

## ВІДНОВЛЕННЯ ГЕОМЕТРІЇ ДОВІЛЬНО СПОТВОРЕНИХ ЗОБРАЖЕНЬ ДОКУМЕНТІВ ШЛЯХОМ СЕГМЕНТАЦІЇ

*Анотація.* Довільно деформовані зображення документів можуть мати декілька змін нахилу вздовж рядків тексту, а також слів того ж рядку тексту. У даній роботі представлено спосіб на основі сегментації рядків (визначення послідовності маркованих зв'язних компонент) для ефективного відновлення довільно викривлених зображень документа.

*Ключові слова.* Зв'язні компоненти, відновлення деформованого зображення.

*Abstract.* Randomly deformed images of documents may have several changes of slope along the lines of text, and words of the same line of text. This paper presents a method based on the segmentation lines (labeled sequencing connected components) for the efficient recovery of arbitrarily warped document images.

*Keywords.* Connected components, restoring the deformed image.

### Вступ

Для підвищення швидкості та якості оптичного розпізнавання зображень текстових документів важливого значення набуває проблема розпізнавання зображень, отриманих за допомогою цифрових фотокамер. Тексти на таких фото не є прямими та горизонтальними. Спотворення рядків тексту призводять до сильного погіршення якості розпізнавання, що вимагає попередньої геометричної корекції [1-5].

Методи корекції нелінійних геометричних спотворень зображень документів можна поділити на такі 2 групи:

- 1) методи аналізу і випрямлення ліній рядків,
- 2) методи реконструкції поверхні сторінки.

Розглянутий далі метод відноситься до 1 групи методів.

### Виявлення геометрії рядків тексту і слів

Запропонований спосіб відновлення геометрії документів, побудований на таких трьох етапах:

1. виявлення рядків тексту і слів із використанням сегментації (маркування зв'язних компонент),

---

<sup>1</sup> Uniwersytet Warmińsko-Mazurski w Olsztynie

<sup>2</sup> Українська академія друкарства

<sup>3</sup> Львівський Національний університет ім.Івана Франка

2. обертання (поворот) слів і поміщення їх відповідно між верхніми і нижніми базовими лініями слова,
3. відновлення первинного деформованого зображення, керуючись результатом виправлення бінарного зображення.

### **Формування послідовностей маркованих зв'язних компонент**

Для маркування зв'язних компонент зображення документа  $I_g$  необхідно спочатку провести бінаризацію зображення на основі відповідного алгоритму, наприклад, описаного в [1], створюючи чорно-біле зображення, яке позначимо через  $I_b$ .

Усі слова на зображенні  $I_b$  виділяються за допомогою відповідного згладжування (розфокусування) зображення; потім горизонтально сусідні слова послідовно з'єднуються з метою визначення текстових рядків. Опишемо цей процес наступним алгоритмом:

*Крок 1:* Маркування зв'язних компонент.

*Крок 2:* Розраховується гістограма, складена з висот всіх виявлених зв'язних компонент. Максимальне значення гістограми відповідає середній висоті символу  $H$ .

*Крок 3:* Видалення шуму і нетекстових компонент висотою більше  $3H$  або менше  $H/4$  або шириною меншою, ніж  $H/4$ .

*Крок 4:* Застосування горизонтального згладжування (RLSA [11]) з порогом  $H$  і подальшим маркуванням зв'язних компонент з метою виявлення слова.

*Крок 5:* Виявлення першої зв'язної компоненти (виявлення першого «горизонтального» слова) в переборі зверху вниз (рис.1г – верхній рядок). Ця компонента позначається як слово  $W$  з обмежувальними координатами  $(x_1, x_2, y_1, y_2)$  і призначається до першого рядка тексту  $L$ .

*Крок 6:* Пошук слова  $W_r$  з обмежувальними координатами  $(x'_1, x'_2, y'_1, y'_2)$ , що знаходиться на невеликій відстані з правої сторони від слова  $W$ . Це здійснюється наступним чином: з усіх пов'язаних компонент, які задовольняють умові  $[y_1, y_2] \cap [y'_1, y'_2] \neq \emptyset$ , вибирається одна, з найменшою відстанню  $D = x'_1 - x_2$ , але так, щоб виконувалась умова  $0 < D < 5H$ .

