

Tomasz GAĆCIARZ, Krzysztof CZAJKOWSKI  
Politechnika Krakowska, Wydział Fizyki, Matematyki i Informatyki Stosowanej,  
Katedra Teleinformatyki

## KLASYFIKACJA ZDJĘĆ W MULTIMEDIALNYCH BAZACH DANYCH

**Streszczenie.** Praca dotyczy zagadnień szybkiej klasyfikacji zdjęć w oparciu o informacje w nich zawarte. Brane są pod uwagę takie cechy, jak występowanie linii prostych, średnia jasność w poszczególnych obszarach zdjęć, występowanie tekstów oraz obecność twarzy. Na podstawie tych cech, wykorzystując m.in. sieć neuronową, obrazy są klasyfikowane do jednej z kilku kategorii. W pracy przedstawiono implementację omawianego podejścia oraz wyniki przeprowadzonych eksperymentów.

**Słowa kluczowe:** przetwarzanie obrazu, integral image, transformata Hougha, klasyfikator Haara, sieć neuronowa, multimedialna baza danych

## PICTURES CLASSIFICATION IN MULTIMEDIA DATABASES

**Summary.** This paper concerns pictures fast classification on the basis of information included in them. Features as occurrence of straight lines, average brightness in particular areas of picture, text detection as well as human face occurrence are taken into consideration. On the basis of these features, images are classified to one of several categories, by using neural network. Chapter describes implementation and results obtained using mentioned approach.

**Keywords:** image processing, integral image, Hough transform, Haar classifier, neural networks, multimedia databases.

### 1. Wstęp

Wraz ze wzrostem ilości gromadzonych zdjęć oraz możliwościami przechowywania bardzo dużych zbiorów obrazów narasta potrzeba sprawnego ich kategoryzowania i katalogowania. W niniejszej pracy przedstawiono projekt i implementację oprogramowania wspomagającego klasyfikowanie zdjęć. Wyszukiwanie zdjęć w multimedialnych bazach

gającego klasyfikowanie zdjęć. Wyszukiwanie zdjęć w multimedialnych bazach danych jest problemem bardzo złożonym [11, 12, 13]. Wymagane są odpowiednie rozwiązania, które podniosą skuteczność realizacji zapytań do dużych zbiorów tego typu danych oraz zwiększą szybkość całego procesu selekcji. Jedną z możliwości jest analiza zdjęć przeprowadzana w momencie wstawiania ich do bazy, umożliwiająca ich klasyfikację.

Z segregacją zdjęć człowiek radzi sobie bardzo dobrze, jest to jednak proces bardzo czasochłonny. Pomocne wydaje się być wykorzystanie aplikacji automatyzujących tę czynność. W tym celu opracowano odpowiedni zestaw cech, które wyszukiwane są w obrazach. Ich wartości są podstawą podejmowania decyzji o przynależności danego zdjęcia do jednej ze zdefiniowanych kategorii. Decyzja ta podejmowana jest m.in. z pomocą sztucznej sieci neuronowej, którą uczymy na przygotowanym wcześniej zbiorze treningowym. W artykule opisano techniki wykorzystywane do szybkiej ekstrakcji cech z obrazów. Starano się tak dobrać zestaw cech, aby w miarę możliwości efektywnie wykorzystać wszelkie informacje zebrane na każdym kolejnym etapie przetwarzania obrazu. Z uwagi na próbę wyodrębnienia stosunkowo uniwersalnych cech skupiono się na przetwarzaniu zdjęć w skali szarości. Kolor stanowi ważny atrybut opisywanego zagadnienia i oczywiście należy go wziąć pod uwagę podczas realizacji takiego systemu, niemniej jednak my staraliśmy się dobrać cechy niezależne od niego. Bardzo ważnym kryterium doboru cech był również czas przetwarzania obrazów.

## 2. Kategorie zdjęć

Na podstawie analizy przykładowego zbioru, składającego się z kilku tysięcy zdjęć, wyodrębniono zestaw kategorii, do których można było zaliczyć większość spośród nich (w nawiasach podano angielskie nazwy zastosowane w aplikacji):

- Miasto (*city*).
- Zdjęcia robione wewnątrz budynków (*indoor*).
- Zdjęcia krajobrazowe – plaża, morze, góry (*landscape*).
- Dokument (*document*).
- Portret (*portrait*).

