

МОДЕЛЬ ГЕОМЕТРИЧНИХ СПОТВОРЕНЬ ЗОБРАЖЕНЬ ТЕКСТОВИХ ДОКУМЕНТІВ

Анотація. У зв'язку із стрімким розвитком цифрової техніки, важливого значення набуває проблема розпізнавання зображень текстових документів, отриманих за допомогою цифрових фотокамер. Тексти на таких фотографіях погано підходять для розпізнавання і не дуже зручні для читання. Більшість сучасних систем розпізнавання тексту розраховані на те, що рядки тексту на зображенні будуть прямими і горизонтальними. Найменші спотворення рядків тексту призводять до сильного погіршення якості розпізнавання. Такі зображення потребують попередньої геометричної корекції існуючих спотворень. У даній статті подається математична постановка задачі розпрямлення рядків тексту на зображенні.

Abstract. The problem of image recognition text documents obtained by digital cameras becomes important in view of the rapid development of digital technology. Texts on these pictures are not suitable for recognition and are not very easy to read. Modern optical character recognition systems can recognize only images in which the text lines are straight and horizontal. Any distortion of text lines lead to severe degradation of recognition. Geometric image correction is required. This paper is devoted to a problem of de-warping text lines of document images intended to recognition.

Вступ

Процес переведення паперових документів в електронний (цифровий) вигляд здійснюється шляхом сканування чи фотографування. Оптичне розпізнавання символів (англ. Optical Character Recognition, OCR) - конвертація зображень символів і букв в текст, який можна редагувати на комп'ютері [1]. На вхід система оптичного розпізнавання тексту отримує цифрове растрове зображення сканованого або сфотографованого документа, на виході система повинна сформуванати текст, який містить це зображення, у вигляді, придатному для збереження в одному з форматів електронних текстових документів.

Зображення, отримані за допомогою сканера, сучасні системи розпізнавання текстів (такі як FineReader, OmniPage, ReadIris) можуть розпізнати з достатньо високою точністю (яка перевищує 99, 9%). Проте якість отриманих зображень не завжди є високою, тоді виникають проблеми при розпізнаванні документів. На якість документів може впливати дуже

¹ Uniwersytet Warmińsko-Mazurski w Olsztynie

² Українська академія друкарства

³ Львівський Національний університет ім.Івана Франка

багато факторів. Ми розглянемо лише геометричні спотворення зображень.

За останні декілька років з'явилася достатньо велика кількість систем для оцифровки документів, в основі яких лежать цифрові фотоапарати. Система на основі цифрового фотоапарата працює в кілька разів швидше звичайного сканера, так як в ньому не витрачається час на фізичне пересування скануючої планки. Однак використання цифрових фотоапаратів створює багато нових труднощів при розпізнаванні тексту. Геометричні спотворення при фотографуванні можуть з'явитися по декількох незалежних один від одного причин [3]. Це може бути недоліком використовуваної оптики - так званий «ефект бочки». Інша причина полягає в тому, що документ при фотографуванні розташований не під прямим кутом до осі камери, що породжує перспективні спотворення. Спотворення також може бути викликане нерівністю вихідного документа - наприклад, вигином в центрі книги. Всі вищезгадані спотворення можуть комбінуватися між собою в будь-якому поєднанні. В цьому випадку вже неможливо точно описати і передбачити модель спотворення сфотографованого документа. Потрібно зауважити, що проблема геометричних спотворень на зображенні документа існувала ще до початку широкого використання цифрових фотоапаратів. Зокрема, при скануванні товстих книг дуже часто виникають спотворення на зображенні сторінки в області згину книги.

Постановка задачі

Припустимо, що ми маємо паперовий документ з надрукованим на ньому текстом. Позначимо такий документ – PD.

Двоколірне (чорно-біле) растрове зображення – це прямокутна матриця A розміру $m \times n$, кожен елемент a_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) якої (один піксель зображення) являє собою один з двох кольорів (чорний чи білий).

