



## ВИЗНАЧЕННЯ ЕЛЕМЕНТІВ ТЕКСТУ НА ЗОБРАЖЕННІ З ВИКОРИСТАННЯМ НЕЧІТКИХ КОГНІТИВНИХ МОДЕЛЕЙ

Олена П'ятикоп

Приазовський державний технічний університет  
вул. Університетська, 7, Маріуполь 87500 Україна, E-mail: Pjatikopelena@rambler.ru

**Резюме:** в даній публікації описуються результати моделювання знань з когнітивної психології про первинну обробку інформації в зоровій системі для аналізу зображення з метою визначення елементів тексту. Обробка зображення на рівні клітин первинної зорової кори формалізована за допомогою нечітких множин.

**Ключові слова:** зорова система, клітини первинної зорової кори, орієнтаційні колонки, зображення, рядки літери.

### ВСТУП

У галузі штучного інтелекту актуальним завданням залишається проектування систем автоматичної обробки графічної інформації. Одним з напрямків застосування таких систем є розпізнавання тексту. Вирішення цієї задачі актуально при перетворенні паперових схем, креслень та інших документів в їх електронний еквівалент. На сьогоднішній день вже є досвід розробки методів розпізнавання друкованих символів із застосуванням різного виду класифікаторів, нейронних мереж [1-2]. Але обсяг цифрової графічної інформації збільшується, змінюються умови розпізнавання, і від сучасних систем потрібно більш глибокий інтелектуальний аналіз. Тому для обробки зображення досліджується можливість застосування знань про будову зорової системи та процеси, що відбуваються в неї [3-5]. Сприйняття зорової та звукової інформації, а також процеси мислення і пам'яті описує когнітивна психологія [6-7].

Ці знання дають можливість моделювати сприйняття ока людини і застосовувати отримані моделі для обробки зображення.

У даній статті розглядають використання когнітивних моделей, зокрема моделі гангліозних клітин та моделі клітин первинної зорової кори для вирішення задачі визначення елементів тексту на зображенні.

### 1. ПОСТАНОВКА ЗАДАЧІ

Нехай є зображення з текстом в градаціях

сірого 8 біт. Текст звичайно формується з рядків, а рядки, в свою чергу, складаються з слів. Тому спочатку необхідно визначити місце розташування передбачуваних рядків, далі розбити рядки на можливі слова, а кінцевим етапом виконати виділення букв. Розташування рядків на зображенні можливо не тільки горизонтальне, а під будь-яким кутом. Крім цього невідомий розмір букв тексту, тобто висота рядків. У зв'язку з цим, початковий аналіз зображення спрямований на визначення кута нахилу і висоти рядків тексту. Після визначення цих параметрів передбачуваний рядок розташовуємо горизонтально і виконуємо подальший аналіз.

Для дослідження зображення використовувати нечітку модель клітин первинної зорової кори.

#### Когнітивні знання

Якщо кинути побіжний погляд на зображення, не фіксуючи його на деталях, то відбувається сприйняття найбільш простих (укрупнених) паттернів. У розглянутій задачі це будуть рядки тексту, де рядки – це чергуються смуги одного напрямку і приблизно однієї ширини. Так відбувається, якщо зображення потрапляє в область периферичного зору, а також, якщо дивитися на зображення побіжно або здалеку. Після проєцювання зображення на сітчатку в ній відбуваються такі процеси:

1) Першими образ на сітчатці обробляють фоторецептори палички і колбочки, які через проміжні біполярні клітини активізують певні рецептивні поля гангліозних клітин. Розрізняють

гангліозні клітини двох типів: "ON" клітини активізуються, коли світло потрапляє на деякий чутливий центр, і не реагують, коли світло падає на область навколо цього центру; клітини "OFF" типу реагують навпаки. Тому ці клітини ще називають релейними [6-7]. Безліч активних гангліозних клітин утворює першу "проекцію" зображення.

2) Далі, більшість аксонів гангліозних клітин утворюють синаптичні зв'язки з клітинами латерального колінчастого тіла (ЛКТ), рецептивні поля яких дуже схожі на рецептивні поля гангліозних клітин сітчатки: мають центральні "зони включення" і периферійні "зони виключення" чи навпаки. Крім цього клітини ЛКТ відрізняються розмірами і діляться на дві групи: парвоцелюлярні і магноцелюлярні клітини.

