

Georgy KUKHAREV¹, Ryszard SAŁUDA¹, Marcin MIKŁASZ^{2,3}, Grzegorz KAWKA³

¹ZACHODNIOPOMORSKI UNIWERSYTET TECHNOLOGICZNY, WYDZIAŁ INFORMATYKI

²ZACHODNIOPOMORSKA SZKOŁA BIZNESU

³P.H.U. TELSAT

Metoda ekstrakcji cech orientowana na sprzętową realizację w zadaniach rozpoznawania obrazów

Prof. dr hab. inż. Georgy KUKHAREV

Obronił pracę doktorską w 1979 r. w Instytucie Mechaniki Precyzyjnej i Optyki w Leningradzie. W 1986 r. obronił pracę habilitacyjną, wyższa komisja atestacji przy Radzie Ministrów ZSRR nadała mu tytuł doktora habilitowanego. Od 1993 r. jest profesorem nadzwyczajnym na Politechnice Szczecińskiej (Wydział Informatyki), gdzie jest kierownikiem Zakładu Przetwarzania i Rozpoznawania Obrazów. Jego obecne zainteresowania naukowe to biometryka oraz rozpoznawanie i detekcja twarzy.

e-mail: gekukharev@wi.ps.pl



Inż. Ryszard SAŁUDA

Student ostatniego roku informatyki. Pracę inżynierską z rozpoznawania wzorców na rynkach finansowych obronił w maju 2007 roku. Jego zainteresowania naukowe związane są z metodami przetwarzania i rozpoznawania obrazów oraz technicznym prognozowaniem zachowania rynków finansowych.

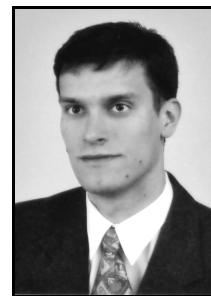
e-mail: rychusa@gmail.com



Dr inż. Marcin MIKŁASZ

Ukończył studia na Wydziale Informatyki Politechniki Szczecińskiej w 2002 roku. Pracę doktorską dotyczącą wyszukiwania podobnych obrazów twarzy w dużych zbiorach danych obronił w marcu 2008 roku. Jego zainteresowania naukowe obejmują przetwarzanie i rozpoznawanie różnorodnych klas obrazów, techniki multimedialne i biometryczne. W chwili obecnej jest adiunktem w Zachodniopomorskiej Szkole Biznesu a jednocześnie pracuje jako specjalista ds. rozpoznawania obrazów w firmie Telsat.

e-mail: miklasz.telsat@555.pl



Mgr inż. Grzegorz KAWKA

W 2005 r. ukończył studia inżynierskie na wydziale elektrycznym Politechniki Szczecińskiej w specjalności elektronika i telekomunikacja. Prowadzi prace wdrożeniowe w dziedzinie nowoczesnych technologii rozpoznawania obrazów w obszarach systemów bezpieczeństwa oraz inteligentnych systemów transportowych (ITS). Od 19 lat prowadzi własną działalność gospodarczą PHU TELSAT, w ramach której realizuje wdrożenia zaawansowanych technologicznie projektów.

e-mail: kawka@555.pl



Streszczenie

Rozpoznawanie obrazów to zadanie realizowane najczęściej przez skomplikowane i złożone metody. Jednak wykorzystanie zestawu prostych i szybkich metod pozwala na dorównanie skutecznością systemom używającym skomplikowanych podejść. Rozwiązanie to ma dodatkowy plus – łatwość implementacji sprzętowej. W artykule przedstawiono podejście analizujące lokalną symetrię obrazu, które pomimo swojej prostoty, wykazało się dużą skutecznością. Przeprowadzone eksperymenty pokazały, że omawiana metoda ekstrakcji cech z obrazu może mieć bezpośrednie zastosowanie w systemach rozpoznawania, a jej prostota pozwala na sprzętową realizację. Dodatkową zaletą prezentowanej metody jest jej inwariantność od oświetlenia twarzy. Dzięki temu istnieje możliwość znaczącej poprawy wydajności całego systemu rozpoznawania.