Kategorie te dobrano kierując się częstością ich występowania oraz przydatnością dla potencjalnego użytkownika.

### 3. Cechy charakteryzujące zdjęcia

Jak już wcześniej wspomniano zrezygnowano, z cech bazujących na kolorach. Kolejne założenie, jakie przyjęto w związku z doбором cech, to ich "odporność" na rozmiar, rozdzielczość obrazu i duże wahania oświetlenia.

Analizując zdjęcia przedstawiające obszary miejskie zwrócono uwagę na występowanie dużej ilości linii prostych. Znaleźć je można w krawędziach budynków, lampach ulicznych, krawędziach chodników i innych obiektów. Ważna okazuje się być ich liczba, długość, koncentracja i orientacja.

Zdjęcia przedstawiające krajobraz – plażę, morze itp. – zwykle nie zawierają dużej ilości linii. Przy detekcji takich zdjęć pomocny może być rozkład jasności. Zdjęcia wykonane na plaży są szczególnie jasne. Dla większości zdjęć typu *landscape* zwykle najjaśniejszy jest obszar górny. Zdjęcia wykonane wewnątrz pomieszczeń są ciemniejsze, przy czym rozkład jasności jest bardziej równomierny niż dla tych typu robionych na zewnątrz. Z analizy zbiorów zdjęć wynika, że nie tylko średnia jasność w poszczególnych obszarach może mieć znaczenie, ale również stopień jej zmienności. Z tych względów zdecydowano się na wstępny podział obrazu na trzy strefy – górną, środkową i dolną, a następnie na niezależną analizę średniej jasności oraz jej zmienności (w postaci odchylenia standardowego) w tych obszarach. Jasny i jednolity obszar górny wskazuje zwykle na zdjęcia typu *landscape* z widocznym horyzontem. Jeżeli na danym obszarze znajduje się wiele zróżnicowanych obiektów, to odchylenie standardowe jasności będzie wysokie.



Rys. 1. Podział obrazu na trzy obszary

Fig. 1. Image divide into three areas

Kolejną cechą, która dostarcza ważnych informacji o zdjęciach, jest występowanie na nich osób. Obecnie znaleźć można już rozwiązania wyszukujące sylwetki ludzkie w obrazach, jednak w naszej pracy skupiliśmy się na wyszukiwaniu twarzy ze względu na dostępność algorytmów (w przyszłości planujemy wzbogacić to podejście o detekcję sylwetki). Znaczenie ma położenie, wielkość i orientacja znalezionych twarzy. Jeśli znaleziona zostanie tylko jedna twarz i zajmować będzie ponad połowę powierzchni obrazu, to najprawdopodobniej będzie to portret.

Obecnie coraz częściej spotkać można zdjęcia dokumentów. W tym przypadku, po uprzedniej binaryzacji adaptacyjnej, obraz poddawany jest procesowi OCR. Rezultaty otrzymane w trakcie tego etapu pozwalają na podjęcie decyzji, czy mamy do czynienia z dokumentem, jeśli znaleziono fragmenty tekstu zorganizowane w linie i odpowiednio jednolite tło, czy też z obrazem, na którym tekst pojawia się sporadycznie, a jego wielkość lub orientacja nie wskazują na dokument. W tym drugim przypadku będzie to najprawdopodobniej zdjęcie typu *city*.

## 4. Wyznaczanie cech

Biorąc pod uwagę mnogość cech, jakie należy znaleźć chociażby w przypadku detekcji twarzy, czy procesu OCR, dużym wyzwaniem jest odpowiednia kolejność przetwarzania obrazu i ich wyodrębniania. Starano się tak dobrać metody wyszukiwania cech, aby na kolejnych etapach wykorzystać maksimum informacji z poprzedniego. Ma to wpływ przede wszystkim na szybkość całego systemu oraz upraszcza kod aplikacji, co przekłada się na mniejszą ilość błędów, czas pracy i koszty budowy całego systemu.

### 4.1. Integral Images

Czynnikiem, który ma decydujący wpływ na szybkość działania opisywanego systemu, jest sposób obliczania średniej jasności oraz zmiany jasności pikseli w dowolnym wybranym obszarze obrazu. Jak wiadomo, te dwa parametry są najczęściej wykorzystywane w aplikacjach związanych z analizą obrazów w skali szarości. Dzięki zastosowaniu tablic sumowalnych (ang. *integral image* [3]) jesteśmy w stanie obliczać średnią jasność pikseli oraz odchylenie standardowe dla dowolnie dużych, prostokątnych obszarów w stałym czasie.