3) В ЛКТ перша "проекція" зображення стає пульсуючою [8]. Відразу ж після саккади<sup>1</sup> в ЛКТ формується ретинотопічна карта, яка передається в первинну зорову кору. Діаметр рецептивних полів зменшується, і на основі нього формуються нові ретинотопічні карти. До чергового стрибка зорова кора переробляє дані, отримані з ЛКТ.

4) Для збудження клітин зорової кори потрібно більш тонкий механізм, що відображається на анатомічних особливостях самих клітин і на сигналах, які необхідні для їх збудження. Існує декілька типів цих клітин: *прості* клітини реагують тільки на лінійні сегменти, орієнтовані певним чином; *складні* клітини вимагають руху в певному напрямку; *гіперскладні* клітини вимагають, щоб стимули, що потрапляють у середину їх рецептивних полів, були певної довжини. Окрім, цього відомо [7,9], що клітини, які реагують на одну орієнтацію, формують, собою колонку зорової кори (рис. 1). При цьому рецептивні поля прилеглих клітин мають інші орієнтаційні уподобання, які змінюються поступово. Таким чином, серед безлічі орієнтаційних колонок, на рядки тексту найбільш активно проявить себе одна колонка (або група суміжних) певної орієнтації. В цьому випадку множина клітин цієї колонки буде відповідати, в залежності від довжини, або рядку тексту, або вже конкретному слову. Тоді кут нахилу рядків буде відповідати куту обраної колонки, а висота рядка – розмірам ядра клітини.

Формальний опис моделі гангліозних клітин та можливість її застосування докладно розглянуто в [9]. В даній публікації вже

розглядається застосування моделі клітин первинної зорової кори.

## 2. МОДЕЛЬ КЛІТИН ПЕРВИННОЇ ЗОРОВОЇ КОРИ

Оскільки для активізації кліток цього рівня необхідна активність групи суміжних гангліозних кліток [8], то введемо поняття детектора. Сукупність суміжних кліток показана на рис. 1.

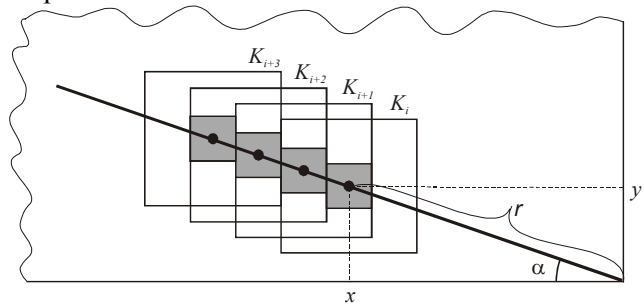


Рис. 1 – Сукупність суміжних клітин

Тоді модель детектора це вираз (1):

$$D^\alpha = \langle K^{\alpha,r}, \alpha, \omega, \ell^*, \delta(D^\alpha) \rangle, \quad (1)$$

де  $K^{\alpha,r} = \langle K^{\alpha,r+2Rk}, K^{\alpha,r+4Rk} \dots K^{\alpha,r+\ell^* \cdot 2Rk} \rangle$ ;

$\alpha$  – кут орієнтації від  $0^0$  до  $175^0$ ;

$\omega$  – ширина детектора, що визначається  $\omega = Rk$ ;

$\ell^*$  – довжина детектора, що визначається кількістю клітин суміжних,  $K^{\alpha,r}$ , але не менш 4;

$\delta(D^\alpha)$  – функція впевненості наявності детектора, обчислюється за формулою (2):

$$\delta(D^\alpha) = \begin{cases} 0, & \text{якщо } \prod_{i \in 1 \dots \ell^*} \mu(K_i) = 0 \\ \frac{1}{\ell^*} \cdot \sum_{i \in 1 \dots \ell^*} \mu(K_i), & \text{якщо } \prod_{i \in 1 \dots \ell^*} \mu(K_i) > 0 \end{cases} \quad (2)$$

де  $\mu(K_i)$  де обчислюється за формулами [9].

На одному проведеному промені може бути кілька детекторів. Нехай кількість цих детекторів буде равно  $n_\alpha$ . Тоді множина всіх детекторів буде  $D^\alpha = \{D_j^\alpha\}$ ,  $j \in 1 \dots n_\alpha$ , що визначаються виразом (1).

На основі поняття детектора опишемо модель різних видів клітин зорової кори.