*Крок 7:* Повторення кроку 6 для нового слова  $W_r$ , аж поки не буде знайдено нове слово справа.

*Крок 8:* Повторюються кроки 6 і 7 для лівої сторони від слова  $W$ .

*Крок 9:* Позначаються всі слова, знайдені в порядку зліва направо і призначаються до першого рядка тексту. Крім того, встановлюється мітка для того, щоб вказати, що ці слова не будуть брати участь у подальших розрахунках.

Крок 10: Повторяються кроки з 5 по 9 для інших рядків тексту, поки всім словам не буде призначено текстових рядків.

У кінці цієї процедури, кожен рядок тексту розглядається як  $L_i$ , а кожне слово  $j$ , що належить лінії  $L_i$  як  $W_{ij}$ . Приклад функціонування процедури виявлення текстового рядка і слів зображений на рис. 1.

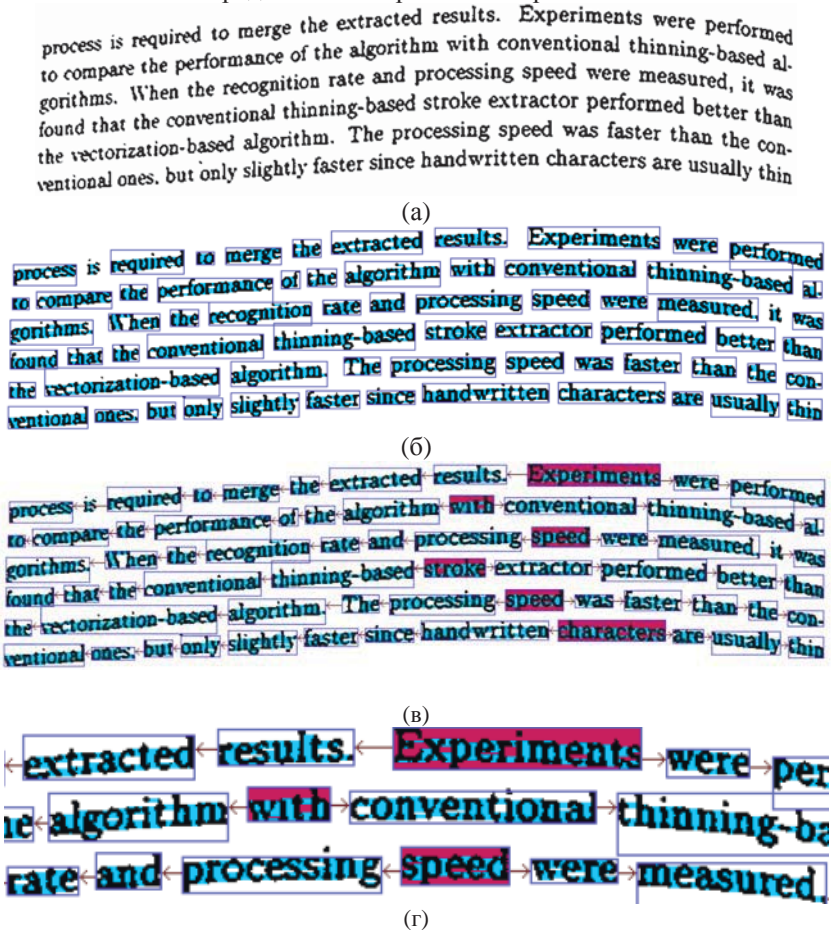


Рис.1 Приклад виявлення текстових ліній та слів: (а) початкове викривлене зображення; (б) результат після горизонтального згладжування для виявлення слів (крок 4); (в) виявлено текстовий рядок після вилучення послідовних правих і лівих сусідніх слів першого виявленого слова (позначено зафарбованим прямокутником) після перебору зверху вниз; (г) – те ж в збільшеному масштабі

Останнім завданням цього етапу є об'єднання перших двох слів (ліворуч) кожного текстового рядка, якщо ширина першого слова менше  $5H$

та ігнорування слів з шириною меншою, ніж  $2H$ . Ці умови є необхідними, оскільки для коротких слів важко виявити нахил слова і тому неможливо перейти до наступного кроку. З іншого боку, важливо мати точне визначення нахилу для першого слова з лівого боку, яке дає напрямок вирівнювання для цілого рядка тексту.

