

Б.М. Гавриш<sup>1</sup>, к.т.н., ст. викл, УАД, О.В. Тимченко<sup>1, 2</sup>, д.т.н, професор,  
О.О. Тимченко<sup>1</sup>, аспірант

## ПОБУДОВА ГРАМАТИК ДЛЯ СТРУКТУРНОГО РОЗПІЗНАВАННЯ ТЕКСТОВИХ ЗОБРАЖЕНЬ

**Анотація.** Розглядається метод побудови систем оптичного розпізнавання текстів, формування ознак зображень та досліджується метод побудови граматики для структурного розпізнавання текстових зображень. Зроблено висновок про доцільність використання в методах розпізнавання текстових зображень двовимірних контекстно-вільних граматики незважаючи на їх велику обчислювальну складність.

**Ключові слова:** систем оптичного розпізнавання текстів, структурне розпізнавання, контекстно-вільні граматики.

**Abstract.** The method of constructing optical character recognition systems, the formation of signs of images and the method of constructing grammar for the structural recognition of text images is investigated. A conclusion is made on the expediency of using two-dimensional context-free grammars in the methods for recognizing text images despite their high computational complexity.

**Key words:** systems of optical character recognition, structural recognition, context-free grammar.

### Вступ

Оптичне розпізнавання тексту (англ. optical character recognition, OCR) — це механічне або електронне переведення зображень рукописного, машинописного або друкованого тексту в послідовність кодів, що використовуються для представлення в текстовому редакторі.

Проблема ефективного розпізнавання тексту посідає важливе місце в сферах інформатизації різних процесів людської діяльності. Текстове представлення інформації, порівняно із графічним, дозволяє істотно скоротити витрати на зберігання та передачу інформації, а також дозволяє реалізувати всі методи використання та аналізу електронних документів. Тому найбільший інтерес з практичної точки зору представляє саме переклад паперових носіїв в текстовий електронний документ [1].

**Мета роботи:** визначити особливості систем оптичного розпізнавання текстів та дослідити метод побудови граматики для структурного розпізнавання текстових зображень.

### Загальна характеристика систем розпізнавання текстів.

Для переведення зображень в текстовий електронний документ

---

<sup>1</sup>, Українська академія друкарства

<sup>2</sup> Uniwersytet Warmińsko-Mazurski w Olsztynie

використовуються системи оптичного розпізнавання (наприклад, FineReader, OmniPage, ReadIris чи інші). Система OCR отримує цифрове зображення сканованого або сфотографованого документа і повинна сформувати текст, який містить це зображення, у вигляді, придатному для збереження в одному з форматів електронних текстових документів. Загальна структурна системи показана на рис. 1. [2].



Рис.1. Структурна схема системи розпізнавання символів

Призначення блоків системи наступне. **Блок зчитування вхідних даних** (сканування, фотографування) проводить перетворення енергії світла в електричні сигнали. Розмір матриці сканування залежить від точності процесу отримання зображення та кольору зображення або його фрагментів. При розпізнаванні букв в характеристику кольору зображення вводиться лише чорний і білий кольори, які значно зменшують розміри матриці ознак.

**Блок попередньої обробки** переводить зображення у чорно біле з метою економії пам'яті і зменшення витрат на обробку інформації. Також відбувається зменшення рівня шумів у вигляді дрібних ізольованих цяток, ліквідації так званої бахроми, дрібних прогалин (порожнеч) всередині ліній і, якщо необхідно, операції потоншення ліній. Також цей блок може виконувати функції виправлення геометричних спотворень зображень, отриманих, наприклад, шляхом фотографування.

**Блок виділення ознак** може мати різну складність в залежності від природи зображення, яке розпізнається. Кожен підданий попередній обробці об'єкт повинен бути представлений у вигляді структури мовного типу, наприклад ланцюжка. Цей процес складається з двох підпроцесів: з сегментації і з виділення непохідних елементів. Основною метою процесу сегментації є виділення з зображення окремих елементів для виділення певних ознак або порівняти відмінності зі стандартними зразками в пам'яті. Будь-яка помилка в процесі сегментації неминуче призведе до збою у виконанні наступних етапів. Так, при розпізнаванні друкованого тексту, в якому символи відокремлені один від іншого, сегментація дає результати, які дозволяють в результаті розпізнавання отримати текст, практично повністю ідентичний оригінальному. При розпізнаванні ж рукописних текстів або друкованих символів, що зливаються або накладаються, рівень відповідності

отриманого тексту оригінальному значною мірою залежить від можливостей системи. Проблеми з точністю процесу сегментації виникають і тоді, коли символи пов'язані з графіками, таблицями чи малюнками.