Так, проста клітина  $S^\alpha$ , що реагує тільки на певну орієнтацію, описується виразом (3):

$$S^\alpha = \langle K^{\alpha,r}, \alpha, \omega, \ell, \delta(S^\alpha) \rangle, \quad (3)$$

<sup>1</sup> рухи очей, які переводять точку фіксації з однієї ділянки зображення на інший, та використовуються переважно для обстеження та вивчення поля зору

де  $\ell \geq \ell^*$ , а  $\delta(S^\alpha) = \min_{j \in 1 \dots n_\alpha} \{\delta(D_j^\alpha)\}$ .

Складні клітини розглянуті не будуть, оскільки вони реагують на рух, а в даній роботі розглядається обробка статичного зображення.

Гіперскладні клітини зорової кори, також як і прості, чутливі до напрямку сигналу, але суттєвої їх особливістю є реакція на певну довжину  $\ell$ . Тому гіперскладна клітина  $G^{\alpha, \ell}$  може бути описана виразом (4):

$$G^{\alpha, \ell} = \langle K^{\alpha, r}, \alpha, \omega, \ell, \delta(G^{\alpha, \ell}) \rangle, \quad (4)$$

де  $\delta(G^{\alpha, \ell}) = \delta(D^\alpha)$ .

В одному напрямку  $\alpha$  гіперскладних клітин однієї довжини може бути кількість  $m_\alpha$ . Тоді множина всіх гіперскладних клітин буде описано  $G^{\alpha, \ell} = \{G_j^{\alpha, \ell}\}$ ,  $j \in 1 \dots m_\alpha$ . Але також може бути довжини  $\ell_1, \ell_2, \dots, \ell_z$ . Тоді множина гіперскладних клітин, реагуючих на певні довжини, висловлюється так

$$G^{\alpha, \ell_i} = \{G_j^{\alpha, \ell_i}\}, \quad i \in 1 \dots z, \quad j \in 1 \dots m_\alpha$$

У підсумку можна побудувати модель зображення у вигляді безлічі орієнтаційних колонок, де кожна колонка описується виразом (5):

$$C^\alpha = \{S^\alpha, \{G_j^{\alpha, \ell_i}\}_{\ell_i}\}, \quad \alpha = 0, 5, 10 \dots 175 \quad (5)$$

Тоді модель "проекції" зображення, тобто ретинотопічної карти описується виразом (6):

$$M^\omega = \langle \{C^\alpha\}_\omega, \omega, \lambda(M^\omega) \rangle, \quad (6)$$

де  $\alpha$  – кут орієнтації від  $0^0$  до  $175^0$ ;

$\omega$  – ширина детектора, що визначається  $\omega = Rk$ ;

$\lambda(M^\omega)$  – функція впевненості наявності найбільшої кількості гіперсложних клітин найбільшої довжини, обчислюється за формулою (7):

$$\lambda(M^\omega) = \left[ \frac{1}{2} (\lambda_1 + \lambda_2) \right]_{\alpha^*}, \quad (7)$$

де  $\lambda_1$  – величина відображення наскільки всі активні релейні клітини сприяли активізації гіперсложних клітин;

$\lambda_2$  – середнє значення впевненості всіх

гіперсложних клітин шириною  $\omega$ ;

$\alpha^*$  – кут орієнтації від 0 до 175, обраний на множині орієнтаційних колонок  $C^\alpha$ .

Тоді  $\lambda_1$  визначається виразом (8):

$$\lambda_1 = \frac{n^{K^\omega}}{n^{G^{\alpha, \ell, \omega}}}, \quad (8)$$

де  $n^{K^\omega}$  – кількість активних гангліозних клітин

$K^\omega$  з розміром ядра  $Rk = \omega$ ;

$n^{G^{\alpha, \ell, \omega}}$  – кількість активних гангліозних клітин з розміром ядра  $Rk = \omega$ , які активізували гіперскладні клітини  $\{G^{\alpha, \ell_i}\}_{\ell_i}$ .

У свою чергу  $\lambda_2$  визначається як (9):

$$\lambda_2 = \frac{1}{m_\alpha} \sum_j \sum_i \delta(G_i^{\alpha, \ell_j, \omega}), \quad (9)$$

де  $m_\alpha$  – загальна кількість гіперскладних клітин

одного напрямку, а кут  $\alpha^*$  визначається з виразу  $\alpha^* = \max_\omega \{ \max_\alpha \{ l_\alpha \} \}$ , де визначається для  $\alpha$  від 0 до 175.

