode='wrap') + rng.normal(0, sigma, f.shape)

# оцінка СКВ шуму за однорідною ділянкою (фрагмент неба)
roi = L[20:80, 380:480]
print("ROI:", roi.std(ddof=1))
# робастна оцінка за вейвлет-коефіцієнтами (3.26)
_, (_, _, dHH) = pywt.dwt2(L, 'db1')
print("MAD:", np.median(np.abs(dHH)) / 0.6745)
# енергія шуму (3.25)
print("eps^2 =", L.size * sigma**2)

print("PSNR =", metrics.peak_signal_noise_ratio(f, np.clip(L, 0, 1), data_range=1))
print("SSIM =", metrics.structural_similarity(f, np.clip(L, 0, 1), data_range=1))
```

Зверніть увагу: оцінка за ділянкою неба дещо завищена, оскільки на «однорідній» ділянці реального зображення завжди присутні слабкі градієнти яскравості; оцінка (3.26) також може бути трохи завищеною через залишкові деталі в субсмузі HH. Обидві оцінки слід порівняти з істинним значенням $\sigma = 0{,}02$.

3.9. Приклади розв’язання задач

Задача 3.1. Обчисліть лінійну і циклічну згортки образу $f = [0,1,1,1,0,0]$ з імпульсною характеристикою $g = [0{,}5;\ 0{,}3;\ 0{,}2]$.

Розв'язання. Лінійна згортка за (3.10) має $6 + 3 - 1 = 8$ відліків:

$L(0) = 0;\ L(1) = 0{,}5 \cdot 1 = 0{,}5;\ L(2) = 0{,}5 + 0{,}3 = 0{,}8;\ L(3) = 0{,}5 + 0{,}3 + 0{,}2 = 1{,}0;\ L(4) = 0{,}3 + 0{,}2 = 0{,}5;\ L(5) = 0{,}2;\ L(6) = L(7) = 0.$

Циклічна згортка з періодом $N = 6$ (множення на циркулянтну матрицю (3.14)) дає $[0;\ 0{,}5;\ 0{,}8;\ 1{,}0;\ 0{,}5;\ 0{,}2]$. У цьому прикладі результати на інтервалі $0 \ldots 5$ збігаються, оскільки останні відліки образу нульові і «загортання» не відбувається. Якби одиниці стояли наприкінці образу, частина розмитого сигналу з'явилася б на його початку — крайовий ефект циклічної моделі.

Задача 3.2. Середньоквадратична похибка відновленого 8-розрядного зображення дорівнює 25. Обчисліть PSNR.

Розв'язання. За (3.28) $\mathrm{PSNR} = 10\lg(255^2/25) = 10\lg 2601 \approx 34{,}2$ дБ, що відповідає добрій візуальній якості.

Задача 3.3. На однорідній ділянці зображення розміром 20×20 пікселів обчислена вибіркова дисперсія яскравості $\hat{\sigma}^2 = 4{,}0$ (у рівнях квантування). Оцініть енергію шуму для всього зображення 256×256.

Розв'язання. За гіпотезами однорідності та ергодичності дисперсія шуму однакова в усіх точках, тому за (3.25) $\varepsilon^2 \approx M\sigma^2 = 65\,536 \cdot 4{,}0 = 262\,144$. Відносна похибка оцінки дисперсії за $K = 400$ незалежними відліками для гаусівського шуму становить приблизно $\sqrt{2/K} \approx 7\,\%$; якщо на «однорідній» ділянці є слабкий градієнт яскравості, оцінка буде завищеною.

Висновки до розділу 3

1. Спотворення образів поділяють на геометричні, точкові (неоднорідність поля), локальні (неточкова імпульсна характеристика), шумові та інші. Найбільш загальною і широко вживаною моделлю локальних спотворень є лінійна модель деградації з адитивним шумом.
2. Лінійна система повністю характеризується імпульсною характеристикою (функцією розсіювання точки). Для ізопланарних систем модель деградації зводиться до згортки ідеального образу з імпульсною характеристикою.
3. Дискретна згортка записується як добуток матриці на вектор. Лінійній згортці відповідає теплицева матриця, циклічній — циркулянтна. Циркулянтна структура є ключем до ефективних методів реставрації в області ортогональних перетворень. Для двовимірних образів матриці мають розміри $N^2 \times N^2$, що унеможливлює пряме обернення.
4. Імпульсну характеристику визначають аналітично (дифракція, гаусівська модель, турбулентність, розфокусування, рух) або експериментально (точкові джерела, метод краю).
5. Основними моделями шуму є адитивний гаусівський, імпульсний, пуасонівський та мультиплікативний. Для опису шуму використовують середнє, дисперсію, автокореляційну функцію або спектральну густину потужності; на практиці приймають гіпотези однорідності, ергодичності та білого шуму. Мінімально необхідною для умовної деконволюції характеристикою шуму є його енергія.

6. Для оцінювання якості використовують MSE, PSNR, SNR, а також SSIM, що краще узгоджується зі сприйняттям.
7. Задача реставрації є некоректно поставленою: пряме обернення оператора розмиття катастрофічно підсилює шум. Тому її розв'язання вимагає залучення апріорної інформації про шукану оцінку.

Питання для обговорення та самоперевірки

1. Назвіть основні групи спотворень образів та відповідні їм групи методів корекції.
2. Чим об'єктивна реставрація відрізняється від суб'єктивного поліпшення зображення?
3. Сформулюйте принцип суперпозиції. Що таке імпульсна характеристика (функція розсіювання точки)?
4. Яку систему називають ізопланарною? Чому реальні оптичні системи часто анізопланарні?
5. Запишіть модель деградації образу в інтегральній, дискретній і матричній формах.
6. Чим відрізняються лінійна і циклічна згортки? Як це відображається на структурі матриці оператора згортки?
7. Що таке крайові ефекти циклічної згортки і як їх зменшити?
8. Яку структуру має матриця оператора двовимірної згортки? Оцініть кількість операцій для її обернення при розмірі образу 512×512.
9. Запишіть гаусівську модель імпульсної характеристики та модель для атмосферної турбулентності. Чим вони відрізняються?
10. Як отримати частотну характеристику системи для розмиття, спричиненого рухом?
11. Чому імпульсні характеристики розфокусування і рівномірного руху ускладнюють реставрацію?
12. Опишіть метод краю для визначення характеристик системи відображення.
13. Охарактеризуйте основні типи шуму. Який з них залежить від сигналу?
14. Які характеристики використовують для опису випадкового поля шуму? Які гіпотези приймають на практиці?
15. Як оцінити енергію шуму за одним спотвореним образом?
16. Чому медіанна оцінка (3.26) є робастною?
17. Чому PSNR погано узгоджується з візуальним сприйняттям? Які компоненти поєднує SSIM?
18. Що таке число обумовленості? Як воно пов'язане з чутливістю розв'язку до шуму?
19. Поясніть за допомогою сингулярного розкладу, чому пряме обернення оператора розмиття підсилює шум.
20. Яку задачу називають некоректно поставленою? Який загальний підхід до розв'язання таких задач?

Завдання для самостійної роботи

1. Сформуйте теплицеву і циркулянтну матриці згортки для $N = 8$ та імпульсної характеристики $g = [0{,}5;\ 0{,}3;\ 0{,}2]$. Перевірте чисельно, що добуток циркулянтної матриці на вектор збігається з циклічною згорткою, обчисленою через ШПФ.
2. Змоделюйте розмиття тестового зображення чотирма імпульсними характеристиками (3.16)–(3.19). Обчисліть і порівняйте модулі їх частотних характеристик. Знайдіть положення нулів для розфокусування і руху.
3. Додайте до зображення гаусівський шум відомої дисперсії. Оцініть дисперсію трьома способами: за однорідною ділянкою, за вейвлет-коефіцієнтами, за різницею сусідніх пікселів. Порівняйте точність оцінок для різних рівнів шуму.
4. Для одновимірного гаусівського розмиття обчисліть залежність числа обумовленості матриці G від параметра σ при $N = 64{,}128{,}256$.
5. Згенеруйте зображення краю, розмийте його відомою гаусівською імпульсною характеристикою, додайте шум і оцініть σ методом краю. Дослідіть вплив рівня шуму на точність.
6. Для кількох типів спотворень з однаковим PSNR обчисліть SSIM і поясніть отримані результати.

РОЗДІЛ 4. ОРТОГОНАЛЬНІ ПЕРЕТВОРЕННЯ. ДИСКРЕТНЕ ПЕРЕТВОРЕННЯ ФУР'Є

Навчальні цілі. Після вивчення розділу ви зможете: пояснити зміст ортогонального перетворення як розкладу образу за базисом; записувати пряме й обернене перетворення в матричній формі та перевіряти їх ортогональність і повноту; обчислювати і тлумачити дискретне перетворення Фур'є одновимірних і двовимірних образів; застосовувати теорему про згортку для діагоналізації циркулянтної матриці; оцінювати вплив розтікання спектра та ефекту Гіббса; виконувати фільтрацію, кореляційне суміщення та формування спектральних ознак у частотній області.

4.1. Образ як вектор і його розклад за базисом

4.1.1. Основна ідея ортогональних перетворень

У розділі 2 було показано, що дискретний образ з N відліками зручно розглядати як вектор $\mathrm{x} \in \mathbb{R}^N$ (або $\mathbb{C}^N$). Звичайне подання образу — значеннями відліків у точках $n = 0,1,\dots,N-1$ — називають поданням у натуральних координатах. Формально воно відповідає розкладу вектора за стандартним базисом з одиничних векторів e_n, кожний з яких «вмикає» лише один відлік:

$$\mathrm{x} = \sum_{n=0}^{N-1} x\,(n)\,\mathrm{e}_n. \tag{4.1}$$

Цей базис зручний для відображення, але далеко не завжди зручний для аналізу. Відліки природних образів сильно корельовані, енергія розподілена між ними більш-менш рівномірно, а важливі властивості образу (періодичність, гладкість, наявність різких перепадів) не виражені явно. Ідея ортогональних перетворень полягає в тому, щоб подати той самий вектор у вигляді розкладу за *іншим* набором базисних векторів φ_k, підібраних так, щоб коефіцієнти розкладу краще відображали суттєві для задачі властивості:

$$\mathrm{x} = \sum_{k=0}^{N-1} X\,(k)\,\varphi_k. \tag{4.2}$$

Коефіцієнти $X(k)$ називають трансформантами (спектральними коефіцієнтами, амплітудами базисних функцій), а сукупність $\{X(k)\}$ — спектром образу в обраному базисі. Ці терміни традиційні для вітчизняної радіотехнічної школи [1, 2, 22]. Самі базисні вектори φ_k (відліки базисних функцій) також часто називають трансформантами; зміст терміна зазвичай зрозумілий з контексту.

4.1.2. Ортогональність, нормованість і повнота

Систему векторів $\{\varphi_k\}$ називають ортогональною, якщо скалярний добуток будь-яких двох різних векторів дорівнює нулю:

$$\langle \varphi_k, \varphi_l \rangle = \sum_{n=0}^{N-1} \varphi_k(n)\, \varphi_l^*(n) = \begin{cases} E_k, & k = l, \\ 0, & k \neq l, \end{cases} \tag{4.3}$$

де E_k — «енергія» k-ї базисної функції (у неперервному випадку — енергія за період). Якщо всі $E_k = 1$, систему називають ортонормованою. Для неперервних періодичних базисних функцій $w(k,t)$ з періодом T умова ортогональності має аналогічний вигляд [1]:

$$\int_0^T w(k,t)\, w^*(l,t)\, dt = \begin{cases} E_k, & k = l, \\ 0, & k \neq l. \end{cases} \tag{4.4}$$

Ортогональність дозволяє легко знаходити коефіцієнти розкладу. Помноживши (4.2) скалярно на φ_l, отримаємо з урахуванням (4.3)

$$X(k) = \frac{1}{E_k} \langle \mathrm{x}, \varphi_k \rangle = \frac{1}{E_k} \sum_{n=0}^{N-1} x(n)\, \varphi_k^*(n). \tag{4.5}$$

Отже, кожна трансформанта — це міра «схожості» образу з відповідною базисною функцією, нормована на енергію цієї функції. Це спостереження буде ключовим для методів розпізнавання за еталоном (розділ 10): якщо одна з базисних функцій збігається за формою з еталоном, то її трансформанта безпосередньо вимірює ступінь присутності еталона в образі.

Повнотою перетворення називають його здатність забезпечити взаємно однозначну відповідність між образом у натуральних координатах і його спектром [1]. Для дискретних перетворень формату N повнота означає, що N ортогональних векторів утворюють базис простору, тобто для оператора прямого перетворення існує обернений оператор.

4.1.3. Матрична форма перетворення

Запишемо рядками матриці W розміром $N \times N$ спряжені відліки базисних функцій. Тоді пряме перетворення (4.5) для ортонормованого базису набуває вигляду

$$\mathrm{X} = \mathrm{W}\,\mathrm{x}, \tag{4.6}$$

а обернене (4.2) —

$$\mathrm{x} = \mathrm{W}^{-1}\,\mathrm{X}. \tag{4.7}$$

Матрицю W називають матричним оператором перетворення. Для ортонормованого базису обернення матриці зводиться до транспонування зі спряженням:

$$\mathrm{W}^{-1} = \mathrm{W}^{*T} = \mathrm{W}^H, \qquad \mathrm{W}\,\mathrm{W}^H = \mathrm{E}, \tag{4.8}$$

де E — одинична матриця. Такі матриці називають унітарними (для дійсних — ортогональними). Якщо базис ортогональний, але не нормований, то $\mathrm{W}\mathrm{W}^H = \mathrm{diag}\{E_0, \ldots, E_{N-1}\}$, і для отримання оберненого оператора кожний рядок W слід поділити на $\sqrt{E_k}$ — процедура нормування перетворення [1]. Для деяких перетворень (наприклад, для дискретного косинусного з несиметричним матричним оператором) енергії E_k різні і залежать від формату N, тому

нормування є обов'язковим кроком побудови пари прямого й оберненого перетворень.

Найважливішою властивістю унітарного перетворення є збереження енергії (рівність Парсеваля):

$$\sum_{n=0}^{N-1}|x(n)|^2 = \mathrm{x}^H\mathrm{x} = \mathrm{X}^H\mathrm{W}\mathrm{W}^H\mathrm{X} = \sum_{k=0}^{N-1}|X(k)|^2. \qquad (4.9)$$

Геометрично унітарне перетворення — це поворот (можливо, разом із дзеркальним відображенням) системи координат у N-вимірному просторі: довжини векторів і кути між ними не змінюються. Зокрема, евклідова відстань між двома образами однакова в натуральних координатах і в будь-якому ортонормованому базисі. Отже, *перехід до іншого ортонормованого базису сам по собі не додає і не втрачає інформації*; його користь полягає в тому, що в новому базисі енергія образу може бути зосереджена в невеликій кількості трансформант, а деякі операції (згортка, диференціювання) набувають простішого вигляду.

4.1.4. Перетворення матриць операторів

Якщо в натуральних координатах деяке лінійне перетворення образу описується матрицею A ($\mathrm{y} = \mathrm{Ax}$), то в області трансформант той самий оператор описується матрицею

$$\mathrm{A}_W = \mathrm{W}\,\mathrm{A}\,\mathrm{W}^H, \qquad (4.10)$$

оскільки $\mathrm{Y} = \mathrm{Wy} = \mathrm{WAW}^H\mathrm{Wx} = \mathrm{A}_W\mathrm{X}$. Перетворення (4.10) є перетворенням подібності: воно не змінює власних чисел матриці, але може суттєво змінити її *структуру*. Головна мета вибору базису в задачах фільтрації і реставрації — знайти такий базис, у якому матриця оператора деградації стає діагональною або хоча б розрідженою (блочно-діагональною). Для дискретного перетворення Фур'є і циркулянтних матриць це досягається точно (підрозділ 4.4), для перетворення Адамара — з точністю до блочно-діагональної структури (розділи 5 і 9).

4.2. Неперервне перетворення Фур'є

Нагадаємо основні відомості про неперервне перетворення Фур'є, яке є історичною та концептуальною основою спектральних методів [16]. Пара перетворень Фур'є для сигналу $s(t)$ має вигляд

$$S(j\omega) = \int_{-\infty}^{\infty} s\,(t)\,e^{-j\omega t}\,dt, \qquad s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S\,(j\omega)\,e^{j\omega t}\,d\omega. \qquad (4.11)$$

Базисними функціями тут є комплексні гармоніки $e^{j\omega t}$, які є власними функціями будь-якої лінійної стаціонарної системи: на гармонічну дію частоти ω така система реагує гармонікою тієї самої частоти, змінюючи лише амплітуду і фазу. З цього випливають дві теореми, які роблять перетворення Фур'є основним інструментом аналізу лінійних систем [13, 16].

Теорема про згортку оригіналів:

$$g(t) * f(t) \;\leftrightarrow\; G(j\omega)\,F(j\omega), \tag{4.12}$$

тобто згортці в часовій (просторовій) області відповідає добуток спектрів.

Теорема про диференціювання оригіналу:

$$\frac{ds(t)}{dt} \;\leftrightarrow\; j\omega\,S(j\omega), \tag{4.13}$$

тобто диференціюванню відповідає множення спектра на $j\omega$. Завдяки цьому лінійні диференціальні рівняння, що описують системи, перетворюються на алгебраїчні.

Для задачі реставрації в моделі без шуму $L(t) = g(t) * f(t)$ теорема (4.12) дає формально простий розв'язок

$$F(j\omega) = \frac{L(j\omega)}{G(j\omega)}, \qquad f(t) = \mathcal{F}^{-1}\{F(j\omega)\}, \tag{4.14}$$

який набагато простіший, ніж розв'язання тієї самої задачі в натуральних координатах [2]. Проте обчислення інтегралів (4.11) навіть для простих аналітичних виразів пов'язане зі значними складнощами, а для експериментальних даних можливе лише наближено — через дискретизацію. Тому на практиці використовують дискретне перетворення Фур'є.

Інші важливі властивості перетворення Фур'є:

- *зсув:* $s(t - t_0) \leftrightarrow S(j\omega)\,e^{-j\omega t_0}$ — зсув змінює лише фазовий спектр, амплітудний залишається незмінним;
- *зміна масштабу:* $s(at) \leftrightarrow \frac{1}{|a|}S(j\omega/a)$ — стиснення сигналу в часі розширює його спектр, і навпаки;
- *теорема про диференціювання спектра:* $-jt\,s(t) \leftrightarrow dS(j\omega)/d\omega$;
- *рівність Парсеваля:* $\int |s(t)|^2 dt = \frac{1}{2\pi}\int |S(j\omega)|^2 d\omega$.

Властивість зміни масштабу пов'язана з принципом невизначеності: добуток ефективної тривалості сигналу на ефективну ширину його спектра не може бути меншим за деяку сталу. Цей принцип визначає межі частотно-часового аналізу і детально розглядається в розділі 6.

4.3. Дискретне перетворення Фур'є

4.3.1. Визначення

Дискретним перетворенням Фур'є (ДПФ) послідовності $x(n)$, $n = 0,1,\dots,N-1$, називають послідовність

$$X(k) = \sum_{n=0}^{N-1} x\,(n)\,e^{-j\frac{2\pi}{N}kn}, \qquad k = 0,1,\dots,N-1. \tag{4.15}$$

Обернене ДПФ має вигляд

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X\,(k)\,e^{j\frac{2\pi}{N}kn}. \tag{4.16}$$

Базисними функціями ДПФ є дискретні комплексні гармоніки $e^{j2\pi kn/N}$, які на інтервалі з N відліків вміщують рівно k періодів. На рис. 4.1 показано дійсні (косинусні) та уявні (синусні) частини базисних функцій для $N = 8$. Зверніть увагу на симетрію: функції з номерами k і $N - k$ мають однакові дійсні частини і протилежні за знаком уявні; функція з номером $k > N/2$ відповідає «від'ємній» частоті $k - N$.

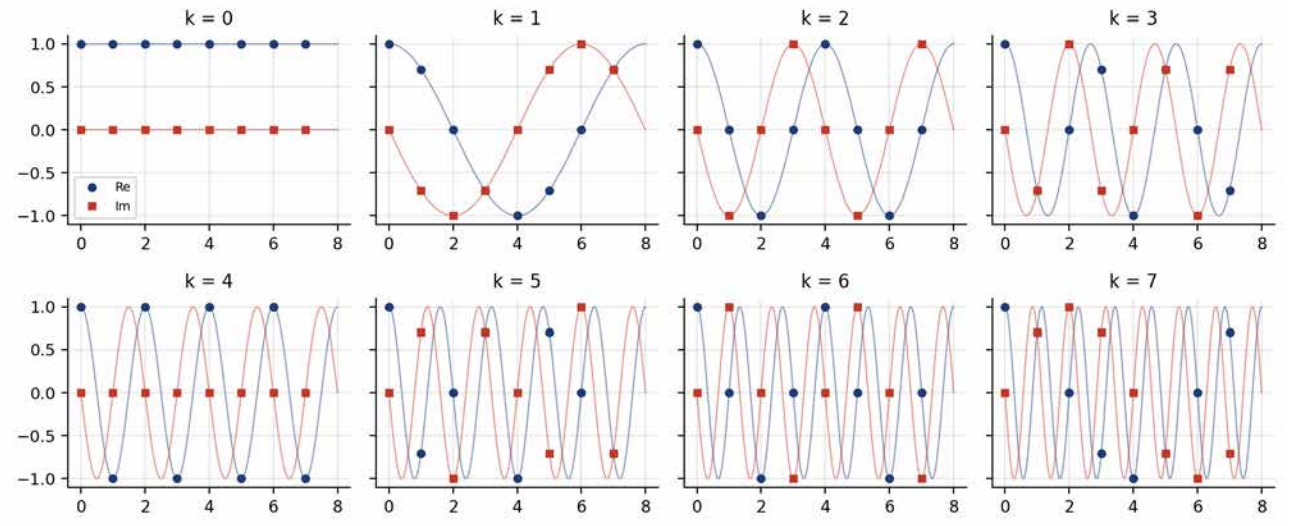


Рис. 4.1. Базисні функції ДПФ при N = 8 (кружки — дійсна частина, квадрати — уявна)

Позначивши $w = e^{-j2\pi/N}$ (корінь N-го степеня з одиниці), можна записати матричний оператор ДПФ:

$$\mathrm{F} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{N-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{N-1} & w^{2(N-1)} & \cdots & w^{(N-1)^2} \end{bmatrix}, \qquad F_{kn} = w^{kn}. \tag{4.17}$$

Матриця (4.17) симетрична, її рядки ортогональні з енергією $E_k = N$. Тому нормований оператор $\mathrm{F_н} = \mathrm{F}/\sqrt{N}$ є унітарним: $\mathrm{F_н^{-1}} = \mathrm{F_н^*}$. Залежно від джерела, множник $1/N$ відносять до прямого, оберненого або симетрично розподіляють між ними ($1/\sqrt{N}$); на результати аналізу це не впливає, але при порівнянні формул з різних джерел слід уважно стежити за нормуванням. Зокрема, у працях з умовної деконволюції [2, 3] у формулах для двовимірних спектрів з'являються множники M^2, зумовлені саме вибраним способом нормування ДПФ формату $M \times M$.

4.3.2. Фізичний зміст номера трансформанти

Якщо сигнал дискретизовано з частотою $f_д$, то k-та трансформанта ДПФ відповідає частоті

$$f_k = k\,\frac{f_д}{N} = \frac{k}{N\Delta t}, \qquad k = 0,1,\ldots,N/2, \tag{4.18}$$

а трансформанти з номерами $k > N/2$ — від'ємним частотам $f_k - f_д$. Крок за частотою $\Delta f = f_д/N = 1/T$ (де $T = N\Delta t$ — тривалість інтервалу аналізу) називають роздільною здатністю ДПФ за частотою. Отже, для розрізнення близьких частотних складових необхідно збільшувати тривалість спостереження, а не частоту дискретизації. Найбільша частота, що може бути представлена, дорівнює $f_д/2$ (частота Найквіста), у повній відповідності з теоремою відліків.

Нульова трансформанта $X(0) = \sum_n x\,(n)$ пропорційна середньому значенню сигналу (постійній складовій).

4.3.3. Основні властивості ДПФ

1. Лінійність: $a\,x_1(n) + b\,x_2(n) \leftrightarrow a\,X_1(k) + b\,X_2(k)$.
2. Періодичність: $X(k+N) = X(k), x(n+N) = x(n)$. ДПФ неявно розглядає послідовність як один період періодичного сигналу. Цей факт має принципове значення: усі операції в області ДПФ (зсув, згортка) є *циклічними*.
3. Спряжена симетрія для дійсних сигналів: $X(N-k) = X^*(k)$. Отже, для дійсного сигналу незалежними є лише $N/2+1$ трансформант, а амплітудний спектр $|X(k)|$ симетричний відносно $N/2$.
4. Циклічний зсув: $x\big((n-m) \bmod N\big) \leftrightarrow X(k)\,w^{km}$ — амплітудний спектр інваріантний до циклічного зсуву.
5. Модуляція: $x(n)\,w^{-ln} \leftrightarrow X\big((k-l) \bmod N\big)$.
6. Рівність Парсеваля: $\sum_n |x(n)|^2 = \frac{1}{N}\sum_k |X(k)|^2$.
7. Теорема про циклічну згортку — розглядається нижче.

На рис. 4.2, а, б наведено приклад: сигнал, що складається з двох гармонік з частотами 8 і 20 Гц і шуму, та його амплітудний спектр. Гармоніки, які в часовій області важко розрізнити на фоні шуму, у спектрі проявляються як чіткі піки, а енергія білого шуму рівномірно розподілена по всіх трансформантах. Саме ця здатність концентрувати енергію регулярних складових у небагатьох коефіцієнтах робить спектральне подання цінним для виявлення та розпізнавання.

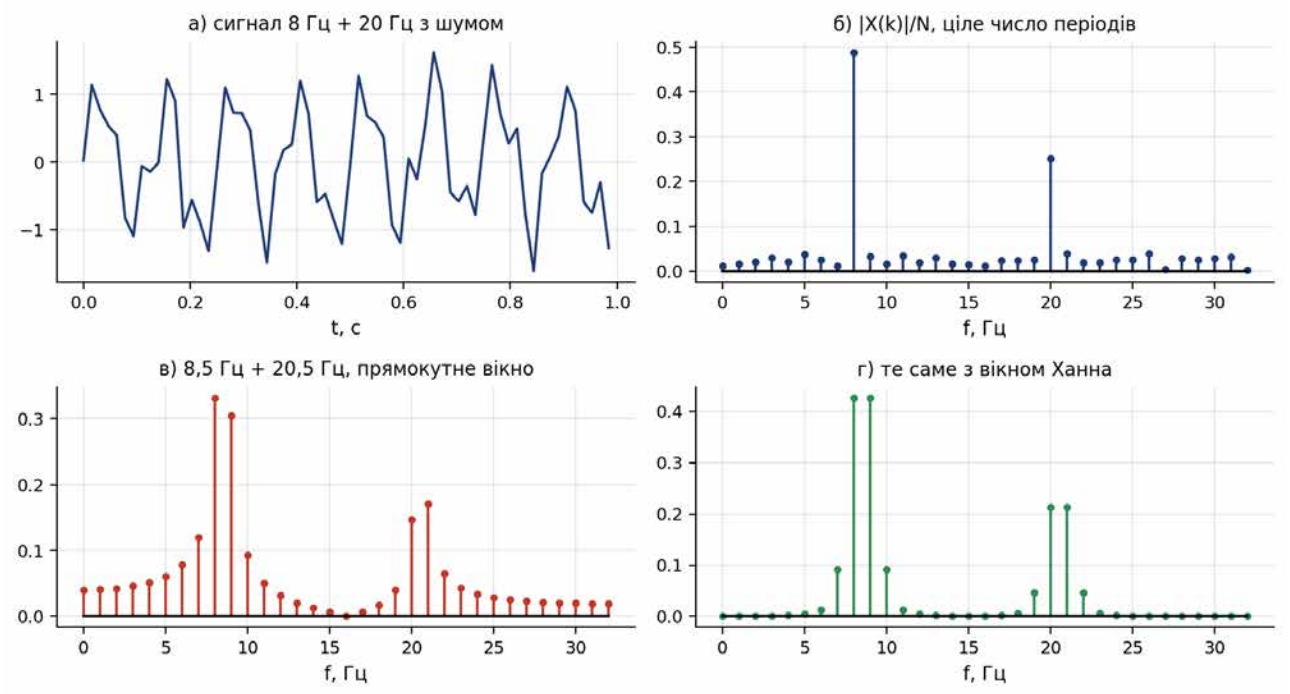


Рис. 4.2. Амплітудні спектри ДПФ: ціле і нецiле число періодів на інтервалі аналізу

4.3.4. Розтікання спектра і вікна

На рис. 4.2, в показано спектр сигналу з частотами 8,5 і 20,5 Гц, тобто таких, що на інтервалі аналізу вміщують нецiле число періодів. Замість двох чітких піків бачимо розмиті максимуми і «хвости», що поширюються на весь діапазон частот. Це явище називають розтіканням спектра (англ. *spectral leakage*). Його причина в тому, що ДПФ неявно продовжує сигнал періодично, а при нецілому числі періодів на межах інтервалу виникають розриви, які в спектрі проявляються широкосмуговими складовими. Еквівалентно, обмеження сигналу інтервалом тривалістю T — це множення на прямокутне вікно, а в частотній області — згортка зі спектром прямокутного вікна, що має форму $\sin(x)/x$ з повільно спадаючими бічними пелюстками.

Для зменшення розтікання сигнал перед обчисленням ДПФ множать на вікно $w(n)$ — функцію, що плавно спадає до нуля на краях інтервалу і тим самим усуває розриви [16]. Найпоширеніші вікна:

- Ханна: $w(n) = 0{,}5 - 0{,}5\cos\big(2\pi n/(N-1)\big)$;
- Хеммінга: $w(n) = 0{,}54 - 0{,}46\cos\big(2\pi n/(N-1)\big)$;
- Блекмана: $w(n) = 0{,}42 - 0{,}5\cos\big(2\pi n/(N-1)\big) + 0{,}08\cos\big(4\pi n/(N-1)\big)$.

Форми вікон та їх амплітудні спектри наведено на рис. 4.3. Вибір вікна завжди є компромісом: зменшення рівня бічних пелюсток (з −13 дБ для прямокутного до −31 дБ для Ханна, −43 дБ для Хеммінга і −58 дБ для Блекмана) супроводжується розширенням головної пелюстки, тобто погіршенням частотної роздільної здатності. Результат застосування вікна Ханна показано на рис. 4.2, г: бічні складові практично зникли, хоча кожний пік тепер займає кілька сусідніх трансформант.

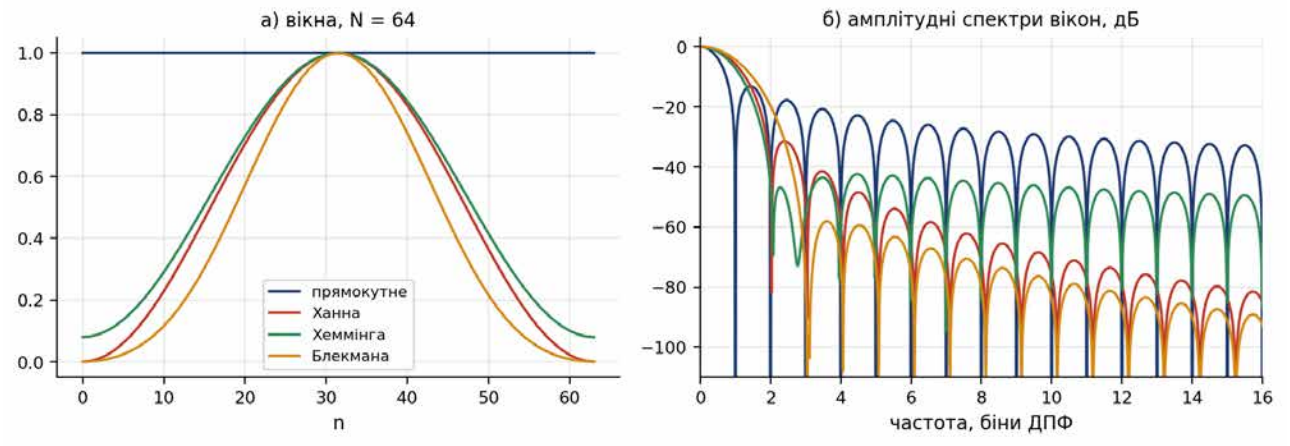


Рис. 4.3. Віконні функції та їх амплітудні спектри

4.3.5. Ефект Гіббса

Тісно пов’язаним з розтіканням спектра є ефект Гіббса — поява осциляцій поблизу розривів сигналу при його відновленні за обмеженою кількістю гармонік. На рис. 4.4 показано часткові суми ряду Фур’є для меандру при різній кількості врахованих гармонік K. Зі збільшенням K осциляції стискаються до точки розриву, але амплітуда першого викиду не зменшується, прямуючи до приблизно 9 % від величини стрибка (значення 1,09 на рис. 4.4, б). Цей ефект принципово не усувається збільшенням кількості членів ряду: усіченням ряду Фур’є неможливо рівномірно наблизити розривну функцію.

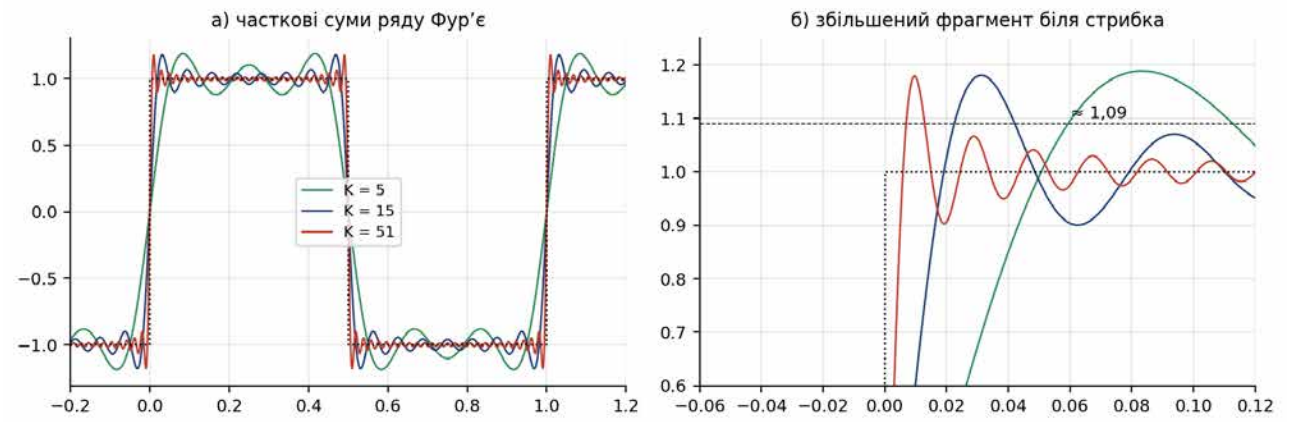


Рис. 4.4. Ефект Гіббса: часткові суми ряду Фур’є для меандру

Для оброблення зображень ефект Гіббса має велике практичне значення. Межі об’єктів на зображенні — це саме розриви яскравості. Будь-яка операція, що «обрізає» спектр Фур’є (ідеальна низькочастотна фільтрація, стиснення з відкиданням високочастотних коефіцієнтів, реставрація з обмеженням спектра через шум), породжує біля контурів хибні світлі та темні смуги — так званий «дзвін» (англ. *ringing*), що візуально сприймається як розмивання меж графоелементів. Під час реставрації образів в області трансформант Фур’є цей ефект є одним з основних чинників, що погіршують якість результату [2]. Саме тому значна увага приділяється базисам, функції яких краще пристосовані до подання розривів: перетворенням з прямокутними базисними функціями (Уолша — Адамара, Хаара) та вейвлетам (розділи 5, 6, 9).

4.4. Теорема про циклічну згортку і діагоналізація циркулянтних матриць

4.4.1. Теорема про циклічну згортку

Циклічною згорткою послідовностей $g(n)$ і $f(n)$ довжиною N називають

$$y(n) = \sum_{m=0}^{N-1} g\left((n-m) \bmod N\right) f(m). \tag{4.19}$$

Для неї справедлива дискретна теорема про згортку:

$$y(n) = g(n) \circledast f(n) \;\leftrightarrow\; Y(k) = G(k)\,F(k). \tag{4.20}$$

Доведення безпосередньо випливає з властивості циклічного зсуву: підставивши (4.19) у (4.15) і змінивши порядок підсумовування, отримаємо $Y(k) = \sum_m f\,(m) \sum_n g\left((n-m) \bmod N\right) w^{kn} = \sum_m f\,(m) w^{km} G(k) = G(k)F(k)$.

4.4.2. Матрична форма: діагоналізація

У матричних позначеннях циклічна згортка (4.19) — це множення вектора f на циркулянтну матрицю G_C (3.14). Теорема (4.20) стверджує, що в області трансформант Фур'є цій операції відповідає поелементне множення на вектор $G(k)$, тобто множення на діагональну матрицю. За формулою (4.10) це означає, що

$$\mathrm{F_н}\,\mathrm{G}_C\,\mathrm{F}_\mathrm{н}^H = \mathrm{diag}\{G(0), G(1), \ldots, G(N-1)\}, \tag{4.21}$$

де $G(k)$ — ДПФ першого стовпця матриці G_C (тобто імпульсної характеристики). Іншими словами, будь-яка циркулянтна матриця діагоналізується дискретним перетворенням Фур'є, а її власні числа є значеннями частотної характеристики системи, власні вектори — дискретними гармоніками. Чисельну перевірку (4.21) ілюструє рис. 4.5: усі позадіагональні елементи матриці $\mathrm{F_н G}_C \mathrm{F}_\mathrm{н}^H$ мають порядок 10^{-16}, тобто є похибками округлення.

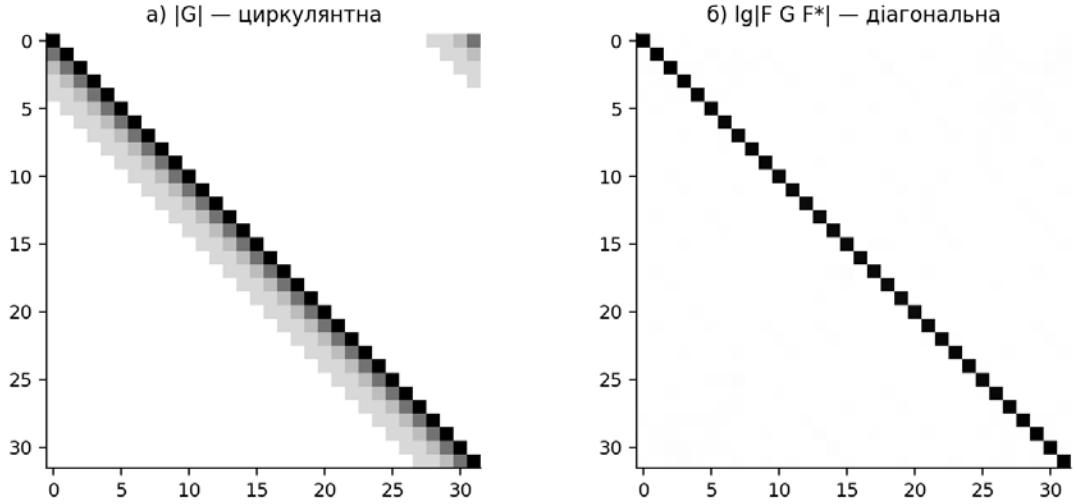


Рис. 4.5. Діагоналізація циркулянтної матриці перетворенням Фур'є

Для двовимірного випадку аналогічно: блочно-циркулянтна матриця з циркулянтними блоками діагоналізується двовимірним ДПФ, причому на

діагоналі розташовуються значення двовимірної частотної характеристики $G(m,n)$, впорядковані рядок за рядком [2].

Наслідки (4.21) для задачі реставрації важко переоцінити. Обернення циркулянтної матриці розміром $N^2 \times N^2$ (для двовимірного образу $N \times N$), що вимагало б порядку N^6 операцій, замінюється обчисленням двох ДПФ (по $O(N^2 \log N)$ операцій кожне) та N^2 поелементних ділень. Для образу 1024×1024 це різниця між 10^{18} і приблизно 10^8 операцій. Як показано в [2], саме це робить метод умовної деконволюції, сформульований у натуральних координатах, практично реалізованим в області трансформант Фур'є (розділ 8).

4.4.3. Обчислення лінійної згортки через ДПФ

Фізичні системи виконують лінійну, а не циклічну згортку. Щоб обчислити лінійну згортку послідовностей довжиною N_1 і N_2 за допомогою ДПФ, обидві послідовності доповнюють нулями до довжини $N \geq N_1 + N_2 - 1$. Тоді «загортання» відліків за модулем N не відбувається, і циклічна згортка збігається з лінійною. Для зображень аналогічно доповнюють нулями обидва виміри. Якщо ж доповнення не виконати, циклічна модель вноситься свідомо, і на межах образу з'являються крайові ефекти.

4.5. Швидке перетворення Фур'є

Безпосереднє обчислення ДПФ за формулою (4.15) вимагає N^2 комплексних множень. Алгоритми швидкого перетворення Фур'є (ШПФ), систематизовані Кулі і Тьюкі [46], зменшують цю кількість до порядку $N \log_2 N$.

Розглянемо алгоритм з проріджуванням у часі для $N = 2^p$. Розіб'ємо послідовність на парні та непарні відліки:

$$X(k) = \sum_{m=0}^{N/2-1} x\,(2m)\, w_N^{2mk} + w_N^k \sum_{m=0}^{N/2-1} x\,(2m+1)\, w_N^{2mk}. \qquad (4.22)$$

Оскільки $w_N^2 = w_{N/2}$, кожна з сум є ДПФ половинної довжини: $X(k) = X_{\text{п}}(k) + w_N^k X_{\text{н}}(k)$. З урахуванням періодичності $X_{\text{п}}$, $X_{\text{н}}$ з періодом $N/2$ і того, що $w_N^{k+N/2} = -w_N^k$, маємо

$$X(k) = X_{\text{п}}(k) + w_N^k X_{\text{н}}(k), \qquad X(k + N/2) = X_{\text{п}}(k) - w_N^k X_{\text{н}}(k), \qquad (4.23)$$

$k = 0, \ldots, N/2 - 1$. Пару операцій (4.23) називають «метеликом». Рекурсивно застосовуючи розбиття, отримуємо $\log_2 N$ етапів по $N/2$ метеликів, тобто $(N/2)\log_2 N$ комплексних множень. Для $N = 1024$ це 5120 множень замість понад мільйона — виграш у 200 разів, а для $N = 2^{16}$ — понад 4000 разів (рис. 4.6). Існують також алгоритми з проріджуванням за частотою, для довжин, що не є степенем двійки, та спеціалізовані алгоритми для дійсних даних.

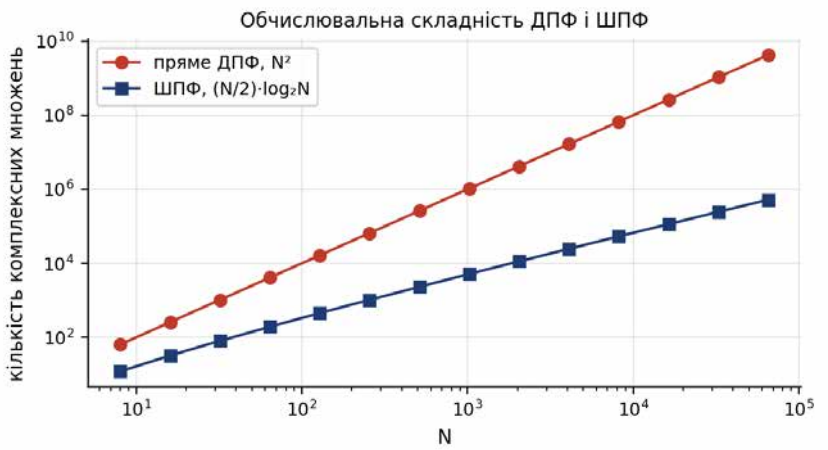


Рис. 4.6. Обчислювальна складність прямого ДПФ і ШПФ

Широко поширена думка про «трудомісткість» перетворення Фур'є, яке нібито ускладнює оброблення, не відповідає дійсності: завдяки ШПФ додаткові витрати на перехід в область трансформант настільки малі порівняно з виграшем від діагоналізації операторів, що перетворення Фур'є, навпаки, на кілька порядків спрощує процедури фільтрації та реставрації [2].

4.6. Двовимірне дискретне перетворення Фур'є

4.6.1. Визначення і сепарабельність

Для зображення $f(m,n)$ розміром $M \times N$ двовимірне ДПФ визначається як

$$F(u,v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f\,(m,n)\, e^{-j2\pi\left(\frac{um}{M}+\frac{vn}{N}\right)}, \tag{4.24}$$

а обернене —

$$f(m,n) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F\,(u,v)\, e^{j2\pi\left(\frac{um}{M}+\frac{vn}{N}\right)}. \tag{4.25}$$

Базисними функціями є двовимірні гармоніки — «хвилі» яскравості, що поширюються в напрямку вектора (u,v) з просторовою частотою $\sqrt{(u/M)^2+(v/N)^2}$. Ядро перетворення (4.24) сепарабельне: $e^{-j2\pi(um/M+vn/N)} = e^{-j2\pi um/M}\, e^{-j2\pi vn/N}$. Тому двовимірне ДПФ обчислюється як послідовність одновимірних ДПФ — спочатку всіх рядків, потім усіх стовпців результату [13, 14]. У матричній формі

$$\mathrm{F}_{2D} = \mathrm{F}_M\, \mathrm{f}\, \mathrm{F}_N^T, \tag{4.26}$$

де f — матриця зображення, F_M, F_N — матричні оператори одновимірних ДПФ. Така сепарабельність властива всім двовимірним перетворенням, що розглядаються в посібнику (косинусному, Уолша — Адамара, Хаара), і дозволяє обчислювати їх за $O\big(MN\log(MN)\big)$ операцій.

4.6.2. Подання і тлумачення двовимірного спектра

Для наочного відображення спектр зазвичай центрують: переставляють квадранти так, щоб нульова частота опинилася в центрі (це еквівалентно множенню зображення на $(-1)^{m+n}$ перед перетворенням). Оскільки динамічний діапазон амплітудного спектра дуже великий (постійна складова може на кілька порядків перевищувати високочастотні), відображають логарифм $\log(1 + |F(u,v)|)$.

На рис. 4.7 наведено кілька зображень та їх центровані амплітудні спектри. Гармонічна решітка (вертикальні смуги) дає пару точок на горизонтальній осі на відстані, що відповідає її частоті. Похила решітка дає пару точок на прямій, перпендикулярній до смуг; хрестоподібні лінії в її спектрі — прояв розтікання через нецілe число періодів у кадрі. Прямокутник дає спектр у вигляді добутку двох функцій $\sin(x)/x$, причому вужчий бік прямокутника відповідає ширшому спектру. Спектр реального зображення має максимум у центрі і швидко спадає з частотою; яскраві промені, що проходять через центр, відповідають протяжним прямим контурам на зображенні і перпендикулярні до них. Поворот зображення повертає спектр на той самий кут.

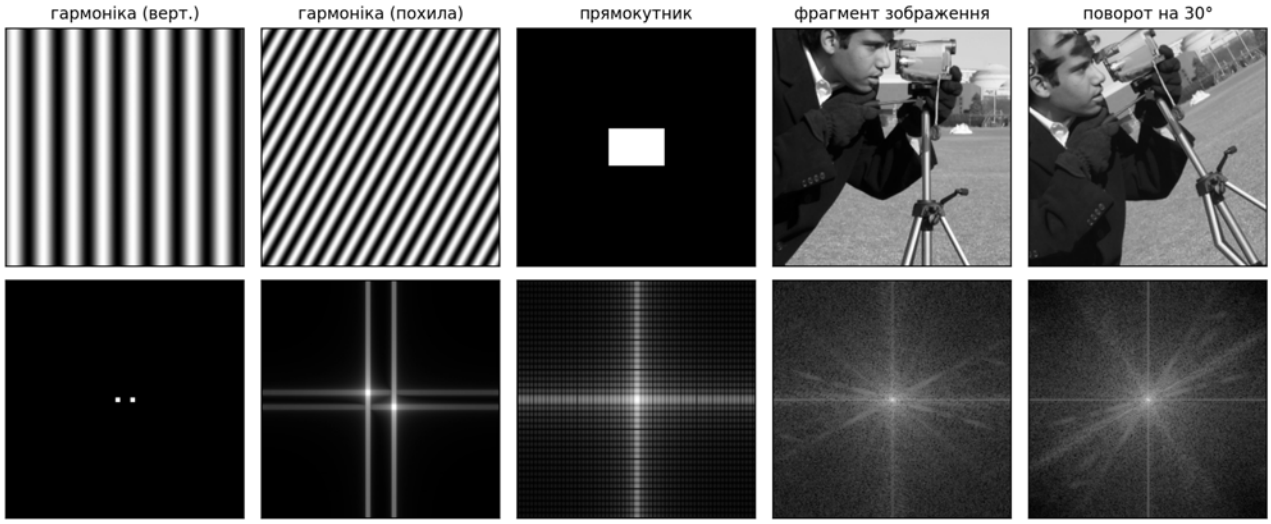


Рис. 4.7. Зображення (вгорі) та їх центровані логарифмічні амплітудні спектри (внизу)

4.6.3. Амплітуда і фаза

Двовимірний спектр комплексний: $F(u,v) = |F(u,v)|\, e^{j\phi(u,v)}$. Інтуїтивно здається, що основна інформація міститься в амплітудному спектрі, проте насправді для зображень важливішою є фаза. Це демонструє класичний експеримент (рис. 4.8): якщо обчислити спектри двох зображень A і B, «обміняти» їх фази і виконати обернене перетворення, то зображення, синтезоване з амплітуди A і фази B, візуально нагадуватиме B, і навпаки [13, 16]. Причина полягає в тому, що положення контурів і деталей закодоване саме у взаємних фазових співвідношеннях гармонік; амплітудні спектри природних зображень досить схожі між собою (спадають з частотою приблизно як $1/f$).

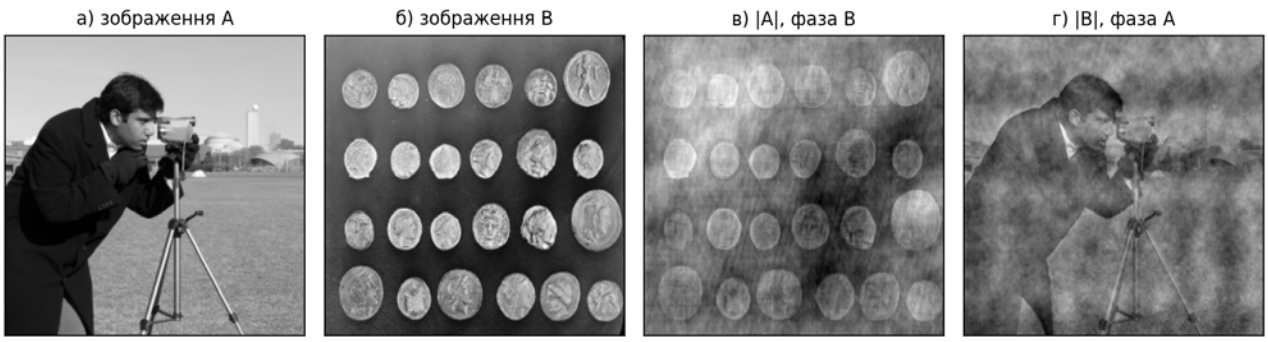


Рис. 4.8. Експеримент з обміном амплітудних і фазових спектрів

З цього випливають два висновки для розпізнавання. По-перше, амплітудний спектр інваріантний до зсуву (властивість 4 ДПФ), тому він може бути основою ознак, незалежних від положення об’єкта, — проте при цьому втрачається значна частина інформації про форму. По-друге, методи, що працюють з фазою (наприклад, фазова кореляція, підрозділ 4.8), дозволяють точно визначати взаємне положення образів.

4.7. Фільтрація зображень у частотній області

4.7.1. Загальна схема

Згідно з теоремою про згортку, лінійну інваріантну до зсуву фільтрацію зображення можна виконати в частотній області [13]:

1. обчислити ДПФ зображення $F(u, v)$ (за потреби — з доповненням нулями);
2. помножити на частотну характеристику фільтра $H(u, v)$;
3. обчислити обернене ДПФ і взяти дійсну частину.

Фільтрацію в частотній області доцільно застосовувати, коли імпульсна характеристика фільтра велика (тоді вона швидша за пряму згортку), коли фільтр природно задається в частотах (наприклад, для придушення періодичних завад) або коли потрібно обернути оператор згортки (реставрація). Позначимо $D(u, v)$ відстань від точки (u, v) до початку координат у центрованій частотній площині, D_0 — частоту зрізу.

4.7.2. Фільтри нижніх частот

Ідеальний ФНЧ пропускає без змін усі частоти в колі радіусом D_0 і повністю придушує решту:

$$H(u, v) = \begin{cases} 1, & D(u, v) \le D_0, \\ 0, & D(u, v) > D_0. \end{cases} \tag{4.27}$$

Внаслідок різкого обрізання спектра імпульсна характеристика ідеального фільтра має вигляд двовимірної функції sinc з повільно спадаючими осциляціями, і результат фільтрації містить виражений «дзвін» біля контурів — прояв ефекту Гіббса (рис. 4.9, ліворуч).

Фільтр Баттерворта порядку n має плавну частотну характеристику [13, 16]

$$H(u,v) = \frac{1}{1 + [D(u,v)/D_0]^{2n}}. \tag{4.28}$$

При малих порядках ($n = 1{,}2$) дзвін практично відсутній; зі збільшенням порядку фільтр наближається до ідеального.

Гаусівський ФНЧ

$$H(u,v) = e^{-D^2(u,v)/\left(2D_0^2\right)} \tag{4.29}$$

має гаусівську імпульсну характеристику без бічних пелюсток і тому взагалі не створює дзвону. Ціною є менш крутий спад частотної характеристики.

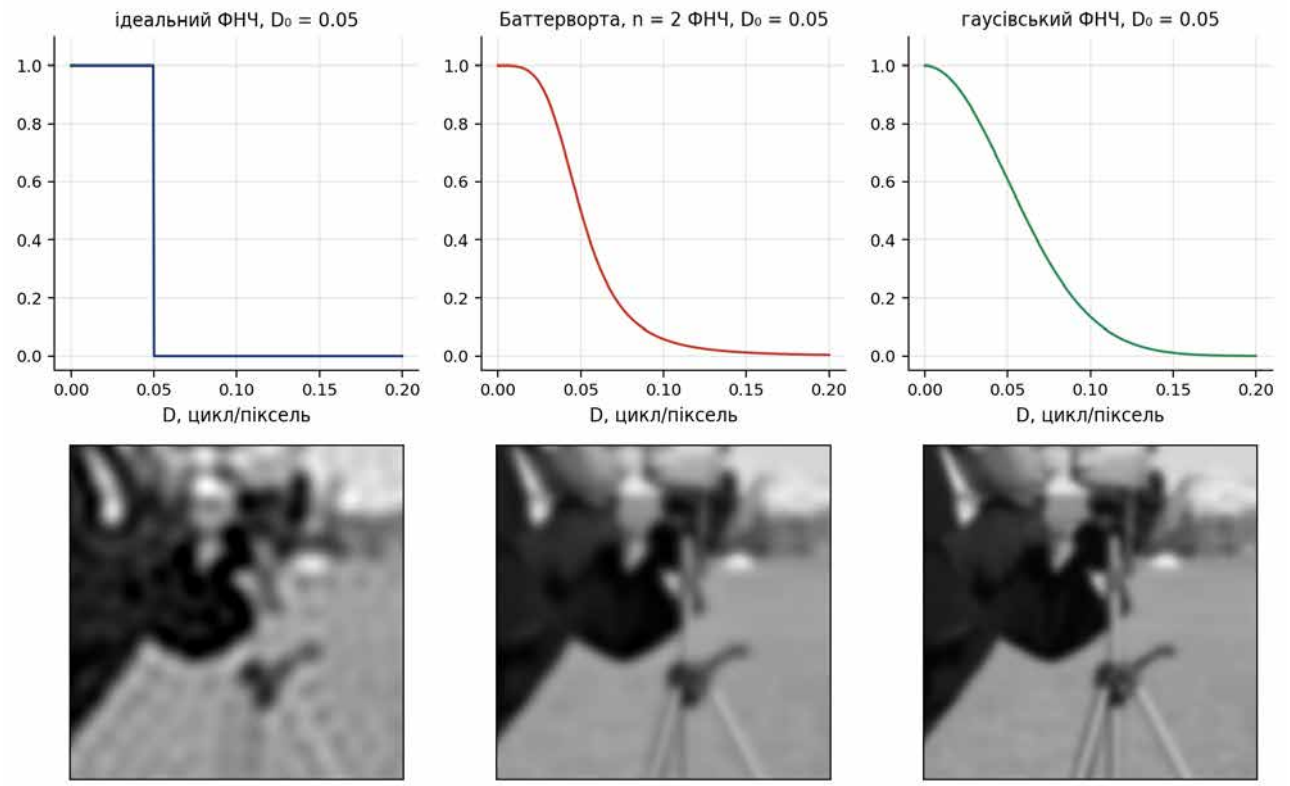


Рис. 4.9. Частотні характеристики фільтрів нижніх частот (вгорі) та результати фільтрації (внизу)

4.7.3. Фільтри верхніх частот і підвищення різкості

Фільтри верхніх частот отримують з фільтрів нижніх частот як $H_{\text{Вч}} = 1 - H_{\text{Нч}}$. Вони придушують повільні зміни яскравості і підкреслюють контури та дрібні деталі (рис. 4.10, б). Додавши до вихідного зображення його високочастотну складову з деяким коефіцієнтом k, отримаємо зображення з підвищеною різкістю (рис. 4.10, в):

$$f_{\text{р}}(m,n) = f(m,n) + k\,\mathcal{F}^{-1}\{H_{\text{Вч}}(u,v)\,F(u,v)\}. \tag{4.30}$$

У просторовій області аналогічна операція відома як нерізке маскування (розділ 7).



Рис. 4.10. Високочастотна фільтрація та підвищення різкості

Гомоморфна фільтрація ґрунтується на моделі (2.3) формування зображення як добутку освітленості і коефіцієнта відбиття. Після логарифмування $\ln f = \ln i + \ln r$ складові стають адитивними; фільтр, що послаблює низькі частоти (повільні зміни освітленості) і підсилює високі (деталі відбиття), з подальшим потенціюванням дозволяє одночасно стиснути динамічний діапазон і підвищити локальний контраст [13].

Фільтри для придушення періодичних завад (наприклад, наведень від мережі живлення або растру друку) задаються як режекторні — з нулями в околах точок, де в спектрі спостерігаються піки завади [13, 14].

4.8. Кореляція в частотній області і суміщення образів

4.8.1. Взаємна кореляція

Взаємна кореляційна функція двох образів f і s

$$R_{fs}(m,n) = \sum_i \sum_j f\,(i+m, j+n)\, s(i,j) \tag{4.31}$$

відрізняється від згортки лише відсутністю дзеркального відображення одного з операндів. Тому вона також ефективно обчислюється за допомогою ДПФ:

$$R_{fs} \;\leftrightarrow\; F(u,v)\, S^*(u,v). \tag{4.32}$$

Максимум взаємної кореляційної функції вказує на зсув, при якому образи найкраще збігаються. Це основа кореляційних методів виявлення об’єкта заданої форми (погодженої фільтрації), що детально розглядаються в розділі 10. Проте для природних зображень, енергія яких зосереджена на низьких частотах, кореляційна функція має широкий і пологий максимум (рис. 4.11, б), положення якого визначається неточно, особливо при шумі.

4.8.2. Фазова кореляція

Метод фазової кореляції [47] нормує взаємний спектр на його модуль, залишаючи лише фазову інформацію:

$$P(u,v) = \frac{F(u,v)\,S^*(u,v)}{|F(u,v)\,S^*(u,v)|}, \qquad p(m,n) = \mathcal{F}^{-1}\{P(u,v)\}. \tag{4.33}$$

Якщо $f(m,n) = s(m-m_0, n-n_0)$, то за властивістю зсуву $P(u,v) = e^{-j2\pi(um_0/M+vn_0/N)}$, а її обернене ДПФ — дельта-функція в точці (m_0, n_0). На практиці (за наявності шуму і неповного перекриття) функція фазової кореляції має гострий, добре локалізований пік на рівномірному фоні (рис. 4.11, в). Метод стійкий до змін освітленості і контрасту (оскільки амплітуда нормується) і широко застосовується для суміщення кадрів відеопослідовностей, стабілізації зображень, побудови панорам, реєстрації медичних зображень. Поєднуючи фазову кореляцію з переходом до логарифмічно-полярних координат, можна визначати також поворот і зміну масштабу.

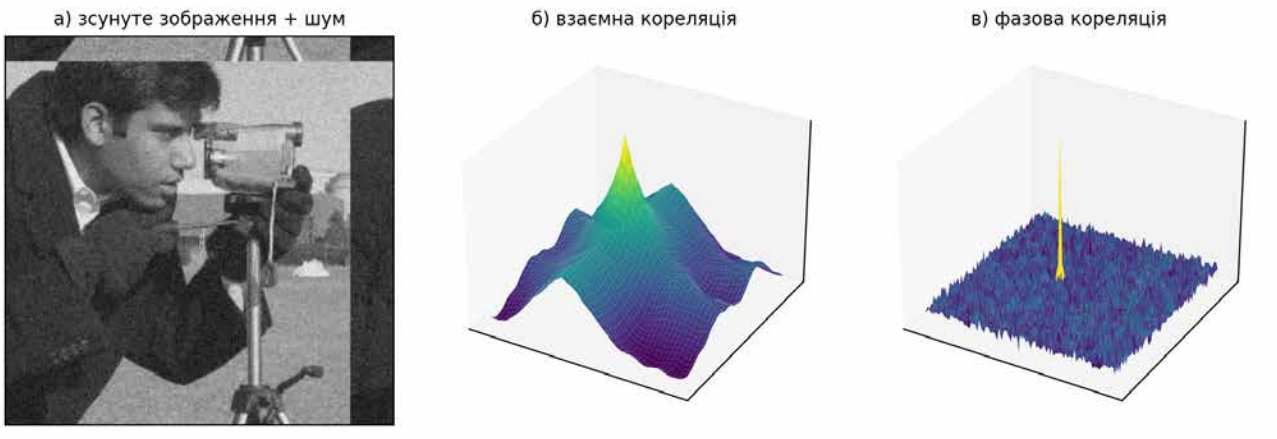


Рис. 4.11. Визначення зсуву методами взаємної та фазової кореляції

4.9. Спектральні ознаки образів

Спектральне подання є джерелом ознак для розпізнавання, інваріантних до певних перетворень образу.

Ознаки на основі амплітудного спектра зображення. Амплітудний спектр інваріантний до зсуву. Розділивши частотну площину на концентричні кільця і обчисливши енергію спектра в кожному з них, отримаємо ознаки, що характеризують розподіл деталей за розміром (масштабом) і є інваріантними до повороту. Енергії в секторах (клинах) характеризують переважну орієнтацію деталей і є інваріантними до масштабу. Такі ознаки ефективні для розпізнавання текстур і класифікації ділянок зображень дистанційного зондування.

Дескриптори Фур'є контуру [48]. Якщо контур об'єкта подано послідовністю N точок (x_n, y_n), утворимо комплексну послідовність $z(n) = x_n + j\,y_n$ і обчислимо її ДПФ $Z(k)$. Коефіцієнти $Z(k)$ називають дескрипторами Фур'є. Низькочастотні дескриптори описують загальну форму контуру, високочастотні — дрібні деталі; на рис. 4.12 показано відновлення контуру за K парами дескрипторів. Інваріантність досягається простими операціями:

- до *зсуву* — відкиданням $Z(0)$ (центру ваги контуру);
- до *масштабу* — нормуванням усіх дескрипторів на $|Z(1)|$;

- до *повороту і вибору початкової точки* — використанням лише модулів $|Z(k)|$ (поворот на кут θ множить усі дескриптори на $e^{j\theta}$, зсув початкової точки — на $e^{j2\pi k n_0/N}$).

Отже, вектор $[|Z(2)|/|Z(1)|, \dots, |Z(K)|/|Z(1)|, |Z(-1)|/|Z(1)|, \dots]$ є компактною ознакою форми, інваріантною до зсуву, повороту і масштабу.

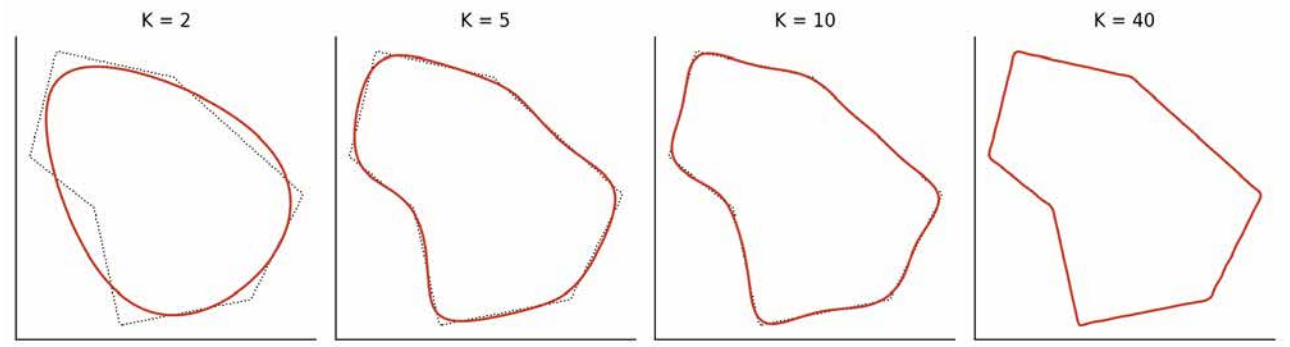


Рис. 4.12. Відновлення контуру за K парами дескрипторів Фур'є

4.10. Практична реалізація

Наведемо приклади, що ілюструють основні положення розділу: побудову матричного оператора ДПФ і перевірку його унітарності, діагоналізацію циркулянтної матриці, фільтрацію в частотній області та визначення зсуву методом фазової кореляції.

