AUTOMATYKA • 2011 • Tom 15 • Zeszyt 3

Piotr Pawlik*, Zbigniew Bubliński*

# Rozpoznawanie twarzy za pomocą deskryptorów punktów charakterystycznych**

## 1. Wprowadzenie

Inspiracją niniejszych badań były doniesienia psychologiczne, informujące że do rozpoznania twarzy wystarczają człowiekowi dwa ruchy fiksacyjne. W publikacji donoszącej o tym odkryciu [5] wykazano, że człowiek koncentruje się na stosunkowo blisko położonych punktach rejonu nosa. Dalsze badania (np. [1, 2]) potwierdziły tę tezę. W niniejszym artykule zaprezentowano wyniki testów sprawdzających, czy stosując deskryptory punktów charakterystycznych, można rozpoznać twarz, bazując na co najmniej jednym punkcie.

## 2. Przesłanki psychologiczne

We wspominanych badaniach Hsiao [5] zadaniem testowanych osób było rozpoznawanie 32 twarzy wcześniej pokazanych (przez 3 sekundy każda) spośród 64 twarzy. Na wstępie koncentrowano uwagę za pomocą markera na centrum obrazu, a następnie powyżej lub poniżej prezentowano „uśrednioną twarz" jako swego rodzaju maskę przysłaniającą twarz testowaną. W momencie wykrycia przeniesienia uwagi na maskę pokazywano testowaną twarz na czas określonej liczby fiksacji, po którym znowu przesłaniano ją maską. Testy wykazały, że wystarczało pokazanie twarzy do momentu drugiej fiksacji, aby testowana osoba prawidłowo zdecydowała, czy jest to „nowa" twarz, czy też jedna z wcześniej pokazanych.

Kolejnym interesującym doniesieniem są badania Bindemanna [2], w których nie koncentrowano się na obrazach twarzy *en face*, ale także sprawdzono punkty fiksacji dla

* AGH Akademia Górniczo-Hutnicza, Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki, Katedra Automatyki, al. A. Mickiewicza 30, 30-059 Kraków

** Praca realizowana w ramach Projektu SIMPOZ z Ministerstwa Nauki i Szkolnictwa Wyższego nr 0128/R/t00/2010/12

półprofili i profili. W efekcie stwierdzono, że pierwsza fiksacja ma miejsce w okolice środka ciężkości twarzy, a następne przemieszczają się w najbliższą okolicę – czyli obszar między oczami dla obrazów *en face* i lepiej widocznego oka w przypadku półprofili. Natomiast w obrazach profili druga fiksacja przemieszczała się albo w stronę widocznego oka, albo w stronę ucha. Również G. van Belle w pracy [1] potwierdza, iż początkowe punkty skupienia są umieszczone w okolicach nasady nosa.

Warto w tym miejscu zwrócić uwagę, że badania nad położeniem punktów fiksacji nie obejmują tematyki rozmiaru obszaru twarzy, który po fiksacji jest podstawą do rozpoznania. Innymi słowy, mogą one jedynie pokazać, gdzie człowiek koncentruje uwagę, ale nie potrafią precyzyjnie określić na czym (np. czy jest to cała twarz, czy też np. okolice oczu i nosa z pominięciem ust).

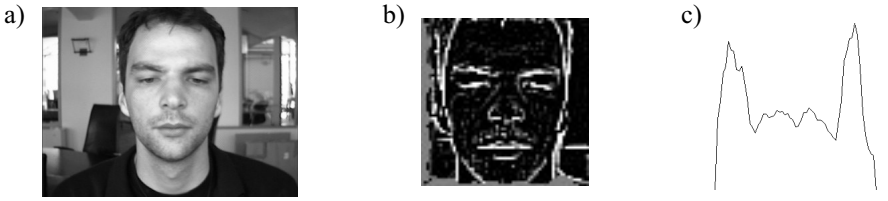
Tak więc można przyjąć, że do rozpoznania twarzy wystarcza wskazanie kilku (być może nawet tylko jednego) punktów koncentracji uwagi i punkty te są umieszczone w okolicach „styku" oczu i nosa. Otwarte pozostaje pytanie, jaka część obszaru twarzy wystarcza do rozpoznania. Umiejscowienie punktu koncentracji bliżej linii oczu sugeruje, że rejon ust jest tutaj mniej istotny. Niniejsza praca miała na celu zbadanie, czy kierując się doniesieniami badań psychologicznych, opis otoczeń punktów fiksacji za pomocą popularnych deskryptorów punktów charakterystycznych (jakimi są deskryptory SIFT [7] i HOG [4]) nie pozwoliłby na uzyskanie mechanizmu automatycznego rozpoznawania twarzy. Deskryptory punktów charakterystycznych dobrze sprawdzają się w zagadnieniach detekcji wybranych obiektów (np. wspomniany HoG jako detektor postaci człowieka, czy też liczne zastosowania detektora SIFT do wyszukiwania na obrazach pojazdów czy broni). Są one w pewnym zakresie niewrażliwe na przesunięcia, zmiany rozmiaru i oświetlenia. Powstaje pytanie, czy kierując się przesłankami psychologicznymi, można ich użyć w opisie twarzy w zakresie wystarczającym do jej rozpoznania.

