

Marcin PAZIO

POLITECHNIKA GDAŃSKA, WYDZIAŁ ELEKTRONIKI, TELEKOMUNIKACJI I INFORMATYKI,
KATEDRA SYSTEMÓW AUTOMATYKI

Lokalizacja tekstu w obrazie

Mgr inż. Marcin PAZIO

W 1989r. ukończył studia na Wydziale Elektroniki Politechniki Gdańskiej. Pracuje w Katedrze Systemów Automatyki wydziału Elektroniki, Telekomunikacji i Informatyki PG. Zajmuje się technikami przetwarzania i analizy obrazów oraz ich zastosowaniami w rozwiązaniach mogących nieść pomoc osobom niewidomym.



e-mail: Marcin.Pazio@eti.pg.gda.pl

Streszczenie

W naturalnym otoczeniu człowieka znajduje się duża ilość łatwo rozpoznawalnej informacji przedstawionej w postaci znaków graficznych i tekstu. Informacja taka jest bardzo przydatna w poruszaniu się w środowisku miejskim. Niestety, osoby z upośledzonymi funkcjami wzroku w sposób oczywisty pozbawione są możliwości korzystania z tego rodzaju przekazu. Istniejące na rynku systemy rozpoznawania tekstu (OCR) nie są niestety dostosowane do rozpoznawania tekstu zawartego w obrazach zarejestrowanych kamerą czy aparatem cyfrowym. W artykule przedstawiono, opracowane w Katedrze Systemów Automatyki Wydziału ETI PG, algorytmy wyszukiwania tekstu w obrazie oraz jego przetwarzania do postaci umożliwiającej rozpoznanie przez system OCR i odczytania treści za pomocą syntezatora mowy.

Słowa kluczowe: analiza obrazu, algorytmy lokalizacji tekstu.

Localization of text in images**Abstract**

The system capable of localizing and reading aloud text embedded in natural scene images can be very helpful for blind and visually impaired persons - providing information useful in everyday life and increasing their confidence and autonomy. Even though the currently available optical character recognition (OCR) programs are fast and accurate, most of them fail to recognize text embedded in natural scene images. The goal of the algorithm described in this paper is to localize text-like image regions and pre-process them in a way that will make OCR work more reliably. The approach described in the paper is based on color image segmentation and segment shape analysis. Preliminary tests have shown that the proposed algorithm offers satisfactory detection rate and is pretty robust to typical text distortions, such as slant, tilt and bend.

Keywords: image analysis, text detection algorithms.

1. Wstęp

Człowiek wprowadza w swoje środowisko, szczególnie miejskie, dużą ilość informacji przedstawionej w postaci znaków graficznych i tekstu. Komunikaty te są pomocne w określeniu punktu w którym się znajdujemy lub pozwalają odnaleźć interesujące nas miejsca. Niejednokrotnie, informacja podawana jest pośrednio – na przykład poszukujemy sklepu znajdującego się obok reklamy, zawierającej określony tekst lub symbol.

Osoby z upośledzonymi funkcjami wzroku w sposób oczywisty pozbawione są możliwości korzystania z tych udogodnień a więc bardzo pomocne dla nich byłoby urządzenie pozwalające na zlokalizowanie napisów oraz znaków i symboli a następnie przekazywanie ich treść za pomocą syntezatora mowy.

Przy projektowaniu systemu odczytującego napisy w naturalny sposób nasuwa się koncepcja połączenia urządzenia (w postaci kamery lub aparatu cyfrowego) rejestrującego obraz, oprogramowania do rozpoznawania pisma OCR (ang. Optical Recognition System) oraz syntezatora mowy. Niestety, w praktyce okazuje się, iż programy OCR, powszechnie stosowane w skanerach, są nie-

skuteczne przy rozpoznawaniu tekstu zawartego w obrazach scen naturalnych, ponieważ tekst taki zwykle umieszczony jest na niejednorodnym tle, ma niedostatecznie duży kontrast w stosunku do tła, ustawiony jest ukośnie w stosunku do krawędzi obrazu i może być zdeformowany geometrycznie na skutek odwzorowania optycznego z przestrzeni trójwymiarowej na dwuwymiarową. Dlatego, dla skutecznego odczytania tekstu, przed jego analizą przy użyciu programu OCR, konieczna jest automatyczna lokalizacja tekstu, jego ekstrakcja oraz korekcja.

2. Metody lokalizacji tekstu w obrazie

W procesie poszukiwania napisów w obrazie możemy założyć, że stanowią one grupy obiektów „literopodobnych” o pewnej regularności przestrzennej oraz wspólnych dla grupy cechach, takich jak barwa, położenie w obrazie i wysokość.