```
import numpy as np
from scipy.linalg import circulant
from skimage import data, img_as_float

N = 16
n = np.arange(N)
F = np.exp(-2j * np.pi * np.outer(n, n) / N) / np.sqrt(N)   # (4.17), нормований
print(np.allclose(F @ F.conj().T, np.eye(N)))               # унітарність (4.8)

x = np.random.default_rng(1).normal(size=N)
print(np.allclose(F @ x, np.fft.fft(x) / np.sqrt(N)))       # збіг з ШПФ

g = np.zeros(N); g[:4] = [0.4, 0.3, 0.2, 0.1]
G = circulant(g)                                            # (3.14)
D = F @ G @ F.conj().T                                      # (4.21)
off = D - np.diag(np.diag(D))
print(np.abs(off).max(), np.allclose(np.diag(D), np.fft.fft(g)))

# гаусівський ФНЧ у частотній області (4.29)
f = img_as_float(data.camera())
M = f.shape[0]
u = np.fft.fftfreq(M)[:, None]; v = np.fft.fftfreq(M)[None, :]
H = np.exp(-(u**2 + v**2) / (2 * 0.05**2))
f_lp = np.real(np.fft.ifft2(np.fft.fft2(f) * H))
```

```
# фазова кореляція (4.33)
s = f[100:356, 100:356]
fm = np.roll(s, (17, -29), axis=(0, 1))
R = np.fft.fft2(fm) * np.conj(np.fft.fft2(s))
p = np.real(np.fft.ifft2(R / (np.abs(R) + 1e-12)))
dy, dx = np.unravel_index(np.argmax(p), p.shape)
print("зсув:", dy, dx - s.shape[1])                    # (17, -29)
```

4.11. Приклади розв'язання задач

Задача 4.1. Обчисліть ДПФ послідовності $x = [1,2,3,4]$ і перевірте рівність Парсеваля.

Розв'язання. Для $N = 4$: $w = e^{-j\pi/2} = -j$. За (4.15):

$X(0) = 1 + 2 + 3 + 4 = 10;$

$X(1) = 1 + 2(-j) + 3(-1) + 4(j) = -2 + 2j;$

$X(2) = 1 - 2 + 3 - 4 = -2;$

$X(3) = 1 + 2j - 3 - 4j = -2 - 2j.$

Спряжена симетрія виконується: $X(3) = X^*(1)$. Рівність Парсеваля: $\sum|x|^2 = 1 + 4 + 9 + 16 = 30$; $\frac{1}{N}\sum|X|^2 = \frac{1}{4}(100 + 8 + 4 + 8) = 30$.

Задача 4.2. Обчисліть циклічну згортку послідовностей $a = [1,2,0,0]$ і $b = [1,1,1,0]$ безпосередньо і за допомогою ДПФ.

Розв'язання. Безпосередньо за (4.19): $y(0) = a_0b_0 + a_1b_3 = 1$; $y(1) = a_0b_1 + a_1b_0 = 3$; $y(2) = a_0b_2 + a_1b_1 = 3$; $y(3) = a_0b_3 + a_1b_2 = 2$. Отже, $y = [1,3,3,2]$. Через ДПФ: $A = [3,\ 1 - 2j,\ -1,\ 1 + 2j]$, $B = [3,\ -j,\ 1,\ j]$; поелементний добуток $Y = [9,\ -2 - j,\ -1,\ -2 + j]$; обернене ДПФ дає $[1,3,3,2]$, що збігається з безпосереднім обчисленням.

Задача 4.3. Сигнал дискретизовано з частотою 8 кГц, ДПФ обчислюється для фрагментів по 256 відліків. Визначте частотну роздільну здатність і частоту, що відповідає трансформанті з номером 40.

Розв'язання. За (4.18) $\Delta f = 8000/256 = 31{,}25$ Гц; $f_{40} = 40 \cdot 31{,}25 = 1250$ Гц. Трансформанта з номером $216 = 256 - 40$ відповідає частоті -1250 Гц; для дійсного сигналу її модуль дорівнює модулю трансформанти 40.

Висновки до розділу 4

1. Ортогональне перетворення є розкладом образу за базисом ортогональних функцій; трансформанти вимірюють подібність образу до базисних функцій. Унітарне перетворення зберігає енергію і відстані між образами, тому не змінює інформаційного змісту, але може зосередити енергію в малій кількості коефіцієнтів і спростити структуру операторів.
2. Матриця лінійного оператора в області трансформант отримується перетворенням подібності WAW^H. Вибір базису, в якому матриця оператора деградації стає діагональною або розрідженою, — основний засіб спрощення задач фільтрації і реставрації.

3. ДПФ розкладає образ за дискретними комплексними гармоніками. Воно періодичне, для дійсних сигналів має спряжено-симетричний спектр, а його амплітудний спектр інваріантний до циклічного зсуву.
4. Теорема про циклічну згортку в матричній формі означає, що будь-яка циркулянтна матриця діагоналізується ДПФ, а її власні числа — значення частотної характеристики. Це зводить обернення оператора згортки до поелементного ділення спектрів.
5. Швидке перетворення Фур'є зменшує складність з N^2 до $(N/2)\log_2 N$ операцій, завдяки чому перехід в область трансформант є економічним.
6. Обмеження інтервалу аналізу спричиняє розтікання спектра, а усічення спектра — ефект Гіббса (осциляції біля розривів з викидом близько 9 %). Для зображень ефект Гіббса проявляється як «дзвін» і розмивання меж графоелементів; він є суттєвим недоліком методів, що обрізають спектр Фур'є.
7. Двовимірне ДПФ сепарабельне. Фаза спектра несе основну інформацію про розташування деталей зображення, амплітуда — про їх енергетичний розподіл за частотами і орієнтаціями.
8. У частотній області ефективно реалізуються фільтрація, кореляційне виявлення і суміщення (фазова кореляція), а також формування ознак, інваріантних до зсуву, повороту і масштабу (кільцеві та секторні енергії, дескриптори Фур'є).

Питання для обговорення та самоперевірки

1. Що таке трансформанта ортогонального перетворення? Як її обчислити для ортогонального, але не нормованого базису?
2. Сформулюйте умови ортогональності і повноти дискретного перетворення.
3. Яку матрицю називають унітарною? Як знайти обернений оператор для унітарного перетворення?
4. Запишіть рівність Парсеваля і поясніть її геометричний зміст.
5. Як перетворюється матриця лінійного оператора при переході в область трансформант?
6. Сформулюйте теореми про згортку та диференціювання для неперервного перетворення Фур'є.
7. Запишіть пряме й обернене ДПФ. Побудуйте матричний оператор ДПФ для $N = 4$.
8. Якій фізичній частоті відповідає k-та трансформанта ДПФ? Від чого залежить частотна роздільна здатність?
9. Чому для дійсного сигналу незалежними є лише $N/2 + 1$ трансформант?
10. Доведіть теорему про циклічну згортку.
11. Поясніть, чому будь-яка циркулянтна матриця діагоналізується ДПФ. Що стоїть на її діагоналі?
12. Як обчислити лінійну згортку за допомогою ДПФ?
13. Опишіть принцип алгоритму ШПФ з проріджуванням у часі. Що таке «метелик»?

14. Що таке розтікання спектра? Як його зменшити? Який компроміс пов’язаний з вибором вікна?
15. У чому полягає ефект Гіббса? Чому він не зникає зі збільшенням кількості гармонік?
16. Як проявляється ефект Гіббса при обробленні зображень?
17. Що означає сепарабельність двовимірного ДПФ? Яку перевагу вона дає?
18. Як змінюється спектр зображення при зсуві, повороті, зміні масштабу?
19. Яку інформацію несуть амплітудний і фазовий спектри зображення?
20. Порівняйте ідеальний, Баттерворта і гаусівський фільтри нижніх частот.
21. У чому полягає метод фазової кореляції? Чому він точніший за звичайну взаємну кореляцію?
22. Як отримати дескриптори Фур’є контуру, інваріантні до зсуву, повороту і масштабу?

Завдання для самостійної роботи

1. Побудуйте матричний оператор ДПФ для $N = 8$, перевірте ортогональність його рядків і обчисліть енергії E_k. Сформуйте нормований оператор і перевірте рівність Парсеваля для випадкового вектора.
2. Реалізуйте рекурсивний алгоритм ШПФ з проріджуванням у часі. Порівняйте час виконання з прямим обчисленням ДПФ та бібліотечною функцією для $N = 2^6 \dots 2^{14}$.
3. Для сигналу, що є сумою двох гармонік із близькими частотами, дослідіть, при якій тривалості спостереження вони розрізняються в спектрі для прямокутного вікна і вікна Ханна.
4. Обчисліть часткові суми ряду Фур’є для прямокутного імпульсу і визначте амплітуду викиду для $K = 11{,}51{,}201$. Порівняйте з теоретичним значенням.
5. Перевірте чисельно співвідношення (4.21) для циркулянтної і для теплицевої матриць. Наскільки теплицева матриця «відхиляється» від діагональної форми в базисі Фур’є?
6. Виконайте фільтрацію зображення ідеальним ФНЧ, фільтрами Баттерворта ($n = 1{,}2{,}5$) та гаусівським для однакової частоти зрізу. Порівняйте результати за PSNR відносно вихідного зображення та візуально.
7. Повторіть експеримент з обміном фаз (рис. 4.8) для власних зображень.
8. Реалізуйте визначення зсуву методом фазової кореляції з субпіксельною точністю (параболічною інтерполяцією піка). Дослідіть точність залежно від рівня шуму.
9. Для набору бінарних силуетів кількох класів (наприклад, ключів, листків рослин) обчисліть інваріантні дескриптори Фур’є і побудуйте класифікатор за мінімальною відстанню до еталона.

РОЗДІЛ 5. ДИСКРЕТНІ ОРТОГОНАЛЬНІ ПЕРЕТВОРЕННЯ З ДІЙСНИМ ЯДРОМ

Навчальні цілі. Після вивчення розділу ви зможете: будувати матриці перетворень Уолша — Адамара, Хаара та дискретного косинусного перетворення; переходити між різними впорядкуваннями функцій Уолша; реалізовувати швидке перетворення Уолша — Адамара; пояснювати причини високої енергетичної компактності косинусного перетворення та його зв'язок з перетворенням Карунена — Лоева; формулювати теорему про диференціювання оригіналу для довільного дискретного ортогонального перетворення; порівнювати перетворення за обчислювальною складністю, компактністю та структурою операторів.

5.1. Вступні зауваження

Перетворення Фур'є і споріднені з ним перетворення Лапласа і Z-перетворення є найпоширенішим засобом аналізу лінійних систем: вони перетворюють диференціальні та різницеві рівняння на алгебраїчні і діагоналізують оператори згортки. Проте для задач аналізу, стиснення, передачі і розпізнавання сигналів перетворення Фур'є з його комплексним ядром $e^{-j\omega t}$ не завжди є зручним або оптимальним [1, 22].

По-перше, комплексна арифметика подвоює обсяг пам'яті та збільшує кількість операцій, а для дійсних образів половина трансформант ДПФ є надлишковою (спряженою). По-друге, базисні функції Фур'є — гармоніки — нескінченно гладкі і не локалізовані в просторі; вони погано пристосовані до подання розривів, що призводить до ефекту Гіббса (підрозділ 4.3.5). По-третє, перетворення Фур'є оптимальне для періодичних сигналів, а більшість образів, з якими працює система розпізнавання, неперіодичні.

Тому поряд з ДПФ широкого вжитку набули перетворення з дійсним ядром. Серед них насамперед виділяють перетворення Уолша (і його окремий випадок — Адамара), що поширилося в техніці зв'язку і цифровій обробці завдяки виключно простій реалізації; перетворення Хаара, яке є першим прикладом вейвлет-перетворення; та косинусне перетворення, що стало стандартом у стисненні зображень і широко використовується для аналізу біосигналів [1, 22, 37]. Усі вони мають дійсні ортогональні матриці, допускають швидкі алгоритми і сепарабельне двовимірне узагальнення.

5.2. Функції та перетворення Уолша

5.2.1. Функції Радемахера і Уолша

Функції Уолша, введені Дж. Уолшем у 1923 р. [23], утворюють повну ортонормовану систему прямокутних функцій, що набувають лише значень $+1$ і -1. Найпростіше їх побудувати через функції Радемахера

$$r_p(t) = \text{sign}[\sin(2^p \pi t)], \qquad 0 \le t < 1, \quad p = 1,2, \ldots, \tag{5.1}$$

які являють собою меандри з 2^{p-1} періодами на одиничному інтервалі. Функції Радемахера ортонормовані, але не утворюють повної системи. Повну

систему отримують, утворюючи всеможливі добутки функцій Радемахера: функція Уолша з номером k, двійковий запис якого $k = \left(k_{p-1} \dots k_1 k_0\right)_2$, визначається як

$$\mathrm{wal}(k,t) = \prod_{i=0}^{p-1} [r_{i+1}(t)]^{g_i}, \tag{5.2}$$

де g_i — розряди коду Грея числа k (для впорядкування за Уолшем) або безпосередньо розряди k у тій чи іншій послідовності (для інших впорядкувань).

Важливою характеристикою функції Уолша є секвента — половина середньої кількості перетинів нуля (змін знака) на одиничному інтервалі. Секвента є узагальненням поняття частоти: для гармонік вона збігається з частотою, але має сенс і для неперіодичних прямокутних функцій. Застосування функцій Уолша для передачі інформації детально розглянуто в монографії Х. Хармута [23].

5.2.2. Матриця Адамара і кронекерів добуток

Для дискретного перетворення Уолша формату $N = 2^p$ зручніше спочатку побудувати матричний оператор, а потім, за потреби, отримати з нього неперервні функції [1]. Базовою є матриця Адамара другого порядку

$$\mathrm{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{5.3}$$

Матриці вищих порядків отримують за допомогою кронекерового добутку $\otimes$, у якому кожний елемент першого множника замінюється добутком цього елемента на всю другу матрицю:

$$\mathrm{H}_{2N} = \mathrm{H}_2 \otimes \mathrm{H}_N = \begin{bmatrix} \mathrm{H}_N & \mathrm{H}_N \\ \mathrm{H}_N & -\mathrm{H}_N \end{bmatrix}. \tag{5.4}$$

За цим простим рекурсивним правилом (запропонованим Дж. Сильвестром) порядок матриці подвоюється на кожному кроці. Наприклад,

$$\mathrm{H}_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}. \tag{5.5}$$

Елементи матриці Адамара в природному впорядкуванні можна обчислити і безпосередньо:

$$h(k,n) = (-1)^{\sum_{i=0}^{p-1} k_i n_i}, \tag{5.6}$$

де k_i, n_i — двійкові розряди номерів рядка і стовпця. Кожен рядок матриці — це відліки однієї функції Уолша. Інтервал між точками n і $n+1$ на осі часу відповідає значенню трансформанти в точці n, що дозволяє перейти від дискретних рядків до неперервних функцій (рис. 5.1, а).

Властивості матриці Адамара:

1. Елементи дорівнюють ± 1; отже, множення на матрицю зводиться до додавань і віднімань.
2. Матриця симетрична: $\mathrm{H}_N = \mathrm{H}_N^T$.
3. Рядки ортогональні, енергія кожного рядка дорівнює N: $\mathrm{H}_N \mathrm{H}_N^T = N\,\mathrm{E}$.
4. Звідси обернена матриця $\mathrm{H}_N^{-1} = \frac{1}{N}\mathrm{H}_N$, а нормована матриця $\mathrm{Had}_\mathrm{н} = \mathrm{H}_N/\sqrt{N}$ є ортогональною, симетричною і інволютивною — збігається зі своєю оберненою: $\mathrm{Had}_\mathrm{н}^{-1} = \mathrm{Had}_\mathrm{н}$.

Отже, пряме й обернене дискретне перетворення Уолша — Адамара (ДПУА) мають однаковий вигляд:

$$\mathrm{X} = \frac{1}{\sqrt{N}}\mathrm{H}_N\,\mathrm{x}, \qquad \mathrm{x} = \frac{1}{\sqrt{N}}\mathrm{H}_N\,\mathrm{X}. \tag{5.7}$$

5.2.3. Впорядкування функцій Уолша

Залежно від нумерації базисних функцій (порядку рядків матриці) розрізняють кілька варіантів перетворення, відомих як перетворення Уолша — Качмажа, Уолша — Пейлі та Уолша — Адамара [1, 23]:

- природне впорядкування (за Адамаром) — рядки розташовані так, як вони отримуються за рекурсією (5.4). Це впорядкування найзручніше для формування матриці і для математичних операцій;
- секвентне впорядкування (за Уолшем, або Уолша — Качмажа) — рядки розташовані за зростанням секвенти, тобто кількості змін знака, подібно до того, як функції Фур'є впорядковані за частотою. Це впорядкування найзручніше для спектрального аналізу та стиснення, оскільки енергія гладких образів зосереджується в трансформантах з малими номерами;
- впорядкування за Пейлі (діадне) — номер функції в цьому впорядкуванні отримується з номера за Адамаром двійковою інверсією (запис розрядів у зворотному порядку).

Існують прості алгоритми перенумерації: номер функції в природному впорядкуванні, що відповідає секвенті s, знаходять, переводячи s у код Грея і виконуючи двійкову інверсію отриманого числа. На рис. 5.1 наведено функції Уолша для $N = 8$ в природному і секвентному впорядкуваннях із зазначенням секвенти s кожної функції.

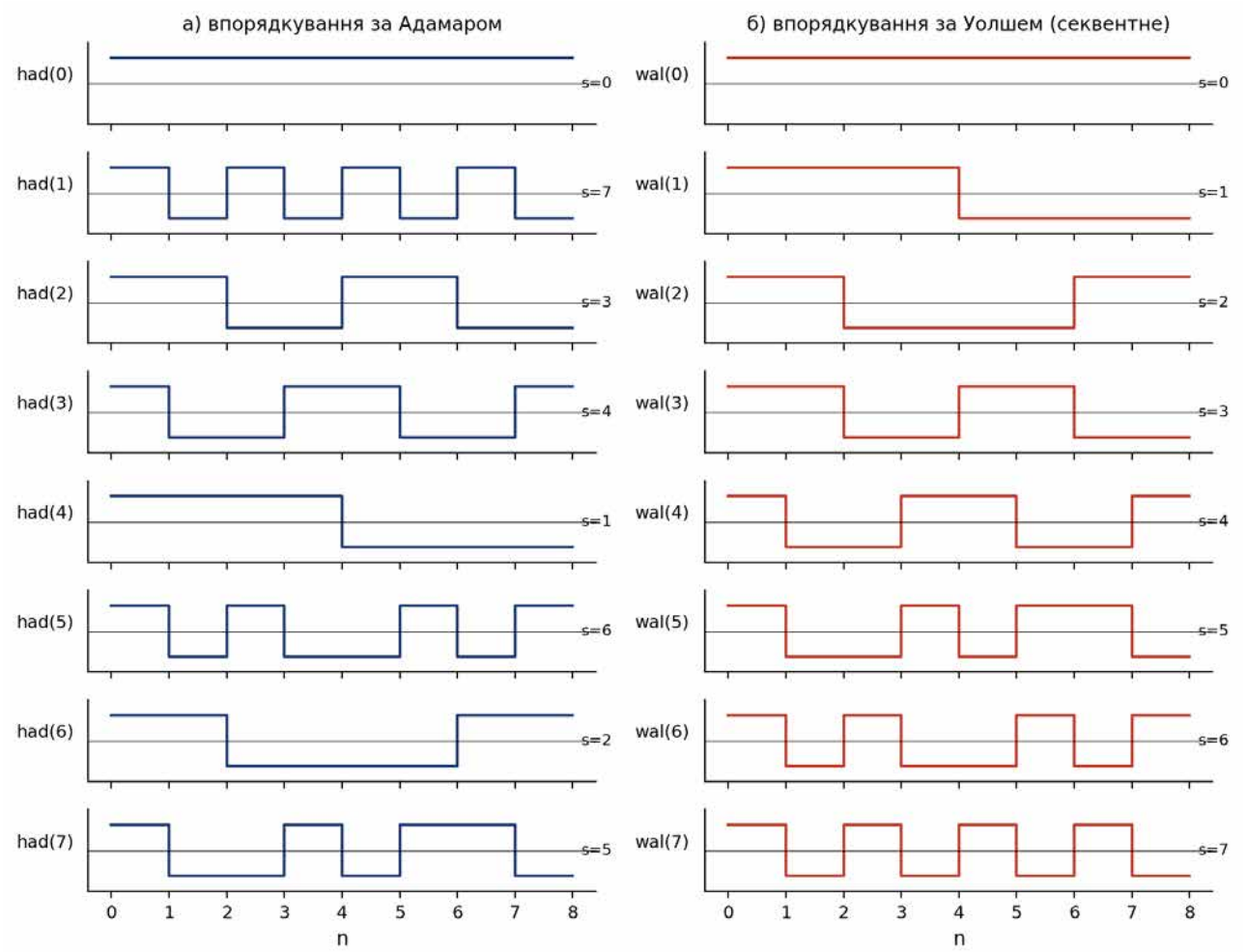


Рис. 5.1. Функції Уолша для N = 8: природне (а) та секвентне (б) впорядкування

Приклад 5.1. Обчислимо ДПУА вектора x=[4,2,2,0]^T за допомогою ненормованої матриці H_4:

H_4 x=[4+2+2+0, 4-2+2-0, 4+2-2-0, 4-2-2+0]^T=[8,4,4,0]^T.

Нульова трансформанта (8) пропорційна середньому значенню (2). Перевіримо рівність Парсеваля для нормованого перетворення: ‖x‖^2=16+4+4+0=24; ‖X‖^2=(64+16+16+0)/4=24. Обернене перетворення: 1/4 H_4 [8,4,4,0]^T=1/4 [16,8,8,0]^T=x.

5.3. Швидке перетворення Уолша — Адамара

Рекурсивна структура (5.4) дозволяє факторизувати матрицю Адамара в добуток $p = \log_2 N$ розріджених матриць, кожна з яких має в кожному рядку лише два ненульових елементи ± 1:

$$\mathrm{H}_N = \prod_{i=1}^{p} \left(\mathrm{E}_{2^{i-1}} \otimes \mathrm{H}_2 \otimes \mathrm{E}_{N/2^i} \right). \tag{5.8}$$

Множення на кожну з цих матриць — це один етап обчислень, на якому виконуються $N/2$ «метеликів» вигляду

$$a' = a + b, \qquad b' = a - b. \tag{5.9}$$

На відміну від ШПФ, метелики ПУА не містять множень на комплексні коефіцієнти. Отже, швидке перетворення Уолша — Адамара (ШПУА) потребує лише $N\log_2 N$ додавань і віднімань і жодного множення (не рахуючи кінцевого нормування). Для $N = 1024$ це 10 240 операцій додавання, що робить ПУА найпростішим у реалізації з усіх нетривіальних ортогональних перетворень [1, 23]. Граф ШПУА для $N = 8$ наведено на рис. 5.2.

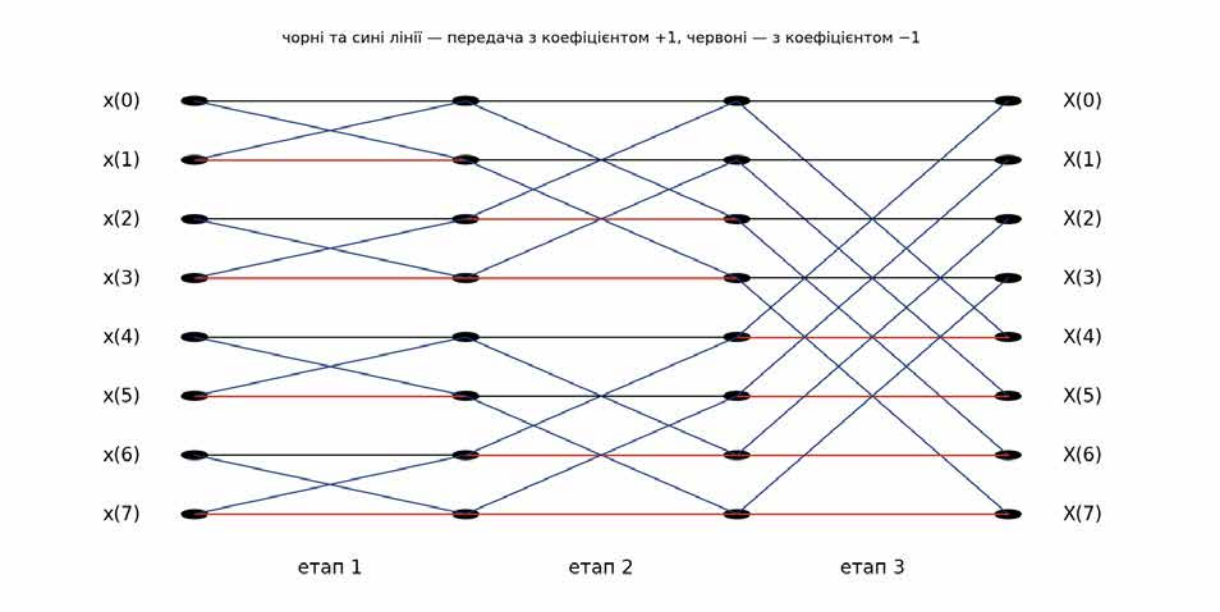


Рис. 5.2. Граф швидкого перетворення Уолша — Адамара для N = 8

Програмна реалізація ШПУА «на місці» (без додаткової пам’яті) надзвичайно компактна:

```
def fwht(a):
    """Швидке перетворення Уолша-Адамара (природне впорядкування),
    ненормоване; довжина a — степінь двійки."""
    a = a.astype(float).copy()
    h = 1
    while h < len(a):
        for i in range(0, len(a), 2 * h):
            for j in range(i, i + h):
                x, y = a[j], a[j + h]
                a[j], a[j + h] = x + y, x - y        # метелик (5.9)
        h *= 2
    return a
```

Цікаво, що через інволютивність того самого алгоритму достатньо і для оберненого перетворення (з діленням результату на N).

5.4. Згортка і перетворення Уолша — Адамара

5.4.1. Діадна згортка

Для перетворення Фур'є фундаментальною є теорема про циклічну згортку. Аналогічна теорема для ПУА існує, але стосується іншої операції — діадної (логічної) згортки

$$y(n) = \sum_{m=0}^{N-1} g\,(n \oplus m)\, f(m), \tag{5.10}$$

де $\oplus$ — порозрядне додавання за модулем 2 (операція «виключне АБО»). Для діадної згортки справедливо: $\mathrm{Y}_{Had} = \sqrt{N}\,\mathrm{G}_{Had} \circ \mathrm{F}_{Had}$, де $\circ$ — поелементний добуток, тобто діадна згортка діагоналізується перетворенням Уолша — Адамара так само, як циклічна згортка — перетворенням Фур'є. Діадний зсув $n \oplus m$ відповідає не зсуву в часі, а «перестановці» відліків, і тому для опису фізичних лінійних систем діадна згортка безпосередньо не придатна.

5.4.2. Циклічна згортка в базисі Адамара: блочно-діагональна структура

Фізичні системи відображення описуються звичайною (лінійною чи циклічною) згорткою з циркулянтною матрицею G. Що відбувається з такою матрицею при переході в базис Адамара? Згідно з (4.10), у новому базисі оператор має вигляд

$$\mathrm{G}_{1Had} = \mathrm{Had}_{\mathrm{н}}\,\mathrm{G}\,\mathrm{Had}_{\mathrm{н}}^{T}. \tag{5.11}$$

Як показано в [2, 4, 7], матриця G_{1Had} при природному впорядкуванні функцій Адамара є блочно-діагональною: її головна діагональ складається з квадратних блоків, розміри яких зростають зі збільшенням номерів рядків — $1,1,2,4,8,\ldots,N/2$. Порівняння структури оператора згортки в базисах різних перетворень наведено на рис. 5.3 (чорним позначено ненульові елементи). У базисі Фур'є матриця діагональна; у базисі Адамара — блочно-діагональна; при секвентному впорядкуванні (за Уолшем) ті самі ненульові елементи розподіляються по матриці за складнішим, але регулярним законом; для перетворень Хаара та косинусного структура ще менш розріджена.

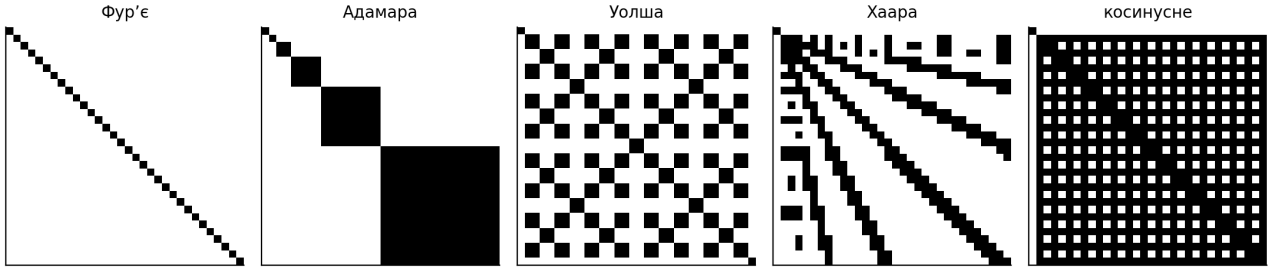


Рис. 5.3. Структура матриці циклічної згортки в базисах різних перетворень (N = 32)

Елементи перших блоків мають простий вигляд. Перші два діагональні елементи — це сума всіх відліків імпульсної характеристики та їх знакозмінна сума:

$$g_{H1,00} = g_0 + g_1 + g_2 + \cdots + g_{N-1}, \qquad g_{H1,11} = g_0 - g_1 + g_2 - g_3 + \cdots, \tag{5.12}$$

а блок другого порядку (рядки і стовпці 2, 3) має однакові діагональні елементи і протилежні за знаком позадіагональні [2]:

$$g_{H1,22} = g_{H1,33} = g_0 - g_2 + g_4 - g_6 + \cdots, \quad g_{H1,23} = -g_{H1,32} = g_1 - g_3 + g_5 - g_7 + \cdots \tag{5.13}$$

Наступний блок має четвертий порядок (рядки і стовпці 4–7), далі — восьмий (8–15) тощо. Блочно-діагональна структура є прямим аналогом діагоналізації в базисі Фур'є, хоча й менш «досконалим»: найбільший блок має порядок $N/2$, і обернення такої матриці в загальному випадку не простіше за обернення в натуральних координатах. Проте завдяки жорсткій регулярній структурі блоків (кожний блок порядку M складається з чотирьох підблоків порядку $M/2$, з яких два однакові, а два відрізняються знаком) розроблено ефективні символьні алгоритми формування та обернення цих матриць, які обчислюють лише ненульові елементи і не накопичують похибок округлення [2, 7, 8, 11]. Ці алгоритми та їх застосування до реставрації образів детально розглядаються в розділі 9.

5.4.3. Двовимірне перетворення Уолша — Адамара

Двовимірне ПУА зображення f розміром $N \times N$ обчислюється, як і двовимірне ДПФ, за сепарабельною схемою:

$$\mathrm{F}_{Had} = \mathrm{Had}_{\mathrm{н}}\, \mathrm{f}\, \mathrm{Had}_{\mathrm{н}}^{T}. \tag{5.14}$$

Базисними зображеннями є зовнішні добутки пар функцій Уолша — «шахові» візерунки з прямокутних клітинок, що набувають значень ± 1 (рис. 5.4, а). Двовимірне ПУА використовувалося в ранніх системах стиснення телевізійних сигналів, у пристроях обробки зображень на основі простих апаратних засобів, а також як джерело ознак у розпізнаванні символів.

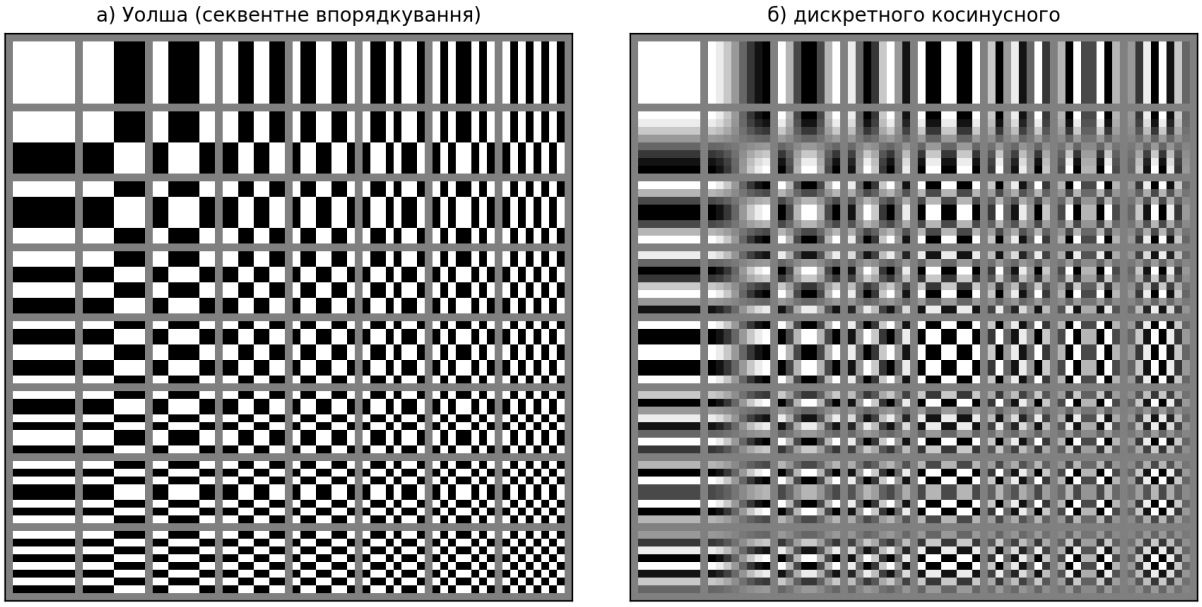


Рис. 5.4. Базисні зображення 8×8 двовимірних перетворень Уолша (а) і косинусного (б)

5.5. Перетворення Хаара

Функції Хаара, запропоновані А. Хааром у 1910 р. [22, 25], також є прямокутними, але, на відміну від функцій Уолша, відмінні від нуля лише на частині інтервалу визначення. Для $N = 2^p$ система складається з постійної функції та функцій, що нумеруються двома індексами — масштабом q і зсувом r:

$$\mathrm{har}(0,t) = 1; \qquad \mathrm{har}(q,r,t) = \begin{cases} 2^{q/2}, & \dfrac{r}{2^q} \le t < \dfrac{r+1/2}{2^q}, \\ -2^{q/2}, & \dfrac{r+1/2}{2^q} \le t < \dfrac{r+1}{2^q}, \\ 0, & \text{інакше,} \end{cases} \tag{5.15}$$

де $q = 0,1,\dots,p-1$; $r = 0,1,\dots,2^q - 1$. Функції Хаара для $N = 8$ показано на рис. 5.5. Матриця перетворення Хаара для $N = 4$ (з нормуванням):

$$\mathrm{Har}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}. \tag{5.16}$$

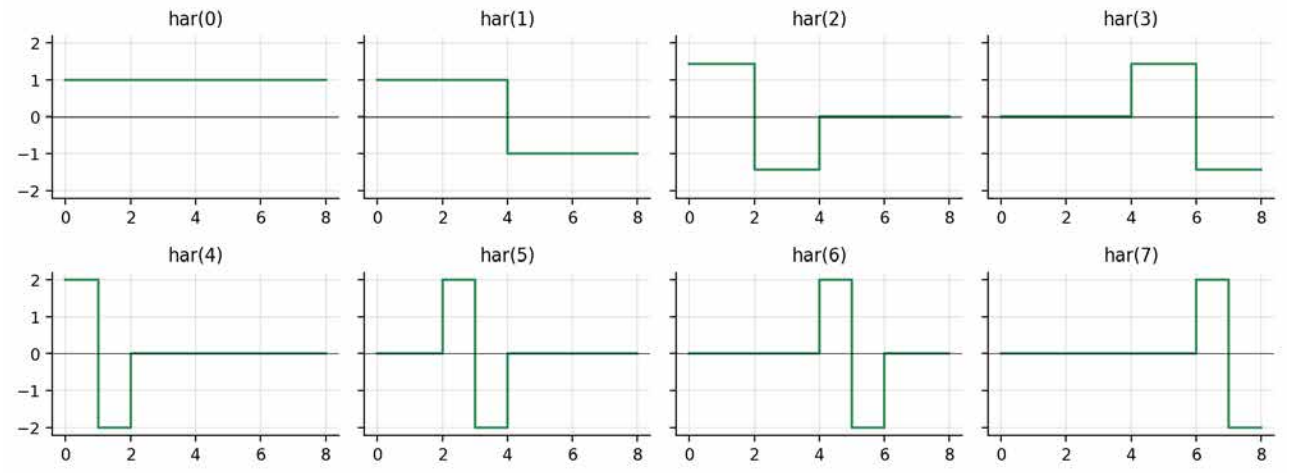


Рис. 5.5. Функції Хаара для N = 8

Ключова відмінність перетворення Хаара від Фур'є та Уолша полягає в локальності базисних функцій: функції з великими значеннями q «бачать» лише малу ділянку образу. Тому трансформанти Хаара несуть інформацію одночасно про масштаб (просторову частоту) і положення деталі. Різкий перепад яскравості в образі впливає лише на ті коефіцієнти Хаара, носії яких його містять, — по одному на кожному масштабі, тобто лише на $\log_2 N$ коефіцієнтів, тоді як у перетворенні Фур'є перепад «розмазується» по всіх коефіцієнтах. Перетворення Хаара — найпростіший приклад дискретного вейвлет-перетворення (розділ 6) [24, 25].

Швидкий алгоритм перетворення Хаара потребує лише близько $2N$ операцій: на кожному кроці послідовність ділиться на пари, для кожної пари обчислюються нормовані сума і різниця, різниці зберігаються як коефіцієнти поточного масштабу, а суми передаються на наступний крок. Це найшвидше з розглянутих перетворень.

5.6. Дискретне косинусне перетворення

5.6.1. Від косинусного ряду до дискретного перетворення

Косинусне перетворення, запропоноване в 1974 р. Н. Ахмедом, Т. Натараджаном і К. Рао [22], стало найпоширенішим перетворенням для стиснення зображень (стандарти JPEG, MPEG, H.264 використовують його або його цілочисельні апроксимації). Широкого застосування воно набуло і в аналізі біосигналів, що навіть призвело до створення спеціалізованих косинусних процесорів [1].

Ідея косинусного перетворення ґрунтується на уявному парному (дзеркальному) продовженні сигналу. ДПФ неявно продовжує сигнал періодично, і якщо значення на початку і в кінці інтервалу відрізняються, на межах періоду виникають розриви (рис. 5.6, а), які породжують розтікання спектра і повільне спадання коефіцієнтів. Якщо ж сигнал спочатку дзеркально відобразити, тобто утворити послідовність подвійної довжини $[x(0), \ldots, x(N-1), x(N-1), \ldots, x(0)]$, а потім продовжити періодично, розриви зникають (рис.

5.6, б). Спектр Фур'є такої парної послідовності дійсний і містить лише косинуси; він і дає коефіцієнти косинусного перетворення.

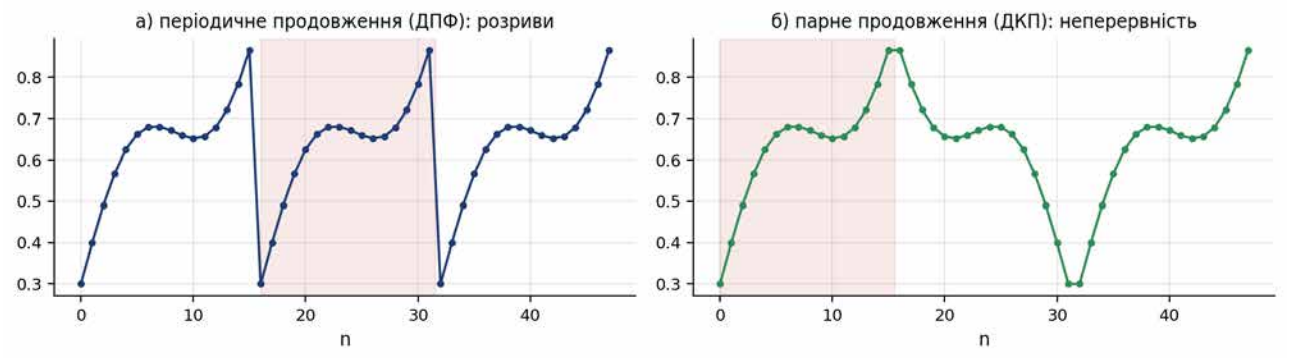


Рис. 5.6. Періодичне (а) та парне (б) продовження сигналу

Розглянемо побудову матричного оператора докладніше, дотримуючись [1]. Неперервні косинусні трансформанти $Cs(k,t) = \cos(\pi kt/T)$ на інтервалі $[0,T]$ мають цілу кількість півперіодів. Якщо взяти дискретні відліки в точках $t_n = nT/N$, $n = 0,\dots,N-1$ (як для ДПФ чи ДПУА), то відліки функції $Cs(1,n)$ не матимуть нульового середнього: відлік у точці $n = 0$ нічим не компенсується, тоді як усі інші компенсуються попарно. Отже, трансформанта $Cs(1,n)$ не буде ортогональною до $Cs(0,n)$, і умова ортогональності порушуватиметься також для інших пар (крім парних). Для забезпечення ортогональності відліки слід зсунути на половину кроку дискретизації — взяти їх у точках $t_n = (n+1/2)T/N$. Так отримуємо найуживанішу форму — ДКП-II:

$$X(k) = c_k \sum_{n=0}^{N-1} x\,(n)\cos\frac{\pi(2n+1)k}{2N}, \qquad k = 0,1,\dots,N-1, \tag{5.17}$$

де нормувальні множники $c_0 = \sqrt{1/N}$, $c_k = \sqrt{2/N}$ при $k > 0$. Обернене перетворення (ДКП-III):

$$x(n) = \sum_{k=0}^{N-1} c_k\; X(k)\cos\frac{\pi(2n+1)k}{2N}. \tag{5.18}$$

Множники c_k — результат нормування: ненормовані рядки матриці косинусного перетворення мають різні енергії (N для нульового рядка і $N/2$ для решти), тому, як зазначалося в розділі 4, для отримання ортогональної матриці кожний рядок ділять на корінь з його енергії [1]. Після нормування матриця ДКП ортогональна, і обернений оператор дорівнює транспонованому. Існують ще сім варіантів ДКП (I–VIII), що відрізняються способом парного продовження і положенням відліків; вони використовуються в спеціальних застосуваннях.

5.6.2. Двовимірне ДКП і стиснення зображень

Двовимірне ДКП сепарабельне і обчислюється за формулою, аналогічною (5.14). Базисні зображення ДКП 8×8 наведено на рис. 5.4, б: це плавні косинусні «хвилі» зростаючої частоти по горизонталі і вертикалі.

Схема стиснення JPEG у спрощеному вигляді така: зображення ділиться на блоки 8×8; для кожного блоку обчислюється двовимірне ДКП; коефіцієнти квантуються з кроком, що зростає з частотою (високочастотні деталі менш помітні для ока); квантовані коефіцієнти впорядковуються «зигзагом» від низьких частот до високих (при цьому утворюються довгі серії нулів) і кодуються ентропійним кодом. Ефективність схеми ґрунтується на енергетичній компактності ДКП, що розглядається далі.

5.7. Енергетична компактність і перетворення Карунена — Лоева

5.7.1. Перетворення Карунена — Лоева

Яке ортогональне перетворення є найкращим для стиснення або для зменшення розмірності ознак? Нехай образи є реалізаціями випадкового вектора x з нульовим середнім і коваріаційною матрицею $\mathrm{R} = \mathrm{M}\{\mathrm{x}\mathrm{x}^T\}$. Потрібно знайти ортонормований базис, у якому апроксимація образу першими K трансформантами дає мінімальну середньоквадратичну похибку. Розв'язком є перетворення Карунена — Лоева (КЛП), базисними векторами якого є власні вектори коваріаційної матриці:

$$\mathrm{R}\,\varphi_k = \lambda_k\,\varphi_k, \qquad \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1} \geq 0. \tag{5.19}$$

Коефіцієнти КЛП некорельовані, а їх дисперсії дорівнюють власним числам λ_k; середньоквадратична похибка апроксимації першими K коефіцієнтами дорівнює сумі відкинутих власних чисел $\sum_{k=K}^{N-1} \lambda_k$. Перетворення Карунена — Лоева оптимальне за двома критеріями одночасно: воно повністю декорелює коефіцієнти і максимально концентрує енергію. У статистиці та машинному навчанні те саме перетворення відоме як метод головних компонент (PCA) [49]; його застосування для зменшення розмірності ознак розглядається в розділі 11.

Недоліки КЛП як перетворення для оброблення сигналів: базис залежить від статистики конкретного класу образів і має обчислюватися заново для кожного класу; для його обчислення потрібна оцінка коваріаційної матриці, що вимагає великої навчальної вибірки (див. обговорення прокляття розмірності в розділі 1); швидкого алгоритму в загальному випадку не існує [1, 22].

5.7.2. Чому ДКП близьке до оптимального

Для багатьох природних сигналів і зображень добре підходить модель марковського процесу першого порядку, коваріаційна функція якого спадає експоненційно: $R(n,m) = \sigma^2 \rho^{|n-m|}$, де ρ — коефіцієнт кореляції сусідніх відліків (для зображень зазвичай $\rho \approx 0{,}9 \ldots 0{,}98$). Відомо, що при $\rho \to 1$ власні вектори коваріаційної матриці такого процесу асимптотично збігаються з базисними векторами ДКП-II [22]. На рис. 5.7 наведено порівняння базисних векторів КЛП для $\rho = 0{,}95$, $N = 8$ і базисних векторів ДКП: вони практично ідентичні. Саме тому ДКП, маючи фіксований базис і швидкий алгоритм, досягає майже оптимальної компактності для широкого класу зображень.

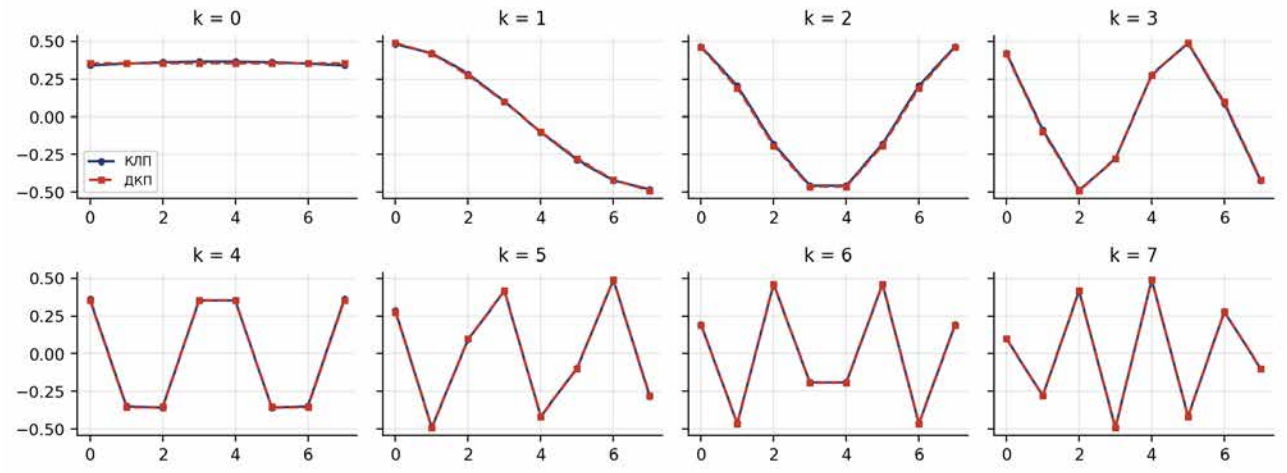


Рис. 5.7. Базисні вектори КЛП (марковський процес, ρ = 0,95) та ДКП для N = 8

5.7.3. Порівняльний експеримент

Порівняємо енергетичну компактність розглянутих перетворень на реальному зображенні. Зображення розбито на блоки 8 × 8; для кожного блоку обчислено двовимірні перетворення Фур'є, косинусне, Уолша, Хаара та (сепарабельне) Карунена — Лоева, базис якого оцінено за рядками самого зображення. На рис. 5.8, а показано середню частку енергії змінної складової блоку, що припадає на K найбільших коефіцієнтів. Найкращу компактність мають КЛП і ДКП, далі йдуть перетворення Хаара та Уолша, найгіршу — ДПФ (частково через те, що коефіцієнти ДПФ комплексні і кожна гармоніка «займає» два коефіцієнти, частково через розриви на межах блоку). Рис. 5.8, б показує PSNR зображення, відновленого після збереження лише 6 найбільших коефіцієнтів з 64 у кожному блоці; на рис. 5.9 наведено фрагменти відновлених зображень. Прямокутні базисні функції Уолша і Хаара створюють характерні «сходинкові» артефакти, тоді як ДКП дає плавніше відновлення.

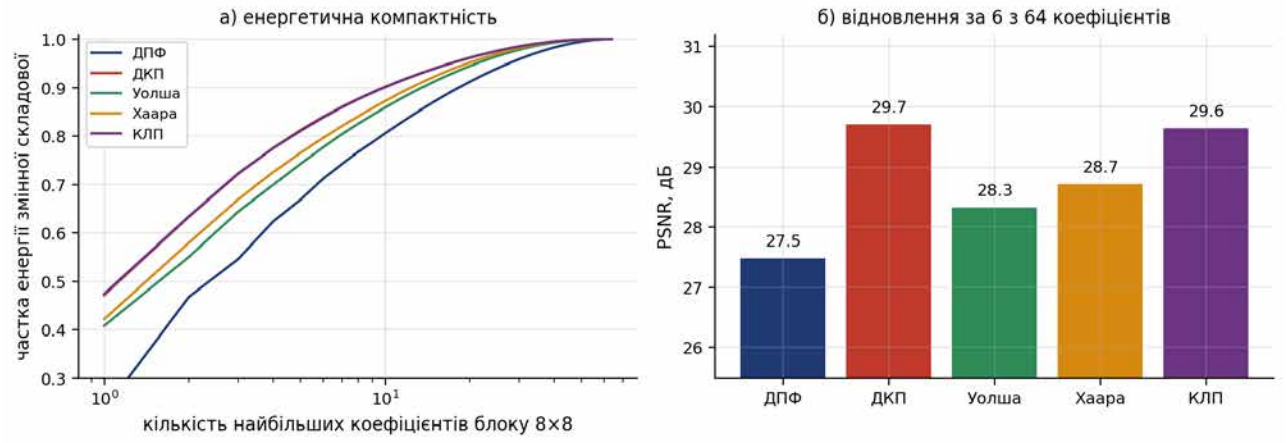


Рис. 5.8. Енергетична компактність (а) та якість відновлення зображення за 6 з 64 коефіцієнтів блоку 8×8 (б)

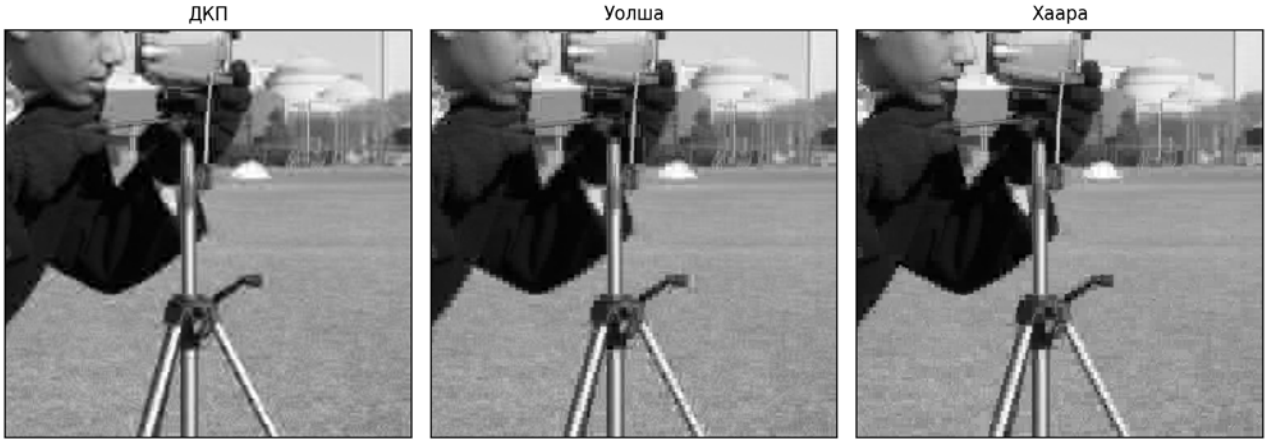


Рис. 5.9. Фрагменти зображення, відновленого за 6 найбільших коефіцієнтів у блоках 8×8

Слід звернути увагу на те, що висока енергетична компактність гладких базисів (ДКП) і локальна пристосованість прямокутних базисів до розривів (Хаар, Уолш) — це дві різні якості, корисні в різних задачах. Для стиснення природних зображень важливіша перша; для точного відтворення різких меж графоелементів (наприклад, у реставрації або при аналізі штрихових зображень) — друга.

5.8. Теорема про диференціювання оригіналу для дискретних ортогональних перетворень

5.8.1. Загальне формулювання

Теорема про диференціювання оригіналу (4.13) разом із тим фактом, що лінійна система відповідає на гармоніку гармонікою тієї самої частоти, є основою спектральних методів аналізу лінійних систем: диференціальні рівняння перетворюються на алгебраїчні [2, 12]. Наявність великої кількості відмінних від Фур'є ортогональних перетворень робить актуальним питання: чи існує аналог цієї теореми для них? Відповідь дає підхід, розроблений у працях [2, 5, 12] для дискретних перетворень.

У дискретних методах похідна замінюється скінченною різницею. Для циклічної (періодичної) послідовності різниці «назад» $x(n) - x(n-1)$ відповідає множення вектора відліків на циркулянтну матрицю

$$\Delta = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ -1 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \qquad \mathrm{x}_d = \Delta\,\mathrm{x}, \tag{5.20}$$

(крок дискретизації нормовано до одиниці). Нехай W — нормований матричний оператор довільного дискретного ортогонального перетворення формату $N = 2^p$, $\mathrm{X} = \mathrm{W}\mathrm{x}$. Тоді спектр похідної

$$\mathrm{X}_d = \mathrm{W}\Delta\mathrm{x} = \mathrm{W}\Delta\mathrm{W}^H\,\mathrm{X} = \mathrm{A}_W\,\mathrm{X}. \tag{5.21}$$

Отже, теорему про диференціювання оригіналу для довільного дискретного ортогонального перетворення можна сформулювати так: якщо $\mathrm{x} \leftrightarrow \mathrm{X}$, то $\mathrm{x}_d \leftrightarrow \mathrm{A}_W\mathrm{X}$, де $\mathrm{A}_W = \mathrm{W}\Delta\mathrm{W}^H$ [2]. Матриця Δ у натуральних координатах однакова для всіх перетворень, тоді як матриця A_W різна для різних перетворень. Власні числа A_W (як і будь-якої матриці, отриманої перетворенням подібності) не залежать від вибору базису, а власні вектори — залежать.

5.8.2. Розв'язання лінійних диференціальних рівнянь в області трансформант

Розглянемо лінійне диференціальне рівняння зі сталими коефіцієнтами, що описує лінійну систему:

$$a_m \frac{d^m y}{dt^m} + \cdots + a_1 \frac{dy}{dt} + a_0 y = b_n \frac{d^n x}{dt^n} + \cdots + b_1 \frac{dx}{dt} + b_0 x. \qquad (5.22)$$

Замінюючи похідні i-го порядку множенням спектрів на A_W^i, отримуємо систему лінійних алгебраїчних рівнянь в області трансформант:

$$(a_m \mathrm{A}_W^m + \cdots + a_1 \mathrm{A}_W + a_0 \mathrm{E})\mathrm{Y} = (b_n \mathrm{A}_W^n + \cdots + b_1 \mathrm{A}_W + b_0 \mathrm{E})\mathrm{X}, \qquad (5.23)$$

або, позначивши суми в дужках A_Σ і B_Σ,

$$\mathrm{Y} = \mathrm{A}_\Sigma^{-1}\mathrm{B}_\Sigma\,\mathrm{X} = \mathrm{D}_\Sigma\,\mathrm{X}. \qquad (5.24)$$

Структура матриці D_Σ визначається структурою A_W, тобто вибором перетворення [2]:

- для перетворення Фур'є D_Σ — діагональна, і (5.24) зводиться до поелементного множення спектра дії на частотну характеристику системи;
- для перетворення Адамара (природне впорядкування) — блочно-діагональна з тією самою структурою блоків, що й на рис. 5.3;
- для інших перетворень матриця має іншу, але також певну регулярну структуру.

У натуральних координатах аналогічний вираз $\mathrm{y} = \mathrm{D}_{\Sigma t}\mathrm{x}$ містить циркулянтну матрицю $\mathrm{D}_{\Sigma t}$ — матричний оператор дискретної згортки, у кожному рядку якого записано імпульсну характеристику системи в зворотному порядку, зсунуту на один крок праворуч при переході до наступного рядка. Таким чином, аналіз лінійної системи в довільному ортогональному базисі зводиться до обчислення спектра дії, множення на матрицю системи в цьому базисі та оберненого перетворення. Детально алгоритми аналізу лінійних систем в області трансформант Уолша — Адамара, у тому числі з розкладом оператора диференціювання на власні вектори і власні числа, викладено в [1, 12].

5.8.3. Власні числа різницевого оператора і перетворення RTF

Знайдемо власні числа матриці Δ. Оскільки вона циркулянтна, її власні вектори — дискретні гармоніки, а власні числа — ДПФ першого стовпця $[1, -1, 0, \ldots, 0]^T$:

$$\beta_k = 1 - e^{-j\frac{2\pi}{N}k} = 2j\sin\left(\frac{\pi}{N}k\right)e^{-j\frac{\pi}{N}k}, \qquad k = 0, 1, \ldots, N-1. \qquad (5.25)$$

Числа β_k лежать на колі одиничного радіуса з центром у точці (1,0) комплексної площини (рис. 5.10): $\beta_0 = 0, \beta_{N/2} = 2, \beta_{N/4} = 1 + j$ тощо [2, 5]. Для порівняння, для неперервного диференціювання в базисі Фур'є відповідні множники $j\omega_k = j2\pi k/N$ лежать на уявній осі. При малих k (низьких частотах) $\beta_k \approx j\omega_k$, оскільки $1 - e^{-j\omega} \approx j\omega$ при $\omega \ll 1$; з наближенням до частоти Найквіста розбіжність зростає.

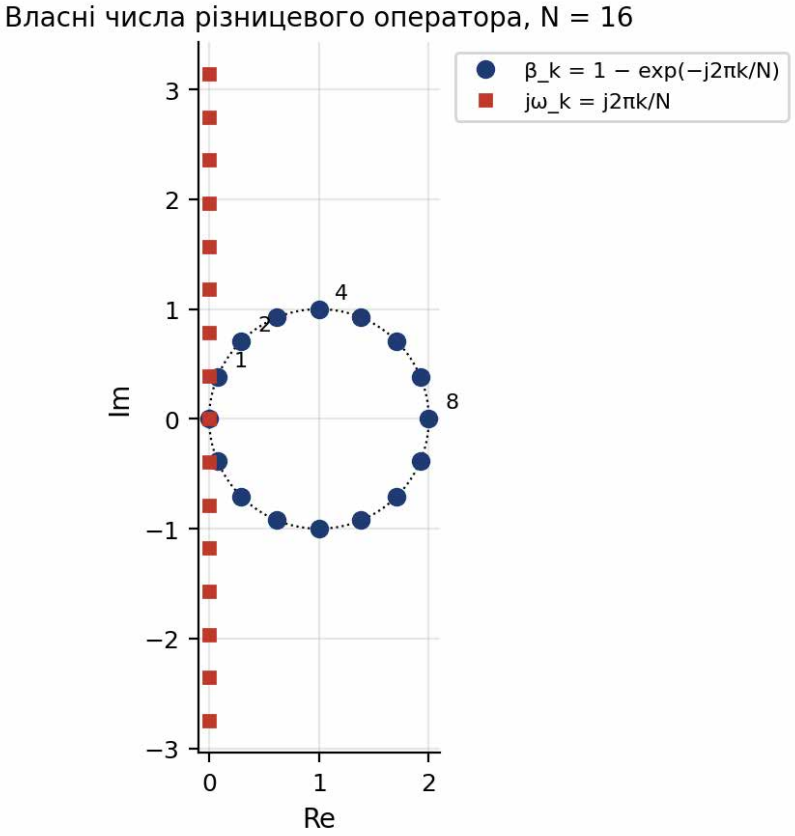


Рис. 5.10. Власні числа різницевого оператора β_k та частоти jω_k класичного ДПФ

Перетворення, у якому роль «частот» відіграють власні числа β_k різницевого оператора, отримало назву перетворення RTF [2, 5]. Його теорема про диференціювання оригіналу має ту саму просту форму, що й для неперервного перетворення Фур'є, але точно відповідає різницевій схемі:

$$\mathrm{X}_d = \mathrm{diag}\{\beta_0, \dots, \beta_{N-1}\}\,\mathrm{X}. \qquad (5.26)$$

Відмінність між β_k і $j\omega_k$ пояснюється тим, що в класичному перетворенні Фур'є похідна обчислюється аналітично в точці, а в перетворенні RTF — як різниця сусідніх відліків, поділена на крок дискретизації. Наслідки цієї відмінності добре видно на рис. 5.11. Якщо обчислювати похідну сигналу з різкими перепадами множенням його ДПФ на $j\omega_k$ (тобто «точним» спектральним диференціюванням тригонометричного інтерполянта), результат містить значні паразитні осциляції — прояв ефекту Гіббса через обмежену кількість трансформант. Множення на β_k дає точну різницеву похідну без будь-яких осциляцій.

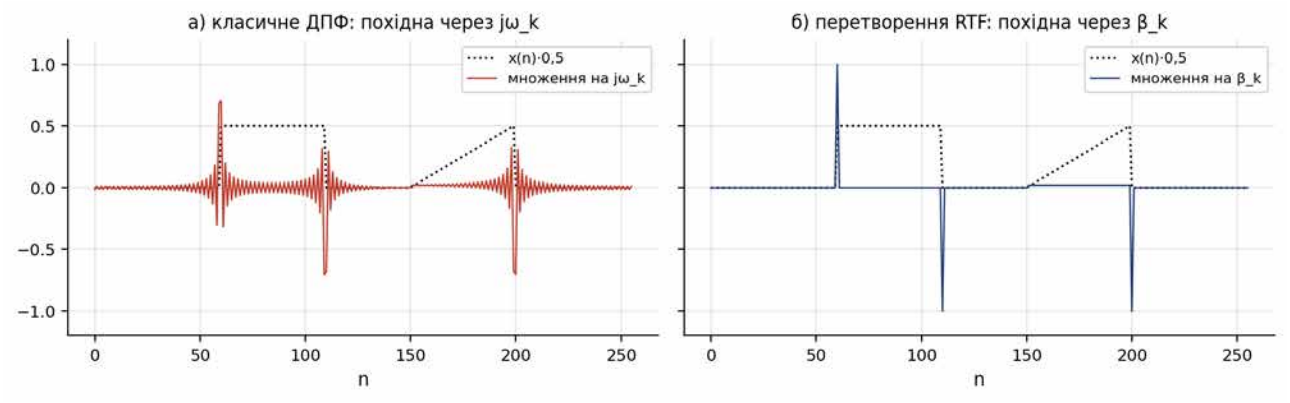


Рис. 5.11. Диференціювання сигналу в базисі ДПФ (а) і в базисі RTF (б)

Платою за відсутність осциляцій є додаткове згасання на високих частотах: модуль $|\beta_k| = 2|\sin(\pi k/N)|$ менший за $|\omega_k|$, тобто перехід від Фур'є до RTF еквівалентний додатковій низькочастотній фільтрації. Ця властивість корисна для придушення осциляцій, але при надто великому кроці дискретизації може спотворювати швидкі коливальні процеси. Тому дискретизацію слід вибирати так, щоб основна частина спектра сигналу потрапляла в область відносно низьких частот [2, 5]. Застосування перетворення RTF для реставрації образів розглянуто в розділі 9.

5.9. Інші перетворення з дійсним ядром

Крім розглянутих, у цифровій обробці сигналів відомі й інші ортогональні перетворення з дійсним ядром [22]:

- перетворення Хартлі з ядром $\mathrm{cas}(x) = \cos x + \sin x$, яке для дійсних сигналів містить ту саму інформацію, що й ДПФ, але не потребує комплексної арифметики;
- похиле (slant) перетворення, базис якого містить лінійно зростаючу функцію, що дозволяє ефективно подавати ділянки з лінійною зміною яскравості;
- дискретне синусне перетворення, близьке до оптимального для процесів з малою кореляцією;
- перетворення Габора та інші перетворення з локалізованими базисними функціями, що ведуть до вейвлет-аналізу (розділ 6).

Особливе місце посідають перетворення, базис яких будується під конкретний еталонний сигнал так, щоб одна з базисних функцій збігалася з еталоном. Такі нормальні перетворення [1] безпосередньо пов'язані з розпізнаванням образів і розглядаються в розділі 10.

5.10. Порівняльна характеристика перетворень

Основні характеристики розглянутих перетворень зведено в табл. 5.1 [1, 22, 23].

Таблиця 5.1

Порівняльна характеристика дискретних ортогональних перетворень

Перетворення	Ядро	Складність швидкого алгоритму	Структура оператора циклічної згортки	Енергетична компактність	Локальність базису
Фур'є (ДПФ)	комплексне	$(N/2)\log_2 N$ компл. множень	діагональна	середня	відсутня
Уолша — Адамара	± 1	$N\log_2 N$ додавань	блочно-діагональна	середня	відсутня
Хаара	$0, \pm 2^{q/2}$	$\approx 2N$	розріджена	середня	висока
Косинусне (ДКП)	дійсне	$O(N\log_2 N)$	щільна (для циркулянта)	висока	відсутня
Карунена — Лоева	дійсне, залежить від даних	N^2	залежить від даних	оптимальна	залежить від даних

Вибір перетворення визначається задачею. Для аналізу лінійних систем і реставрації ключовою є структура оператора згортки (найкраща — у ДПФ, близька — у ПУА). Для стиснення і формування ознак — енергетична компактність (ДКП, КЛП). Для простих апаратних реалізацій — відсутність множень (ПУА). Для аналізу локальних особливостей і нестаціонарних образів — локальність базису (Хаар, вейвлети) [22, 24].

5.11. Програмна реалізація

Нижче наведено функції для побудови матриць розглянутих перетворень та перевірки їх ортогональності, а також експеримент з енергетичної компактності для двовимірних блоків зображення.

