

Andrzej MAJKOWSKI, Marcin KOŁODZIEJ, Remigiusz J. RAK, Michał NASTERNAK
POLITECHNIKA WARSZAWSKA, INSTYTUT ELEKTROTECHNIKI TEORETYCZNEJ I SYSTEMÓW INFORMACYJNO-POMIAROWYCH,
ul. Koszykowa 75, 00-661 Warszawa

Przetwarzanie wstępne i analiza obrazu na użytek lokalizacji twarzy

Dr inż. Andrzej MAJKOWSKI

Absolwent Wydziału Elektrycznego Politechniki Warszawskiej. Od 1995 roku pracownik Instytutu Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych Wydziału Elektrycznego. Stopień doktora nauk technicznych uzyskał w roku 2000. Zainteresowania naukowe: cyfrowe przetwarzanie sygnałów, projektowanie i oprogramowanie systemów pomiarowych.



e-mail: andrzej.majkowski@iem.pw.edu.pl

Prof. dr hab. inż. Remigiusz J. RAK

Absolwent Wydziału Elektroniki Politechniki Warszawskiej (specjalność radiotechnika). Od 1974 roku pracownik Instytutu Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych Wydziału Elektrycznego PW. Uzyskał stopień doktora nauk technicznych (1982), doktora habilitowanego (1998), tytuł naukowy profesora (2004), stanowisko profesora zwyczajnego (2009). Zainteresowania naukowe: cyfrowe przetwarzanie sygnałów, projektowanie systemów pomiarowych, wirtualne laboratorium.



e-mail: remigiusz.rak@ee.pw.edu.pl

Dr inż. Marcin KOŁODZIEJ

Adiunkt w Instytucie Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych na Wydziale Elektrycznym Politechniki Warszawskiej. Ukończony kierunek elektrotechnika o specjalności Automatyka i Inżynieria Komputerowa na Wydziale Elektrycznym Politechniki Warszawskiej. Przewód doktorski: Przetwarzanie analiza i klasyfikacja sygnału EEG na potrzeby interfejsu mózg-komputer. Zainteresowania naukowe: cyfrowe przetwarzanie sygnałów, data-mining, systemy ekspertowe.



e-mail: marcin.kolodziej@ee.pw.edu.pl

Inż. Michał NASTERNAK

Student kierunku elektrotechnika o specjalności Automatyka i Inżynieria Komputerowa na Wydziale Elektrycznym Politechniki Warszawskiej. Praca inżynierska: Detekcja twarzy w obrazie. Zainteresowania naukowe: przetwarzanie i rozpoznawanie obrazów, wizja komputerowa.



e-mail: michnast@wp.pl

Streszczenie

W artykule zaprezentowany jest algorytm automatycznej detekcji twarzy w obrazie statycznym. Detektor ma osiągać najwyższą skuteczność przy znajdowaniu twarzy możliwie niepochyłonych i patrzących na wprost kamery. Wielkość wykrywanych twarzy musi być (z pewnymi odchyleniami) zgodna z rozmiarem twarzy zawartych na obrazach zastosowanych do uczenia klasyfikatora. Obrazy wejściowe mogą być kolorowe lub czarno-białe. Nie ma limitu co do liczby twarzy znajdujących się na obrazie.

Słowa kluczowe: detekcja twarzy, przetwarzanie obrazów, klasyfikacja.

Automatic face detection method

Abstract

The aim of this work is to design and implement a face detection algorithm in static images. The detector have to achieve the best results in finding possible not inclined faces of people looking directly at the camera. The authors have proposed an algorithm which operation is based on the appearance (features) of the face. Block diagram of the proposed face detector is given in Fig. 1. In the first stage, the image containing the face is subjected to preprocessing in which normalization is the most important. Normalization aims to unify a variety of analyzed images. We have used here a conversion of colors to gray levels and stretching and equalization of image histogram. Thus prepared image is processed by the appropriate face detection algorithm, which consists of pre-selection and classification. In order to train the classifier the authors created a database of images consisting of two major categories: containing faces and do not contain faces. As a collection of images that include faces there have been used Olivetti DB ORL database [1]. Final processing step is to get rid of the multiple detection of the same faces. As a result of the algorithm we obtain the location of all faces in the input image (Fig. 4). The size of detected faces should be (with some variations) in accordance with the size of images used to train the classifier. Input images can be color or black and white. There is no limit to the number of faces in an image.