Представимо зображення як функцію на площині в двовимірній декартовій системі координат. Для цього на дійсній площині введемо ортогональну систему координат з осями x і y . Растрове зображення на цій площині можна представити у вигляді двовимірної функції кольору: $C: R^2 \rightarrow \{0; 1\}$:

$$C(x, y) = \begin{cases} 1, & x < 0 \cup x > m \cup y < 0 \cup y > n, \\ 0, & a_{ij} = \text{білий}; i = [x], j = [y], \\ 1, & a_{ij} = \text{чорний}; i = [x], j = [y]. \end{cases}$$

Початок системи координат лежить у лівому нижньому кутку зображення, а координатні осі направлені уздовж вертикальної і горизонтальної межі зображення. Одиниці виміру по осях взяті так, щоб відповідати одному пікселю зображення. Функція зображення дорівнює нулю, якщо значення аргументів функції відповідають білому пікселю зображення і одиниці, якщо чорному. Вважається, що функція зображення дорівнює одиниці, якщо точка лежить поза прямокутником вихідного зображення, тобто зображення знаходиться всередині чорного нескінченного

поля.

Процес перетворення документа в цифровий вигляд включає в себе сканування чи фотографування документа та бінаризацію отриманого зображення. Позначимо такий процес перетворення через ξ ; відображення ξ з паперового документа в функцію двох змінних можна визначити наступним чином:

$$\xi: PD \rightarrow I = \{C: R^2 \rightarrow \{0; 1\}; D = [0; m] \times [0; n] \subseteq R^2\}.$$

Нехай $T_{PD} = \{w_i | i = 1, \dots, n\}$ – текст, написаний на документі PD; w_i – один символ (буква, цифра, пробіл, розділовий знак).

Нехай відображення:

$$\psi: I \rightarrow T_\psi = \{\bar{w}_i | i = 1, \dots, \bar{n}\} \text{ – процес розпізнавання зображення.}$$

Для оцінки ступеня помилки системи розпізнавання введемо поняття редакційної відстані між текстами або відстані Левенштейна [2]. Для цього визначимо наступні операції над текстом:

1. *ADD* (w, i) – додавання символу w в текст на позицію i .
2. *DELETE* (i) – видалення з тексту символу на позиції i .
3. *REPLACE* (w, i) – заміна символу в тексті на позиції i на символ w .

Будь-який текст T_2 можна представити у вигляді суперпозиції операцій над текстом:

$$T_1: T_2 = v_n \dots v_1(T_1); v_i \in \{ADD, DELETE, REPLACE\}.$$

Редакційною відстанню між двома наборами тексту називається мінімальна кількість операцій, необхідна для приведення першого вигляду тексту до другого:

$$ED(T_1, T_2) = \min(n | T_2 = v_n \dots v_1(T_1)).$$

Якістю роботи системи розпізнавання на зображенні називається редакційна відстань між розпізнаним текстом $T_{I\psi}$ і вихідним текстом документу T_{PD} :

$$\delta = ED(T_{I\psi}, T_{PD}).$$

Щоб системи обробки і розпізнавання могли якісно розпізнати текст на зображенні документа, зображення має бути приведенне до деякого стандартного вигляду (повинно бути чорно-білим, на ньому повинен бути текст, написаний чорним на білому фоні, крім тексту не повинно бути ніяких інших чорних об'єктів, текст повинен бути розташований в одну колонку, рядки тексту у цій колонці розташовані горизонтально і не мають бути викривлені). Такий процес попередньої обробки зображення називається

нормалізацією [3]. Всі сучасні системи розпізнавання документів можуть привести зображення документа до необхідного вигляду.

Проте, якщо рядки тексту містять більш складні моделі викривлення, то системи розпізнавання не можуть справитись з цією задачею. Виникає необхідність попередньої обробки зображення.

Нехай $\Phi(t, u): R^2 \rightarrow R^2$ – функція перетворення зображення:

$$\Phi(t, u) = \begin{cases} \Phi_x(t, u) \\ \Phi_y(t, u) \end{cases}; \Phi_x(t, u): R^2 \rightarrow R, \Phi_y(t, u): R^2 \rightarrow R.$$

За допомогою функції перетворення зображення можна з вихідного зображення I отримати змінене зображення \bar{I} :

$$\bar{I} = \Phi^{-1}(I) = \{\bar{C}(x, y) = C(\Phi_x(x, y), \Phi_y(x, y)); \bar{D} = \Phi^{-1}(D)\}$$

Розділимо умовно всю множину зображень документів на дві підмножини – зображення з викривленими рядками $\mathcal{L}_{damaged}$ і зображення без спотворень рядків \mathcal{L}_{normal} .