Після визначення для кожної ретинотопічної карти  $M^{\omega_i}$  її функції впевненості потрібно вибрати карту з найбільшою упевненістю. Тоді клітини цієї карти будуть відповідати або смугам, або словам висотою  $\omega$  під кутом  $\alpha$  на зображенні.

### 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Розглянемо приклад застосування наведених моделей за допомогою зображення розміром  $860 \times 227$ , яке наводиться на рис. 2. На рис. 3 показана ретинотопічна карта, отримана в результаті автоматичного визначення оптимального кута нахилу  $\alpha = 5^0$  і ширини  $\omega = 12$  пікселів передбачуваних рядків тексту.

Детальний процес вибору оптимальних параметрів приводиться в роботі [10].

Оскільки моделювання клітин прив'язано до піксельного поля, то знаючи координати розміщення клітин, можна локалізувати і елементи, які вони утворюють, що показано на рис. 4.

Завдяки класичним формулам комп'ютерної графіки (10) та визначеному куту можна виконати поворот зображення.

$$\begin{aligned} x' &= x \cdot \cos(\alpha) - y \cdot \sin(\alpha) \\ y' &= x \cdot \sin(\alpha) + y \cdot \cos(\alpha) \end{aligned} \quad (10) \quad \begin{array}{l} \text{Результат зображення після повороту} \\ \text{наведено на рис.5.} \end{array}$$

де  $(x, y)$  – початкові, а  $(x', y')$  – кінцеві координати.

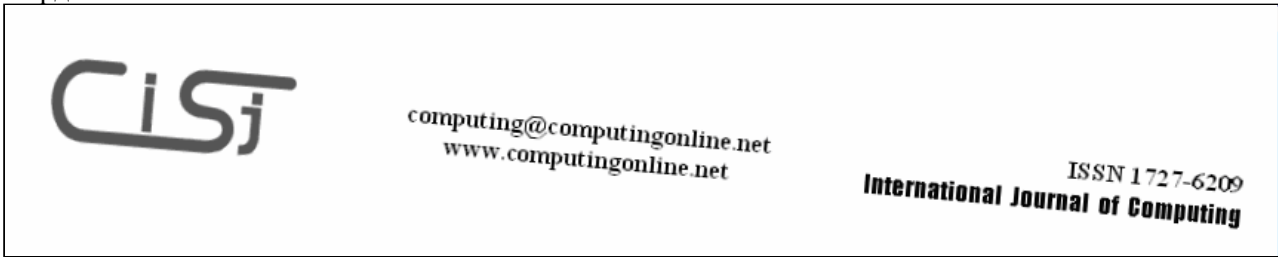


Рис. 2 – Початкове зображення

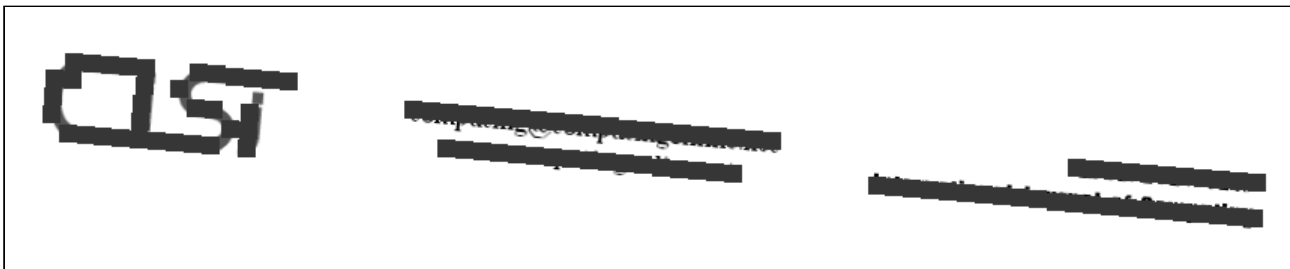


Рис. 3 – Ретинотопічна карта, отримана в результаті автоматичного визначення оптимального кута нахилу  $\alpha=5^{\circ}$  і ширини  $\omega=12$  пікселів можливих рядків тексту.

