Про одну модель представлення даних в задачах розпізнавання та аналізу цифрових зображень

Козлов М. В.

Кафедра моделювання складних систем
Київський націонаціональний університет
імені Тараса Шевченка
nikita.kozlov@gmail.com

Анотація

В статті наводено спосіб представлення зв'язних об'єктів монохромного зображення у вигляді комбінованого кодування довжин серій. Наводяться алгоритми побудови вказаного представлення в поєднанні з адаптованими формулами обчислення моментних та секторних ознак об'єктів заданих у запропонований спосіб.

1. Вступ

На практиці, в процесі роботи з цифровим зображенням, виникає цілий спектр питань: визначення способів зберігання, передачі по каналам зв'язку, розробка алгоритмів синтезу та систем для отримання зображень об'єктів зовнішнього світу. Кожен з цих аспектів є добре розвиненим як в плані теоретичних настанов так і цілим різноманіттям готових технічних приладів і рішень.

Найскладнішими і, в той же час, найцікавішими для досліджень з ціллю впровадження в сучасні програмно-алгоритмічні комплекси є задачі розпізнавання зображень. Саме практичні потреби задач розпізнавання, їх тісний зв'язок з проблемами реалізації алгоритмів класифікації, стали причиною інтенсивного дослідження способів опису форми об'єктів на цифровому зображенні.

Статтю організовано наступним чином: в 2-ому розділі наведено короткий огляд способів кодування цифрових зображень, в 3-му наведено опис представлення зв'язних об'єктів монохромного зображення у вигляді комбінованого кодування довжин відрізків (ККДВ) та алгоритм перетворення монохромного растру в ККДВ-представлення, в 4-му розглянуто реалізацію декількох алгоритмів структурного опису об'єктів в ККДВ-представленні, далі наведено висновки і список використаних джерел.

2. Огляд способів представлення зв'язних об'єктів монохромного зображення

З практичних міркувань в багатьох прикладних задачах в якості опису форми об'єкта обирають його представлення у вигляді монохромного растрового зображення. При цьому один з кольорів відповідає точкам (пікселям) об'єкта, а інший – точкам фону. В залежності від поставлених цілей, таких як: компактне зберігання, зручність реалізації алгоритмів обробки зображень, застосування (перенесення) певного математичного апарату до задач цифрової обробки відрізняються й відомі способи цифрового представлення таких зображень.

Граф суміжності ліній (LAG) являє собою метод кодування довжин серій для кодування об'єктів [1]. "Медіанна вісь" (median axis) [3] є представленням, що не містить інформації про границі об'єктів. Ланцюговий код [3] є результатом квантування траєкторії обходу об'єкта по контуру. Факс-стиснуте представлення [3] кодує відмінність поточної стрічки сканування від попередньої. Стандарт кодування зображень JBIG відстежує форму, а не довжину серій чи зміну однієї стрічки зображення на іншу. Стандарти JBIG2/JB2[3] визначають формат стисненого бітового потоку об'єктів монохромного зображення. В роботі [2] пропонується спосіб кодування монохромних цифрових зображень за допомогою списку особливих місць з вказанням їх характеристик. Для формування такого представлення вводиться двомірна граматика, яка за основу приймає алфавіт, елементами якого є 16 різних кодувань четвірок пікселів з координатами (i, j), $(i, j-1)$, $(i-1, j)$, $(i-1, j-1)$. Таке представлення допускає однозначне відтворення вхідного зображення.

Хоча наведені вище схеми кодування дуже ефективні для передачі даних, вони не зручні

для використання в системах обробки зображень, де форма об'єктів може змінюватись динамічно. Однак, і для таких високоефективних в плані стиснення способів представлення зображень відомі роботи по адаптації алгоритмів фільтрації, що працюють безпосередньо з стисненим потоком бітів, або довжинами серій [3].

3. Представлення об'єктів монохромного зображення у вигляді комбінованого кодування довжин відрізків

Класичним способом отримання даних про розташування та форму об'єктів дослідження на цифровому зображенні є попередня обробка, що перетворює вхідний растр за допомогою набору операцій фільтрації у монохромне зображення. При цьому шуканим об'єктам відповідають області одного з кольорів (чорного або білого) на вихідному зображенні.

Операція заливки ({seed,flood} fill)[3] є одним з фундаментальних алгоритмів обробки зображень, і являє собою процес присвоєння міток пікселям деякої області зображення. Початковим є один піксель або набір пікселів, з яких починається зміна кольору. Операції заливки можуть бути використані для визначення зв'язних компонент, послідовно заливаюючи їх, тобто шляхом зміни їх кольору, або їх видалення.