Задача розпрямлення рядків тексту на зображенні формулюється наступним чином:

Дано зображення I , визначене на прямокутній області $m \times n$, зображення може належати множині зображень з викривленими рядками $\mathcal{L}_{damaged}$ або множині зображень без спотворень рядків \mathcal{L}_{normal} :

$$I = \{C: R^2 \rightarrow \{0; 1\}; D = [0; m] \times [0; n] \subseteq R^2\}, \\ I \in \mathcal{L}_{damaged} \cup I \in \mathcal{L}_{normal}; \mathcal{L}_{damaged} \cap \mathcal{L}_{normal} = \emptyset.$$

Необхідно побудувати таке відображення ϕ початкового зображення I , щоб на виході отримати таке зображення \bar{I} , для якого виконувалися наступні умови:

- якщо на вихідному зображенні були викривлені рядки, то на перетвореному зображенні, де рядки практично рівні якість розпізнавання має бути набагато кращою, ніж на вихідному;
- якщо на вихідному зображенні були прямі рядки, то на перетвореному зображенні якість розпізнавання має бути не гіршою, ніж на вихідному.

Тобто, будується функція ϕ :

$$\phi: I \rightarrow \bar{I}$$

так, щоб

$$I \in \mathcal{L}_{damaged} \Rightarrow$$

$$\delta(I) = ED(\psi(I), T_{PD}) \gg \delta(\bar{I}) = ED(\psi(\bar{I}), T_{PD}) \rightarrow \min,$$

$$I \in \mathcal{L}_{normal} \Rightarrow \delta(I) = ED(\psi(I), T_{PD}) \geq \delta(\bar{I}) = ED(\psi(\bar{I}), T_{PD}).$$

Оцінкою ефективності алгоритму випрямлення є величина похибки, що дорівнює різниці між величиною, що визначає якість розпізнавання тексту до випрямлення рядків тексту

$$\delta = \frac{1}{n} \sum_i \delta_i, \delta_i = ED(T_i, \bar{T}_i); \bar{T}_i = \psi(I_i)$$

і величиною, що визначає якість розпізнавання тексту після випрямлення рядків тексту

$$\hat{\delta} = \frac{1}{n} \sum_i \hat{\delta}_i, \hat{\delta}_i = ED(T_i, \hat{T}_i).$$

Тут $\hat{T}_i = \psi(\hat{I}_i)$ – результат розпізнавання випрямленого зображення \hat{I}_i . Таким чином, маємо, що ефективність алгоритму випрямлення рядків тексту визначається наступною формулою:

$$\Delta = \delta - \hat{\delta}.$$

Задача полягає в тому, щоб мінімізувати величину оцінки похибки $\hat{\delta}$.

Висновки

Задача виправлення геометричних спотворень, наявних у растровому зображенні тексту, полягає в побудові нового бінарного зображення, в якому рядки тексту будуть горизонтальними і всі початкові символи вирівняні по вертикалі. Математична задача полягає у пошуку відповідного перетворення двовимірного простору сторінки в себе, що забезпечує виконання цих умов.

На жаль, неможливо точно оцінити ступінь випрямлення зображення текстового документу, оскільки неможливо отримати вихідний ідеальний документ. Тому для оцінки якості алгоритму доцільно використовувати якість розпізнавання документу до і після випрямлення рядків тексту. Таким же способом можна оцінювати різні методи виправлення геометричних спотворень.

1. *Arms, William Y.*, "Digital libraries", Cambridge, Massachusetts, USA, MIT Press, 2000 <http://www.cs.cornell.edu/wya/diglib/>
2. *Левенштейн В. И.* Двоичные коды с исправлением выпадений, вставок и замещений символов // Доклады Академии Наук СССР, –1965. 163.4:845-848.
3. *Масалович А.А.* Численные методы детектирования и исправления геометрических искажений в изображениях текстовых документов. Диссертация на соискание ученой степени кандидата физико-математических наук / Антон Андреевич Масалович // МГУ имени М.В. Ломоносова. – 2010.

Поступила 7.02.2013р.