Idea budowy *integral image* polega na obliczeniu dodatkowego obrazu  $i$ , w którym dla każdego piksela o współrzędnych  $(x,y)$  przechowywana jest wartość sumy jasności pikseli na lewo i powyżej niego, łącznie z nim samym (rys. 2a):

$$i(x, y) = \sum_{i=0}^x \sum_{j=0}^y g(i, j), \quad (1)$$

gdzie  $g(i, j)$  są wartościami jasności pikseli obrazu wejściowego typu greyscale.

Wartości *integral image* mogą być obliczone rekurencyjnie podczas jednego przebiegu przez punkty obrazu wejściowego wiersz po wierszu:

$$i(x, y) = g(x, y) + i(x-1, y) + i(x, y-1) - i(x-1, y-1) \quad (2)$$

pamiętając, że  $i(x, -1) = i(-1, y) = 0$ .

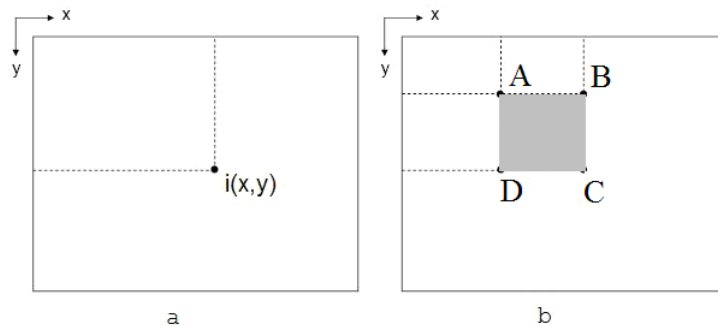
Dysponując raz utworzonym *integral image* w bardzo efektywny sposób otrzymać można sumę wartości jasności pikseli w dowolnym prostokątnym obszarze obrazu wejściowego (rys. 2b – obszar zaciemniony) zgodnie ze wzorem:

$$\sum_{A(x) < x \leq C(x), A(y) < y \leq C(y)} g(x, y) = i(C) + i(A) - i(B) - i(D) \quad (3)$$

Dzieląc otrzymany wynik przez  $n$  (ilość pikseli w obszarze ABCD) otrzymujemy średnią jasność:

$$m_{ABCD} = (i(C) + i(A) - i(B) - i(D)) / n, \quad (4)$$

gdzie  $m_{ABCD}$  to średnia jasność pikseli, a  $n$  liczba pikseli w tym obszarze.



Rys. 2. Sposób obliczania wartości integral image  
Fig. 2. integral image calculation

Analogicznie postępujemy w przypadku obliczania wariancji jasności pikseli w dowolnym obszarze tworząc wcześniej pomocniczy obraz  $ii$  typu *integral image* dla kwadratów jasności pikseli:

$$ii(x, y) = \sum_{i=0}^x \sum_{j=0}^y g^2(i, j) \quad (5)$$

i korzystając ze wzoru na wariancję otrzymujemy:

$$v_{ABCD} = \left( \frac{1}{n} \sum_{A(x) < x \leq C(x), A(y) < y \leq C(y)} g^2(x, y) \right) - m_{ABCD}^2 \quad (6)$$

## 4.2. Detekcja linii prostych - szybka transformata Hougha

W celu znalezienia linii prostych zastosowano zmodyfikowaną transformatę Hougha wzorując się na pracy [1]. Zastosowano w niej odmienny sposób akumulacji przestrzeni parametrów Hougha [8, 9]. Operuje ona na klastrach współliniowych pikseli (rys. 3c), a nie jak w przypadku klasycznej metody na pojedynczych pikselach. Dla każdego z klastrów proces głosowania odbywa się za pomocą eliptycznych kerneli Gaussa, które modelują niedokładności związane z dopasowaniem prostej do klastra.