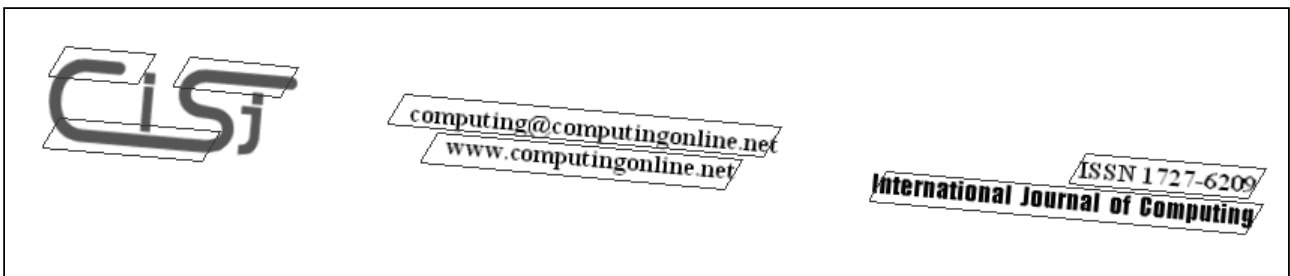


Рис. 4 – Результат локалізації елементів можливих строк тексту.

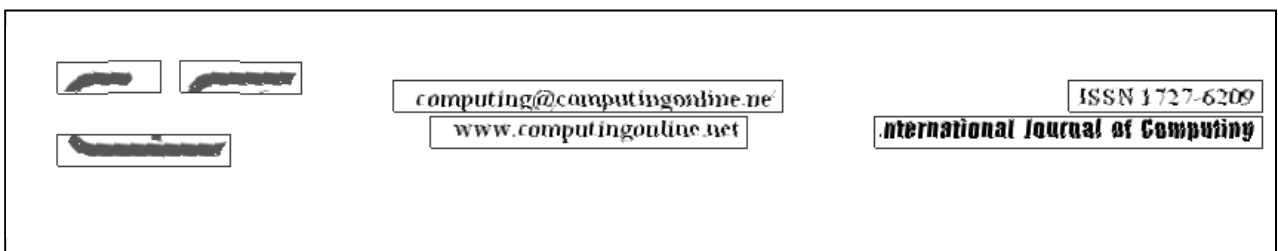


Рис. 5 – Зображення після повороту.

Тепер, коли основні крупні елементи зображення локалізовані, можна виконати більш детальний аналіз, тобто наступна саккада стосується кожного окремого елемента. Можна спробувати розбити кожний із елементів на літери, а можна попередньо проаналізувати, чи має цей елемент характерні для слів особливості. Під характерними особливостями слів мається на увазі наявність послідовності вертикальних паттернів, та низки горизонтальних [11]. Для цього треба знову представити зображення за

допомогою моделі клітин первинної зорової кори тільки зі зменшеним рецептивним полем.

На рис. 6 наведено приклад наявності таких характерних особливостей, а на рис. 7 – їх відсутність. На рис. 8 показаний наступний крок –розбивка на літери, який виконується теж на основі представлення зображення за допомогою клітин первинної зорової кори. Як видно з рис. 8 в трьох крайніх лівих елементах розбивка на літери не виконувалась. Подальша обробка їх також не стосується.