### **Оцінка нижньої і верхньої базової лінії слова**

На цьому кроці використовується лінійна регресія, яка будується на множині точок, які є найнижчими (для нижньої межі слова) чи верхніми (для верхньої межі) чорними пікселями для кожного зображення у колонці. Після цієї процедури, верхня базова лінія слова  $W_{ij}$  визначається наступним чином:

$$y = a_{ij}x + b_{ij},$$

нижня базова лінія слова  $W_{ij}$  визначається як

$$y = a'_{ij}x + b'_{ij}.$$

### **Оцінка виправлення деформованого зображення**

На цьому етапі всі виявлені слова повертаються і впорядковуються так, щоб отримати першу оцінку бінаризованого виправленого зображення.

Нахил кожного слова отримується з відповідних нахилів базових ліній. Нахил верхньої і нижньої базової лінії  $\theta_{ij}^u$  і  $\theta_{ij}^l$  слова  $W_{ij}$  визначається як:

$$\theta_{ij}^u = \arctan(a_{ij}), \quad \theta_{ij}^l = \arctan(a'_{ij}).$$

Оскільки найбільш характерним є, як правило, менший абсолютний нахил, то нахил самого лівого слова може бути визначений як:

$$\theta_{i0} = \begin{cases} \theta_{i0}^u, & \text{якщо } |\theta_{i0}^u| < |\theta_{i0}^l| \\ \theta_{i0}^l, & \text{в іншому випадку} \end{cases}$$

Нахил усіх інших слів є одним з найближчих значень до нахилу попереднього слова:

$$\theta_{ij} = \begin{cases} \theta_{ij}^u, & \text{якщо } |\theta_{ij}^u - \theta_{ij-1}| < |\theta_{i0}^l - \theta_{ij-1}| \\ \theta_{ij}^l, & \text{в іншому випадку,} \end{cases}$$

де  $j > 0$ .

Поворот слова  $W_{ij}(x, y)$  обчислимо наступним чином:

$$\begin{cases} y^r = (x - x_{ij}^{min}) * \sin(-\theta_{ij}) + y * \cos(\theta_{ij}) \\ x^r = x \end{cases}$$

де  $W_{ij}^r(x^r, y^r)$  – повернуте слово,

$x_{ij}^{min}$  – ліва обмежуюча координата слова  $W_{ij}$ .

Після повертання слів, всі слова кожного текстового рядка, окрім самого лівого, повинні бути переміщені вертикально з метою відновлення горизонтального вирівнювання. Повертання і переміщення слова  $W_{ij}(x, y)$  робиться таким чином:

$$\begin{cases} y^{rs} = y^r + d_{ij} \\ x^{rs} = x \end{cases}$$

де  $W_{ij}^{rs}(x^{rs}, y^{rs})$  – повернуте і переміщене слово,

$y^r$  визначається з попередньої системи,

$d_{ij}$  відповідає вертикальному зсуву слова і визначається за наступною формулою:

$$d_{ij} = \begin{cases} y_{i0}^{ru} - y_{ij}^{ru}, & \text{якщо } |\theta_{ij}^u - \theta_{ij-1}| < |\theta_{i0}^l - \theta_{ij-1}|, \\ y_{i0}^{rl} - y_{ij}^{rl}, & \text{в іншому випадку} \end{cases},$$

де  $y_{ij}^{ru} = (a_{ij}x_{ij}^{min} + b_{ij}) * \cos(\theta_{ij})$

і  $y_{ij}^{rl} = (a'_{ij}x_{ij}^{min} + b'_{ij}) * \cos(\theta_{ij})$ .

Причиною наявності двох варіантів переміщення є те, що кожне слово може бути посунуте або до його нижньої базової лінії або до схилу верхньої базової лінії. Отже, воно має бути зміщеним так, щоб його нижня чи верхня базова лінія сумістилась з нижньою чи верхньою базовою лінією лівого слова рядка тексту.

Далі ми зберігаємо всі коефіцієнти перетворення ( $T_{xy}, \theta_{xy}, X_{xy}$ ) для кожного пікселя.

