

Porównanie i ocena skuteczności detekcji i rozpoznawania twarzy w sekwencjach wideo

Streszczenie. Artykuł przedstawia badania związane z identyfikacją osób w sekwencji wideo rejestrowanej za pomocą kamery IP monitoringu wizyjnego. Celem prac było zbudowanie systemu umożliwiającego rozpoznawanie twarzy przy wykorzystaniu minikomputera pracującego pod kontrolą systemu Linux. Przyjęto także założenie, że współpraca użytkowników systemu jest ograniczona, czyli akwizycja zdjęć twarzy nie odbywa się frontalnie zgodnie z zaleceniami norm biometrycznych. Przeprowadzono badania z wykorzystaniem bazy danych ChokePoint w celu dobrania najlepszych parametrów metod wykorzystujących techniki PCA, LDA oraz LBPH. Oprogramowanie przygotowano w języku Python z użyciem bibliotek OpenCV.

Abstract. The paper presents studies related to the people identification in video sequences recorded with a video surveillance IP camera. The aim of this work was to build a system for face recognition using a minicomputer running with Linux. A premise was also a limited collaboration of the users with the system, i.e. that face acquisitions are not necessarily made frontally in accordance with recommendations of the biometric standards. The research was conducted using the ChokePoint database in order to select the best parameters for the analyzed methods: PCA, LDA, and LBPH. The software was prepared in Python using the OpenCV libraries. (**Comparison and assessment of effectiveness of face detection and recognition in video sequences**).

Słowa kluczowe: rozpoznawanie twarzy, identyfikacja biometryczna, PCA, LDA, LBPH.

Keywords: face recognition, biometric identification, PCA, LDA, LBPH.

Wprowadzenie

Współczesne interfejsy człowiek-komputer HMI (ang. *human-machine interface*) opierają się coraz częściej na wykorzystaniu naturalnych kanałów komunikacji człowieka w celu porozumiewania się z maszyną. Wymiana informacji takimi kanałami powinna być szybsza i bardziej naturalna dla użytkownika niż komunikacja z wykorzystaniem paneli sterowniczych. Informacja może być przekazywana mową, gestem i mimiką twarzy jednocześnie [1].

Obraz twarzy człowieka jest cechą biometryczną, której akwizycji łatwo dokonać a następnie wykorzystać do procesu sterowania urządzeniem. Aktualnie typowym zastosowaniem obrazu twarzy w systemach automatyki jest wykorzystanie do celów identyfikacji. Pomimo iż to zadanie jest już realizowane przez różnego rodzaju systemy komercyjne i było wielokrotnie analizowane w publikacjach naukowych za pomocą rozmaitych technik [2], to należy pamiętać, że uzyskanie wysokiej skuteczności działania jest często trudne, co wynika z wrażliwości takich systemów na jakość akwizycji obrazu twarzy – oświetlenie, ustawienie głowy, czy też rozdzielczość [3]. W przypadku analizy sekwencji wideo współpraca użytkownika systemu (czyli osoby która ma być rozpoznana) może być dość ograniczona i zapewnienie akwizycji obrazu twarzy zgodnie z normą ISO/IEC 19792-5 jest często niemożliwe [3]. Warto więc, nawet przy stosowaniu standardowych technik rozpoznawania twarzy, przeprowadzić badania eksperymentalne i dobrać najkorzystniejsze parametry działania wykorzystywanych algorytmów.

W przypadku identyfikacji biometrycznej za pomocą systemów wbudowanych (np. minikomputera Raspberry Pi 2) istotna jest także analiza czasu wykonywania procesu rozpoznawania [4]. Czas ten nie powinien być dłuższy niż odstęp czasowy pomiędzy poszczególnymi ramkami rejestrowanej sekwencji wideo. Jedynie wtedy system będzie umożliwiał pracę w czasie rzeczywistym.

Detekcja i rozpoznawanie twarzy

Wyróżnia się dwie podstawowe koncepcje rozpoznawania twarzy [1, 2]:

- wykorzystanie cech szczególnych twarzy: kształtów poszczególnych elementów twarzy oraz odległości i

kątów pomiędzy punktami charakterystycznymi twarzy, takimi jak oczy, usta, nos, czy kości policzkowe,

- ujęcie holistyczne – polegające na traktowaniu twarzy jako wektora cech, zawierającego informacje analizowane następnie z wykorzystaniem metod algorytmów liniowej.