```
import numpy as np
from scipy.linalg import hadamard
from scipy.fft import dct

def had(N):                    # природне впорядкування, нормоване
    return hadamard(N) / np.sqrt(N)

def walsh(N):                  # секвентне впорядкування
    H = had(N)
    seq = [np.sum(np.abs(np.diff(np.sign(r)))) // 2 for r in H]
    return H[np.argsort(seq)]

def haar(N):
    H = np.zeros((N, N)); H[0] = 1 / np.sqrt(N); k, q = 1, 0
```

```
    while k < N:
        L = N // 2**q
        for r in range(2**q):
            h = np.zeros(N); h[r*L:r*L + L//2] = 1; h[r*L + L//2:(r+1)*L] = -1
            H[k] = h / np.sqrt(L); k += 1
        q += 1
    return H

def dctm(N):                    # ДКП-II, ортонормоване
    return dct(np.eye(N), norm='ortho', axis=0)

for name, W in [('had', had(8)), ('walsh', walsh(8)),
                ('haar', haar(8)), ('dct', dctm(8))]:
    print(name, np.allclose(W @ W.T, np.eye(8)))

# двовимірне перетворення блоку 8x8 і частка енергії 6 найбільших
коефіцієнтів
from skimage import data, img_as_float
blk = img_as_float(data.camera())[200:208, 200:208]
for name, W in [('dct', dctm(8)), ('walsh', walsh(8)), ('haar', haar(8))]:
    C = W @ (blk - blk.mean()) @ W.T
    e = np.sort(C.ravel()**2)[::-1]
    print(name, e[:6].sum() / e.sum())
```

5.12. Приклади розв'язання задач

Задача 5.1. Для вектора $\mathrm{x} = [1,3,5,7]^T$ обчисліть нормовані перетворення Уолша — Адамара, Хаара і косинусне. Порівняйте розподіл енергії.

Розв'язання. Нормоване ПУА: $\mathrm{X}_{Had} = \frac{1}{2}\mathrm{H}_4\mathrm{x} = \frac{1}{2}[16, -4, -8,0]^T = [8, -2, -4,0]^T$.

Перетворення Хаара за матрицею (5.16): $\mathrm{X}_{Har} = \left[8,\ -4,\ -\sqrt{2},\ -\sqrt{2}\right]^T \approx [8;\ -4;\ -1{,}414;\ -1{,}414]^T$.

ДКП-II за (5.17): $\mathrm{X}_{DCT} \approx [8;\ -4{,}461;\ 0;\ -0{,}317]^T$.

Енергія вектора $\| \mathrm{x} \|^2 = 84$ однакова в усіх трьох базисах (перевірте: $64 + 4 + 16 = 84$; $64 + 16 + 2 + 2 = 84$; $64 + 19{,}9 + 0 + 0{,}1 = 84$). Енергія змінної складової (20) зосереджена: у ДКП — на 99,5 % в одному коефіцієнті; у Хаара — на 80 %; в Адамара — на 80 % (але в коефіцієнті з номером 2 природного впорядкування, що відповідає секвенті 1). Для лінійно зростаючого сигналу ДКП найкомпактніше, оскільки його перша базисна функція — півперіод косинуса — найкраще наближає лінійну зміну.

Задача 5.2. Імпульсна характеристика має відліки $g = [0{,}5;\ 0{,}3;\ 0{,}2]$, формат $N = 8$. Знайдіть елементи перших трьох блоків матриці G_{1Had}.

Розв'язання. За (5.12), (5.13): $g_{H1,00} = 0{,}5 + 0{,}3 + 0{,}2 = 1{,}0$; $g_{H1,11} = 0{,}5 - 0{,}3 + 0{,}2 = 0{,}4$; $g_{H1,22} = g_{H1,33} = g_0 - g_2 = 0{,}3$; $g_{H1,23} = -g_{H1,32} = g_1 - g_3 =$

0,3. Обчислення добутку $\mathrm{Had_н G Had_н^T}$ дає ті самі значення; усі елементи поза блоками дорівнюють нулю.

Задача 5.3. Знайдіть власні частоти перетворення RTF для $N = 8$.

Розв'язання. За (5.25): $\beta_0 = 0$; $\beta_1 = 1 - e^{-j\pi/4} \approx 0{,}293 + 0{,}707j$; $\beta_2 = 1 + j$; $\beta_3 \approx 1{,}707 + 0{,}707j$; $\beta_4 = 2$; $\beta_5 = \beta_3^*$, $\beta_6 = \beta_2^*$, $\beta_7 = \beta_1^*$. Усі вони лежать на колі $|\beta - 1| = 1$. Для порівняння, $j\omega_1 = j\pi/4 \approx 0{,}785j$: модуль $|\beta_1| \approx 0{,}765$ дещо менший, що відповідає додатковому згасанню перетворення RTF.

Висновки до розділу 5

1. Перетворення з дійсним ядром (Уолша — Адамара, Хаара, косинусне) не потребують комплексної арифметики, мають швидкі алгоритми і в багатьох задачах аналізу, стиснення та розпізнавання сигналів є зручнішими за перетворення Фур'є.
2. Матриця Адамара будується рекурсивно за допомогою кронекерового добутку; вона симетрична, ортогональна з точністю до множника і інволютивна. Розрізняють природне (за Адамаром), секвентне (за Уолшем) і діадне (за Пейлі) впорядкування функцій Уолша; для них існують прості алгоритми перенумерації. Швидке перетворення Уолша — Адамара потребує лише $N\log_2 N$ додавань.
3. ПУА діагоналізує діадну згортку, а оператор звичайної циклічної згортки в базисі Адамара набуває блочно-діагональної форми з блоками розмірів $1,1,2,4,\ldots,N/2$ і жорсткою регулярною структурою. Ця властивість лежить в основі методів реставрації в базисі Адамара (розділ 9).
4. Функції Хаара локалізовані і несуть інформацію одночасно про масштаб і положення деталей; перетворення Хаара — найпростіше вейвлет-перетворення.
5. ДКП ґрунтується на парному продовженні сигналу, що усуває розриви на межах інтервалу. Для забезпечення ортогональності відліки беруться зі зсувом на половину кроку, а рядки матриці нормуються. ДКП асимптотично збігається з оптимальним перетворенням Карунена — Лоева для марковських процесів з високою кореляцією і тому має майже оптимальну енергетичну компактність.
6. Теорему про диференціювання оригіналу можна сформулювати для будь-якого дискретного ортогонального перетворення: диференціюванню (різниці) відповідає множення спектра на матрицю $\mathrm{W\Delta W}^H$, діагональну для Фур'є і блочно-діагональну для Адамара.
7. Власні числа різницевого оператора $\beta_k = 1 - e^{-j2\pi k/N}$ визначають перетворення RTF, у якому диференціювання не породжує осциляцій Гіббса, але супроводжується додатковим згасанням на високих частотах.

Питання для обговорення та самоперевірки

1. Які недоліки перетворення Фур'є зумовлюють інтерес до перетворень з дійсним ядром?
2. Як утворюються функції Уолша з функцій Радемахера? Що таке секвента?
3. Побудуйте матрицю Адамара H_8 за допомогою кронекерового добутку.
4. Перелічіть властивості матриці Адамара. Чому пряме й обернене ПУА мають однаковий вигляд?
5. Чим відрізняються впорядкування функцій Уолша за Адамаром, за Уолшем і за Пейлі? Для яких задач зручне кожне з них?
6. Скільки операцій потребує швидке ПУА? Чому в ньому немає множень?
7. Що таке діадна згортка? Чому вона безпосередньо не описує фізичні системи?
8. Яку структуру має матриця циклічної згортки в базисі Адамара? Запишіть елементи перших трьох блоків.
9. Чим функції Хаара принципово відрізняються від функцій Уолша і Фур'є?
10. Поясніть ідею парного продовження сигналу в ДКП. Чому ДКП має менше розтікання спектра, ніж ДПФ?
11. Чому для ортогональності дискретних косинусних трансформант відліки беруть зі зсувом на половину кроку? Навіщо нормувати рядки матриці?
12. У чому полягає оптимальність перетворення Карунена — Лоева? Які його практичні недоліки?
13. Чому ДКП є хорошим наближенням до КЛП для природних зображень?
14. Сформулюйте теорему про диференціювання оригіналу для довільного дискретного ортогонального перетворення.
15. Як розв'язується лінійне диференціальне рівняння в області трансформант? Яку структуру має матриця системи для Фур'є і для Адамара?
16. Що таке перетворення RTF? Де на комплексній площині розташовані його «частоти» β_k?
17. Чому спектральне диференціювання з множенням на $j\omega_k$ дає осциляції, а множення на β_k — ні?
18. За якими критеріями слід обирати ортогональне перетворення для конкретної задачі?

Завдання для самостійної роботи

1. Реалізуйте функцію перенумерації функцій Уолша з природного впорядкування в секвентне через код Грея і двійкову інверсію. Перевірте результат обчисленням кількості змін знака.
2. Реалізуйте швидке ПУА та швидке перетворення Хаара. Перевірте збіг з матричними формами і порівняйте час виконання з ШПФ для $N = 2^{10} \ldots 2^{16}$.
3. Для циркулянтної матриці згортки з $N = 16$ і імпульсною характеристикою з 5 відліків обчисліть $\mathrm{Had}_\mathrm{н}\mathrm{G}\mathrm{Had}_\mathrm{н}^T$. Перевірте формули (5.12), (5.13) і структуру блоків порядку 4 та 8.

4. Порівняйте енергетичну компактність ДПФ, ДКП, ПУА, Хаара і КЛП для одновимірних реалізацій марковського процесу з $\rho = 0{,}5;\ 0{,}9;\ 0{,}99$.
5. Реалізуйте спрощену схему стиснення зображення на основі блочного ДКП 8×8 з рівномірним квантуванням коефіцієнтів. Побудуйте залежність PSNR від частки збережених коефіцієнтів. Повторіть для ПУА і Хаара.
6. Розв'яжіть рівняння $\tau y' + y = x$ (RC-фільтр) для прямокутного імпульсу дії в області трансформант Фур'є (з $j\omega_k$), RTF (з β_k) та Адамара (з матрицею A_W). Порівняйте результати з аналітичним розв'язком.
7. Побудуйте матрицю $\mathrm{A}_W = \mathrm{W}\Delta\mathrm{W}^T$ для ПУА при $N = 16$ і порівняйте її власні числа з β_k (5.25).

РОЗДІЛ 6. ЧАСТОТНО-ЧАСОВИЙ АНАЛІЗ: ВІКОННЕ ПЕРЕТВОРЕННЯ ФУР'Є ТА ВЕЙВЛЕТИ

Навчальні цілі. Після вивчення розділу ви зможете: пояснити, чому глобальні ортогональні перетворення недостатні для аналізу нестаціонарних образів; будувати та тлумачити спектрограми віконного перетворення Фур'є; формулювати принцип невизначеності і вибирати ширину вікна; обчислювати неперервне і дискретне вейвлет-перетворення; реалізовувати кратномасштабний аналіз за алгоритмом Малла; застосовувати вейвлети для придушення шуму, оцінювання його рівня і формування текстурних ознак; використовувати фільтри Габора.

6.1. Стаціонарні та нестаціонарні образи

Перетворення, розглянуті в розділах 4 і 5, є глобальними: кожна трансформанта обчислюється за всіма відліками образу, а кожна базисна функція (крім функцій Хаара) відмінна від нуля на всьому інтервалі. Глобальний спектр відповідає на запитання «які частотні (секвентні) складові присутні в образі», але не відповідає на запитання «де саме» вони присутні. Для стаціонарних сигналів, статистичні властивості яких не змінюються в часі, таке питання і не виникає. Проте більшість образів, що підлягають розпізнаванню, нестаціонарні: мовлення складається з послідовності звуків різного спектрального складу; електрокардіограма містить короткі комплекси на фоні повільних хвиль; зображення складається з однорідних ділянок, контурів і текстур, локалізованих у різних місцях.

Розглянемо простий приклад (рис. 6.1). Перший сигнал — сума двох гармонік з частотами 50 і 120 Гц, що звучать одночасно протягом усього інтервалу. Другий — ті самі гармоніки, але звучать послідовно: спочатку 50 Гц, потім 120 Гц. Сигнали принципово різні, і для системи розпізнавання мовлення вони означали б різні «слова». Проте їх амплітудні спектри майже однакові: два піки на тих самих частотах. Інформація про *послідовність* подій міститься лише у фазовому спектрі в дуже неявній формі.

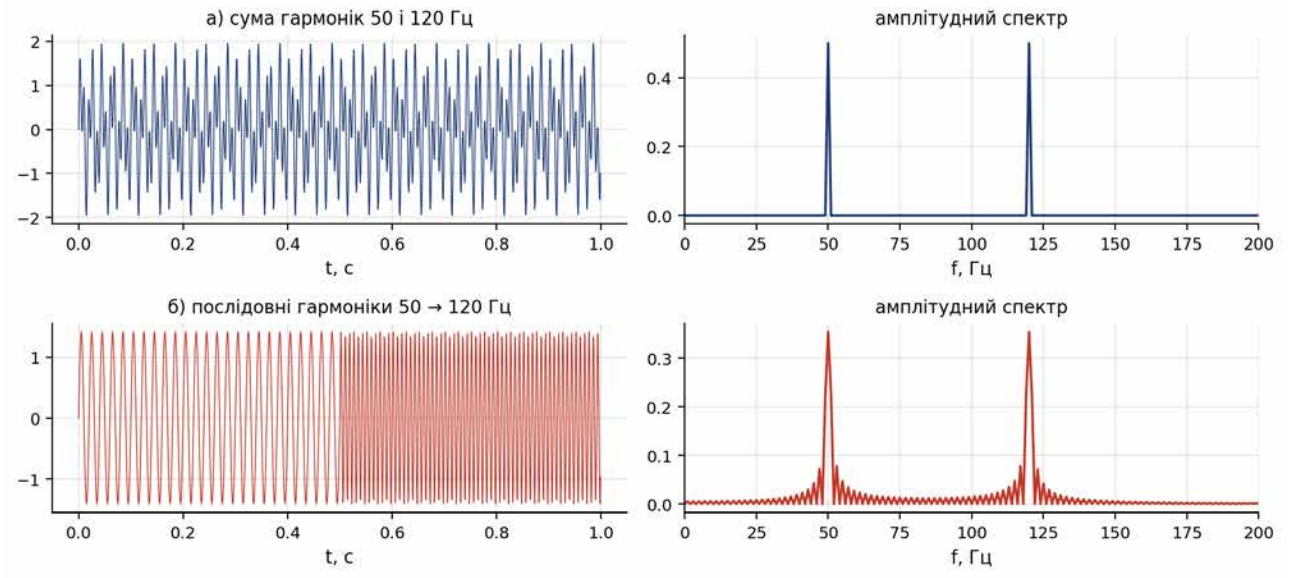


Рис. 6.1. Два різні сигнали з подібними амплітудними спектрами

Для аналізу таких образів потрібні методи частотно-часового (для зображень — просторово-частотного) аналізу, що дозволяють одночасно визначати частотний склад і локалізацію складових [1, 24]. У цьому розділі розглянуто два основні підходи: локальне (віконне) перетворення Фур'є з постійною роздільною здатністю та вейвлет-перетворення зі змінною роздільною здатністю.

6.2. Локальне (віконне) перетворення Фур'є

6.2.1. Визначення

Ідея віконної обробки полягає в тому, щоб розглядати не весь сигнал, а його фрагмент, «вирізаний» вікном $w(t)$, зосередженим поблизу моменту τ, і обчислювати спектр саме цього фрагмента. Переміщуючи вікно вздовж осі часу, отримуємо локальне (віконне, короткочасне) перетворення Фур'є (англ. *Short-Time Fourier Transform*, STFT) [1, 50]:

$$S(\tau, \omega) = \int_{-\infty}^{\infty} s\,(t)\, w(t-\tau)\, e^{-j\omega t}\, dt. \qquad (6.1)$$

Результат є функцією двох змінних — часу τ і частоти ω. Квадрат модуля $|S(\tau, \omega)|^2$, відображений як півтонове зображення на площині «час — частота», називають спектрограмою. У дискретному варіанті сигнал ділять на фрагменти довжиною L відліків (зазвичай з перекриттям), кожний фрагмент множать на вікно (Ханна, Хеммінга, гаусівське) і обчислюють ДПФ.

Вираз (6.1) можна тлумачити двояко. З одного боку, це спектр Фур'є фрагмента сигналу. З іншого — це скалярний добуток сигналу з базисною функцією $w(t-\tau)e^{j\omega t}$, тобто з гармонікою, обмеженою вікном. При гаусівському вікні такі функції називають функціями Габора, а відповідне перетворення — перетворенням Габора [50].

6.2.2. Роздільна здатність і принцип невизначеності

Для заданого вікна $w(t)$ з перетворенням Фур'є $W(j\omega)$ визначимо ефективну тривалість Δt і ефективну ширину смуги $\Delta\omega$ через другі моменти:

$$\Delta t^2 = \frac{\int t^2 |w(t)|^2 dt}{\int |w(t)|^2 dt}, \qquad \Delta\omega^2 = \frac{\int \omega^2 |W(j\omega)|^2 d\omega}{\int |W(j\omega)|^2 d\omega}, \tag{6.2}$$

де знаменники (за теоремою Парсеваля) пропорційні енергії вікна [1]. Дві гармоніки можна розрізнити за частотою, лише якщо їх частоти відрізняються більш ніж на $\Delta\omega$; два імпульси можна розрізнити в часі, лише якщо інтервал між ними перевищує Δt.

Ці величини пов'язані фундаментальною нерівністю Гейзенберга — Габора:

$$\Delta t \cdot \Delta\omega \geq \frac{1}{2}. \tag{6.3}$$

Отже, часову і частотну роздільну здатність неможливо одночасно необмежено поліпшувати: вузьке вікно добре локалізує події в часі, але погано розрізняє частоти, і навпаки. Мінімальну площу $\Delta t \cdot \Delta\omega$ на частотно-часовій площині називають коміркою невизначеності: усередині неї точніше визначити часову і частотну координату принципово неможливо. Рівність у (6.3) досягається лише для гаусівського вікна, перетворення Фур'є якого також є гаусівською функцією.

Вплив вибору вікна ілюструє рис. 6.2. Тестовий сигнал містить лінійно-частотно-модульоване коливання (частота зростає від 20 до 140 Гц), гармоніку 80 Гц і два короткі імпульси. При вузькому вікні (32 відліки) імпульси чітко локалізовані у вигляді вертикальних ліній, але частотні складові розмиті у широкі смуги. При широкому вікні (256 відліків) лінія модуляції частоти та гармоніка видно чітко, але імпульси розмазані в часі і майже зникли.

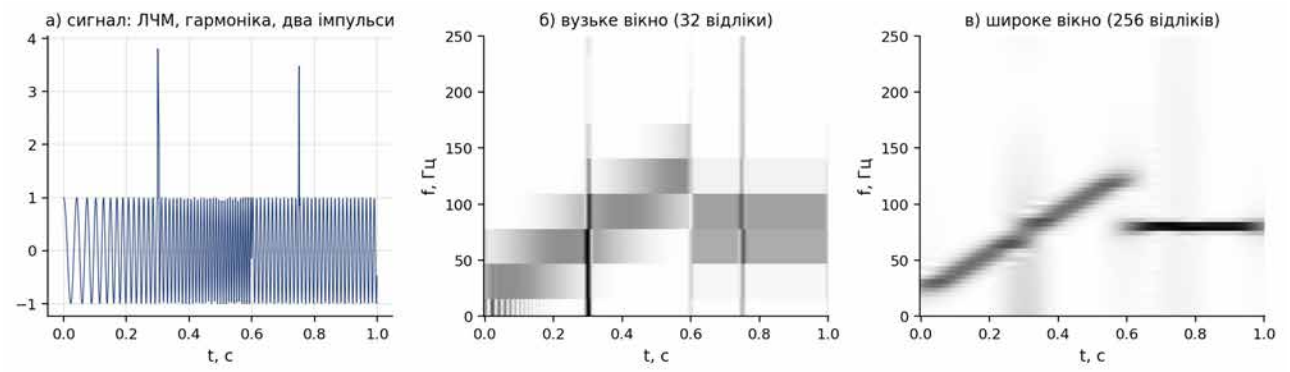


Рис. 6.2. Спектрограми віконного перетворення Фур'є з вузьким і широким вікнами

6.2.3. Обмеження віконного перетворення

Головний недолік віконного перетворення полягає в тому, що роздільна здатність однакова на всій частотно-часовій площині: вікно вибирається один раз для всіх частот. Проте природні сигнали зазвичай мають іншу структуру: низькочастотні складові тривають довго і потребують хорошої частотної роздільної здатності, а високочастотні є короткочасними подіями (перепади,

імпульси, контури), для яких важливіша часова локалізація. Як зазначено в [1], для графоелементів різної тривалості ширину вікна потрібно підлаштовувати, і можна було б виконувати серію віконних перетворень з різними вікнами — але це надлишково і незручно.

Цю проблему розв'язує вейвлет-перетворення. Корисно провести аналогію з радіотехнічним поняттям добротності коливального контуру чи смугового фільтра — відношення центральної частоти до ширини смуги пропускання [1]. Віконне перетворення — це набір фільтрів з *однаковою абсолютною* шириною смуги, тобто зі змінною добротністю. Вейвлет-перетворення — це набір фільтрів з *однаковою добротністю*: ширина смуги пропорційна центральній частоті. Із зростанням частоти частотна роздільна здатність погіршується, але часова — поліпшується, причому добуток $\Delta t \cdot \Delta\omega$ для всіх фільтрів однаковий і задовольняє (6.3).

Схематично розбиття частотно-часової площини різними перетвореннями показано на рис. 6.3: натуральні координати дають ідеальну локалізацію в часі без жодної частотної інформації; перетворення Фур'є — навпаки; віконне перетворення — рівномірну сітку однакових комірок; вейвлет-перетворення — комірки, що звужуються за часом і розширюються за частотою з її зростанням.

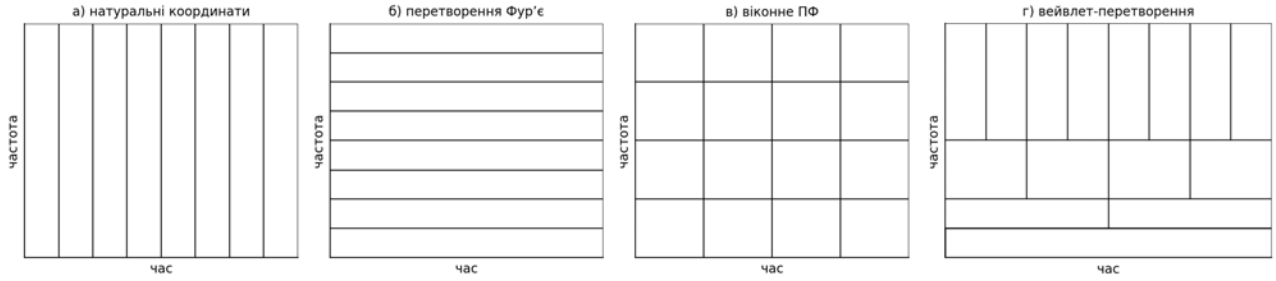


Рис. 6.3. Розбиття частотно-часової площини різними перетвореннями

6.3. Неперервне вейвлет-перетворення

6.3.1. Визначення

Нехай $\psi(t)$ — функція, що має нульове середнє і швидко спадає на нескінченності — «хвилька» (англ. *wavelet*). Усі інші базисні функції отримують з неї, яку називають материнським вейвлетом, зсувом на b і зміною масштабу $a > 0$:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{t-b}{a}\right). \tag{6.4}$$

Множник $1/\sqrt{a}$ зберігає енергію функції при зміні масштабу. Неперервне вейвлет-перетворення (НВП) сигналу $s(t)$ визначається як

$$W_s(a,b) = \int_{-\infty}^{\infty} s\,(t)\,\psi^*_{a,b}(t)\,dt, \tag{6.5}$$

а обернене —

$$s(t) = \frac{1}{C_\psi}\int_0^{\infty}\int_{-\infty}^{\infty} W_s\,(a,b)\,\psi_{a,b}(t)\,\frac{db\,da}{a^2}, \tag{6.6}$$

де стала $C_\psi = \int_0^\infty |\Psi(j\omega)|^2 / \omega \, d\omega$ залежить від обраного вейвлета і має бути скінченною (умова допустимості). Для цього, зокрема, необхідно, щоб $\Psi(0) = \int \psi(t)\, dt = 0$, тобто вейвлет повинен мати нульове середнє, — він «коливається».

Параметр масштабу a (англ. *scale*) обернено пропорційний частоті: при малих a вейвлет стиснутий і «бачить» дрібні, високочастотні деталі; при великих — розтягнутий і аналізує повільні зміни. Зображення $|W_s(a, b)|$ на площині «зсув — масштаб» називають скейлограмою (на відміну від спектрограми віконного перетворення Фур'є) [1].

6.3.2. Материнські вейвлети

Вибір материнського вейвлета визначається задачею. Найпоширенішими є (рис. 6.4):

- вейвлет Хаара — прямокутна функція з двома сходинками; найпростіший, добре локалізований у часі, але погано — за частотою;
- «мексиканський капелюх» — друга похідна гаусівської функції з оберненим знаком, $\psi(t) = (1 - t^2)e^{-t^2/2}$; добре виявляє локальні екстремуми та особливості;
- вейвлет Морле — гармоніка, обмежена гаусівським вікном, $\psi(t) = e^{j\omega_0 t}e^{-t^2/2}$ (зазвичай $\omega_0 \approx 5 \ldots 6$); має добру частотну локалізацію і є прямим аналогом функцій Габора, але, на відміну від віконного перетворення Фур'є, при зміні частоти масштабується разом із вікном;
- вейвлети Добеші db$N$ [25] — ортогональні вейвлети з компактним носієм і N нульовими моментами, що визначаються не аналітичною формулою, а коефіцієнтами відповідних цифрових фільтрів.

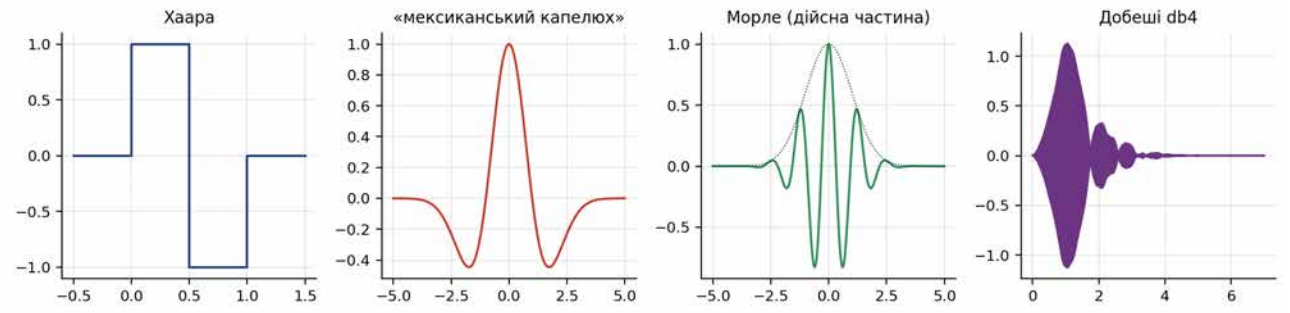


Рис. 6.4. Материнські вейвлети

Розглянутий у [1] підхід дозволяє будувати вейвлети як вікно, заповнене трансформантами деякого ортогонального перетворення: материнською функцією є вікно (гаусівське, Хеммінга, Ханна, Бартлетта, прямокутне) з «заповненням» трансформантою Фур'є, а при зміні масштабу a ширина вікна змінюється разом із періодом заповнення. Так вейвлет Морле можна розглядати як «вейвлет-Фур'є»; аналогічно будуються вейвлети на основі функцій Уолша.

Скейлограму тестового сигналу (рис. 6.2, а), обчислену за допомогою вейвлета Морле, наведено на рис. 6.5. На відміну від спектрограм, тут одночасно

добре відображено і плавну зміну частоти ЛЧМ-коливання на низьких частотах, і короткі імпульси, які на високих частотах локалізовані у вузькі «конуси».

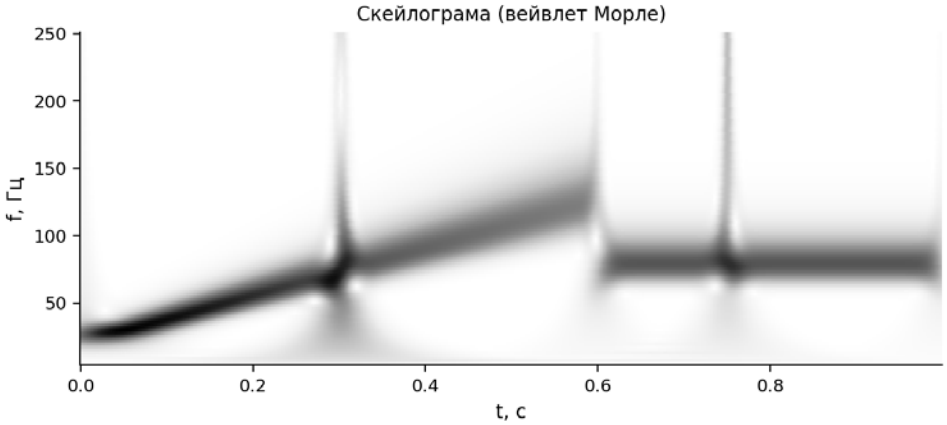


Рис. 6.5. Скейлограма тестового сигналу

6.3.3. Вейвлети і особливості сигналів

Важлива властивість вейвлетів пов'язана з нульовими моментами. Кажуть, що вейвлет має p нульових моментів, якщо $\int t^k \psi(t)\, dt = 0$ для $k = 0,1, \dots, p-1$. Тоді вейвлет «не помічає» поліноміальні складові сигналу степеня, меншого за p: на гладких ділянках вейвлет-коефіцієнти малі, а великі значення виникають лише поблизу особливостей — розривів, зламів, піків. Більше того, швидкість спадання коефіцієнтів $|W_s(a,b)|$ при $a \to 0$ поблизу точки b характеризує ступінь гладкості (регулярність за Ліпшицем) сигналу в цій точці [24]. Ця властивість лежить в основі використання вейвлетів для виявлення контурів, стиснення та придушення шуму: енергія гладкого сигналу з поодинокими особливостями зосереджується в небагатьох великих коефіцієнтах.

6.4. Дискретне вейвлет-перетворення і кратномасштабний аналіз

6.4.1. Від неперервного до дискретного перетворення

Неперервне вейвлет-перетворення надлишкове: одновимірний сигнал відображається у двовимірну функцію $W_s(a,b)$. Для обчислень і стиснення важливо обмежитися дискретною множиною масштабів і зсувів. Найпоширенішою є діадна сітка: $a = 2^j$, $b = k\,2^j$, $j,k \in \mathbb{Z}$. Для спеціально побудованих вейвлетів функції

$$\psi_{j,k}(t) = 2^{-j/2}\,\psi\left(2^{-j}t - k\right) \tag{6.7}$$

утворюють ортонормований базис, і сигнал однозначно подається рядом коефіцієнтів $d_{j,k} = \langle s, \psi_{j,k} \rangle$. Найпростішим прикладом є система Хаара (5.15). Загальна теорія побудови таких базисів — кратномасштабний аналіз (КМА) — розроблена С. Маллатом та І. Добеші [24, 25].

6.4.2. Кратномасштабний аналіз

Ідея КМА полягає в послідовному поданні сигналу з дедалі грубішою роздільною здатністю. Вводиться масштабувальна функція $\varphi(t)$, зсуви якої

$\varphi(t-k)$ утворюють базис простору V_0 «грубих» наближень. Простори V_j наближень з масштабом 2^j вкладені один в одний: $\ldots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots$. Різниця між сусідніми просторами — простір деталей W_j, базис якого утворюють вейвлети $\psi_{j,k}$:

$$V_{j-1} = V_j \oplus W_j. \tag{6.8}$$

Тобто сигнал з роздільною здатністю 2^{j-1} — це його грубіше наближення з роздільною здатністю 2^j плюс деталі, втрачені при огрубленні.

Оскільки φ і ψ належать простору V_{-1}, вони виражаються через стиснуті вдвічі масштабувальні функції — масштабувальне рівняння:

$$\varphi(t) = \sqrt{2}\sum_k h\,(k)\,\varphi(2t-k), \qquad \psi(t) = \sqrt{2}\sum_k g\,(k)\,\varphi(2t-k). \tag{6.9}$$

Коефіцієнти $h(k)$ утворюють імпульсну характеристику фільтра нижніх частот, а $g(k) = (-1)^k h(L-1-k)$ — фільтра верхніх частот (квадратурно-дзеркальна пара). Саме ці фільтри, а не аналітичні формули для φ і ψ, використовуються в обчисленнях. Для вейвлета Хаара $h = [1,1]/\sqrt{2}$, $g = [1,-1]/\sqrt{2}$; для вейвлета Добеші db2

$$h = \frac{1}{4\sqrt{2}}\left[1+\sqrt{3},\ 3+\sqrt{3},\ 3-\sqrt{3},\ 1-\sqrt{3}\right]. \tag{6.10}$$

Графіки функцій φ і ψ отримують з (6.9) ітераційно (каскадним алгоритмом); так побудовано графік вейвлета db4 на рис. 6.4.

6.4.3. Алгоритм Малла

Обчислення дискретного вейвлет-перетворення (ДВП) зводиться до рекурсивної фільтрації з проріджуванням — алгоритму Малла [24]. Позначимо $a_0(n) = x(n)$ відліки сигналу. На кожному рівні j:

$$a_j(k) = \sum_n h\,(n-2k)\,a_{j-1}(n), \qquad d_j(k) = \sum_n g\,(n-2k)\,a_{j-1}(n), \tag{6.11}$$

тобто послідовність наближень пропускається через фільтри нижніх і верхніх частот, а результати проріджуються вдвічі (залишається кожний другий відлік). Коефіцієнти d_j (деталі) зберігаються, а наближення a_j передаються на наступний рівень (рис. 6.6). Після J рівнів сигнал подається набором $\{a_J, d_J, d_{J-1}, \ldots, d_1\}$, загальна кількість коефіцієнтів якого дорівнює кількості відліків сигналу. Обернене перетворення виконується у зворотному порядку: вставлення нулів між відліками, фільтрація синтезуючими фільтрами і підсумовування.

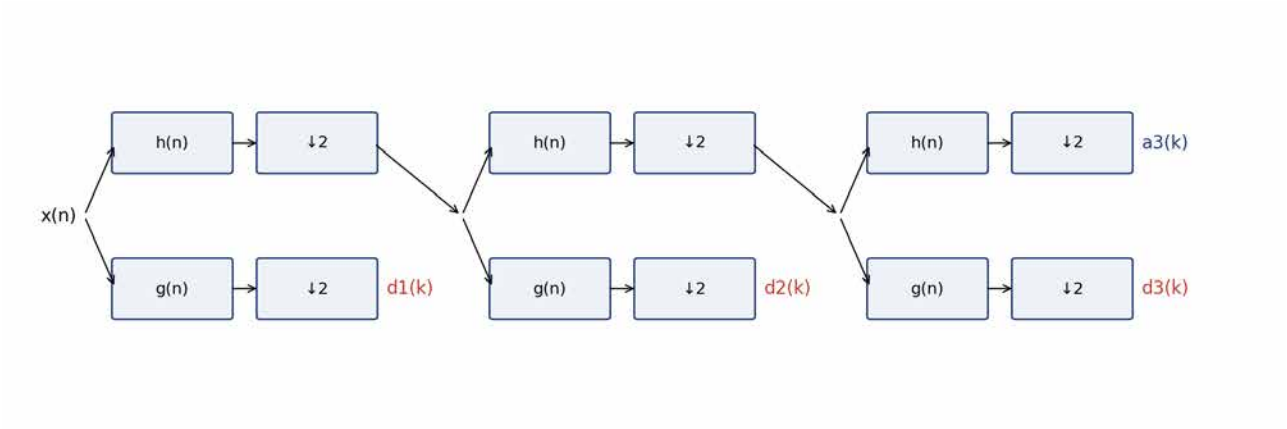


Рис. 6.6. Банк фільтрів алгоритму Малла (три рівні розкладу)

Обчислювальна складність алгоритму Малла становить $O(N)$: на першому рівні виконується $2LN/2 = LN$ множень (L — довжина фільтрів), на кожному наступному — вдвічі менше, тож загалом — не більше $2LN$. Це швидше за ШПФ.

На рис. 6.7 показано чотирирівневий розклад тестового сигналу з вейвлетом db4. Деталі першого рівня d_1 (найвищі частоти) відображають в основному імпульси та різкі зміни; деталі нижчих рівнів — коливальні складові відповідних частотних смуг; наближення a_4 — повільну складову. Кожна смуга займає октаву, тобто має однакову відносну ширину, — у повній відповідності з принципом сталої добротності.

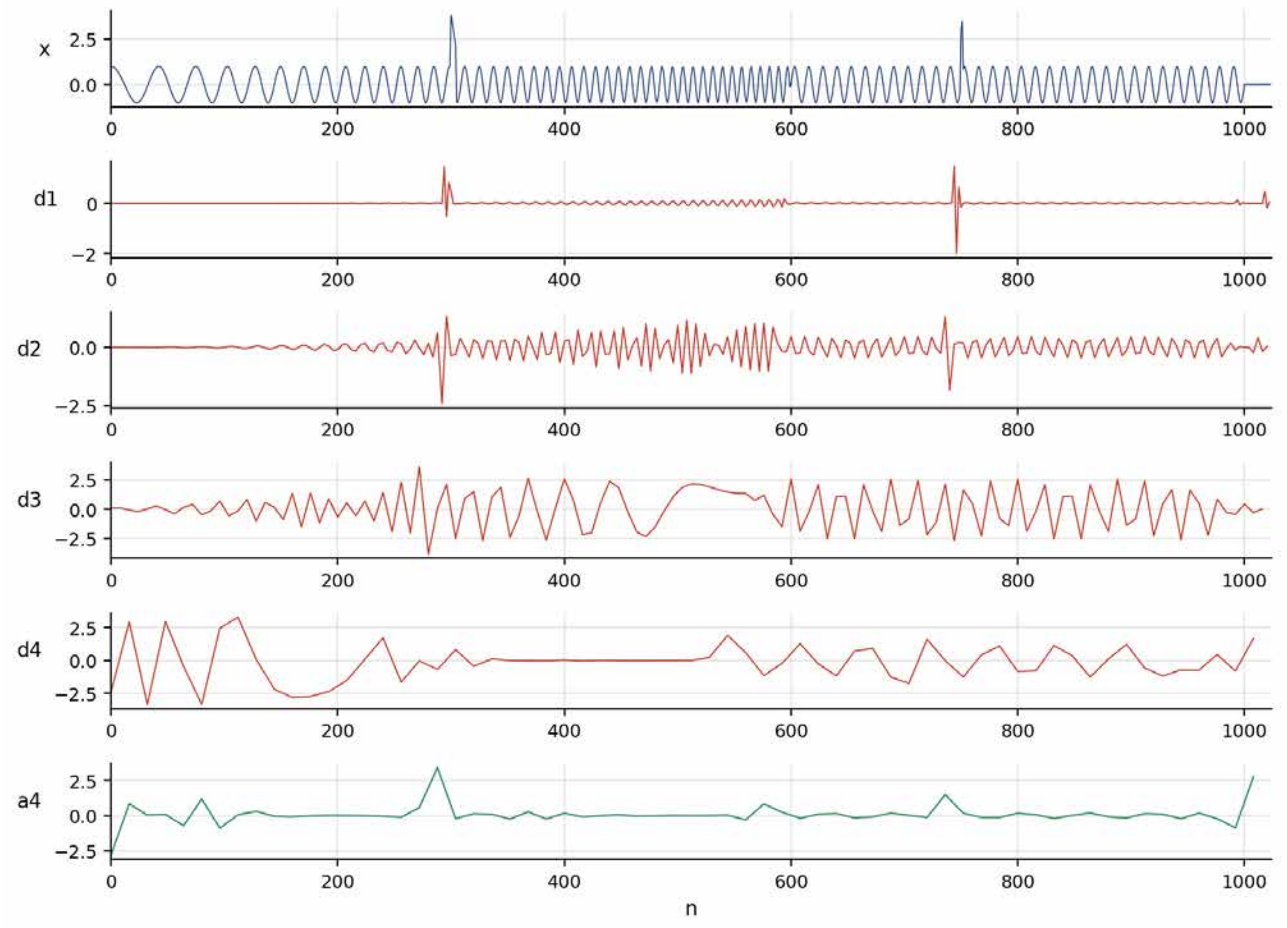


Рис. 6.7. Багаторівневий вейвлет-розклад тестового сигналу (db4)

6.4.4. Двовимірне дискретне вейвлет-перетворення

Для зображень ДВП виконується сепарабельно: на кожному рівні фільтрація з проріджуванням застосовується спочатку до рядків, потім до стовпців. У результаті утворюються чотири субсмуги: LL — наближення (низькі частоти по обох напрямках); LH — горизонтальні деталі (контури, витягнуті горизонтально); HL — вертикальні деталі; HH — діагональні деталі. Наступний рівень розкладу застосовується лише до субсмуги LL. На рис. 6.8, а наведено дворівневий розклад тестового зображення: субсмуги деталей містять контури відповідної орієнтації, причому більшість коефіцієнтів близька до нуля. Розподіл енергії між субсмугами (рис. 6.8, б) показує, що понад 99 % енергії зосереджено в наближенні; це основа вейвлет-стиснення (стандарт JPEG 2000).

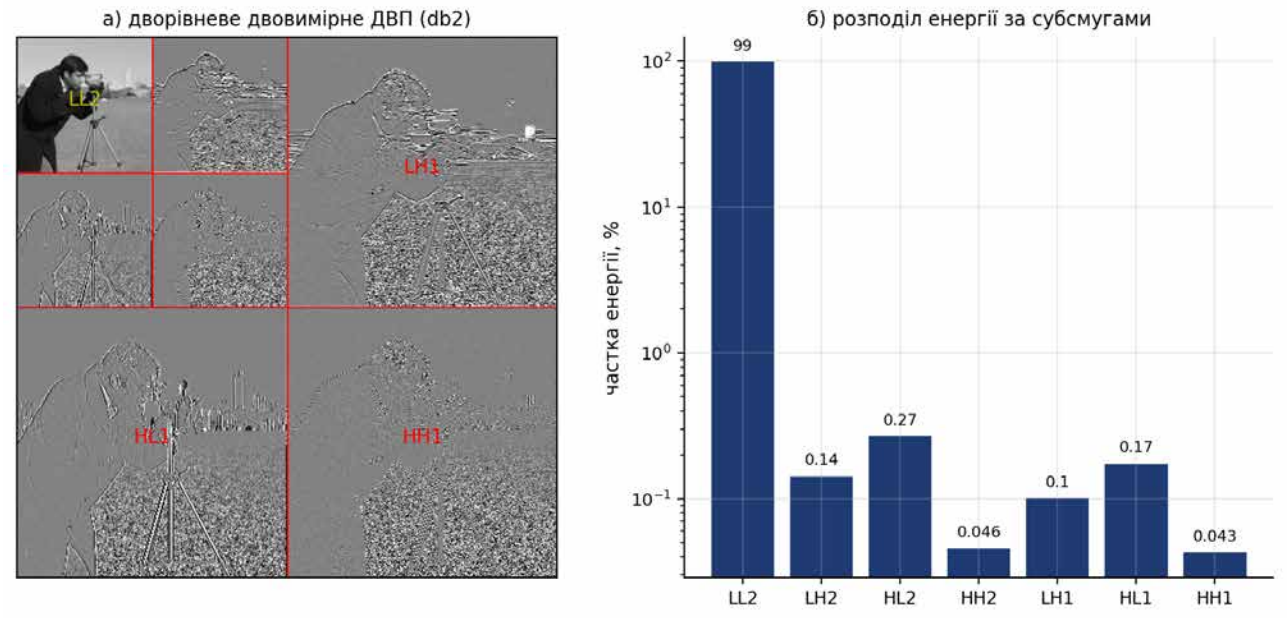


Рис. 6.8. Двовимірне дискретне вейвлет-перетворення зображення

Для задач розпізнавання недоліком ДВП є відсутність інваріантності до зсуву: через проріджування зсув образу на один відлік може суттєво змінити коефіцієнти. Тому для формування ознак часто використовують непроріджене (стаціонарне) вейвлет-перетворення, у якому проріджування не виконується, а фільтри на кожному рівні «розріджуються» вставлянням нулів, або усереднюють результати за кількома зсувами (метод циклічних зсувів).

6.5. Придушення шуму методом вейвлет-порогової обробки

6.5.1. Принцип

Нехай спостерігається сигнал $y(n) = s(n) + v(n)$, де v — білий гаусівський шум з дисперсією σ^2. Оскільки ортонормоване ДВП є унітарним перетворенням, білий шум переходить у білий шум з тією самою дисперсією: шум рівномірно розподіляється по *всіх* вейвлет-коефіцієнтах. Натомість корисний сигнал, гладкий на більшій частині області визначення, подається *небагатьма* великими коефіцієнтами. Отже, коефіцієнти, що за модулем значно перевищують рівень шуму, з високою ймовірністю несуть інформацію про сигнал, а малі —

переважно шум. Звідси метод вейвлет-порогової обробки (англ. *wavelet shrinkage*) [24]:

1. обчислити ДВП зашумленого сигналу;
2. застосувати до коефіцієнтів деталей порогову функцію з порогом T;
3. виконати обернене ДВП.

Застосовують дві основні порогові функції:

$$\text{жорсткий поріг:}\quad \eta_H(d) = \begin{cases} d, & |d| > T, \\ 0, & |d| \le T; \end{cases} \qquad \text{м'який поріг:}\quad \eta_S(d) = \operatorname{sign}(d)\max(|d| - T, 0). \tag{6.12}$$

Жорсткий поріг зберігає амплітуду значущих коефіцієнтів, але створює розриви в залежності «вхід — вихід», що іноді породжує поодинокі артефакти. М'який поріг дає плавніші результати, але зменшує амплітуди всіх значущих коефіцієнтів на T, що призводить до згладжування різких перепадів.

6.5.2. Вибір порогу

Д. Донохо та І. Джонстон [24] запропонували універсальний поріг

$$T = \sigma\sqrt{2\ln N}, \tag{6.13}$$

де N — кількість відліків. Обґрунтування таке: максимум модуля N незалежних гаусівських величин з дисперсією σ^2 з імовірністю, що прямує до одиниці, не перевищує $\sigma\sqrt{2\ln N}$. Тому поріг (6.13) майже напевно «прибирає» всі коефіцієнти, зумовлені лише шумом. Невідоме значення σ оцінюють робастно за коефіцієнтами першого рівня деталей за формулою (3.26) [24]. Існують і адаптивні правила вибору порогу для кожного рівня (SureShrink, BayesShrink), які зазвичай дають кращі результати, ніж універсальний поріг.

6.5.3. Приклади

На рис. 6.9 наведено результати придушення шуму для кусково-сталого сигналу з різкими перепадами і вузькими піками. Лінійний гаусівський фільтр згладжує шум, але одночасно розмиває перепади і зменшує амплітуду вузьких піків — це принципова властивість лінійної фільтрації, яка не розрізняє шум і корисні високочастотні деталі. Вейвлет-обробка з жорстким порогом (вейвлет Хаара, добре пристосований до кусково-сталих сигналів) зберігає різкість перепадів і амплітуди піків, забезпечуючи значно вище відношення сигнал/шум. М'який поріг дає гладкий результат, але помітно зменшує амплітуди.

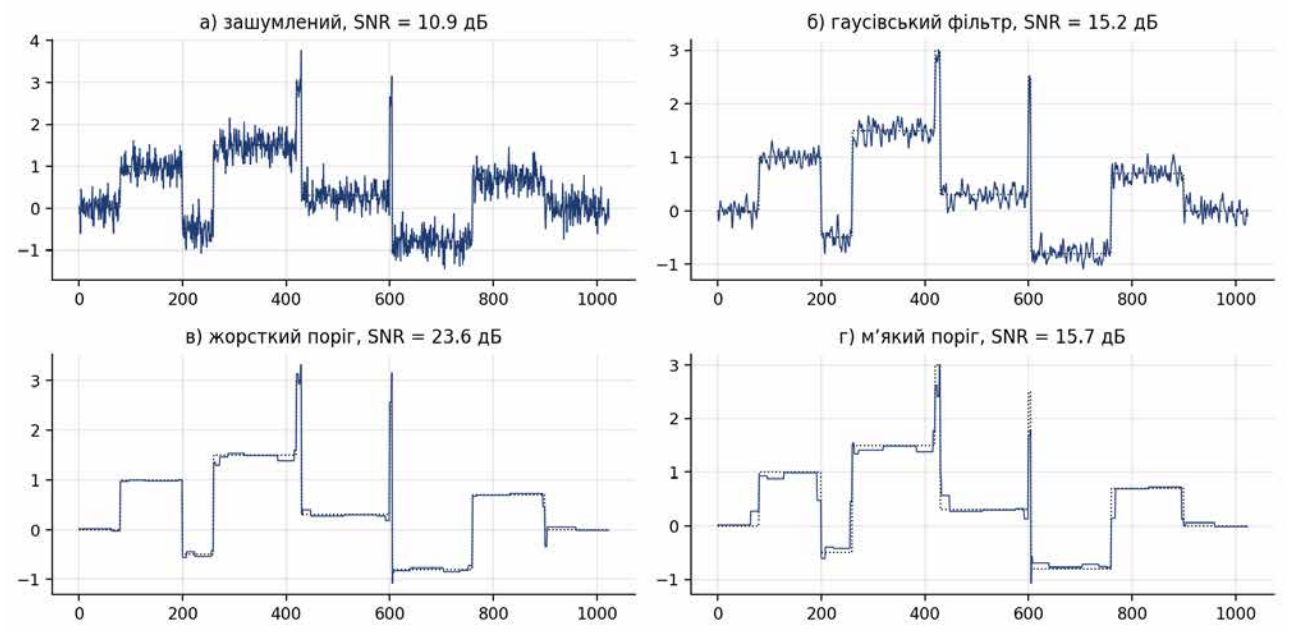


Рис. 6.9. Придушення шуму в кусково-сталому сигналі

Для зображень (рис. 6.10) простий поріговий метод на основі ортогонального ДВП дає блочні артефакти через відсутність інваріантності до зсуву. Їх значно зменшує метод циклічних зсувів (англ. *cycle spinning*): обробку повторюють для кількох циклічних зсувів зображення, результати зсувають назад і усереднюють. Отримане зображення має вище значення PSNR, ніж після гаусівської фільтрації, і краще збережені контури.



Рис. 6.10. Придушення шуму на зображенні

Вейвлет-порогова обробка є нелінійним методом: вона адаптується до локальної структури сигналу, «вмикаючи» згладжування на однорідних ділянках і «вимикаючи» його поблизу особливостей. Саме тому вона перевершує лінійні фільтри для образів з різкими межами. Варто зазначити, що порогова обробка в області ортогонального перетворення — загальний принцип, застосовний і до інших базисів (ДКП, Уолша), проте найефективніший він для базисів, у яких корисний сигнал подається найбільш розріджено.

6.6. Вейвлети і фільтри Габора у формуванні ознак

6.6.1. Вейвлет-ознаки текстури

Текстура — це просторова організація повторюваних елементів яскравості, що характеризується їх розміром, орієнтацією, контрастом і регулярністю. Вейвлет-розклад природно розділяє текстуру за масштабами (рівнями) і орієнтаціями (субсмуги LH, HL, HH). Простий і ефективний набір ознак утворюють енергії субсмуг [24, 48]:

$$E_{j,o} = \frac{1}{N_j}\sum_k \left|d_{j,o}(k)\right|^2 \quad \text{або} \quad E_{j,o} = \frac{1}{N_j}\sum_k \left|d_{j,o}(k)\right|, \tag{6.14}$$

де j — рівень, $o \in \{LH, HL, HH\}$ — орієнтація, N_j — кількість коефіцієнтів у субсмузі. Для трирівневого розкладу отримуємо 9 ознак деталей плюс середнє і дисперсію наближення. Вектор таких ознак, обчислений у ковзному вікні, використовується для сегментації текстурних зображень (аерофотознімків, медичних зображень тканин, зображень поверхонь у контролі якості). Ознаки на основі ДВП з проріджуванням чутливі до зсуву вікна, тому для ознак краще використовувати непроріджене перетворення.

6.6.2. Фільтри Габора

Двовимірний фільтр Габора — це гармонічна хвиля заданої довжини λ та орієнтації θ, обмежена гаусівським вікном [50]:

$$g(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(\frac{2\pi x'}{\lambda} + \phi\right), \tag{6.15}$$

де $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$; σ — ширина гаусівської огинаючої; γ — коефіцієнт еліптичності; ϕ — фаза. Фільтри Габора оптимально локалізовані в просторі і частоті (досягають межі (6.3)), а їх форма близька до рецептивних полів простих клітин зорової кори ссавців. Банк фільтрів Габора з кількома масштабами і орієнтаціями (рис. 6.11) є класичним засобом формування текстурних ознак, ознак для розпізнавання облич, відбитків пальців (підсилення папілярних ліній), райдужної оболонки ока. Цікаво, що фільтри першого шару навчених згорткових нейронних мереж часто нагадують саме фільтри Габора [26].

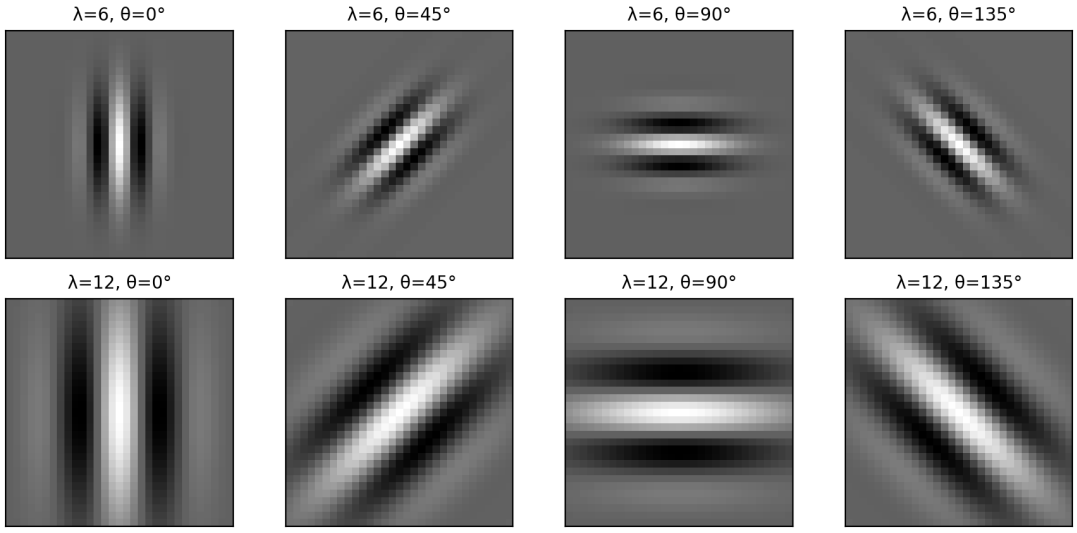


Рис. 6.11. Ядра фільтрів Габора для двох довжин хвилі і чотирьох орієнтацій

Ознаками зазвичай слугують середні значення та дисперсії модулів відгуків фільтрів (для комплексних фільтрів — модуль суми парної і непарної компонент) у межах області інтересу.

6.6.3. Частотно-часові ознаки одновимірних сигналів

Для одновимірних образів (мовлення, біосигнали, вібросигнали) частотно-часове подання є основним джерелом ознак. Спектрограма, обчислена в коротких вікнах (зазвичай 20–30 мс для мовлення), з подальшим групуванням спектральних коефіцієнтів у смуги, логарифмуванням і косинусним перетворенням дає мел-частотні кепстральні коефіцієнти (MFCC) — стандартні ознаки в системах розпізнавання мовлення. Вейвлет-розклад використовується для виявлення і класифікації комплексів електрокардіограми, епілептиформної активності на електроенцефалограмі, дефектів підшипників за вібросигналами [24, 37]. Як ознаки використовують енергії рівнів розкладу, статистики коефіцієнтів, положення і амплітуди локальних максимумів модуля вейвлет-перетворення.

6.7. Вейвлет-стиснення зображень

Висока енергетична компактність вейвлет-подання природних зображень (рис. 6.8) робить його ефективним засобом стиснення. Найпростіша схема стиснення з втратами полягає у збереженні лише заданої частки найбільших за модулем вейвлет-коефіцієнтів (наближення найгрубішого рівня зберігається повністю) з наступним квантуванням і ентропійним кодуванням [24, 25]. На відміну від блочного ДКП (розділ 5), вейвлет-перетворення застосовується до всього зображення і не створює блочних артефактів; помилки стиснення проявляються як помірне розмиття і слабкий «дзвін» біля контурів.

На рис. 6.12 наведено результати відновлення зображення після збереження 10, 3 і 1 % вейвлет-коефіцієнтів (вейвлет db2, чотири рівні розкладу) та для порівняння — значення PSNR для блочного ДКП $8 times 8$ з тією самою часткою коефіцієнтів (зі збереженням постійних складових блоків). При помірному

стисненні обидва підходи дають близькі результати (34,3 і 34,2 дБ при 10 % коефіцієнтів), а при сильному стисненні вейвлет-подання суттєво виграє (26,0 проти 22,4 дБ при 1 %), оскільки блочне ДКП не може адаптуватися до вмісту поза межами блоку і створює помітні блочні артефакти. Саме ці переваги зумовили використання вейвлет-перетворення в стандарті JPEG 2000 [13, 24]. Для задач розпізнавання з цього випливає корисний висновок: вейвлет-коефіцієнти компактно описують зображення на кількох масштабах, і невелика кількість найбільших коефіцієнтів може слугувати ознаковим описом.

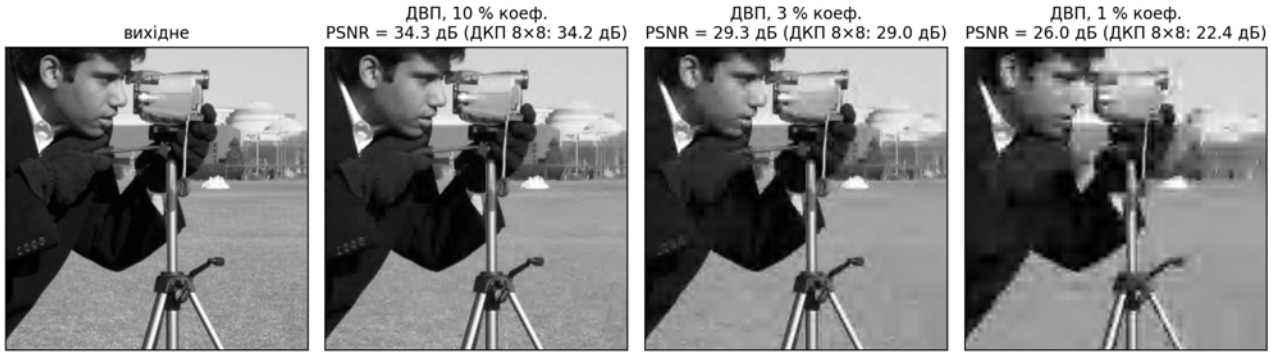


Рис. 6.12. Відновлення зображення за частиною найбільших вейвлет-коефіцієнтів

6.8. Програмна реалізація

Для вейвлет-аналізу в Python використовується бібліотека PyWavelets (pywt), а для віконного перетворення Фур'є — модуль scipy.signal. Нижче наведено приклади обчислення спектрограми, багаторівневого ДВП, придушення шуму з універсальним порогом та обчислення вейвлет-ознак текстури.

```
import numpy as np
import pywt
from scipy import signal

fs = 1000
t = np.arange(0, 1, 1 / fs)
x = signal.chirp(t, 20, 1, 200) * (t < 0.6) + np.sin(2*np.pi*80*t) * (t >= 0.6)

# спектрограма (6.1): вікно Ханна, 64 відліки, перекриття 56
f, tau, Z = signal.stft(x, fs, window='hann', nperseg=64, noverlap=56)
S = np.abs(Z) ** 2

# багаторівневе ДВП (алгоритм Малла) з вейвлетом db4
coeffs = pywt.wavedec(x, 'db4', level=4)        # [a4, d4, d3, d2, d1]

# придушення шуму: оцінка sigma (3.26) і універсальний поріг (6.13)
y = x + 0.3 * np.random.default_rng(0).normal(size=x.size)
c = pywt.wavedec(y, 'db4', level=5)
sigma = np.median(np.abs(c[-1])) / 0.6745
```

```
T = sigma * np.sqrt(2 * np.log(y.size))
c_thr = [c[0]] + [pywt.threshold(d, T, mode='hard') for d in c[1:]]
y_den = pywt.waverec(c_thr, 'db4')[:y.size]