В результаті виконання операції пошуку зв'язних компонент [3] на цифровому зображенні вдається отримати дані про форму та просторове розташування об'єктів у вигляді множин координат пікселів. Таким чином, встановивши відповідність між об'єктами зображення та наборами пікселів, що їм належать, можна звести задачу цифрового представлення об'єктів монохромного зображення до задачі компактного безвтратного зберігання набору координат пікселів.

3.1. Комбіноване кодування довжин відрізків

Далі будемо називати *гліфами* зв'язні об'єкти монохромного зображення. Для опису гліфів використаємо структуру даних, що балансує між компактністю представлення та зручністю програмної реалізації, враховує аспекти адаптації математичного апарату обчислення моментних характеристик [7] та можливість застосування в алгоритмах структурного опису зображень.

Виходячи з того, що пікселі цифрового зображення мають квадратну форму, очевидно, що будь-який гліф монохромного растру можна представити у вигляді множини попарно неперетинних прямокутників, зі сторонами паралельними координатним вісям. Якщо висота всіх таких прямокутників дорівнює одиниці, отримаємо так зване горизонтальне кодування довжин відрізків [1], яке є простою модифікацією факс-представлення. Якщо на зображенні переважна кількість гліфів "розтягнута" вздовж вертикальної вісі, зручно використовувати вертикальне кодування довжин відрізків. Таке представлення відрізняється від горизонтального лише тим, що довжини серій пікселів гліфа задано набором прямокутників шириною в 1 піксель.

При цьому, як для вертикального, так і для горизонтального представлення достатньо зберігати стрічки пікселів гліфа у вигляді множини трійок значень $\{x, y, r\}$, де (x, y) – координата початку стрічки, а r – її довжина, вказавши тип кодування e, такий що $e \in \{v, h\}$. Тут h означає горизонтальне кодування, v – вертикальне. Поєднавши горизонтальне та вертикальне кодування отримаємо комбіноване кодування довжин відрізків (ККДВ) для представлення даних гліфів. Для повного відтворення положення гліфу на зображенні та його форми ККДВ-представлення містить наступні дані: координати прямокунтика, в який вписано гліф, тип кодування (вертикальне або горизонтальне), множина відрізків у вигляді набору трійок $\{x, y, r\}$.

3.2. Алгоритм побудови ККДВ-представлення

Для отримання даних гліфа використовується модифікований алгоритм заливки [5, 3]. При цьому, обхід околу пікселя представлено у вигляді послідовності команд замість циклічного обходу параметризованого околу, це дозволяє зменшити кількість умовних переходів в операціях перевірки належності поточних значень x та y області зображення. Відомо, що об'єм памя'ті необхідний для збереження масиву даних лінійно залежить від розміру типу елементів. З цієї причини доцільно намагатися зменшити абсолютні значення початкових точок відрізків (x, y), та їх довжин r, тому вони зберігаються відносно верхнього лівого кута оточуючого гліф прямокутника.

3.3. Аналіз характеристик ККДВ-представлення

Розмір ККДВ-представлення набору зв'язних компонент залежить від складності зображення. Чим більше об'єкти монохромного зображення нагадують "шахову дошку", тим гірше стиснення. Лінії сканування (горизонтальні чи вертикальні) проходять через одну точку по периметру гліфа і

виходять через іншу точку, і ці дві точки не використовуються для кодування будь-якого іншого відрізку представлення. З цього випливає, що кількість відрізків представлення гліфа приблизно дорівнює половині довжини (вимірюється в пікселях) його периметру. Тому оціночний коефіцієнт стисненя гліфу у ККДВ-представленні приблизно дорівнює відношенню

$$c \approx \frac{2 \times \text{напівпериметра}}{\text{кількість пікселів}} = \frac{\text{периметр}}{\text{площа}}. \qquad (1)$$

ККДВ-представлення в багатьох випадках буде займати менше пам'яті ніж растрове представлення. Лише тоді, коли форма гліфа буде заповнювати оточуючий прямокутник більш повно, або у випадку коли коефіцієнт стиснення гліфу c (1) задовольнятиме нерівності $c \geq 1$, растр передаватиме інформацію про форму гліфа ефективніше. Результати порівняльного аналізу параметрів стиснення ККДВ-представлення гліфів ψ, ω, α наведено у таблиці 1.