Słowa kluczowe: rozpoznawanie twarzy, LIFE, ekstrakcja cech.

Feature extraction method directed on hardware realization for image recognition tasks

Abstract

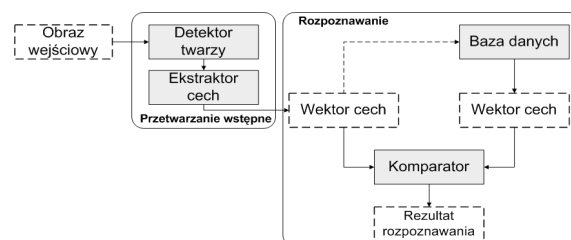
The paper presents the results of investigations concerning face recognition systems based on a simple, fast and efficient feature extractor method. It is based on analysis of the local image symmetry. An additional advantage of the described method is the fact that it is light invariant feature extractor – so it is called LIFE. This benefit (robust on the light conditions) makes it possible to use the method practically as the hardware implementation in real monitoring systems. The idea of LIFE operation is described in Section 2 of the paper. The performed experiments, presented in Section 3 show that LIFE is very efficient in comparison with other simple feature extractor methods – the results of recognition are given in Table 1. In spite of the method simplicity, the proposed approach proved high effectiveness which may be further increased by joining LIFE into a parallel structure with another simple feature extractor (it is described in Section 4). The presented feature extractor enables implementation in hardware system (simplicity and efficiency) such as cameras of the monitoring system. This idea is discussed in the conclusions.

Keywords: LIFE, feature extractor, face recognition.

1. Wstęp

Zadania automatycznej identyfikacji lub weryfikacji tożsamości na podstawie obrazów twarzy są realizowane przez Systemy Rozpoznawania Twarzy (ang. Face Recognition Systems - FaReS) [1, 10]. Liczne praktyczne zastosowania FaReS [1,2,3, 10] powodują ich nieustający rozwój. Coraz częściej metody rozpoznawania twarzy wykorzystywane są w połączeniu z systemami monitoringu. Jednak poważną przeszkodą w rozwoju jest ich duże skomplikowanie i złożoność obliczeniowa. Dlatego ważne jest udoskonalanie i wspomaganie automatycznych systemów rozpoznawania twarzy rozwiązaniami sprzętowymi, pozwalającymi na szybką i skuteczną realizację postawionego zadania.

Na pierwszym etapie FaReS dokonywany jest proces detekcji twarzy na obrazie wejściowym. Uzyskane fragmenty obrazu wejściowego (zawierające twarze) są poddawane procesowi ekstrakcji cech. Ekstrakcja cech jest procesem złożonym obliczeniowo, wymagającym czasochłonnego przetwarzania wyodrębnionych twarzy. Uzyskane cechy poddaje się procesowi porównywania z cechami twarzy przetworzonymi w ten sam sposób i przechowywanymi w bazie systemu. Proces porównywania odbywa się w komparatorze. Strukturę systemu ilustruje rysunek 1:



Rys. 1. Schemat działania FaReS.

Fig. 1. Typical FaReS system.

Systemy monitoringu (miejskiego, stadionowego, w przedsiębiorstwie itp.) to zestawy od kilku do nawet kilkuset kamer połączonych w jedną strukturę. Są one nadzorowane przez operatorów

w centralach obsługi monitoringu. Na przykład monitoring miejski miasta Warszawy składa się z 353 kamer zamontowanych na terenie całego miasta. Nadzór nad nimi pełni Zakład Obsługi Systemu Monitoringu miasta stołecznego Warszawy [4]. W skład systemu wchodzi także 16 Centrów Oglądowych, zlokalizowanych na terenie Komend Rejonowych Policji i Komisariatów Policji, wyposażonych w sprzęt umożliwiający obserwację przez operatorów obrazu z kamer i rejestrację tych obrazów. W Systemie Monitoringu Wizyjnego rejestrowane jest około 10 000 zdarzeń w ciągu roku. Wyposażenie go w system rozpoznawania twarzy znacząco poprawiłoby skuteczność pracy operatorów.