Rozwiązanie to znacznie zwiększa szybkość procesu (nawet dziesięciokrotnie) oraz przejrzystość utworzonej przestrzeni parametrów, co zwiększa precyzję w wyznaczeniu prostych. Algorytm operuje na obrazie binarnym zawierającym punkty tworzące krawędzie. Do wygenerowania takiego obrazu zastosowano metodę Canny [2]. Kolejne obrazy generowane podczas znajdowania linii prostych przedstawia rysunek 3.

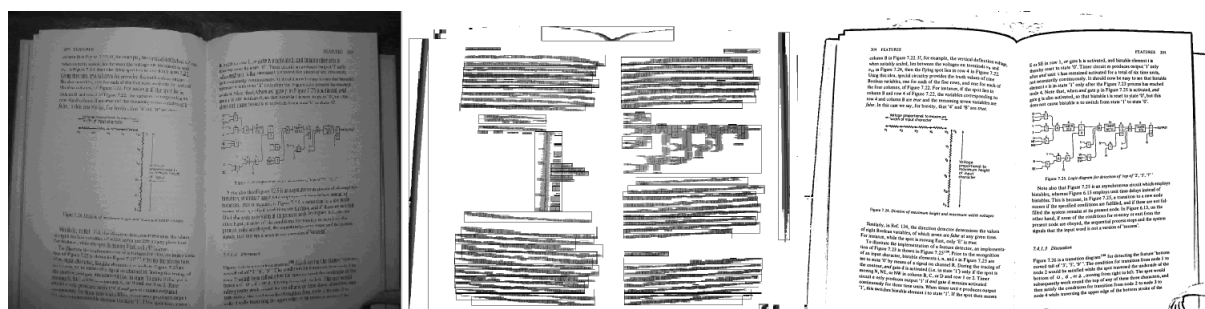


Rys. 3. Kolejne obrazy generowane podczas znajdowania linii prostych  
Fig. 3. Successive images generated during line detection process

## 4.3. Detekcja dokumentów

W celu stwierdzenia, czy zdjęcie zawiera obraz dokumentu, podejmowana jest próba jego rozpoznania w procesie OCR. Pierwszym etapem tego procesu jest segmentacja, w trakcie której staramy się znaleźć wiersze tekstu. To warunkuje przejście do kolejnych etapów.

W celu szybkiego znalezienia wierszy tekstu wykorzystano obliczone wcześniej *integral images*. Dzięki możliwości szybkiej identyfikacji obszarów o dużej wariancji jasności z dobrym przybliżeniem znajduwane są linie tekstu. Po ich wykryciu można przejść do procesu binaryzacji adaptacyjnej. BINARYZACJA przeprowadzana jest tylko w obszarach, w których wykryto linie (rys. 4). Szczegóły metody opisano wcześniej w artykułach [5, 6].



Rys. 4. Kolejne etapy związane z segmentacją i binaryzacją zdjęcia zawierającego dokument  
Fig. 4. Successive steps connected with segmentation and binarization of documents

Po przeprowadzeniu segmentacji znaków w poszczególnych liniach, można przystąpić do ich rozpoznawania. W tym celu zaimplementowano szybki, dedykowany silnik OCR. Bazuje on na prostych drzewach decyzyjnych biorących pod uwagę cechy Haara, obliczane z przygotowanych i opisanych wcześniej *integral images*. Pozwala to na rozpoznanie większości znaków tekstowych (obecnie tylko z alfabetu łacińskiego). Rozpoznawanie zorientowane jest przede wszystkim na szybkość, a nie na precyzję. Czas binaryzacji, segmentacji i rozpoznawania zdjęcia o rozmiarach 3300x2500 (na komputerze z Intel® Penium® M 2.13GHz) wyniósł ok. 1,5 sek.

Jeśli proces OCR potwierdził występowanie pól tekstowych na obrazie, zdjęcie zaliczane jest do kategorii dokument.

#### 4.4. Detekcja twarzy

W celu wykrycia twarzy wykorzystano metodę opisaną w [3, 4], nazywaną detektorem Viola-Jones. Polega ona na przechodzeniu po obrazie wejściowym oknem detektora, w którym badane jest "zadziałanie" zestawu klasyfikatorów zorganizowanych w postaci kaskady.

Bazuje on na prostych cechach Haara [10], które można szybko obliczać mając do dyspozycji *integral image*. Cechy te wykorzystywane są przy budowie kaskady klasyfikatorów trenowanej za pomocą algorytmu Adaboost [4].