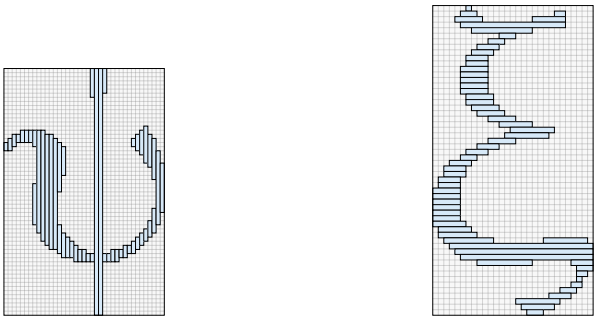

Рис. 1. ККДВ-представлення гліфів ψ, ξ.

Табл. 1. Порівнняня параметрів стиснення для ККДВ-представлення гліфів ψ, ω, α.

#	растр	гліф	відрізків	стиснення (1)
ψ	2340	463	50	0.107991
ω	1287	461	46	0.0997831
α	1120	372	58	0.155914

4. Алгоритми структурного опису об'єктів заданих комбінованим кодуванням довжин відрізків

Одним з широковживаних підходів до опису об'єктів в задачах розпізнавання є встановлення бієкції між властивостими об'єкта та його формою [1]. Ми також будемо притримуватися цого підходу і запропонуємо декілька способів опису об'єктів, що задані ККДВ-представленням.

4.1. Мінімальне описане коло

Важливою характеристико гліфа є його спрощена геометрична версія, або контейнер. Таким контейнером є мінімальний оточуючий прямокутник, що є результатом роботи алгоритму побудови ККДВ-представлення. Для обчислення ознак гліфів, що будуть розглянуті далі, використовується мінімальне описане коло.

Визначення параметрів мінімального кола, якому належить заданий гліф виконується за алгоритмом Везля [4] з декількома модифікаціями. Суть модифікацій полягає в тому, щоб зменшити кількість точок площини, з якими працює алгоритм. Якщо хоча б 2 протилежні кути оточуючого прямокутника є точками гліфа, тоді мінімальне оточуюче коло співпадає з колом, що містить прямокутник. "Точки підтримки" мінімального оточуючого кола, або множина таких точок, що має дві або три точки, через які пройде шукане коло [4], задовольняє одній з наведених нерівностей:

$$x_{lt}y_{lt} - yx_{lt} - xy_{lt} > 0, \qquad (2)$$
$$x(y - y_e) > (x - x_{lb})(y - y_{lb}), \qquad (3)$$
$$(x - x_{rb})(y - y_{rb}) > (x - x_e)(y - y_e), \qquad (4)$$
$$(x - x_e)y > (x - x_{rt})(y - y_{rt}), \qquad (5)$$

тут x, y – координати точки, що перевіряється на

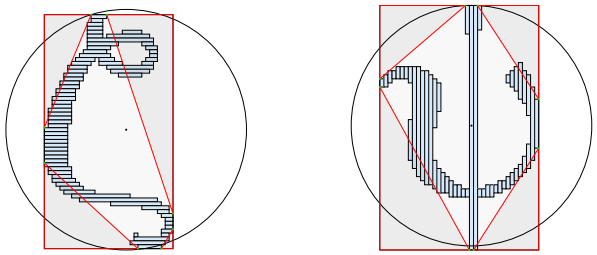

Рис. 2. Області належності "точки підтримки" мінімального оточуючого кола та побудоване коло для гліфів ζ, ψ.

належність до множини «точок підтримки» мінімального оточуючого кола; x_e – найбільша з точок гліфа x-координата; y_e – найбільша з точок гліфа y-координата; Множина точок підтримки не зазнає випадкового перемішування, на відміну від оригінального алгоритму [4]. Точки підтримки йдуть в порядку за годинниковою стрілкою для горизонтального кодування, та проти годинникової стрілки для вертикального. Англійські літери в нижніх індексах означають l – ліва (left), t – верхня (top), r – права (right), b – нижня (bottom), e – остання (end).

4.2. Моментні ознаки

Зрозуміло, що геометричні моменти m_{pq} [6] об'єктів в ККДВ-представленні можна обчислити у вигляді суми відповідних моментів стрічок m^i_{pq}:

$$m^i_{pq} = \left[\left(r_i + \frac{1}{2}\right)^{p+1} - \left(l_i + \frac{1}{2}\right)^{p+1}\right] \times$$
$$\times \left[\left(b_i + \frac{1}{2}\right)^{q+1} - \left(t_i + \frac{1}{2}\right)^{q+1}\right]; \quad (6)$$
$$m_{pq} = \frac{1}{(p+1)(q+1)} \sum_i m^i_{pq}.$$

де l_i, r_i – ліва та права границі поточної стрічки пікселів; t_i, b_i – верхня та нижня границі відповідно; i – індекс сумування, що пробігає по всім стрічкам.