W chwili obecnej znanych jest wiele metod wyszukiwania tekstu w obrazie lub sekwencji wideo. Metody te można podzielić na trzy podstawowe grupy:

- metody działające w oparciu o analizę krawędziową [1],
- metody wykorzystujące analizę tekstur [2],
- metody analizujące segmenty przetworzonego obrazu [3].

W systemie opracowanym w Katedrze Systemów Automatyki zastosowane zostało trzecie podejście, w którym poszukiwanie segmentów spełniających określone kryteria uwzględniające kolorystykę i wzajemne geometryczne relacje, pozwala na lokalizację krótkich sekwencji, składających się z kilku liter.

3. Algorytm wyszukiwania tekstu w obrazie

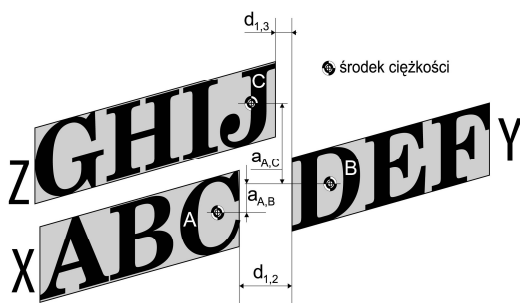
Konstrukcja algorytmu wyszukiwania tekstu oparta została na założeniu, iż napisy rozmieszczone są wzdłuż linii prostych, napisane zostały czcionką o takiej samej wysokości, oraz że wszystkie znaki w obrębie napisu mają ten sam kolor. W praktyce, założenia te spełnione są przez większość tablic informacyjnych i oznaczeń miejsc w postaci nazwy sklepów, punktów usługowych itp.

Opracowany algorytm lokalizacji napisów w obrazie składa się z następujących kroków:

- segmentacja obrazu,
- filtracja segmentów,
- klasteryzacja,
- obróbka końcowa. [4]

Segmentacja obrazu

Zadaniem segmentacji jest podział obrazu na zbiór rozłącznych obszarów zwanych segmentami. W opracowanym algorytmie zastosowana została metoda segmentacji przez rozrost obszaru, do przeprowadzenia której niezbędne jest wskazanie tak zwanych pikseli zarodkowych. Typowanie pikseli zarodkowych odbywa się poprzez odnajdywanie takich pikseli, których parametry odpowiadają maksymalnym wartościom trójwymiarowego histogramu obrazu. Przetwarzanie obrazu oraz analiza histogramu wykorzystuje model barw CIE LAB. Zastosowanie tego modelu barw



Rys. 2. Tworzenie łańcuchów
Fig. 2. Formation of text chains

Aby dwa sąsiednie klastry elementarne zostały połączone w łańcuch, odległości środków ciężkości przylegających elementów sklejanych klastrów d (w poziomie) oraz a (w pionie) muszą być dostatecznie małe w porównaniu ze średnią wysokością segmentów w klastrze. W przykładzie pokazanym na rys. 2. w łańcuch połączone zostaną klastry X oraz Y zaś klastro Z zostanie odrzucony, ponieważ odległość $a_{A,C}$ jest porównywalna z wysokością segmentu. Proces łączenia klastrów trwa do momentu, gdy wartość funkcji kryterialnej, wiążącej wymienione cechy, przekroczy wyznaczoną doświadczalnie wartość progową. Dalszej obróbce poddawane są klastry co najmniej dwuelementowe. Średnia wartość współczynników kierunkowych elementów klastra wyznacza linię bazową tekstu oraz jego nachylenie.

Obróbka końcowa

Zadaniem obróbki końcowej jest takie przetworzenie zlokalizowanych obszarów „tekstopodobnych”, aby umożliwić poprawne działanie programu OCR. Wymaga to:

- korekty położenia napisu w stosunku do krawędzi obrazu,
- ponownej binaryzacji obszarów zawierających tekst,
- eliminacji zniekształceń geometrycznych powstałych w wyniku skrótów perspektywicznych.

Korekta położenia w stosunku do krawędzi obrazu jest możliwa w oparciu o informacje o współczynnikach kierunkowych prostych, wokół których leżą środki ciężkości segmentów stanowiących klastry. Współczynniki te obliczane są jako średnie wartości współczynników kierunkowych wyznaczonych dla klastrów elementarnych.

Ponowna binaryzacja dokonuje się w oparciu o uściślone w procesie klasteryzacji parametry kolorystyczne elementów klastra i jego tła. Lokalnie określona barwa pozwala na zdecydowanie lepsze odtworzenie kształtu znaków, niż przyjęte w procesie segmentacji globalne oszacowanie dla całego obrazu.