Keywords: face detection, image processing, classification.

1. Wstęp

Detekcja twarzy w obrazie jest zadaniem dość złożonym. Istnieje wiele różnych metod rozwiązania tego problemu. W standardowych,

automatycznych systemach rozpoznawania twarzy najczęściej przetwarzane są obrazy pochodzące z kamery, zawierające centralnie zlokalizowaną twarz jednej osoby. Detekcja i wyodrębnienie z obrazu samej twarzy w takim przypadku nie jest zadaniem trudnym, system „spodziewa się” co będzie na wejściu. Wyzwanie stanowią natomiast obrazy, zarejestrowane w różnych warunkach oświetleniowych, na których znajduje się wiele osób [2]. Mimo wielu lat badań i eksperymentów problem detekcji twarzy do tej pory nie został całkowicie rozwiązany. Nie istnieje idealny detektor twarzy.

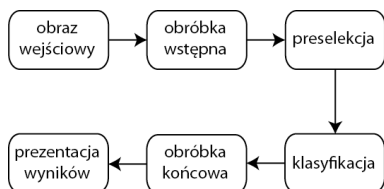
Wyróżnia się cały szereg czynników, które mogą powodować problemy przy detekcji twarzy [3]:

- kąt ustawienia twarzy – zależy od ustawienia człowieka względem kamery, ma duży wpływ na charakterystykę twarzy, zmienia się widoczność i wygląd różnych części (np. oczy, nos) i cech twarzy,
- pochylenie twarzy – wszelkiego rodzaju odchylenie twarzy od pionu,
- wyraz twarzy – twarz to jedno z największych skupisk mięśni w ludzkim ciele, wizerunek osoby może znacznie się zmienić w zależności od wyrażanych emocji (np. uśmiech, smutek),
- przesłonięcie – twarz może być przesłonięta wieloma przeszkodami, (np. ręka, okulary),
- jakość obrazu – zdjęcia niskiej jakości stanowią problem dla detektorów twarzy; brak szczegółów może utrudnić odróżnienie twarzy od elementów tła,
- rasa człowieka – istnieje wiele odcieni koloru skóry, co ma duży wpływ na detekcję metodami segmentacji kolorów.

2. Opracowany algorytm detekcji twarzy

Autorzy zaproponowali algorytm, którego działanie bazuje wyłącznie na wyglądzie (rysach) twarzy. Schemat blokowy zaproponowanego detektora twarzy zamieszczony jest na rys. 1. W pierwszym etapie, obraz zawierający twarze jest poddawany obróbce wstępnej (*preprocessing*), w której jest on normalizowany. Normalizacja ma na celu ujednoczenie analizowanych, różnorodnych obrazów. Autorzy zastosowali tu konwersję barw na odcienie szarości oraz rozciągnięcie i wyrównanie histogramu obrazu. Tak przygotowany obraz przetwarzany jest przez właściwy algorytm

detekcji twarzy, który składa się z *preselekcji i klasyfikacji*. Etap obróbki końcowej ma za zadanie pozbycie się wielokrotnych detekcji tej samej twarzy. Wynik detekcji zawiera lokalizacje wszystkich twarzy zawartych w analizowanym obrazie wejściowym.