Problemem jest fakt, że z każdej kamery pozyskiwany obraz wymaga przetworzenia (normalizacja obrazu, detekcja twarzy, ekstrakcja cech). Realizowanie tych zadań programowo wymaga budowy dużych klastrów komputerowych dlatego lepszym rozwiązaniem jest wykorzystanie do tego dedykowanego sprzętu zintegrowanego bezpośrednio z kamerą. Dzięki realizacji sprzętowej można również uzyskać przenośne rozwiązanie, którego praktyczne zastosowanie jest znacznie szersze niż stacjonarne, centralnego rozwiązania programowego. W tym kontekście celem prac opisanych w artykule był wybór i sprawdzenie metod ekstrakcji cech twarzy, które możliwe są do zaimplementowania w układach sprzętowych. Przy tym cała uwaga skierowana została na metody możliwe do zastosowania w rzeczywistych, niekontrolowanych warunkach oświetleniowych.

2. Ekstrakcja cech twarzy

Metody wykorzystywane przy ekstrakcji cech w systemach FaReS można podzielić na trzy grupy podejść [1, 2, 3]:

- holistyczne – oparte na wyglądzie, nazywane też globalnymi;
- wykorzystujące trójwymiarowy model głowy;
- analityczne – oparte na mierzalnych cechach fizycznych (antropometrycznych), nazywane też geometrycznymi.

Obecnie najczęściej spotykane są w praktyce rozwiązania, które można zaliczyć do pierwszego typu podejść. Początkowe (nisko-poziomowe) cechy obrazów twarzy są przedstawione przez wartości jasności każdego piksela. Istnieje kilka głównych zalet wykorzystania luminancji w FaReS [1]:

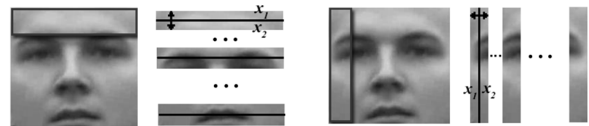
- cechy jasności w naturalny sposób przedstawiają określone obszary i są w dużej mierze niezmiennie przy obrotach i zmianach skali w płaszczyźnie obrazu.
- na podstawie cech jasności można łatwo określić obszary z ostrymi zmianami tej jasności, które odpowiadają granicom charakterystycznych obszarów twarzy – rejonowi oczu, centralnej części oczu, kącików oczu, obszarów brwi, nosa i ust, granic włosów, dolnej części owalu twarzy.
- używając wartości jasności pikseli można zbudować modele obrazów twarzy na podstawie ich aproksymacji w bazie własnych twarzy (ang. EigenFace [1]).

Wektor cech reprezentujący obraz wejściowy w FaReS można uzyskać stosując między innymi metody ekstrakcji cech, które mogą być łatwo zaimplementowane sprzętowo, takie jak [1, 5, 9]: zmniejszenie rozmiaru obrazu (**Skala**) do rozmiaru $m \times n$ i konkatencja wartości wszystkich mn pikseli w postaci wektora cech, pseudolosowy wybór pikseli (**Punkty losowe**) z wejściowego obrazu oraz utworzenie z nich nowego wektora cech (jasności), wykonanie ortogonalnych transformacji obrazu początkowego połączone z tworzeniem wektora cech z zestawu wybranych komponentów widmowych (**widmo DFT**, **widmo DCT**), obliczanie histogramu jasności (**Histogram**) obrazu początkowego i zestawienie jego wartości w wektor cech (histogramowych).

Główną wadą tego typu metod ekstrakcji cech jest znaczący wpływ oświetlenia twarzy na uzyskiwany wektor cech, a tym samym obniżenie skuteczności rozpoznawania osób rejestrowanych w niekontrolowanych warunkach oświetleniowych, np. na podstawie obrazu z monitoringu miejskiego. Dlatego ciągle poszukiwane są metody, które zminimalizują wpływ oświetlenia na skuteczność rozpoznawania, a jednocześnie będą na tyle proste i szybkie, że umożliwią ich sprzętową realizację bezpośrednio w kamerze.

Jedną z takich metod jest metoda analizująca lokalną symetryczność obrazu, będąca przy okazji ekstraktorem cech niezależnym od oświetlenia – metoda LIFE (ang. *light invariant feature extractor*).