Podejście wykorzystujące cechy szczególne jest podatne na zmiany wyrazu twarzy oraz kąta robienia zdjęcia, natomiast metody holistyczne są w dużej mierze uzależnione od natężenia i kąta padania światła. W wykorzystywanej w badaniach bibliotece OpenCV stosuje się metody holistyczne.

Proces identyfikacji biometrycznej na podstawie obrazu z twarzą składa się z dwóch etapów: detekcji twarzy oraz klasyfikacji (rozpoznawania). Proces detekcji można przeprowadzić poprzez wykrywanie za pomocą filtru koloru skóry [5], zastosowanie modeli geometrycznych [6] czy też przy użyciu zmodyfikowanej metody Haara (*Haar-like*) [7]. Każda z tych technik posiada określone zalety [3], jednak największą uniwersalnością i skutecznością charakteryzuje się ostatnia wymieniona metoda, która jest wykorzystywana w algorytmie Viola-Jonesa, zaimplementowanym w klasie `CascadeClassifier` biblioteki OpenCV [8]. Algorytm ten bazuje na nadzorowanym uczeniu maszynowym, do którego jest wykorzystywana duża liczba obrazów uczących pozytywnych (zawierających zdjęcie obiektu, na przykład twarzy) oraz negatywnych. Tak uzyskany klasyfikator, w połączeniu z charakterystycznymi dla omawianej metody sposobami przetwarzania obrazu, jest narzędziem pozwalającym na wyszukiwanie obiektów na zdjęciach. Warto pamiętać, że wraz z biblioteką dostarczane są pliki z danymi wyuczonych klasyfikatorów, przygotowane przez twórców narzędzia, dzięki czemu nie ma potrzeby przygotowywania ich we własnym zakresie [8].

Proces klasyfikacji twarzy wyodrębnionych z ramki wideo może być dokonywany za pomocą wielu technik [1, 2]. Jak już wcześniej wspomniano ze względu na użycie bibliotek OpenCV badania eksperymentalne ograniczono do następujących metod:

- *Eigenfaces* – w tej metodzie jest wykorzystywana analiza składowych głównych PCA (ang. *principal component analysis*), którą wykorzystuje się do

zmniejszenia wymiarowości reprezentacji obrazów. Głównym celem jest tutaj przekształcenie dużego zbioru skorelowanych danych w niewielki zestaw parametrów o zróżnicowanej wariancji, opisujących dany obraz.

- Fisherfaces – głównym mechanizmem wykorzystywanym w tej metodzie jest liniowa analiza dyskryminacyjna LDA (ang. *linear discriminant analysis*), która w tym przypadku jest używana do redukcji wymiarowości reprezentacji obrazu z uwzględnieniem podziału na klasy. W przeciwieństwie do poprzedniego algorytmu, kombinacja cech pozwalających w najlepszy sposób rozdzielić zbiory jest poszukiwana przez maksymalizację stosunku rozrzutu międzyklasowego do wewnątrzklasowego a nie przez zwiększanie ogólnej wariancji [8].
- Histogramy lokalnych wzorców binarnych LBPH (ang. *local binary patterns histograms*), które umożliwiają wydobywanie cech z obrazów w formie cyfrowej i służą do opisu kształtów czy pokrycia powierzchni obiektu znajdującego się w przetwarzanym zdjęciu [9]. Metoda ta w przeciwieństwie do dwóch poprzednich nie wykorzystuje metod opartych o statystykę, wobec czego do uzyskania zadowalających wyników

klasyfikacji nie potrzeba dużej liczby obrazów twarzy, wykorzystywanych do tworzenia danych dla klasyfikatora [8]. Jest to pomocne w sytuacjach, gdy trudno zgromadzić dostatecznie dużo odpowiednich zdjęć twarzy osoby, która ma być rozpoznawana.

O ile wybór techniki detekcji twarzy jest jednoznaczny (algorytm Viola-Jonesa), to decyzja o stosowaniu określonego algorytmu klasyfikacji (rozpoznawania) twarzy wymaga analiz eksperymentalnych ze względu na charakter danych wejściowych. W przeprowadzonych badaniach eksperymentalnych zakładano, że rozpoznawanie twarzy będzie odbywać się z wykorzystaniem sekwencji wideo.