# вейвлет-ознаки текстури (6.14) для фрагмента зображення
from skimage import data, img_as_float
patch = img_as_float(data.brick())[:128, :128]
feats = []
for (cH, cV, cD) in pywt.wavedec2(patch, 'db2', level=3)[1:]:
    feats += [np.mean(cH**2), np.mean(cV**2), np.mean(cD**2)]
print(np.round(feats, 5))
```

6.9. Приклади розв’язання задач

Задача 6.1. Виконайте дворівневий розклад Хаара сигналу $x = [4,6,10,12,8,6,5,5]$.

Розв’язання. Для вейвлета Хаара $h = [1,1]/\sqrt{2}$, $g = [1,-1]/\sqrt{2}$. Перший рівень (6.11): наближення $a_1 = [(4+6),(10+12),(8+6),(5+5)]/\sqrt{2} \approx [7{,}07;\ 15{,}56;\ 9{,}90;\ 7{,}07]$; деталі $d_1 = [(4-6),(10-12),(8-6),(5-5)]/\sqrt{2} \approx [-1{,}41;\ -1{,}41;\ 1{,}41;\ 0]$. Другий рівень: $a_2 = [(7{,}07+15{,}56),(9{,}90+7{,}07)]/\sqrt{2} = [16{,}12]$; $d_2 = [-6{,}2]$. Перевірка збереження енергії: $\sum x^2 = 16+36+100+144+64+36+25+25 = 446$; $|a_2|^2 + |d_2|^2 + |d_1|^2 = (256+144)+(36+4)+(2+2+2+0) = 446$. Основна частина енергії (90 %) зосереджена в наближенні a_2, а найбільший коефіцієнт деталей (-6 на другому рівні) відповідає перепаду між першою і другою половинами першої четвірки відліків.

Задача 6.2. Сигнал довжиною $N = 1024$ відліки містить білий шум з $\sigma = 0{,}1$. Обчисліть універсальний поріг.

Розв’язання. За (6.13) $T = 0{,}1\sqrt{2\ln 1024} = 0{,}1\sqrt{13{,}86} \approx 0{,}372$. Коефіцієнти деталей з модулем, меншим за 0,372, обнуляються. Для гаусівського шуму ймовірність того, що окремий шумовий коефіцієнт перевищить $3{,}72\sigma$, становить близько $2 \cdot 10^{-4}$, тобто серед 1024 коефіцієнтів у середньому залишиться лише 0,2 «шумового» коефіцієнта.

Задача 6.3. Мовний сигнал дискретизовано з частотою 8 кГц і аналізується вікном 256 відліків. Визначте часову і частотну роздільну здатність віконного перетворення Фур’є.

Розв’язання. Тривалість вікна $T = 256/8000 = 32$ мс; крок частотної сітки ДПФ $\Delta f = 1/T = 31{,}25$ Гц (з урахуванням ширини головної пелюстки вікна Ханна реальна роздільна здатність приблизно вдвічі гірша). Події, розділені інтервалом менше ніж близько 30 мс, не розрізняються. Для аналізу основного тону жіночого голосу (200–300 Гц) така роздільна здатність достатня; для виявлення коротких вибухових приголосних (тривалістю 5–10 мс) потрібне вужче вікно — і, відповідно, гірша частотна роздільна здатність.

Висновки до розділу 6

1. Глобальні ортогональні перетворення не відображають локалізацію частотних складових у часі (просторі), тому для нестаціонарних образів необхідні методи частотно-часового аналізу.
2. Віконне перетворення Фур'є обчислює спектри фрагментів сигналу, виділених ковзним вікном. Його роздільна здатність за часом і частотою стала і обмежена нерівністю Гейзенберга — Габора $\Delta t\,\Delta\omega \geq 1/2$; вибір ширини вікна є компромісом.
3. Вейвлет-перетворення використовує базисні функції, отримані з материнського вейвлета зсувом і масштабуванням; це еквівалентно банку фільтрів зі сталою добротністю. На високих частотах воно забезпечує кращу часову, на низьких — кращу частотну роздільну здатність.
4. Дискретне вейвлет-перетворення на діадній сітці реалізується алгоритмом Малла — рекурсивною фільтрацією парою квадратурно-дзеркальних фільтрів з проріджуванням — за $O(N)$ операцій. Двовимірне ДВП розкладає зображення на субсмуги наближення і деталей трьох орієнтацій.
5. Завдяки нульовим моментам вейвлетів енергія гладких образів з поодинокими особливостями зосереджується в небагатьох коефіцієнтах. На цьому ґрунтуються вейвлет-стиснення, робастне оцінювання рівня шуму і нелінійне придушення шуму пороговою обробкою, яке зберігає різкі перепади краще за лінійні фільтри.
6. Енергії вейвлет-субсмуг і відгуки банку фільтрів Габора є ефективними ознаками текстури; частотно-часові подання — основне джерело ознак одновимірних нестаціонарних сигналів.

Питання для обговорення та самоперевірки

1. Чому амплітудний спектр Фур'є недостатній для аналізу нестаціонарних образів? Наведіть приклад.
2. Запишіть визначення віконного перетворення Фур'є. Що таке спектрограма?
3. Як визначаються ефективна тривалість і ширина смуги вікна?
4. Сформулюйте принцип невизначеності Гейзенберга — Габора. Для якого вікна досягається рівність?
5. Як впливає ширина вікна на вигляд спектрограми?
6. Поясніть аналогію між вейвлет-перетворенням і банком фільтрів зі сталою добротністю.
7. Запишіть визначення неперервного вейвлет-перетворення. Що таке умова допустимості?
8. Як пов'язані масштаб вейвлета і частота?
9. Охарактеризуйте вейвлети Хаара, «мексиканський капелюх», Морле, Добеші.
10. Що дає вейвлету наявність нульових моментів?
11. Поясніть зміст кратномасштабного аналізу. Що таке масштабувальна функція?

12. Опишіть алгоритм Малла. Яка його обчислювальна складність?
13. Які субсмуги утворюються при двовимірному ДВП? Що вони містять?
14. Чому ДВП не інваріантне до зсуву? Як цей недолік усувають?
15. Опишіть метод придушення шуму пороговою обробкою вейвлет-коефіцієнтів. Чим відрізняються жорсткий і м’який пороги?
16. Обґрунтуйте вибір універсального порогу.
17. Чому вейвлет-обробка краще зберігає контури, ніж лінійна фільтрація?
18. Як формуються вейвлет-ознаки текстури?
19. Запишіть вираз для фільтра Габора. Які його параметри?

Завдання для самостійної роботи

1. Побудуйте спектрограми мовного сигналу (запис власного голосу) з вікнами 10, 25 і 100 мс. Поясніть відмінності.
2. Обчисліть ефективні тривалість і ширину смуги для прямокутного, Ханна і гаусівського вікон і перевірте нерівність (6.3).
3. Реалізуйте неперервне вейвлет-перетворення з вейвлетом Морле через згортку і побудуйте скейлограму тестового сигналу.
4. Реалізуйте алгоритм Малла для вейвлетів Хаара і db2 за формулами (6.10), (6.11) та обернене перетворення. Перевірте точність відновлення і збереження енергії.
5. Дослідіть залежність ефективності придушення шуму від вибору вейвлета (Хаар, db2, db4, sym8), кількості рівнів розкладу і типу порогу для кусково-сталого і гладкого сигналів.
6. Сформуйте вейвлет-ознаки (6.14) для фрагментів кількох текстур (цегла, трава, гравій) і побудуйте класифікатор за мінімальною відстанню до середнього вектора ознак класу.
7. Побудуйте банк фільтрів Габора (3 масштаби, 6 орієнтацій), обчисліть середні модулі відгуків для тих самих текстур і порівняйте якість класифікації з вейвлет-ознаками.

РОЗДІЛ 7. ПОПЕРЕДНЄ ОБРОБЛЕННЯ ЗОБРАЖЕНЬ У ПРОСТОРОВІЙ ОБЛАСТІ

Навчальні цілі. Після вивчення розділу ви зможете: застосовувати точкові перетворення яскравості та методи перетворення гістограми; обирати і реалізовувати лінійні та нелінійні просторові фільтри для згладжування і підвищення різкості; виділяти контури градієнтними операторами, операторами Лапласа і детектором Канні; виконувати порогову сегментацію методом Оцу та адаптивними методами; застосовувати операції математичної морфології для оброблення бінарних образів; виконувати геометричне нормування образів з інтерполяцією.

7.1. Завдання попереднього оброблення

Попереднє оброблення займає проміжне місце між реєстрацією образу та формуванням ознак (рис. 1.2). Його мета — перетворити первинне зображення на вигляд, найбільш придатний для подальшого аналізу: зменшити вплив шуму, вирівняти освітленість і контраст, підкреслити суттєві деталі (контури), виділити об’єкти інтересу з фону, привести образи до стандартного положення і розміру. На відміну від методів реставрації (розділи 8, 9), які спираються на явну модель спотворення, методи цього розділу здебільшого є *евристичними*: вони не потребують знання імпульсної характеристики і параметрів шуму, а їх параметри підбираються емпірично для конкретного класу зображень [13].

Методи попереднього оброблення поділяють на:

- точкові, у яких нове значення пікселя залежить лише від його власного значення: $g(m,n) = T[f(m,n)]$;
- локальні (просторові), у яких нове значення залежить від значень у деякому околі пікселя;
- глобальні, що використовують усе зображення (перетворення в частотній області, розділ 4; вирівнювання гістограми);
- геометричні, що змінюють положення пікселів.

У цьому розділі розглянуто перші дві групи, а також порогову сегментацію, морфологічні операції і геометричні перетворення.

7.2. Точкові перетворення яскравості

7.2.1. Основні функції перетворення

Нехай $r \in [0,1]$ — нормована вхідна яскравість, $s = T(r)$ — вихідна. Найбільш уживані функції перетворення показано на рис. 7.1.

Негатив: $s = 1 - r$. Використовується для зручнішого візуального аналізу деталей на темному фоні (наприклад, на рентгенограмах).

Логарифмічне перетворення: $s = c\log(1 + r)$. Стискає діапазон великих значень і розтягує діапазон малих; застосовується для відображення сигналів з величезним динамічним діапазоном, зокрема амплітудних спектрів Фур’є (рис. 4.7).

Степеневе (гамма) перетворення: $s = c\,r^{\gamma}$. При $\gamma < 1$ темні ділянки освітлюються, при $\gamma > 1$ — затемнюються. Більшість пристроїв відображення мають степеневу передавальну характеристику, яку компенсують гамма-корекцією. Як зазначалося в розділі 3, нелінійність передавальної характеристики системи реєстрації порушує лінійну модель деградації; тому перед реставрацією дані слід лінеаризувати оберненим гамма-перетворенням [13, 15].

Кусково-лінійне розтягнення контрасту: діапазон $[r_1, r_2]$, у якому зосереджена переважна частина значень, лінійно відображається на повний діапазон $[0,1]$:

$$s = \begin{cases} 0, & r < r_1, \\ \dfrac{r - r_1}{r_2 - r_1}, & r_1 \le r \le r_2, \\ 1, & r > r_2. \end{cases} \tag{7.1}$$

Межі r_1, r_2 часто обирають за процентилями гістограми (наприклад, 1 % і 99 %), щоб окремі «гарячі» пікселі не впливали на результат.

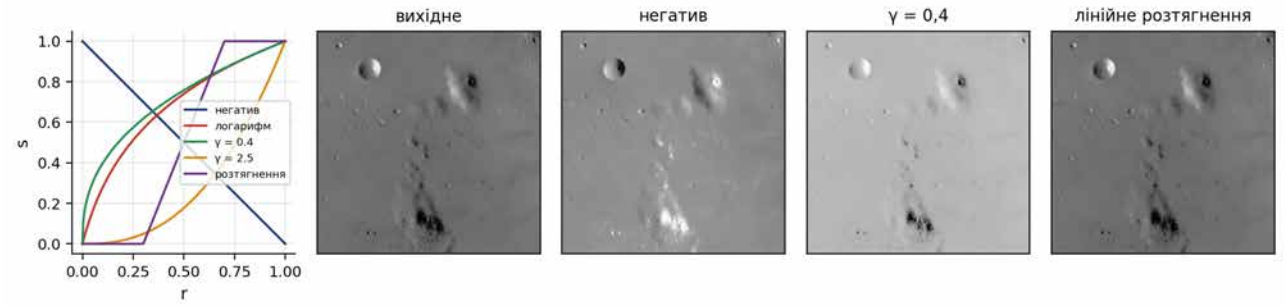


Рис. 7.1. Функції точкових перетворень та їх застосування

7.2.2. Вирівнювання гістограми

Зображення, яскравості якого займають вузький діапазон, мають низький контраст. Метод вирівнювання (еквалізації) гістограми автоматично будує монотонне перетворення, після якого яскравості розподілені якомога рівномірніше [13]. Для неперервної випадкової величини r з густиною $p_r(r)$ перетворення

$$s = T(r) = \int_0^r p_r\,(w)\,dw \tag{7.2}$$

(функція розподілу) дає величину s, рівномірно розподілену на $[0,1]$. Справді, $p_s(s) = p_r(r)\,|dr/ds| = p_r(r)/p_r(r) = 1$. У дискретному випадку для L рівнів

$$s_k = (L-1)\sum_{i=0}^{k} p\,(r_i), \qquad k = 0,1,\dots,L-1, \tag{7.3}$$

з округленням до найближчого цілого. Дискретний результат лише наближено рівномірний: рівні, що трапляються часто, розносяться на більші відстані, а рідкісні — зливаються. На рис. 7.2, а, б показано зображення з низьким

контрастом до і після вирівнювання гістограми; кумулятивна гістограма (червона крива) після вирівнювання близька до прямої.

Специфікація гістограми узагальнює вирівнювання: потрібно перетворити зображення так, щоб його гістограма набула заданої форми. Для цього обидві гістограми — фактичну і бажану — перетворюють на рівномірну за (7.3), а потім застосовують перше перетворення і обернене до другого. Специфікація використовується для нормування яскравості образів, отриманих у різних умовах, перед розпізнаванням.

7.2.3. Адаптивне вирівнювання гістограми

Глобальне вирівнювання може надмірно підсилювати шум на однорідних ділянках і недостатньо підсилювати контраст дрібних деталей у ділянках з іншою середньою яскравістю. Адаптивне вирівнювання виконує перетворення (7.3) окремо для кожного фрагмента (наприклад, 8 × 8 фрагментів зображення), з білінійною інтерполяцією перетворень між сусідніми фрагментами для уникнення видимих меж. У методі CLAHE (*Contrast Limited Adaptive Histogram Equalization*) [13] гістограма кожного фрагмента додатково «обрізається» на заданому рівні, а надлишок рівномірно перерозподіляється між усіма рівнями. Це обмежує нахил функції перетворення і, отже, підсилення шуму. CLAHE широко застосовується в медичній візуалізації та як етап попереднього оброблення перед розпізнаванням облич і символів (рис. 7.2, в).

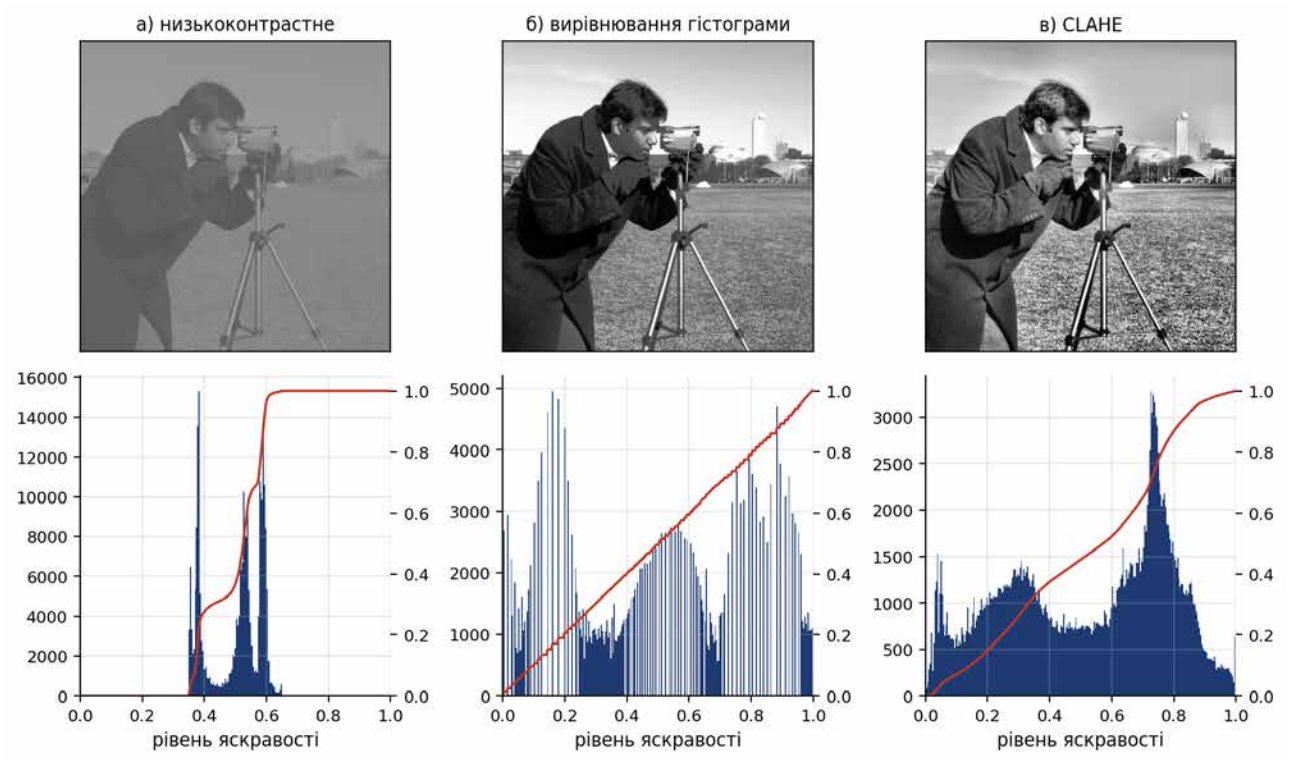


Рис. 7.2. Вирівнювання гістограми: зображення (вгорі), гістограми і кумулятивні гістограми (внизу)

7.3. Просторова фільтрація

7.3.1. Лінійні фільтри

Лінійна просторова фільтрація — це дискретна двовимірна згортка (3.11) зображення з маскою (ядром) w розміром $(2a+1)\times(2b+1)$:

$$g(m,n)=\sum_{i=-a}^{a}\sum_{j=-b}^{b} w\,(i,j)\,f(m+i,n+j). \tag{7.4}$$

(Для симетричних масок згортка і кореляція збігаються, тому знак індексів несуттєвий.) На межах зображення маска виходить за його межі; ці пікселі доповнюють нулями, повторенням крайових значень, дзеркальним відображенням або циклічно — аналогічно до вибору між лінійною і циклічною моделлю згортки в розділі 3.

Усереднювальний фільтр (фільтр ковзного середнього) має маску з однакових коефіцієнтів, наприклад для 3×3:

$$w=\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}. \tag{7.5}$$

Усереднення K незалежних значень білого шуму зменшує його дисперсію в K разів (для маски 5×5 — у 25 разів, тобто середньоквадратичне відхилення — у 5 разів). Однак разом із шумом розмиваються контури і дрібні деталі. Крім того, частотна характеристика прямокутної маски має бічні пелюстки, тому усереднювальний фільтр не повністю придушує високі частоти.

Гаусівський фільтр з ваговими коефіцієнтами $w(i,j)\propto \exp\{-(i^2+j^2)/(2\sigma^2)\}$ надає більшої ваги центральним пікселям, не має бічних пелюсток у частотній характеристиці (див. (4.29)) і є сепарабельним. Параметр σ безпосередньо визначає масштаб згладжування; розмір маски зазвичай вибирають рівним $\lceil 6\sigma\rceil$ (непарне число). Гаусівський фільтр — основа багатомасштабних методів аналізу (детектор Канні, оператор LoG, масштабний простір) [15, 28].

7.3.2. Нелінійні фільтри

Лінійні фільтри погано справляються з імпульсним шумом: одне значення «солі» з амплітудою 1 після усереднення 5×5 перетворюється на пляму з амплітудою $1/25$, яка розмиває деталі, але не зникає. Медіанний фільтр замінює значення пікселя медіаною значень у вікні:

$$g(m,n)=\operatorname{median}\{f(m+i,n+j):\ (i,j)\in \Omega\}. \tag{7.6}$$

Якщо кількість імпульсних завад у вікні менша за половину кількості пікселів, вони повністю усуваються, оскільки опиняються на краях упорядкованого ряду значень. Медіанний фільтр зберігає різкі перепади яскравості (монотонна межа не змінюється), але знищує тонкі лінії і дрібні деталі, ширина яких менша за половину вікна, та закруглює кути об’єктів. Медіанна фільтрація належить до більш загального класу рангових фільтрів:

вибір мінімального елемента ряду дає ерозію, максимального — дилатацію сірої морфології (підрозділ 7.6) [13].

Білатеральний фільтр [51] — зважене середнє, у якому вага сусіднього пікселя залежить не лише від відстані до центра (як у гаусівському фільтрі), а й від різниці яскравостей:

$$g(\mathrm{p}) = \frac{1}{K_\mathrm{p}} \sum_{\mathrm{q}\in\Omega} G_{\sigma_s}(\| \mathrm{p} - \mathrm{q} \|)\, G_{\sigma_r}(|f(\mathrm{p}) - f(\mathrm{q})|)\, f(\mathrm{q}), \tag{7.7}$$

де K_p — нормувальний множник. Пікселі по інший бік контуру мають велику різницю яскравостей і отримують малу вагу, тому контури зберігаються, а однорідні ділянки згладжуються.

Порівняння фільтрів наведено на рис. 7.3. Для гаусівського шуму (верхній ряд) усі фільтри знижують рівень шуму, найкраще зберігає деталі білатеральний [51]. Для імпульсного шуму (нижній ряд) медіанний фільтр 3×3 практично повністю усуває заваду, тоді як лінійні фільтри лише розмазують її.



Рис. 7.3. Згладжування зображень з гаусівським (вгорі) та імпульсним (внизу) шумом

7.3.3. Підвищення різкості

Розмиття зображення — це, по суті, усереднення (інтегрування). Відповідно, підвищення різкості досягається операціями, аналогічними диференціюванню. Найпростішим ізотропним диференціальним оператором другого порядку є лапласіан

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}, \tag{7.8}$$

дискретні апроксимації якого задаються масками

$$w_4 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad w_8 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{7.9}$$

Зверніть увагу, що одновимірна маска $[1, -2, 1]$, з якої складається w_4, — це друга різниця, та сама, що використовується як оператор «гладкості» в методі умовної деконволюції (розділ 8) [2]. Там вона штрафує нерівномірність оцінки, тут — підкреслює її.

Підвищення різкості за допомогою лапласіана виконується за формулою

$$g(m,n) = f(m,n) - c\,\nabla^2 f(m,n), \qquad c > 0. \tag{7.10}$$

Віднімання лапласіана посилює перепади: з темного боку контуру з'являється темніша смуга, зі світлого — світліша, що візуально сприймається як збільшення різкості (рис. 7.4).

Нерізке маскування (англ. *unsharp masking*) — класичний прийом, що походить з фотолабораторної практики [13, 15]: від зображення віднімають його розмиту копію, отримуючи «маску» високочастотних деталей, і додають її з коефіцієнтом k:

$$g = f + k\,(f - f * h_\sigma), \tag{7.11}$$

де h_σ — гаусівське ядро. При $k > 1$ метод називають підсиленням високих частот. Аналогічна операція в частотній області — (4.30).

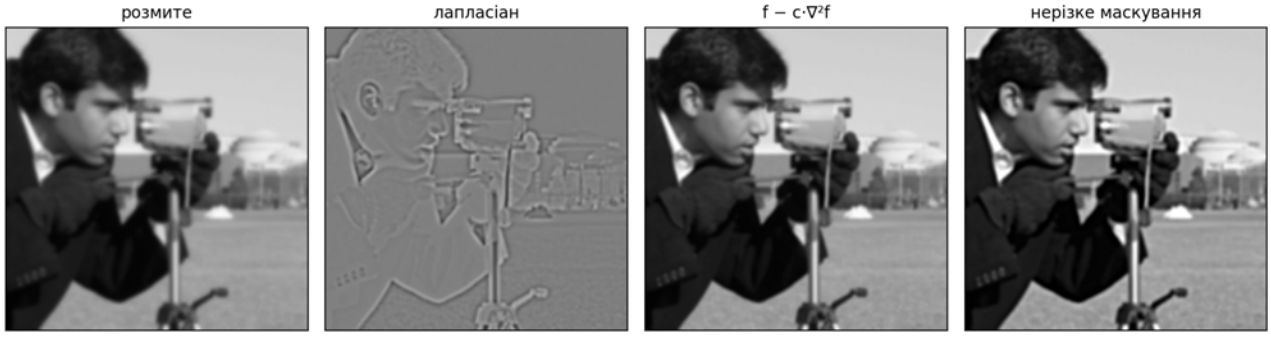


Рис. 7.4. Підвищення різкості розмитого зображення

Важливо розуміти обмеження цих методів. Підвищення різкості лише підкреслює деталі, які *вже присутні* в зображенні з помітним контрастом; воно не відновлює деталей, втрачених через розмиття, і неминуче підсилює шум. Відновлення втрачених деталей з урахуванням моделі спотворення і рівня шуму — задача реставрації, що розглядається в розділах 8 і 9.

7.4. Виділення контурів

7.4.1. Градієнтні оператори

Контури (межі об'єктів) — найбільш інформативні елементи зображення для розпізнавання форми. Контуру відповідає різка зміна яскравості, тобто велике значення модуля градієнта

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]^T, \qquad |\nabla f| = \sqrt{G_x^2 + G_y^2}, \qquad \theta = \arctan\frac{G_y}{G_x}. \tag{7.12}$$

Напрямок градієнта θ перпендикулярний до контуру. Похідні апроксимують різницями, які обчислюються згорткою з невеликими масками:

- оператор Робертса використовує діагональні різниці в околі 2×2: $G_1 = f(m,n) - f(m+1,n+1)$, $G_2 = f(m+1,n) - f(m,n+1)$; він найпростіший, але найбільш чутливий до шуму;
- оператор Превітт поєднує різницю з усередненням у перпендикулярному напрямку:

$$w_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \qquad w_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}; \tag{7.13}$$

- оператор Собеля надає центральному рядку (стовпцю) подвійної ваги, що забезпечує кращу ізотропність і стійкість до шуму [13, 15]:

$$w_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad w_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{7.14}$$

Маска Собеля сепарабельна: $w_x = [1,2,1]^T \cdot [-1,0,1]$, тобто це згладжування в одному напрямку і диференціювання в іншому.

7.4.2. Оператор Лапласа — Гаусса

Лапласіан (7.8) ізотропний, але дуже чутливий до шуму (друга похідна підсилює високі частоти сильніше, ніж перша). Д. Марр і Е. Хілдрет запропонували спочатку згладжувати зображення гаусівським фільтром, а потім обчислювати лапласіан. Завдяки лінійності ці операції можна об'єднати в один оператор LoG (*Laplacian of Gaussian*) [13, 28]:

$$\nabla^2 G_\sigma(x,y) = \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6}\, e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{7.15}$$

Профіль цього оператора з оберненим знаком — «мексиканський капелюх» (рис. 6.4). Контури відповідають переходам через нуль відгуку LoG, причому параметр σ задає масштаб деталей, що виділяються. Оператор LoG часто наближають різницею двох гаусівських функцій (DoG), яка лежить в основі популярного детектора особливих точок SIFT [52].

7.4.3. Детектор Канні

Дж. Канні [15] сформулював три критерії оптимального детектора контурів: висока ймовірність виявлення справжніх контурів і низька — хибних; точна локалізація; єдиний відгук на кожен контур. Розв'язання відповідної оптимізаційної задачі для ступінчастого перепаду на фоні білого шуму дає фільтр, близький до першої похідної гаусівської функції. Алгоритм Канні складається з етапів:

1. згладжування гаусівським фільтром з параметром σ;
2. обчислення модуля і напрямку градієнта (наприклад, операторами Собеля);

3. придушення немаксимумів: піксель залишається кандидатом на контур, лише якщо модуль градієнта в ньому не менший, ніж у двох сусідах уздовж напрямку градієнта; це стоншує контури до одного пікселя;
4. подвійна порогова обробка з гістерезисом: пікселі з модулем градієнта вище верхнього порогу вважаються контурними; пікселі між нижнім і верхнім порогами — контурними лише за умови зв'язності з «сильними» пікселями.

Детектор Канні дає тонкі, неперервні контури і є стандартом де-факто в задачах виділення меж. Порівняння розглянутих операторів наведено на рис. 7.5.

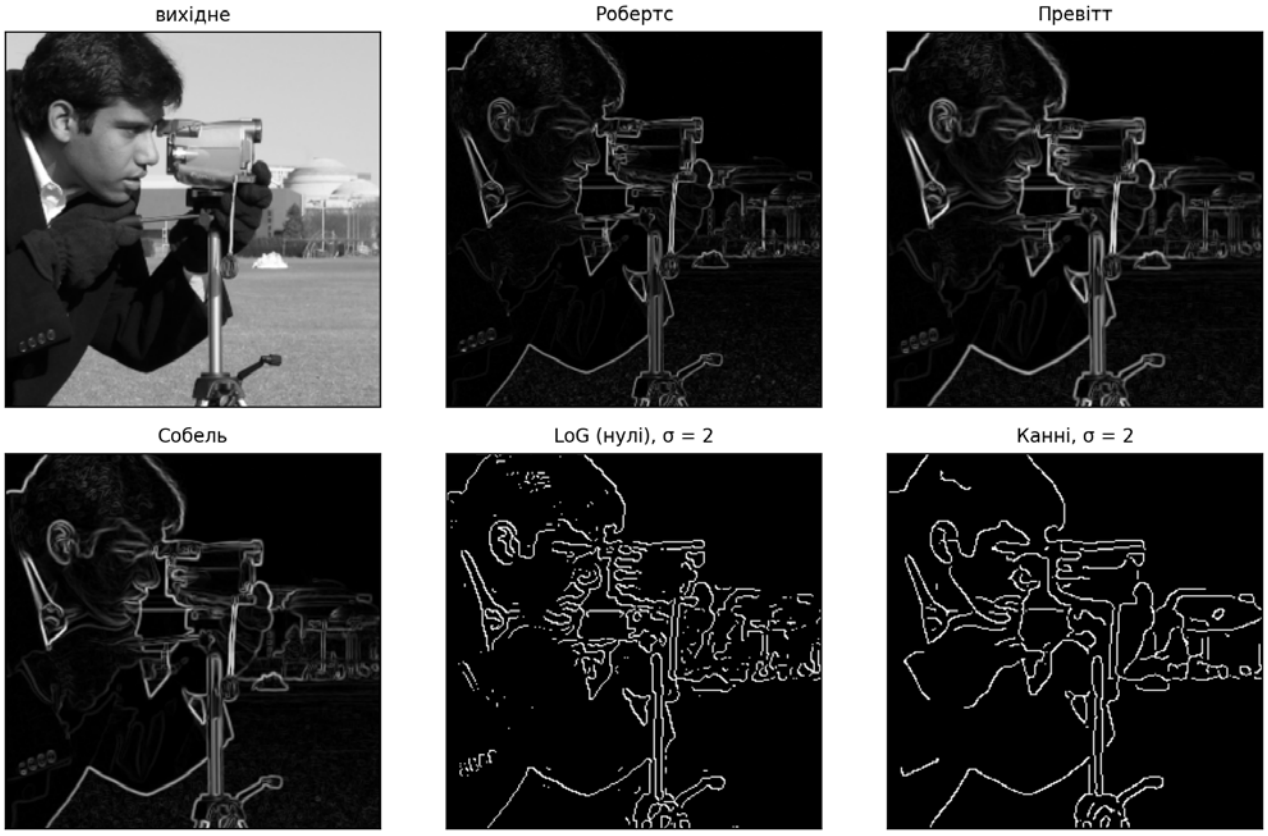


Рис. 7.5. Виділення контурів різними операторами

Виділені контури є основою для обчислення ознак форми (розділ 11), пошуку прямих і кіл перетворенням Хафа [13], суміщення зображень за характерними точками [47, 52].

7.5. Порогова сегментація

7.5.1. Глобальний поріг

Сегментація — поділ зображення на області, що відповідають окремим об'єктам або їх частинам, — є, мабуть, найважливішим і найскладнішим етапом попереднього оброблення: від її якості безпосередньо залежать ознаки форми і, отже, результат розпізнавання. Найпростіший метод сегментації — порогова обробка:

$$b(m,n) = \begin{cases} 1, & f(m,n) > T, \\ 0, & f(m,n) \le T, \end{cases} \tag{7.16}$$

яка ефективна, коли об’єкти і фон суттєво відрізняються за яскравістю, а гістограма має два виражені максимуми (моди), розділені западиною. Ключове питання — вибір порогу T.

7.5.2. Метод Оцу

Метод Оцу [13] обирає поріг, що максимізує розділення двох класів пікселів у статистичному сенсі. Нехай поріг T ділить пікселі на класи C_0 ($r \le T$) і C_1 ($r > T$) з імовірностями $\omega_0(T)$, $\omega_1(T) = 1 - \omega_0(T)$ і середніми $\mu_0(T)$, $\mu_1(T)$. Загальна дисперсія σ^2 розкладається на внутрішньокласову і міжкласову:

$$\sigma^2 = \sigma_W^2(T) + \sigma_B^2(T), \qquad \sigma_B^2(T) = \omega_0(T)\,\omega_1(T)[\mu_0(T) - \mu_1(T)]^2. \tag{7.17}$$

Оскільки загальна дисперсія від порогу не залежить, мінімізація внутрішньокласової дисперсії еквівалентна максимізації міжкласової:

$$T^* = \arg\max_T \sigma_B^2(T). \tag{7.18}$$

Обчислення σ_B^2 для всіх L можливих порогів виконується за один прохід гістограми з використанням кумулятивних сум. Метод Оцу по суті розв’язує задачу кластеризації яскравостей на два класи і є прикладом того, як критерій розділення класів (подібний до критерію Фішера, розділ 11) застосовується вже на етапі попереднього оброблення. Приклад наведено на рис. 7.6, а–в.

7.5.3. Адаптивні пороги

Глобальний поріг непридатний, якщо освітленість нерівномірна: на рис. 7.6, д частина тексту, що потрапила в тінь, перетворилася на суцільну темну пляму. Адаптивний (локальний) поріг обчислюється для кожного пікселя за статистиками його околу, наприклад

$$T(m,n) = \mu_\Omega(m,n) - C \quad \text{або} \quad T(m,n) = \mu_\Omega(m,n)\left[1 + k\left(\frac{\sigma_\Omega(m,n)}{R} - 1\right)\right], \tag{7.19}$$

де μ_Ω, σ_Ω — локальні середнє і середньоквадратичне відхилення у вікні; C, k, R — параметри (друга формула — метод Саволи, поширений для документів [15]). Розмір вікна має бути більшим за розміри символів, але меншим за масштаб зміни освітленості. Результат адаптивної обробки (рис. 7.6, е) практично не залежить від нерівномірності освітлення. Альтернативою є попередня компенсація освітленості — віднімання фону, оціненого сильним згладжуванням або морфологічним закриттям, або гомоморфна фільтрація (розділ 4).

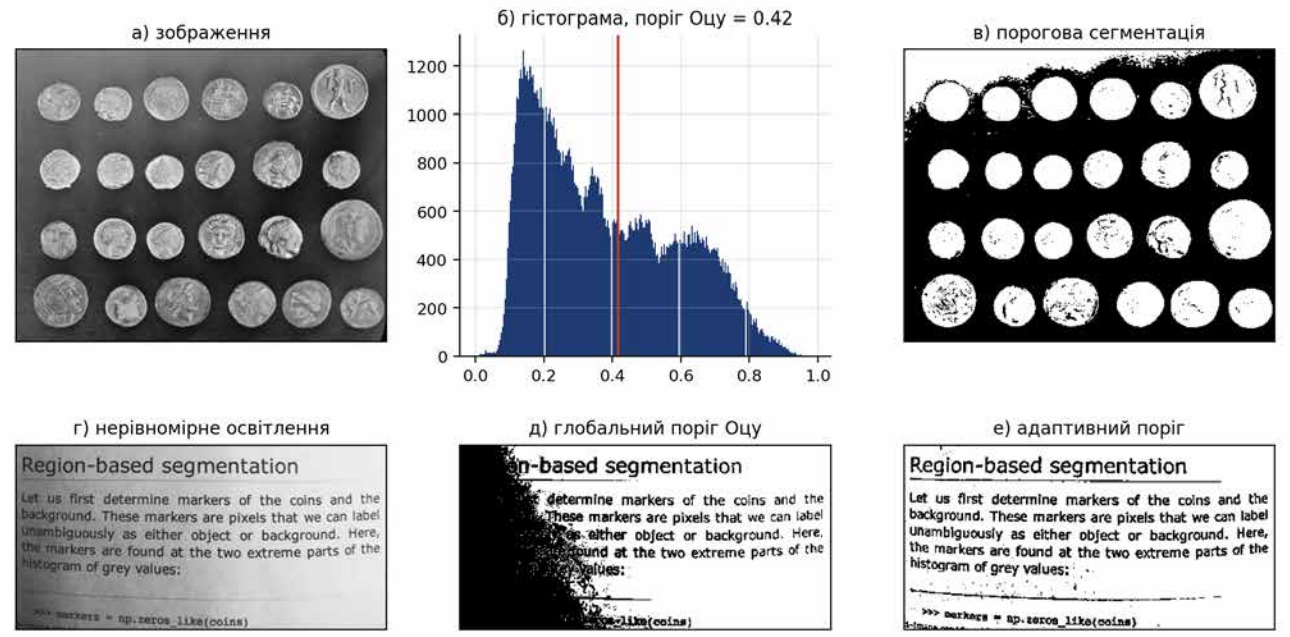


Рис. 7.6. Порогова сегментація: метод Оцу (вгорі) та адаптивний поріг при нерівномірному освітленні (внизу)

Крім порогових, існують методи сегментації нарощуванням областей (від «зерен» до однорідних областей), водорозділу (англ. *watershed*), що розглядає зображення як рельєф, кластеризації ознак пікселів (яскравості, кольору, текстури), а також методи, засновані на нейронних мережах [13, 28]. Вони виходять за межі цього розділу.

7.6. Математична морфологія

7.6.1. Базові операції над бінарними образами

Результат порогової сегментації зазвичай містить дефекти: дрібні хибні об’єкти від шуму, дірки всередині об’єктів, розриви тонких з’єднань, злиття сусідніх об’єктів. Для їх виправлення використовують операції математичної морфології [53], що аналізують форму об’єктів за допомогою невеликого еталонного образу — структурного елемента B (зазвичай диск, квадрат або хрест).

Нехай A — множина пікселів об’єкта, B_z — структурний елемент, зсунутий у точку z. Основні операції:

- ерозія — множина точок, у яких структурний елемент повністю вміщується в об’єкт:

$$A \ominus B = \{z:\ B_z \subseteq A\}; \tag{7.20}$$

- дилатація — множина точок, у яких (віддзеркалений) структурний елемент перетинається з об’єктом:

$$A \oplus B = \{z:\ \hat{B}_z \cap A \neq \varnothing\}. \tag{7.21}$$

Ерозія «стоншує» об’єкт, видаляє дрібні об’єкти і тонкі перемички, збільшує дірки; дилатація «потовщує» об’єкт, заповнює дрібні дірки і щілини, може злити близькі об’єкти. Їх комбінації:

- відкриття $A \circ B = (A \ominus B) \oplus B$ — видаляє дрібні об’єкти і виступи, менші за структурний елемент, згладжує контури зовні, розриває тонкі перемички, практично не змінюючи розмірів великих об’єктів;
- закриття $A \bullet B = (A \oplus B) \ominus B$ — заповнює дрібні дірки і заглибини, з’єднує близькі фрагменти.

Дію операцій на результат порогової сегментації зображення монет показано на рис. 7.7. Відкриття видаляє шумові фрагменти, закриття разом із заповненням дірок дає суцільні силуети монет, після чого маркування зв’язних компонент (розділ 2) дозволяє полічити об’єкти і обчислити ознаки кожного з них. Зауважимо, що об’єкти, які стикаються, після закриття можуть злитися; для їх розділення використовують ерозію з подальшим маркуванням або метод водорозділу на перетворенні відстані.

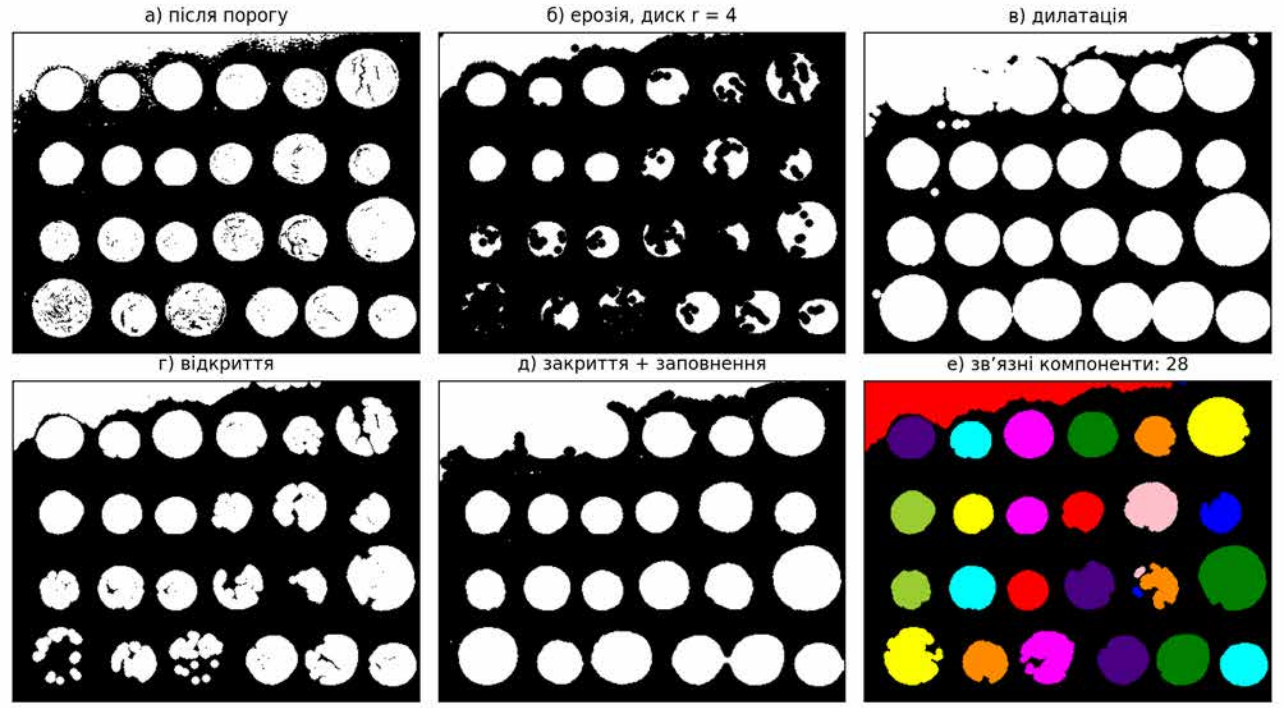


Рис. 7.7. Морфологічні операції над бінарним образом

7.6.2. Перетворення відстані і скелет

Перетворення відстані зіставляє кожному пікселю об’єкта відстань до найближчого пікселя фону (у метриці D_E, D_4 або D_8, розділ 2). Скелет (серединна вісь) об’єкта — множина центрів максимальних вписаних кіл; він зберігає топологію і загальну форму об’єкта, будучи лінією завтовшки в один піксель (рис. 7.8). Скелет використовується для аналізу форми тонких об’єктів (символів, судин, тріщин), обчислення довжини, кількості гілок і кінцевих точок як ознак [13, 53].



Рис. 7.8. Перетворення відстані та скелет бінарного об’єкта

7.6.3. Морфологія півтонових зображень

Операції ерозії та дилатації узагальнюються на півтонові зображення: ерозія замінює значення пікселя мінімумом у вікні структурного елемента, дилатація — максимумом. Різниця між зображенням і його відкриттям (перетворення «циліндр», англ. *top-hat*) виділяє світлі деталі, менші за структурний елемент, на нерівномірному фоні; різниця між дилатацією і ерозією (морфологічний градієнт) підкреслює контури [53].

7.7. Геометричні перетворення та нормування образів

7.7.1. Афінні перетворення

Для того щоб ознаки не залежали від положення, орієнтації і розміру об’єкта, образ часто приводять до стандартного вигляду — нормують. Найважливішим класом геометричних перетворень є афінні [13, 28]:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \tag{7.22}$$

що включають зсув, поворот, зміну масштабу і скіс. Типова процедура нормування образу об’єкта:

1. зсув центра ваги об’єкта в центр кадру (центр ваги обчислюється через моменти першого порядку, розділ 11);
2. поворот так, щоб головна вісь інерції (власний вектор матриці центральних моментів другого порядку) збігалася з горизонтальною віссю;
3. масштабування до стандартного розміру (за площею або за розміром охоплюючого прямокутника).

7.7.2. Інтерполяція

При геометричному перетворенні координати вихідних пікселів, як правило, потрапляють між вузлами вихідної сітки, тому значення яскравості доводиться інтерполювати. Розрахунок ведуть «назад»: для кожного пікселя результату обчислюють відповідну точку вихідного зображення за оберненим перетворенням і інтерполюють значення в ній. Методи інтерполяції (рис. 7.9) [15]:

- найближчого сусіда — найпростіший, зберігає вихідні значення (важливо для бінарних зображень і зображень міток), але дає «сходинки»;

- білінійна — зважене середнє чотирьох сусідніх пікселів з вагами, пропорційними площам протилежних прямокутників; дає неперервний результат, дещо розмиває;
- бікубічна — використовує 16 сусідів і кубічні сплайни; дає гладкіший і різкіший результат, але може створювати невеликі викиди біля контурів.

З точки зору теорії відліків (розділ 2) ідеальною є інтерполяція функцією sinc (2.7); перелічені методи є її наближеннями з короткими ядрами [43]. При зменшенні зображення перед проріджуванням необхідна низькочастотна фільтрація, інакше виникає аліасинг (рис. 2.3).

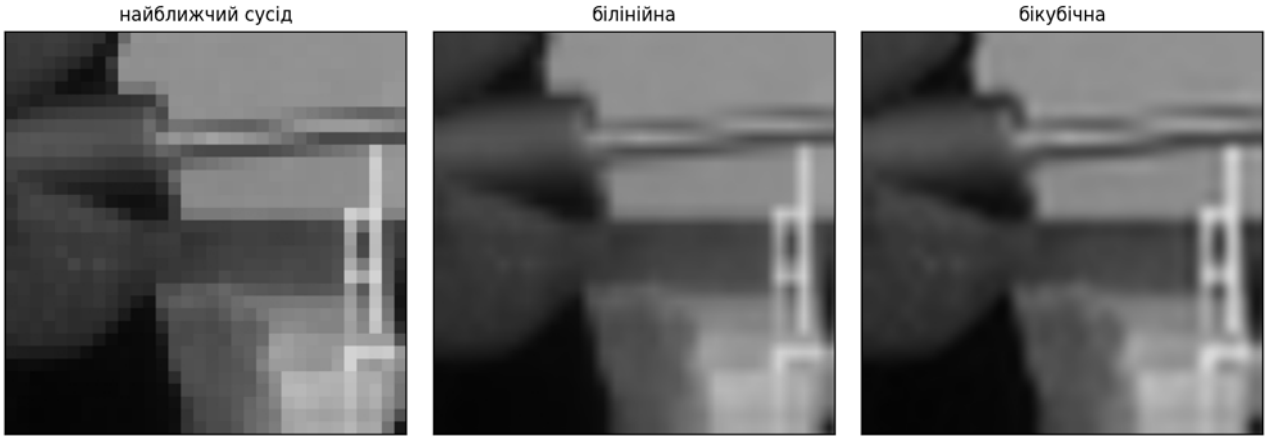


Рис. 7.9. Збільшення фрагмента зображення у 8 разів з різними методами інтерполяції

7.8. Програмна реалізація

Нижче наведено типовий конвеєр попереднього оброблення півтонового зображення перед розпізнаванням об’єктів: корекція контрасту, придушення шуму, порогова сегментація методом Оцу, морфологічне очищення, маркування зв’язних компонент та обчислення найпростіших ознак кожного об’єкта.

```
import numpy as np
from scipy import ndimage as ndi
from skimage import data, exposure, filters, morphology, measure, feature,
img_as_float

f = img_as_float(data.coins())

# 1. корекція контрасту (CLAHE)
f1 = exposure.equalize_adapthist(f, clip_limit=0.02)
# 2. придушення шуму: медіанний фільтр 3x3
f2 = ndi.median_filter(f1, size=3)
# 3. порогова сегментація: метод Оцу (7.18)
T = filters.threshold_otsu(f2)
b = f2 > T
# 4. морфологічне очищення: відкриття, закриття, заповнення дірок
se = morphology.disk(3)
b = morphology.opening(b, se)
```

```
b = morphology.closing(b, se)
b = ndi.binary_fill_holes(b)
b = morphology.remove_small_objects(b, min_size=200)
# 5. маркування і ознаки об'єктів
lab = measure.label(b, connectivity=2)
for r in measure.regionprops(lab):
    print(r.label, r.area, round(r.eccentricity, 3), r.centroid)

# контури: Собель і Канні
grad = filters.sobel(f)
edges = feature.canny(f, sigma=2, low_threshold=0.05, high_threshold=0.15)
```

7.9. Приклади розв'язання задач

Задача 7.1. Зображення має 8 рівнів яскравості з кількостями пікселів $h = [10,20,40,15,10,5,0,0]$. Виконайте вирівнювання гістограми.

Розв'язання. Загальна кількість пікселів — 100. Нормована гістограма $p = [0{,}10;\ 0{,}20;\ 0{,}40;\ 0{,}15;\ 0{,}10;\ 0{,}05;\ 0;\ 0]$, кумулятивна сума $[0{,}10;\ 0{,}30;\ 0{,}70;\ 0{,}85;\ 0{,}95;\ 1;\ 1;\ 1]$. За (7.3) $s_k = \mathrm{round}(7 \cdot \mathrm{cdf}_k)$: $[1,2,5,6,7,7,7,7]$. Отже, рівні 0, 1, 2, 3 відображаються в 1, 2, 5, 6, а рівні 4 і 5 зливаються в 7. Найчастіший рівень 2 «відсунувся» від сусідніх, що підвищило контраст у найбільш заселеній частині діапазону.

Задача 7.2. У вікні 3×3 значення пікселів: 12, 15, 255, 14, 13, 0, 16, 14, 15 (центральний — 13). Порівняйте результати усереднювального і медіанного фільтрів.

Розв'язання. Середнє значення $(12 + 15 + 255 + 14 + 13 + 0 + 16 + 14 + 15)/9 = 354/9 \approx 39{,}3$ — імпульсні завади 255 і 0 сильно спотворили результат. Упорядкований ряд: 0, 12, 13, 14, 14, 15, 15, 16, 255; медіана дорівнює 14, що близько до справжнього рівня фону. Медіанний фільтр повністю ігнорує обидва викиди.

Задача 7.3. Обчисліть відгук горизонтальної маски Собеля на вертикальному перепаді яскравості від 10 до 50.

Розв'язання. Якщо маска (7.14) розташована так, що лівий стовпець вікна має значення 10, а правий — 50, то $G_x = (-1 - 2 - 1) \cdot 10 + (1 + 2 + 1) \cdot 50 = -40 + 200 = 160$. Такий самий відгук отримуємо, якщо центральний стовпець вікна належить як до лівої, так і до правої ділянки (перепад між центральним і правим або між лівим і центральним стовпцями). Отже, ширина відгуку — два пікселі, а його величина в 4 рази більша за величину перепаду (сума ваг одного стовпця маски). Тому перед застосуванням порогу до модуля градієнта необхідно враховувати нормувальний множник маски.

Висновки до розділу 7

1. Попереднє оброблення в просторовій області включає точкові, локальні і геометричні перетворення; більшість його методів евристичні і не потребують моделі спотворення.

2. Точкові перетворення (негатив, логарифмічне, гамма, розтягнення) змінюють розподіл яскравостей; вирівнювання гістограми автоматично будує перетворення, що наближає розподіл до рівномірного; адаптивне вирівнювання з обмеженням контрасту (CLAHE) враховує локальні особливості і обмежує підсилення шуму.
3. Лінійні фільтри згладжування (усереднювальний, гаусівський) зменшують гаусівський шум, але розмивають контури; нелінійні — медіанний, білатеральний — краще зберігають межі, а медіанний ефективно усуває імпульсний шум.
4. Підвищення різкості (віднімання лапласіана, нерізке маскування) підкреслює наявні деталі, але не відновлює втрачені і підсилює шум.
5. Контури виділяють градієнтними операторами (Робертса, Превітт, Собеля), оператором Лапласа — Гаусса або оптимальним детектором Канні.
6. Порогова сегментація з вибором порогу методом Оцу максимізує міжкласову дисперсію; при нерівномірному освітленні необхідні адаптивні пороги.
7. Операції математичної морфології (ерозія, дилатація, відкриття, закриття, скелетизація) виправляють дефекти сегментації і служать для аналізу форми.
8. Геометричне нормування з інтерполяцією приводить образи до стандартного положення, орієнтації і розміру, що спрощує подальше розпізнавання.

Питання для обговорення та самоперевірки

1. Чим відрізняються точкові, локальні і глобальні методи оброблення зображень?
2. Для чого використовують гамма-корекцію? Чому її слід враховувати перед реставрацією?
3. Виведіть перетворення, що вирівнює гістограму. Чому в дискретному випадку гістограма лише наближено рівномірна?
4. Що таке специфікація гістограми і де вона застосовується?
5. Чим CLAHE відрізняється від глобального вирівнювання гістограми?
6. Як обробляються межі зображення при просторовій фільтрації?
7. У скільки разів усереднювальний фільтр 5×5 зменшує середньоквадратичне відхилення білого шуму?
8. Чому гаусівський фільтр кращий за усереднювальний?
9. Поясніть принцип роботи медіанного фільтра. Чому він ефективний для імпульсного шуму?
10. Як працює білатеральний фільтр?
11. Запишіть маски лапласіана. Як з їх допомогою підвищують різкість?
12. Що таке нерізке маскування?
13. Запишіть маски операторів Превітт і Собеля. Чому маска Собеля сепарабельна?
14. Що таке оператор LoG? Як за його допомогою знаходять контури?

15. Опишіть етапи алгоритму Канні. Навіщо потрібні придушення немаксимумів і гістерезис?
16. Сформулюйте критерій вибору порогу в методі Оцу.
17. Коли необхідні адаптивні пороги? Як вони обчислюються?
18. Дайте визначення ерозії, дилатації, відкриття і закриття. Яку дію вони мають на бінарний образ?
19. Що таке скелет об’єкта? Для чого він використовується?
20. Які методи інтерполяції використовують при геометричних перетвореннях? Порівняйте їх.

Завдання для самостійної роботи

1. Реалізуйте вирівнювання гістограми за формулою (7.3) і порівняйте з бібліотечною функцією. Виконайте специфікацію гістограми одного зображення під гістограму іншого.
2. Додайте до зображення імпульсний шум різної щільності (2–30 %). Порівняйте PSNR після медіанної фільтрації вікнами 3×3, 5×5 і адаптивної медіанної фільтрації.
3. Реалізуйте білатеральний фільтр за формулою (7.7). Дослідіть вплив параметрів σ_s і σ_r.
4. Реалізуйте оператори Робертса, Превітт, Собеля через згортку і порівняйте їх стійкість до шуму на синтетичному зображенні з відомими контурами.
5. Реалізуйте метод Оцу за формулами (7.17), (7.18) і порівняйте з бібліотечною функцією.
6. Для зображення документа з нерівномірним освітленням порівняйте глобальний поріг Оцу, адаптивний поріг (7.19) та поріг Оцу після віднімання фону, оціненого морфологічним закриттям.
7. Побудуйте конвеєр підрахунку об’єктів (клітин, монет, зерен) на зображенні. Оцініть, як помилки на кожному етапі впливають на кінцевий результат.
8. Реалізуйте геометричне нормування бінарного силуету (зсув центра ваги, поворот за головною віссю, масштабування) і перевірте його на повернутих і масштабованих копіях образу.

РОЗДІЛ 8. РЕСТАВРАЦІЯ ОБРАЗІВ: ІНВЕРСНА І ВІНЕРІВСЬКА ФІЛЬТРАЦІЯ, УМОВНА ДЕКОНВОЛЮЦІЯ

Навчальні цілі. Після вивчення розділу ви зможете: формулювати задачу реставрації як обернену задачу; пояснювати нестійкість інверсної фільтрації і способи її подолання; виводити фільтр Вінера; формулювати і розв'язувати задачу умовної деконволюції в натуральних координатах і в області трансформант Фур'є; ітераційно визначати параметр регуляризації за енергією шуму; порівнювати фільтр умовної деконволюції з фільтром Вінера і регуляризацією Тихонова; записувати узагальнений фільтр найменших квадратів у довільному ортогональному базисі; пояснювати метод деконволюції за критерієм форми результуючої імпульсної характеристики.

8.1. Постановка задачі реставрації

У розділі 3 було сформульовано лінійну модель деградації образу з адитивним шумом. Для дискретних образів у матричній формі вона має вигляд

$$\mathrm{L} = \mathrm{G}\,\mathrm{f} + \mathrm{V}, \tag{8.1}$$

де f — стовпець відліків вихідного (неспотвореного) образу; G — матричний дискретний оператор згортки, утворений відліками імпульсної характеристики системи відображення; V — стовпець відліків реалізації адитивного шуму; L — стовпець відліків спостережуваного (деградованого) образу. Для двовимірних образів вектори утворюються розгортанням матриць рядок за рядком.

Задача реставрації полягає в тому, щоб за відомими (апріорі) L, G і деякими характеристиками шуму V знайти оцінку $\hat{\mathrm{f}}$, яка за обраним критерієм найближча до f [2, 33]. Оскільки реставрація є операцією, оберненою до згортки (конволюції), її називають деконволюцією.

У розділі 3 (підрозділ 3.7) показано, що ця задача некоректно поставлена: формальне обернення оператора G катастрофічно підсилює шум. Тому всі практичні методи реставрації так чи інакше залучають додаткову апріорну інформацію — про статистику образів і шуму, про гладкість шуканої оцінки, про енергію шуму, про форму бажаної результуючої імпульсної характеристики. Методи відрізняються саме видом і обсягом цієї інформації, а також тим, наскільки точно описуються окремі члени моделі (8.1) [2]. У цьому розділі ми розглянемо послідовно:

- інверсну і псевдоінверсну фільтрацію, що не використовують апріорної інформації про шум;
- фільтрацію Вінера, яка потребує знання спектрів потужності образу і шуму;
- метод умовної деконволюції, що потребує лише енергії шуму та загального припущення про гладкість образу;
- узагальнену фільтрацію за методом найменших квадратів у довільному ортогональному базисі;
- деконволюцію за критерієм форми результуючої імпульсної характеристики.

8.2. Інверсна фільтрація

8.2.1. Реставрація в області трансформант Фур'є

Нехай оператор згортки G циркулянтний (циклічна модель). Тоді згідно з (4.21) перехід в область трансформант ДПФ діагоналізує модель (8.1) [2, 13]:

$$L(k) = G(k)\,F(k) + V(k), \tag{8.2}$$

де $L(k)$, $F(k)$, $V(k)$ — ДПФ відповідних векторів, $G(k)$ — частотна характеристика системи (ДПФ імпульсної характеристики). Для двовимірних образів k означає пару частот (u, v). Інверсний фільтр обчислює оцінку спектра як

$$\hat{F}(k) = \frac{L(k)}{G(k)} = F(k) + \frac{V(k)}{G(k)}. \tag{8.3}$$

За відсутності шуму інверсна фільтрація точно відновлює образ (за умови, що $G(k) \neq 0$ для всіх k). На рис. 8.1, б наведено такий результат для гаусівського розмиття: PSNR перевищує 99 дБ, тобто відновлення точне з точністю до похибок округлення. Проте вже шум із середньоквадратичним відхиленням 0,001 (0,1 % від діапазону яскравостей, тобто значно менший за шум квантування 8-розрядного зображення) повністю руйнує результат (рис. 8.1, в).

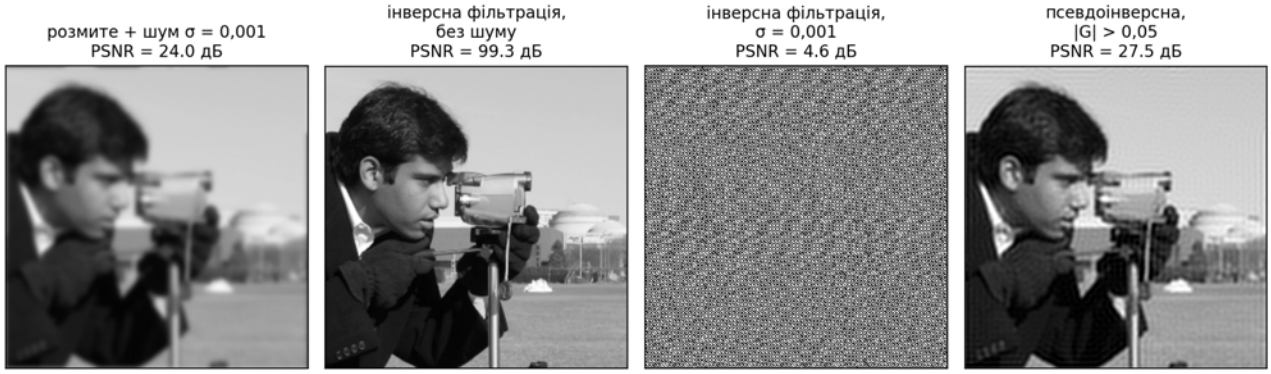


Рис. 8.1. Інверсна та псевдоінверсна фільтрація розмитого зображення

Причина зрозуміла з рис. 8.2, а: модуль частотної характеристики гаусівського розмиття спадає з частотою дуже швидко, на високих частотах він становить 10^{-5} і менше. Модуль коефіцієнта передачі інверсного фільтра $1/|G(k)|$ відповідно зростає до 10^5 і більше. Спектр білого шуму рівномірний, тому на високих частотах доданок $V(k)/G(k)$ у (8.3) у тисячі разів перевищує корисну складову $F(k)$, спектр якої для природних образів, навпаки, спадає з частотою. Це спектральний аналог явища, описаного через сингулярні числа в розділі 3 [33].

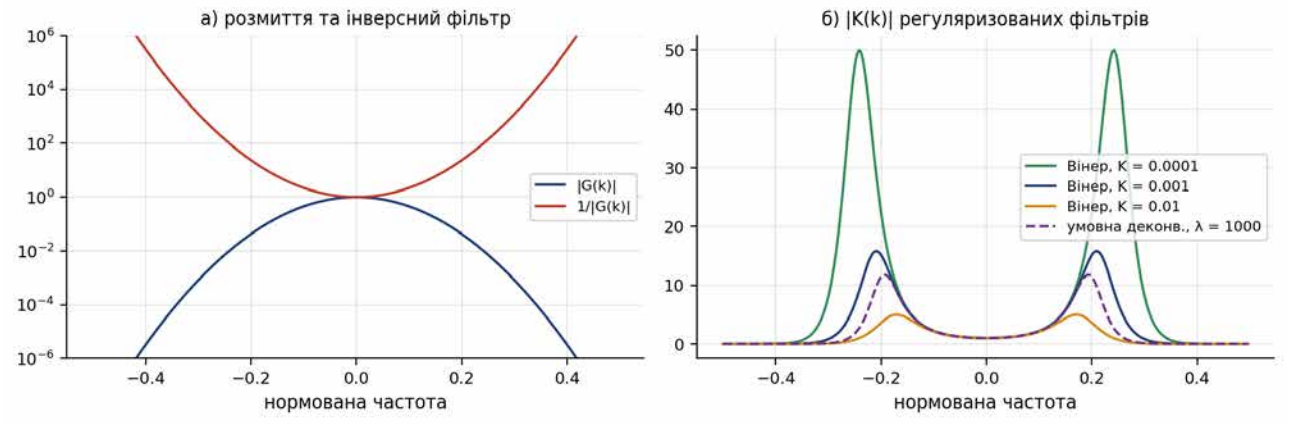


Рис. 8.2. Частотні характеристики розмиття, інверсного фільтра (а) та регуляризованих фільтрів (б)

8.2.2. Псевдоінверсна фільтрація

Найпростіший спосіб обмежити підсилення шуму — не обертати частотну характеристику там, де вона мала:

$$K_{PI}(k) = \begin{cases} 1/G(k), & |G(k)| > \delta, \\ 0, & |G(k)| \le \delta, \end{cases} \tag{8.4}$$

або обмежити частоти деякою смугою. Це псевдоінверсний фільтр (у матричній формі — псевдообернення з відкиданням малих сингулярних чисел, тобто усічений сингулярний розклад) [45]. Як видно з рис. 8.1, г, результат уже придатний для використання, проте має два недоліки. По-перше, поріг δ доводиться підбирати вручну, і оптимальне його значення залежить від рівня шуму і від образу. По-друге, різке обрізання спектра породжує ефект Гіббса — характерний «дзвін» біля контурів і текстуру з хвиль на однорідних ділянках (див. рис. 8.6, в). Ці недоліки обумовлюють потребу в методах, що плавно і обґрунтовано зменшують коефіцієнт передачі там, де шум переважає сигнал.

8.3. Фільтр Вінера

8.3.1. Виведення

Фільтр Вінера [13] мінімізує середньоквадратичну похибку оцінки за умови, що образ і шум є реалізаціями некорельованих однорідних випадкових полів з відомими спектральними густинами потужності $S_{ff}(k)$ і $S_{VV}(k)$. Шукатимемо оцінку у вигляді лінійного фільтра в частотній області: $\hat{F}(k) = K(k)\,L(k)$. Середньоквадратична похибка для частоти k:

$$\mathrm{M}\{|F - KL|^2\} = \mathrm{M}\{|F - KGF - KV|^2\} = |1 - KG|^2 S_{ff} + |K|^2 S_{VV}, \tag{8.5}$$

де використано некорельованість образу і шуму (перехресні члени мають нульове середнє). Мінімізуючи (8.5) по K (прирівнюючи до нуля похідну за K^*), отримуємо

$$K_W(k) = \frac{G^*(k)\,S_{ff}(k)}{|G(k)|^2 S_{ff}(k) + S_{VV}(k)} = \frac{1}{G(k)} \cdot \frac{|G(k)|^2}{|G(k)|^2 + \dfrac{S_{VV}(k)}{S_{ff}(k)}}. \tag{8.6}$$

Друга форма запису наочно показує властивості фільтра Вінера. На частотах, де відношення сигнал/шум S_{ff}/S_{VV} велике, другий співмножник близький до одиниці, і фільтр діє як інверсний. На частотах, де шум переважає, другий співмножник прямує до нуля, і фільтр «закривається», не пропускаючи шум. Перехід між цими режимами плавний, що усуває різке обрізання спектра і, відповідно, значно послаблює ефект Гіббса [13, 14].

8.3.2. Практичні варіанти

На практиці спектр потужності неспотвореного образу невідомий. Тому часто використовують спрощену форму з постійним відношенням шум/сигнал K:

$$K_W(k) \approx \frac{G^*(k)}{|G(k)|^2 + K}. \tag{8.7}$$

Параметр K підбирають емпірично, що повертає нас до проблеми вибору параметра. На рис. 8.3 показано результати фільтрації Вінера з різними значеннями K: при малому K результат містить підсилений шум, при великому — залишкове розмиття. Найкращий результат (рис. 8.3, г) отримано з «оракульним» фільтром, у якому використано точний спектр потужності вихідного образу, — на практиці недосяжна ситуація, що слугує лише верхньою межею якості.

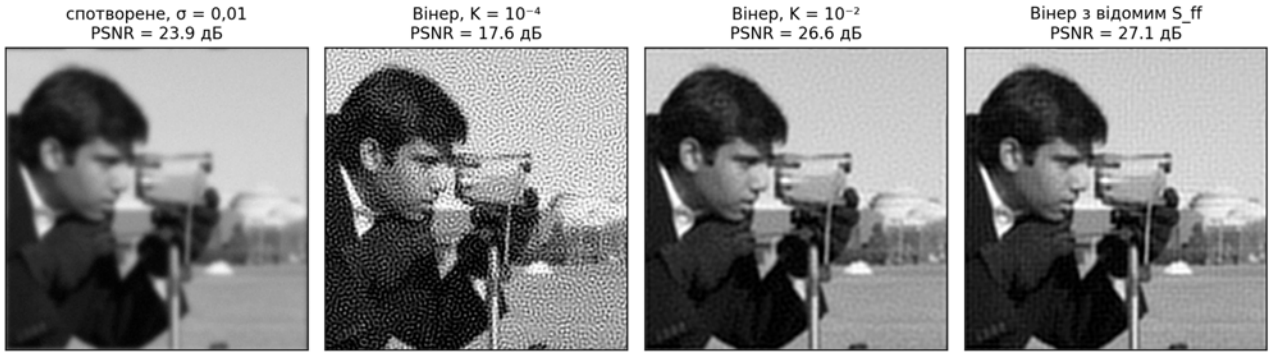


Рис. 8.3. Фільтрація Вінера з різними значеннями параметра K

Зазначимо принципову особливість фільтра Вінера [2]: він налаштований на *клас* образів із заданим спектром потужності і шум із заданим спектром. Для всіх образів цього класу фільтр однаковий і перебудови не потребує. Платою є необхідність знати (або оцінити за ансамблем) усереднені спектри потужності, а також вимога однорідності і некорельованості полів образу і шуму.

8.4. Метод умовної деконволюції в натуральних координатах

8.4.1. Апріорна інформація

Метод умовної деконволюції — це метод реставрації, що потребує мінімальної апріорної інформації: відомими вважають спотворений образ L, оператор згортки G, а про шум (крім того, що він утворює ергодичне, стаціонарне, ізопланарне, однорідне і адитивне поле) відома лише його енергія [2, 3]

$$\varepsilon^2 = \mathrm{V}^T\mathrm{V}. \tag{8.8}$$

Як показано в розділі 3, для образу великого формату (кількість відліків $M \sim 10^3$ і більше) енергію шуму досить точно оцінюють за ділянкою образу, де об'єкт відсутній, а при нульовому середньому шуму вона пропорційна дисперсії: $\varepsilon^2 = M\sigma^2$. Логічний і фізичний зміст методу найзрозуміліший у натуральних координатах, де він історично і з'явився; тому почнемо саме з цього випадку, розглядаючи для простоти одновимірний образ.

8.4.2. Умова на енергію шуму

Оскільки шум $\mathrm{V} = \mathrm{L} - \mathrm{Gf}$, його енергію можна виразити як

$$\varepsilon^2 = (\mathrm{L} - \mathrm{Gf})^T(\mathrm{L} - \mathrm{Gf}). \tag{8.9}$$

Через обмеженість інформації про шум точне значення f обчислити неможливо; можна знайти лише оцінку $\hat{\mathrm{f}}$, яка за певним критерієм наближається до f. Перший критерій полягає в тому, що оцінка має бути узгодженою з даними в межах рівня шуму: якщо підставити $\hat{\mathrm{f}}$ замість f, енергія «залишку» (нев'язки) має дорівнювати енергії шуму:

$$\left(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}\right)^T\left(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}\right) = \varepsilon^2. \tag{8.10}$$

Справді, вимагати точнішого збігу ($\mathrm{G}\hat{\mathrm{f}} = \mathrm{L}$) немає сенсу — це означало б «пояснювати» шум деталями образу, що й призводить до катастрофи інверсної фільтрації. Менш точний збіг означав би втрату інформації, яка в даних є.

Рівняння (8.10) нелінійне (квадратичне) відносно $\hat{\mathrm{f}}$. Розкриємо його для найпростішого випадку матриці третього порядку:

$$\begin{aligned}(L_0 - g_{00}f_0 - g_{01}f_1 - g_{02}f_2)^2 + (L_1 - g_{10}f_0 - g_{11}f_1 - g_{12}f_2)^2 \\ + (L_2 - g_{20}f_0 - g_{21}f_1 - g_{22}f_2)^2 = \varepsilon^2.\end{aligned} \tag{8.11}$$

Це рівняння поверхні (еліпсоїда) у тривимірному просторі невідомих (f_0, f_1, f_2): розв'язків нескінченно багато [2]. Отже, потрібна ще одна умова, яка вибрала б з цієї множини єдину оцінку. Ця умова має бути простою, щоб не підвищувати ступінь нелінійності задачі, і відображати фізичні властивості, притаманні образам досліджуваних об'єктів.

8.4.3. Умова гладкості

Такою умовою в методі умовної деконволюції є критерій гладкості оцінки: серед усіх оцінок, що задовольняють (8.10), обирається найбільш гладка. Функцію «гладкості» задають у вигляді суми квадратів других різниць [2]:

$$\psi\left(\hat{\mathrm{f}}\right) = \sum_{i=1}^{M-2}\left(\hat{f}_{i-1} - 2\hat{f}_i + \hat{f}_{i+1}\right)^2 \to \min. \tag{8.12}$$

Зміст критерію пояснює рис. 8.4. Якщо три послідовні значення оцінки лежать на одній прямій, друга різниця дорівнює нулю (рис. 8.4, а). Якщо середнє значення суттєво відхиляється від прямої, що з'єднує сусідні, друга різниця різко зростає (рис. 8.4, б, в), і такі «викиди» критерій (8.12) штрафує. Друга різниця —

це дискретний аналог другої похідної (і одновимірний аналог лапласіана (7.9)), тому критерій (8.12) мінімізує «кривину» оцінки.

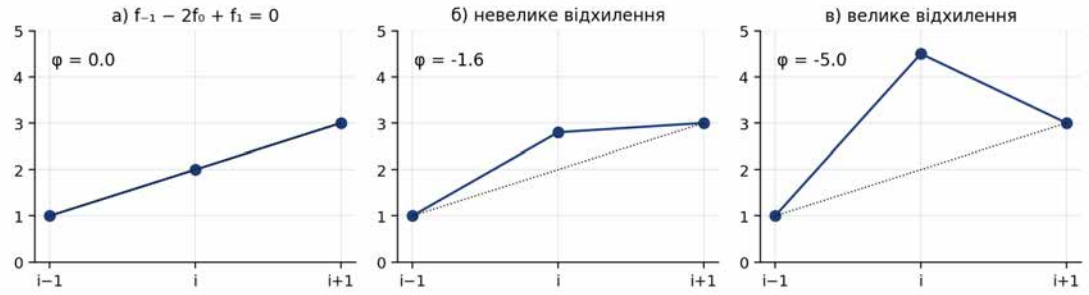


Рис. 8.4. Зміст критерію гладкості

Критерій (8.12) є найпростішим з можливих, оскільки враховує мінімальну кількість послідовних точок. Можна використовувати і складніші критерії з більшою кількістю точок або критерії, що допускають розриви (наприклад, на основі перших різниць або повної варіації); вибір критерію — одне з джерел модифікацій методу [2].

Якщо кожну з других різниць вважати елементом стовпця φ, то $\psi = \varphi^T\varphi$, а сам стовпець отримують множенням оцінки на дискретний оператор C, побудований так само, як оператор згортки (3.13), (3.14), але з «імпульсної характеристики гладкості» $c = [1, -2, 1]$:

$$\varphi = \mathrm{C}\,\hat{\mathrm{f}}, \qquad \psi(\hat{\mathrm{f}}) = (\mathrm{C}\hat{\mathrm{f}})^T(\mathrm{C}\hat{\mathrm{f}}). \tag{8.13}$$

Для двовимірних образів умова гладкості описується двовимірною імпульсною характеристикою — лапласіаном

$$c(m,n) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{8.14}$$

8.4.4. Метод множників Лагранжа і розв’язок

Отже, оцінка має задовольняти дві умови:

$$(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}})^T(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}) - \varepsilon^2 = 0, \qquad (\mathrm{C}\hat{\mathrm{f}})^T(\mathrm{C}\hat{\mathrm{f}}) = \min. \tag{8.15}$$

Обидві умови мають розмірність і зміст енергії шуму: перша — енергії адитивного шуму, яка після реставрації не повинна бути більшою, ніж у спотвореному образі; друга — енергії «негладкості» оцінки [2]. Для об’єднання умов скористаємося методом невизначених множників Лагранжа: зважимо першу умову додатним коефіцієнтом λ і будемо шукати мінімум суми

$$\beta(\hat{\mathrm{f}}) = \lambda\left[(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}})^T(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}) - \varepsilon^2\right] + (\mathrm{C}\hat{\mathrm{f}})^T(\mathrm{C}\hat{\mathrm{f}}) \to \min. \tag{8.16}$$

У точці мінімуму всі часткові похідні $\partial\beta/\partial\hat{f}_i$, $i = 0,1,\ldots,M-1$, дорівнюють нулю. Для обчислення похідної квадратичної форми зручно скористатися правилами матричного диференціювання: $\partial(\mathrm{a} - \mathrm{B}\mathrm{x})^T(\mathrm{a} - \mathrm{B}\mathrm{x})/\partial\mathrm{x} = -2\mathrm{B}^T(\mathrm{a} - \mathrm{B}\mathrm{x})$ і $\partial(\mathrm{B}\mathrm{x})^T(\mathrm{B}\mathrm{x})/\partial\mathrm{x} = 2\mathrm{B}^T\mathrm{B}\mathrm{x}$. (У [2] ці правила виведено покоординатно для матриць третього порядку, що корисно виконати самостійно як вправу.) Отримуємо

$$\frac{\partial \beta}{\partial \hat{\mathrm{f}}} = 2\lambda\left(\mathrm{G}^T\mathrm{G}\hat{\mathrm{f}} - \mathrm{G}^T\mathrm{L}\right) + 2\,\mathrm{C}^T\mathrm{C}\hat{\mathrm{f}} = 0. \tag{8.17}$$

Таким чином, нелінійна задача з квадратичними формами завдяки диференціюванню зведена до *системи лінійних рівнянь* M-го порядку відносно $\hat{\mathrm{f}}$, хоча і з невідомим коефіцієнтом λ. Її розв'язок:

$$\hat{\mathrm{f}} = \left[\mathrm{G}^T\mathrm{G} + \frac{1}{\lambda}\mathrm{C}^T\mathrm{C}\right]^{-1} \mathrm{G}^T\mathrm{L}. \tag{8.18}$$

Параметр λ визначається з умови (8.10): потрібно знайти таке λ, при якому енергія нев'язки оцінки (8.18) дорівнює енергії шуму. Звідси ітераційна процедура умовної деконволюції [2]:

1. сформувати матриці G, C і стовпець L (для одновимірного образу з $M = 1024$ відліками матриці мають розмір 1024×1024);
2. задати початкове значення λ;
3. обчислити $\hat{\mathrm{f}}$ за (8.18), для чого обернути (або розв'язати систему з) матрицю в квадратних дужках;
4. обчислити енергію нев'язки $\hat{\varepsilon}^2 = \left(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}\right)^T\left(\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}\right)$;
5. перевірити, чи виконується з потрібною точністю $\hat{\varepsilon}^2 \approx \varepsilon^2$ (принаймні $\hat{\varepsilon}^2 \leq \varepsilon^2$); якщо так — закінчити обчислення;
6. змінити λ: якщо $\hat{\varepsilon}^2 > \varepsilon^2$, збільшити λ (послабити згладжування), якщо $\hat{\varepsilon}^2 < \varepsilon^2$ — зменшити; перейти до кроку 3.

Енергія нев'язки монотонно спадає зі зростанням λ (рис. 8.5, б), тому для пошуку λ ефективний метод поділу навпіл у логарифмічному масштабі.

8.4.5. Обчислювальний приклад

На рис. 8.5 наведено приклад реставрації одновимірного образу, що містить прямокутні імпульси різної ширини, два близько розташовані вузькі імпульси і лінійно зростаючу ділянку. Образ спотворено експоненційною імпульсною характеристикою і білим шумом з $\sigma = 0{,}02$ (рис. 8.5, а); два вузькі імпульси в спотвореному образі майже злилися. Залежність енергії нев'язки від λ і знайдене значення λ^*, при якому $\hat{\varepsilon}^2 = \varepsilon^2 = M\sigma^2$, показано на рис. 8.5, б. Оцінка при $\lambda = \lambda^*$ (рис. 8.5, в) розділяє близькі імпульси, відновлює крутизну перепадів і водночас не містить підсиленого шуму. При значенні λ, у 100 разів меншому, оцінка надто гладка: вузькі деталі занижені. При значенні, у 100 разів більшому (рис. 8.5, г), регуляризація недостатня: перепади відновлено, але оцінка містить значний шум, який система «намагається пояснити» деталями образу.

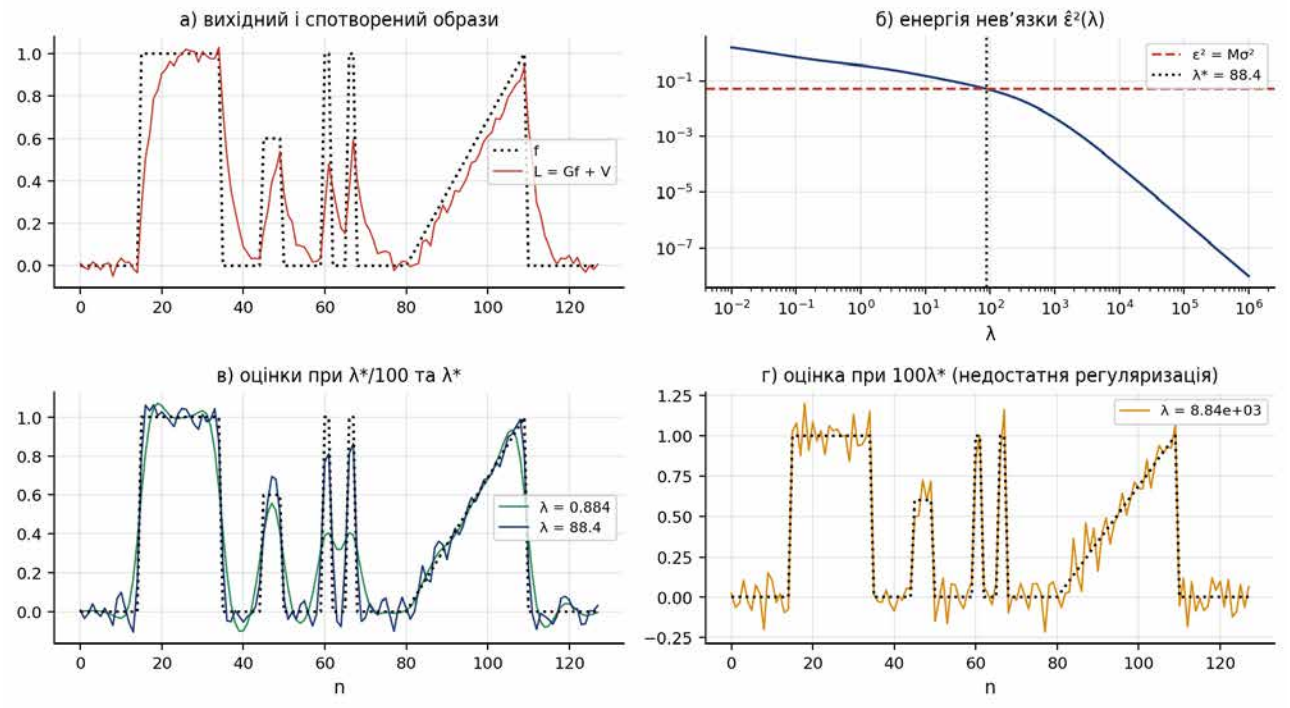


Рис. 8.5. Умовна деконволюція одновимірного образу в натуральних координатах

8.4.6. Обчислювальна складність

З виразу (8.18) та ітераційної процедури зрозуміло, що метод умовної деконволюції в натуральних координатах вимагає на кожній ітерації обернення матриці порядку M, тобто порядку M^3 операцій, великих обсягів пам'яті і призводить до накопичення операційних похибок. Навіть для одновимірних образів великого формату ($M = 512,1024,2048$) процедура є надзвичайно громіздкою [2].

Для двовимірних образів ситуація погіршується катастрофічно. Формули і ітераційна процедура залишаються тими самими, змінюється лише умова гладкості (8.14). Але для образу 1024×1024 розгортанням за рядками утворюється стовпець розміром $(1024)^2 \times 1$, а матриці G і $[\mathrm{G}^T\mathrm{G} + \lambda^{-1}\mathrm{C}^T\mathrm{C}]$ мають розмір $10^6 \times 10^6$. Для обернення такої матриці потрібно близько $(10^6)^3 = 10^{18}$ операцій на кожній ітерації [2]. Тому в натуральних координатах метод у такому вигляді на практиці не реалізують. Зате при переході в область трансформант Фур'є він стає не лише реалізованим, а й досить швидким.

8.5. Умовна деконволюція в області трансформант Фур'є

8.5.1. Перехід в частотну область

Нехай оператори G і C циркулянтні (для двовимірних образів — блочно-циркулянтні з циркулянтними блоками). Тоді, згідно з теоремою про згортку (4.21), нормований оператор ДПФ діагоналізує обидві матриці одночасно, і всі операції в (8.16)–(8.18) стають поелементними [2, 3]. Позначимо $L(k)$, $\hat{F}(k)$, $G(k)$, $C(k)$ — ДПФ стовпця спостережень, оцінки, імпульсної характеристики системи та «імпульсної характеристики гладкості» (доповненої нулями до

формату образу). Для двовимірного образу $M \times M$ індекс k пробігає всі M^2 пар просторових частот (m, n), а $C(m, n)$ — двовимірне ДПФ від маски (8.14). Завдяки рівності Парсеваля умови (8.15) в області трансформант набувають вигляду

$$\sum_k \left|L(k) - G(k)\hat{F}(k)\right|^2 = \varepsilon^2, \qquad \sum_k \left|C(k)\hat{F}(k)\right|^2 = \min. \tag{8.19}$$

(Тут використано унітарне ДПФ. Якщо ДПФ не нормоване, як у [2, 3], у формулах з'являються множники M^2 для двовимірного образу формату $M \times M$; на кінцевий результат це не впливає.)

Функціонал Лагранжа розпадається на суму незалежних доданків для кожної частоти:

$$\beta = \sum_k \left\{\lambda\left(\left|L(k) - G(k)\hat{F}(k)\right|^2 - \varepsilon_k^2\right) + \left|C(k)\hat{F}(k)\right|^2\right\} \to \min. \tag{8.20}$$

Такий простий вигляд є безпосереднім наслідком діагоналізації матриці дискретної згортки при переході в область трансформант Фур'є [2]. Диференціювання кожного доданка за $\hat{F}^*(k)$ (з урахуванням того, що $|z|^2 = z^* z$) і прирівнювання до нуля дає

$$\hat{F}(k) = \frac{G^*(k)\, L(k)}{|G(k)|^2 + \dfrac{1}{\lambda}|C(k)|^2} = K_{DF}(k)\, L(k), \tag{8.21}$$

де

$$K_{DF}(k) = \frac{1}{G(k)} \cdot \frac{|G(k)|^2}{|G(k)|^2 + \dfrac{1}{\lambda}|C(k)|^2} \tag{8.22}$$

— коефіцієнт передачі фільтра умовної деконволюції з підстроюваним параметром λ [2, 3]. Вираз (8.21) — це просто поелементна форма матричного розв'язку (8.18): обернення матриці порядку M^2 замінилося M^2 діленнями.

8.5.2. Рівняння для параметра λ

Підставивши (8.21) в умову на енергію нев'язки (8.19), отримаємо рівняння, яке містить лише одну невідому — λ:

$$\hat{\varepsilon}^2(\lambda) = \sum_k \frac{\dfrac{1}{\lambda^2}\,|L(k)|^2\,|C(k)|^4}{\left[\dfrac{1}{\lambda}|C(k)|^2 + |G(k)|^2\right]^2} = \varepsilon^2. \tag{8.23}$$

Отже, на відміну від реставрації в натуральних координатах, ітераційно уточнюється лише скалярний параметр λ, а кожна ітерація потребує $O(M^2)$ операцій. Алгоритм умовної деконволюції в області трансформант Фур'є [2, 3]:

1. обчислити за допомогою ШПФ спектри деградованого образу $L(k)$, частотну характеристику системи $G(k)$ і частотну характеристику оператора гладкості $C(k)$;

2. задати початкове значення λ;
3. обчислити $\hat{\varepsilon}^2(\lambda)$ за (8.23);
4. якщо $|\hat{\varepsilon}^2(\lambda) - \varepsilon^2|$ достатньо мале, перейти до кроку 6; інакше змінити λ (збільшити, якщо $\hat{\varepsilon}^2 > \varepsilon^2$, зменшити — якщо навпаки);
5. перейти до кроку 3;
6. обчислити спектр оцінки за (8.21) та обернене ШПФ.

Рівність нулю похідних означає екстремум, тому формально слід перевірити, що знайдене значення відповідає саме мінімуму функціоналу. У [2] з аналізу других похідних виведено умову $\lambda > \lambda_{min} = \min_k |C(k)|^2/|G(k)|^2$, виконання якої рекомендовано перевіряти перед завершенням циклу і, у разі невиконання, змінювати початкове значення λ.

Підсумовуючи, можна стверджувати, що досі поширена думка про трудомісткість перетворення Фур'є як бази переходу в область трансформант не відповідає дійсності: додаткові витрати на ШПФ настільки малі порівняно з трудомісткістю деконволюції в натуральних координатах, що перетворення Фур'є, навпаки, на кілька порядків спрощує процедуру реставрації [2].

8.5.3. Порівняння з фільтром Вінера

Фільтр умовної деконволюції (8.22) за формою дуже нагадує фільтр Вінера (8.6). Відмінність полягає в коригуючому доданку знаменника: замість відношення спектрів потужності шуму і образу $S_{VV}(k)/S_{ff}(k)$ у ньому стоїть доданок $\lambda^{-1}|C(k)|^2$, що відображає умову гладкості з вагою $1/\lambda$. Оскільки $|C(k)|^2$ для оператора другої різниці зростає з частотою як $\sim (2\pi k/M)^4$, а спектр потужності природних образів спадає з частотою, якісна поведінка обох доданків подібна: обидва фільтри «закриваються» на високих частотах (рис. 8.2, б).

Проте між фільтрами існує принципова різниця [2]:

- фільтр Вінера налаштований на *клас* образів і шумів з відомими спектрами потужності і для образів цього класу перебудови не потребує; фільтр умовної деконволюції має додатковий параметр λ, який обчислюється за рівнянням (8.23) для *кожного конкретного* деградованого образу;
- початкова апріорна інформація для умовної деконволюції значно менша: потрібна лише енергія шуму (яку можна оцінити за тим самим образом), і немає потреби обчислювати усереднені спектри потужності образу і шуму;
- відсутня вимога, щоб усі образи і реалізації шуму належали до однорідних некорельованих полів з однаковими спектрами.

Обраний критерій оптимальності — гладкість у вигляді лапласіана — є лише одним з багатьох можливих, що відкриває можливості для модифікацій методу [2].

8.5.4. Зв'язок з регуляризацією Тихонова

Вираз (8.18) збігається з розв'язком задачі регуляризації за Тихоновим [33, 45]

$$\hat{\mathrm{f}}_{\alpha} = \underset{\mathrm{f}}{\operatorname{argmin}}\{\| \mathrm{L} - \mathrm{Gf} \|^2 + \alpha \ \| \mathrm{Cf} \|^2\} \tag{8.24}$$

з параметром регуляризації $\alpha = 1/\lambda$ і стабілізатором $\| \mathrm{Cf} \|^2$, а правило вибору λ з умови $\| \mathrm{L} - \mathrm{G}\hat{\mathrm{f}} \|^2 = \varepsilon^2$ — з принципом нев'язки (принципом Морозова), відомим у теорії некоректних задач. Таким чином, метод умовної деконволюції можна розглядати як конкретну, фізично вмотивовану реалізацію загальної схеми регуляризації некоректних задач з вибором параметра за рівнем шуму. Таке тлумачення корисне з двох причин: воно пояснює, чому метод працює (регуляризація перетворює некоректну задачу на коректну), і дозволяє використовувати розвинений апарат теорії регуляризації — інші стабілізатори, інші правила вибору параметра (метод L-кривої, узагальнену перехресну перевірку) [45].

8.5.5. Обчислювальний експеримент

На рис. 8.6 наведено приклад реставрації зображення, спотвореного гаусівським розмиттям ($\sigma = 2$ пікселі) і білим шумом ($\sigma = 0{,}01$). Параметр λ визначено автоматично з рівняння (8.23) методом поділу навпіл за відомою енергією шуму $\varepsilon^2 = M^2\sigma^2$. Порівняно з псевдоінверсною фільтрацією результат умовної деконволюції має меншу кількість артефактів у вигляді «дзвону» та хвилястої текстури на однорідних ділянках при подібному значенні PSNR.



Рис. 8.6. Реставрація зображення методом умовної деконволюції в області трансформант Фур'є

Кількісне порівняння методів для трьох рівнів шуму наведено в табл. 8.1. Для псевдоінверсного фільтра і фільтра Вінера зі сталим K наведено результати при *найкращому* значенні параметра, підібраному перебором з використанням еталонного зображення, — на практиці недоступна можливість. Фільтр Вінера з точним спектром потужності еталона («оракул») дає верхню межу якості лінійної фільтрації. Метод умовної деконволюції використовує лише енергію шуму і не потребує ні підбору параметрів, ні знання еталона, проте забезпечує якість, близьку до найкращої досяжної для оптимально налаштованих фільтрів.

Таблиця 8.1

PSNR (дБ) результатів реставрації зображення, спотвореного гаусівським розмиттям (σ = 2) і білим шумом

СКВ шуму	Спотворене	Псевдоінверсний (найкращий поріг)	Вінер, стале K (найкраще K)	Вінер з точним спектром	Умовна деконволюція (λ за енергією шуму)
0,002	24,0	27,9	28,2	28,5	27,7
0,01	23,9	26,3	26,7	27,1	26,2
0,03	23,1	25,0	25,0	25,9	25,0

8.5.6. Крайові ефекти і ефект Гіббса

Реставрація в області трансформант Фур'є неявно використовує циклічну модель згортки. Якщо реальне спотворення є лінійною згорткою (що завжди так для фізичних систем), на межах образу модель не відповідає даним: циклічна модель «очікує», що розмиття переносить яскравість з лівого краю на правий, чого насправді не відбувається. Результатом є сильні артефакти біля меж (рис. 8.7, б), що поширюються всередину образу у вигляді осциляцій — прояв того самого ефекту Гіббса, пов'язаного з розривами періодичного продовження (розділ 4). Ефективними засобами боротьби є плавне «зважування» країв образу перед ДПФ (рис. 8.7, в), дзеркальне продовження образу, доповнення нулями з подальшим відкиданням крайових зон, а також використання базисів, вільних від ефекту Гіббса.



Рис. 8.7. Крайові ефекти при реставрації за циклічною моделлю

Ефект Гіббса проявляється і всередині образу: обмеження спектра оцінки (фільтр (8.22) практично «закриває» високі частоти) породжує осциляції біля різких меж графоелементів, що візуально сприймається як їх розмивання. Саме цей недолік реставрації в базисі Фур'є став мотивом для розроблення модифікацій методу умовної деконволюції в базисах перетворень RTF і Адамара [2, 4, 6, 9], які розглядаються в розділі 9.

8.6. Узагальнена дискретна фільтрація за методом найменших квадратів

Методи, розглянуті вище, спираються на модель деградації для однорідного поля образу. У багатьох випадках неоднорідністю поля образу знехтувати не можна; тоді слід використовувати анізопланарну модель (3.3), для якої теорема про згортку не виконується, і перехід в область Фур'є втрачає свої переваги [2]. Водночас це розширює інструментарій дослідника: якщо базис Фур'є не має особливих переваг, можна використовувати будь-яке ортогональне перетворення, зокрема з дійсним ядром.

Нехай модель деградації має вигляд (8.1), а A — унітарна матриця довільного ортогонального перетворення (Адамара, Хаара, косинусного, похилого, Фур'є тощо). Алгоритм фільтрації в області трансформант складається з трьох кроків [2]:

1. пряме перетворення спотвореного образу: $\mathrm{L}_A = \mathrm{AL}$;
2. корекція в «частотній» області: $\mathrm{L}'_A = \mathrm{M}\,\mathrm{L}_A$, де M — матриця коригуючого фільтра, яку і потрібно знайти;
3. обернене перетворення: $\hat{\mathrm{f}} = \mathrm{A}^{-1}\mathrm{MAL}$.

Для однорідного поля і базису Фур'є матриця M діагональна. Для неоднорідного поля (і для перетворень, для яких теорема про згортку не виконується) вона недіагональна: кожна відфільтрована «частотна» складова залежить не лише від однойменної вхідної складової, а від багатьох або всіх складових.

Визначимо M за критерієм мінімуму середньоквадратичної похибки лінійної оцінки. За принципом ортогональності похибка оцінки має бути некорельованою з даними:

$$\mathrm{M}\left\{\left(\mathrm{f}-\hat{\mathrm{f}}\right)\mathrm{L}^{T}\right\} = 0. \tag{8.25}$$

Підставивши сюди $\hat{\mathrm{f}} = \mathrm{A}^{-1}\mathrm{MAL}$ і $\mathrm{L} = \mathrm{Gf} + \mathrm{V}$, вважаючи образ і шум некорельованими і позначивши кореляційні матриці $\mathrm{R}_{ff} = \mathrm{M}\{\mathrm{ff}^{T}\}$, $\mathrm{R}_{VV} = \mathrm{M}\{\mathrm{VV}^{T}\}$, отримаємо

$$\mathrm{M} = \mathrm{A}\,\mathrm{R}_{ff}\mathrm{G}^{T}\left(\mathrm{G}\mathrm{R}_{ff}\mathrm{G}^{T} + \mathrm{R}_{VV}\right)^{-1}\mathrm{A}^{-1}. \tag{8.26}$$

Введемо за аналогією зі спектром потужності (теорема Вінера — Хінчина) матриці, перетворені в обраний базис:

$$\mathrm{S}_{ff} = \mathrm{A}\mathrm{R}_{ff}\mathrm{A}^{-1}, \quad \mathrm{S}_{VV} = \mathrm{A}\mathrm{R}_{VV}\mathrm{A}^{-1}, \quad \mathrm{K} = \mathrm{AGA}^{-1}, \quad \mathrm{K}^{*} = \mathrm{AG}^{T}\mathrm{A}^{-1}, \tag{8.27}$$

де K^{*} — аналог «спряження» спектра. Вставивши між співмножниками в (8.26) одиничні матриці $\mathrm{A}^{-1}\mathrm{A}$, отримаємо [2]

$$\mathrm{M} = \mathrm{S}_{ff}\,\mathrm{K}^{*}\left(\mathrm{K}\,\mathrm{S}_{ff}\,\mathrm{K}^{*} + \mathrm{S}_{VV}\right)^{-1}. \tag{8.28}$$

Вираз (8.28) за формою дуже нагадує фільтр Вінера, але є справедливим для довільного «частотного» базису і неоднорідних полів. Для базису Фур'є і однорідного поля всі матриці в (8.28) діагональні, і вираз зводиться до (8.6). У

загальному випадку матриця M має N^2 ненульових елементів замість N для діагональної.

Властивості методу [2]:

1. узагальнена фільтрація значно універсальніша за вінерівську, оскільки не вимагає однорідності поля образу (а однорідність практично ніколи точно не виконується); фільтр Вінера є її окремим випадком для базису Фур'є і однорідного поля;
2. обчислення кореляційних матриць є окремою проблемою: неоднорідні поля не є ергодичними, тому усереднення потрібно виконувати за ансамблем реалізацій, що унеможливлює застосування методу при малій кількості реалізацій образу і шуму;
3. матриця A може бути матрицею будь-якого глобального ортогонального перетворення; при використанні, наприклад, перетворення Адамара трудомісткість обчислень може бути суттєво зменшена;
4. у деяких випадках алгоритм спрощується: якщо деградація зумовлена лише адитивним шумом (G = E), то $\mathrm{M} = \mathrm{S}_{ff}\left(\mathrm{S}_{ff} + \mathrm{S}_{VV}\right)^{-1}$.

8.7. Деконволюція за критерієм форми результуючої імпульсної характеристики

8.7.1. Ідея методу

Цей метод також належить до лінійних, але критерієм реставрації в ньому є не близькість оцінки до вихідного образу, а відхилення результуючої імпульсної характеристики системи «деградація — реставрація» від дельта-функції [2]. Шукається коригуючий фільтр з імпульсною характеристикою $\xi(x, y)$, який перетворює імпульсну характеристику деградуючої системи $g(x, y)$ на деяку $c(x, y)$, що за обраним критерієм наближає дельта-функцію:

$$c(x, y) = g(x, y) * \xi(x, y) \to \delta(x, y). \tag{8.29}$$

Отримання точно $c = \delta$ відповідало б інверсній фільтрації. Але за наявності шуму

$$\xi * L = \xi * (g * f + V) = c * f + \xi * V, \tag{8.30}$$

і останній доданок — профільтрований шум — може замаскувати образ сильніше, ніж у вихідних даних. Тому результуюча імпульсна характеристика не повинна бути дельта-імпульсом; вона повинна мати таку форму, щоб енергія шуму після реставрації не перевищувала енергії шуму в спотвореному образі.

8.7.2. Формалізація

Мірою відхилення від дельта-функції є величина, обернена до гостроти результуючої імпульсної характеристики:

$$r^2 = \frac{\int w(x)\, c^2(x)\, dx}{\int c^2(x)\, dx}, \tag{8.31}$$

де $w(x)$ — вагова (штрафна) функція, яка «штрафує» енергію $c(x)$ за межами околу початку координат. Найчастіше використовують $w(x) = x^2$ або $w(x) = 0$ при $|x| < D/2$ і $w(x) = 1$ поза цим околом [2].

Нехай коригуючий фільтр має $2M + 1$ відліків $\xi = [\xi_{-M}, \dots, \xi_M]^T$. Тоді $c(x) = \sum_i \xi_i \; g(x - x_i)$, і чисельник і знаменник (8.31) є квадратичними формами:

$$r^2 = \frac{\xi^T \mathrm{A}\, \xi}{\xi^T \mathrm{B}\, \xi}, \quad A_{ij} = \int w(x)\, g(x - x_i)\, g\big(x - x_j\big)\, dx, \quad B_{ij} = \int g(x - x_i)\, g\big(x - x_j\big)\, dx. \tag{8.32}$$

Енергія шуму після фільтрації для однорідного шуму з кореляційною матрицею R_{VV}:

$$\varepsilon_0^2 = \xi^T \mathrm{R}_{VV}\, \xi. \tag{8.33}$$

Задача полягає в мінімізації r^2 за умови $\varepsilon_0^2 \le \varepsilon^2$. Методом невизначених множників (з множниками $\lambda^{(1)}$ і $\lambda^{(2)}$) і диференціюванням за ξ отримуємо [2]

$$\left[\mathrm{B}^{-1}\left(\mathrm{A} + \lambda^{(2)}\mathrm{R}_{VV}\right) - \lambda^{(1)}\mathrm{E}\right]\xi = 0, \tag{8.34}$$

тобто $\lambda^{(1)}$ — власне число, а ξ — власний вектор матриці $\mathrm{D} = \mathrm{B}^{-1}\left(\mathrm{A} + \lambda^{(2)}\mathrm{R}_{VV}\right)$. Глобальному мінімуму відповідає найменше власне число. Алгоритм:

1. задати $\lambda^{(2)}$ і сформувати матрицю D;
2. знайти всі власні числа D і вибрати найменше;
3. знайти відповідний власний вектор ξ;
4. виконати реставрацію набору тестових образів (фантомів) згорткою з ξ;
5. якщо результати задовільні — закінчити; інакше змінити $\lambda^{(2)}$ і повернутися до кроку 2.

Параметр $\lambda^{(2)}$ регулює компроміс між гостротою результуючої імпульсної характеристики і підсиленням шуму. На рис. 8.8 наведено результати для гаусівської імпульсної характеристики. При $\lambda^{(2)} = 0$ (шум не враховується) результуюча характеристика найвужча, але коригуючий фільтр має великі знакозмінні коефіцієнти, що сильно підсилюють шум. Зі збільшенням $\lambda^{(2)}$ коефіцієнти фільтра зменшуються, а результуюча характеристика розширюється, наближаючись до вихідної.

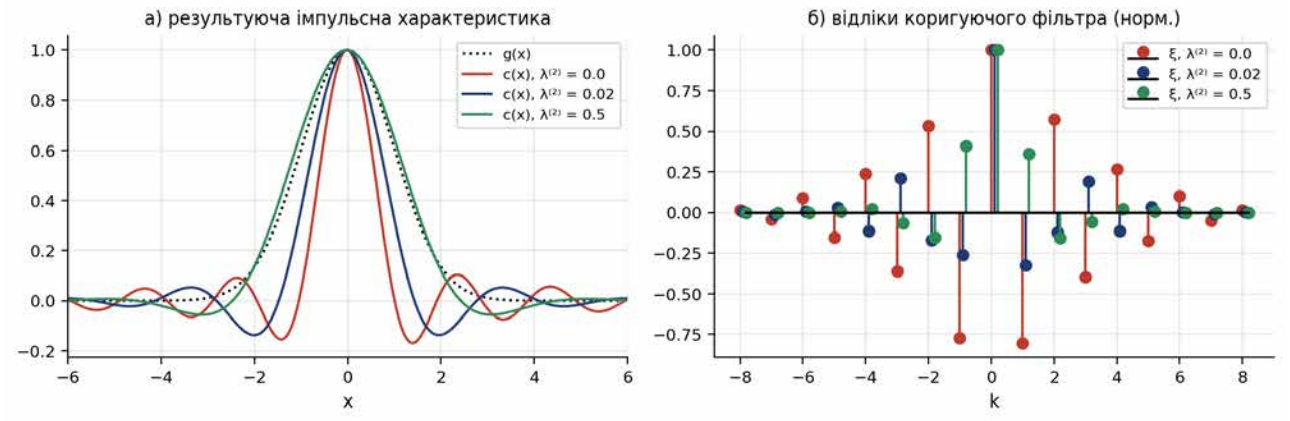


Рис. 8.8. Деконволюція за критерієм форми результуючої імпульсної характеристики

Перевагою методу є те, що коригуючий фільтр для даної системи обчислюється один раз: імпульсна характеристика системи і характеристики її внутрішніх шумів не змінюються і є спільними для всіх образів [2]. Критерій якості — мінімізація відхилення результуючої імпульсної характеристики від дельта-функції — фізично зрозумілий і наочний. Обмеженнями є вимога однорідності полів образу і шуму та лінійність методу. Напрямками вдосконалення є спрощення вибору $\lambda^{(2)}$ і перенесення методу в частотну область для зменшення трудомісткості.

8.8. Програмна реалізація

Нижче наведено реалізацію умовної деконволюції в області трансформант Фур'є для двовимірного образу з автоматичним визначенням параметра λ за енергією шуму, а також варіант у натуральних координатах для одновимірного образу.