**Блок класифікації** формує ознаки відповідно до результатів, які надає блок виділення ознак, формує ознаки відповідності елементів зображення тим чи іншим еталонним зразкам. Оцінка ефективності класифікації залежить від якості виконання попередніх етапів і оцінюється кількістю ознак, за якими буде встановлена ця відповідність. Слід зазначити, що процес класифікації досить часто вимагає повторних циклів аналізу подання вхідного зображення.

Для розпізнавання символів у системах розпізнавання текстів використовуються такі класифікатори: растровий, ознаковий, структурний та їх поєднання. Використовуються вони в залежності від змісту документа, вхідних параметрів зображення і задачі розпізнавання [3].

Растровий класифікатор порівнює символ з набором еталонів, по черзі накладаючи зображення одне на одне. Еталонами виступають спеціально підготовлені зображення, кожне з них об'єднує в собі контури багатьох варіантів написання того або іншого символу.

Ознаковий класифікатор висуває гіпотези, виходячи із ступеня схожості параметрів символу з еталонними значеннями. Оперує певними числовими ознаками, такими, наприклад, як довжина периметра, кількість чорних точок в різних областях або уздовж різних напрямів і т. д.

Структурний класифікатор проводить структурний аналіз символу, розкладаючи останній на елементарні складові (відрізки, дуги, кола, точки) і формує точну схему аналізованого знаку. Потім отримана схема (структурний опис букви) порівнюється з еталоном. Цей класифікатор працює

повільніше растрового і ознакового, зате відрізняється високою точністю.

**Блок навчання** модифікує представлення елементів таким чином, щоб допомогти якнайшвидше класифікувати елементи зображення на наступних етапах. Він так само надає можливість збереження системою методу, стилю та інших результатів класифікації, які після успішного завершення процесу виділення ознак можуть бути використані для більш точного і швидкого розпізнавання зображень.

### **Характеристика зображення графічних документів**

Особливістю цих зображень є те [4], що вони складаються з великої кількості взаємозалежних частин. Аналіз таких зображень доцільно проводити за допомогою методів структурного розпізнавання. Результатом такого розпізнавання є не віднесення зображення символу до прототипу, а перелік символів і відношень між ними. Складність, характерна для розпізнавання саме таких зображень, полягає у тому, що символ не визначається однозначно своїм зображенням. Ім'я зображення символу залежить не тільки від зображення цього фрагмента, а і від його місця, оточення, тобто контексту.

Ряд методів такого розпізнавання базується на евристичних прийомах, тобто розпізнавання за контекстом. Як правило, вказані методи залежать від предметної області і досвіду. Програмні продукти цього типу дуже ефективно розв'язують конкретну задачу. Таким чином, в блоці навчання (рис.1) доцільно реалізовувати можливість збереження системою методу, стилю та інших результатів класифікації.

Процедури розпізнавання, тобто функціонування вже навченої системи, є, як правило, достатньо простими і простими в реалізації. Тому, коли структурне розпізнавання застосовується до розв'язання проблем розпізнавання рукописного тексту, актуальність процедур навчання різко зростає. Однак проблеми навчання в структурному розпізнаванні значно складніші і значно менш досліджені порівняно з навчанням в неструктурному розпізнаванні.

Структурне розпізнавання символів в умовах випадкових шумів зводиться до пошуку оптимального зображення символів, які розпізнаються. Функцію якості полягає у відшуканні найімовірнішої сукупності прихованих параметрів зображення. Наприклад, при розпізнаванні рядка тексту вимога знайти найімовірнішу послідовність літер рівнозначна вимозі мінімізації кількості неправильно розпізнаних символів.

Зображення які містять тексти, таблиці, рисунки, створюються і читаються згідно з певними правилами, набір яких формалізувати як певну граматику. Очевидно, що алгоритми розпізнавання таких зображень повинні базуватись на використанні правил цієї граматики.

**Методи створення таких граматик наступні:**

- використання теорії графів. Зображення тексту представлено у вигляді розміченого графа. Задачі розпізнавання ставляться як задачі знаходження ізоморфізму еталонного та вхідного графів, або ізоморфізму їх підграфів.

- методи теорії формальних мов та граматик. Зображення розглядається як слово у деякій формальній мові, яка задається за допомогою конструкцій, що є узагальненнями граматик Хомського. Розпізнавання полягає у відшуканні найкращого в певному значенні виводу зображення у заданій граматиці [5].