Послідовність кроків може бути описана наступним алгоритмом:

Крок 1: Визначення виправленого зображення:

$$I_{b_{dew}}(x, y) = 0, x \in [1, x_{max}], y \in [1, y_{max}].$$

Крок 2: Визначення коефіцієнтів перетворення

$$T_{xy} = \theta_{xy} = X_{xy} = NULL, x \in [1, x_{max}], y \in [1, y_{max}].$$

Крок 3: Обчислення виправленого бінарного зображення:

$$\forall (x, y) W_{ij}(x, y) = 1 \rightarrow I_{b_{dew}}(x, y') = 1,$$

де  $y' = (x - x_{ij}^{min}) * \sin(-\theta_{ij}) + y * \cos(\theta_{ij}) + d_{ij}$ ,

$T_{xy'} = d_{ij}$ ,

$\theta_{xy'} = \theta_{ij}$ ,

$X_{xy'} = x_{ij}^{min}$ ,

$x_{ij}^{min}$  – ліва обмежуюча межа слова  $W_{ij}$ ,

$\theta_{ij}$  – нахил слова  $W_{ij}$ , який визначається з нахилу верхньої і нижньої базової лінії слова.

### **Заключне відновлення деформованого зображення**

На цьому етапі відбувається повне відновлення оригінального сірого деформованого зображення на основі результатів попереднього етапу виправлення бінаризованого зображення. У зворотному порядку наносяться сірі пікселі для того, щоб отримати остаточне сіре виправлене зображення.

Для всіх пікселів, для яких коефіцієнти перетворення не були виділені, будуть використовуватися коефіцієнти перетворення найближчого пікселя.

Алгоритм:

Крок 1: Отримання виправленого сірого зображення:

$$I_{g_{dew}}(x, y) = 0, x \in [1, x_{max}], y \in [1, y_{max}]$$

Крок 2: Заміна невизначених значень коефіцієнтів перетворення відповідними значеннями найближчих пікселів:

$$\begin{aligned} \text{якщо } I_{b_{dew}}(x, y) = 0 \Rightarrow \\ T_{xy} = T_{x'y'}, \theta_{xy} = \theta_{x'y'}, X_{xy} = X_{x'y'} \end{aligned}$$

де  $x', y'$ :  $\arg \min(|x - x'| + 2 * |y - y'|)$ ,

$$I_{b_{dew}}(x', y') = 1.$$

Мінімальна відстань використовуватися для того, щоб визначити найближчий піксель по осі  $x$  з метою визначення пікселів, що належать до тієї ж лінії тексту.

Крок 3: Остаточне відновлення неспотвореного зображення в градаціях сірого:

$$I_{g_{dew}}(x, y) = I_{g_{dew}}(x, y'),$$

$$\text{де } y' = \frac{y - T_{xy} - (x - X_{xy}) \sin(-\theta_{xy})}{\cos(\theta_{xy})}.$$

**Висновок.** Основним обмежуючим фактором якості виправлення геометричних спотворень в розглянутому алгоритмі є правильне виявлення нахилу окремих слів або виділення базових ліній тексту, що безпосередньо пов'язане з роздільною здатністю зображення. Іншим обмеженням при значних спотвореннях є помилки у визначенні маркованих послідовностей (віднесення слів до сусідніх рядків). Тому необхідні додаткові дослідження для визначення області застосування розробленого методу.

1. *Кульчицька І.О., Тимченко О.В.* Виділення зв'язних компонент на бінарному зображенні в задачах розпізнавання тексту // Зб. наук. пр. ПІМЕ НАН України. – Вип.69. – К.: 2013. – С.165-170.
2. *B. Gatos, I. Pratikakis & S.J. Perantonis*, “Adaptive Degraded Document Image Binarization”, Pattern Recognition, 39, 2006, pp. 317-327.
3. *Федоров А.* Бинаризация черно-белых изображений: состояние и перспективы развития: [Електронний ресурс]. – Режим доступу: <http://itclaim.ru/Library/Books/ITS/wwwbook/ist4b/its4/fyodorov.htm>
4. *Бутаков Е. А. и др.* Обработка изображений на ЭВМ / Е. А. Бутаков, В. И. Островский, И. Л. Фадеев. — М.: Радио и связь, 1987.
5. *Мазуров В. Д.* Математические методы распознавания образов / В. Д. Мазуров // Уч. пособ. 2-е изд., доп. и перераб. – Екатеринбург: Урал. ун-та . – 2010.

*Поступила 14.10.2013р.*