```
import numpy as np
from scipy.linalg import circulant

def psf2otf(h, shape):
    """ДПФ імпульсної характеристики, центрованої в (0, 0), формату shape."""
    H = np.zeros(shape); r, c = h.shape
    H[:r, :c] = h
    H = np.roll(H, (-(r // 2), -(c // 2)), axis=(0, 1))
    return np.fft.fft2(H)

def cond_deconv_fourier(L, h, eps2, lo=-8.0, hi=12.0, iters=60):
    """Умовна деконволюція (8.21)-(8.23); eps2 — енергія шуму."""
    G = psf2otf(h, L.shape)
    C = psf2otf(np.array([[0, 1, 0], [1, -4, 1], [0, 1, 0]], float), L.shape)
    Lf = np.fft.fft2(L); n = L.size
    def resid(lam):                    # (8.23), унітарна нормалізація
        R = Lf * (np.abs(C)**2 / lam) / (np.abs(G)**2 + np.abs(C)**2 / lam)
        return np.sum(np.abs(R)**2) / n
```

```
    for _ in range(iters):              # поділ навпіл за lg(lambda)
        m = 0.5 * (lo + hi)
        if resid(10**m) > eps2: lo = m
        else: hi = m
    lam = 10**(0.5 * (lo + hi))
    F = np.conj(G) * Lf / (np.abs(G)**2 + np.abs(C)**2 / lam)
    return np.real(np.fft.ifft2(F)), lam

def cond_deconv_natural(L, g, eps2):
    """Умовна деконволюція в натуральних координатах (8.18), 1D."""
    N = L.size
    col = np.zeros(N); col[:g.size] = g; G = circulant(col)
    cv = np.zeros(N); cv[[0, 1, N - 1]] = [-2, 1, 1]; Cm = circulant(cv)
    lo, hi = -4.0, 8.0
    for _ in range(50):
        lam = 10**(0.5 * (lo + hi))
        f = np.linalg.solve(G.T @ G + Cm.T @ Cm / lam, G.T @ L)
        if np.sum((L - G @ f)**2) > eps2: lo = np.log10(lam)
        else: hi = np.log10(lam)
    return f, lam
```

8.9. Приклади розв'язання задач

Задача 8.1. На трьох частотах модуль частотної характеристики розмиття дорівнює $G = 0{,}9;\ 0{,}1;\ 0{,}01$. Порівняйте коефіцієнти передачі інверсного фільтра і фільтра Вінера (8.7) з $K = 0{,}01$ (вважаючи G дійсним).

Розв'язання. Інверсний фільтр: $1/G = 1{,}11;\ 10;\ 100$. Фільтр Вінера $K_W = G/(G^2 + K)$: $0{,}9/(0{,}81 + 0{,}01) = 1{,}10$; $0{,}1/(0{,}01 + 0{,}01) = 5{,}0$; $0{,}01/(0{,}0001 + 0{,}01) \approx 0{,}99$. На частоті, де розмиття слабке, обидва фільтри майже однакові; на частоті з $G = 0{,}1$ фільтр Вінера підсилює вдвічі менше; на частоті з $G = 0{,}01$, де інверсний фільтр підсилив би шум у 100 разів, фільтр Вінера практично не підсилює. Максимальне підсилення фільтра Вінера дорівнює $1/\left(2\sqrt{K}\right) = 5$ і досягається при $G = \sqrt{K}$.

Задача 8.2. Задано матрицю згортки третього порядку $\mathrm{G} = \begin{bmatrix} 0{,}6 & 0 & 0{,}4 \\ 0{,}4 & 0{,}6 & 0 \\ 0 & 0{,}4 & 0{,}6 \end{bmatrix}$, спостереження $\mathrm{L} = [1{,}0;\ 0{,}6;\ 0{,}2]^T$, оператор гладкості — циркулянтна матриця, утворена з $[-2,1,1]$. Дослідіть залежність оцінки (8.18) від λ.

Розв'язання. Точне обернення дає $\mathrm{G}^{-1}\mathrm{L} \approx [1{,}457;\ 0{,}029;\ 0{,}314]^T$ — оцінку з великим перепадом між першим і другим відліками. Розв'язуючи систему (8.17) для різних λ, отримуємо:

Таблиця 8.2

Оцінки умовної деконволюції для різних λ

λ	$\hat{f}_0$	$\hat{f}_1$	$\hat{f}_2$	Енергія нев'язки
1	0,626	0,583	0,591	0,301

λ	$\hat{f}_0$	$\hat{f}_1$	$\hat{f}_2$	Енергія нев'язки
10	0,803	0,464	0,532	0,186
100	1,249	0,168	0,384	0,019
10^6	1,457	0,029	0,314	$3 \cdot 10^{-10}$

При малих λ домінує умова гладкості, і оцінка наближається до сталої; енергія нев'язки велика. Зі зростанням λ оцінка прямує до точного обернення, а нев'язка — до нуля. Якщо, наприклад, енергія шуму в спостереженнях $\varepsilon^2 = 0{,}02$, то принцип нев'язки вибере $\lambda \approx 100$, і оцінка буде компромісом між гладкістю і узгодженістю з даними.

Задача 8.3. Обчисліть коефіцієнт передачі фільтра умовної деконволюції (8.22) на частоті, де $G = 0{,}2$ (дійсне), $|C|^2 = 4$, при $\lambda = 100$.

Розв'язання. $K_{DF} = G/(G^2 + |C|^2/\lambda) = 0{,}2/(0{,}04 + 0{,}04) = 2{,}5$. Для порівняння, інверсний фільтр дав би $1/G = 5$. Доданок $|C|^2/\lambda = 0{,}04$, що дорівнює G^2, удвічі зменшує підсилення на цій частоті.

Висновки до розділу 8

1. Реставрація — обернена до згортки некоректна задача; її розв'язання потребує апріорної інформації, вид і обсяг якої визначають метод.
2. Інверсна фільтрація точна лише за відсутності шуму; за наявності навіть слабкого шуму вона непридатна через підсилення високочастотних складових. Псевдоінверсна фільтрація обмежує підсилення, але потребує ручного вибору порогу і спричиняє ефект Гіббса.
3. Фільтр Вінера мінімізує середньоквадратичну похибку для класу образів і шумів з відомими спектрами потужності; він плавно переходить від інверсного режиму на частотах з високим відношенням сигнал/шум до придушення на частотах, де переважає шум.
4. Метод умовної деконволюції потребує лише енергії шуму і припущення про гладкість образу. Задача з двома квадратичними умовами розв'язується методом множників Лагранжа і зводиться до системи лінійних рівнянь з одним невідомим скалярним параметром λ, що визначається ітераційно з умови рівності енергії нев'язки енергії шуму.
5. У натуральних координатах метод потребує обернення матриць порядку M (для двовимірного образу — M^2), що для реальних форматів практично неможливо. В області трансформант Фур'є завдяки діагоналізації циркулянтних матриць розв'язок записується поелементно, і реставрація образу 1024×1024 стає швидкою задачею.
6. Фільтр умовної деконволюції за формою подібний до фільтра Вінера, але замість відношення спектрів потужності містить зважений спектр оператора гладкості і налаштовується для кожного конкретного образу. Математично метод є регуляризацією за Тихоновим з вибором параметра за принципом нев'язки.
7. Узагальнений фільтр найменших квадратів у довільному ортогональному базисі (8.28) поширює ідею фільтра Вінера на неоднорідні поля і на базиси, відмінні від Фур'є, ціною недіагональної структури матриць.

8. Деконволюція за критерієм форми результуючої імпульсної характеристики зводиться до задачі на власні значення; коригуючий фільтр обчислюється один раз для системи.
9. Реставрація в базисі Фур'є чутлива до крайових ефектів і ефекту Гіббса, що мотивує пошук базисів, вільних від цього недоліку.

Питання для обговорення та самоперевірки

1. Сформулюйте задачу реставрації образу. Яку апріорну інформацію використовують різні методи?
2. Чому інверсна фільтрація непридатна за наявності шуму? Поясніть, використовуючи частотну характеристику розмиття.
3. Що таке псевдоінверсний фільтр? Які його недоліки?
4. Виведіть фільтр Вінера. Як він поводиться на частотах з високим і низьким відношенням сигнал/шум?
5. Яку апріорну інформацію потребує метод умовної деконволюції? Як її отримати?
6. Чому одна умова на енергію шуму не визначає оцінку однозначно?
7. Запишіть критерій гладкості. Чому для нього використовують другі різниці?
8. Як за допомогою методу Лагранжа отримують розв'язок (8.18)?
9. Опишіть ітераційну процедуру визначення λ. Як змінюється оцінка зі зміною λ?
10. Оцініть обчислювальну складність умовної деконволюції в натуральних координатах для одновимірного і двовимірного образів.
11. Чому перехід в область трансформант Фур'є радикально спрощує метод?
12. Запишіть коефіцієнт передачі фільтра умовної деконволюції і рівняння для λ.
13. У чому подібність і принципова відмінність фільтра умовної деконволюції від фільтра Вінера?
14. Як пов'язаний метод умовної деконволюції з регуляризацією Тихонова і принципом нев'язки?
15. Чому виникають крайові ефекти при реставрації за циклічною моделлю? Як їх зменшити?
16. Як ефект Гіббса проявляється при реставрації в базисі Фур'є?
17. Запишіть узагальнений фільтр найменших квадратів у довільному базисі. Коли він зводиться до фільтра Вінера?
18. Сформулюйте критерій форми результуючої імпульсної характеристики. Чому результуюча характеристика не повинна бути дельта-функцією?
19. Чому задача деконволюції за критерієм форми зводиться до задачі на власні значення?

Завдання для самостійної роботи

1. Перевірте чисельно, що розв’язок (8.18) з циркулянтними матрицями збігається з розв’язком (8.21) у частотній області.
2. Виконайте покоординатне виведення похідних функціоналу (8.16) для матриць третього порядку і переконайтеся, що отримується вираз (8.17).
3. Реалізуйте умовну деконволюцію в натуральних координатах для одновимірного образу. Дослідіть залежність часу обчислень від формату $M = 64,128, \dots ,2048$ і порівняйте з реалізацією через ШПФ.
4. Побудуйте залежності PSNR результату від λ для кількох рівнів шуму і перевірте, наскільки значення λ^*, отримане за принципом нев’язки, близьке до оптимального за PSNR.
5. Дослідіть вплив похибки оцінки енергії шуму ($\pm 20\,\%$, $\pm 50\,\%$) на результат умовної деконволюції.
6. Замініть оператор гладкості (лапласіан) оператором перших різниць і порівняйте результати для образів з різкими межами і з плавними переходами.
7. Реалізуйте метод L-кривої для вибору параметра регуляризації і порівняйте з принципом нев’язки.
8. Змоделюйте лінійне (нециклічне) розмиття і порівняйте способи зменшення крайових ефектів: доповнення нулями, дзеркальне продовження, згладжування країв.
9. Реалізуйте деконволюцію за критерієм форми результуючої імпульсної характеристики для гаусівського розмиття з двома видами штрафної функції. Застосуйте отриманий фільтр до тестового образу і порівняйте з умовною деконволюцією.

РОЗДІЛ 9. РЕСТАВРАЦІЯ ОБРАЗІВ У БАЗИСАХ ПЕРЕТВОРЕНЬ RTF ТА АДАМАРА

Навчальні цілі. Після вивчення розділу ви зможете: пояснити мотивацію переходу від базису Фур'є до базисів RTF та Адамара в задачі реставрації; формувати частотні характеристики систем у базисі RTF і переходити від спектрів Фур'є до спектрів RTF; записувати метод умовної деконволюції в довільному ортогональному базисі; пояснювати блочно-діагональну структуру матриць деградації в базисі Адамара і символьний метод обчислення їх елементів; оцінювати обчислювальну складність алгоритмів реставрації в різних базисах; розуміти, які властивості результату реставрації залежать від вибору базису, а які — ні.

9.1. Вступні зауваження

Метод умовної деконволюції в області трансформант Фур'є (розділ 8) поєднує мінімальні вимоги до апріорної інформації з високою швидкодією. Проте він має недоліки, зумовлені природою базису Фур'є: обмежена кількість трансформант призводить до ефекту Гіббса, який проявляється осциляціями і «розмиванням» меж графоелементів реставрованого образу, а циклічна модель згортки — до крайових ефектів [2]. Природно постає питання про пошук координатних базисів, вільних від цих недоліків, але таких, що зберігають позитивні риси перетворення Фур'є — передусім просту форму теорем про згортку і диференціювання.

У дисертаційному дослідженні [2] і пов'язаних публікаціях [4–11] запропоновано два напрямки розв'язання цієї задачі:

- перехід до перетворення RTF, у якому «частотами» є власні числа різницевого оператора (підрозділ 5.8.3); воно є «калькою» різницевих методів у частотній області і вільне від осциляцій, пов'язаних з обмеженою кількістю членів ряду Фур'є;
- перехід до базису перетворення Адамара, функції якого прямокутні і краще пристосовані до подання різких меж; ціною є блочно-діагональна (а не діагональна) структура матриць деградації, для роботи з якою розроблено економічні символьні алгоритми.
- У цьому розділі розглядаються обидва підходи.

9.2. Перетворення RTF

9.2.1. Теорема про диференціювання оригіналу в базисі RTF

Нагадаємо результати підрозділу 5.8. Для дискретних методів аналізу похідна замінюється різницею, а різницевий оператор Δ (5.20) є циркулянтною матрицею, яку нормований оператор ДПФ $F_н$ приводить до діагонального вигляду:

$$F_н \, \Delta \, F_н^* = \beta = \mathrm{diag}\{\beta_0, \beta_1, \dots, \beta_{N-1}\}, \qquad (9.1)$$

де власні числа

$$\beta_k = 1 - e^{-j\frac{2\pi}{N}k} = 2j\sin\left(\frac{\pi}{N}k\right)e^{-j\frac{\pi}{N}k} \tag{9.2}$$

розташовані на колі одиничного радіуса з центром у точці $(1,0)$ (рис. 5.10) [2, 5]. Власні числа будь-якої матриці $\mathrm{A}_W = \mathrm{W}\Delta\mathrm{W}^H$ для будь-якого ортогонального перетворення W однакові; власні вектори різні, а у випадку $\mathrm{W} = \mathrm{F}_\mathrm{н}$ матриці власних векторів одиничні. Звідси теорема про диференціювання оригіналу для перетворення RTF:

$$\mathrm{X}_d = \beta\,\mathrm{X}, \tag{9.3}$$

тобто диференціюванню (різниці) в натуральних координатах відповідає множення спектра на діагональну матрицю β [2]. Аналогічно формулюється парна теорема про диференціювання спектра: для неї множником є β^*, а $\beta_1 = 1 - e^{j2\pi/N} \approx -j\,2\pi/N$ має зміст нормованого часу.

Нормовані частоти класичного ДПФ дорівнюють $\omega_k = 2\pi k/N$. Розбіжність між β_k і $j\omega_k$ пояснюється тим, що в класичному перетворенні Фур'є диференціювання виконується для аналітичних функцій у точці, тоді як у перетворенні RTF похідні обчислюються як різниці відліків, поділені на інтервал між ними (у нашому випадку одиничний). Для першої трансформанти $e^{-j2\pi/N} - 1 \approx -j\omega_1$ при $2\pi/N \ll 1$; для більших частот у розкладі (9.2) в ряд Тейлора дедалі суттєвішими стають члени вищих порядків, і похибка зростає з наближенням номера k до $N/2$ [2].

9.2.2. Аналіз лінійних систем у базисі RTF

Застосовуючи (9.3) до лінійного диференціального рівняння (5.22), отримуємо в області трансформант RTF алгебраїчне рівняння з діагональними матрицями, а в натуральних координатах —

$$\mathrm{y} = \mathrm{F}_\mathrm{н}^*\,\mathrm{K}(\beta)\,\mathrm{F}_\mathrm{н}\,\mathrm{x}, \tag{9.4}$$

де $\mathrm{K}(\beta)$ — діагональна матриця значень функції системи, у якій змінна p операторного запису замінена на β_k. Для порівняння, класичний спектральний метод дає

$$\mathrm{y} = \mathrm{F}_\mathrm{н}^*\,\mathrm{K}(j\omega)\,\mathrm{F}_\mathrm{н}\,\mathrm{x}. \tag{9.5}$$

Вираз (9.4) описує в матричній формі аналіз лінійної системи в базисі перетворення RTF [2, 5]. Перетворення RTF вносить додаткове згасання зі зростанням частоти порівняно з класичним ДПФ ($|\beta_k| < \omega_k$). Проте при правильному виборі кроку дискретизації це згасання може стати перевагою: при диференціюванні сигналів з розривами класичне ДПФ дає паразитні осциляції, зумовлені обмеженою кількістю трансформант, тоді як перетворення RTF дає задовільний результат (рис. 5.11).

9.2.3. Формування частотної характеристики в базисі RTF

Для практичного використання перетворення RTF необхідно вміти переходити від спектрів Фур'є до спектрів RTF і навпаки. Залежно від наявної інформації можливі різні алгоритми [2, 5, 6].

Відома схемна модель системи. Закон Ома для комплексних амплітуд $Z_R = R$, $Z_C = 1/(j\omega C)$, $Z_L = j\omega L$ у базисі RTF набуває вигляду

$$Z_{R\beta} = R, \qquad Z_{C\beta} = \frac{1}{\beta C}, \qquad Z_{L\beta} = \beta L. \tag{9.6}$$

Тому, якщо відома електрична схема (або інша фізична модель) пристрою, відомими методами теорії кіл знаходять операторну функцію системи

$$K(p) = \frac{b_n p^n + \cdots + b_1 p + b_0}{a_m p^m + \cdots + a_1 p + a_0}, \tag{9.7}$$

і замінюють у ній символ p на $j\omega_k$ (для перетворення Фур'є) або на β_k (для перетворення RTF) [2].

Відома лише частотна характеристика. Складніший випадок виникає, коли схемна модель невідома, а є лише дискретні відліки амплітудно- і фазо-частотної характеристики в базисі Фур'є. Тоді можна скористатися розкладом функції системи в ряд Тейлора за степенями p:

$$K(p) = K_{T0} + K_{T1}\, p + K_{T2}\, p^2 + \cdots + K_{Tb}\, p^b + \cdots, \tag{9.8}$$

де коефіцієнти виражаються через похідні функції системи. Похідні обчислюють як різниці сусідніх значень частотної характеристики, поділені на інтервал $2\pi/N$. Для переходу від аргументу $j\omega_k$ до β_k кожний k-й відлік b-ї похідної слід помножити на коригуючий множник $(\beta_k/j\omega_k)^b$; для зворотного переходу — на $(j\omega_k/\beta_k)^b$ [2, 5].

Модулі коригуючих множників для $N = 16$ наведено в табл. 9.1 і на рис. 9.1. Аргумент множника при $b = 1$ дорівнює $-\pi k/N$ (від $-11{,}25°$ при $k = 1$ до $-90°$ при $k = 8$), а для b-ї похідної — у b разів більший.

Таблиця 9.1

Модулі коригуючих множників $|\beta_k/j\omega_k|^b$ при N = 16

b	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
1	0,9936	0,9745	0,9432	0,9003	0,8469	0,7842	0,7136	0,6366
2	0,9872	0,9496	0,8896	0,8106	0,7173	0,6150	0,5092	0,4053
4	0,9746	0,9018	0,7913	0,6570	0,5145	0,3782	0,2593	0,1643
8	0,9498	0,8133	0,6262	0,4317	0,2647	0,1430	0,0672	0,0270

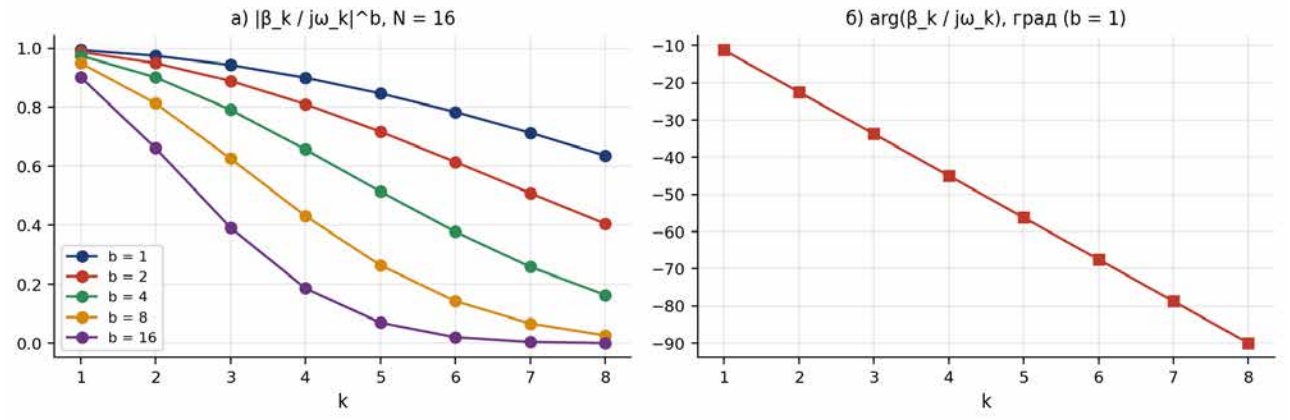


Рис. 9.1. Коригуючі множники переходу від перетворення Фур'є до перетворення RTF

З таблиці видно, що на найвищій частоті ($k = N/2$) загасання для першої похідної становить $1/0{,}6366 \approx 1{,}57$ раза, а для восьмої — понад 35 разів; на першій частоті загасання майже відсутнє для всіх порядків. Оскільки коефіцієнти ряду (9.8) у загальному випадку невідомі, невідомі і ваги, з якими слід підсумовувати частотні залежності похідних; проте в будь-якому разі перехід від Фур'є до RTF еквівалентний множенню спектра на характеристику деякого фільтра нижніх частот [2]. Це корисно, бо придушує паразитні осциляції, але може бути шкідливим, бо на високих частотах спотворюються частотні характеристики, наприклад, коливальних контурів. Для запобігання цьому сигнали нормують так, щоб головна частина їх спектрів потрапляла в область відносно низьких частот — тобто вибирають крок дискретизації, що забезпечує відносно велику кількість відліків на період коливання.

9.3. Умовна деконволюція в базисі RTF

Фільтрацію спотвореного образу фільтром умовної деконволюції (8.22) у матричній формі (для одновимірного випадку) можна записати як

$$\hat{\mathrm{f}} = \mathrm{F}_{\mathrm{н}}^{*}\,\mathrm{K}_{\omega}\,\mathrm{F}_{\mathrm{н}}\,\mathrm{L}, \tag{9.9}$$

де K_{ω} — діагональна матриця коефіцієнтів передачі фільтра на частотах ω_k. При переході від різницевих рівнянь до рівнянь у частотній області можливий перехід до перетворення RTF [2, 6]; тоді (9.9) набуває вигляду

$$\hat{\mathrm{f}} = \mathrm{F}_{\mathrm{н}}^{*}\,\mathrm{K}_{\beta}\,\mathrm{F}_{\mathrm{н}}\,\mathrm{L}, \tag{9.10}$$

де в діагональній матриці K_{β} аргументами частотних характеристик системи і оператора гладкості є не частоти $j\omega_k$, а власні частоти β_k. Отже, для реставрації в базисі RTF достатньо сформувати відповідні характеристики як функції β_k, обчислити оцінку за (8.21) для поточного λ і енергію нев'язки, і далі діяти за тим самим ітераційним алгоритмом, що й у базисі Фур'є.

Зауважимо, що для стабілізуючого оператора другої різниці його ДПФ точно дорівнює $-\beta_k\beta_k^{*}$, тобто $|C(k)|^2 = |\beta_k|^4$: оператор гладкості в методі умовної деконволюції *від початку* є різницевим і природно описується в термінах перетворення RTF.

Розглянемо обчислювальний приклад, що ілюструє, коли саме перетворення RTF дає перевагу. Нехай система відображення описується диференціальним рівнянням першого порядку $\tau\, y' + y = x$ (RC-ланка), а її дискретна реалізація — відповідним різницевим рівнянням. Тоді точна дискретна частотна характеристика системи дорівнює $G(\beta_k) = 1/(1 + \tau\beta_k)$. Якщо ж під час реставрації використати «класичну» модель $G(j\omega_k) = 1/(1 + \tau j\omega_k)$, отриману підстановкою частот ДПФ у функцію неперервної системи, то модель деградації буде неточною, і ця невідповідність буде тим більшою, чим вища частота. На рис. 9.2 наведено результати умовної деконволюції одновимірного образу з прямокутними елементами, спотвореного такою системою з $\tau = 4$ і шумом. Реставрація з моделлю $G(j\omega_k)$ менш точно відтворює перепади і містить більші залишкові осциляції (рис. 9.2, б); реставрація в базисі RTF з моделлю $G(\beta_k)$,

узгодженою з різницевим описом системи, вільна від них і має суттєво вище відношення сигнал/шум (рис. 9.2, в).

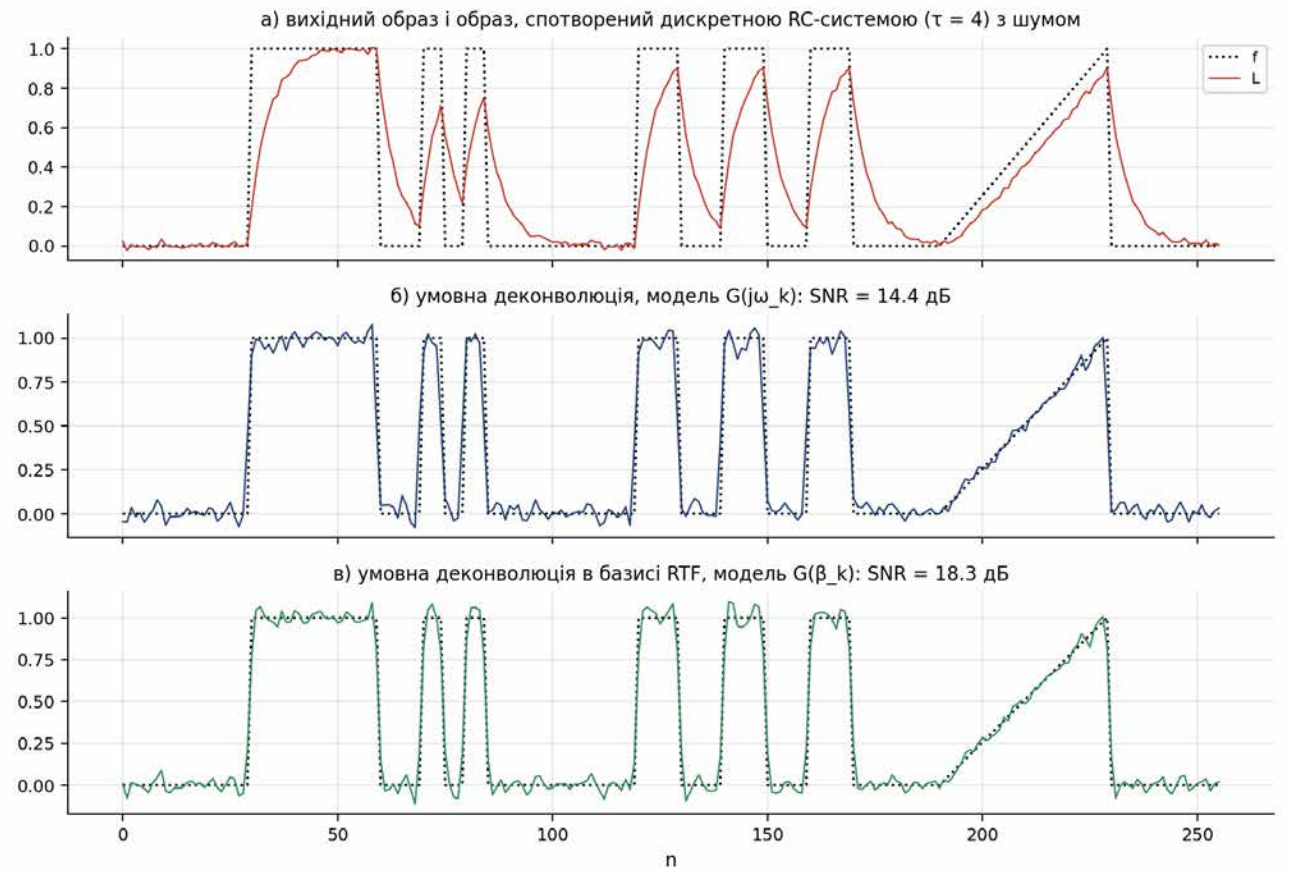


Рис. 9.2. Умовна деконволюція з моделлю системи в базисі Фур'є (б) і RTF (в)

За результатами експериментів, наведених у [2, 6], реставрація в базисі перетворення RTF методом умовної деконволюції дає задовільні результати і для двовимірних образів. Певна зміна динамічного діапазону реставрованих образів може бути скомпенсована подальшою корекцією шкали рівнів сірого. Помітніше згасання амплітудних характеристик у базисі RTF порівняно з Фур'є пояснюється тим, що за великого кроку дискретизації дрібні деталі (зокрема періоди високочастотних коливань) стають сумірними з цим кроком; нормування спектрів або збільшення формату перетворення дозволяє зменшити цей ефект [2].

9.4. Умовна деконволюція в області трансформант Адамара

9.4.1. Загальний вираз для довільного ортогонального базису

Повернімося до розв'язку умовної деконволюції в натуральних координатах (8.18):

$$\hat{\mathrm{f}} = \left[\mathrm{G}^T\mathrm{G} + \frac{1}{\lambda}\mathrm{C}^T\mathrm{C}\right]^{-1}\mathrm{G}^T\mathrm{L}. \tag{9.11}$$

Формальний перехід до області трансформант довільного ортогонального перетворення з нормованим матричним оператором $\mathrm{W_н}$ ($\mathrm{W_н}\mathrm{W_н^{*T}} = \mathrm{E}$)

виконується множенням (9.11) зліва на $\mathrm{W_н}$ і вставлянням між співмножниками одиничних матриць $\mathrm{W_н^{*T}W_н}$ [2]:

$$\hat{\mathrm{f}}_\xi = \left[\mathrm{G}_{2\xi} + \frac{1}{\lambda}\mathrm{C}_{2\xi}\right]^{-1} \mathrm{G}_{1\xi}^{T}\,\mathrm{L}_\xi, \tag{9.12}$$

де $\hat{\mathrm{f}}_\xi = \mathrm{W_н}\hat{\mathrm{f}}$, $\mathrm{L}_\xi = \mathrm{W_н L}$ — спектри оцінки і спотвореного образу; $\mathrm{G}_{1\xi} = \mathrm{W_н G W_н^{*T}}$ — оператор згортки в області трансформант; $\mathrm{G}_{2\xi} = \mathrm{W_н G}^T\mathrm{G W_н^{*T}} = \mathrm{G}_{1\xi}^{*T}\mathrm{G}_{1\xi}$ — «спектр» квадратичної форми оператора згортки; $\mathrm{C}_{2\xi} = \mathrm{W_н C}^T\mathrm{C W_н^{*T}}$ — перетворений оператор гладкості. Усі квадратні матриці мають порядок N.

Для базису Фур'є всі матриці в (9.12) діагональні, і (9.12) переходить у поелементну формулу (8.21). Для базису Адамара (з нормованим симетричним оператором $\mathrm{Had_н}$) вираз (9.12) набуває вигляду [2, 4, 9]

$$\hat{\mathrm{f}}_{Had} = \left[\mathrm{G}_{2Had} + \frac{1}{\lambda}\mathrm{C}_{2Had}\right]^{-1} \mathrm{G}_{1Had}^{T}\,\mathrm{L}_{Had}, \tag{9.13}$$

де

$$\mathrm{G}_{1Had} = \mathrm{Had_н\,G\,Had_н^T}, \quad \mathrm{G}_{2Had} = \mathrm{G}_{1Had}^{T}\mathrm{G}_{1Had}, \quad \mathrm{C}_{2Had} = \mathrm{Had_н\,C}^T\mathrm{C\,Had_н^T}. \tag{9.14}$$

Оскільки G і C — циркулянтні матриці, усі матриці (9.14), як показано в розділі 5, є блочно-діагональними з блоками порядків $1,1,2,4,\ldots,N/2$. Добуток і сума блочно-діагональних матриць з однаковою структурою блоків мають ту саму структуру, тому і матриця в квадратних дужках (9.13), і обернена до неї також блочно-діагональні (рис. 9.3). Отже, обернення матриці порядку N розпадається на обернення незалежних блоків, а матрично-векторні добутки виконуються лише для ненульових блоків.

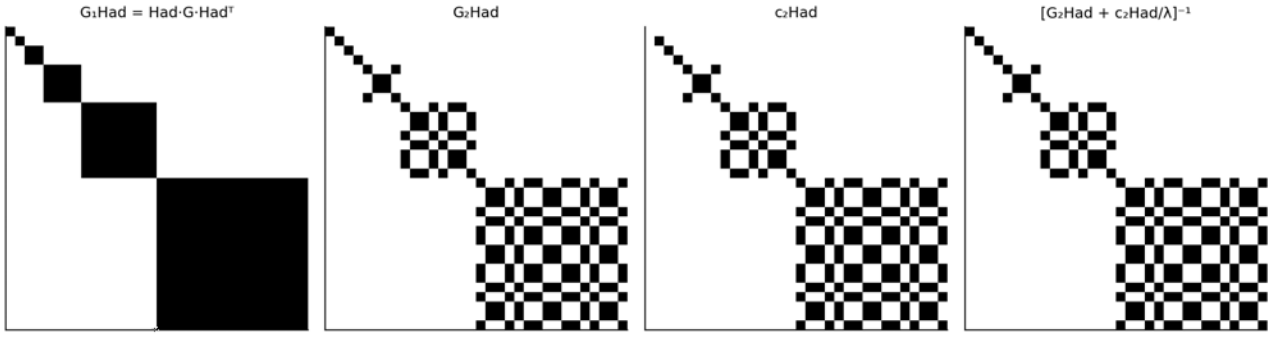


Рис. 9.3. Структура матриць методу умовної деконволюції в базисі Адамара (N = 32)

9.4.2. Структура блоків матриці G₁Had

Розглянемо детальніше матрицю G_{1Had}, оскільки з неї утворюється матриця G_{2Had}. Перші два елементи діагональні:

$$g_{H1,00} = g_0 + g_1 + g_2 + \cdots + g_{N-1}, \qquad g_{H1,11} = g_0 - g_1 + g_2 - g_3 + \cdots + g_{2n} - g_{2n+1} + \cdots, \tag{9.15}$$

де g_i — i-й відлік імпульсної характеристики, $N = 2^p$ — формат перетворення. (Оскільки імпульсна характеристика зазвичай значно коротша за

формат образу, рядки матриці G доповнюються нулями.) Елементи другого і третього рядків утворюють матрицю другого порядку:

$$g_{H1,22} = g_{H1,33} = g_0 - g_2 + g_4 - g_6 + \cdots, \qquad g_{H1,23} = -g_{H1,32} = g_1 - g_3 + g_5 - g_7 + \cdots \tag{9.16}$$

Наступним є блок четвертого порядку (рядки і стовпці 4–7), елементи якого містять уже і половинні коефіцієнти, наприклад [2, 7]:

$$g_{H1,44} = g_0 + 1/2\,(g_1 - g_3) - g_4 - 1/2\,(g_5 - g_7) + g_8 + \cdots, \qquad g_{H1,45} = 1/2\,[(g_1 + g_3) - (g_5 + g_7) + (g_9 + g_{11}) - \cdots]. \tag{9.17}$$

Далі йдуть блоки восьмого порядку (рядки і стовпці 8–15), шістнадцятого (16–31) і так далі. Формули (9.15)–(9.17) легко перевірити чисельно (див. табл. 9.2 і завдання до розділу).

Будь-який блок порядку $M = 2^m$ має однакову регулярну структуру [2]:

$$\mathrm{G}_{(a\ldots b)} = \begin{bmatrix} \mathrm{P} & \mathrm{Q} \\ -\mathrm{Q} & \mathrm{P} \end{bmatrix}, \tag{9.18}$$

де P, Q — підблоки порядку $M/2$. Отже, для формування будь-якого блоку достатньо сформувати лише дві його підматриці, що вдвічі зменшує обсяг обчислень.

9.4.3. Символьний метод формування матриць

Безпосереднє обчислення добутку (9.14) у числовій формі вимагає порядку $2N^3$ операцій і призводить до накопичення операційної похибки: як видно з (9.15)–(9.17), значна частина доданків у процесі обчислень взаємно знищується, а при обмеженій розрядності операндів таке «знищення» неточне [2]. Для форматів одновимірних образів $N = 128 \ldots 512$ і двовимірних $N^2 \approx 2{,}5 \cdot 10^4 \ldots 4 \cdot 10^6$ це суттєво.

У [2, 7, 11] запропоновано символьний метод: кожний ненульовий елемент $g_{H1,kl}$ записується як лінійна комбінація відліків імпульсної характеристики з *цілими (або половинними)* коефіцієнтами, які обчислюються наперед і зберігаються в масивах $\mu_{k,l}$:

$$g_{H1,kl} = \sum_{i=0}^{D-1} \mu_{k,l}\,(i)\; g_i, \tag{9.19}$$

де D — кількість відліків імпульсної характеристики. Коефіцієнти отримують так. Елемент (k, l) добутку $\mathrm{Had\,G\,Had}^T$ дорівнює $\sum_{p,q} h_{kp}\; g_{(p-q) \bmod N}\, h_{lq}$. Утворимо матрицю знаків $\varphi_{kl} = \mathrm{had}_k^T \mathrm{had}_l$ — зовнішній добуток k-го рядка матриці Адамара (як стовпця) на l-й рядок. Порівнюючи її з циркулянтною структурою G, бачимо, що коефіцієнт при відліку g_i дорівнює сумі елементів ± 1 матриці φ_{kl}, розташованих на тих (циклічних) діагоналях, на яких у матриці G стоїть g_i. Наприклад, коефіцієнт при g_0 — сума елементів головної діагоналі φ_{kl}, а коефіцієнт при g_3 — сума елементів третьої нижньої і п’ятої верхньої піддіагоналей (для $N = 8$) [2].

Для $N = 8$ коефіцієнти перших блоків (після ділення на N) наведено в табл. 9.2. Вони точно відтворюють формули (9.15)–(9.17).

Таблиця 9.2

Коефіцієнти $\mu_{k,l}(i)$ символьного подання елементів матриці G_{1Had} (N = 8)

Елемент	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7
$g_{H1,00}$	1	1	1	1	1	1	1	1
$g_{H1,11}$	1	−1	1	−1	1	−1	1	−1
$g_{H1,22} = g_{H1,33}$	1	0	−1	0	1	0	−1	0
$g_{H1,23} = -g_{H1,32}$	0	1	0	−1	0	1	0	−1
$g_{H1,44}$	1	1/2	0	−1/2	−1	−1/2	0	1/2
$g_{H1,45}$	0	1/2	0	1/2	0	−1/2	0	−1/2
$g_{H1,46}$	0	1/2	1	1/2	0	−1/2	−1	−1/2
$g_{H1,55}$	1	−1/2	0	1/2	−1	1/2	0	−1/2

Переваги символьного методу [2]:

1. обчислюються лише ненульові елементи матриці G_{1Had}, для чого перемножуються лише ті рядки і стовпці операторів Адамара, які утворюють відповідні елементи добутку;
2. коефіцієнти при відліках g_i обчислюються підсумовуванням цілих чисел ± 1, що виключає накопичення операційної похибки, пов'язане з обмеженою розрядністю;
3. масиви коефіцієнтів залежать лише від формату N і не залежать від конкретної імпульсної характеристики, тому їх можна обчислити один раз і зберігати;
4. використання регулярної структури блоків (9.18) дозволяє отримати символьний запис будь-якого елемента матриці G_{1Had} довільного порядку.

Алгоритм обчислення блоків матриці G_{1Had}:

1. для заданого формату $N = 2^p$ визначити номери рядків і стовпців матриці Адамара, що відповідають рядкам і стовпцям окремих блоків діагоналі G_{1Had};
2. для кожної пари (k, l) у межах блоку утворити матрицю знаків $\varphi_{kl} = \mathrm{had}_k^T \mathrm{had}_l$ і обчислити коефіцієнти $\mu_{k,l}$;
3. обчислити значення елементів $g_{H1,kl}$ за (9.19) і сформувати блочно-діагональну матрицю, використовуючи властивість (9.18).

Аналогічний символьний підхід застосовано до формування матриць G_{2Had} і C_{2Had} [2, 7].

9.4.4. Обернення матриці деградації

Матриця $[\mathrm{G}_{2Had} + \lambda^{-1}\mathrm{C}_{2Had}]$ блочно-діагональна, і її обернення зводиться до обернення окремих блоків. Однак найбільший блок має порядок $N/2$, тому в загальному випадку обернення не простіше за порядком величини, ніж обернення в натуральних координатах: обсяг обчислень зменшується приблизно на 70–80 % (рис. 9.4), але залишається кубічним за N [2]. Для радикального спрощення в [2, 8, 10] запропоновано використовувати кратні ортогональні перетворення: регулярна структура блоків (9.18) дозволяє застосувати до

кожного блоку додаткове ортогональне перетворення, що приводить його до діагонального (або простішого блочного) вигляду, після чого обернення виконується поелементно. Розроблена методика дозволяє з високою точністю обчислювати обернену матрицю будь-якого порядку.

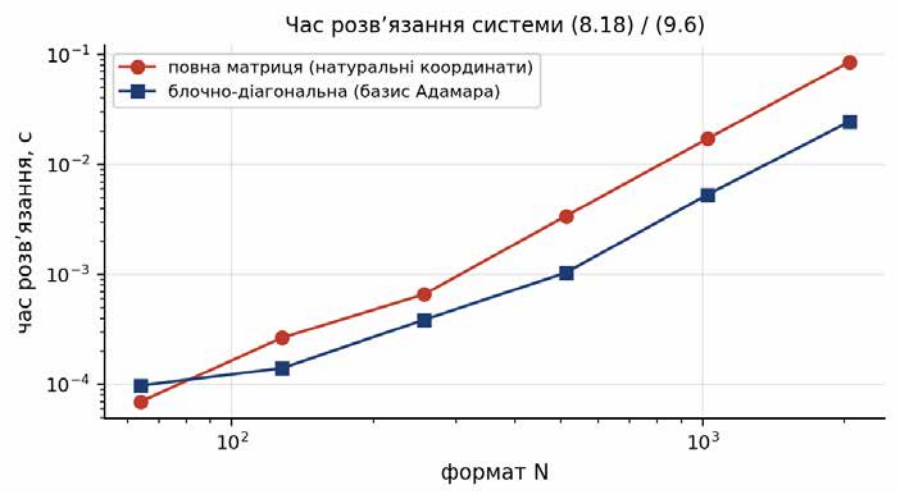


Рис. 9.4. Час розв'язання системи рівнянь умовної деконволюції з повною та блочно-діагональною матрицею

9.4.5. Алгоритм реставрації одновимірного образу в базисі Адамара

З урахуванням розглянутих алгоритмів реставрація одновимірного образу в базисі Адамара виконується так [2, 10]:

1. побудова матриці деградації (за відомою імпульсною характеристикою);
2. формування блочно-діагональних матриць G_{1Had}, G_{2Had}, C_{2Had} символьним методом;
3. обчислення спектра Адамара спотвореного образу (швидким ПУА);
4. для поточного λ: обернення (поблочне) матриці деградації і обчислення спектра оцінки за (9.13);
5. обчислення енергії нев'язки (у базисі Адамара — за рівністю Парсеваля) і корекція λ за тим самим принципом, що й у розділі 8;
6. обернене ПУА спектра оцінки.

9.5. Реставрація двовимірних образів

Для двовимірного образу формату $N \times N$ матриці деградації в натуральних координатах мають порядок N^2 ($N^2 = 65\,536 \ldots 1\,048\,576$), і обернення методом Гаусса вимагає $(N^2)^3 \approx 2{,}8 \cdot 10^{14} \ldots 1{,}2 \cdot 10^{18}$ операцій [2]. У базисі Фур'є ця складність обходиться легко: матриця, що підлягає оберненню, діагональна, і замість обернення матриці слід знайти N^2 обернених величин. Перехід до базису Адамара, викликаний бажанням підвищити якість реставрованого образу, знову приводить до ускладнень, оскільки матриця деградації блочно-діагональна. Проте, як і для одновимірного випадку, коректним є прийом використання кратних матричних ортогональних перетворень [2].

Двовимірна схема будується на основі сепарабельності двовимірного ПУА: $\mathrm{Had}_{2D} = \mathrm{Had}_{\text{н}} \otimes \mathrm{Had}_{\text{н}}$. Якщо двовимірна імпульсна характеристика

сепарабельна, $g(m,n) = g_r(m)\, g_c(n)$, то оператор згортки є кронекеровим добутком $\mathrm{G} = \mathrm{G}_r \otimes \mathrm{G}_c$, а в базисі Адамара

$$\mathrm{G}_{1Had,2D} = (\mathrm{Had}_{\mathrm{н}}\mathrm{G}_r\mathrm{Had}_{\mathrm{н}}^T) \otimes (\mathrm{Had}_{\mathrm{н}}\mathrm{G}_c\mathrm{Had}_{\mathrm{н}}^T), \qquad (9.20)$$

тобто кронекерів добуток двох блочно-діагональних матриць, структура якого повністю визначається одновимірними символьними масивами. Двовимірний оператор гладкості (лапласіан (8.14)) є сумою двох кронекерових добутків, і його перетворена матриця має ту саму блочну структуру. Алгоритм формування матриць деградації двовимірного образу, їх обернення та чисельні експерименти з реставрації двовимірних образів детально наведено в [2, 9, 10].

Для оцінки ефективності наведемо результати порівняння кількості трудомістких арифметичних операцій, отримані в [2]: для образу 256×256 реставрація модифікованим методом умовної деконволюції в базисах Адамара і RTF потребує близько 10^4 і 10^6 операцій відповідно (з урахуванням символьного формування і структури матриць), тоді як відомий метод реставрації в натуральних координатах — близько 10^8 операцій.

9.6. Порівняння результатів реставрації в різних базисах

9.6.1. Результати експериментальних досліджень

За результатами машинних експериментів, наведених у [2, 9], реставрація методом умовної деконволюції в базисі Адамара для розглянутих тестових образів (смугові та «шахові» структури, геометричні фігури, текстові фрагменти, спотворені неточковими імпульсними характеристиками типу $e^{-\alpha x}$ і $1/(1+\alpha x)$ та адитивним шумом різного рівня) дала якісно кращі результати, ніж у базисі Фур'є: на результатах, отриманих у базисі Фур'є, спостерігалося розмиття меж графоелементів, характерне для ефекту Гіббса. Кількісно, за даними [2], використання базису Адамара дозволило покращити відношення сигнал/шум реставрованих образів на величину від −3 до 7 дБ залежно від образу, а якість реставрованих зображень порівняно з умовною деконволюцією в базисі Фур'є — не менше ніж на 6 дБ.

Вибір «меандрових» тестових сигналів у [2] невипадковий: такий одновимірний сигнал відповідає одному рядку значень яскравості, перпендикулярному до чергування чорних і білих смуг на двовимірному образі, а на одновимірних графіках простежуються деталі, непомітні на двовимірному зображенні. Приклад реставрації одновимірного образу такого типу в області трансформант Адамара наведено на рис. 9.5.

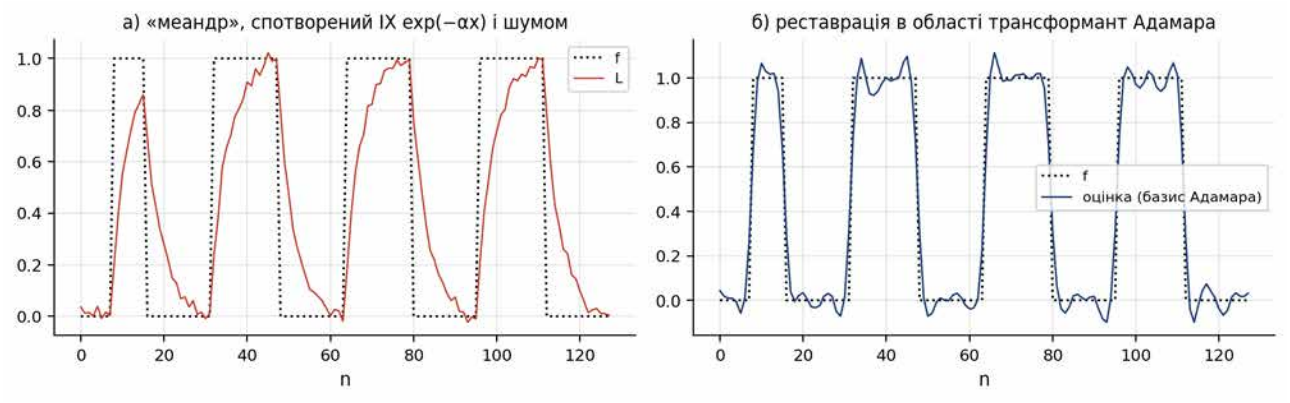


Рис. 9.5. Реставрація одновимірного образу в області трансформант Адамара

9.6.2. Методичне зауваження: що залежить від базису, а що — ні

Під час вивчення методів реставрації в різних базисах важливо чітко розрізняти дві речі.

Математичний розв'язок не залежить від базису. Перехід (9.11) → (9.12) — це перетворення подібності з унітарною матрицею. Якщо модель деградації і оператор гладкості задані *одними й тими самими* матрицями G і C, а обчислення виконуються точно, то оцінка $\hat{\mathrm{f}} = \mathrm{W}_{\mathrm{н}}^{*T}\hat{\mathrm{f}}_{\xi}$ однакова в натуральних координатах, у базисі Фур'є, Адамара чи будь-якому іншому. Це легко перевірити чисельно: для образу на рис. 9.5 максимальна розбіжність між оцінками, отриманими за (9.11) і за (9.13) з подальшим оберненим ПУА, становить близько 10^{-15}, тобто лежить на рівні похибок округлення.

Від базису залежить реалізація і, отже, практичний результат. Відмінності між результатами реставрації в різних базисах, що спостерігаються на практиці, зумовлені тим, *як саме* алгоритм реалізується в кожному базисі:

- структурою матриць і, відповідно, обчислювальною складністю та накопиченням похибок округлення (діагональна, блочно-діагональна, щільна);
- тим, як враховуються крайові умови і чи узгоджена циклічна модель з реальною (лінійною) згорткою;
- тим, у якому вигляді задаються характеристики системи: як показує приклад з перетворенням RTF (рис. 9.2), модель $G(\beta_k)$, узгоджена з різницевим описом системи, і модель $G(j\omega_k)$ — це *різні* моделі, і вони дають різні результати;
- наближеннями, які зручно робити в тому чи іншому базисі (усічення спектра, наближення блоків, поелементна регуляризація в області трансформант), — ефект Гіббса виникає саме при усіченні або різкій зміні спектра Фур'є, тоді як прямокутні функції Адамара інакше реагують на подібні наближення.

Саме тому дослідження методів реставрації в різних базисах має практичний сенс, а порівняння результатів слід проводити за одним і тим самим об'єктивним критерієм, з однаковими моделями деградації і способами врахування меж образу.

9.7. Програмна реалізація

Нижче наведено реалізацію умовної деконволюції одновимірного образу в базисі Адамара з поблочним розв'язанням системи рівнянь, а також функцію для обчислення символьних коефіцієнтів $\mu_{k,l}$.