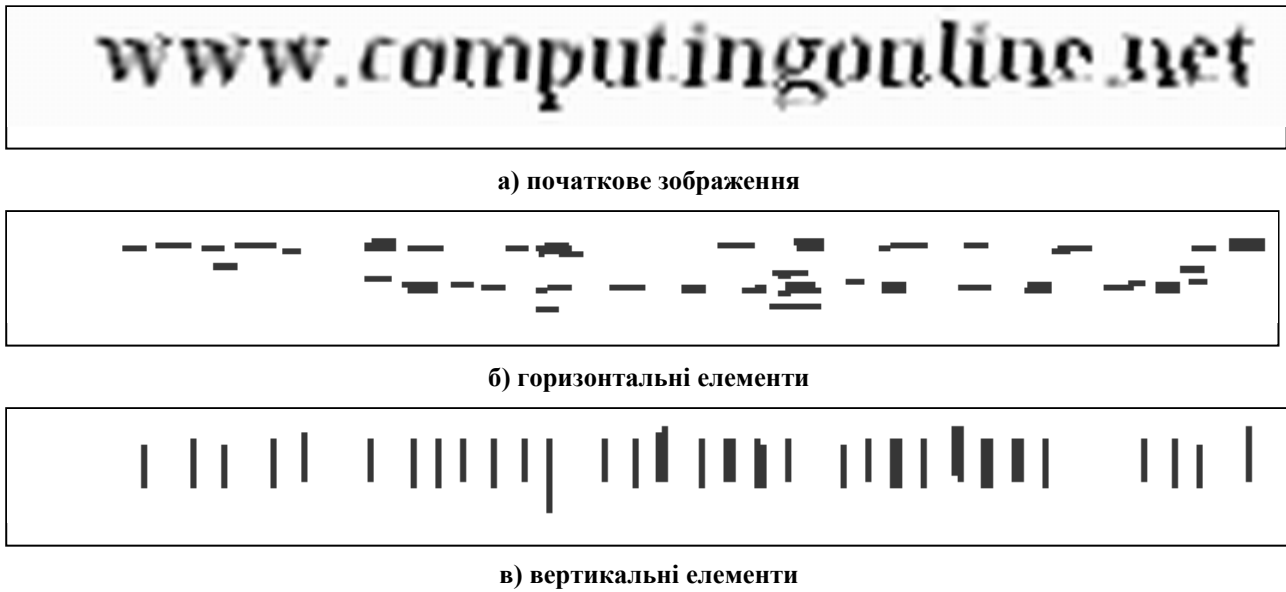


Рис. 6 – Наявність характерних для слів особливостей

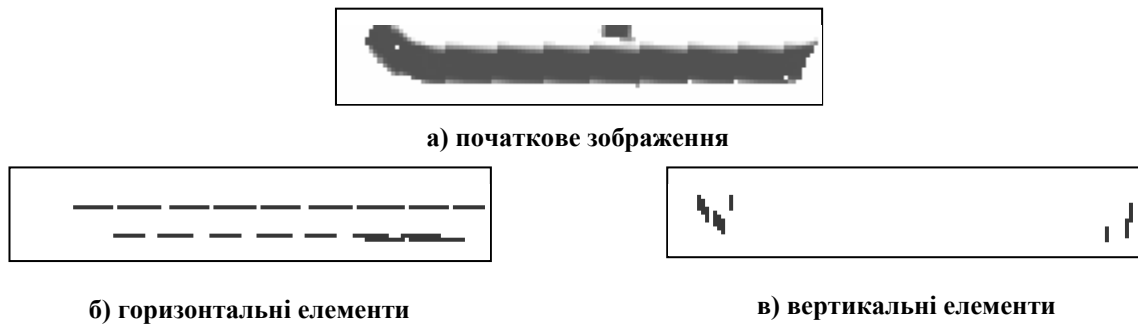


Рис. 7 – Відсутність характерних для слів особливостей

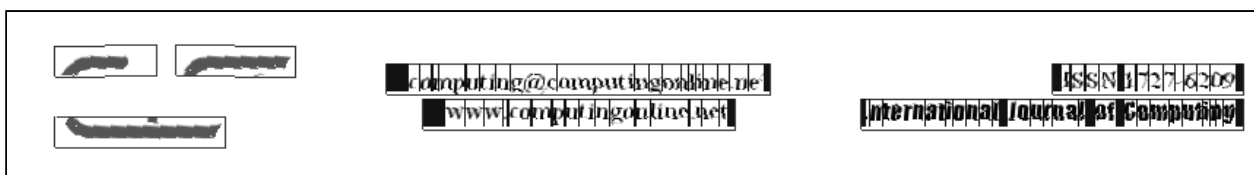


Рис. 8 – Розбивка на літери

Далі для кожного з отриманих блоків необхідно виконати процедуру розширення області для нарядкових або підрядкових елементів, а також процедуру обрізки зайвого фону. Після цього, блок-обводка буде чітко відповідати ширині і висоті символу (символів). На сьогоднішній день досить досліджені ці співвідношення між висотою і шириною для різних символів та шрифтів. На підставі цього прийемо:

1) якщо це співвідношення більше 0,85 то, швидше за все, це один символ;

2) якщо співвідношення висоти до ширини більше 0,65, але менше 0,85, то припустимо, що це може бути один із символів ш, щ, м, ж, ю, або група символів;

3) якщо співвідношення висоти до ширини менше або дорівню 0,65, то це група декількох символів.