Badania eksperymentalne

Baza nagrań testowych

Testy wybranych algorytmów wykrywania twarzy przeprowadzono, wykorzystując bazę nagrań *ChokePoint Dataset*, przeznaczoną do eksperymentów z wykrywaniem i rozpoznawaniem osób [10]. W nagraniu uczestniczyło 19 mężczyzn i 6 kobiet (zestaw G1) oraz 23 mężczyzn i 6 kobiet (zestaw G2). System akwizycji składający się z 3-ch kamer i przykładowe ramki z sekwencji wideo pokazano na rysunku 1.



Rys. 1. System akwizycji 3 kamer i przykładowe ramki z sekwencji wideo (źródło: <http://arma.sourceforge.net/chokepoint/>)

Baza składa się z 48 sekwencji wideo, nagrywanych z szybkością 30 ramek na sekundę. Sumarycznie baza zawiera 176875 ramek wideo, przy czym twarz pojawia się w 64204 ramkach. Każda ramka jest zapamiętana jako plik JPEG o rozdzielczości 800×600 pikseli. Dla każdej sekwencji przygotowano pliki XML z dodatkowymi informacjami, dotyczącymi każdej ramki:

- identyfikator osoby, której twarz wystąpiła w danej ramce,
- współrzędne lewego i prawego oka danej twarzy.

Uzupełnieniem bazy nagrań są też pliki z obrazami twarzy wyciętymi z sekwencji wideo, podzielonymi na katalogi o nazwach odpowiadających numerycznym identyfikatorom osób. Szczegóły dotyczące budowy bazy danych można znaleźć pod adresem [11].

Ekstrakcja twarzy ze zdjęcia

Wykadrowanie obrazu twarzy składa się z następujących etapów:

- poszukiwania obszaru występowania twarzy w analizowanej ramce wideo,

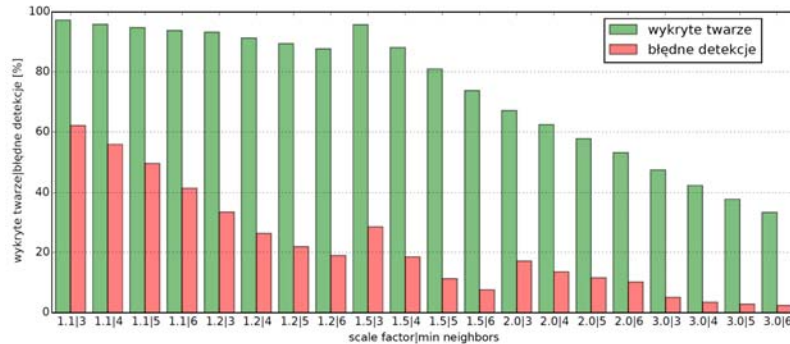
- wyszukiwania par oczu we wszystkich znalezionych fragmentach obrazu, przy czym do dalszego przetwarzania są brane tylko te obszary, w których wykryto dokładnie jedną parę oczu,
- obliczania wymiarów i położenia kadru na podstawie wielkości prostokąta otaczającego znaną parę oczu i parametrów wywołania klasy dotyczących kadrowania,
- wycinania obliczonego kadru z obrazu wejściowego i skalowania go do określonych wymiarów.

Wykrywanie twarzy wykorzystuje obiekt klasy *CascadeClassifier* [8]. Pozyskiwanie danych o obiektach występujących w analizowanej ramce odbywa się poprzez użycie metody *detectMultiScale* z parametrami:

- *image* – obraz wejściowy przeznaczony do analizy, w skali szarości (8 bitów na piksel)
- *scaleFactor* – współczynnik skalowania masek, używanych do wyszukiwania cech w kolejnych iteracjach algorytmu

- `minNeighbors` – współczynnik określający dokładność poszukiwań, przyjmuje wartości całkowite ≥ 1 , im jest większy tym większa jest dokładność
- `minSize`, `maxSize` – parametry określające minimalną i maksymalną wielkość rozpoznawanych obiektów, wyrażone w pikselach
- `objects` – tablica wektorów opisujących prostokątne obszary, otaczające znalezione obiekty.