```
import numpy as np
from scipy.linalg import hadamard, circulant

def mu_coeffs(N, k, l):
    """Коефіцієнти символьного подання елемента (k, l) матриці G1Had (9.19)."""
    H = hadamard(N)
    phi = np.outer(H[k], H[l])              # матриця знаків
    mu = np.zeros(N)
    for p in range(N):
        for q in range(N):
            mu[(p - q) % N] += phi[p, q]    # діагоналі циркулянта
    return mu / N

def block_slices(N):                        # блоки 1, 1, 2, 4, ..., N/2
    s, out = 0, []
    for size in [1, 1] + [2**m for m in range(1, int(np.log2(N)))]:
        out.append(slice(s, s + size)); s += size
    return out

def cond_deconv_hadamard(L, g, eps2):
    N = L.size
    Hn = hadamard(N) / np.sqrt(N)
    col = np.zeros(N); col[:g.size] = g; G = circulant(col)
    cv = np.zeros(N); cv[[0, 1, N - 1]] = [-2, 1, 1]; C = circulant(cv)
    G1 = Hn @ G @ Hn.T; G2 = G1.T @ G1; C2 = Hn @ C.T @ C @ Hn.T  # (9.14)
    Lh = Hn @ L; rhs = G1.T @ Lh
    def estimate(lam):
        F = np.zeros(N)
        for s in block_slices(N):                       # поблочно
            F[s] = np.linalg.solve(G2[s, s] + C2[s, s] / lam, rhs[s])
        return F
    lo, hi = -4.0, 8.0
    for _ in range(50):                     # принцип нев'язки
        m = 0.5 * (lo + hi); F = estimate(10**m)
        if np.sum((Lh - G1 @ F)**2) > eps2: lo = m   # Парсеваль
        else: hi = m
    return Hn.T @ estimate(10**(0.5 * (lo + hi)))

# перевірка формул (9.15)-(9.17)
print(mu_coeffs(8, 4, 4))   # [1, .5, 0, -.5, -1, -.5, 0, .5]
```

У наведеному коді для наочності матриці (9.14) обчислюються безпосереднім множенням; у повноцінній реалізації їх слід формувати символьним методом лише для ненульових блоків.

9.8. Приклади розв'язання задач

Задача 9.1. Система описується рівнянням $\tau y' + y = x$ з $\tau = 2$ (у кроках дискретизації). Обчисліть значення частотної характеристики на трансформанті $k = 2$ при $N = 8$ у базисах Фур'є і RTF.

Розв'язання. Операторна функція $K(p) = 1/(1 + \tau p)$. У базисі Фур'є $j\omega_2 = j \cdot 2\pi \cdot 2/8 = j\pi/2 \approx 1{,}571j$, тому $K(j\omega_2) = 1/(1 + 3{,}142j) \approx 0{,}092 - 0{,}289j$, $|K| \approx 0{,}303$. У базисі RTF $\beta_2 = 1 - e^{-j\pi/2} = 1 + j$, тому $K(\beta_2) = 1/(1 + 2 + 2j) = 1/(3 + 2j) \approx 0{,}231 - 0{,}154j$, $|K| \approx 0{,}277$. Різниця модулів невелика, але фази суттєво різні: $-72{,}3°$ і $-33{,}7°$. Саме значення $K(\beta_k)$ точно відповідає різницевій моделі системи; використання $K(j\omega_k)$ для такої системи вносить похибку моделі, яка під час реставрації проявляється у вигляді залишкових осциляцій (рис. 9.2).

Задача 9.2. Для імпульсної характеристики $g = [0{,}5;\ 0{,}3;\ 0{,}2]$ при $N = 8$ обчисліть елементи блоку четвертого порядку матриці G_{1Had}, використовуючи табл. 9.2.

Розв'язання. За коефіцієнтами табл. 9.2 (враховуються лише g_0, g_1, g_2, оскільки інші відліки нульові):

$g_{H1,44} = 1 \cdot 0{,}5 + 0{,}5 \cdot 0{,}3 + 0 \cdot 0{,}2 = 0{,}65$;

$g_{H1,45} = 0 + 0{,}5 \cdot 0{,}3 + 0 = 0{,}15$;

$g_{H1,46} = 0 + 0{,}5 \cdot 0{,}3 + 1 \cdot 0{,}2 = 0{,}35$;

$g_{H1,55} = 0{,}5 - 0{,}5 \cdot 0{,}3 + 0 = 0{,}35$.

Ці значення збігаються з результатом безпосереднього обчислення $\mathrm{Had_н G Had_н^T}$. Символьне обчислення потребує лише кількох множень на коефіцієнти $\pm 1, \pm 1/2$ замість повного матричного добутку порядку N^3 операцій.

Задача 9.3. Оцініть кількість операцій для обернення блочно-діагональної матриці порядку $N = 1024$ у базисі Адамара порівняно з оберненням повної матриці (вважаючи, що обернення матриці порядку n потребує n^3 операцій).

Розв'язання. Розміри блоків: 1, 1, 2, 4, …, 512. Сума кубів $\approx 512^3(1 + 1/8 + 1/64 + \cdots) \approx 512^3 \cdot 8/7 \approx 1{,}53 \cdot 10^8$. Для повної матриці $1024^3 \approx 1{,}07 \cdot 10^9$. Виграш приблизно в 7 разів. Як зазначено в підрозділі 9.4.4, для радикального зменшення складності блоки додатково спрощують за допомогою кратних перетворень.

Висновки до розділу 9

1. Реставрація в базисі Фур'є, попри високу швидкодію, супроводжується проявами ефекту Гіббса і крайових ефектів, що мотивує пошук інших координатних базисів зі збереженням простої форми теорем про згортку і диференціювання.
2. У перетворенні RTF роль частот відіграють власні числа різницевого оператора $\beta_k = 1 - e^{-j2\pi k/N}$; теорема про диференціювання оригіналу в

ньому має таку саму просту діагональну форму, як для Фур'є, але точно відповідає різницевій схемі. Перехід від Фур'є до RTF еквівалентний низькочастотній фільтрації, що придушує паразитні осциляції; для збереження високочастотних деталей необхідна достатня частота дискретизації.

3. Частотні характеристики систем у базисі RTF отримують заміною $p \to \beta_k$ в операторній функції системи або перерахунком похідних частотної характеристики з коригуючими множниками $(\beta_k / j\omega_k)^b$.
4. Метод умовної деконволюції записується в довільному ортогональному базисі виразом (9.12); у базисі Адамара всі матриці блочно-діагональні з блоками порядків $1,1,2,\ldots,N/2$ і регулярною внутрішньою структурою.
5. Символьний метод формування матриць зберігає коефіцієнти при відліках імпульсної характеристики як цілі (половинні) числа, обчислює лише ненульові елементи і виключає накопичення похибок; кратні ортогональні перетворення дозволяють спростити обернення блоків.
6. За результатами досліджень [2] модифікації методу умовної деконволюції в базисах RTF і Адамара забезпечують якісно кращі результати для образів з різкими межами графоелементів, ніж у базисі Фур'є, при порівнянних витратах пам'яті і часу.
7. Математичний розв'язок задачі реставрації з однаковими моделями не залежить від ортонормованого базису; від базису залежать структура обчислень, точність, спосіб урахування меж і наближень, а отже — практичний результат.

Питання для обговорення та самоперевірки

1. Які недоліки реставрації в базисі Фур'є мотивують перехід до інших базисів?
2. Що таке власні частоти β_k перетворення RTF? Як вони пов'язані з різницевим оператором?
3. Сформулюйте теорему про диференціювання оригіналу для перетворення RTF.
4. Чому перехід від перетворення Фур'є до перетворення RTF еквівалентний фільтрації нижніх частот?
5. Як записується закон Ома для комплексних амплітуд у базисі RTF?
6. Як перейти від частотної характеристики системи в базисі Фур'є до характеристики в базисі RTF, якщо схемна модель невідома?
7. Що показують коригуючі множники $(\beta_k / j\omega_k)^b$? Як вони залежать від номера частоти і порядку похідної?
8. У яких випадках модель $G(\beta_k)$ точніша за $G(j\omega_k)$?
9. Запишіть розв'язок задачі умовної деконволюції в довільному ортогональному базисі.
10. Яку структуру мають матриці G_{1Had}, G_{2Had}, C_{2Had}? Чому обернена матриця також блочно-діагональна?
11. Запишіть елементи перших трьох блоків матриці G_{1Had}.

12. У чому полягає символьний метод формування матриць? Які його переваги?
13. Як за допомогою матриці знаків φ_{kl} визначити коефіцієнти при відліках імпульсної характеристики?
14. Чому поблочне обернення не змінює кубічний порядок складності? Як цю проблему розв'язують кратні перетворення?
15. Як будується схема реставрації двовимірних образів у базисі Адамара для сепарабельної імпульсної характеристики?
16. Чи залежить математичний розв'язок задачі реставрації від вибору ортонормованого базису? Від чого залежать практичні відмінності результатів?

Завдання для самостійної роботи

1. Обчисліть власні числа β_k для $N = 16$ і перевірте чисельно, що вони є власними числами циркулянтної матриці різниці «назад».
2. Обчисліть коригуючі множники $(\beta_k/j\omega_k)^b$ для $N = 32$ і $b = 1, \dots, 8$; побудуйте їх залежності від k і b.
3. Для RC-ланки ($\tau = 2,4,8$) змоделюйте спотворення образу різницевою моделлю і виконайте реставрацію з моделями $G(j\omega_k)$ та $G(\beta_k)$. Як залежить виграш від τ і від рівня шуму?
4. Для $N = 16$ обчисліть матрицю $\mathrm{Had_н G Had_н^T}$ для випадкової імпульсної характеристики довжини 16 і перевірте формули (9.15)–(9.17) та структуру (9.18).
5. Реалізуйте символьне формування матриці G_{1Had} для довільного $N = 2^p$ з обчисленням лише ненульових блоків. Порівняйте точність з прямим множенням матриць для $N = 512$ при обчисленнях з одинарною точністю.
6. Реалізуйте умовну деконволюцію одновимірного образу в базисах Фур'є, Адамара та в натуральних координатах з однаковою циркулянтною моделлю; переконайтеся в збігу результатів. Потім повторіть експеримент з лінійною (нециклічною) згорткою при моделюванні спотворення і порівняйте результати.
7. Оцініть кількість операцій для поблочного обернення матриці в базисі Адамара і для обернення повної матриці при $N = 2^6 \dots 2^{12}$.

РОЗДІЛ 10. РОЗПІЗНАВАННЯ ЗА ЕТАЛОНОМ: КОРЕЛЯЦІЙНІ МЕТОДИ, ПОГОДЖЕНА ФІЛЬТРАЦІЯ ТА НОРМАЛЬНЕ ПЕРЕТВОРЕННЯ

Навчальні цілі. Після вивчення розділу ви зможете: застосовувати кореляційні методи і нормовану взаємну кореляцію для пошуку еталона на зображенні; виводити лінійний погоджений фільтр з умови максимуму відношення сигнал/шум; пояснювати обмеження класичного погодженого фільтра для нестаціонарних образів і обирати альтернативні детектори подібності; нормалізувати ортогональні перетворення за кроком і за рівнем; будувати нормальне ортогональне перетворення для заданого еталона і використовувати коефіцієнт трансформант як міру відмінності; реалізовувати алгоритм аутентифікації особи за динамічним підписом і оцінювати його помилки.

10.1. Задача розпізнавання за еталоном

Найпростіший підхід до розпізнавання, згаданий у розділі 1, полягає в порівнянні невідомого образу з еталонами (шаблонами) класів. Попри простоту, він широко використовується на практиці: у пошуку відомих об’єктів на зображеннях (англ. *template matching*), виявленні сигналів відомої форми на фоні шуму в радіолокації та зв’язку, суміщенні зображень, контролі якості виробів за еталонним зразком, біометричній верифікації [13, 28, 54]. Він також є основою багатьох складніших методів: класифікатор за найближчим сусідом порівнює образ з усіма навчальними прикладами як з еталонами; згортковий шар нейронної мережі обчислює кореляцію образу з набором навчених «еталонів»-фільтрів.

У задачі розпізнавання за еталоном можна виділити два аспекти:

1. виявлення і локалізація — чи присутній еталон (сегмент) s у спостережуваному сигналі або зображенні і де саме; розв’язується обчисленням міри подібності для всіх можливих положень еталона і пошуком її максимуму;
2. порівняння (верифікація) — наскільки поданий образ відрізняється від еталона; розв’язується обчисленням міри відмінності і порівнянням з порогом.

Ключове питання обох задач — вибір міри подібності, яка була б чутливою до відмінностей форми і водночас нечутливою до несуттєвих змін (масштабу, постійної складової, шуму). Розглянемо основні підходи до її побудови [1].

10.2. Кореляційні методи

10.2.1. Взаємна кореляція

Нехай еталон $s(n)$, $n = 0, \dots, L-1$, шукається в сигналі $x(n)$. Позначимо $\mathrm{x}_k = [x(k), x(k+1), \dots, x(k+L-1)]^T$ — фрагмент сигналу довжиною L, що починається з k-го відліку. Кореляційний детектор обчислює скалярний добуток фрагмента з еталоном для кожного зсуву:

$$R(k) = \mathrm{x}_k^T \mathrm{s} = \sum_{n=0}^{L-1} x\,(k+n)\,s(n). \qquad (10.1)$$

Для зображень аналогічно обчислюється двовимірна взаємна кореляція (4.31). Згідно з (4.32), для великих еталонів її ефективно обчислювати через ДПФ. Максимум $R(k)$ вказує на положення, у якому фрагмент найбільше схожий на еталон.

Проте ця «схожість» розуміється у специфічному сенсі. Розкладемо фрагмент на постійну складову і центровану частину: $\mathrm{x}_k = \bar{x}_k 1 + \tilde{\mathrm{x}}_k$. Тоді

$$R(k) = \tilde{\mathrm{x}}_k^T \mathrm{s} + \bar{x}_k\, 1^T \mathrm{s} = \tilde{\mathrm{x}}_k^T \mathrm{s} + L\,\bar{x}_k\,\bar{s}. \qquad (10.2)$$

Другий доданок, пропорційний локальному середньому значенню сигналу, не несе жодної інформації про подібність форм [1]. Якщо середнє значення еталона ненульове, а сигнал нестаціонарний (локальне середнє змінюється), максимум кореляції зміщується до ділянок з найбільшою яскравістю. Саме це видно на рис. 10.1: на зображенні з градієнтом освітленості максимум взаємної кореляції потрапив на найсвітлішу ділянку неба, а не на місце, з якого було взято еталон.



Рис. 10.1. Пошук еталона на зображенні за допомогою взаємної та нормованої взаємної кореляції

10.2.2. Нормована взаємна кореляція

Усунути вплив локального середнього і локального контрасту дозволяє нормована взаємна кореляція (NCC) — коефіцієнт кореляції Пірсона (1.9) між фрагментом і еталоном:

$$\rho(k) = \frac{\sum_n (x(k+n) - \bar{x}_k)\,(s(n) - \bar{s})}{\sqrt{\sum_n (x(k+n) - \bar{x}_k)^2}\,\sqrt{\sum_n (s(n) - \bar{s})^2}}. \qquad (10.3)$$

Значення $\rho(k) \in [-1{,}1]$ інваріантне до лінійних перетворень яскравості фрагмента $x \to ax + b$ $(a > 0)$, тобто до змін контрасту і яскравості. Для зображень формула аналогічна з підсумовуванням по двовимірному вікну. Чисельник (10.3) обчислюється через ДПФ, а локальні суми і суми квадратів у знаменнику — за допомогою інтегральних зображень (кумулятивних сум), що дозволяє обчислити NCC для всіх положень за $O\big(MN\log(MN)\big)$ операцій незалежно від розміру еталона [54]. На рис. 10.1, в карта NCC має чіткий максимум у правильному положенні.

Недоліком кореляційних методів є чутливість до геометричних перетворень: поворот на кілька градусів або зміна масштабу на 10–20 % суттєво знижують максимум кореляції. Тому їх використовують або в умовах контрольованої геометрії, або в поєднанні з геометричним нормуванням (розділ 7), або перебирають набір повернутих і масштабованих еталонів.

10.3. Лінійний погоджений фільтр

10.3.1. Виведення

Розглянемо задачу виявлення сигналу відомої форми $s(t)$ на фоні адитивного білого шуму $v(t)$ зі спектральною густиною потужності $N_0/2$. Шукатимемо лінійний фільтр з імпульсною характеристикою $h(t)$, який у деякий момент t_0 забезпечує максимальне відношення сигнал/шум на виході [55]:

$$q = \frac{y_s^2(t_0)}{\mathrm{M}\{y_v^2(t)\}}, \tag{10.4}$$

де $y_s(t_0) = \int h(\tau)\, s(t_0 - \tau)\, d\tau$ — значення корисної складової на виході в момент t_0, а $\mathrm{M}\{y_v^2\} = \frac{N_0}{2}\int h^2(\tau)\, d\tau$ — дисперсія шуму на виході. Згідно з нерівністю Коші — Буняковського

$$\left[\int h(\tau)\, s(t_0 - \tau)\, d\tau\right]^2 \le \int h^2(\tau)\, d\tau \cdot \int s^2(t_0 - \tau)\, d\tau, \tag{10.5}$$

причому рівність досягається тоді й лише тоді, коли $h(\tau) = c\, s(t_0 - \tau)$. Звідси

$$h(t) = c\, s(t_0 - t), \qquad q_{max} = \frac{2E}{N_0}, \tag{10.6}$$

де $E = \int s^2(t)\, dt$ — енергія сигналу. Фільтр (10.6) називають погодженим (узгодженим, англ. *matched*): його імпульсна характеристика є дзеркальною копією сигналу, зсунутою на t_0, а частотна характеристика — комплексно спряженим спектром сигналу, $H(j\omega) = c\, S^*(j\omega) e^{-j\omega t_0}$. Максимальне відношення сигнал/шум залежить лише від енергії сигналу і спектральної густини шуму, але не від форми сигналу.

Вихід погодженого фільтра є взаємною кореляційною функцією прийнятого сигналу і еталона:

$$y(t) = \int x(\tau)\, s(\tau - t + t_0)\, d\tau, \tag{10.7}$$

тому погоджений фільтр і кореляційний детектор (10.1) — це дві реалізації однієї й тієї самої операції [1].

10.3.2. Цифровий погоджений фільтр

Для дискретних сигналів погоджений фільтр — це КІХ-фільтр з імпульсною характеристикою $h(n) = s(L - 1 - n)$, тобто відліки еталона в зворотному порядку. Його вихід

$$y(k) = \sum_{n=0}^{L-1} h\,(n)\, x(k - n) = \sum_{n=0}^{L-1} s\,(n)\, x(k - L + 1 + n) \tag{10.8}$$

збігається з (10.1) з точністю до зсуву. На рис. 10.2 наведено приклад: два радіоімпульси, практично непомітні на фоні шуму в прийнятому сигналі, на виході погодженого фільтра дають чіткі піки у відповідних положеннях. Виграш у відношенні сигнал/шум дорівнює кількості відліків сигналу (для білого шуму — $10\lg L$ дБ при однакових енергіях).

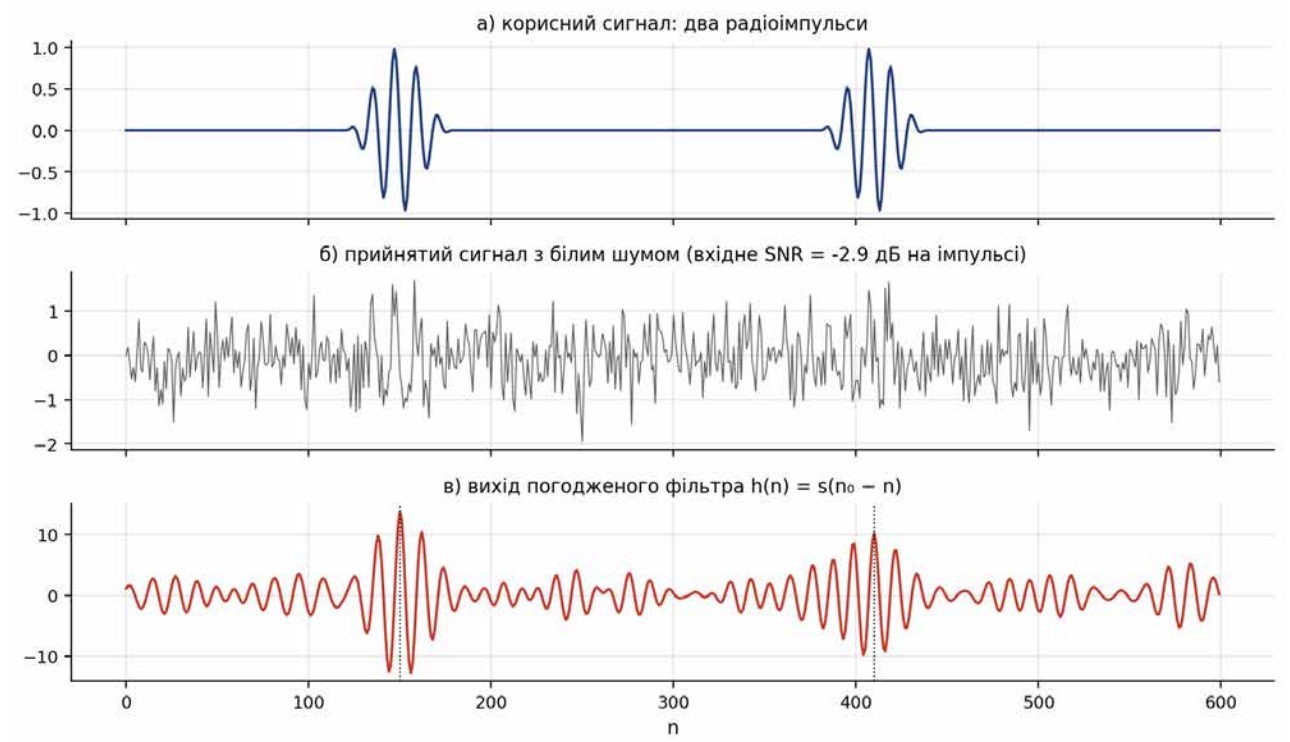


Рис. 10.2. Виявлення сигналу відомої форми погодженим фільтром

Якщо шум не білий, а має кореляційну матрицю R_{VV}, оптимальний фільтр має вигляд $\mathrm{h} = \mathrm{R}_{VV}^{-1}\mathrm{s}$: спочатку шум «вибілюється», а потім застосовується погоджений фільтр. Цей результат безпосередньо пов'язаний з відстанню Махаланобіса (1.7).

10.4. Недоліки класичного погодженого фільтра та альтернативні детектори

10.4.1. Обмеження оптимальності

Погоджений фільтр оптимальний за дуже жорстких припущень: стаціонарність сигналу і шуму, білий гаусівський шум з нульовим середнім, відома форма і положення сигналу, велика кількість відліків [1]. Часто вважають, що класичний погоджений фільтр завжди надійно визначає, чи присутній заданий фрагмент у сигналі. Насправді це не так навіть для сигналів з нульовим шумом, якщо сигнал нестаціонарний: як показує розклад (10.2), рівень вихідного сигналу визначається не лише подібністю форм, а й локальним середнім. У задачах аналізу структури зображень умови оптимальності погодженого фільтра, як правило, не виконуються.

10.4.2. Детектори подібності

У [1] розглянуто кілька детекторів, які усувають ці недоліки різними способами.

Коваріаційний детектор замінює еталон його центрованою версією $\tilde{\mathrm{s}} = \mathrm{s} - \bar{s}1$, чим знищує другий доданок у (10.2):

$$R_{cov}(k) = \mathrm{x}_k^T\tilde{\mathrm{s}} = \tilde{\mathrm{x}}_k^T\tilde{\mathrm{s}}. \tag{10.9}$$

Евклідовий детектор базується на мінімізації норми різниці векторів фрагмента і еталона:

$$\| \mathrm{x}_k - \mathrm{s} \|^2 = \| \mathrm{x}_k \|^2 - 2\,\mathrm{x}_k^T\mathrm{s} + \| \mathrm{s} \|^2. \tag{10.10}$$

Після відкидання сталого члена $\| \mathrm{s} \|^2$ детектор зводиться до класичного погодженого фільтра, від якого віднімається залежна від часу «енергія» вхідного сигналу $\| \mathrm{x}_k \|^2/2$. Основний його недолік — чутливість до абсолютних значень сигналу.

Косинусний детектор використовує кут між векторами, який не залежить від їх довжини:

$$\cos\theta_k = \frac{\mathrm{x}_k^T\mathrm{s}}{\| \mathrm{x}_k \| \| \mathrm{s} \|}. \tag{10.11}$$

Нормований косинусний детектор обчислює косинус кута між *центрованими* векторами, тобто коефіцієнт кореляції (10.3).

Ефективність детекторів суттєво залежить від моделі сигналу і завад. У дослідженнях, наведених у [1], для одно- і двовимірних сигналів з адитивними компонентами у вигляді білого гаусівського шуму, пропущеного через фільтр нижніх частот, гармонічного сигналу і лінійного тренду отримано середні похибки детектування, наведені в табл. 10.1. За цими даними найкращим виявився косинусний детектор, найгіршим — класичний погоджений фільтр; якщо рішення приймається за положенням глобального максимуму відгуку в досить широкому вікні, найнадійніше працюють кутові (косинусні) детектори, а також евклідовий [1]. Водночас при використанні простого порогового критерію без пам'яті косинусний детектор менш зручний: його максимум, хоча й більш надійний, значно менш виражений.

Таблиця 10.1

Середні похибки детектування сегмента сигналу різними детекторами (за даними [1])

Тип детектора	Середня похибка
Класичний погоджений фільтр	81 %
Коваріаційний фільтр	52 %
Евклідовий фільтр	10 %
Косинусний детектор	0,5 %
Нормований косинусний детектор	15 %

Корисно самостійно змоделювати подібну ситуацію і переконатися, що результат залежить від властивостей і еталона, і завад. На рис. 10.3 наведено вихідні сигнали п'яти детекторів для сигналу, що містить лінійний тренд, повільні коливання і вставлений еталон — коливання з гаусівською огинаючою,

середнє значення якого близьке до нуля. У цьому прикладі детектори, що використовують центрування (коваріаційний і нормований косинусний), а також кореляційний (бо середнє еталона майже нульове, і доданок з локальним середнім у (10.2) малий), правильно локалізують еталон; евклідовий і нецентрований косинусний детектори, чутливі до локального рівня сигналу, помиляються. Отже, *універсально найкращого* детектора немає, і вибір міри подібності має враховувати характер нестаціонарності конкретних образів.

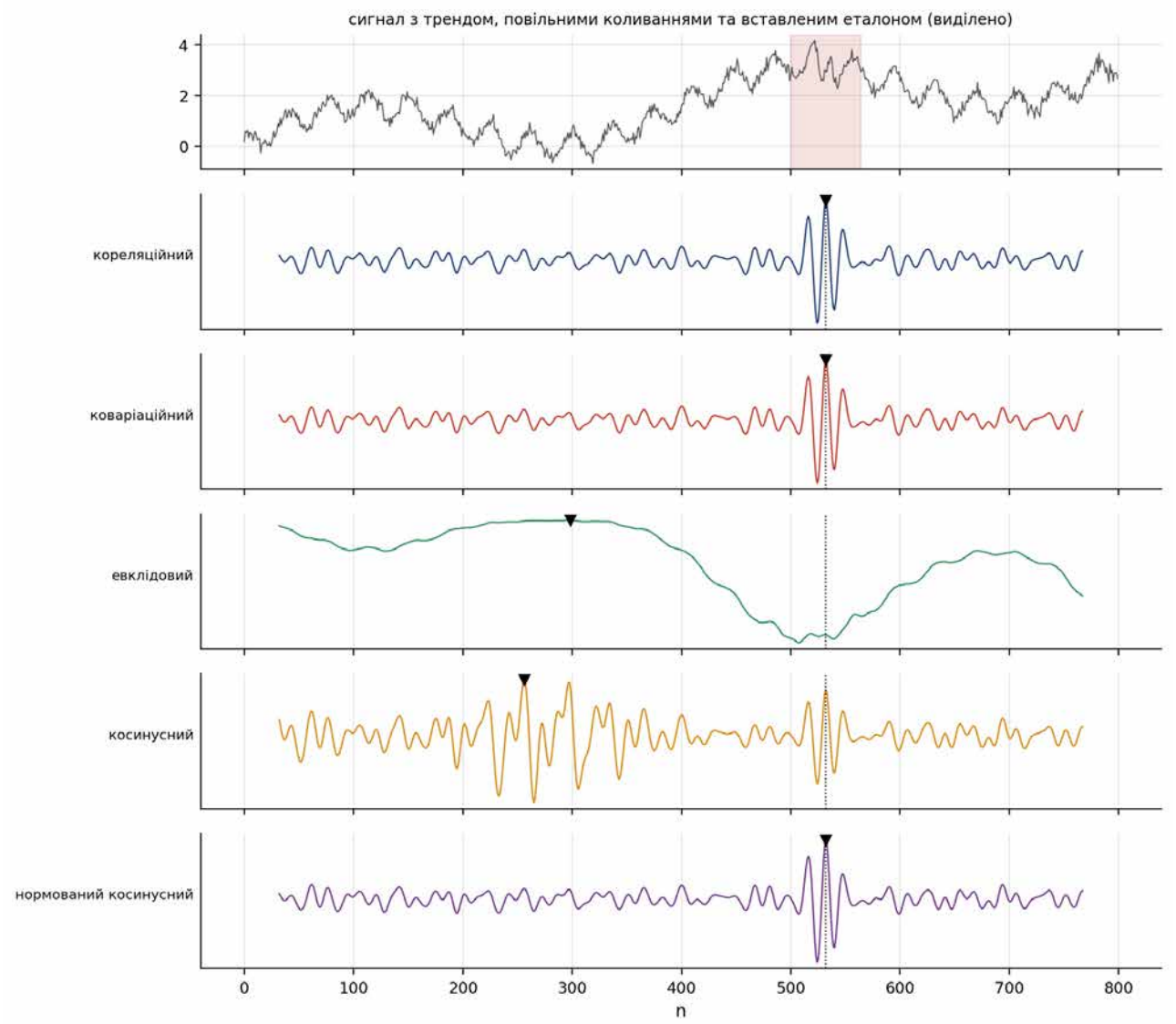


Рис. 10.3. Відгуки різних детекторів подібності для нестаціонарного сигналу

10.5. Нормалізація ортогональних перетворень

10.5.1. Ідея

Ортогональні перетворення часто використовують для виявлення змін у графоелементах сигналів, які малопомітні в натуральних координатах [1]. Наприклад, спотворення форми гармонічного струму «на око» можуть бути непомітними, тоді як у спектрі Фур'є наявність складових кратних частот добре видно і може бути чисельно оцінена коефіцієнтом гармонік

$$k_{\text{г}} = \frac{\sqrt{\sum_{k\geq 2} A_k^2}}{A_1}, \tag{10.12}$$

де A_k — амплітуди гармонік. Зручність такої оцінки зумовлена тим, що форма ідеального сигналу (синусоїда) збігається з формою однієї з базисних функцій перетворення: спектр ідеального сигналу містить одну складову, і будь-яке спотворення одразу проявляється появою інших складових.

На жаль, еталони реальних радіотехнічних, біомедичних та інших сигналів зазвичай не збігаються за формою з базисними функціями жодного зі стандартних перетворень. Оскільки стандартні перетворення універсальні, характерні графоелементи конкретних сигналів розподіляються між багатьма трансформантами і в спектрі втрачають виразність. Відмінності між сигналами в області трансформант виявляються тим чіткіше, чим менше домінантних трансформант містить спектр еталона (чим вужчий його ефективний спектр) [1]. Звідси ідея: побудувати перетворення, одна з базисних функцій якого збігається (у певному сенсі) з еталоном досліджуваних сигналів. Така процедура називається нормалізацією перетворення.

10.5.2. Нормалізація за кроком

Нехай еталонний сигнал нормовано так, що його максимальне і мінімальне значення дорівнюють $+1$ і -1, тобто його динамічний діапазон збігається з діапазоном базисної функції обраного перетворення. Розглянемо, наприклад, трикутний еталон і першу трансформанту дискретного косинусного перетворення формату $N = 8$ [1]. Значення відліків косинусної трансформанти можна спроєктувати на трикутний еталон і знайти моменти часу, в яких еталон набуває тих самих значень. Ці моменти утворюють нееквідистантну сітку відліків. Якщо тепер дискретизувати досліджуваний сигнал саме в цих точках, то для сигналу, що відрізняється від еталона лише масштабом і постійною складовою, відмінними від нуля будуть лише нульова і перша трансформанти нормалізованого перетворення; усі інші дорівнюватимуть нулю. Незначні відмінності форми породжують малі ненульові вищі трансформанти, суттєві — великі. Нормалізація полягає в переставлянні стовпців матричного оператора відповідно до нової нумерації відліків і в збереженні таблиці нееквідистантних кроків.

Єдиною незручністю цього методу є нееквідистантний крок дискретизації, що ускладнює використання стандартних пристроїв; крім того, для перетворень з двома рівнями значень базисних функцій (Уолша) нормалізацію за кроком виконати неможливо.

10.5.3. Нормалізація за рівнем

Для функцій Уолша і загалом для довільних перетворень застосовують нормалізацію за рівнем [1]. Нехай обрано базисну трансформанту $\psi(n)$ (наприклад, одну з функцій Уолша — Качмажа) і тестовий сигнал $s(n)$ того самого формату. Обчислимо коригуючі коефіцієнти

$$\kappa(n) = \frac{s(n)}{\psi(n)}, \qquad n = 0, \dots, N-1. \tag{10.13}$$

Поділивши відліки тестового сигналу на коригуючі коефіцієнти, отримаємо точно базисну трансформанту. Алгоритм нормалізації за рівнем:

1. для тестового сигналу (відповідно нормованого) обрати ортогональне перетворення і базисну трансформанту;
2. для кожного з N відліків тестового сигналу і трансформанти знайти коригуючі коефіцієнти (10.13);
3. поділити відліки досліджуваного сигналу на коригуючі коефіцієнти (якщо сигнал відрізняється від тестового лише амплітудою, отримаємо трансформанту тієї самої амплітуди);
4. виконати глобальне перетворення скоригованого сигналу;
5. якщо спектр містить лише амплітуду базисної трансформанти, досліджуваний сигнал з точністю до амплітуди тотожний тестовому; якщо спектр багатий і містить багато трансформант, порівнянних з базисною, — досліджуваний сигнал не схожий на тестовий.

Якщо досліджуваний сигнал відрізняється від тестового масштабним множником і постійною складовою (для зображень — зміною контрасту і яскравості), обидва сигнали перед нормалізацією слід центрувати — відняти їх середні значення [1]. Стаціонарний адитивний шум з нульовим середнім на масштабування коефіцієнтів нормалізації не впливає.

10.5.4. Коефіцієнт трансформант і гострота

Ступінь відмінності форми досліджуваного сигналу від еталонного кількісно оцінюють коефіцієнтом трансформант — узагальненням коефіцієнта гармонік (10.12) [1]:

$$k_{\text{тр}} = \frac{\sqrt{\sum_{k \neq k_0} X_k^2 \; E_k}}{|X_{k_0}| \sqrt{E_{k_0}}}, \tag{10.14}$$

де X_k — амплітуда k-ї трансформанти, E_k — енергія k-ї трансформанти одиничної амплітуди, k_0 — номер базисної трансформанти, за якою проведено нормалізацію. Формула (10.14) зручна для порівняння спотворень при аналізі тих самих сигналів різними перетвореннями. Якщо всі сигнали порівнюються за одним нормалізованим перетворенням з нормованими базисними функціями, використовують спрощену форму

$$k_{\text{тр}} = \frac{\sqrt{\sum_{k \neq k_0} X_k^2}}{|X_{k_0}|}. \tag{10.15}$$

Для пошуку сигналів, подібних до еталона, зручно використовувати обернену величину — гостроту $1/k_{\text{тр}}$: перевищення гостротою заданого порогу свідчить про наявність сигналу, близького до тестового [1].

10.6. Нормальне ортогональне перетворення

10.6.1. Визначення

Нормалізація за кроком і за рівнем пристосовує *існуюче* перетворення до еталона. Більш радикальний підхід полягає в побудові *нового* дискретного ортогонального перетворення, одна з трансформант якого збігається з наперед заданим дискретним сигналом довільної форми. Тоді спектр цього сигналу в такому перетворенні міститиме лише одну ненульову складову. Таке ортогональне перетворення називають нормальним [1].

Нехай еталонний сигнал заданий вектором відліків $\mathrm{s} = [s_0, s_1, \dots, s_{N-1}]^T$, формат перетворення збігається з форматом сигналу і дорівнює $N = 2^p$. Потрібно знайти ортогональну матрицю W_Σ, перший рядок якої пропорційний s^T. Тоді

$$\mathrm{W}_\Sigma\,\mathrm{s} = [\,\|\,\mathrm{s}\,\|\,,0,0,\dots,0\,]^T. \tag{10.16}$$

10.6.2. Побудова матричного оператора

У [1] запропоновано будувати матричний оператор нормального перетворення як добуток $p = \log_2 N$ послідовних ортогональних перетворень, при кожному з яких кількість ненульових трансформант еталона зменшується вдвічі. Розглянемо цю процедуру в зручній для обчислень формі.

Перший етап. Розіб'ємо відліки еталона на пари $(s_0, s_1), (s_2, s_3), \dots$ Для кожної пари (a, b) утворимо ортогональний блок другого порядку

$$\mathrm{B}(a,b) = \frac{1}{\sqrt{a^2+b^2}}\begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \qquad \mathrm{B}(a,b)\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+b^2} \\ 0 \end{bmatrix}. \tag{10.17}$$

Блочно-діагональна матриця $\mathrm{S}_1 = \mathrm{diag}\{\mathrm{B}(s_0, s_1), \mathrm{B}(s_2, s_3), \dots\}$ ортогональна і перетворює еталон на вектор, у якому ненульовими є лише $N/2$ компонент (на парних позиціях). Для $N = 8$ ця матриця має вигляд

$$\mathrm{S}_1 = \begin{bmatrix} \mathrm{B}_{01} & & & \\ & \mathrm{B}_{23} & & \\ & & \mathrm{B}_{45} & \\ & & & \mathrm{B}_{67} \end{bmatrix}. \tag{10.18}$$

Наступні етапи. До ненульових компонент результату знову застосовуються повороти пар, і їх кількість зменшується до $N/4$, потім до $N/8$, і так далі, доки не залишиться одна ненульова компонента. Добуток матриць усіх етапів

$$\mathrm{W}_\Sigma = \mathrm{S}_p\,\mathrm{S}_{p-1}\cdots\mathrm{S}_1 \tag{10.19}$$

є ортогональною матрицею, яка переводить еталон у вектор з єдиною ненульовою першою компонентою. Оскільки $\mathrm{W}_\Sigma^{-1} = \mathrm{W}_\Sigma^T$, з (10.16) випливає, що перший *рядок* W_Σ дорівнює $\mathrm{s}^T/\|\,\mathrm{s}\,\|$, тобто перша трансформанта нормального перетворення з точністю до множника збігається з еталоном. Інші рядки — ортогональні до нього функції, які утворюють базис ортогонального доповнення (рис. 10.4).

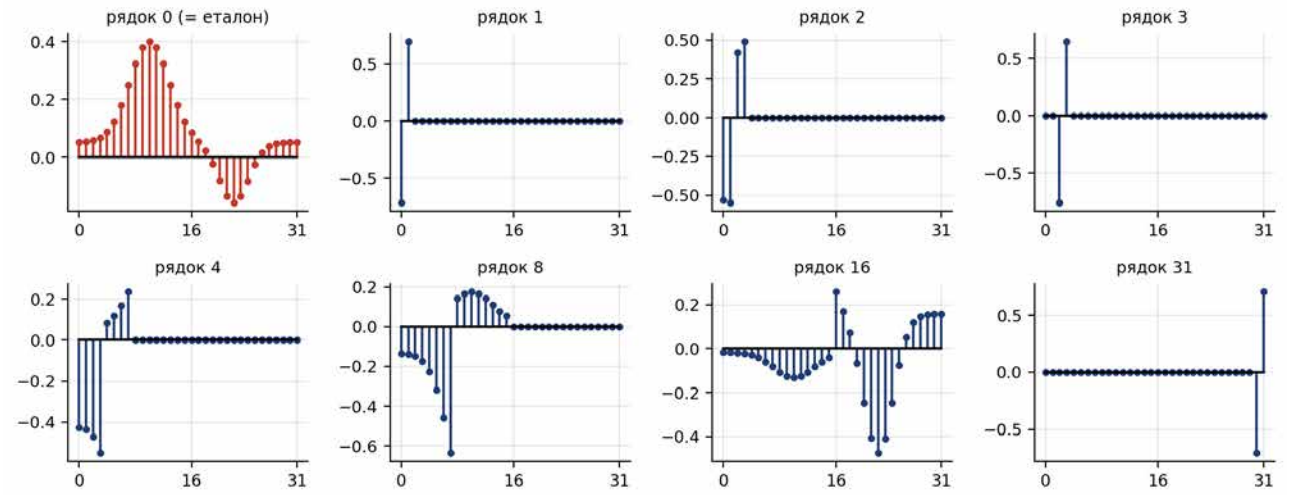


Рис. 10.4. Рядки (трансформанти) нормального перетворення, побудованого для еталона (N = 32)

Структура алгоритму нагадує швидкі алгоритми ортогональних перетворень (ШПФ, ШПУА): $\log_2 N$ етапів по $N/2$ «метеликів». Тому множення вектора на W_{Σ} можна виконувати за $O(N)$ операцій, зберігаючи лише коефіцієнти блоків. У [1] описано також алгоритм безпосереднього формування матричного оператора і нормальні перетворення двох видів, а в обчислювальній математиці подібні конструкції відомі як послідовності поворотів Гівенса; альтернативним способом побудови ортогональної матриці з заданим першим рядком є відбиття Хаусхолдера [56].

10.6.3. Властивості нормального спектра

Якщо досліджуваний сигнал x (центрований разом з еталоном) відрізняється від еталона лише масштабом, його нормальний спектр містить тільки першу трансформанту і $k_{\text{тр}} = 0$. Незначні відмінності форми і шум дають малі вищі трансформанти; суттєві відмінності — великі. На рис. 10.5 наведено нормальні спектри кількох сигналів у перетворенні, побудованому для еталона: сам еталон і його масштабована копія мають $k_{\text{тр}} = 0$; зашумлений еталон — мале значення; зсунутий еталон — помітно більше; сигнал іншої форми — велике значення.

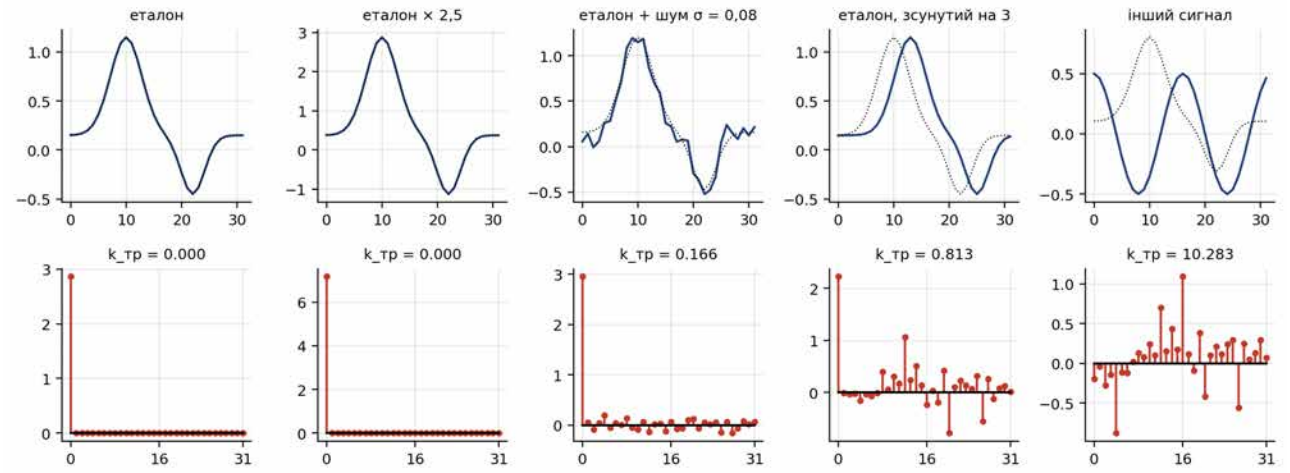


Рис. 10.5. Сигнали (вгорі) та їх нормальні спектри з коефіцієнтами трансформант (внизу)

Встановимо зв'язок коефіцієнта трансформант із мірами подібності розділу 1. Оскільки перетворення ортогональне, $\sum_k X_k^2 = \| \mathrm{x} \|^2$, а перша трансформанта $X_0 = \mathrm{s}^T\mathrm{x}/\| \mathrm{s} \|$. Отже,

$$k_{\text{тр}} = \frac{\sqrt{\| \mathrm{x} \|^2 - X_0^2}}{|X_0|} = \frac{\sqrt{1-\cos^2\theta}}{|\cos\theta|} = |\tan\theta|, \qquad \cos\theta = \frac{\mathrm{s}^T\mathrm{x}}{\| \mathrm{s} \| \| \mathrm{x} \|}. \tag{10.20}$$

Тобто коефіцієнт трансформант нормального перетворення є тангенсом кута між вектором образу і вектором еталона, а гострота — котангенсом. Для центрованих сигналів $\cos\theta$ — це коефіцієнт кореляції (10.3). Із (10.20) випливає кілька важливих висновків.

1. Рішення за порогом на $k_{\text{тр}}$ еквівалентне рішенню за порогом на модулі косинусної подібності (або коефіцієнта кореляції для центрованих сигналів). Нормальне перетворення дає *геометрично наочну* і *чисельну* інтерпретацію косинусної міри: енергія образу розкладається на «паралельну» до еталона частину (X_0^2) і «перпендикулярну» ($\sum_{k\geq 1} X_k^2$).
2. Коефіцієнт трансформант нелінійно залежить від сигналу, тому нормальний фільтр, на відміну від класичного погодженого, є нелінійним [1].
3. Нормальний спектр несе більше інформації, ніж одне число: розподіл енергії відмінностей між трансформантами з номерами $k \geq 1$ показує, *які саме* компоненти образу відрізняються від еталона. Ця інформація використовується в алгоритмі класифікації з уточненням (підрозділ 10.9).

10.6.4. Нормальне перетворення для двовимірних образів

Нормальне перетворення природно поширюється на двовимірні образи. Найпростіший спосіб — розгорнути еталонне зображення $MimesN$ (з $MN = 2^p$ пікселів) у вектор за рядками і побудувати нормальне перетворення для цього вектора за алгоритмом (10.17)–(10.19). Перша трансформанта такого перетворення — розгорнутий еталон, а коефіцієнт трансформант досліджуваного фрагмента дорівнює тангенсу кута між фрагментом і еталоном у просторі розмірності MN; для центрованих фрагментів це еквівалентно

нормованій взаємній кореляції (10.3) у нульовому зсуві. Отже, пошук еталона на зображенні з обчисленням $k_{\text{тр}}$ для всіх положень ковзного вікна дає ту саму карту подібності, що й NCC (з точністю до монотонного перетворення $ an(rccos r)$), але додатково дозволяє аналізувати нормальний спектр відхилень у кожному положенні.

Другий спосіб використовує сепарабельність. Якщо еталон подається зовнішнім добутком двох векторів $\mathrm{s} = \mathrm{s}_r \mathrm{s}_c^T$ (або добре наближається таким добутком, що можна перевірити сингулярним розкладом матриці еталона), то двовимірне нормальне перетворення будується як кронекерів добуток одновимірних нормальних перетворень для s_r і s_c, а спектр обчислюється сепарабельно: $\mathrm{X} = \mathrm{W}_r \mathrm{F} \mathrm{W}_c^T$, де F — матриця фрагмента. Для еталонів, що не є сепарабельними, можна використовувати суму кількох сепарабельних складових, отриманих із сингулярного розкладу.

Обчислювальні витрати при розгорнутому поданні становлять $O(MN)$ операцій на одне положення вікна (завдяки поетапній структурі (10.19)), тобто стільки ж, скільки і для прямого обчислення кореляції; для пошуку в усіх положеннях вигідніше обчислювати NCC через ДПФ (розділ 4), а нормальний спектр — лише в кількох положеннях-кандидатах, відібраних за максимумами NCC.

10.7. Аутентифікація особи за динамічним підписом

10.7.1. Постановка задачі

Розглянемо застосування нормального перетворення для аутентифікації особи за динамічно введеним підписом [1]. Такі системи використовуються для контролю доступу до інформації: особа ставить підпис на планшеті, і реєструється не лише кінцеве зображення підпису, а *процес* його введення — проєкції миттєвого положення пера на координатні осі площини підпису як функції часу $x(t)$ і $y(t)$. Динамічні характеристики підпису (швидкість, послідовність штрихів, ритм) значно важче підробити, ніж статичне зображення.

Як і будь-яка система класифікації, система аутентифікації має пройти етап навчання на навчальній множині — статистично репрезентативній вибірці підписів особи, яку надалі потрібно аутентифікувати [1].

10.7.2. Навчання

1. Усі підписи навчальної множини нормують у часі (приводять до однакової кількості відліків N) та за амплітудою (з метою усунення впливу масштабу); компоненти центрують.
2. Обчислюють оцінки математичного сподівання кожної з траєкторій:

$$\bar{\mathrm{x}} = \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{x}_q \,, \qquad \bar{\mathrm{y}} = \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{y}_q, \tag{10.21}$$

де Q — кількість підписів у навчальній множині. Як показує досвід, для надійного результату достатньо усереднити 20–30 підписів [1]. 3. На основі

еталонних сигналів $\bar{\mathrm{x}}$ і $\bar{\mathrm{y}}$ формують дві матриці нормального перетворення $\mathrm{W}_{\Sigma x}$ і $\mathrm{W}_{\Sigma y}$. 4. Для всіх реалізацій навчальної множини обчислюють нормальні спектри $\mathrm{X}_q = \mathrm{W}_{\Sigma x}\mathrm{x}_q$, $\mathrm{Y}_q = \mathrm{W}_{\Sigma y}\mathrm{y}_q$ і коефіцієнти трансформант $k_{\text{тр},x}(q)$, $k_{\text{тр},y}(q)$. 5. Порогові значення коефіцієнтів трансформант для кожної компоненти визначають як максимальні з отриманих на навчальній вибірці:

$$\theta_x = \max_q k_{\text{тр},x}(q), \qquad \theta_y = \max_q k_{\text{тр},y}(q). \tag{10.22}$$

У базі даних системи для кожного зареєстрованого користувача зберігаються: його ідентифікатор (логін), вибірки нормованих реалізацій компонент підпису, дві матриці нормального перетворення (або коефіцієнти їх блоків) і два порогові значення.

10.7.3. Аутентифікація

При спробі отримати доступ користувач вводить логін і динамічний підпис. Кожна компонента підпису нормується так само, як під час навчання; обчислюються коефіцієнти трансформант у нормальних перетвореннях, що відповідають заявленому логіну. Якщо обидва коефіцієнти не перевищують порогових значень, приймається рішення про відповідність підпису особі. У цьому разі нормований підпис доповнює вибірку в базі даних, коригуються математичні сподівання компонент, а з ними — матриці нормального перетворення і пороги [1]. Така адаптація дозволяє враховувати повільні зміни психофізичних характеристик особи з часом.

10.7.4. Експериментальні дослідження і обмеження методу

Дослідження ймовірнісних характеристик алгоритму можливе лише за наявності великої кількості реальних підписів. У дослідженні [1] брали участь 20 осіб; кожна з них протягом кількох днів ввела по 100 підписів (10–15 за день). За результатами обчислень коефіцієнтів трансформант визначено порогові значення і ймовірнісні характеристики алгоритму: ймовірність правильного розпізнавання, помилок першого роду (відмова особі, яка має право доступу) і другого роду (надання доступу особі, яка такого права не має). Зроблено висновок, що класифікатор на основі нормального перетворення математичних сподівань проєкцій дає надійні результати, а сам алгоритм простий і не потребує значних обчислювальних ресурсів.

Для самостійного відтворення принципу роботи алгоритму на рис. 10.6 наведено результати моделювання на *синтетичних* даних: траєкторії «підписів» трьох умовних осіб задано гладкими параметричними кривими з випадковими нелінійними деформаціями часової осі, амплітуди і шумом. Розподіли нормованих коефіцієнтів трансформант для «своїх» і «чужих» підписів добре розділені (зверніть увагу на логарифмічну шкалу); при порозі, що дорівнює максимуму на навчальній вибірці, частка помилок першого роду становить близько 5 %, а другого роду — практично нуль. Зміщуючи поріг, можна обрати інше співвідношення помилок відповідно до вимог до системи.

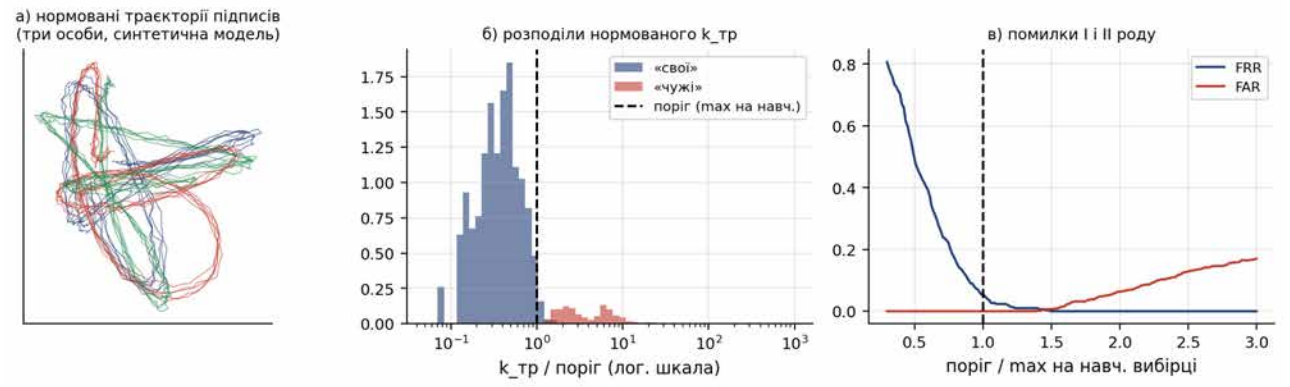


Рис. 10.6. Моделювання аутентифікації за динамічним підписом з використанням нормального перетворення

Для інших класів сигналів класифікатор на основі нормального перетворення може виявитися недостатньо чутливим [1]. Наприклад, для фонограм українських голосних «а», «о», «у», «е» відмінності між графоелементами настільки незначні, що нормальні перетворення їх математичних сподівань мало відрізняються між собою, і класифікатор, побудований так само, як для підписів, ці звуки не розрізняє. Причина зрозуміла з (10.20): коли еталони різних класів утворюють малі кути між собою, а розкид усередині класу великий, однієї кутової міри недостатньо. Для таких сигналів потрібні тонші, хоча й трудомісткіші методи, насамперед методи, що враховують *статистику відхилень* усередині класу, — класифікація за критерієм максимальної правдоподібності.

10.8. Порівняння лінійної погодженої та нормальної фільтрації

Узагальнимо порівняння класичного погодженого фільтра і фільтра на основі нормального перетворення (нормального фільтра) [1].

1. Класичний лінійний погоджений фільтр (так само, як і нормальний) обчислює скалярний добуток спостережуваного образу з еталоном; максимум його виходу пропорційний енергії образу і косинусу кута між ним і еталоном.
2. На відміну від нормального, класичний погоджений фільтр не дає чисельної оцінки ступеня подібності, незалежної від енергії (контрасту) образу; нормальний фільтр дозволяє чисельно оцінювати ступінь відмінності через $k_{\text{тр}}$.
3. Нормальний фільтр нелінійний.
4. Між нормальним і класичним погодженим фільтрами існує жорсткий зв’язок: перша трансформанта нормального спектра є виходом погодженого фільтра, нормованим на енергію еталона, а сума енергій решти трансформант дорівнює енергії образу мінус квадрат цього виходу (10.20). У цьому сенсі фільтри інформаційно споріднені: нормальний фільтр додає до виходу погодженого фільтра нормування на енергію образу і розклад залишку.

5. Обидва фільтри будуються за одним еталоном і мають однакову обчислювальну складність порядку N операцій на відлік (для нормального фільтра — завдяки структурі (10.19)).

10.9. Класифікація з уточненням

Класифікатор на основі нормального перетворення простий і швидкий, але, як показано вище, не завжди достатньо надійний. Більш надійним (і трудомістким) є статистичний класифікатор за критерієм максимальної правдоподібності, який для нормального розподілу відліків образу в класі ω_k обчислює дискримінантне число — квадрат відстані Махаланобіса до математичного сподівання класу:

$$d_k(\mathrm{x}) = (\mathrm{x} - \bar{\mathrm{x}}_k)^T \mathrm{R}_k^{-1} (\mathrm{x} - \bar{\mathrm{x}}_k), \qquad (10.23)$$

де R_k — кореляційна матриця відхилень реалізацій класу від математичного сподівання [1]. Основна складність його реалізації полягає в поганій обумовленості кореляційної матриці: при малому кроці дискретизації сусідні відліки майже повністю корельовані, і матриця має багато нульових або дуже малих власних чисел. У [1] запропоновано модифікований метод, що використовує розклад $\mathrm{R}_k = \Pi\Lambda\Pi^T$ на власні вектори і власні числа і не потребує обернення матриці: для досліджуваного образу утворюють матрицю «миттєвої» кореляції відхилення, перетворюють її власними векторами класу і оцінюють відмінність отриманої матриці від діагональної матриці власних чисел за деякою нормою.

Показано також, що кореляційна матриця може бути обчислена в області нормального спектра відхилень, і перехід між двома класифікаторами не потребує додаткових обчислень. Звідси алгоритм класифікації з уточненням [1]:

1. для досліджуваного образу виконати дискретизацію і нормування;
2. обчислити нормальний спектр і коефіцієнт трансформант;
3. якщо різниця між отриманим $k_{\text{тр}}$ і пороговим значенням велика, прийняти рішення: класифікація з високою ймовірністю правильна;
4. інакше (у «сумнівній» зоні) обчислити дискримінантне число статистичного класифікатора і прийняти рішення за ним.

Детально статистичні класифікатори, розклад кореляційних матриць і методи боротьби з їх поганою обумовленістю розглядатимуться в другій частині посібника.

10.10. Програмна реалізація

Нижче наведено функції побудови нормального перетворення за алгоритмом (10.17)–(10.19), обчислення коефіцієнта трансформант, а також пошуку еталона на зображенні методом нормованої взаємної кореляції.