Rys. 1. Schemat blokowy detektora twarzy
Fig. 1. Block diagram of the face detector

W systemach detekcji twarzy, w których wykorzystuje się metody bazujące na wyglądzie twarzy, istnieje potrzeba użycia bazy danych do nauczania klasyfikatora. Autorzy stworzyli bazę obrazów składającą się z dwóch głównych kategorii: zawierających twarze oraz nie zawierających twarzy. Jako zbiór obrazów zawierających twarze wykorzystano bazę Olivetti DB ORL [1]. Znajduje się w niej 400 zdjęć, po 10 na każdą z 40 osób (rys.2). Zdjęcia mają wymiary 92×112 pikseli i są zapisane w 256 odcieniach szarości. Sekwencja zdjęć tej samej twarzy zawiera różne pochyleńcia twarzy i różną mimikę. Niektóre twarze zawierają okulary, pokryte są brodą itp. Zaletami tej bazy są: wielkość twarzy (około 80% całego obrazu), centralna lokalizacja oraz jednolite tło. Jednak bardzo niewielka różnorodność kątów obrotu twarzy oraz jednakowe warunki oświetleniowe wszystkich zdjęć wprowadzają znaczące ograniczenie co do ich różnorodności, a co za tym idzie przydatności do uczenia klasyfikatora.



Rys. 2. Przykład zdjęć twarzy z bazy Olivetti DB ORL dla jednej osoby
Fig. 2. Example of face images from Olivetti DB ORL database for one person

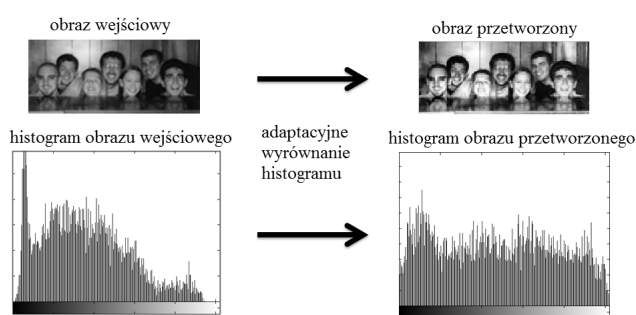
Niewątpliwie, pewien kłopot może sprawić dobór zbioru obrazów niezawierających twarzy. Zbiór ten może zawierać obrazy przedstawiające niemal wszystko, od krajobrazów, poprzez konkretne obiekty, aż po abstrakcyjne grafiki. Istnieją pewne gotowe zbiory, takie chociażby jak w bazie *MIT CBCL FACE DATABASE #1* [4], która oprócz 2429 zdjęć twarzy zawiera aż 4548 obrazów niezawierających twarzy. Jednak dobór odpowiedniego zestawu, spośród elementów drugiego zbioru, zależy od konkretnego detektora, a ściślej rzecz ujmując od zastosowanej metody klasyfikacji klas „twarz” oraz „nie-twarz”. Zbyt duża liczba obrazów niezawierających twarzy, użytych w procesie uczenia, może obniżyć

skuteczność detekcji. Autorzy stworzyli własny zestaw obrazów niezawierających twarzy składający się z 600 zdjęć.

2.1. Przetwarzanie wstępne

Proces przetwarzania wstępnego (*obróbka wstępna*) ma na celu normalizację obrazów wejściowych oraz obrazów uczących. Przebiega on dwuetapowo. Pierwszym etapem jest zamiana obrazu kolorowego na obraz w skali odcieni szarości. W opracowanym algorytmie wszystkie pozostałe procesy przetwarzania obrazu odbywają się na obrazach w skali szarości.

Drugi etap obróbki wstępnej ma na celu uniezależnienie się (w pewnym stopniu) od warunków oświetleniowych zdjęcia. W tym celu przeprowadzane jest adaptacyjne wyrównywanie histogramu (rys. 3). Histogram obrazu tworzony jest dla pełnego zakresu 256-ciu odcieni szarości.