W aplikacji wykorzystano już wytrenowaną kaskadę dostępną w ramach przykładów związanych z biblioteką IPP firmy Intel®. Rozwiązanie to ma jednak pewne wady. Cechuje się wysoką skutecznością znajdowania twarzy, ale z drugiej strony znajdują się również obiekty, które nimi nie są (znaczna ilość błędów typu "false positive"). Aby zminimalizować ten negatywny efekt na przyszłość należałoby tak wytrenować klasyfikatory na poszczególnych warstwach kaskady, aby uzyskać możliwie mały błąd fałszywej detekcji, nawet kosztem braku detekcji w niektórych przypadkach. Innym praktycznym rozwiązaniem tego problemu może być wykorzystanie informacji związanych z kolorem na zdjęciach. Na jego podstawie można eliminować przypadkowe rejony znalezione jako twarze, a nie posiadające koloru skóry.

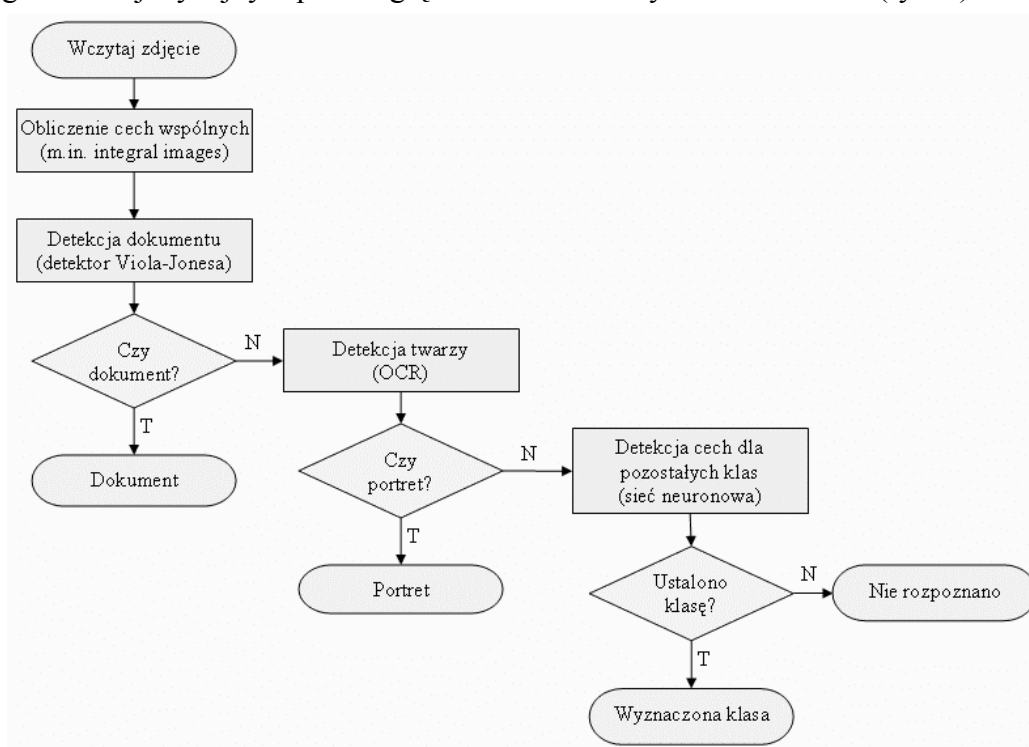
## 5. Wyniki eksperymentów

W celu przeprowadzenia eksperymentów przygotowano zestaw próbek, w którego skład wchodziło po 300 zdjęć należących do każdej kategorii. Przyjęto, że wszystkie zdjęcia mają prawidłową orientację.

Z uwagi na bardzo dobry stopień detekcji obrazów zawierających dokumenty oraz portrety (odpowiednio 99% i 98%) zdecydowano się rozważać je oddzielnie od trzech pozostałych.

Dodatkowym faktem przemawiającym za tym podejściem jest wykorzystanie niezależnych metod ich detekcji. Pozytywnym aspektem zastosowanego podejścia w kontekście opisywanej aplikacji jest jednak fakt, że algorytmy detekcji twarzy oraz procesu OCR korzystają wspólnie z przygotowanych obrazów typu *integral image*. Z obrazów tych korzystamy również wyznaczając cechy służące do rozróżnienia trzech pozostałych kategorii.