Pierwszy krok metody polega na wyodrębnieniu z obrazu twarzy pewnej ilości jego wycinków, ograniczonych odpowiednio przesuwanymi ramkami, zarówno poziomymi (ramka przesuwana z góry na dół obrazu), jak i pionowymi (ramka przesuwana z lewej do prawej strony obrazu). Szerokość ramki poziomej wynosi tyle samo co szerokość obrazu twarzy, natomiast wysokość ok. 15% wysokości obrazu twarzy. Krok przesunięcia ramki wynosi jeden piksel, ale w celu redukcji ilości uzyskanych danych można go zwiększyć. Wycięty za pomocą ramki obszar obrazu dzielony jest na dwa symetryczne obszary. Obrazuje to rysunek 1.



Rys. 2. Idea działania metody LIFE

Fig. 2. Idea of the LIFE method

Jedną z uzyskanych w ten sposób macierzy intensywności pikseli obszaru jest odbijana wierszami (dla ramki poziomej) lub kolumnami (dla ramki pionowej) – ostatni wiersz wędruje na samą górę, przedostatni na drugie miejsce itd. Następnie macierze są odejmowane od siebie i wyliczane w następujący sposób:

$$R(X_1, X_2) = \sqrt{\sum (X_1 - X_2)^2}, \quad (1)$$

gdzie: X_1, X_2 - odpowiednie obszary połowy ramki.

Przy czym wynik odejmowania dwóch pikseli zapisywany jest w zakresie 0...255, tzn.

$$(x_{1(i,j)} - x_{2(i,j)})^2 \in (0...255) = 0, 1, 32, 243, 255 \quad (2)$$

gdzie: $x_{1(i,j)}, x_{2(i,j)}$ - wartości intensywności pikseli z pierwszej i drugiej połowy ramki o współrzędnych i, j .

Uzyskana w ten sposób wartość stanowi jedną z cech (wartości) wektora cech. Z wyliczonych wartości dla wszystkich ramek (z przebiegu poziomego i pionowego) tworzony jest jeden wektor cech analizowanego obrazu. Zapisywany jest on w bazie jako wektor cech opisujący daną twarz i charakteryzuje się znacznie mniejszą wymiarowością od oryginalnego obrazu twarzy, którą opisuje.

Warto zauważyć, że operacje dokonywane w ramach opisywanej metody są prostymi operacjami arytmetycznymi dokonywanymi na macierzach intensywności obrazu. Są one powtarzane przy każdym przesunięciu ramki. W związku z czym istnieje mocne wskazanie na realizację sprzętową opisywanego ekstraktora cech. Implementacja wszystkich wymienionych ekstraktorów cech możliwa jest w układach FPGA. Możliwe jest także wykorzystanie innych procesorów specjalizowanych, na przykład przez procesory przetwarzania obrazu firmy Harris lub Sumitomo [6].

Należy podkreślić, że powyższa metoda umożliwia pracę z dynamiczną, zmienną w czasie bazą danych i nie wymaga procedury rekonstrukcji bazy danych podczas dodawania lub usuwania pojedynczych obrazów z bazy. Po ekstrakcji cech wynikowe wektory cech są bezpośrednio zapisywane w bazie FaReS.

3. Eksperymenty

W prezentowanych eksperymentach rozpoznawania użyte zostały obrazy ze standardowych (i dostępnych publicznie) baz twarzy: ORL [7] i FERET [8] a także utworzoną specjalnie na potrzeby zadania bazę osób zarejestrowanych w rzeczywistych warunkach monitoringu miejskiego. W pracach skupiono się głównie na ekstraktorze cech kosztem uproszczenia komparatora (prosta metryka minimalnoodległościowa L_1 i L_2), jednak uzyskane wyniki rozpoznawania są w pełni akceptowalne.