```
import numpy as np
from skimage import data, feature, img_as_float

def normal_transform(s):
    """Ортогональна матриця, перший рядок якої пропорційний еталону s.
```

```
    len(s) = 2**p; log2(N) етапів поворотів пар компонент (10.17)-(10.19)."""
    s = np.asarray(s, float); N = s.size
    W = np.eye(N); v = s.copy(); idx = list(range(N))
    while len(idx) > 1:
        S = np.eye(N); keep = []
        for a, b in zip(idx[0::2], idx[1::2]):
            r = np.hypot(v[a], v[b])
            c, sn = (v[a] / r, v[b] / r) if r > 0 else (1.0, 0.0)
            S[a, a], S[a, b], S[b, a], S[b, b] = c, sn, -sn, c
            keep.append(a)
        v = S @ v; W = S @ W; idx = keep
    if v[0] < 0: W[0] *= -1
    return W

def k_tr(W, x):                 # (10.15); дорівнює |tg(theta)| (10.20)
    X = W @ x
    return np.sqrt(np.sum(X[1:]**2)) / abs(X[0])

t = np.arange(32)
s = np.exp(-((t - 10) / 4)**2) - 0.6 * np.exp(-((t - 22) / 3)**2)
s -= s.mean()                   # центрування
W = normal_transform(s)
print(np.allclose(W @ W.T, np.eye(32)), np.round(W @ s, 6)[:4])
x = 2.5 * s + 0.05 * np.random.default_rng(0).normal(size=32)
x -= x.mean()
cos = x @ s / np.linalg.norm(x) / np.linalg.norm(s)
print(k_tr(W, x), np.sqrt(1 - cos**2) / abs(cos))      # однакові

# пошук еталона на зображенні: нормована взаємна кореляція (10.3)
img = img_as_float(data.camera())
tpl = img[95:135, 230:270]
ncc = feature.match_template(img, tpl)
print(np.unravel_index(np.argmax(ncc), ncc.shape))    # (95, 230)
```

10.11. Приклади розв'язання задач

Задача 10.1. Еталон $\mathrm{s} = [1,2,3,2,1]^T$. Обчисліть косинусну подібність і коефіцієнт кореляції для фрагментів $\mathrm{x}_1 = [3,5,7,5,3]^T$ і $\mathrm{x}_2 = [2,3,2,3,2]^T$.

Розв'язання. Для $\mathrm{x}_1 = 2\mathrm{s} + 1$ (подвоєний контраст і зсув рівня): $\mathrm{s}^T\mathrm{x}_1 = 3 + 10 + 21 + 10 + 3 = 47$, $\| \mathrm{s} \| = \sqrt{19}$, $\| \mathrm{x}_1 \| = \sqrt{117}$, $\cos\theta_1 = 47/\sqrt{19 \cdot 117} \approx 0{,}997$. Центровані вектори $\tilde{\mathrm{s}} = [-0{,}8;\ 0{,}2;\ 1{,}2;\ 0{,}2;\ -0{,}8]$, $\tilde{\mathrm{x}}_1 = 2\tilde{\mathrm{s}}$, тому коефіцієнт кореляції $r_1 = 1$ — фрагмент точно збігається з еталоном з точністю до лінійного перетворення яскравості.

Для x_2: $\mathrm{s}^T\mathrm{x}_2 = 2 + 6 + 6 + 6 + 2 = 22$, $\| \mathrm{x}_2 \| = \sqrt{30}$, $\cos\theta_2 = 22/\sqrt{19 \cdot 30} \approx 0{,}921$. Незважаючи на зовсім іншу форму (коливання замість

«горба»), косинусна подібність досить висока через велику постійну складову обох векторів. Коефіцієнт кореляції: $\tilde{x}_2 = [-0{,}4;\ 0{,}6;\ -0{,}4;\ 0{,}6;\ -0{,}4]$, $\tilde{s}^T\tilde{x}_2 = 0{,}32 + 0{,}12 - 0{,}48 + 0{,}12 + 0{,}32 = 0{,}4$, $\| \tilde{s} \| = \sqrt{2{,}8}$, $\| \tilde{x}_2 \| = \sqrt{1{,}2}$, $r_2 = 0{,}4/\sqrt{3{,}36} \approx 0{,}218$ — центрування чітко виявляє розбіжність форм. Коефіцієнт трансформант (10.20) для нецентрованих векторів: $k_{\text{тр}} = \tan(\arccos 0{,}921) \approx 0{,}42$.

Задача 10.2. Побудуйте нормальне перетворення для еталона $s = [3,4,0,5]^T$.

Розв'язання. Перший етап (10.17): пара $(3,4)$, $r = 5$, блок $\frac{1}{5}\begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$; пара $(0,5)$, $r = 5$, блок $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Після першого етапу вектор $[5,0,5,0]^T$. Другий етап: пара $(5,5)$ у позиціях 0 і 2, $r = 5\sqrt{2}$, блок $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Результат $\left[5\sqrt{2}, 0,0,0\right]^T = [\| s \|, 0,0,0]^T$. Добуток матриць етапів:

$$W_\Sigma \approx \begin{bmatrix} 0{,}424 & 0{,}566 & 0 & 0{,}707 \\ -0{,}8 & 0{,}6 & 0 & 0 \\ -0{,}424 & -0{,}566 & 0 & 0{,}707 \\ 0 & 0 & -1 & 0 \end{bmatrix}. \qquad (10.24)$$

Перший рядок дорівнює $s^T/\| s \| = [3,4,0,5]/\sqrt{50}$; рядки взаємно ортогональні й нормовані.

Задача 10.3. Погоджений фільтр для сигналу з $L = 60$ відліків і однаковою енергією у кожному відліку. У скільки разів (у децибелах) зростає відношення сигнал/шум на виході?

Розв'язання. Вхідне відношення сигнал/шум (на один відлік) $q_{\text{вх}} = s^2/\sigma^2$; вихідне за (10.6) — $E/\sigma^2 = Ls^2/\sigma^2$. Виграш дорівнює $L = 60$, тобто $10\lg 60 \approx 17{,}8$ дБ. Саме тому на рис. 10.2 імпульси, непомітні на вході, чітко видно на виході погодженого фільтра.

Висновки до розділу 10

1. Розпізнавання за еталоном полягає в обчисленні міри подібності образу з еталоном і пошуку її максимуму (виявлення і локалізація) або порівнянні з порогом (верифікація). Вибір міри подібності визначає, до яких змін образу система буде інваріантною.
2. Взаємна кореляція чутлива до локального середнього значення сигналу; нормована взаємна кореляція інваріантна до змін контрасту і яскравості та ефективно обчислюється через ДПФ та інтегральні зображення.
3. Лінійний погоджений фільтр з імпульсною характеристикою, дзеркальною до сигналу, максимізує відношення сигнал/шум на виході при білому шумі; його вихід — взаємна кореляційна функція прийнятого сигналу і еталона.
4. Оптимальність погодженого фільтра обмежена жорсткими припущеннями (стаціонарність, білий шум, нульове середнє); для нестаціонарних образів доцільно використовувати коваріаційний, евклідовий, косинусний та нормований косинусний детектори; їх відносна ефективність залежить від моделі сигналу і завад.

5. Нормалізація перетворень за кроком або за рівнем пристосовує базисну функцію перетворення до еталона; ступінь відмінності оцінюється коефіцієнтом трансформант або гостротою.
6. Нормальне перетворення має першу трансформанту, що збігається з еталоном; його матриця будується як добуток $\log_2 N$ етапів поворотів пар компонент. Коефіцієнт трансформант нормального перетворення дорівнює тангенсу кута між образом і еталоном, що пов'язує його з косинусною мірою і коефіцієнтом кореляції.
7. Алгоритм аутентифікації за динамічним підписом на основі нормального перетворення математичних сподівань траєкторій простий, не потребує значних обчислень і дає надійні результати; для класів з «тонкими» відмінностями потрібні статистичні класифікатори, а поєднання обох підходів дає алгоритм класифікації з уточненням.

Питання для обговорення та самоперевірки

1. У чому полягають задачі виявлення і верифікації при розпізнаванні за еталоном?
2. Чому максимум взаємної кореляції може не відповідати положенню еталона?
3. Запишіть нормовану взаємну кореляцію. До яких перетворень яскравості вона інваріантна?
4. Як ефективно обчислити NCC для всіх положень еталона?
5. Виведіть імпульсну характеристику погодженого фільтра. Чому дорівнює максимальне відношення сигнал/шум?
6. Чому погоджений фільтр і кореляційний детектор еквівалентні?
7. Як будується оптимальний фільтр для небілого шуму?
8. За яких умов погоджений фільтр є оптимальним? Чому ці умови рідко виконуються при аналізі зображень?
9. Запишіть вирази для коваріаційного, евклідового, косинусного і нормованого косинусного детекторів. Які недоліки усуває кожний з них?
10. Що таке нормалізація перетворення? Чим нормалізація за кроком відрізняється від нормалізації за рівнем?
11. Запишіть коефіцієнт трансформант. Як він пов'язаний з коефіцієнтом гармонік?
12. Що таке нормальне ортогональне перетворення? Опишіть алгоритм побудови його матриці.
13. Доведіть, що коефіцієнт трансформант нормального перетворення дорівнює тангенсу кута між образом і еталоном.
14. Чому нормальний фільтр нелінійний?
15. Опишіть етапи навчання і роботи системи аутентифікації за динамічним підписом.
16. Як визначаються порогові значення коефіцієнтів трансформант? Як можна змінити співвідношення помилок першого і другого роду?
17. Чому класифікатор на основі нормального перетворення не розрізняє голосні звуки?

18. У чому полягає складність класифікатора за критерієм максимальної правдоподібності? Як її обходять?
19. Опишіть алгоритм класифікації з уточненням.

Завдання для самостійної роботи

1. Реалізуйте нормовану взаємну кореляцію з використанням ДПФ та інтегральних зображень і порівняйте швидкодію з прямим обчисленням для еталонів $16 \times 16 \ldots 128 \times 128$.
2. Дослідіть стійкість NCC до повороту і зміни масштабу еталона: побудуйте залежність максимуму NCC від кута повороту $(0 \ldots 20°)$ і масштабу $(0{,}7 \ldots 1{,}3)$.
3. Змоделюйте виявлення радіоімпульсу погодженим фільтром для різних відношень сигнал/шум; побудуйте залежність імовірності правильного виявлення від вхідного відношення сигнал/шум.
4. Повторіть експеримент рис. 10.3 для еталонів з ненульовим середнім і для різних видів нестаціонарності (тренд, змінна дисперсія, гармонічна завада). Складіть таблицю похибок детекторів, аналогічну табл. 10.1.
5. Реалізуйте нормалізацію за рівнем для функцій Уолша — Качмажа і перевірте алгоритм на тестовому сигналі з шумом.
6. Реалізуйте побудову нормального перетворення і перевірте співвідношення (10.20). Побудуйте нормальне перетворення через відбиття Хаусхолдера і порівняйте отримані базиси.
7. Зберіть (за допомогою графічного планшета або миші) по 30 динамічних підписів трьох–п’яти осіб. Реалізуйте алгоритм аутентифікації, оцініть FAR, FRR і EER.
8. Реалізуйте класифікатор за дискримінантним числом (10.23) з регуляризацією кореляційної матриці (додаванням малої константи до діагоналі) і порівняйте з класифікатором на основі нормального перетворення для тих самих даних.

РОЗДІЛ 11. ФОРМУВАННЯ, НОРМУВАННЯ ТА ВІДБІР ОЗНАК

Навчальні цілі. Після вивчення розділу ви зможете: обчислювати геометричні ознаки бінарних образів і моментні інваріанти; описувати контури ланцюговими кодами і дескрипторами Фур'є; формувати статистичні та текстурні ознаки (матриці суміжності, локальні бінарні шаблони, вейвлет- і габорівські ознаки); обчислювати гістограми орієнтованих градієнтів; нормувати ознаки; зменшувати розмірність методом головних компонент; оцінювати інформативність ознак критерієм Фішера та обирати стратегію відбору ознак.

11.1. Місце формування ознак у системі розпізнавання

Формування ознак — заключний етап першої частини конвеєра розпізнавання (рис. 1.2), на якому попередньо оброблений і, за потреби, реставрований та сегментований образ перетворюється на вектор ознак (1.1). Від вдалості цього етапу значною мірою залежить, наскільки простим може бути класифікатор і наскільки надійними будуть його рішення. У розділі 1 сформульовано вимоги до ознак: інформативність, стійкість (інваріантність) до несуттєвих перетворень, незалежність, обчислювальна доступність та інтерпретованість.

Ознаки, що використовуються в розпізнаванні образів, можна згрупувати так [18, 48]:

- геометричні — характеризують розмір і форму об'єкта (площа, периметр, компактність, моменти);
- ознаки контуру — описують межу об'єкта (ланцюгові коди, кривина, дескриптори Фур'є);
- яскравісні (фотометричні) — статистики розподілу яскравості чи кольору в межах об'єкта;
- текстурні — характеризують просторову організацію яскравості (матриці суміжності, локальні бінарні шаблони, енергії в частотних і вейвлет-субсмугах, відгуки фільтрів Габора);
- спектральні — коефіцієнти ортогональних перетворень (Фур'є, косинусного, Уолша, вейвлет), у тому числі нормальних (розділ 10);
- локальні (точкові) — дескриптори околів характерних точок (кутів, плям), гістограми орієнтованих градієнтів.

Окремо розглядаються процедури, що працюють з уже сформованими векторами: нормування, зменшення розмірності (перехід до нових, синтезованих ознак) і відбір (вибір підмножини найінформативніших ознак).

11.2. Геометричні ознаки бінарних образів

Нехай після сегментації об'єкт поданий зв'язною областю R на бінарному зображенні (розділ 7). Найпростіші геометричні ознаки [13, 48]:

- площа A — кількість пікселів області;
- периметр P — довжина межі (обчислюється за ланцюговим кодом: горизонтальні і вертикальні кроки мають довжину 1, діагональні — $\sqrt{2}$);
- охоплюючий прямокутник, його розміри і відношення сторін;

- заповненість (англ. *extent*) — відношення площі об’єкта до площі охоплюючого прямокутника;
- опуклість (англ. *solidity*) — відношення площі об’єкта до площі його опуклої оболонки;
- число Ейлера — кількість зв’язних компонент мінус кількість дірок (для літери «В» дорівнює -1, для «О» — 0, для «С» — 1);
- компактність (коефіцієнт форми)

$$K = \frac{4\pi A}{P^2}, \tag{11.1}$$

яка дорівнює одиниці для круга і менша за одиницю для будь-якої іншої фігури; вона інваріантна до зсуву, повороту і масштабу; - ексцентриситет еквівалентного еліпса, що має ті самі центральні моменти другого порядку (див. нижче).

Значення цих ознак для кількох простих фігур (рис. 11.1) наведено в табл. 11.1. Зверніть увагу, що для дискретних зображень компактність круга менша за одиницю через похибку оцінювання периметра на растрі; для малих об’єктів ця похибка може бути суттєвою.

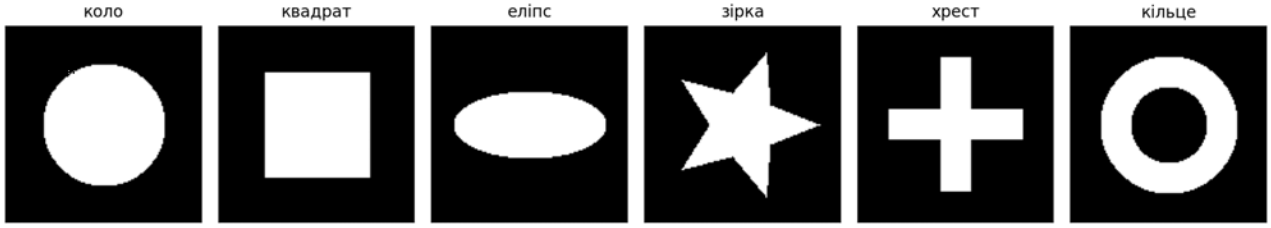


Рис. 11.1. Тестові фігури для обчислення геометричних ознак

Таблиця 11.1

Геометричні ознаки тестових фігур

Фігура	Площа	Периметр	Компактність	Ексцентриситет	Опуклість	Число Ейлера
Коло	5013	260,5	0,929	0,000	0,987	1
Квадрат	4761	272,0	0,809	0,000	1,000	1
Еліпс	3437	242,5	0,734	0,899	0,985	1
Зірка	3239	355,3	0,322	0,096	0,544	1
Хрест	3120	345,7	0,328	0,000	0,582	1
Кільце	4408	460,6	0,261	0,000	0,686	0

Навіть така невелика кількість ознак дозволяє розрізнити всі шість фігур: еліпс вирізняється ексцентриситетом, зірка і хрест — низькою опуклістю (при близькій компактності їх розрізняє, наприклад, ексцентриситет або кількість кутів контуру), кільце — числом Ейлера.

11.3. Моменти та моментні інваріанти

11.3.1. Геометричні моменти

Геометричним моментом порядку $(p + q)$ півтонового або бінарного образу $f(x, y)$ називають

$$m_{pq} = \sum_x \sum_y x^p \; y^q \, f(x, y). \tag{11.2}$$

Момент m_{00} — це «маса» образу (для бінарного — площа); координати центра ваги $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$. Центральні моменти, обчислені відносно центра ваги,

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \; (y - \bar{y})^q \, f(x, y), \tag{11.3}$$

інваріантні до зсуву. Центральні моменти другого порядку утворюють матрицю

$$\mathrm{J} = \frac{1}{\mu_{00}} \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}, \tag{11.4}$$

власні вектори якої визначають напрямки головних осей інерції образу, а власні числа $\lambda_1 \geq \lambda_2$ — квадрати півосей еквівалентного еліпса (з точністю до множника 4). Звідси орієнтація $\theta = \frac{1}{2}\arctan\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}$ та ексцентриситет $e = \sqrt{1 - \lambda_2/\lambda_1}$. Ці величини використовуються для геометричного нормування (розділ 7).

Нормовані центральні моменти

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \qquad \gamma = \frac{p+q}{2} + 1, \tag{11.5}$$

інваріантні також до зміни масштабу.

11.3.2. Інваріанти Ху

М.-К. Ху [57] побудував із нормованих моментів другого і третього порядків сім комбінацій, інваріантних до зсуву, масштабу і повороту. Перші чотири з них:

$$\phi_1 = \eta_{20} + \eta_{02}, \qquad \phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \tag{11.6}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \qquad \phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2. \tag{11.7}$$

Інваріанти ϕ_5, ϕ_6, ϕ_7 мають громіздкіший вигляд і наводяться в [13, 57]. Інваріант ϕ_7 змінює знак при дзеркальному відображенні, що дозволяє розрізняти дзеркально симетричні образи. Оскільки значення інваріантів різняться на багато порядків, на практиці використовують величини $-\mathrm{sign}(\phi_i)\lg|\phi_i|$.

На рис. 11.2 і в табл. 11.2 наведено інваріанти Ху для п’ятикутника та його повернутої, зменшеної, зсунутої та віддзеркаленої копій. Значення практично збігаються (незначні розбіжності для поворотів і масштабування зумовлені дискретизацією); при віддзеркаленні змінює знак лише ϕ_7.

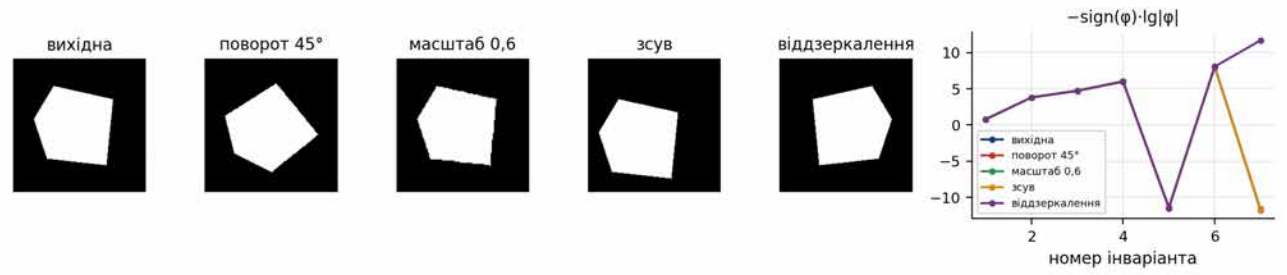


Рис. 11.2. Інваріантність моментів Ху до зсуву, повороту та масштабу

Таблиця 11.2

Значення $-\text{sign}(\phi_i)\lg|\phi_i|$ для варіантів фігури

Варіант	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
Вихідна	0,786	3,801	4,714	5,991	−11,399	8,049	−11,668
Поворот 45°	0,786	3,800	4,743	6,037	−11,477	8,099	−11,765
Масштаб 0,6	0,785	3,795	4,737	5,973	−11,401	8,003	−11,598
Зсув	0,786	3,801	4,714	5,991	−11,399	8,049	−11,668
Віддзеркалення	0,786	3,801	4,714	5,991	−11,399	8,049	11,668

Моменти вищих порядків дуже чутливі до шуму і дрібних деталей на межі. Кращі властивості мають ортогональні моменти (Цернике, Лежандра), що є проєкціями образу на системи ортогональних поліномів: вони мають меншу надлишковість і дозволяють відновлювати образ за моментами — ще одне застосування ідеї ортогонального розкладу.

11.4. Ознаки контуру

Ланцюговий код Фрімена [13, 48] описує контур послідовністю напрямків переходу між сусідніми пікселями межі (4 напрямки для 4-зв'язності, 8 — для 8-зв'язності). Різницевий код (різниці сусідніх напрямків за модулем 8) інваріантний до повороту на кути, кратні 45°, а нормалізація вибору початкової точки (циклічний зсув, що дає мінімальне число) — до вибору початку. Гістограма напрямків ланцюгового коду є простою ознакою форми.

Кривина контуру — швидкість зміни напрямку дотичної — дозволяє виділити кути (точки максимальної кривини) і прямолінійні ділянки. Кількість кутів, їх розподіл і кути між сторонами є ефективними ознаками для розпізнавання багатокутних об'єктів [48].

Дескриптори Фур'є розглянуто в підрозділі 4.9: модулі нормованих коефіцієнтів ДПФ комплексної послідовності точок контуру утворюють вектор ознак, інваріантний до зсуву, масштабу, повороту і вибору початкової точки. Перевагою дескрипторів Фур'є є можливість регулювати деталізацію опису кількістю коефіцієнтів (рис. 4.12): низькочастотні коефіцієнти стійкі до шуму на межі і описують загальну форму.

Для контурів, заданих як функції кута або довжини дуги, можна використовувати і інші ортогональні перетворення — косинусне, Уолша, вейвлет-перетворення. Останнє особливо зручне для багатомасштабного аналізу форми: великі масштаби описують загальну форму, малі — локальні нерівності [24, 48].

11.5. Яскравісні та статистичні ознаки

Якщо об'єкт виділено на півтоновому або кольоровому зображенні, ознаками слугують статистики яскравості (кольору) в межах області: середнє значення, дисперсія, асиметрія, ексцес, ентропія, процентилі, а також гістограма в цілому (розділ 2) [13]. Для кольорових зображень — середні і дисперсії в кожному каналі зручної колірної моделі (HSV, Lab), а також колірні гістограми. Гістограмні ознаки інваріантні до зсуву і повороту об'єкта, але чутливі до умов освітлення; тому перед їх обчисленням застосовують нормування яскравості (розділ 7) або використовують інваріантні до освітлення колірні координати (тон, нормовані хроматичні координати $r = R/(R+G+B)$, $g = G/(R+G+B)$).

Порівняння гістограм виконують спеціальними мірами: перетином гістограм $\sum_i \min(h_i, g_i)$, відстанню χ^2, відстанню Бхаттачарії, відстанню Вассерштейна (переміщення «маси» між рівнями) [17, 28].

11.6. Текстурні ознаки

11.6.1. Матриці суміжності рівнів яскравості

Текстура характеризується не розподілом окремих значень яскравості, а їх *взаємним розташуванням*. Класичним інструментом її опису є матриця суміжності (матриця спільної появи рівнів яскравості, англ. *gray-level co-occurrence matrix*, GLCM), запропонована Р. Хараліком [48]. Для заданого вектора зсуву $\mathrm{d} = (d\cos\varphi, d\sin\varphi)$ елемент $P_{\mathrm{d}}(i,j)$ матриці дорівнює (нормованій) кількості пар пікселів, розташованих на відстані d один від одного, перший з яких має рівень i, а другий — j. Зазвичай яскравість попередньо квантують до 8–32 рівнів, а матриці обчислюють для кількох напрямків $(0°, 45°, 90°, 135°)$ і усереднюють для отримання інваріантності до повороту.

За матрицею суміжності обчислюють ознаки Хараліка, зокрема:

$$\text{контраст} = \sum_{i,j} (i-j)^2 P(i,j), \qquad \text{однорідність} = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}, \tag{11.8}$$

$$\text{енергія} = \sum_{i,j} P^2(i,j), \qquad \text{кореляція} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}, \tag{11.9}$$

а також ентропію, дисперсію, ознаки суми і різниці тощо. Контраст великий для текстур з різкими перепадами між сусідніми пікселями; кореляція — для текстур з протяжними структурами вздовж напрямку зсуву; енергія — для регулярних текстур з невеликою кількістю переважаючих пар рівнів.

На рис. 11.3 наведено три природні текстури та розподіл ознак «контраст — кореляція» для їх фрагментів 64×64. Навіть у двовимірному просторі ознак класи текстур утворюють добре розділені хмари точок.

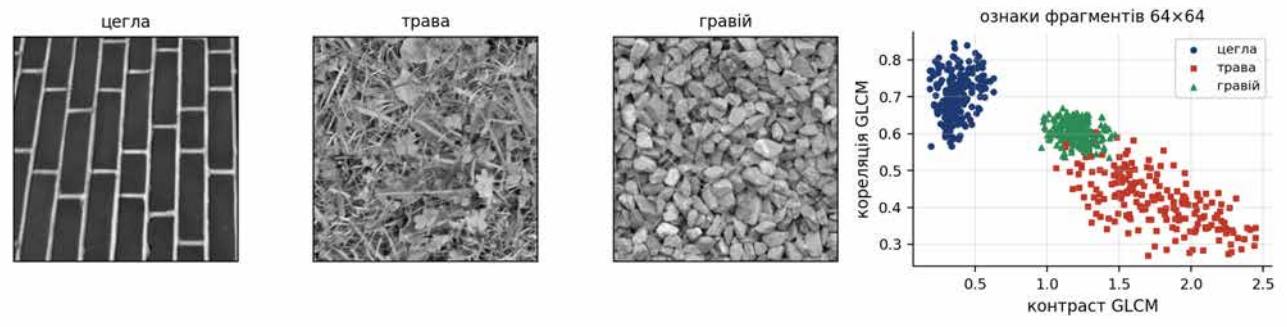


Рис. 11.3. Текстури та їх ознаки на основі матриць суміжності

11.6.2. Локальні бінарні шаблони

Локальний бінарний шаблон (LBP) [58] кодує локальну структуру околу пікселя: значення P сусідів, розташованих на колі радіусом R, порівнюються зі значенням центрального пікселя, і результати порівнянь утворюють двійкове число:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) 2^p, \qquad s(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \tag{11.10}$$

Оскільки використовується лише знак різниці, LBP інваріантний до будь-яких монотонних змін яскравості. Особливу роль відіграють рівномірні шаблони, у яких при обході кола кількість переходів між 0 і 1 не перевищує двох (вони відповідають плямам, кутам, краям і лініям); для $P = 8$ таких шаблонів 58, а з урахуванням інваріантності до повороту — 9 (плюс одна категорія для всіх нерівномірних). Ознакою текстури є гістограма кодів LBP у межах області (рис. 11.4). LBP-ознаки обчислюються дуже швидко, стійкі до змін освітлення і широко використовуються в розпізнаванні облич і текстур [58].

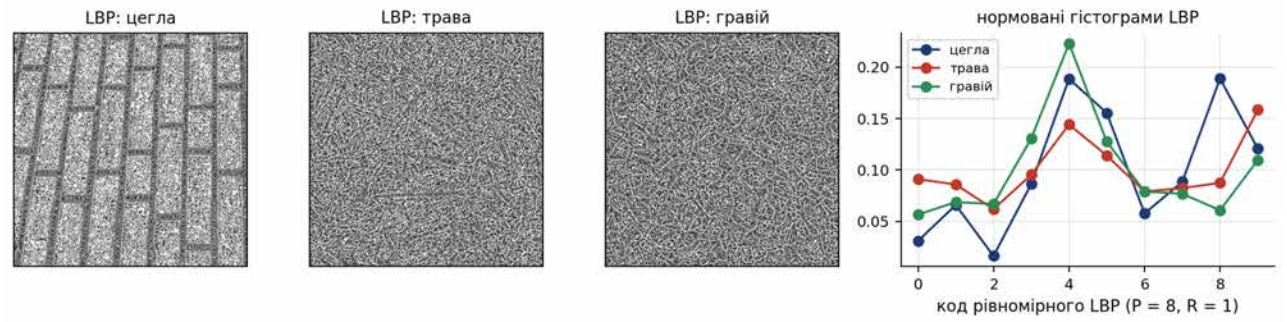


Рис. 11.4. Зображення кодів LBP та їх гістограми для трьох текстур

11.6.3. Спектральні та багатомасштабні текстурні ознаки

Інший підхід до опису текстур ґрунтується на частотному аналізі: енергії спектра Фур'є в кільцях і секторах (підрозділ 4.9), енергії субсмуг вейвлет-розкладу (6.14), середні значення і дисперсії модулів відгуків банку фільтрів Габора (підрозділ 6.6). Такі ознаки безпосередньо характеризують масштаб (розмір елементів) та орієнтацію текстури і добре узгоджуються з механізмами зорового сприйняття [24, 50].

11.7. Гістограми орієнтованих градієнтів і локальні ознаки

11.7.1. Гістограма орієнтованих градієнтів

Форма об’єкта значною мірою визначається розподілом напрямків контурів. Цю ідею реалізує дескриптор гістограм орієнтованих градієнтів (HOG) [59]:

1. обчислити градієнт у кожному пікселі (модуль і напрямок, розділ 7);
2. розбити зображення на невеликі комірки (наприклад, 8×8 пікселів) і для кожної комірки побудувати гістограму напрямків градієнта (зазвичай 9 інтервалів у діапазоні $0 \ldots 180°$), зважену модулями градієнта;
3. об’єднати сусідні комірки в блоки (наприклад, 2×2 комірки) з перекриттям і нормувати гістограми в межах кожного блоку — це забезпечує стійкість до локальних змін освітлення і контрасту;
4. об’єднати нормовані гістограми всіх блоків у єдиний вектор ознак.

Візуалізацію HOG-дескриптора показано на рис. 11.5: у кожній комірці зображено «зірочку» з відрізків, довжини яких пропорційні вагам напрямків. Дескриптор HOG у поєднанні з лінійним класифікатором став класичним методом виявлення пішоходів і багатьох інших класів об’єктів.



Рис. 11.5. Гістограма орієнтованих градієнтів

11.7.2. Характерні точки і їх дескриптори

Для розпізнавання об’єктів при частковому перекритті, суттєвих змінах ракурсу і масштабу використовують локальні ознаки: на зображенні знаходять характерні (особливі) точки — кути, плями, екстремуми масштабного простору — і для околу кожної обчислюють дескриптор. Детектор кутів Харріса [28] аналізує власні числа матриці, утвореної добутками похідних яскравості в околі точки (структурного тензора): кут — це точка, у якій обидва власні числа великі. Метод SIFT [52] знаходить екстремуми різниці гаусіанів у масштабному просторі (розділ 7), визначає домінуючу орієнтацію околу і формує дескриптор з гістограм орієнтованих градієнтів у сітці 4×4 підвікон (128 ознак), інваріантний до масштабу, повороту і стійкий до змін освітлення. Образ при цьому описується не одним вектором, а *множиною* дескрипторів; розпізнавання

зводиться до зіставлення множин (пошуку відповідностей) або до побудови гістограми «візуальних слів».

11.8. Нормування ознак

Ознаки, отримані різними способами, мають різні фізичні розмірності і діапазони значень: площа вимірюється в тисячах пікселів, компактність — у частках одиниці, інваріанти Ху — величини порядку 10^{-3} ... 10^{-12}. Більшість класифікаторів, що використовують відстані (розділ 1), без нормування фактично ігноруватимуть ознаки з малими діапазонами. Тому перед класифікацією ознаки нормують, використовуючи статистики, оцінені лише за навчальною вибіркою (застосування статистик тестової вибірки призводить до завищених оцінок якості) [35, 42]:

- z-нормування (1.6) — приведення до нульового середнього і одиничної дисперсії;
- мінімаксне нормування $\tilde{x} = (x - x_{min})/(x_{max} - x_{min})$ — приведення до діапазону $[0,1]$; чутливе до викидів;
- робастне нормування з використанням медіани і міжквартильного розмаху замість середнього і середньоквадратичного відхилення;
- нелінійні перетворення (логарифмування, степеневі) для ознак з сильно асиметричним розподілом, наприклад площ або енергій;
- вибілювання (англ. *whitening*) — лінійне перетворення, після якого коваріаційна матриця ознак стає одиничною; евклідова відстань після вибілювання дорівнює відстані Махаланобіса до центру вибірки [34].

Категоріальні (номінальні) ознаки кодують унітарним кодом (англ. *one-hot*): ознаці з C можливими значеннями відповідає C бінарних ознак, з яких одиничною є лише одна [42].

11.9. Зменшення розмірності

11.9.1. Метод головних компонент

Великі вектори ознак (сотні і тисячі компонент, як у HOG чи у випадку використання самих пікселів як ознак) створюють проблеми, пов'язані з прокляттям розмірності (розділ 1): потрібні величезні навчальні вибірки, коваріаційні матриці вироджені, обчислення громіздкі. Метод головних компонент (PCA), математично тотожний перетворенню Карунена — Лоева (розділ 5), знаходить ортонормований базис, у якому перші K компонент найкраще (у середньоквадратичному сенсі) подають навчальні вектори [49].

Нехай $\mathrm{x}_1, \ldots, \mathrm{x}_M$ — навчальні вектори розмірності n, $\bar{\mathrm{x}}$ — їх середнє, а

$$\mathrm{R} = \frac{1}{M-1}\sum_{i=1}^{M}(\mathrm{x}_i - \bar{\mathrm{x}})\,(\mathrm{x}_i - \bar{\mathrm{x}})^T \qquad (11.11)$$

— вибіркова коваріаційна матриця. Її власні вектори $\varphi_1, \ldots, \varphi_n$, впорядковані за спаданням власних чисел $\lambda_1 \geq \cdots \geq \lambda_n$, — головні компоненти. Нові ознаки — проєкції центрованого вектора на перші K компонент:

$$z_k = \varphi_k^T(\mathrm{x} - \bar{\mathrm{x}}), \qquad k = 1, \ldots, K. \tag{11.12}$$

Частка дисперсії, «пояснена» першими K компонентами, дорівнює $\sum_{k=1}^{K} \lambda_k \,/\, \sum_{k=1}^{n} \lambda_k$; кількість компонент зазвичай обирають так, щоб ця частка становила 90–99 %. На практиці головні компоненти обчислюють через сингулярний розклад центрованої матриці даних, що чисельно стійкіше, ніж розклад матриці (11.11) [49, 56].

На рис. 11.6 наведено результати PCA для набору зображень рукописних цифр розміром 8×8 пікселів (вектор ознак розмірності 64): 90 % дисперсії зосереджено в 21 компоненті (рис. 11.6, а); головні компоненти (рис. 11.6, б) нагадують узагальнені «штрихи» цифр; уже в проєкції на дві перші компоненти (рис. 11.6, в) центри більшості класів помітно розділені, хоча класи перекриваються.

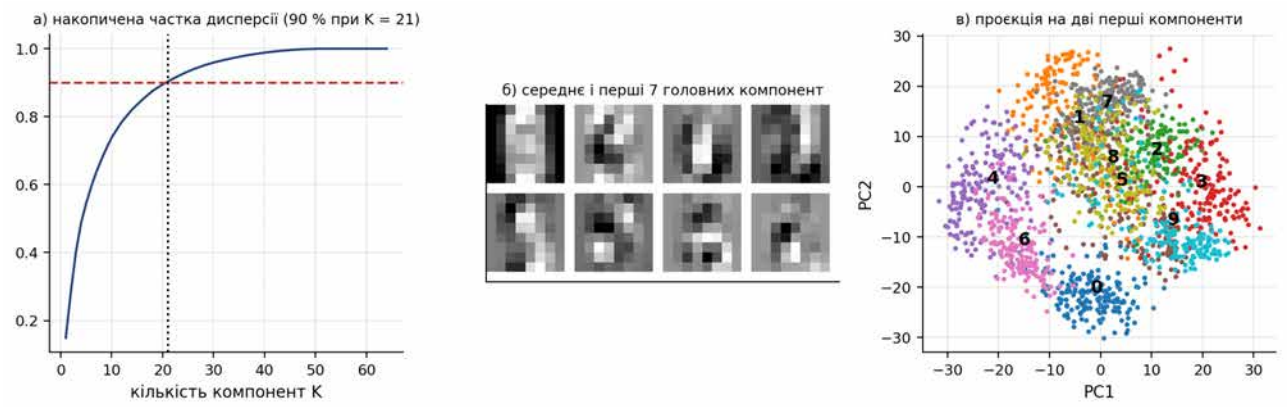


Рис. 11.6. Метод головних компонент для зображень рукописних цифр

Слід пам’ятати, що PCA — метод без учителя: він максимізує дисперсію даних, а не розділення класів. Напрямки найбільшої дисперсії не обов’язково є найбільш інформативними для класифікації. Наприклад, якщо дві витягнуті хмари класів розташовані паралельно і розділені в поперечному напрямку, перша головна компонента буде спрямована вздовж хмар і зовсім не розділятиме класи [17, 18].

11.9.2. Лінійний дискримінантний аналіз

Метод, що враховує мітки класів, запропонував Р. Фішер [35]. Лінійний дискримінантний аналіз (LDA) шукає напрямки w, що максимізують відношення міжкласового розсіювання до внутрішньокласового:

$$J(\mathrm{w}) = \frac{\mathrm{w}^T \mathrm{S}_B \mathrm{w}}{\mathrm{w}^T \mathrm{S}_W \mathrm{w}}, \quad \mathrm{S}_B = \sum_k M_k \,(\bar{\mathrm{x}}_k - \bar{\mathrm{x}})(\bar{\mathrm{x}}_k - \bar{\mathrm{x}})^T, \quad \mathrm{S}_W = \sum_k \sum_{\mathrm{x} \in \omega_k} (\mathrm{x} - \bar{\mathrm{x}}_k)\,(\mathrm{x} - \bar{\mathrm{x}}_k)^T, \tag{11.13}$$

де M_k — кількість образів класу ω_k, $\bar{\mathrm{x}}_k$ — середнє класу. Розв’язком є власні вектори узагальненої задачі $\mathrm{S}_B\mathrm{w} = \lambda \mathrm{S}_W\mathrm{w}$; для K класів існує не більше $K - 1$ інформативних напрямків. Зауважимо, що функціонал (11.13) — це відношення двох квадратичних форм, так само як критерій гостроти (8.32) у методі

деконволюції за критерієм форми імпульсної характеристики, і задача так само зводиться до задачі на власні значення. LDA одночасно є методом зменшення розмірності і лінійним класифікатором; детально він розглядатиметься в другій частині посібника.

11.10. Відбір ознак

11.10.1. Критерії інформативності

На відміну від зменшення розмірності, відбір ознак не синтезує нових ознак, а вибирає підмножину наявних. Це зберігає інтерпретованість і зменшує витрати на обчислення ознак у працюючій системі. Для оцінювання інформативності окремої ознаки x використовують:

- критерій Фішера (відношення міжкласової дисперсії до внутрішньокласової для однієї ознаки):

$$F(x) = \frac{\sum_k M_k \, (\bar{x}_k - \bar{x})^2}{\sum_k M_k \; \sigma_k^2}, \tag{11.14}$$

де $\bar{x}_k$, σ_k^2 — середнє і дисперсія ознаки в класі ω_k. Велике значення означає, що класи «далеко» один від одного порівняно з розкидом усередині класів [17, 20];

- взаємну інформацію між ознакою і міткою класу $I(x; y) = H(y) - H(y \mid x)$, яка враховує і нелінійні залежності [21];
- статистичні критерії (дисперсійний аналіз, χ^2 для категоріальних ознак).

На рис. 11.7, а показано значення критерію Фішера для кожного з 64 пікселів зображень цифр: найменш інформативні крайні пікселі (вони майже завжди порожні), найбільш інформативні — центральна область, де розташовані характерні для різних цифр штрихи.

11.10.2. Стратегії відбору

Критерії для окремих ознак не враховують надлишковості: дві сильно корельовані ознаки можуть обидві мати високий критерій Фішера, але разом дати не більше інформації, ніж кожна окремо. Тому застосовують стратегії відбору підмножин [20, 35]:

- фільтрові методи — ранжування ознак за критерієм і вибір кращих (найшвидші, але ігнорують взаємодію ознак);
- обгорткові методи (англ. *wrapper*) — оцінювання якості підмножини безпосередньо за точністю класифікатора на валідаційній вибірці; перебір організовують жадібно: послідовне додавання (почати з порожньої множини і на кожному кроці додавати ознаку, що найбільше підвищує точність) або послідовне вилучення (почати з повної множини і вилучати найменш корисну);
- вбудовані методи — відбір ознак у процесі навчання класифікатора (наприклад, регуляризація, що «обнуляє» ваги неінформативних ознак, або оцінки важливості ознак у деревах рішень) [35, 42].

11.10.3. Порівняльний експеримент

На рис. 11.7, б наведено залежність точності найпростішого класифікатора — за найближчим центром класу (див. (1.3)) — від кількості ознак для двох способів зменшення розмірності: відбору пікселів за критерієм Фішера і переходу до головних компонент. Синтезовані ознаки PCA значно ефективніші при малій кількості ознак: 8 головних компонент дають точність близько 85 %, тоді як 8 найкращих за Фішером пікселів — лише близько 72 %. Це пояснюється тим, що головні компоненти «збирають» інформацію з багатьох корельованих пікселів, тоді як окремі пікселі несуть багато надлишкової інформації. Починаючи з 20–30 ознак, обидва підходи досягають точності близько 90 %, що відповідає можливостям самого класифікатора за найближчим центром; подальше її підвищення потребує складніших вирішальних правил, які будуть розглянуті в другій частині посібника.

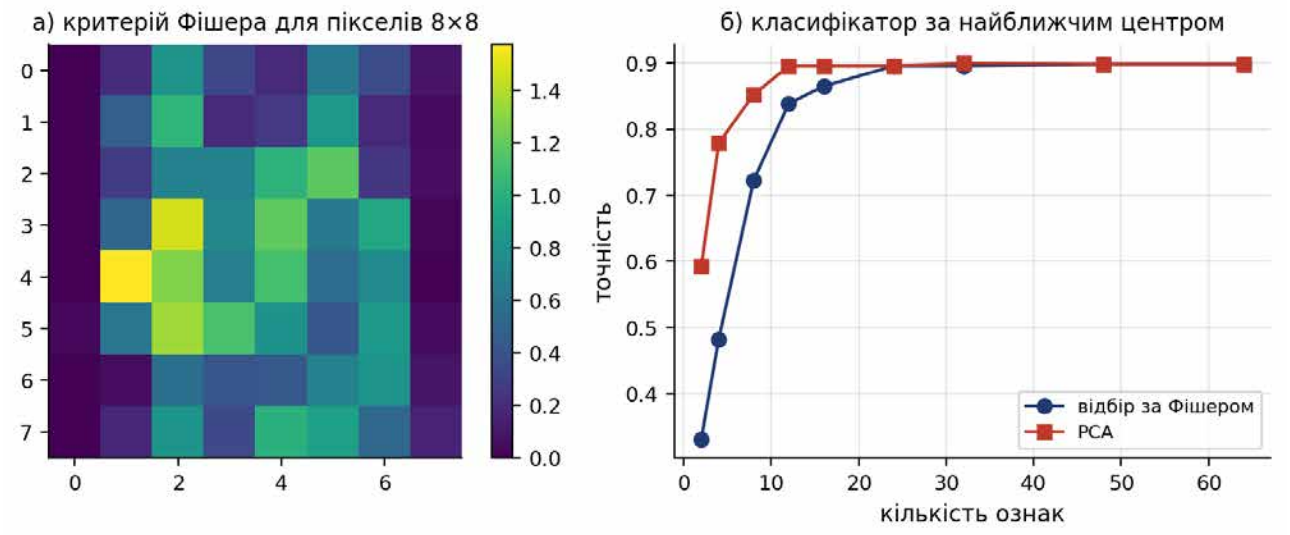


Рис. 11.7. Критерій Фішера для пікселів (а) і точність класифікації залежно від кількості ознак (б)

11.11. Практичні рекомендації щодо формування ознак

Узагальнюючи матеріал розділу і попередніх розділів, можна сформулювати такі рекомендації [18, 35, 48].

1. Починайте з аналізу того, *які чинники мають і які не мають* впливати на рішення (положення, орієнтація, масштаб, освітлення, шум). Це визначає, які інваріантності мають забезпечувати ознаки або етапи нормування.
2. Віддавайте перевагу простим, фізично зрозумілим ознакам; складні ознаки додавайте тоді, коли прості виявилися недостатніми.
3. Перевіряйте розподіли ознак за класами (гістограми, діаграми розсіювання) — це дозволяє виявити неінформативні ознаки, викиди і помилки обчислень ще до навчання класифікатора.
4. Обов’язково нормуйте ознаки за статистиками навчальної вибірки.
5. При великій кількості ознак застосовуйте PCA або відбір ознак, контролюючи результат на валідаційній вибірці.

6. Пам’ятайте про зв’язок ознак з попереднім обробленням: шум і розмиття найбільше впливають на ознаки, що використовують високі частоти (контури, моменти вищих порядків, текстурні ознаки малого масштабу). Реставрація (розділи 8, 9) і придушення шуму (розділи 6, 7) підвищують стабільність саме цих ознак.

11.12. Програмна реалізація

```
import numpy as np
from skimage import measure, data, img_as_ubyte
from skimage.feature import graycomatrix, graycoprops, local_binary_pattern, hog
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.neighbors import NearestCentroid

# геометричні ознаки та інваріанти Ху для об'єктів бінарного зображення
b = data.coins() > 110
for r in measure.regionprops(measure.label(b)):
    if r.area < 300: continue
    comp = 4 * np.pi * r.area / r.perimeter**2          # (11.1)
    hu = r.moments_hu
    hu_log = -np.sign(hu) * np.log10(np.abs(hu) + 1e-30)
    print(r.label, r.area, round(comp, 3), np.round(hu_log[:4], 2))

# текстурні ознаки GLCM (11.8)-(11.9) і гістограма LBP (11.10)
tex = img_as_ubyte(data.brick())[:128, :128] // 32      # 8 рівнів
P = graycomatrix(tex, [1, 2], [0, np.pi/4, np.pi/2, 3*np.pi/4], levels=8,
                 symmetric=True, normed=True)
glcm = [graycoprops(P, p).mean() for p in
        ('contrast', 'homogeneity', 'energy', 'correlation')]
lbp = local_binary_pattern(data.brick()[:128, :128], 8, 1, 'uniform')
h_lbp, _ = np.histogram(lbp, bins=np.arange(11), density=True)

# HOG-дескриптор
fd = hog(data.camera(), orientations=9, pixels_per_cell=(16, 16),
         cells_per_block=(2, 2))

# нормування, PCA і класифікатор за найближчим центром
X, y = load_digits(return_X_y=True)
Xtr, Xte, ytr, yte = train_test_split(X, y, test_size=0.4,
                                      stratify=y, random_state=0)
sc = StandardScaler().fit(Xtr)                  # статистики лише навч. вибірки
pca = PCA(n_components=0.9).fit(sc.transform(Xtr))
```

clf = NearestCentroid().fit(pca.transform(sc.transform(Xtr)), ytr)
print(pca.n_components_, clf.score(pca.transform(sc.transform(Xte)), yte))

11.13. Приклади розв'язання задач

Задача 11.1. Бінарний образ складається з пікселів з координатами (x, y): $(1,1), (2,1), (3,1), (1,2), (2,2), (1,3), (2,3)$ (x — номер стовпця, y — номер рядка). Обчисліть площу, центр ваги, центральні моменти другого порядку та інваріанти Ху ϕ_1, ϕ_2.

Розв'язання. Площа $m_{00} = 7$. Центр ваги: $\bar{x} = (1 + 2 + 3 + 1 + 2 + 1 + 2)/7 = 12/7 \approx 1{,}714$; $\bar{y} = (1 + 1 + 1 + 2 + 2 + 3 + 3)/7 = 13/7 \approx 1{,}857$. Центральні моменти (11.3): $\mu_{20} = \sum(x - \bar{x})^2 \approx 3{,}429$; $\mu_{02} = \sum(y - \bar{y})^2 \approx 4{,}857$; $\mu_{11} = \sum(x - \bar{x})(y - \bar{y}) \approx -1{,}286$. Нормовані моменти (11.5) з $\gamma = 2$: $\eta_{20} = 3{,}429/49 \approx 0{,}0700$, $\eta_{02} \approx 0{,}0991$, $\eta_{11} \approx -0{,}0262$. Інваріанти (11.6): $\phi_1 = 0{,}0700 + 0{,}0991 \approx 0{,}169$; $\phi_2 = (0{,}0700 - 0{,}0991)^2 + 4 \cdot 0{,}0262^2 \approx 0{,}0036$. Від'ємне значення μ_{11} показує, що головна вісь інерції нахилена: зі зростанням x координата y у середньому зменшується.

Задача 11.2. Ознака має в класі ω_1 значення 2,3,4, у класі ω_2 — 6,7,8. Обчисліть критерій Фішера.

Розв'язання. Середні класів $\bar{x}_1 = 3$, $\bar{x}_2 = 7$, загальне середнє $\bar{x} = 5$. Чисельник (11.14): $3(3 - 5)^2 + 3(7 - 5)^2 = 24$. Внутрішньокласові дисперсії (зміщені) $\sigma_1^2 = \sigma_2^2 = 2/3$, знаменник $3 \cdot 2/3 + 3 \cdot 2/3 = 4$. Критерій $F = 24/4 = 6$. Класи за цією ознакою розділяються без помилок порогом 5.

Задача 11.3. Для набору даних власні числа коваріаційної матриці дорівнюють 5,0; 2,0; 1,5; 0,8; 0,4; 0,2; 0,1. Скільки головних компонент потрібно зберегти, щоб пояснити не менше 90 % дисперсії?

Розв'язання. Сума власних чисел — 10,0. Накопичені частки: 50 %, 70 %, 85 %, 93 %, … Отже, потрібно зберегти чотири компоненти (93 % дисперсії); три компоненти пояснюють лише 85 %.

Висновки до розділу 11

1. Ознаки поділяють на геометричні, контурні, яскравісні, текстурні, спектральні та локальні; вибір ознак визначається тим, до яких змін образу рішення має бути інваріантним.
2. Прості геометричні ознаки (площа, периметр, компактність, ексцентриситет, опуклість, число Ейлера) інтерпретовані і часто достатні для розрізнення об'єктів простої форми.
3. Центральні моменти інваріантні до зсуву, нормовані — до масштабу; інваріанти Ху — до зсуву, масштабу і повороту; моменти другого порядку визначають орієнтацію і ексцентриситет.
4. Контур описують ланцюговими кодами, кривиною і дескрипторами Фур'є.
5. Текстуру описують матрицями суміжності та ознаками Харалiка, гістограмами локальних бінарних шаблонів, а також енергіями частотних і вейвлет-субсмуг і відгуками фільтрів Габора.

6. Гістограми орієнтованих градієнтів описують форму через розподіл напрямків контурів; локальні дескриптори характерних точок забезпечують стійкість до перекриття і змін ракурсу.
7. Ознаки необхідно нормувати за статистиками навчальної вибірки. Метод головних компонент (тотожний перетворенню Карунена — Лоева) зменшує розмірність без учителя; лінійний дискримінантний аналіз враховує мітки класів. Відбір ознак виконують за критеріями інформативності (Фішера, взаємної інформації) або за якістю класифікатора.

Питання для обговорення та самоперевірки

1. Наведіть класифікацію ознак, що використовуються в розпізнаванні образів.
2. Як обчислюються площа, периметр, компактність, опуклість, число Ейлера? До яких перетворень вони інваріантні?
3. Чому компактність дискретного круга менша за одиницю?
4. Запишіть визначення геометричних, центральних і нормованих центральних моментів. До яких перетворень інваріантні кожні з них?
5. Як за моментами другого порядку визначити орієнтацію і ексцентриситет образу?
6. Запишіть перші два інваріанти Ху. Який інваріант змінює знак при віддзеркаленні?
7. Що таке ланцюговий код Фрімена? Як забезпечити його інваріантність до повороту і вибору початкової точки?
8. Як будується матриця суміжності рівнів яскравості? Які ознаки Хараліка ви знаєте?
9. Як обчислюється код LBP? Чому LBP інваріантний до монотонних змін яскравості? Що таке рівномірні шаблони?
10. Опишіть етапи обчислення дескриптора HOG. Навіщо виконується нормування в блоках?
11. Що таке характерні точки? Як працює детектор Харріса?
12. Навіщо нормувати ознаки? Чому статистики для нормування слід обчислювати лише за навчальною вибіркою?
13. Сформулюйте задачу методу головних компонент. Як він пов’язаний з перетворенням Карунена — Лоева?
14. Чому PCA може бути неефективним для розділення класів?
15. Запишіть критерій лінійного дискримінантного аналізу. Скільки інформативних напрямків він дає для K класів?
16. Запишіть критерій Фішера для окремої ознаки. Чому ранжування ознак за ним не враховує їх надлишковість?
17. Порівняйте фільтрові, обгорткові та вбудовані методи відбору ознак.

Завдання для самостійної роботи

1. Сегментуйте зображення з кількома об’єктами (монети, зерна, клітини), обчисліть для кожного об’єкта геометричні ознаки та інваріанти Ху. Побудуйте діаграми розсіювання пар ознак і визначте найбільш інформативні.
2. Перевірте інваріантність інваріантів Ху до повороту для кутів $0 \ldots 90°$ з кроком $5°$ і до масштабу $0{,}3 \ldots 2{,}0$. При яких розмірах об’єкта похибки дискретизації стають суттєвими?
3. Реалізуйте обчислення ланцюгового коду і різницевого коду контуру; побудуйте гістограми напрямків для кількох класів фігур.
4. Сформуйте набір фрагментів 5–6 текстур (наприклад, з тестових зображень бібліотеки scikit-image). Порівняйте якість класифікації за найближчим центром для ознак GLCM, гістограм LBP, вейвлет-енергій (6.14) та енергій Габора.
5. Обчисліть HOG-дескриптори для зображень цифр і порівняйте точність класифікації з використанням пікселів як ознак.
6. Реалізуйте PCA через сингулярний розклад і перевірте збіг з бібліотечною реалізацією. Відновіть зображення цифр за $K = 5{,}10{,}20{,}40$ компонентами і обчисліть похибку.
7. Реалізуйте LDA для набору цифр і порівняйте проєкцію на дві перші дискримінантні компоненти з проєкцією на дві головні компоненти (рис. 11.6, в).
8. Реалізуйте послідовне додавання ознак (обгортковий метод) з класифікатором за найближчим центром і порівняйте результат з ранжуванням за критерієм Фішера.

РОЗДІЛ 12. НАСКРІЗНИЙ ПРИКЛАД: РОЗПІЗНАВАННЯ СПОТВОРЕНИХ СИМВОЛІВ

Навчальні цілі. Після вивчення розділу ви зможете: спланувати обчислювальний експеримент з дослідження системи розпізнавання; поєднати в одному конвеєрі моделювання деградації, оцінювання параметрів шуму, реставрацію, формування ознак і класифікацію за еталоном; обґрунтовано інтерпретувати результати експерименту і робити висновки щодо доцільності окремих етапів попереднього оброблення.

12.1. Мета і постановка задачі

У попередніх розділах окремі етапи системи розпізнавання розглядалися відокремлено: подання і дискретизація образів (розділ 2), моделі спотворень (розділ 3), ортогональні перетворення (розділи 4–6), попереднє оброблення (розділ 7), реставрація (розділи 8, 9), розпізнавання за еталоном (розділ 10) і формування ознак (розділ 11). Мета цього розділу — показати, як ці етапи взаємодіють у цілісній системі, і на конкретному обчислювальному експерименті відповісти на практичне запитання: чи підвищує реставрація образів точність їх розпізнавання і наскільки.

Як об’єкт дослідження оберемо задачу, що зустрічається, наприклад, при розпізнаванні номерних знаків, маркування виробів чи показів приладів [28, 31]: розпізнавання окремих друкованих цифр на зображеннях, отриманих у неідеальних умовах. Зображення кожної цифри має розмір 48×48 пікселів і зазнає спотворень трьох видів:

1. геометричних — випадковий поворот у межах $\pm 6°$, зміна масштабу в межах $\pm 8\,\%$, зсув на ± 2 пікселі;
2. локальних (розмиття) — гаусівське розмиття, що моделює розфокусування і аберації оптики, або розмиття рухом, що моделює переміщення об’єкта чи камери під час експозиції;
3. шумових — адитивний білий гаусівський шум.

Крім того, образи подано *різними шрифтами*: еталони (і навчальна вибірка) сформовані з трьох шрифтів, а тестова вибірка — з п’яти, тобто система має узагальнювати на шрифти, яких не було під час навчання (рис. 12.1). Ця обставина наближає модельну задачу до реальних умов, де форма символів відрізняється від еталонної.

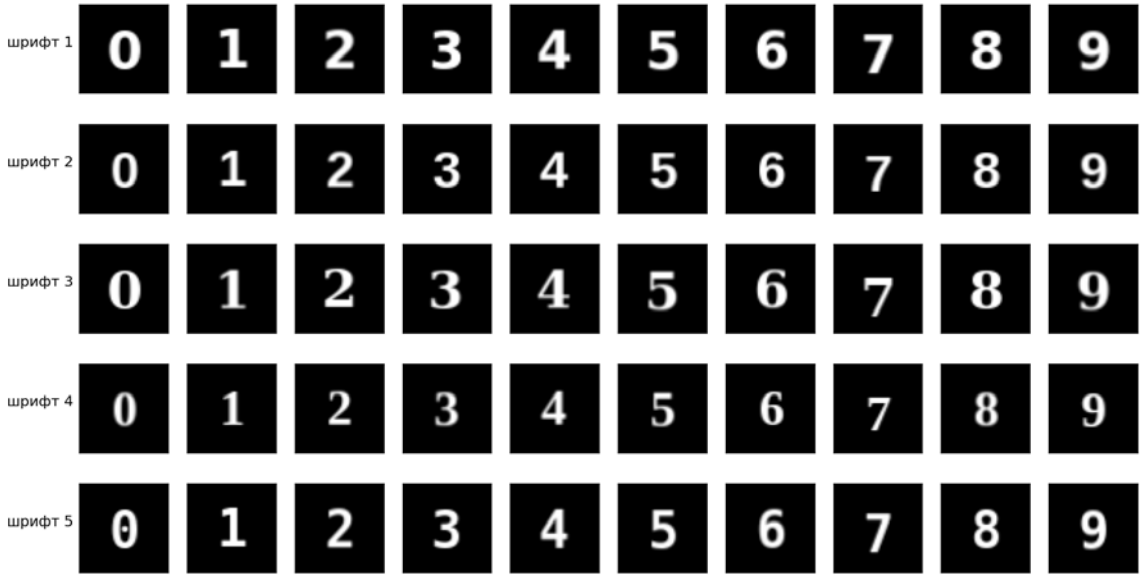


Рис. 12.1. Образи цифр, сформовані п’ятьма шрифтами (шрифти 1–3 використано для еталонів і навчання)

12.2. Структура досліджуваної системи

Досліджувана система складається з таких етапів (порівняйте з рис. 1.2):

1. Реєстрація (моделювання): еталонне зображення цифри піддається випадковому афінному перетворенню, згортці з імпульсною характеристикою системи відображення (циклічна модель, розділ 3) і додаванню шуму:
$$\mathrm{L} = \mathrm{G}\,\mathcal{A}\{\mathrm{f}\} + \mathrm{V}, \tag{12.1}$$
де $\mathcal{A}$ — афінне перетворення з випадковими параметрами.
2. Оцінювання рівня шуму за спостережуваним образом робастним методом (3.26) [24]: середньоквадратичне відхилення оцінюється за медіаною модулів коефіцієнтів діагональної субсмуги першого рівня вейвлет-розкладу Хаара, а енергія шуму — за формулою $\hat{\varepsilon}^2 = M^2\hat{\sigma}^2$. Важливо, що *істинне значення* σ системі не повідомляється.
3. Реставрація методом умовної деконволюції в області трансформант Фур’є (розділ 8) з оператором гладкості у вигляді лапласіана (8.14) і автоматичним визначенням параметра λ з рівняння (8.23) [2, 3]. Імпульсна характеристика системи вважається відомою (ідентифікованою заздалегідь, розділ 3).
4. Розпізнавання двома способами:
   – за еталоном — нормованою взаємною кореляцією (10.3) з еталонами всіх класів (по три шрифти на клас) з пошуком максимуму в околі ± 4 пікселі; образ відноситься до класу еталона з найбільшим значенням NCC [54];
   – за ознаками — обчислення дескриптора HOG (розділ 11) [59] і класифікація за найближчим центром класу (1.3); центри класів обчислено за навчальною вибіркою з 15 випадково геометрично

перетворених, але не розмитих і не зашумлених копій кожного еталона.

Для кожного способу розпізнавання точність визначається двічі: для спотвореного образу без реставрації і для реставрованого образу. Таким чином, порівнюються чотири варіанти системи.

12.3. Методика експерименту

Під час планування обчислювальних експериментів з системами розпізнавання слід дотримуватися кількох принципів.

Незалежність тестових даних. Тестова вибірка не повинна використовуватися ні для навчання класифікатора, ні для вибору параметрів оброблення [35]. У нашому експерименті параметри реставрації визначаються автоматично для кожного образу, а класифікатори побудовано лише за трьома «навчальними» шрифтами.

Однакові умови для порівнюваних варіантів. Усі чотири варіанти системи обробляють *одні й ті самі* реалізації спотворених образів. Це усуває випадкові відмінності, зумовлені різними реалізаціями шуму та геометричних перетворень, і дозволяє коректно порівнювати варіанти навіть на вибірках помірного обсягу.

Контроль відтворюваності. Генератор випадкових чисел ініціалізується фіксованим значенням, що дозволяє точно відтворити результати.

Варіювання умов. Досліджуються кілька рівнів розмиття і шуму та два типи імпульсної характеристики. Це дозволяє зробити висновки не для однієї випадково обраної ситуації, а для діапазону умов.

Достатній обсяг вибірки. Для кожної комбінації умов сформовано 200 тестових образів (10 цифр × 5 шрифтів × 4 реалізації спотворень). Стандартна похибка оцінки точності p за n незалежними випробуваннями дорівнює $\sqrt{p(1-p)/n}$; для $p = 0{,}9$ і $n = 200$ вона становить близько 2 %. Отже, відмінності в кілька відсотків слід тлумачити обережно, а відмінності в десятки відсотків є статистично значущими [18, 35].

Досліджено п’ять комбінацій умов спотворення: гаусівське розмиття з параметром $\sigma_b = 1{,}5$ і $2{,}5$ пікселі, розмиття рівномірним рухом на 11 пікселів під кутом 20° та два рівні шуму $\sigma = 0{,}02$ і $0{,}05$ (при діапазоні яскравостей $[0,1]$).

12.4. Результати

Приклади спотворених і реставрованих образів наведено на рис. 12.2. При слабкому розмитті і низькому шумі реставрація повертає символу чіткі контури. При сильному гаусівському розмитті ($\sigma_b = 2{,}5$) і підвищеному шумі реставрований образ залишається розмитим: метод умовної деконволюції «чесно» не відновлює деталей, повністю поглинутих шумом, і обмежується тим рівнем різкості, який узгоджується з енергією шуму [2]. Для розмиття рухом реставрація особливо ефективна: характерне «подвоєння» штрихів практично усувається.

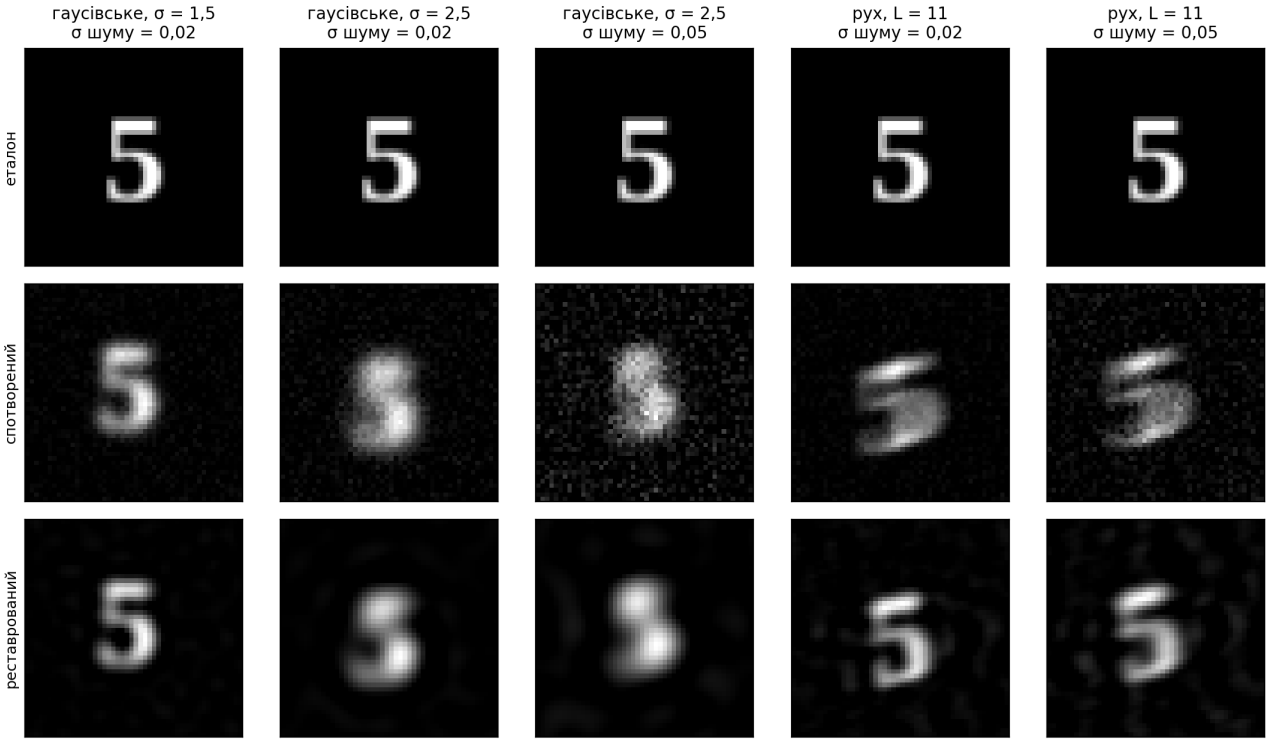


Рис. 12.2. Еталон (вгорі), спотворені (посередині) та реставровані (внизу) образи цифри «5» (шрифт 4) для різних умов

Кількісні результати наведено в табл. 12.1 і на рис. 12.3.

Таблиця 12.1

Точність розпізнавання (%) спотворених цифр

Розмиття	СКВ шуму	NCC без реставрації	NCC з реставрацією	HOG без реставрації	HOG з реставрацією
Гаусівське, $\sigma_b = 1{,}5$	0,02	97,5	100,0	84,5	93,0
Гаусівське, $\sigma_b = 2{,}5$	0,02	83,5	95,0	42,0	71,0
Гаусівське, $\sigma_b = 2{,}5$	0,05	84,5	89,5	33,0	57,0
Рух, $L = 11$	0,02	77,5	97,0	67,5	87,5
Рух, $L = 11$	0,05	81,0	92,5	48,5	80,0

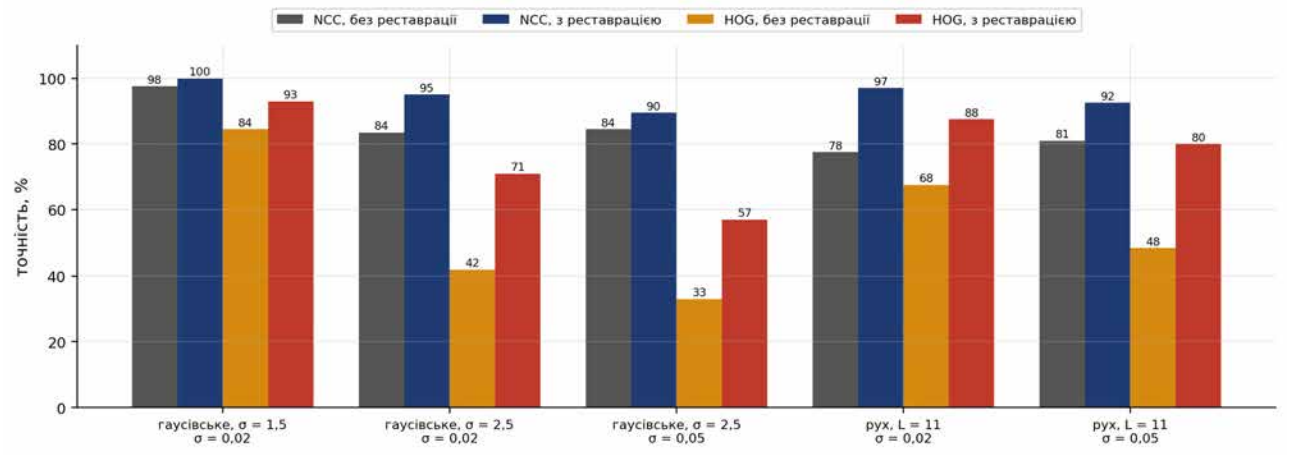


Рис. 12.3. Точність розпізнавання для різних умов спотворення і варіантів системи

12.5. Аналіз результатів

Реставрація підвищує точність розпізнавання в усіх досліджених умовах. Для класифікатора за еталоном (NCC) приріст становить від 2,5 до 19,5 процентного пункту, для класифікатора на основі ознак HOG — від 8,5 до 31,5 пункту. Найбільший виграш спостерігається для розмиття рухом: анізотропне розмиття змінює форму символу так, що він стає схожим на інші цифри (наприклад, «5» з «подвоєними» штрихами нагадує «6» або «8»), а реставрація усуває саме це спотворення форми.

Класифікатор за еталоном стійкіший до розмиття, ніж класифікатор за ознаками HOG. Це закономірно: NCC порівнює образ з еталоном цілісно, а розмиття змінює загальний розподіл яскравості відносно повільно. Натомість HOG ґрунтується на напрямках градієнтів, а розмиття і шум руйнують саме локальні градієнти: розмиття зменшує їх модулі і «розпливає» напрямки, шум додає хаотичні градієнти. До того ж центри класів для HOG обчислено за нерозмитими образами, тобто модель класу не враховувала розмиття. Цей результат ілюструє загальне правило, сформульоване в розділі 11: ознаки, що використовують високочастотну інформацію (контури, градієнти, моменти вищих порядків), найбільше страждають від розмиття і шуму, і саме для них попереднє оброблення найважливіше.

Збільшення шуму зменшує ефект реставрації. Для гаусівського розмиття з $\sigma_b = 2{,}5$ підвищення шуму з 0,02 до 0,05 зменшило приріст точності NCC з 11,5 до 5 пунктів. Це прямо випливає з природи некоректної оберненої задачі (розділ 3) [33, 45]: чим вищий рівень шуму, тим менша частина спектра може бути відновлена, і тим ближчим до спотвореного є реставрований образ. Невелике зростання точності NCC без реставрації при підвищенні шуму (з 83,5 до 84,5 %) лежить у межах статистичної похибки.

Автоматичне визначення параметрів працює. Енергія шуму оцінювалася за самим спотвореним образом, а параметр регуляризації — за принципом нев'язки [2, 45]. Отже, система не потребує ручного налаштування для кожного рівня шуму, що є важливою практичною перевагою методу умовної деконволюції (розділ 8).

12.6. Порівняння методів реставрації в складі системи розпізнавання

Метод умовної деконволюції в розглянутій системі використовувався з автоматичним вибором параметра. Корисно порівняти його з іншими методами реставрації (розділ 8), параметри яких зафіксовано заздалегідь: псевдоінверсним фільтром з порогом $|G| > 0{,}1$ і фільтром Вінера зі сталим $K = 10^{-3}$ та $K = 10^{-2}$. Порівняння виконано для двох найскладніших умов на новій незалежній тестовій вибірці (по 200 образів) з класифікатором NCC.

Таблиця 12.2

Точність розпізнавання (%) з різними методами реставрації (класифікатор NCC)

Умови спотворення	Без реставрації	Псевдоінверсний	Вінер, $K = 10^{-3}$	Вінер, $K = 10^{-2}$	Умовна деконволюція
Гаусівське, $\sigma_b = 2{,}5$; $\sigma = 0{,}02$	83,5	99,5	100,0	99,0	94,5
Рух, $L = 11$; $\sigma = 0{,}05$	79,5	97,5	99,0	97,5	95,0

Результати заслуговують на уважне обговорення. Усі методи реставрації суттєво підвищують точність розпізнавання. Водночас для класифікації за еталоном фільтри з фіксованими параметрами виявилися дещо ефективнішими за умовну деконволюцію з автоматичним вибором λ. Причина полягає в різниці критеріїв. Метод умовної деконволюції обирає найгладкішу оцінку, *узгоджену з даними в межах рівня шуму* (принцип нев'язки); відомо, що такий вибір дає дещо надмірно згладжені оцінки [45]. Для класифікатора NCC, який порівнює реставрований образ з різкими еталонами, вигідніша менш гладка оцінка, навіть якщо вона містить більше залишкового шуму: кореляційна міра сама частково усереднює шум. Інакше кажучи, *оптимальність реставрації за критерієм відновлення образу не тотожна оптимальності за критерієм розпізнавання*.

З цього випливає практична рекомендація: якщо реставрація є етапом системи розпізнавання, її параметри доцільно уточнювати за кінцевим критерієм — точністю розпізнавання на валідаційній вибірці. Для методу умовної деконволюції природним параметром налаштування є множник при енергії шуму в умові (8.10). Проте зменшувати цей множник слід обережно. У додатковому експерименті зменшення цільової енергії нев'язки вдвічі для розмиття рухом підвищило точність до 99 %, а для сильного гаусівського розмиття призвело до втрати стійкості і падіння точності до 10 %: частотна характеристика гаусівського розмиття настільки мала на високих частотах, що нев'язку меншу за енергію шуму можна отримати лише ціною практично інверсної фільтрації. Цей приклад ще раз ілюструє зміст некоректності оберненої задачі (розділ 3) і важливість точної оцінки енергії шуму. Перевагою умовної деконволюції залишається те, що вона не потребує ні налаштування на кожен рівень шуму, ні знання спектрів образів, і забезпечує стійкий результат в усьому діапазоні умов.

12.7. Обмеження і напрямки розвитку

Результати модельного експерименту слід тлумачити з урахуванням його обмежень.

1. Імпульсна характеристика вважалася точно відомою. На практиці її доводиться ідентифікувати (розділ 3), і похибки ідентифікації знижують ефективність реставрації. Корисним продовженням є дослідження чутливості результатів до похибок задання параметрів розмиття.
2. Використано циклічну модель згортки; для образів, де символ не займає всього кадру, вона прийнятна, але в загальному випадку потрібні заходи проти крайових ефектів (розділ 8).
3. Класифікатори є найпростішими. Сучасні класифікатори (зокрема нейромережеві), навчені на спотворених даних, можуть частково «навчитися» компенсувати розмиття [26, 30]. Проте і для них реставрація чи інше попереднє оброблення зазвичай зменшує потрібний обсяг навчальних даних і підвищує стійкість до умов, не представлених у навчальній вибірці.
4. Умовну деконволюцію застосовано лише в базисі Фур'є. Порівняння з реставрацією в базисах RTF і Адамара (розділ 9) є змістовним завданням для самостійної роботи.

12.8. Програмна реалізація експерименту

Нижче наведено основні фрагменти програми експерименту. Функції psf2otf і cond_deconv_fourier наведено в розділі 8.