Розглянемо методику створення відповідної граматики. Для цих граматик характерним є розгляд зображення як об'єкта, що складається за певними правилами з великої кількості елементарних частин. Ці частини і правила можуть значно відрізнитись між собою (наприклад, розпізнавання нот [6] та розпізнавання математичних формул [7]). В роботі [5] введено формалізм, названий загальною структурною конструкцією. Окрім власне визначення загальної структурної конструкції, в [5] наведено загальну постановку та алгоритм розв'язання задачі розпізнавання в рамках введеного формалізму.

Нехай  $T$  – певна прямокутна підмножина двовимірної цілочисельної решітки:  $T = \{(i, j) \mid 0 \leq i < H, 0 \leq j < W\}$ . Множину  $T$  називатимемо полем зору, числа  $H$  та  $W$  – його висотою та шириною відповідно, а елементи поля зору називатимемо пікселями. Кольори пікселів вибираються з певної

скінченної множини  $Y$ . Відображення  $x: T \rightarrow Y$  називатимемо зображенням, величина  $x(t)$  визначає колір зображення  $x$  в пікселі  $t \in T$ .

Двовимірною контекстно-вільною граматику буде  $G = \langle E, V, P, \varepsilon \rangle$ , де  $E$  – множина термінальних зображень містить однопіксельні зображення (зображення, визначені на полі зору розміром 1 піксел) всіх кольорів з множини  $Y$ ,  $V$  – множина нетермінальних імен (метасимволи), що присвоюються частинам зображення в процесі його генерування за допомогою граматики,  $\varepsilon \in V$  – аксіома, використовується для іменування всього, повністю згенерованого зображення,  $P$  – множина правил виводу містить правила трьох типів: правила горизонтальної конкатенації, вертикальної конкатенації (об'єднання блоків зображення) та правила заміни. Кожну окрему множину цих правил позначатимемо відповідно  $P_h, P_v, P_s$ .

Правила заміни  $P_s$  мають вигляд  $v \rightarrow e$ , де  $v \in V$  – деяке нетермінальне ім'я, а  $e \in E$  – термінальне зображення.

Множина правил горизонтальної конкатенації  $P_h$  містить трійки нетермінальних імен вигляду  $v \rightarrow v_l/v_r$ . Наявність такого правила вказує на те, що будь-який прямокутний фрагмент зображення може отримати ім'я  $v$ , якщо його можна розбити вертикальною лінією на такі два прямокутні фрагменти, що лівий вже має ім'я  $v_l$ , а правий –  $v_r$ . Символ  $|$  використовується для розділення пари нетермінальних імен у правилах горизонтальної конкатенації, а не для альтернативного запису кількох правил, як це прийнято формальних граматиках.

Аналогічно множина правил вертикальної  $P_v$  конкатенації містить трійки нетермінальних імен вигляду  $v \rightarrow \frac{v_t}{v_b}$ .

Наявність такого правила вказує на те, що будь-який прямокутний фрагмент зображення може отримати ім'я  $v$ , якщо його можна розбити горизонтальною лінією на такі два прямокутні фрагменти, що верхній вже має ім'я  $v_t$ , а нижній –  $v_b$ .

Мова граматики  $G$  складається з таких зображень, яким може бути присвоєне ім'я  $\varepsilon$ . Послідовність правил, застосованих до зображення  $x$ , результатом чого є присвоєння імені  $\varepsilon$  всьому зображенню, називається виводом зображення  $x$  в граматиці  $G$ .

Природнім чином для двовимірних контекстно-вільних граматик формулюється задача на точний збіг: для вхідного зображення  $x$  та граматики  $G$  визначити, чи належить  $x$  до мови граматики  $G$ , тобто чи існує вивід зображення  $x$  і якщо так, то знайти цей вивід.

Алгоритм (див. [5]) розв'язання задачі є прямим узагальненням алгоритму Кока-Янгера-Касамі [8] для визначення належності слова мові певної контекстно-вільної граматики Хомського. Він полягає в тому, що при послідовному перегляді всіх прямокутних фрагментів зображення для кожного з них визначається, які імена можуть бути присвоєні йому в даній граматиці. При цьому для перегляду фрагменти впорядковуються за розмірами.

Введене означення двовимірної контекстно-вільної граматики та постановка задачі на точний збіг, вочевидь мають ряд недоліків, що суттєво обмежують їх практичне застосування:

- з точки зору розпізнавання зображень формалізм двовимірних контекстно-вільних граматик не завжди дозволяє знайти реальне зображення, яке можна розділити на прямокутні фрагменти і при цьому дані фрагменти не перетинаються;
- надмірна деталізація правил граматики: для кожного зображення слід вказувати, яким чином воно складається з менших частин аж до рівня окремих пікселів. Очевидно слід оперувати більшими фрагментами зображення без вказання, яким чином вони можуть бути виведені з пікселів, наприклад, фрагментами, що відповідають окремим літерам тексту;
- постановка задачі на точний збіг вимагає точного збігу значень кольорів окремих пікселів зображення з кольорами термінальних зображень. За наявності шуму, що є характерним для розпізнавання, ця умова не виконується.