Baza ORL zawiera po 10 obrazów twarzy o rozdzielczości 112×92 pikseli dla 40 osób. Z bazy FERET wykorzystano 2000 wyodrębnionych obrazów twarzy (po 2 zdjęcia na osobę) – zbiory „Fa” i „Fb”, które przeskalowano do rozdzielczości 112 x 92 pikseli. Baza z monitoringu miejskiego zawierała 85 osób zarejestrowanych w różnym czasie i miejscu (3 różne miejsca, nagrania dokonano na przestrzeni jednego miesiąca). W eksperymentach wykorzystano bazę, której jednemu zdjęciu bazowemu danej osoby odpowiada jedno zdjęcie testowe (z innego okresu).

Wektor cech z testowego obrazu porównywany był z wektorami obrazów innych twarzy na podstawie metryki L_1 :

$$L_1(X, Y) = \sum_{i=1}^n |X_i - Y_i| \quad (3)$$

gdzie: X, Y – porównywane obrazy; X_i, Y_i – wartość i -tego elementu wektora cech obrazu X, Y .

Tabela 1 pokazuje procentową ilość poprawnie rozpoznanych twarzy. Dla bazy ORL podano średnie wyniki uzyskane dla wszystkich możliwości kombinacji 1 obrazu bazowego i 9 testowych (10 kombinacji). Dla wariantu 5:5 podano wyniki dla podziału bazy – 5 obrazów bazowych, 5 testowych. Uzyskane wyniki są lepsze od osiągniętych dla 5 innych metod ekstrakcji cech wspomnianych wcześniej (Skala, Punkty losowe, Widmo DFT, Widmo DCT, Histogram), które przedstawiono w [5]. Parametry poszczególnych ekstraktorów porównawczych zostały dobrane w drodze licznych eksperymentów, opisanych w poprzednich pracach [1, 3, 5, 9]. Testy wykonano dla następujących modeli:

$$\begin{array}{l} \text{FERET}(1000/1/1) \\ \text{ORL}(40/1/9/\text{CV}) \\ \text{ORL}(40/5/5) \end{array} \rightarrow \begin{cases} \{\text{LIFE}(\text{Width} = 12) / L_1 / \text{rank} = 1\} \\ \{\text{DFT} : \text{DIM} = 20 / \text{zigzag} + 1 / L_2 / \text{rank} = 1\} \\ \{\text{Scale} : \text{Wavelet}(14 \times 12) / L_2 / \text{rank} = 1\} \\ \{\text{HIS} : \text{BIN} = 32 / L_2 / \text{rank} = 1\} \\ \{\text{DCT} : \text{DIM} = 8 / \text{zigzag} + 1 / L_2 / \text{rank} = 1\} \\ \{\text{Random}(400) / L_2 / \text{rank} = 1\} \end{cases}$$

Tab. 1. Wyniki przeprowadzonych eksperymentów
Tab. 1. Results of executed experiments

Metoda	BAZA		
	FERET	ORL 1/9	ORL 5/5
LIFE	91%	61%	88%
Skala	74%	47%	86%
Punkty losowe	74%	38%	86%
Widmo DFT	76%	57%	92%
Widmo DCT	79%	52%	90%
Histogram	67%	57%	90%

Dodatkowe badania przeprowadzone dla rzeczywistej bazy wykazały, że metoda LIFE, jest w bardzo wysokim stopniu odporna na zmiany warunków oświetlenia oraz mimiki twarzy, natomiast gorzej sobie radzi z przesunięciami oraz zmianą skali. Dla tej bazy uzyskano 90% poprawnie rozpoznanych twarzy.

4. Zwiększenie skuteczności systemu

Przeprowadzone eksperymenty potwierdziły, że stworzenie skutecznego systemu rozpoznawania twarzy wykorzystującego omówione podejście jest możliwe. Aby zwiększyć skuteczność rozpoznawania a także podnieść pewność uzyskanych wyników sprawdzono system wykorzystujący równoległe kilka ekstraktorów cech. Po licznych eksperymentach wybrano kombinację 2 ekstraktorów cech, charakteryzującą się najwyższą skutecznością: LIFE oraz histogramy wyliczane dla każdej ćwiartki obrazu (4 histogramy z jednego obrazu). Dla każdego z 2 kanałów uzyskano wynik rozpoznawania. Ostateczny wynik rozpoznawania został

wyliczony na podstawie wzoru (4), gdzie każda z metod wnosi inną istotność do ostatecznego wyniku:

$$W(X, Y) = \text{Life} \cdot \text{Hist}^{\frac{2}{5}}, \quad (4)$$

gdzie: *Life* – wyniki osiągnięte dla metody LIFE, *Hist* – wyniki osiągnięte dla metody wykorzystującej histogram.