З аналізу символів першої групи можна встановити середнє значення ширини символів, і використовуючи ці дані умовно розбити блоки третьої групи. Кінцевий результат локалізації літер показаний на рис. 9.

computing@computingonline.net  
www.computingonline.net

ISSN 1727-6209  
International Journal of Computing

Рис. 9 – Локалізація літер

## 5. ВИСНОВКИ

Проведені дослідження підтвердили можливість використання знань про будову зорової системи, а саме про первинну обробку інформації на рівні клітин зорової кори, для обробки зображення з метою виявлення елементів тексту. Використання розробленої моделі клітин первинної зорової кори дозволяє представити зображення локалізованого символу у вигляді множини елементів, подальше групування яких дозволить виконати ідентифікацію.

## 6. СПИСОК ЛІТЕРАТУРИ

- [1] В. Л. Арлазаров. *Адаптивное распознавание символов* [Электронный ресурс]/ В.Л. Арлазаров, В.В. Троянker, Н.В. Котович //Режим доступа к статье: <http://www.cuneiform.ru/tech/adaptive.html>
- [2] А. С. Васюра. *Моделирование нейросети для решения задачи идентификации символов* [Электронный ресурс]/ А.С. Васюра, Т. Б. Мартынюк, Л. М. Куперштейн // Наукові праці ВНТУ – 2007 – № 1 – Режим доступа к статье: [www.nbu.gov.ua/e-journals/VNTU/2007-1/ru/07vasosi\\_ru.pdf](http://www.nbu.gov.ua/e-journals/VNTU/2007-1/ru/07vasosi_ru.pdf)
- [3] Е. И. Ярмошевич. *Функциональная спектральная пространственно-временная модель формирования изображений объектов зрительной системой человека* [Электронный ресурс]/ Е. И. Ярмошевич, Е. Е. Михайлова, М. А. Пономаренко // Вестник ВГУ, серия: системный анализ и информационные технологии. – 2008. – №1 – стр. 74-78. – Режим доступа к статье: – [www.vestnik.vsu.ru/pdf/analiz/2008/01/yarmoshevich.pdf](http://www.vestnik.vsu.ru/pdf/analiz/2008/01/yarmoshevich.pdf)
- [4] J.A. Bednar. *Scaling Self-Organizing Maps To Model Large Cortical Networks* [Электронный ресурс], *Neuroinformatics*, 2:275-302, 2004. Режим доступа: <http://nn.cs.utexas.edu/keyword?bednar:neuroinformatics04> – название экрана
- [5] С.А. Гладилин. *Нейронная сеть, воспроизводящая выходной сигнал ганглиозной клетки* [Электронный ресурс]/ С.А. Гладилин, Д.Г. Лебедев // Информационные процессы – 2005.– том 5, №3.– стр. 258-264. – Режим доступа к статье: <http://www.jip.ru/2005/258-264.pdf>
- [6] Р.Л. Солсо. *Когнитивная психология*, СПб.: “Питер”, 2002. – 592с.
- [7] Х.Р. Шиффман. *Ощущение и восприятие*, СПб.: “Питер”, 2003. – 928с., ил.
- [8] В.Е. Демидов. *Как мы видим то, что видим*, М.: Знание, 1987– 240 с.
- [9] О. П'ятикоп. Застосування когнітивних моделей сприйняття зображення в задачах розпізнавання тексту, *Proceedings of the III Conference on Computer Science and information Technologies (CSIT'2008), Lviv, Ukraine, September 25-27,2008, pp.168-171*
- [10] Е.Е. Пятикоп. Некоторые результаты компьютерных экспериментов локализации строк текста на основе когнитивных моделей восприятия изображения, *Вісник донецького університету, Сер. А: Природничі науки*, Донецьк, 2008., № 2, ч.2, стр. 527-532.
- [11] Zhang L, Lu Y, and Tan C L. Italic Font Recognition Using Stroke Pattern Analysis on Wavelet Decomposed Word Images, *The 17th International Conference on Pattern Recognition, ICPR2004, Cambridge, United Kingdom, 23-26 August 2004, vol.4, pp. 835-838* // Режим доступа к статье [www.comp.nus.edu.sg/~tancl/Papers/ICPR04/Italic\\_DIAL2004.pdf](http://www.comp.nus.edu.sg/~tancl/Papers/ICPR04/Italic_DIAL2004.pdf)



**Олена П'ятикоп**, в 2001 році отримала диплом з спеціальності “Інформатика” Приазовського державного технічного університету м. Маріуполь, працює асистентом на кафедрі інформатики цього навчального закладу, навчається в аспірантурі.

**Наукові інтереси:** методи штучного інтелекту, нечітка логіка, комп'ютерна графіка.