```
import numpy as np
from skimage import transform, feature
from skimage.feature import hog
from sklearn.neighbors import NearestCentroid
import pywt

S = 48

def degrade(a, h, sigma, rng):
    """Модель (12.1): афінне перетворення, циклічне розмиття, шум."""
    ang, sc = rng.uniform(-6, 6), rng.uniform(0.92, 1.08)
    c = np.array([S / 2, S / 2])
    tf = (transform.AffineTransform(translation=-c)
          + transform.AffineTransform(scale=sc, rotation=np.radians(ang))
          + transform.AffineTransform(translation=c + rng.uniform(-2, 2, 2)))
    w = transform.warp(a, tf.inverse, order=1)
    b = np.real(np.fft.ifft2(np.fft.fft2(w) * psf2otf(h, a.shape)))
    return b + rng.normal(0, sigma, a.shape)

def sigma_mad(im):                          # оцінка (3.26)
    _, (_, _, hh) = pywt.dwt2(im, 'haar')
```

```
    return np.median(np.abs(hh)) / 0.6745

def restore(L, h):
    eps2 = sigma_mad(L)**2 * L.size              # енергія шуму
    f_hat, lam = cond_deconv_fourier(L, h, eps2)    # розділ 8
    return f_hat

def ncc_classify(im, templates):
    """templates: {клас: [еталон1, еталон2, ...]}"""
    p = np.pad(im, 4, mode='edge')
    scores = {k: max(feature.match_template(p, t).max() for t in ts)
              for k, ts in templates.items()}
    return max(scores, key=scores.get)

def hog_feat(im):
    return hog(im, orientations=9, pixels_per_cell=(8, 8),
               cells_per_block=(2, 2))

# навчання: X_train - HOG-ознаки геометрично перетворених еталонів
# clf = NearestCentroid().fit(X_train, y_train)
# тестування: для кожного образу L = degrade(...) обчислити
#   ncc_classify(L, ...), ncc_classify(restore(L, h), ...),
#   clf.predict([hog_feat(L)]), clf.predict([hog_feat(restore(L, h))])
```

12.9. Приклад оцінювання значущості відмінностей

Задача 12.1. У табл. 12.1 для розмиття рухом при $\sigma = 0{,}05$ точність класифікатора NCC зросла з 81,0 % до 92,5 %. Чи є ця відмінність статистично значущою?

Розв'язання. Обидва варіанти випробовувалися на одних і тих самих 200 образах, тому коректно використовувати критерій для зв'язаних вибірок (критерій Мак-Немара), який враховує лише образи, на яких рішення варіантів різняться. Наближену оцінку можна отримати і без нього. Стандартні похибки окремих оцінок: $\sqrt{0{,}81 \cdot 0{,}19/200} \approx 2{,}8\,\%$ і $\sqrt{0{,}925 \cdot 0{,}075/200} \approx 1{,}9\,\%$. Для незалежних вибірок стандартна похибка різниці була б $\sqrt{2{,}8^2 + 1{,}9^2} \approx 3{,}4\,\%$; спостережувана різниця 11,5 % перевищує її більш ніж утричі. Для зв'язаних вибірок похибка різниці ще менша. Отже, ефект реставрації статистично значущий.

Задача 12.2. Скільки тестових образів потрібно, щоб стандартна похибка оцінки точності близько 95 % не перевищувала 1 %?

Розв'язання. З умови $\sqrt{p(1-p)/n} \le 0{,}01$ при $p = 0{,}95$ отримуємо $n \ge 0{,}95 \cdot 0{,}05 / 0{,}0001 = 475$. Отже, для впевненого розрізнення варіантів, що відрізняються на 2–3 %, потрібні тестові вибірки обсягом кілька сотень — тисяча образів.

Висновки до розділу 12

1. Наскрізний обчислювальний експеримент поєднує моделювання деградації, оцінювання параметрів шуму, реставрацію, формування ознак і класифікацію; його результати залежать від усіх етапів системи.
2. Для коректного порівняння варіантів системи необхідно використовувати однакові реалізації тестових даних, незалежні від навчання тестові вибірки, достатній обсяг вибірок і варіювати умови експерименту.
3. Реставрація методом умовної деконволюції з автоматичним визначенням параметрів підвищила точність розпізнавання спотворених цифр в усіх досліджених умовах — до 19,5 процентного пункту для класифікатора за еталоном і до 31,5 пункту для класифікатора на основі ознак HOG.
4. Ознаки, що спираються на високочастотну інформацію, найбільш чутливі до розмиття і шуму; для них попереднє оброблення і реставрація найважливіші.
5. Ефект реставрації зменшується зі зростанням рівня шуму, що відповідає природі некоректної оберненої задачі.
6. Оптимальність реставрації за критерієм відновлення образу не тотожна оптимальності за критерієм розпізнавання; якщо реставрація є етапом системи розпізнавання, її параметри доцільно уточнювати за точністю розпізнавання на валідаційній вибірці, контролюючи стійкість розв’язку.

Питання для обговорення та самоперевірки

1. З яких етапів складається система, досліджувана в цьому розділі? Яку роль відіграє кожен з них?
2. Чому в експерименті рівень шуму оцінюється за спотвореним образом, а не задається?
3. Які принципи слід враховувати при плануванні обчислювального експерименту із системою розпізнавання?
4. Як оцінити статистичну похибку оцінки точності класифікатора?
5. Чому класифікатор на основі HOG виявився більш чутливим до розмиття, ніж класифікатор за еталоном?
6. Чому реставрація особливо ефективна для розмиття рухом?
7. Чому ефект реставрації зменшується з підвищенням рівня шуму?
8. Які обмеження має проведений модельний експеримент? Як їх можна усунути?

Завдання для самостійної роботи

1. Відтворіть експеримент, використовуючи наведені фрагменти програм. Перевірте, наскільки змінюються результати при іншій ініціалізації генератора випадкових чисел.
2. Дослідіть чутливість результатів до похибки задання параметра гаусівського розмиття (± 10 %, ± 25 %) і довжини розмиття рухом (± 1, ± 2 пікселі).

3. Порівняйте умовну деконволюцію з фільтром Вінера зі сталим K (при різних K) і з псевдоінверсним фільтром у тому самому експерименті.
4. Замініть гаусівське розмиття розфокусуванням (3.18) і повторіть експеримент.
5. Додайте до конвеєра етапи порогової сегментації та геометричного нормування (розділ 7) і з'ясуйте, як вони впливають на точність.
6. Використайте замість HOG інші ознаки (інваріанти Ху, дескриптори Фур'є контуру, головні компоненти пікселів) і порівняйте їх стійкість до розмиття і шуму.
7. Навчіть класифікатор за найближчим центром на розмитих образах і порівняйте з результатами реставрації. Зробіть висновок про те, у яких умовах доцільніше реставрувати образи, а в яких — навчати класифікатор на спотворених даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Сушко І. О., Лащевська Н. О., Мовчанюк А. В., Антипенко Р. В. Конспект лекцій до кредитного модуля «Спеціальні розділи оброблення сигналів 2» для магістрів радіотехнічного факультету спеціальності 172 «Телекомунікації та радіотехніка». Київ : КПІ ім. Ігоря Сікорського, 2017. 172 с.

2. Іванюк Н. О. Реставрація образів методом умовної деконволюції в базисі перетворення Адамара : дис. канд. техн. наук : 05.12.17. Київ : Нац. техн. ун-т України «Київський політехнічний інститут», 2015. 146 с.

3. Рибін О. І., Корольов В. Ю. Алгоритм умовної деконволюції зображень у частотній області. Вісті вищих учбових закладів. Радіоелектроніка. 2000. Т. 43, № 4. С. 51–55.

4. Рибін О. І., Іванюк Н. О. Реставрація образів за методом умовної деконволюції в області трансформант перетворення Адамара. Вісник НТУУ «КПІ». Серія Радіотехніка. Радіоапаратобудування. 2011. № 46. С. 51–58. DOI: https://doi.org/10.20535/RADAP.2011.46.51-58.

5. Рибін О. І., Іванюк Н. О. Зв’язок між перетвореннями RTF та Фур’є. Вісник НТУУ «КПІ». Серія Радіотехніка. Радіоапаратобудування. 2012. № 48. С. 23–37. DOI: https://doi.org/10.20535/RADAP.2012.48.23-37.

6. Шебеста В., Іванюк Н. О., Солохненко К. С., Рибін О. І. Реставрація зображень методом умовної деконволюції в базисі перетворення RTF. Вісник НТУУ «КПІ». Серія Радіотехніка. Радіоапаратобудування. 2014. № 56. С. 25–33. DOI: https://doi.org/10.20535/RADAP.2014.56.25-33.

7. Рибін О. І., Іванюк Н. О. Алгоритм побудови матриці деградації образу з використанням дискретного перетворення Адамара. Вісник НТУУ «КПІ». Серія Радіотехніка. Радіоапаратобудування. 2012. № 49. С. 18–32. DOI: https://doi.org/10.20535/RADAP.2012.49.18-32.

8. Рибін О. І., Іванюк Н. О. Формування оберненої матриці деградації образу в області трансформант Адамара. Вісник НТУУ «КПІ». Серія Радіотехніка. Радіоапаратобудування. 2013. № 52. С. 29–37. DOI: https://doi.org/10.20535/RADAP.2013.52.29-37.

9. Іванюк Н. О., Рибін О. І. Реставрація образів методом умовної деконволюції в базисі дискретного перетворення Адамара. Вісті вищих учбових закладів. Радіоелектроніка. 2014. Т. 57, № 3. С. 33–43. DOI: https://doi.org/10.20535/S0021347014030042.

10. Іванюк Н. О. Алгоритми реставрації образів за методом умовної деконволюції. Наукові вісті НТУУ «КПІ». 2014. № 6. С. 7–15.

11. Rybin A. I., Ivanyuk N. A. Signal restoration by conventional deconvolution method using symbolic operand representations in Hadamard basis. Statistical Methods of Signal and Data Processing (SMSDP-2013) : proceedings of the International Conference, Kyiv, 16–17 Oct. 2013. Kyiv : NAU, 2013. P. 84–90.

12. Рибін О. І. Аналіз лінійних кіл у базисі перетворень Волша. Вісті вищих учбових закладів. Радіоелектроніка. 2004. Т. 47, № 5. С. 36–41.

13. Gonzalez R. C., Woods R. E. Digital Image Processing. 4th ed. New York : Pearson, 2018. 1168 p.

14. Pratt W. K. Digital Image Processing: PIKS Scientific Inside. 4th ed. Hoboken : Wiley-Interscience, 2007. 807 p. DOI: https://doi.org/10.1002/0470097434.

15. Burger W., Burge M. J. Digital Image Processing: An Algorithmic Introduction. 3rd ed. Cham : Springer, 2022. DOI: https://doi.org/10.1007/978-3-031-05744-1.

16. Oppenheim A. V., Schafer R. W. Discrete-Time Signal Processing. 3rd ed. Upper Saddle River : Pearson, 2010. 1108 p.

17. Webb A. R., Copsey K. D. Statistical Pattern Recognition. 3rd ed. Chichester : Wiley, 2011. 642 p. DOI: https://doi.org/10.1002/9781119952954.

18. Duda R. O., Hart P. E., Stork D. G. Pattern Classification. 2nd ed. New York : Wiley, 2001. 654 p.

19. Bishop C. M. Pattern Recognition and Machine Learning. New York : Springer, 2006. 738 p. DOI: https://doi.org/10.1007/978-0-387-45528-0.

20. Theodoridis S., Koutroumbas K. Pattern Recognition. 4th ed. Burlington : Academic Press, 2009. 984 p.

21. Murphy K. P. Probabilistic Machine Learning: An Introduction. Cambridge : MIT Press, 2022. 864 p.

22. Britanak V., Yip P. C., Rao K. R. Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations. Amsterdam : Academic Press, 2007. 368 p.

23. Agaian S., Sarukhanyan H., Egiazarian K., Astola J. Hadamard Transforms. Bellingham : SPIE Press, 2011. 520 p.

24. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd ed. Burlington : Academic Press, 2009. 832 p. DOI: https://doi.org/10.1016/B978-0-12-374370-1.X0001-8.

25. Jensen A., la Cour-Harbo A. Ripples in Mathematics: The Discrete Wavelet Transform. Berlin : Springer, 2001. 246 p. DOI: https://doi.org/10.1007/978-3-642-56702-5.

26. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 775 p. DOI: https://doi.org/10.7551/mitpress/10243.001.0001.

27. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Hoboken : Pearson, 2020. 1136 p.

28. Szeliski R. Computer Vision: Algorithms and Applications. 2nd ed. Cham : Springer, 2022. 925 p. DOI: https://doi.org/10.1007/978-3-030-34372-9.

29. Prince S. J. D. Understanding Deep Learning. Cambridge : MIT Press, 2023. 544 p.

30. Bishop C. M., Bishop H. Deep Learning: Foundations and Concepts. Cham : Springer, 2024. 649 p. DOI: https://doi.org/10.1007/978-3-031-45468-4.

31. Жученко А. І., Черепанська І. Ю., Сазонов А. Ю., Ковалюк Д. О. Технології штучного інтелекту та основи машинного зору в автоматизації: теорія і практика : підручник. Київ : КПІ ім. Ігоря Сікорського, 2019. 386 с.

32. Шаховська Н. Б., Камінський Р. М., Вовк О. Б. Системи штучного інтелекту : навч. посіб. Львів : Вид-во Львів. політехніки, 2018. 391 с.

33. Bertero M., Boccacci P., De Mol C. Introduction to Inverse Problems in Imaging. 2nd ed. Boca Raton : CRC Press, 2021.

34. De Maesschalck R., Jouan-Rimbaud D., Massart D. L. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems. 2000. Vol. 50, No. 1. P. 1–18. DOI: https://doi.org/10.1016/S0169-7439(99)00047-7.

35. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York : Springer, 2009. 745 p. DOI: https://doi.org/10.1007/978-0-387-84858-7.

36. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006. Vol. 27, No. 8. P. 861–874. DOI: https://doi.org/10.1016/j.patrec.2005.10.010.

37. Rangayyan R. M. Biomedical Signal Analysis. 2nd ed. Hoboken : Wiley-IEEE Press, 2015. 720 p. DOI: https://doi.org/10.1002/9781119068129.

38. Harris C. R., Millman K. J., van der Walt S. J. et al. Array programming with NumPy. Nature. 2020. Vol. 585, No. 7825. P. 357–362. DOI: https://doi.org/10.1038/s41586-020-2649-2.

39. Virtanen P., Gommers R., Oliphant T. E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020. Vol. 17, No. 3. P. 261–272. DOI: https://doi.org/10.1038/s41592-019-0686-2.

40. van der Walt S., Schönberger J. L., Nunez-Iglesias J. et al. Scikit-image: image processing in Python. PeerJ. 2014. Vol. 2. Art. e453. DOI: https://doi.org/10.7717/peerj.453.

41. Lee G. R., Gommers R., Waselewski F., Wohlfahrt K., O'Leary A. PyWavelets: a Python package for wavelet analysis. Journal of Open Source Software. 2019. Vol. 4, No. 36. Art. 1237. DOI: https://doi.org/10.21105/joss.01237.

42. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011. Vol. 12. P. 2825–2830.

43. Unser M. Sampling – 50 years after Shannon. Proceedings of the IEEE. 2000. Vol. 88, No. 4. P. 569–587. DOI: https://doi.org/10.1109/5.843002.

44. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612. DOI: https://doi.org/10.1109/TIP.2003.819861.

45. Hansen P. C. Discrete Inverse Problems: Insight and Algorithms. Philadelphia : SIAM, 2010. 213 p. DOI: https://doi.org/10.1137/1.9780898718836.

46. Rao K. R., Kim D. N., Hwang J. J. Fast Fourier Transform: Algorithms and Applications. Dordrecht : Springer, 2010. 423 p. DOI: https://doi.org/10.1007/978-1-4020-6629-0.

47. Zitová B., Flusser J. Image registration methods: a survey. Image and Vision Computing. 2003. Vol. 21, No. 11. P. 977–1000. DOI: https://doi.org/10.1016/S0262-8856(03)00137-9.

48. Nixon M. S., Aguado A. S. Feature Extraction and Image Processing for Computer Vision. 4th ed. London : Academic Press, 2020.

49. Jolliffe I. T. Principal Component Analysis. 2nd ed. New York : Springer, 2002. 487 p. DOI: https://doi.org/10.1007/b98835.

50. Gröchenig K. Foundations of Time-Frequency Analysis. Boston : Birkhäuser, 2001. 359 p. DOI: https://doi.org/10.1007/978-1-4612-0003-1.

51. Paris S., Kornprobst P., Tumblin J., Durand F. Bilateral filtering: theory and applications. Foundations and Trends in Computer Graphics and Vision. 2009. Vol. 4, No. 1. P. 1–73. DOI: https://doi.org/10.1561/0600000020.

52. Lowe D. G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 2004. Vol. 60, No. 2. P. 91–110. DOI: https://doi.org/10.1023/B:VISI.0000029664.99615.94.

53. Soille P. Morphological Image Analysis: Principles and Applications. 2nd ed. Berlin : Springer, 2003. 391 p. DOI: https://doi.org/10.1007/978-3-662-05088-0.

54. Briechle K., Hanebeck U. D. Template matching using fast normalized cross correlation. Optical Pattern Recognition XII : Proceedings of SPIE. 2001. Vol. 4387. P. 95–102. DOI: https://doi.org/10.1117/12.421129.

55. Richards M. A. Fundamentals of Radar Signal Processing. 2nd ed. New York : McGraw-Hill Education, 2014.

56. Golub G. H., Van Loan C. F. Matrix Computations. 4th ed. Baltimore : Johns Hopkins University Press, 2013. 756 p.

57. Flusser J., Suk T., Zitová B. 2D and 3D Image Analysis by Moments. Chichester : Wiley, 2016. 560 p. DOI: https://doi.org/10.1002/9781119039402.

58. Ojala T., Pietikäinen M., Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. Vol. 24, No. 7. P. 971–987. DOI: https://doi.org/10.1109/TPAMI.2002.1017623.

59. Dalal N., Triggs B. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005). San Diego, 2005. Vol. 1. P. 886–893. DOI: https://doi.org/10.1109/CVPR.2005.177.

60. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 3rd ed. Sebastopol : O'Reilly Media, 2022.

61. Raschka S., Liu Y., Mirjalili V. Machine Learning with PyTorch and Scikit-Learn. Birmingham : Packt Publishing, 2022.

62. Момот А. С., Мамута М. С. Інтелектуальні алгоритми розпізнавання образів. Практикум : навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2022. 221 с. URL: https://ela.kpi.ua/handle/123456789/47696 (дата звернення: 07.10.2026).

ДОДАТОК А. ПРОГРАМНЕ СЕРЕДОВИЩЕ ТА ЛАБОРАТОРНИЙ ПРАКТИКУМ

А.1. Встановлення програмного середовища

Для виконання прикладів і лабораторних робіт рекомендовано використовувати Python версії 3.10 або новішої [38–42]. Найзручніше працювати в інтерактивному середовищі Jupyter (Jupyter Notebook або JupyterLab) чи в інтегрованому середовищі розробки з підтримкою блокнотів (наприклад, Visual Studio Code). Рекомендується створювати для курсу окреме віртуальне середовище, щоб версії бібліотек не конфліктували з іншими проєктами.

Як альтернативу локальному встановленню можна використовувати хмарне середовище Google Colab: більшість зазначених бібліотек у ньому встановлено заздалегідь, відсутні встановлюють командою !pip install безпосередньо в блокноті. Саме в такому форматі (Jupyter Notebook або Google Colab) виконуються лабораторні роботи, описані в підрозділах А.2–А.4.

Послідовність дій:

1. Встановити Python з офіційного сайту або дистрибутив Anaconda/Miniconda.
2. Створити і активувати віртуальне середовище:

python -m venv trokurs
trokurs\Scripts\activate (Windows)
source trokurs/bin/activate (Linux, macOS)

3. Встановити необхідні бібліотеки:

pip install numpy scipy pandas matplotlib seaborn scikit-image scikit-learn PyWavelets jupyter

4. Перевірити встановлення:

```
import numpy, scipy, pandas, matplotlib, skimage, sklearn, pywt
for m in (numpy, scipy, pandas, matplotlib, skimage, sklearn, pywt):
    print(m.__name__, m.__version__)
```

Тестові зображення, використані в посібнику (camera, coins, moon, astronaut, brick, grass, gravel, page, horse тощо), входять до складу бібліотеки scikit-image і доступні через модуль skimage.data без підключення до мережі [40]; набір рукописних цифр — через функцію sklearn.datasets.load_digits [42].

А.2. Загальні вимоги до виконання і оформлення лабораторних робіт

Лабораторний практикум побудовано як наскрізний командний проєкт: кожна лабораторна робота є черговим етапом створення однієї системи розпізнавання (класифікації), яку команда розробляє протягом семестру і захищає на завершальному занятті. Команда складається з 3–4 здобувачів; тему проєкту і набір даних погоджують з викладачем на вступному занятті (лабораторна робота № 0). Перелік робіт і їх зв'язок з розділами посібника наведено в табл. А.1.

Під час виконання робіт слід дотримуватися таких правил:

– роботи виконують у середовищі Jupyter Notebook або Google Colab мовою Python з використанням бібліотек pandas, NumPy, Matplotlib/seaborn і scikit-learn [42, 60, 61];
– усі роботи, починаючи з № 3, виконують на даних командного проєкту, підготовлених у лабораторних роботах № 1–2; блокнот і очищений набір даних зберігають у спільному репозиторії команди (GitHub/GitLab);
– для коректного порівняння методів у всіх роботах використовують один і той самий стратифікований поділ на навчальну і тестову вибірки з фіксованим значенням random_state;
– параметри масштабування, імпутації, векторизації та PCA обчислюють (fit) лише на навчальній вибірці, до тестової застосовують тільки transform, щоб уникнути витоку даних (data leakage);
– результати кожної роботи заносять до зведеної таблиці команди, яка використовується в роботах № 7, 11 і 12.

Звіт (блокнот) про виконання роботи має містити:

1) титульний аркуш (назва і тема роботи, назва команди, склад, академічна група);
2) постановку задачі і стислий опис даних проєкту;
3) програмний код з коментарями для кожної частини роботи;
4) результати у вигляді таблиць і графіків з підписами;
5) порівняння з результатами попередніх робіт (там, де це передбачено завданням);
6) аргументовані висновки і відповіді на контрольні запитання.

Кожну роботу оцінюють за 100-бальною шкалою: бали розподіляються між частинами роботи відповідно до їх складності, 15 балів відводиться на оформлення звіту і захист. Отриману оцінку перераховують у бали, визначені для цієї роботи робочою програмою навчальної дисципліни, пропорційно до максимальної кількості балів: наприклад, якщо на роботу відведено 15 балів, оцінка 80 балів зі 100 відповідає 12 балам рейтингу. Звіт подають до початку наступної лабораторної роботи, якщо інший термін не встановлено викладачем. Підсумкову оцінку за командний проєкт виставляють за результатами захисту (лабораторна робота № 12) з урахуванням індивідуального внеску кожного учасника.

Таблиця А.1

Структура лабораторного практикуму

№	Тема лабораторної роботи	Розділи посібника
0	Вступ до курсу та формування командного AI-проєкту	1.1–1.5, 1.9, 1.10
1	Попередній аналіз та підготовка даних для систем розпізнавання образів	1.2, 1.7, 2.6, 11.5
2	Попередня обробка даних та формування простору ознак	11.1, 11.8, 11.10
3	Класифікатор k найближчих сусідів (k-NN)	1.6, 1.7, 11.8; ч. 2
4	Наївний байєсівський класифікатор	1.5.2; ч. 2
5	Класифікатор на основі дерева рішень	1.7, 11.10.1; ч. 2
6	Класифікатор на основі логістичної регресії	1.5.2, 11.8; ч. 2
7	Комплексна оцінка та порівняння класифікаторів	1.7, 1.8, 12.3
8	Зниження розмірності методом головних компонент (PCA)	5.7, 11.9.1
9	Класифікатор випадкового лісу	11.10; ч. 2
10	Класифікатори на основі бустингу	ч. 2
11	Класифікатор на основі багатошарового персептрона (MLP)	1.5.4; ч. 2
12	Фінальний захист командного проєкту	розд. 12, 1.8

Примітка. «ч. 2» — методи класифікації, що детально розглядаються у другій частині посібника; до її виходу використовують опорні конспекти лекцій і рекомендовані джерела [60–62].

А.3. Перелік лабораторних робіт

Лабораторна робота № 0. Вступ до курсу та формування командного AI-проєкту

Мета: ознайомитися зі структурою курсу, основними поняттями розпізнавання образів і робочим інструментарієм; сформувати команду, обрати тему наскрізного проєкту, сформулювати задачу розпізнавання і знайти набір даних.

Теоретичні відомості: розділ 1 (підрозділи 1.1–1.5, 1.9, 1.10).

Порядок виконання:

1. Сформувати команду з 3–4 здобувачів, визначити назву команди і попередній розподіл ролей (дані, програмна реалізація, моделі, аналіз результатів, захист).
2. Обрати предметну область і тему проєкту (аграрні технології та Smart Farming, біоресурси, екологічний моніторинг, IoT, БПЛА і комп'ютерний зір, Smart Campus тощо) і погодити її з викладачем.
3. Сформулювати задачу розпізнавання: об'єкт, вхідні дані X, ознаки, цільову змінну (класи) Y, тип задачі і очікуваний результат.
4. Знайти і описати набір даних (джерело, тип, обсяг, кількість ознак і класів, наявність розмітки).
5. Заповнити паспорт командного проєкту і перевірити роботу середовища Python/Colab (додаток А.1).

Контрольні запитання: 1) Що таке образ, клас і ознака в задачі вашої команди? 2) До якого типу належить ваша задача (класифікація, кластеризація, бінарна чи багатокласова)? 3) Чому обраний набір даних придатний для навчання класифікатора?

Лабораторна робота № 1. Попередній аналіз та підготовка даних для систем розпізнавання образів

Мета: набути навичок розвідувального аналізу (EDA) і підготовки даних командного проєкту до застосування методів розпізнавання образів.

Теоретичні відомості: підрозділи 1.2, 1.7, 2.6, 11.5.

Порядок виконання:

1. Завантажити набір даних, визначити кількість об'єктів, ознак, їх типи і цільову змінну.
2. Обчислити описові статистики числових ознак і частоти значень категоріальних.
3. Виявити пропуски, дублікати, потенційні викиди; обґрунтувати спосіб їх обробки.
4. Проаналізувати розподіл класів і оцінити ризики дисбалансу.
5. Побудувати 3–5 найінформативніших візуалізацій (розподіли ознак, кореляційна матриця).
6. Сформувати X і y, виконати стратифікований поділ train/test (20 %) з фіксованим random_state і масштабувати ознаки без витоку даних.

Контрольні запитання: 1) Чим небезпечний дисбаланс класів для оцінювання якості? 2) Чому scaler навчають лише на навчальній вибірці? 3) Які проблеми якості даних виявлено у вашому наборі і як їх усунено?

Лабораторна робота № 2. Попередня обробка даних та формування простору ознак

Мета: набути навичок попередньої обробки даних (пропуски, викиди, кодування, нормалізація, відбір ознак) і оцінити її вплив на якість базового класифікатора.

Теоретичні відомості: підрозділи 11.1, 11.8, 11.10.

Порядок виконання:

1. Визначити шкалу вимірювання кожної ознаки.
2. Обробити пропущені значення щонайменше двома способами і порівняти результати.
3. Виявити викиди за правилом IQR (або z-оцінкою) щонайменше для двох числових ознак.
4. Закодувати категоріальні ознаки (Label Encoding або One-Hot Encoding).
5. Виконати нормалізацію або стандартизацію ознак і відбір ознак на основі аналізу кореляції.
6. Навчити базовий класифікатор на «сирих» і оброблених даних, порівняти accuracy, precision, recall, F1.

Контрольні запитання: 1) Чим відрізняються Min-Max нормалізація і z-стандартизація? 2) Коли доцільно застосовувати One-Hot Encoding? 3) Що таке data leakage і як ви його уникнули?

Лабораторна робота № 3. Класифікатор k найближчих сусідів (k-NN)

Мета: набути навичок побудови метричного класифікатора k-NN, дослідити вплив метрики відстані і параметра k на якість класифікації.

Теоретичні відомості: підрозділи 1.6, 1.7, 11.8.

Порядок виконання:

1. Побудувати базовий KNeighborsClassifier (k = 5, евклідова метрика) і обчислити метрики якості.
2. Методом 5-кратної крос-валідації дослідити залежність точності від k (1–25), обрати оптимальне k.
3. Порівняти евклідову, манхеттенську метрики і метрику Мінковського (p = 3), а також зважений k-NN.
4. Дослідити вплив кількості ознак на якість і час передбачення (прокляття розмірності).

Контрольні запитання: 1) Чому для k-NN обов'язкова стандартизація ознак? 2) Як проявляються недо- і перенавчання залежно від k? 3) Чому k-NN називають «лінивим» методом навчання?

Лабораторна робота № 4. Наївний байєсівський класифікатор

Мета: набути навичок побудови наївного байєсівського класифікатора для числових і текстових ознак, дослідити вплив згладжування Лапласа.

Теоретичні відомості: підрозділ 1.5.2.

Порядок виконання:

1. Побудувати GaussianNB на даних проєкту (той самий поділ train/test, що й у роботі № 3), обчислити метрики і час навчання/передбачення.
2. Побудувати MultinomialNB для класифікації текстових повідомлень (спам-фільтр) з векторизацією CountVectorizer або TfidfVectorizer.
3. Дослідити вплив параметра alpha (0; 0,01; 1,0) на якість текстового класифікатора.
4. Порівняти GaussianNB, MultinomialNB і k-NN за якістю і швидкодією.

Контрольні запитання: 1) У чому полягає «наївне» припущення? 2) Як згладжування Лапласа розв'язує проблему нульової ймовірності? 3) Чи виправдане припущення про незалежність ознак для даних вашого проєкту?

Лабораторна робота № 5. Класифікатор на основі дерева рішень

Мета: набути навичок побудови дерева рішень, дослідити перенавчання і підбір гіперпараметрів, оцінити важливість ознак.

Теоретичні відомості: підрозділи 1.7, 11.10.1.

Порядок виконання:

1. Побудувати DecisionTreeClassifier без обмежень і порівняти точність на train і test.

2. Підібрати max_depth і min_samples_leaf за допомогою GridSearchCV (5-кратна крос-валідація).
3. Візуалізувати дерево (plot_tree) і побудувати діаграму важливості ознак.
4. Порівняти дерево рішень з k-NN і GaussianNB за точністю і інтерпретованістю.

Контрольні запитання: 1) Як обчислюються ентропія та індекс Джині? 2) Як ви виявили перенавчання дерева? 3) Які ознаки виявилися найважливішими і чи узгоджується це з відбором ознак у роботі № 2?

Лабораторна робота № 6. Класифікатор на основі логістичної регресії

Мета: набути навичок побудови логістичної регресії, дослідити вплив L1- і L2-регуляризації, інтерпретувати коефіцієнти моделі.

Теоретичні відомості: підрозділи 1.5.2, 11.8.

Порядок виконання:

1. Побудувати LogisticRegression на стандартизованих даних і обчислити метрики якості.
2. Дослідити вплив параметра C і типу штрафу (L1/L2) на якість і кількість ненульових коефіцієнтів.
3. Визначити три найвпливовіші ознаки, обчислити для них відношення шансів.
4. Порівняти логістичну регресію з деревом рішень, k-NN і GaussianNB.

Контрольні запитання: 1) Навіщо в логістичній регресії сигмоїдна функція? 2) Чим L1-регуляризація відрізняється від L2? 3) Що таке відношення шансів і як його обчислити?

Лабораторна робота № 7. Комплексна оцінка та порівняння класифікаторів

Мета: набути навичок повного оцінювання якості класифікаторів (матриця помилок, precision, recall, F1, ROC-AUC) і застосування крос-валідації для обґрунтованого вибору моделі.

Теоретичні відомості: підрозділи 1.7, 1.8, 12.3.

Порядок виконання:

1. Для моделей з робіт № 3–6 обчислити матриці помилок і classification_report.
2. Побудувати ROC-криві всіх моделей на одному графіку і порівняти AUC.
3. Виконати стратифіковану 5-кратну крос-валідацію, навести середнє і стандартне відхилення метрики.
4. Обрати найкращу модель з урахуванням точності, стабільності та інтерпретованості.

Контрольні запитання: 1) Чим precision відрізняється від recall і яка метрика важливіша для вашої задачі? 2) Як інтерпретувати ROC-AUC? 3) Чому порівнювати моделі краще за крос-валідацією, а не за одним поділом даних?

Лабораторна робота № 8. Зниження розмірності методом головних компонент (PCA)

Мета: набути навичок застосування PCA для зниження розмірності і візуалізації даних, оцінити вплив зниження розмірності на якість і швидкодію класифікатора.

Теоретичні відомості: підрозділи 5.7, 11.9.1.

Порядок виконання:

1. Застосувати PCA до стандартизованих даних і побудувати графік кумулятивної поясненої дисперсії.
2. Визначити мінімальну кількість компонент, що пояснює не менше 90–95 % дисперсії.
3. Візуалізувати дані у просторі перших двох головних компонент з розфарбуванням за класами.
4. Навчити найкращий класифікатор (за результатами роботи № 7) на повному і зменшеному наборах ознак, порівняти якість і час навчання.

Контрольні запитання: 1) Чим PCA відрізняється від відбору ознак? 2) Як за поясненою дисперсією обирають кількість компонент? 3) Чому PCA навчають лише на навчальній вибірці?

Лабораторна робота № 9. Класифікатор випадкового лісу

Мета: набути навичок побудови випадкового лісу, дослідити OOB-оцінку, вплив кількості дерев і важливість ознак.

Теоретичні відомості: підрозділ 11.10.

Порядок виконання:

1. Побудувати RandomForestClassifier (n_estimators = 200, oob_score = True) і порівняти OOB-оцінку з точністю на тесті.
2. Дослідити залежність якості і часу навчання від кількості дерев (10–500).
3. Порівняти важливість ознак лісу і одиночного дерева (робота № 5).
4. Порівняти стійкість до перенавчання лісу і дерева рішень.

Контрольні запитання: 1) Як формується бутстреп-вибірка? 2) Що таке OOB-помилка? 3) Чому збільшення кількості дерев майже не призводить до перенавчання?

Лабораторна робота № 10. Класифікатори на основі бустингу

Мета: набути навичок побудови AdaBoost і градієнтного бустингу, дослідити вплив темпу навчання і кількості ітерацій, порівняти стійкість до шуму з випадковим лісом.

Теоретичні відомості: друга частина посібника.

Порядок виконання:

1. Побудувати AdaBoostClassifier і порівняти його з випадковим лісом (робота № 9).
2. Для GradientBoostingClassifier дослідити вплив learning_rate і n_estimators, побудувати криві train/test.

3. Змінити мітки 5–10 % навчальних об'єктів і порівняти падіння точності AdaBoost, градієнтного бустингу і випадкового лісу.

Контрольні запитання: 1) Як обчислюється вага слабкого класифікатора в AdaBoost? 2) У чому ідея градієнтного бустингу? 3) Чим бустинг відрізняється від бегінгу за чутливістю до шуму?

Лабораторна робота № 11. Класифікатор на основі багатошарового персептрона (MLP)

Мета: набути навичок побудови нейромережевого класифікатора, дослідити вплив архітектури і кількості епох на перенавчання.

Теоретичні відомості: підрозділ 1.5.4.

Порядок виконання:

1. Побудувати MLPClassifier (один прихований шар з 10 нейронами) на стандартизованих даних.
2. Порівняти архітектури (5,), (10,), (20,), (10, 10) і побудувати криву навчання для найкращої.
3. Дослідити вплив max_iter на співвідношення точності на train і validation.
4. Звести результати всіх класифікаторів курсу в підсумкову таблицю.

Контрольні запитання: 1) Як обчислюється вихід одного нейрона? 2) Як за кривою навчання виявити перенавчання? 3) Які компроміси (точність, швидкість, інтерпретованість) вплинули на вибір методу?

Лабораторна робота № 12. Фінальний захист командного проєкту

Мета: узагальнити повний цикл роботи над проєктом — від постановки задачі до обґрунтованого вибору фінальної моделі — і представити результати у формі звіту та усної презентації.

Теоретичні відомості: розділ 12, підрозділ 1.8.

Порядок виконання:

1. Зібрати результати робіт № 1–11 у підсумковий звіт: постановка задачі і дані; підготовка даних; побудовані класифікатори; комплексне порівняння і вибір моделі; висновки.
2. Побудувати підсумкову порівняльну таблицю семи класифікаторів за єдиною методикою оцінювання.
3. Обґрунтувати вибір фінальної моделі з урахуванням точності, інтерпретованості, швидкодії і стійкості до шуму.
4. Підготувати презентацію (5–7 хв) і відповісти на запитання (до 3 хв).

Контрольні запитання: 1) Чому саме ця модель обрана фінальною? Які альтернативи розглядалися? 2) Чому обрана методика порівняння моделей є коректною? 3) Які обмеження має ваше рішення і як його можна вдосконалити?

А.4. Шаблон блокнота лабораторної роботи

Нижче наведено шаблон, з якого зручно починати роботи № 3–11. Він забезпечує однаковий поділ даних, захист від витоку даних (масштабування всередині конвеєра), обчислення основних метрик і запис результатів до зведеної

таблиці команди [42, 60–62]. Для кожної роботи замінюють лише блок 2 відповідним класифікатором і додають дослідження, передбачені завданням.

```
import os, time
import pandas as pd
from sklearn.model_selection import (train_test_split, StratifiedKFold,
                                     cross_val_score)
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline
from sklearn.metrics import (classification_report, confusion_matrix,
                             f1_score)
from sklearn.neighbors import KNeighborsClassifier

RANDOM_STATE = 42              # однакове для всіх робіт курсу

# 1. Дані командного проєкту, підготовлені в роботах № 1-2
df = pd.read_csv("data/project_clean.csv")
X, y = df.drop(columns="target"), df["target"]
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=RANDOM_STATE)

# 2. Модель поточної роботи (приклад: k-NN, робота № 3)
model = make_pipeline(StandardScaler(),
                      KNeighborsClassifier(n_neighbors=5))

t0 = time.perf_counter()
model.fit(X_train, y_train)
t_fit = time.perf_counter() - t0
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))

# 3. Крос-валідація для порівняння з попередніми роботами
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=RANDOM_STATE)
cv_f1 = cross_val_score(model, X_train, y_train, cv=cv, scoring="f1_macro")
print(f"F1 (5-fold): {cv_f1.mean():.3f} ± {cv_f1.std():.3f}")

# 4. Запис до зведеної таблиці команди (роботи № 7, 11, 12)
row = pd.DataFrame([{"model": "k-NN",
                     "f1_test": f1_score(y_test, y_pred, average="macro"),
                     "f1_cv": cv_f1.mean(), "t_fit_s": t_fit}])
path = "results/summary.csv"
row.to_csv(path, mode="a", index=False, header=not os.path.exists(path))
```

ДОДАТОК Б. ТЕСТОВІ ЗАВДАННЯ ДЛЯ САМОПЕРЕВІРКИ

У кожному завданні оберіть одну правильну відповідь. Ключ відповідей наведено наприкінці додатка.

Б.1. До розділу 1

1.1. Образом у теорії розпізнавання називають:

а) мітку класу; б) вирішальне правило; в) формалізований опис об’єкта, придатний для оброблення технічною системою; г) фізичний об’єкт, що підлягає класифікації.

1.2. Задача, у якій мітки класів невідомі і потрібно розбити образи на групи схожих, називається:

а) класифікацією; б) детекцією; в) кластеризацією; г) верифікацією.

1.3. Відстань Махаланобіса, на відміну від евклідової, враховує:

а) лише середні значення ознак; б) дисперсії ознак і кореляцію між ними; в) кількість класів; г) апріорні ймовірності класів.

1.4. Косинусна міра подібності інваріантна до:

а) додавання шуму; б) перестановки компонент; в) зсуву рівня (додавання сталої); г) зміни масштабу вектора.

1.5. Помилка першого роду в системі біометричної верифікації (за прийнятою в посібнику термінологією) — це:

а) неправильне визначення класу при ідентифікації; б) збій сенсора; в) відмова справжньому користувачу; г) допуск зловмисника.

1.6. Площа під ROC-кривою випадкового вгадування дорівнює:

а) 1; б) 0,5; в) 0; г) залежить від порогу.

Б.2. До розділу 2

2.1. Сигнал зі спектром, обмеженим частотою 4 кГц, без втрат відновлюється за відліками, якщо частота дискретизації не менша за:

а) 16 кГц; б) 4 кГц; в) 2 кГц; г) 8 кГц.

2.2. Гармоніка 9 Гц, дискретизована з частотою 10 Гц, дає ті самі відліки, що й гармоніка частотою:

а) 19 Гц; б) 0,9 Гц; в) 1 Гц; г) 5 Гц.

2.3. Дисперсія похибки рівномірного квантування з кроком q дорівнює:

а) $q/2$; б) q^2; в) $q^2/12$; г) $q^2/4$.

2.4. Кожний додатковий розряд квантувача покращує відношення сигнал/шум квантування приблизно на:

а) 20 дБ; б) 6 дБ; в) 3 дБ; г) 10 дБ.

2.5. Тон (Hue) у моделі HSV слабко залежить від:

а) насиченості; б) кольору об’єкта; в) освітленості; г) довжини хвилі.

2.6. Ентропія гістограми максимальна, якщо:

а) зображення бінарне; б) гістограма двомодальна; в) усі пікселі однакові; г) розподіл яскравостей рівномірний.

Б.3. До розділу 3

3.1. Ізопланарною називають систему, у якої:

а) імпульсна характеристика не залежить від положення точкового джерела; б) передавальна характеристика нелінійна; в) відсутній шум; г) імпульсна характеристика є дельта-функцією.

3.2. Циклічній згортці відповідає матриця:

а) трикутна; б) циркулянтна; в) теплицева; г) діагональна.

3.3. Модель атмосферної турбулентності містить показник степеня:

а) 3; б) 5/6; в) 1/2; г) 2.

3.4. Для робастної оцінки СКВ шуму за вейвлет-коефіцієнтами медіану модулів коефіцієнтів субсмуги HH ділять на:

а) 0,5; б) 0,6745; в) 1,4826; г) $\sqrt{2}$.

3.5. Пряме обернення оператора розмиття непридатне через:

а) недостатню швидкодію; б) підсилення шуму діленням на малі сингулярні числа; в) неможливість обчислити обернену матрицю; г) ефект Гіббса.

3.6. Індекс SSIM поєднує порівняння:

а) яскравості, контрасту і структури; б) гістограм; в) спектрів Фур’є; г) лише середніх значень.

Б.4. До розділу 4

4.1. Для унітарної матриці W обернена матриця дорівнює:

а) $-\mathrm{W}$; б) W^{H}; в) W; г) W^{T}.

4.2. Власними векторами будь-якої циркулянтної матриці є:

а) функції Хаара; б) функції Уолша; в) дискретні гармоніки; г) власні вектори кореляційної матриці.

4.3. Кількість комплексних множень у ШПФ з основою 2 для $N = 1024$ становить:

а) 5120; б) 10 240; в) 1024; г) 1 048 576.

4.4. Викид ефекту Гіббса прямує до приблизно:

а) 1 % стрибка; б) 9 % стрибка; в) 25 % стрибка; г) нуля зі збільшенням кількості гармонік.

4.5. Амплітудний спектр ДПФ інваріантний до:

а) повороту; б) зміни масштабу; в) додавання шуму; г) циклічного зсуву.

4.6. Метод фазової кореляції використовують для:

а) придушення шуму; б) визначення зсуву між зображеннями; в) сегментації; г) стиснення.

Б.5. До розділу 5

5.1. Множення на матрицю Адамара потребує:

а) обчислення косинусів; б) лише додавань і віднімань; в) комплексних множень; г) обернення матриці.

5.2. Секвента функції Уолша — це:

а) енергія функції; б) номер функції в природному впорядкуванні; в) половина середньої кількості змін знака на інтервалі; г) частота дискретизації.

5.3. Циркулянтна матриця згортки в базисі Адамара (природне впорядкування) має структуру:

а) трикутну; б) блочно-діагональну з блоками $1,1,2,4,\ldots,N/2$; в) щільну без регулярності; г) діагональну.

5.4. Найвищу енергетичну компактність для заданого класу образів має перетворення:

а) Карунена — Лоева; б) Фур'є; в) Уолша; г) Хаара.

5.5. Власні числа різницевого оператора «назад» дорівнюють:

а) $j2\pi k/N$; б) $1-e^{-j2\pi k/N}$; в) $(-1)^k$; г) $\cos(2\pi k/N)$.

5.6. Відліки ДКП-II беруться зі зсувом на половину кроку для того, щоб:

а) прискорити обчислення; б) забезпечити ортогональність трансформант; в) зменшити розмір матриці; г) отримати комплексні коефіцієнти.

Б.6. До розділу 6

6.1. Нерівність Гейзенберга — Габора стверджує, що:

а) $\Delta t+\Delta\omega\geq 1$; б) $\Delta t\,\Delta\omega\geq 1/2$; в) $\Delta t=\Delta\omega$; г) $\Delta t\,\Delta\omega\leq 1$.

6.2. Вейвлет-перетворення еквівалентне банку фільтрів зі сталою:

а) шириною смуги; б) затримкою; в) добротністю; г) центральною частотою.

6.3. Алгоритм Малла має обчислювальну складність порядку:

а) N^3; б) N; в) $N\log N$; г) N^2.

6.4. Універсальний поріг Донохо — Джонстона дорівнює:

а) $\sigma\sqrt{2\ln N}$; б) 3σ; в) σ; г) $\sigma\ln N$.

6.5. Субсмуга HH двовимірного ДВП містить переважно:

а) вертикальні деталі; б) горизонтальні деталі; в) наближення; г) діагональні деталі.

6.6. Фільтр Габора — це:

а) медіанний фільтр; б) прямокутне вікно; в) гармонічна хвиля, обмежена гаусівським вікном; г) маска Собеля.

Б.7. До розділу 7

7.1. Вирівнювання гістограми використовує як функцію перетворення:

а) логарифм; б) лінійну функцію; в) гамма-функцію; г) кумулятивну функцію розподілу.

7.2. Найефективнішим для придушення імпульсного шуму є фільтр:

а) медіанний; б) усереднювальний; в) лапласіан; г) гаусівський.

7.3. Маска $[-1,0,1;-2,0,2;-1,0,1]$ відповідає оператору:

а) Собеля; б) Превітт; в) Лапласа; г) Робертса.

7.4. Метод Оцу обирає поріг, що максимізує:

а) міжкласову дисперсію; б) контраст; в) кількість пікселів об'єкта; г) ентропію.

7.5. Операція відкриття — це:

а) різниця дилатації та ерозії; б) дві ерозії; в) ерозія, потім дилатація; г) дилатація, потім ерозія.

7.6. Інтерполяція найближчого сусіда доцільна для:

а) фотографій; б) зображень міток (сегментації); в) спектрів; г) медичних зображень.

Б.8. До розділу 8

8.1. Метод умовної деконволюції потребує апріорної інформації про шум у вигляді:

а) повної густини ймовірності; б) кількох реалізацій шуму; в) спектра потужності шуму; г) енергії шуму.

8.2. Критерій гладкості в методі умовної деконволюції використовує:

а) другі різниці (оператор $[1, -2, 1]$); б) перші різниці; в) гістограму; г) спектр Фур'є образу.

8.3. Параметр $\boldsymbol{\lambda}$ визначають з умови:

а) максимуму гладкості; б) мінімуму PSNR; в) $\lambda = 1$; г) рівності енергії нев'язки енергії шуму.

8.4. Фільтр Вінера для частот з низьким відношенням сигнал/шум:

а) не змінює спектр; б) «закривається» (коефіцієнт передачі прямує до нуля); в) працює як інверсний; г) підсилює сигнал.

8.5. Математично метод умовної деконволюції еквівалентний:

а) методу головних компонент; б) вирівнюванню гістограми; в) медіанній фільтрації; г) регуляризації Тихонова з вибором параметра за принципом нев'язки.

8.6. У натуральних координатах реставрація двовимірного образу 1024×1024 потребує обернення матриці розміром порядку:

а) $10^9 \times 10^9$; б) $10^3 \times 10^3$; в) $10^6 \times 10^6$; г) 10×10.

Б.9. До розділу 9

9.1. У перетворенні RTF роль частот відіграють:

а) секвенти; б) $j\omega_k$; в) масштаби вейвлета; г) власні числа різницевого оператора β_k.

9.2. Перехід від перетворення Фур'є до RTF еквівалентний:

а) зсуву в часі; б) підсиленню всіх частот; в) фільтрації нижніх частот; г) фільтрації верхніх частот.

9.3. Елемент $g_{H1,11}$ матриці G_{1Had} дорівнює:

а) g_0; б) нулю; в) знакозмінній сумі відліків; г) сумі всіх відліків імпульсної характеристики.

9.4. Символьний метод формування матриць дозволяє:

а) отримати діагональну матрицю; б) уникнути перетворення Адамара; в) не знати імпульсної характеристики; г) обчислювати лише ненульові елементи без накопичення похибок.

9.5. Розв'язок задачі умовної деконволюції з однаковими циркулянтними моделями в різних ортонормованих базисах:

а) завжди різний; б) математично однаковий; в) не існує; г) визначений лише для базису Фур'є.

9.6. Блок порядку M матриці G_{1Had} має структуру:

а) трикутну; б) $\begin{bmatrix} P & Q \\ -Q & P \end{bmatrix}$; в) діагональну; г) циркулянтну.

Б.10. До розділу 10

10.1. Імпульсна характеристика погодженого фільтра:

а) є дельта-функцією; б) дорівнює спектру сигналу; в) є дзеркальною копією сигналу; г) гаусівська.

10.2. Максимальне відношення сигнал/шум на виході погодженого фільтра для білого шуму дорівнює:

а) залежить від форми сигналу; б) E/N_0; в) N_0/E; г) $2E/N_0$.

10.3. Нормована взаємна кореляція інваріантна до:

а) повороту; б) змін контрасту і яскравості; в) масштабу; г) перекриття.

10.4. Нормальне перетворення — це ортогональне перетворення, у якого:

а) перша трансформанта збігається з еталоном; б) усі базисні функції гармонічні; в) матриця діагональна; г) ядро комплексне.

10.5. Коефіцієнт трансформант нормального перетворення дорівнює:

а) нулю для будь-якого образу; б) тангенсу кута між образом і еталоном; в) енергії образу; г) косинусу кута між образом і еталоном.

10.6. Порогове значення коефіцієнта трансформант в алгоритмі аутентифікації за підписом визначається як:

а) максимальне на навчальній вибірці; б) мінімальне на тестовій вибірці; в) одиниця; г) середнє на навчальній вибірці.

Б.11. До розділу 11

11.1. Компактність $4\pi A/P^2$ дорівнює одиниці для:

а) круга; б) зірки; в) еліпса; г) квадрата.

11.2. Інваріант Ху, що змінює знак при віддзеркаленні:

а) ϕ_2; б) ϕ_1; в) ϕ_4; г) ϕ_7.

11.3. LBP інваріантний до:

а) розмиття; б) будь-яких монотонних змін яскравості; в) масштабу; г) шуму.

11.4. Метод головних компонент:

а) є нелінійним; б) максимізує розділення класів; в) потребує міток класів; г) максимізує дисперсію даних (метод без учителя).

11.5. Лінійний дискримінантний аналіз для K класів дає не більше:

а) одного напрямку; б) n напрямків; в) K напрямків; г) $K-1$ напрямків.

11.6. Статистики для нормування ознак слід обчислювати:

а) за тестовою вибіркою; б) за всіма даними разом; в) за навчальною вибіркою; г) не обчислювати.

Б.12. До розділу 12

12.1. Для коректного порівняння варіантів системи розпізнавання їх слід випробовувати:

а) на навчальній вибірці; б) на одних і тих самих тестових реалізаціях; в) на різних випадкових вибірках; г) лише на одному образі.

12.2. Стандартна похибка оцінки точності $p = 0{,}9$ за 200 випробуваннями становить приблизно:

а) 20 %; б) 2 %; в) 0,2 %; г) 10 %.

12.3. Найбільше від реставрації виграв класифікатор на основі:

а) нормованої кореляції; б) обох однаково; в) жодного; г) ознак HOG.

12.4. Зі збільшенням рівня шуму ефект реставрації:

а) зростає; б) не змінюється; в) зменшується; г) стає від’ємним завжди.

12.5. Енергію шуму в експерименті розділу 12 оцінювали:

а) не оцінювали; б) за спотвореним образом робастним методом; в) за істинним значенням σ; г) за тестовою вибіркою.

12.6. Розмиття рухом найбільше погіршує розпізнавання тому, що:

а) додає шум; б) зменшує яскравість; в) анізотропно змінює форму символу; г) змінює розмір зображення.

Б.13. Ключ відповідей

Розділ	1	2	3	4	5	6
1	в	в	б	г	в	б
2	г	в	в	б	в	г
3	а	б	б	б	б	а
4	б	в	а	б	г	б
5	б	в	б	а	б	б
6	б	в	б	а	г	в
7	г	а	а	а	в	б
8	г	а	г	б	г	в
9	г	в	в	г	б	б
10	в	г	б	а	б	а
11	а	г	б	г	г	в
12	б	б	г	в	б	в

ДОДАТОК В. УКРАЇНСЬКО-АНГЛІЙСЬКИЙ СЛОВНИК ОСНОВНИХ ТЕРМІНІВ

Таблиця В.1

Основні терміни

Український термін	Англійський відповідник	Розділ
Автокореляційна функція	autocorrelation function	2, 3
Адаптивний поріг	adaptive threshold	7
Аліасинг (накладання спектрів)	aliasing	2
Аутентифікація (верифікація)	authentication, verification	1, 10
Білий шум	white noise	3
Білатеральний фільтр	bilateral filter	7
Вейвлет-перетворення	wavelet transform	6
Вектор ознак	feature vector	1
Виявлення (детекція)	detection	1, 10
Вирівнювання гістограми	histogram equalization	7
Вирішальне правило	decision rule	1
Відбір ознак	feature selection	11
Відстань Махаланобіса	Mahalanobis distance	1
Віконне перетворення Фур'є	short-time Fourier transform	6
Гістограма орієнтованих градієнтів	histogram of oriented gradients	11
Головні компоненти	principal components	5, 11
Гострота	sharpness (inverse of transform coefficient)	10
Двовимірне ДПФ	two-dimensional DFT	4
Деградація образу	image degradation	3
Деконволюція	deconvolution	8
Дескриптор Фур'є	Fourier descriptor	4, 11
Детектор Канні	Canny edge detector	7
Дилатація	dilation	7
Дискретизація	sampling	2
Дискретне косинусне перетворення	discrete cosine transform	5
Дискретне перетворення Фур'є	discrete Fourier transform	4
Дискримінантне число	discriminant value	10
Ефект Гіббса	Gibbs phenomenon	4
Енергетична компактність	energy compaction	5
Енергія шуму	noise energy	3, 8
Ергодичність	ergodicity	3
Ерозія	erosion	7
Еталон (шаблон)	template	1, 10
Закриття (морфологічне)	closing	7
Згортка	convolution	3
Зменшення розмірності	dimensionality reduction	11
Ідентифікація	identification	1
Ізопланарна система	shift-invariant (isoplanatic) system	3
Імпульсна характеристика	impulse response, point spread function	3
Інваріанти Ху	Hu moment invariants	11
Інверсна фільтрація	inverse filtering	8
Інтерполяція	interpolation	7
Клас	class	1
Класифікація	classification	1

Український термін	Англійський відповідник	Розділ
Кластеризація	clustering	1
Коваріаційна матриця	covariance matrix	1, 5
Коефіцієнт трансформант	transform coefficient (distortion measure)	10
Колірна модель	color model (color space)	2
Кореляційний детектор	correlation detector	10
Косинусна подібність	cosine similarity	1, 10
Кратномасштабний аналіз	multiresolution analysis	6
Крайові ефекти	boundary effects	8
Критерій Фішера	Fisher criterion (Fisher score)	11
Кронекерів добуток	Kronecker product	5
Лапласіан	Laplacian	7, 8
Лінійний дискримінантний аналіз	linear discriminant analysis	11
Локальний бінарний шаблон	local binary pattern	11
Масштабувальна функція	scaling function	6
Матриця Адамара	Hadamard matrix	5
Матриця помилок	confusion matrix	1
Матриця суміжності рівнів яскравості	gray-level co-occurrence matrix	11
Материнський вейвлет	mother wavelet	6
Медіанний фільтр	median filter	7
Метод головних компонент	principal component analysis	11
Міра подібності	similarity measure	1, 10
Модуляційна передавальна функція	modulation transfer function	3
Моменти зображення	image moments	11
Муар	moiré	2
Навчальна вибірка	training set	1
Некоректно поставлена задача	ill-posed problem	3, 8
Нерізке маскування	unsharp masking	7
Нормалізація перетворення	transform normalization	10
Нормальне перетворення	normal (template-adapted) orthogonal transform	10
Нормована взаємна кореляція	normalized cross-correlation	10
Нормування ознак	feature scaling (normalization)	11
Образ	pattern, image	1
Однорідне (гомогенне) поле	homogeneous (stationary) random field	2, 3
Ознака	feature	1, 11
Ортогональне перетворення	orthogonal transform	4
Перенавчання	overfitting	1
Перетворення Карунена — Лоева	Karhunen–Loève transform	5
Перетворення Уолша — Адамара	Walsh–Hadamard transform	5
Перетворення Хаара	Haar transform	5
Погоджений фільтр	matched filter	10
Попереднє оброблення	preprocessing	7
Порогова сегментація	thresholding	7
Принцип невизначеності	uncertainty principle	6
Принцип нев’язки	discrepancy principle	8
Простір ознак	feature space	1
Псевдоінверсний фільтр	pseudo-inverse filter	8
Регуляризація	regularization	8

Український термін	Англійський відповідник	Розділ
Реставрація образів	image restoration	3, 8, 9
Розпізнавання образів	pattern recognition	1
Розтікання спектра	spectral leakage	4
Сегментація	segmentation	1, 7
Секвента	sequency	5
Сингулярний розклад	singular value decomposition	3
Скейлограма	scalogram	6
Скелет	skeleton (medial axis)	7
Спектральна густина потужності	power spectral density	2, 8
Спектрограма	spectrogram	6
Структурний елемент	structuring element	7
Тестова вибірка	test set	1
Теорема відліків	sampling theorem	2
Теорема про згортку	convolution theorem	4
Теплицева матриця	Toeplitz matrix	3
Трансформанта	transform coefficient / basis function	4
Умовна деконволюція	constrained (conditional) deconvolution	8, 9
Фазова кореляція	phase correlation	4
Фільтр Вінера	Wiener filter	8
Фільтр Габора	Gabor filter	6
Функція розсіювання точки	point spread function	3
Циркулянтна матриця	circulant matrix	3, 4
Число обумовленості	condition number	3
Число Ейлера	Euler number	11
Швидке перетворення Фур’є	fast Fourier transform	4
Шум квантування	quantization noise	2

ДОДАТОК Г. ЗВЕДЕННЯ ОСНОВНИХ ФОРМУЛ

Таблиця Г.1

Дискретні ортогональні перетворення

Перетворення	Пряме перетворення	Основні властивості
Загальне унітарне	$\mathrm{X} = \mathrm{W}\mathrm{x}$, $\mathrm{x} = \mathrm{W}^H\mathrm{X}$	$\mathrm{W}\mathrm{W}^H = \mathrm{E}$; $\|\mathrm{x}\| = \|\mathrm{X}\|$; оператор $\mathrm{A} \to \mathrm{W}\mathrm{A}\mathrm{W}^H$
ДПФ	$X(k) = \sum_n x\,(n)e^{-j2\pi kn/N}$	діагоналізує циркулянтні матриці; ШПФ: $(N/2)\log_2 N$ множень
Двовимірне ДПФ	$F(u,v) = \sum_m \sum_n f\,(m,n)e^{-j2\pi(um/M+vn/N)}$	сепарабельне; амплітудний спектр інваріантний до зсуву
Уолша — Адамара	$\mathrm{X} = \frac{1}{\sqrt{N}}\mathrm{H}_N\mathrm{x}$, $\mathrm{H}_{2N} = \mathrm{H}_2 \otimes \mathrm{H}_N$	лише додавання; інволютивне; циркулянт → блочно-діагональна матриця
Хаара	розклад за функціями (5.15)	локальні базисні функції; $\approx 2N$ операцій
ДКП-II	$X(k) = c_k \sum_n x\,(n)\cos\frac{\pi(2n+1)k}{2N}$	парне продовження; близьке до КЛП для марковських процесів
Карунена — Лоева	$\mathrm{X} = \Phi^T(\mathrm{x} - \bar{\mathrm{x}})$, $\mathrm{R}\varphi_k = \lambda_k\varphi_k$	оптимальна компактність; некорельовані коефіцієнти
RTF	«частоти» $\beta_k = 1 - e^{-j2\pi k/N}$	диференціювання ↔ множення на β_k; без осциляцій Гіббса
Нормальне	$\mathrm{W}_\Sigma = \mathrm{S}_p \cdots \mathrm{S}_1$, перший рядок $= \mathrm{s}^T/\|\mathrm{s}\|$	$k_{\text{тр}} = \|\tan\theta\|$

Таблиця Г.2

Методи реставрації образів (модель $\mathrm{L} = \mathrm{G}\mathrm{f} + \mathrm{V}$)

Метод	Розв'язок	Апріорна інформація
Інверсна фільтрація	$\hat{F} = L/G$	g
Псевдоінверсна фільтрація	$\hat{F} = L/G$ при $\|G\| > \delta$, інакше 0	g, поріг δ
Фільтр Вінера	$\hat{F} = \frac{G^*L}{\|G\|^2 + S_{VV}/S_{ff}}$	g, спектри потужності образу і шуму
Умовна деконволюція (натуральні координати)	$\hat{\mathrm{f}} = [\mathrm{G}^T\mathrm{G} + \lambda^{-1}\mathrm{C}^T\mathrm{C}]^{-1}\mathrm{G}^T\mathrm{L}$	g, енергія шуму ε^2
Умовна деконволюція (базис Фур'є)	$\hat{F} = \frac{G^*L}{\|G\|^2+\lambda^{-1}\|C\|^2}$, λ: $\|\mathrm{L} - \mathrm{G}\hat{\mathrm{f}}\|^2 = \varepsilon^2$	g, ε^2
Умовна деконволюція (довільний базис)	$\hat{\mathrm{f}}_\xi = \left[\mathrm{G}_{2\xi} + \lambda^{-1}\mathrm{C}_{2\xi}\right]^{-1}\mathrm{G}_{1\xi}^T\mathrm{L}_\xi$	g, ε^2
Узагальнений фільтр МНК	$\mathrm{M} = \mathrm{S}_{ff}\mathrm{K}^*\left(\mathrm{K}\mathrm{S}_{ff}\mathrm{K}^* + \mathrm{S}_{VV}\right)^{-1}$	кореляційні матриці образу і шуму
Критерій форми IX	$\left[\mathrm{B}^{-1}\left(\mathrm{A} + \lambda^{(2)}\mathrm{R}_{VV}\right) - \lambda^{(1)}\mathrm{E}\right]\xi = 0$	g, R_{VV}, штрафна функція

Таблиця Г.3

Міри подібності та детектори

Міра	Формула	Інваріантність
Евклідова відстань	$\|\mathrm{x}-\mathrm{y}\|$	—
Відстань Махаланобіса	$\sqrt{(\mathrm{x}-\mu)^T \mathrm{R}^{-1}(\mathrm{x}-\mu)}$	до лінійних невироджених перетворень ознак
Скалярний добуток (кореляція)	$\mathrm{x}^T\mathrm{y}$	—
Косинусна подібність	$\mathrm{x}^T\mathrm{y}/(\|\mathrm{x}\|\|\mathrm{y}\|)$	до масштабу
Коефіцієнт кореляції (NCC)	косинус центрованих векторів	до масштабу і зсуву рівня
Коефіцієнт трансформант	$\sqrt{\sum_{k\geq 1} X_k^2}\,/\|X_0\|$ у нормальному базисі	до масштабу (для центрованих — і до зсуву)
Погоджений фільтр	$h(n)=s(L-1-n)$, $q_{max}=2E/N_0$	—

Таблиця Г.4

Показники якості

Показник	Формула
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$TP/(TP+FP)$
Recall (TPR)	$TP/(TP+FN)$
F_1	$2\,\mathrm{Precision}\cdot\mathrm{Recall}/(\mathrm{Precision}+\mathrm{Recall})$
FAR, FRR	частки хибних прийнять і хибних відмов
MSE	$\frac{1}{MN}\sum\left(f-\hat{f}\right)^2$
PSNR	$10\lg(f_{max}^2/\mathrm{MSE})$
SSIM	$\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}$