W pierwszym etapie obróbki końcowej dla każdego klastra wyznaczany jest opisujący go czworokąt. Segmenty znajdujące się wewnątrz wyznaczonego czworokąta i nie będące elementami klastra, są uznawane za elementy tła. Dla tych wszystkich segmentów ponownie obliczany jest średni kolor, ale z uwzględnieniem wyłącznie tych pikseli, które znajdują się wewnątrz wyznaczonego czworokąta.

W drugim etapie, obszar obrazu wyznaczony przez czworokąt jest obracany o kąt wynikający z wartości współczynników kierunkowych prostych, wokół których leżą środki ciężkości segmentów będących elementami klastra.

W trzecim etapie obrócony obszar obrazu poddawany jest binaryzacji. Do binaryzacji wykorzystywane są parametry kolorystyczne segmentów należących do klastra oraz wyznaczone w pierwszym etapie obróbki końcowej średnie kolory segmentów należących do tła. Dla każdego piksela obróconego fragmentu obrazu znajdowany jest segment najbliższy w przestrzeni Lab spośród segmentów znajdujących się w obszarze wyznaczonego czworokąta. Jeżeli najbliższy segment jest elementem tła, to piksel rów-

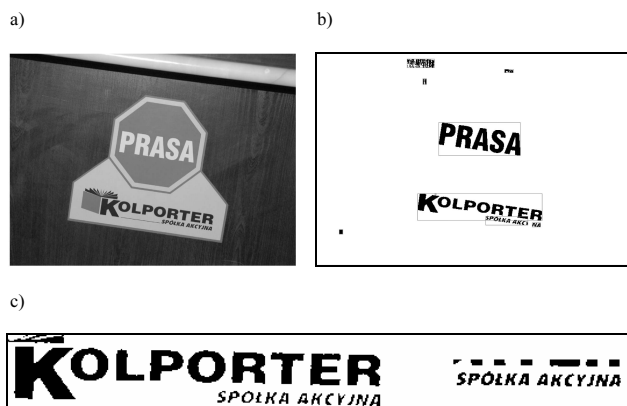
nież uznawany jest za element tła, jeżeli zaś najbliższy segment jest elementem klastra – to piksel uznawany jest za element napisu.

Zniekształcenia geometryczne powstałe w wyniku skrótów perspektywicznych można wyeliminować stosując metodę polegającą na ocenie charakteru zniekształcenia poprzez analizę kształtu obwiedni tekstu oraz analizę histogramów jego rzutów poziomych [5].

4. Wyniki doświadczalne

Opracowane algorytmy przetestowane zostały w procesie wyszukiwania napisów w obrazach niepoddawanych korekcji położenia i zniekształceń, co przedstawiono na rys. 3., rys. 4. i rys. 5. Obszary zawierające zlokalizowane teksty otaczane są ramką a rozpoznane litery i cyfry przedstawiane są w postaci czarnych znaków.

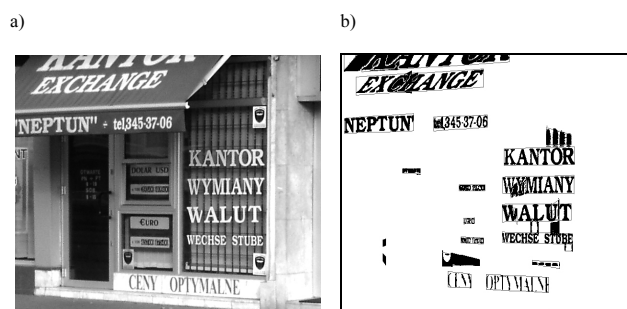
Przykład na zdjęciu z rys. 3. przedstawia dość prostą sytuację - zlokalizowany został napis PRASA oraz KOLPORTER i wyeliminowane zostały obiekty niebędące napisami. Rys. 3c ilustruje wynik działania obróbki końcowej, wraz z prostowaniem tekstu.



Rys. 3. Przykład lokalizacji prostego tekstu na szyldzie. Obraz oryginalny (a), wyodrębniony tekst (b) i obrócone napisy (c)

Fig. 3. Example of a simple board and the result of the text detection. Original image (a) the isolated text (b) and rotated inscriptions

Zawartość obrazu przedstawionego na rys. 4. jest bardziej skomplikowana, ponieważ znajduje się tam stosunkowo dużo tekstu na niejednorodnym tle. Tekst został zlokalizowany i wyodrębniony ale konieczna jest dalsza obróbka, ponieważ nie wszystkie znaki będą mogły być rozpoznane przez system OCR.