Na rysunku 2 pokazano zależność skuteczności detekcji twarzy w zależności od współczynników `scaleFactor` oraz `minNeighbors`. Na podstawie tego wykresu do dalszych etapów przetwarzania wybrano wartości `scaleFactor=1.5` oraz `minNeighbors=6`.



Rys. 2. Wykres detekcji prawidłowych i nieprawidłowych

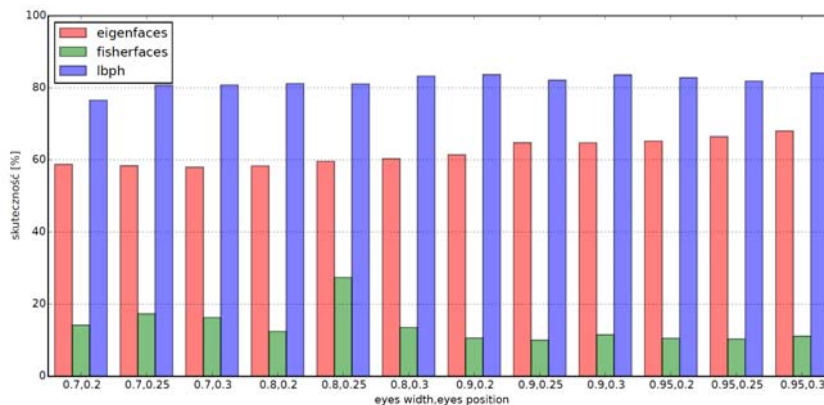
Podobną analizę przeprowadzono dla etapu wyszukiwania par oczu. W tym przypadku najkorzystniejszą parą parametrów okazały się wartości `scaleFactor=1.05` oraz `minNeighbors=3`. W wyniku zastosowania wykrywania par oczu udało się zmniejszyć znacząco liczbę błędnych rozpoznań, ale niestety jednocześnie obniżeniu uległa liczba prawidłowo wykrytych twarzy (spadek z ok. 75% do około 40%) [12].

Rozpoznawanie twarzy

Kluczowym elementem procesu rozpoznawania jest właściwe wykadrowanie wykrytego obrazu twarzy na podstawie parametrów `eyes_position` i `eyes_width`, związanych z położeniem oczu i procentową szerokością twarzy w wykadrowanym obrazie. Na podstawie wstępnych

testów ustalono następujący zestaw wartości parametrów kadrowania sprawdzanych w badaniu: `eyes_position={0.2, 0.25, 0.3}` i `eyes_width={0.7, 0.8, 0.9, 0.95}` [12].

Wyniki eksperymentu pokazano na rysunku 3. Jak nietrudno zauważyć zdecydowanie największą skuteczność rozpoznawania można uzyskać przy zastosowaniu algorytmu LBPH, dla którego wpływ parametrów nie jest znaczący. Najlepsze wyniki (prawie 85%) otrzymano dla zestawu parametrów 0.95 i 0.3. Gorsze wyniki rozpoznawania, w stosunku do LBPH, uzyskano dla algorytmu Eigenfaces. Wyniki dla metody Fisherfaces są natomiast nieakceptowalne.



Rys. 3. Skuteczność rozpoznawania za pomocą 3 metod przy różnych wartościach parametrów kadrowania

Testy z wykorzystaniem własnych nagrań z kamery IP Baza danych *ChokePoint* wykorzystuje obrazy o rozdzielczości 800×600, natomiast dokonano natomiast docelowy system wykorzystuje kamerę IP NBC-255-P pracującą z rozdzielczością 640×480. Sprawdzono więc czy zmniejszenie rozdzielczości wpłynie niekorzystnie na skuteczność rozpoznawania. Po przeskalowaniu obrazów bazy danych *ChokePoint* z rozdzielczości 800×600 na 640×480, skuteczność spadła jedynie o 3%, tj. utrzymuje się na poziomie powyżej 82%. Można więc uznać, że

rozdzielczość standardu VGA jest wystarczająca do budowy systemu.