Ten prosty (kaskadowy) model detekcji kategorii zawęży nam przestrzeń poszukiwań, co czyni go bardziej wydajnym pod względem obliczeniowym i dokładności (rys. 5).



Rys. 5. Ogólny schemat algorytmu klasyfikacji  
Fig. 5. General schema of classification algorithm

W celu rozróżnienia kategorii *city*, *indoor* i *landscape* zastosowano sztuczną sieć neuronową [7]. Jako wejście do sieci służył wektor cech składający się z 15 cech (po 5 dla każdego z trzech obszarów zdjęcia – górnego, środkowego i dolnego). Dla każdego z obszarów brano pod uwagę następujące cechy: liczbę linii prostych dłuższych 1/3 wysokości obrazu, liczbę



linii prostych krótszych od  $1/3$  wysokości obrazu, średnią jasności w obszarze, średnią wariancję jasności w obszarze, liczbę twarzy wykrytych w danym obszarze.

Zaimplementowano jednokierunkową sieć neuronową posiadającą dwie warstwy ukryte, złożone z 30 neuronów i uczoną algorytmem wstecznej propagacji błędów. Warstwa wyjściowa zawierała 3 neurony powiązane z rozpatrywanymi kategoriami, które zwracały wartości od 0 do 1. Zdjęcie klasyfikowano do kategorii, dla której wartość zwracana przez neuron wyjściowy przekraczała 0,5, a wartości pozostałych były niższe od tego progu.

Sieć uczone była na losowo wybranym zbiorze  $2/3$  próbek, a testowana na pozostałej części. Średnie wyniki, jakie otrzymano dla dziesięciu eksperymentów uczenia i rozpoznawania na zbiorze testowym (za każdym razem losowano nowy zbiór treningowy i testowy) przedstawiono w tabeli 1.

Tabela 1

Rozpoznawanie zbiorów testowych (sumaryczne wyniki 10 eksperymentów)

Kategorie \ Rozpoznanie	Sity	Indoor	Landscape
City	917	35	4
Indor	32	887	2
Landscape	3	5	958
Nierozpoznane	48	73	36
Skuteczność	92%	89%	96%

Można zauważyć, że sieć dobrze uczy się rozróżniać rozpatrywane trzy kategorie obrazów. Najwięcej trudności sprawiają jej zdjęcia wykonane wewnątrz budynków. Spowodowane jest to dużymi wahaniami jasności w ramach danej kategorii (np. w przypadku zdjęć robionych z lampą błyskową, gdzie tło jest dużo ciemniejsze od obiektów z pierwszego planu).

## 6. Podsumowanie

W artykule zaprezentowano podejście do katalogowania zdjęć na podstawie analizy obrazów. Skupiono się na opracowaniu odpowiedniego sposobu analizy obrazów w celu szybkiego wydobycia cech umożliwiających prawidłową klasyfikację. Zaproponowano wykorzystanie odpowiednich odmian metod przetwarzania obrazu wykorzystujących często *integral image*. Posłużyły one m.in. do znajdowania linii za pomocą szybkiej transformaty Hougha, algorytmu detekcji twarzy Viola-Jones wykorzystującego cechy Haara oraz binaryzacji adaptacyjnej i szybkiego silnika OCR. W celu klasyfikacji wykorzystano sieć neuronową, której topologie określono po serii eksperymentów. Wyniki eksperymentów wykazały przydatność

wybranych cech i prawidłowość metody ich klasyfikacji. Dalsze prace obejmować będą opracowanie kolejnych cech, jakie mogą być wydobywane z obrazu (szczególnie koloru) oraz rozszerzenie analizy o pomocnicze informacje kontekstowe dostępne w bazie danych lub na zdjęciu (np. pola Exif). Planowane jest również rozszerzenie systemu o możliwość przechowywania informacji uzyskanych w procesie klasyfikowania zdjęć (położenie istotnych linii prostych, lokalizacja i liczba twarzy, obecności fragmentów tekstu i ich treści) w celu wykorzystania ich w procesie wyszukiwania.