Базова контекстно-вільна конструкція, введена в [4] дозволяє усунути вказані недоліки, оскільки:

- вона описує зображення як таке, що складається з фрагментів довільної, а не лише прямокутної форми;
- процес побудови зображень задається до рівня довільного набору термінальних фрагментів, розміри яких можуть значно перевищувати розмір пікселя;
- визначається штраф за присвоєння імен фрагментам зображення, який для заданого фрагмента дорівнює сумі штрафів фрагментів, з яких його було утворено;
- штраф за термінальні фрагменти визначається довільною функцією, яка не пов'язана з формалізмом контекстно-вільних граматик. Це дозволяє використовувати у її побудові довільні, в тому числі, й евристичні міркування. Наприклад, в задачі розпізнавання текстів нею може бути будь-яка функція, яка визначає схожість довільних літери та фрагмента зображення.

### **Висновки**

Алгоритм розв'язання задачі синтаксичного аналізу в такій конструкції є узагальненням відповідного алгоритму для двовимірних контекстно-вільних граматик і полягає в послідовному обчисленні для кожного фрагмента штрафу за присвоєння йому кожного імені.

Спільним недоліком вказаних алгоритмів є їх часова та просторова складність, яка, наприклад, у випадку двовимірних контекстно-вільних граматик становить  $O(H^2W^2(H + W))$  та  $O(H^2W^2)$  відповідно. Ця складність хоч і є поліноміальною від розмірів  $H$  та  $W$  зображення, та все ж є надто великою для безпосереднього застосування алгоритмів на практиці.

Часова та просторова складність цих алгоритмів визначаються перш за все кількістю фрагментів, які переглядаються в ході їх роботи. Тому

зменшення цієї кількості є основним способом зменшення складності алгоритмів розпізнавання.

1. *Анисимов, Б. В.* Распознавание и цифровая обработка изображений: Учебное пособие для вузов / Б. В. Анисимов, В. Д. Курганов, В. К. Злобин. – М.: Высшая школа, 1983. – 295 с.
2. *Тимченко О.В.* Алгоритми та функції інформаційної системи розпізнавання символів на основі методів поліпшення зображень / Тимченко О.В., Кульчицька І.О., Тимченко О.О. // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Вип.69. – К.: 2013. – С.167-173.
3. *Патрик Э.* Основы теории распознавания образов / Эдвард А. Патрик. – М.: Советское радио, 1980; (пер. с англ. под ред. Б.Р. Левина). – 864с.
4. *Савчинський Б. Д.* Контекстно-вільні граматичні конструкції для розпізнавання зображень текстових та графічних документів. Дис. ... канд. техн. наук за спец. 05.13.23 - системи та засоби штучного інтелекту. Київ, 2007.
5. *Шлезингер М., Главач В.* Десять лекций по статистическому и структурному распознаванию. - Киев: Наукова Думка, 2004. - 545 С.
6. *Kopec G. E., Chou P. A., Maltz D. A.* Markov source model for printed music decoding // Journal of Electronic Imaging. January 1996. Vol. 5, no. 1. Pp. 7–14.
7. *Garcia P., Couasnon B.* Using a genetic document recognition method for mathematical formulae recognition // IAPR Intern. Workshop on Graphics Recognition / Ed. by D. Blostein, Y.-B. Kwon. LNCS 2390. Berlin, Heidelberg: Springer-Verlag, 2002. Pp. 236–244.
8. *Kasami T.* An efficient recognition and syntax analysis algorithm for context-free languages: Tech. Rep. AFCLR-65-758. Bedford, Mass., USA: Air Force Cambridge Research Laboratory, 1965.

*Поступила 2.03.2017р.*

УДК 004.032.26

Є.Я. Ваврук, І.Є.Ваврук

## **КОМПРЕСІЯ ЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ**

Розроблено метод компресії зображень з використанням штучної нейронної мережі Кохонена.

**Ключові слова.** Штучна нейронна мережа, мережа Кохонена, компресія кольорових зображень, середньоквадратичне відхилення, пікове співвідношення сигналу до шуму.

Разработан метод компрессии изображений с использованием искусственной нейронной сети Кохонена.

**Ключевые слова.** Искусственная нейронная сеть, сеть Кохонена,