Rys. 4. Obraz oryginalny (a) i obraz z wyodrębnionym tekstem (b)

Fig. 4. Original image (a) and the isolated text (b)

Natomiast na rys. 5 przedstawiony jest bardzo przydatny w życiu codziennym przypadek, a mianowicie zlokalizowanie i rozpoznanie numeru domu i nazwy ulicy na tablicy umieszczonej na budynku.



Rys. 5. Wynik działania algorytmu lokalizacji w przypadku napisów na tablicy z numerem domu

Fig. 5. The result of text detection for the board with the house number and the name of the street

Znaki znajdujące się na zdjęciu tej tablicy mają stosunkowo mały rozmiar oraz zmniejszoną, poprzez zanieczyszczenia, czytelność (w dolnej części tablicy). Zastosowany algorytm ponownej binaryzacji pozwolił na prawidłowe odwzorowanie kształtu znaków.

5. Wnioski

Wstępne badania testowe wykazały, że opracowane algorytmy skutecznie wyszukują i odwzorowują napisy na zarejestrowanych obrazach. Istotnym problemem, który będzie przedmiotem dalszych badań, jest nadmiar pozyskiwanej informacji tekstowej. Dlatego konieczne będzie wprowadzenie odpowiedniej filtracji

danych, tak aby osobie niewidomej przekazywane były tylko te informacje, które są w danej chwili istotne.

Praca naukowa finansowana ze środków na naukę w latach 2005-2007 jako projekt badawczy.

6. Literatura

- [1] N. Ezaki, M. Bulacu and L. Schomaker: Text detection from natural scene images: towards a system for visually impaired persons, Proc. 17th IEEE International Conference on Pattern Recognition, ICPR'04, 2004, pp. 683–686
- [2] S. Ferreira, C. Thilou and B. Gosselin: From picture to speech: an innovative application or embedded environment, Proc. of the 14th ProRISC Workshop on Circuits, Systems and Signal Processing, ProRISC'03, 2003
- [3] M. Pazio, M. Niedźwiecki, R. Kowalik, J. Lebidź: Text detection system for the blind, EUSIPCO 2007, pp. 272-276
- [4] M. Pazio: Analiza przydatności wybranych współczynników kształtu do oceny podobieństwa do litery, Zeszyty Naukowe ETI PG, Nr 5, T. 13 (2007), s. 539-546
- [5] Myers, et al., "Recognition of Text in 3-D Scenes," Proc. 4th Symp on Document Image Understanding Technology, Apr. 23-25, 2001, pp. 85–99

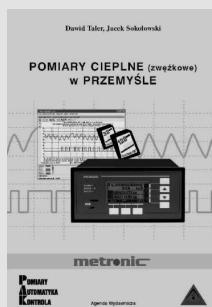
Artykuł recenzowany

INFORMACJE

Książki Wydawnictwa PAK



Książka „Komputerowa technika pomiarowa. Oprogramowanie wirtualnych przyrządów w LabVIEW”, autorstwa Dariusza Świsłowskiego, stanowi kontynuację wydanej w 2002 roku książki „Komputerowa technika pomiarowa w przykładach”. Zasadniczą część książki zajmuje bardzo szczegółowy opis środowiska LabVIEW. Książka przeznaczona jest dla osób rozpoczynających pracę z oprogramowaniem LabVIEW, ale będzie również interesująca dla osób, które już wcześniej używały tego środowiska.



Książka „Pomiary cieplne (zweźkowe) w przemyśle” stanowi nową pozycję literaturową poświęconą pomiarom strumienia masy i ciepła płynów przepływających w przewodach przy użyciu zweźek pomiarowych. Książka przeznaczona jest dla inżynierów i techników zajmujących się zagadnieniami ciepło-przepływowymi w przemyśle, energetyce i ogrzewnictwie. W książce omówiono przyrządy i układy do pomiarów zweźkowych strumienia ciepła, produkowane przez firmę Metronic.



Książka „Regulacja mikroklimatu pomieszczenia” stanowi nowe opracowanie w stosunku do wydanej w 2002 r. książki „Klimat Pomiaru Regulacja”. Prezentuje ona aktualny stan wiedzy na temat mikroklimatu pomieszczeń i nowoczesne rozwiązania systemów pomiarowo - regulacyjno - sterujących oferowanych przez firmę LAB-EL. Rozwiązania te są osiągnięciem polskiej myśli technicznej o standardzie europejskim.

Zamówienia prosimy składać na adresy PAK:

Wydawnictwo PAK
00-050 Warszawa, ul. Świętokrzyska 14A,
tel./fax: 022 827 25 40

Redakcja PAK
44-100 Gliwice, ul. Akademicka 10, p. 30b,
tel./fax: 032 237 19 45
e-mail: wydawnictwo@pak.info.pl