Rys. 3. Poprawa kontrastu za pomocą adaptacyjnego wyrównywania histogramu
Fig. 3. Contrast enhancement by adaptive histogram equalization

2.2. Preselekcja

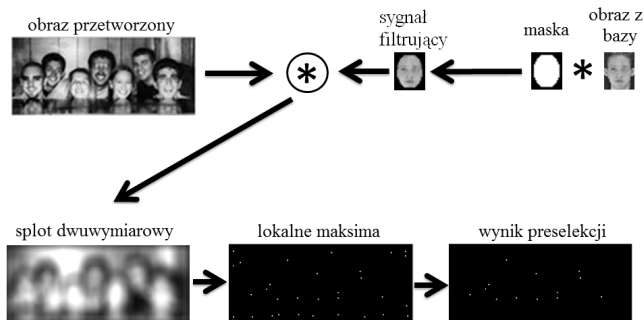
Preselekcja to proces, który ma na celu lokalizację obszarów potencjalnie zawierających twarze. Zwykle algorytm preselekcji „znajduje” takich obszarów dużo więcej niż faktyczna liczba twarzy występujących w obrazie wejściowym. Jednak w sposób znaczący ogranicza to ilość danych podlegających dalszemu przetwarzaniu (detekcji). Głównym zadaniem preselekcji jest redukcja czasu realizacji ostatniego etapu, klasyfikacji.

Schemat działania preselekcji przedstawiony jest na rys. 4. Pierwszym etapem zaimplementowanego algorytmu preselekcji jest wykonanie operacji splotu dwuwymiarowego (obliczenie korelacji) pomiędzy wybranym (uśrednionym) obrazem twarzy z bazy, na który została nałożona maska likwidująca tło, a analizowanym obrazem wejściowym.

W kolejnym etapie dla powstałej macierzy wynikowej odnajdowane są lokalne maksima, których położenie zostaje zapisane w nowej, binarnej macierzy o takich samych rozmiarach jak obraz wejściowy (wartości **1** w miejsce znalezionych maksimów, **0** w pozostałych). Otrzymana macierz odwzorowuje lokalizacje środków obszarów, potencjalnie zawierających twarze.

Autorzy założyli, że jest bardzo mało prawdopodobne, że któraś z szukanych twarzy znajduje się bardzo blisko końców obrazu lub częściowo wychodzi poza obraz. Wyeliminowanie tych środków obszarów uwalnia klasyfikator od pobierania pikseli „spoza” obrazu, co znacząco upraszcza algorytm.

Nie wszystkie twarze na obrazach wejściowych są w równym stopniu podobne do twarzy przygotowanej w preselekcji, zdefiniowanej jako sygnał filtrujący (rys. 4). Mogą nieznacznie różnić się rozmiarem czy rysami (rozmięszczenie oczu, ust itp.). Nawet bardzo niewielkie odchylenia, mimo wskazania przez *preselekcję* jako kandydata, nie gwarantują pozytywnego przypisania do odpowiedniej klasy przez klasyfikator.



Rys. 4. Schemat działania preselekcji

Fig. 4. Diagram of the preselection

Wygodnym rozwiązaniem jest uzyskanie efektu, aby etap *preselekcji* zwracał kilka punktów (lokalnych maksimów) w ramach jednej twarzy. Wystarczy wykorzystany i przygotowany wcześniej sygnał filtrujący przeskalować odpowiednio do poziomu 80% oraz 60% rozmiaru pierwotnego i wykonać jeszcze raz wszystkie pozostałe etapy *preselekcji*. Powoduje to znaczne zwiększenie liczby kandydatów zwracanych przez *preselekcję* i zwiększa prawdopodobieństwo wykrycia twarzy. Niestety takie podejście rodzi problem uzyskania wielu pozytywnych wyników dla tej samej twarzy.

2.3. Klasyfikacja

Najważniejszym etapem działania algorytmu detekcji twarzy jest klasyfikacja. Punkty centralne otrzymane na etapie *preselekcji* to środki obszarów, które będą przypisywane przez klasyfikator do jednej z dwóch klas: „twarz” albo „nie-twarze”. Autorzy zastosowali klasyfikator 3-najbliższych sąsiadów (3-NN).