Testy na własnych sekwencjach wizyjnych, których akwizycji dokonano za pomocą kamery Bosch NBC-255-P (rys. 4), przeprowadzono dla dobranych wcześniej (dla bazy *ChokePoint*) parametrów detekcji twarzy, tj. wyszukiwania par oczu i kadrowania oraz przy użyciu trzech metod rozpoznawania. Prawie frontalne ustawienie kamery spowodowało, że skuteczność rozpoznawania wyniosła ponad 80% dla wszystkich 3 metod. Niemniej jednak, detekcja twarzy wymagała dobrania nieznacznie innych

parametrów skalujących. Zastosowanie komputera z 4-rdzeniowym procesorem AMD pracującym z zegarem 1600 Hz pozwalało uzyskiwać średnią szybkość przetwarzania na poziomie 15 ramek na sekundę (rozdzielczość 640×480 i algorytm rozpoznawania LBPH).



Rys. 4. Akwizycja za pomocą kamery IP Bosch NBC-255-P i przykładowe ramki sekwencji wideo

Podsumowanie

Dobór skutecznych parametrów systemu rozpoznawania twarzy jest zadaniem wymagającym przeprowadzenia testów działania systemu w danej lokalizacji. Krytycznym elementem całego systemu pozostaje etap detekcji twarzy, który, wykorzystując weryfikację na podstawie detekcji oczu, obniża błędną akceptację, ale znacząco obniża także poziom prawidłowego wykrycia.

Należy zauważyć, że w przypadku sekwencji pochodzących z monitoringu wideo największą

skutecznością charakteryzuje się zastosowanie metody LBPH. Niemniej jednak należy pamiętać, że metoda ta jest bardziej czasochłonna niż metody Eigenfaces i Fisherfaces.

Badania zostały sfinansowane ze środków projektu DS-2016.

Autorzy: dr inż. Tomasz Marciniak, prof. dr hab. inż. Adam Dąbrowski, Politechnika Poznańska, Wydział Informatyki, Katedra Sterowania i Inżynierii Systemów, Pracownia Układów Elektronicznych i Przetwarzania Sygnałów, ul. Piotrowo 3a, 60-965 Poznań, E-mail: adam.dabrowski@put.poznan.pl.

LITERATURA

- [1] Daijin Kim, Jaewon Sung, Automated Face Analysis: Emerging Technologies and Research, Premier Reference Source, 2009
- [2] Li S.Z., Jain A. (Eds.), Handbook of Face Recognition, 2nd ed., Wiley, 2011
- [3] Marciniak T., Chmielewska A., Weychan R., et al, Influence of low resolution of images on reliability of face detection and recognition, *Multimedia Tools and Applications*, Volume: 74 Issue: 12 Pages: 4329-4349 Published: JUN 2015.
- [4] Weychan R., Parzych M., Marciniak T., Dąbrowski A., Zagadnienia implementacji metod detekcji osób we wbudowanych systemach CCTV, Materiały XIV Krajowej Konferencji Elektroniki (KKE), Darłówko Wschodnie, 08-12 czerwca 2015, 578-584
- [5] Kapur J.P., Face Detection in Color Images, EE499 Capstone Design Project, University of Washington Department of Electrical Engineering. <http://www.oocities.org/jaykapur/face.html>, dostępne 20.12.2015 r.
- [6] Jesorsky O., Kirchberg K.J., Frischholz R.W., Robust face detection using the Hausdorff distance. BioID AG, Berlin, Germany, 2001
- [7] Haar Feature-based Cascade Classifier for Object Detection, OpenCV 2.0 C Reference, Image Processing and Computer Vision, 2009
- [8] Dokumentacja biblioteki funkcji programistycznych OpenCV, OpenCV dev team, The OpenCV Reference Manual Release 2.4.9.0, 2014
- [9] Ahonen T., Hadid A., Pietikainen M., Description with Local Binary Patterns: Application to Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, (2006), No. 12, 2037-2041
- [10] Wong Y., Chen S., Mau S., Sanderson C., Lovell B.C., Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition, *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, 81-88. IEEE, June 2011
- [11] <http://arma.sourceforge.net/chokepoint/>
- [12] Malina A., Śledzenie i identyfikacja osób na podstawie rozpoznawania twarzy, praca dyplomowa magisterska, promotor: T. Marciniak, Politechnika Poznańska, 2015.