Obie metody analizują wysoce niezależne od siebie cechy obrazu twarzy, przez co uzyskano znaczny wzrost skuteczności rozpoznawania. Uzyskane wyniki przedstawiono w tabeli 2:

Tab. 2. Wyniki przeprowadzonych eksperymentów dla kombinacji 2 metod
Tab. 2. Results of recognition for combination of 2 methods of the feature extraction

	FERET	ORL 1/9	ORL 5/5
LIFE	91%	61%	88%
LIFE + histogram	96%	72%	95%

5. Wnioski

Wykorzystanie zaproponowanych metod ekstrakcji cech w rzeczywistych warunkach może zostać zrealizowane w następującej strukturze: kamera monitoringu miejskiego rejestruje obraz, który poddawany jest przetwarzaniu wstępnemu i przeprowadzana jest na nim detekcja twarzy. Odpowiada za to sprzętowy podzespół 1. Następnie wyekstrahowany obraz twarzy trafia do podzespołu 2, gdzie przeprowadzana jest ekstrakcja cech twarzy wykorzystująca metodę LIFE. Uzyskany w ten sposób wektor cech jest transmitowany do serwera, gdzie zapisywany jest w bazie danych wraz z niezbędnymi informacjami (data, godzina, miejsce).

W wariacie rozszerzonego systemu, pracującego z 2 ekstraktorami cech, każda z metod ekstrakcji cech realizowana jest w niezależnym podzespole. Dzięki temu można znacząco zwiększyć szybkość przetwarzania danych.

W chwili zajścia zdarzenia w monitorowanym obszarze istnieje możliwość wyselekcjonowania osób z bazy danych zarejestrowanych w okolicy miejsca zdarzenia i przesledzenia ich drogi poruszania się (skąd przyszły, dokąd poszły, o której godzinie zostały zarejestrowane w kolejnych punktach kontrolnych itd.). Ułatwia to rekonstrukcję zdarzeń, a także znacząco zwiększa skuteczność działania sił bezpieczeństwa publicznego i działa prewencyjnie.

6. Literatura

- [1] G. Kukharev, A. Kuźmiński, Techniki biometryczne. Część 1 - Metody Rozpoznawania Twarzy, Pracownia Poligraficzna WI PS, Szczecin, 310 p., 2003.
- [2] G. Kukharev, M. Mikłasz, Face Retrieval from Large Database, Polish Journal of Environmental Studies, vol. 15. No. 4C, 2006, str. 111-114, Świnoujście 6-8.IX.2006.
- [3] G. Kukharev, M. Mikłasz, N. The Binh, Strategia budowy systemów klasy "Face Retrieval", Metody Informatyki Stosowanej Nr 2/2007 (Tom 12), str. 61-72, Szczecin 2007.
- [4] <http://www.zosm.pl/> - strona Zakład Obsługi Systemu Monitoringu miasta stołecznego Warszawy.
- [5] P. Forczmański, G. Kukharev, Comparative analysis of simple facial features extractors, Journal Real-Time Image Proc (2007/1), 239-255.
- [6] K. Wiatr, Akceleracja obliczeń w systemach wizyjnych, Wydawnictwa Naukowo-Techniczne. 486s. 2003.
- [7] www.uk.research.att.com - The ORL Database of Faces.
- [8] <http://www.itl.nist.gov/iad/humanid/colorferer> - The Color FERET Database.
- [9] G. Kukharev, P. Forczmański, A. Nowosielski, Wykorzystanie prostych ekstraktorów cech obrazu w programowo-sprzętowych systemach biometrycznych, Pomiary Automatyka Kontrola, 7 bis 2006, str. 77-79.
- [10] S.Li, A.Jain, ed., Handbook of Face Recognition, Springer, 394 p., New York, 2005.