Klasyfikator przypisuje wyłonione w procesie *preselekcji* potencjalne obszary do jednej z dwóch klas, którym odpowiadają wartości liczbowe: **1** dla „twarzy”, **0** dla „nie-twarzy”. W efekcie powstaje macierz binarna o takich samych rozmiarach jak obraz wejściowy, podobna do wynikowej macierzy *preselekcji*, pokazująca lokalizację obszarów zaklasyfikowanych jako twarz. Wynik klasyfikacji zaprezentowano na rys. 5.



Rys. 5. Rezultat klasyfikacji

Fig. 5. The result of classification

Po złożeniu wyniku klasyfikacji z obrazem wejściowym (rys. 6) widać, że wszystkie twarze znajdujące się na obrazie zostały wykryte poprawnie. Widocznym problemem jest kilka przypadków zwielokrotnienia pozytywnych detekcji tej samej twarzy. Problem ten jest rozwiązany w fazie *postprocessingu*.



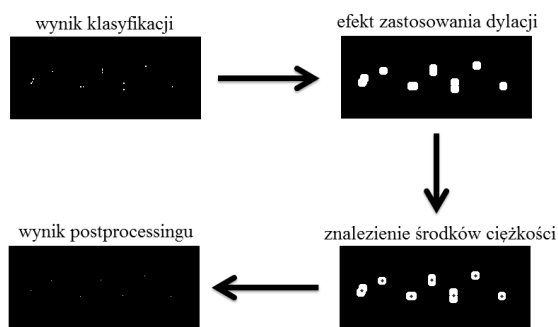
Rys. 6. Wynik detekcji twarzy po procesie klasyfikacji

Fig. 6. Result of face detection after the classification process

2.4. Postprocessing

Pożądanym wynikiem *postprocessingu* jest taka modyfikacja binarnej macierzy zwracanej przez klasyfikator, aby na każdą twarz przypadło dokładnie jedno trafienie oraz by znajdowało się ono możliwie w środku odpowiadającej mu twarzy. Aby osiągnąć taki cel zastosowano jedno z przekształceń morfologicznych obrazu - dylację oraz wyznaczenie środków ciężkości powstałych w wyniku tego przekształcenia obszarów.

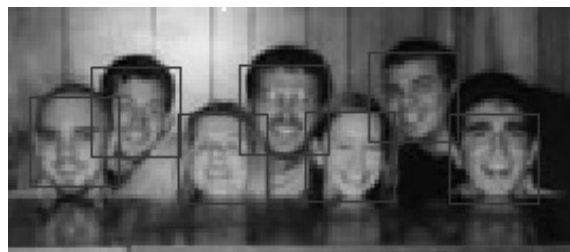
Dylacja to działanie przypisujące każdemu z punktów wartość lokalnego maksimum spośród wszystkich punktów objętych elementem strukturującym [5]. Wykorzystanym elementem strukturującym było koło o promieniu 6 pikseli. Zastosowanie dylacji łączy w jedną figurę „trafienia” (wartości 1 w macierzy) znajdujące się blisko siebie. Kolejnym krokiem jest znalezienie środka ciężkości każdej z tak otrzymanych figur (rys. 7).



Rys. 7. Schemat działania postprocessingu

Fig. 7. Postprocessing diagram

Rezultat działania detektora twarzy z uwzględnieniem bloku *postprocessingu* zaprezentowano na rys. 8.



Rys. 8. Obraz wynikowy całego procesu detekcji twarzy

Fig. 8. The resulting image of the entire face detection process

3. Ocena wyników pracy detektora

W zależności od założonego przeznaczenia detektora twarzy, ocena jego jakości może przebiegać w różnorodny sposób. Czas detekcji oraz ogólną skuteczność uznaje się za najistotniejsze parametry. Skuteczność detekcji twarzy można mierzyć podając:

- błędne trafienia (*false positives*) – liczba obszarów błędnie przypisanych do klasy twarzy,
- błędne odrzucenia (*false negatives*) – liczba rzeczywistych twarzy występujących na obrazie, które nie zostały wykryte przez detektor,
- współczynnik detekcji (*detection rate*) – stosunek liczby twarzy prawidłowo wykrytych przez detektor do rzeczywistej liczby twarzy znajdujących się na obrazie wejściowym.