4.3. Секторні ознаки

Для розпізнавання рукописних символів в роботі [5] в якості кількісних ознак використовуються значення радіального розподілу мас геометричної форми символа відносно його центра ваги. Для опису об'єктів в ККДВ-представленні пропонується використовувати нормалізовані значення ознак [5] відносно центра описаного кола (i_c, j_c):

$$v(n) = \frac{1}{r \cdot pow(V_n)} \sum_{(i,j)\in V_n} \sqrt{(i-i_c)^2 + (j-j_c)^2}, \quad (7)$$

де $pow(V_n)$ – функція, що повертає кількість элементів множини V_n:

$$V_n = \{(i,j) : sector(j-j_c, i-i_c, \frac{2\pi}{n_0}) = n\}. \quad (8)$$

Тут n_0 – кількість секторів, n – номер поточного сектору, $sector(x, y, n_0)$, функція, що повертає номер сектора шириною $\frac{2\pi}{n_0}$ радіан, якому належить точка (x, y), $sector(x, y, n_0) =$

$$\begin{cases} 0, & y=0, x=0, \\ \frac{n_0}{2}, & y=0, x<0, \\ 0, & y=0, x>0, \\ \frac{n_0}{4}, & y>0, x=0, \\ \frac{3n_0}{4}, & y<0, x=0, \\ \frac{n_0}{2\pi}\arctan(\frac{y}{x}), & y>0, x>0, \\ \frac{n_0}{2\pi}\left[\arctan(\frac{y}{x}) + 2\pi\right], & y<0, x>0, \\ \frac{n_0}{2\pi}\left[\arctan(\frac{y}{x}) - \pi\right], & y>0, x<0, \\ \frac{n_0}{2\pi}\left[\arctan(\frac{y}{x}) + \pi\right], & y<0, x<0. \end{cases} \quad (9)$$

де r, (i_c, j_c) – радіус та координати центру мінімального оточуючого гліф кола [4]. Розташування 16-ти секторів мінімального оточуючого кола для гліфів гліфів ξ, α показано на рисунку 3.

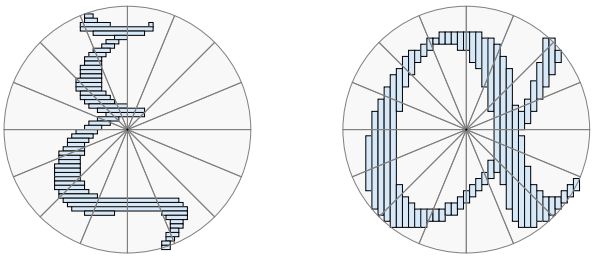

Рис. 3. Області секторів гліфів ξ, α, для $n_0 = 16$.

5. Висновки

Отримані алгоритми та представлення доцільно використовувати в системах з критичними вимогами до пам'яті. ККДВ-представлення є ефективним для зберігання об'єктів, що мають велике значення відношення площі до периметру. Наведені в статті алгоритми реалізовані мовою C++ у бібліотеці розробника OCR-программ, і показали свою придатність в промисловому використанні. Відладка алгоритмів та генерація ілюстрацій до статті виконана за допомогою програми Wolfram Mathematica.

6. Література

[1] *Pavlidis T.* Algorithms for Graphics and Image Processing. – Rockville, MD. Computer Science Press, 1982.

[2] *Шлезингер М. И.* Математические средства обработки изображений. – Киев: Наукова думка, 1989. — 200 с.

[3] *Ritter G. X. Wilson J. N.* Handbook of Computer Vision Algorithms in Image Algebra 2 Ed. – CRC Press, 2001.

[4] *Welzl E.* Smallest enclosing disks (balls and ellipsoids) // LNCS 555, – 1991. – P. 359–370.

[5] *Кириченко Н. Ф. Полищук А. А.* Метод контейнеров в распознавании рукописных математических текстов // Проблемы управления и информатики. — 2003. – № 6. –С. 54–71.

[6] *Flusser J. Suk T. Zitová B.* Moments and Moment Invariants in Pattern Recognition. – Chippenham: John Wiley & Sons Ltd, 2009. – 296 p.

[7] *Гаращенко Ф. Г., Козлов Н. В.* Разработка и реализация эффективных алгоритмов вычисления геометрических моментов // Компьютерная Математика. –2008. – № 1. –С. 88–96.