## BIBLIOGRAFIA

1. Fernandes L. A. F., Oliveira M. M.: Real-time line detection through an improved Hough transform voting scheme. *Pattern Recognition (PR)*, Elsevier, 41:1, 2008, s. 299÷314.
2. Canny J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (1986), s. 679÷714.
3. Viola P., Jones M. J.: Robust real-time face detection. *International Journal of Computer Vision* 57 (2004).
4. Freund Y., Schapire R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory, Eurocolt '95*, Springer-Verlag, 1995, s. 23÷37.
5. Gąciarz T.: Fast adaptive binarization for camera-based document images. *Proceedings of International Conference on Artificial Intelligence, IIS Kraków, 2009*.
6. Gąciarz T., Czajkowski K.: Dewarping of camera based document images based on text lines estimation. *Proceedings of International Conference on Artificial Intelligence, IIS Kraków, 2009*.
7. Tadeusiewicz R., Gąciarz T., Borowik B., Leper B.: Odkrywanie właściwości sieci neuronowych, przy użyciu programów w języku C#. *Polska Akademia Umiejętności*, 2007, s. 426.
8. Duda R. O., Hart P. E.: Use of the Hough transformation to detect lines and curves in pictures. *Communications of the Association for Computing Machinery* 15 (1972), s.11÷15.
9. Matas J., Galambos C., Kittler J.: Robust detection of lines using the progressive probabilistic Hough transform. *Computer Vision Image Understanding* 78 (2000), s. 119÷137.
10. Lienhart R., Maydt J.: An Extended Set of Haar-like Features for Rapid Object Detection. *IEEE ICIP*, 2002.

11. Drabowski M., Czajkowski K.: Zapytania do graficznych baz danych oparte na zawartości – próba klasyfikacji. w: S. Kozielski, B. Małysiak, P. Kasproski, D. Mrozek (red.): Bazy Danych – Modele, technologie, narzędzia – Architektura, metody formalne, bezpieczeństwo, Wydawnictwo Komunikacji i Łączności, Warszawa, 2005, s. 241÷249.
12. Drabowski M., Czajkowski K.: Analiza histogramów dla realizacji zapytań w graficznych bazach danych. w: S. Kozielski, B. Małysiak, P. Kasproski, D. Mrozek (red.): Bazy Danych – Struktury, algorytmy, metody – Architektura, metody formalne i eksploracja danych, Wydawnictwo Komunikacji i Łączności, Warszawa, 2006, s. 145÷152
13. Drabowski M., Czajkowski K.: Analiza Fouriera dla realizacji zapytań w graficznych bazach danych. w: S. Kozielski, B. Małysiak, P. Kasproski, D. Mrozek (red.): Bazy danych – nowe technologie. Bezpieczeństwo, wybrane technologie i zastosowania, Wydawnictwo Komunikacji i Łączności, Warszawa, 2007.

Recenzenci: Dr inż. Adam Duszeńko  
Prof. dr hab. inż. Konrad Wojciechowski

Wpłynęło do Redakcji 21 stycznia 2010 r.

### **Abstract**

Searching images in multimedia databases is a very complex problem. To increase efficiency of realization of queries for large sets of such data and to speedup the whole selection process appropriate solutions are highly demanded.

One possible solution of this problem is to perform image analysis before inserting it to database. This will allow its classification.

It seems to be helpful to use application which automates this process. In order to do this, appropriate set of features has to be extracted from the image as presented in the paper. Based on their values decision about the membership of a given image to one of the defined categories is made. For classification purposes the artificial neural network is used. It is trained on specially selected set of images. This paper focuses first of all on fast features extraction methods. Features such as occurrence of straight lines, parameters connected with illumination and its variance are counted. Fast OCR framework is proposed as well as face detection method. All these methods use the idea of integral images which significantly speed up their calculation.

Experimental results prove usefulness of selected features as well as correctness of classification method. Further studies will concentrate on finding additional feature set, which could be extracted from the image (especially from color domain) as well as extending analysis to additional context information available in the database or in the image (for example Exif fields).

### **Adresy**

Tomasz GAĆIARZ: Politechnika Krakowska, Katedra Teleinformatyki,  
ul. Warszawska 24, 31-155 Kraków, Polska, tga@pk.edu.pl .

Krzysztof CZAJKOWSKI: Politechnika Krakowska, Katedra Teleinformatyki,  
ul. Warszawska 24, 31-155 Kraków, Polska, kc@pk.edu.pl .