Zaprojektowany detektor osiąga współczynnik detekcji ponad 80%, dla zdjęć zawierających możliwie niepochyłone twarze, których rozmiar jest zbliżony do rozmiarów twarzy zawartych w obrazach uczących. Poniżej, zaprezentowano przykładowe wyniki działania detektora dla wybranych obrazów. Na zdjęciu

pierwszym (rys. 9) są cztery osoby o twarzach na wprost kamery, niepochylonych, o rozmiarach zgodnych z rozmiarami obrazów uczących. W takim przypadku detektor jest bezbłędny. Wykryte zostały wszystkie twarze, brak było obszarów błędnie zaklasyfikowanych jako twarz.



Rys. 9. Wynik detekcji twarzy dla obrazu 1
Fig. 9. Result of face detection for image 1



Rys. 10. Wynik detekcji twarzy dla obrazu 2
Fig. 10. Result of face detection for image 2

Zdjęcie drugie (rys. 10) stanowi większe wyzwanie dla detektora (9 twarzy). Niektóre osoby mają obrócone lub pochylone twarze, warunki oświetleniowe zdjęcia też nie są dobre. Twarz jednej osoby jest częściowo przesłonięta przez czapkę. Detektor nie wykrył jednej z nich, właśnie tej pochylonej i częściowo przesłoniętej czapką. Występuje też jedno błędne przypisanie pewnego obszaru na obrazie do klasy twarzy. W tym przypadku współczynnik detekcji wyniósł 89%.

4. Podsumowanie

W artykule zaprezentowano algorytm detekcji twarzy bazujący wyłącznie na jej wyglądzie. Algorytm dobrze radzi sobie z detekcją w przypadku obrazów zawierających możliwie mało pochylone twarze, których rozmiar jest zbliżony do rozmiaru twarzy zawartych na obrazach uczących. Pewien problem stwarzają natomiast twarze znacznie obrócone oraz takie, które nie mają swoich odpowiedników w bazie (np. twarze dzieci). Warunki oświetleniowe, pomimo procesu adaptacyjnego wyrównania histogramu, zaimplementowanego w bloku obróbki wstępnej, mają również wpływ na współczynnik detekcji. Twarze znacznie niedoświetlone mogą nie zostać wykryte.

5. Literatura

- [1] <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [2] Yang M. H., Kriegman D. J., Ahuja N.: Detecting Faces in Images: A Survey. IEEE.
- [3] Gross R., Shi J., Cohn J.: Quo vadis Face Recognition? Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania, Czerwiec 2001.
- [4] <http://cbcl.mit.edu/software-datasets/FaceData2.html>
- [5] Osuna E., Freund R., and Girosi F.: Training Support Vector Machines: An Application to Face Detection. Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997.
- [6] Craw I., Tock D., and Bennett A.: Finding Face Features. Proc. Second European Conf. Computer Vision, pp. 92-96, 1992.

otrzymano / received: 16.08.2013

przyjęto do druku / accepted: 03.02.2014

artykuł recenzowany / revised paper

INFORMACJE

Bezpłatny dostęp do artykułów opublikowanych w PAK

Realizując idee Open Access przez miesięcznik PAK informujemy, że artykuły opublikowane w PAK są dostępne w wersji elektronicznej. Dostęp do artykułów opublikowanych jest bezpłatny, z zachowaniem 1 roku karencji. Artykuły w łatwy sposób można znaleźć korzystając z wyszukiwarki artykułów. Bazę artykułów można przeszukać po nazwisku autora, tytule artykułu lub po słowach kluczowych.

Tadeusz SKUBIS
Redaktor naczelny Wydawnictwa PAK