## 3. Metodyka badań

Zadanie wstępnej detekcji twarzy przed jej właściwym rozpoznaniem zostało już dobrze opracowane w algorytmach przetwarzania obrazu. Klasyczna już kaskada Haara [9] czy też podejście Paisitkriangkrai [8] wykorzystujące *Histogram of Oriented Gradients* rozwiązują ten problem z bardzo dobrym rezultatem. W niniejszej pracy w celu detekcji twarzy wykorzystano kaskadę Haara ze względu na szybkość jej działania.

Po wyszukaniu twarzy za pomocą kaskady Haara obrazy normalizowano do stałej rozdzielczości 64×64. Okazało się jednak, że kaskada Haara niedostatecznie precyzyjnie wyznacza fragmenty obrazu zawierające twarz. Występują tu dość duże różnice w stosunku powierzchni twarzy do całego wyznaczonego fragmentu, co pogarszało efekty normalizacji. Z tego powodu przed normalizacją dokonano doprecyzowania granic obszaru zawierającego twarz. W tym celu dokonano filtracji górnoprzepustowej (laplasjanem) fragmentów wyznaczonych przez kaskadę Haara (por. rys. 1b)). Maksima w połowach rzutu pionowego przefiltrowanego fragmentu przyjęto za granice twarzy (por. rys. 1c)). Odległość między

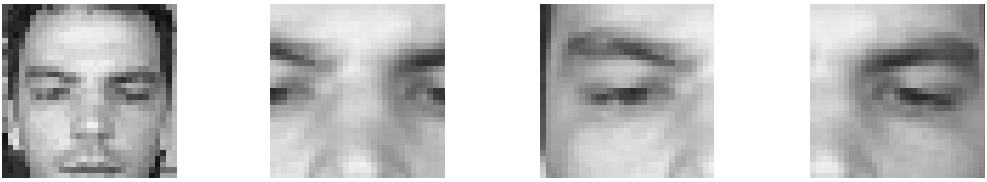
maksimami stała się długością boku kwadratu wycinającego mniejszy fragment obrazu twarzy. W nowym fragmencie najczęściej obcinana była dolna część twarzy, ale nie przeszkadza to dalszej analizie, która skupia się na okolicach oczu. Dodatkowym aspektem takiego ograniczenia się do górnej części twarzy jest znaczne uniezależnienie się od mimiki twarzy, związanej przede wszystkim z okolicą ust.




**Rys. 1.** Przykładowy obraz (a) z bazy BioID [3]; b) laplasjan fragmentu wyszukanego przez kaskadę Haara; c) rzut pionowy obrazu b)

Normalizacja do rozmiaru 64×64 była spowodowana tym, że deskryptory są zależne od skali. W klasycznym podejściu (np. w SIFT-cie) skala jest wyznaczana przy wyborze punktów charakterystycznych. W opisywanym przypadku punkty charakterystyczne są wyznaczane arbitralnie. Zamiast tworzyć opis otoczenia punktu w wielu skalach (uzyskując wiele opisów) zdecydowano znormalizować obraz wejściowy do stałego rozmiaru (co pozwala na opis punktu tylko jednym deskryptorem).

Kierując się opisanymi powyżej doniesieniami psychologicznymi, jako zasadniczy punkt charakterystyczny wybrano punkt między oczami. Jako dodatkowe obszary charakterystyczne wykorzystane w deskrypcji twarzy przyjęto okolice oczu (por. rys. 2). Wybierając położenie oczu jako podstawy wyznaczenia wszystkich trzech punktów charakterystycznych (położenie oczu i punkt między nimi), kierowano się względną łatwością wyznaczenia ich położenia oraz doniesieniami psychologicznymi.



**Rys. 2.** Otoczenia punktów charakterystycznych (w dwukrotnym powiększeniu), dla których liczono deskryptory

Podczas badań podjęto próbę automatycznego wyznaczenia położenia oczu przy użyciu podobnej techniki jak przy opisanym powyżej doprecyzowaniu granic obszaru zawierającego twarz. Jednakże uzyskane wyniki wykazały potrzebę zastosowania lepszej metody. Ze względu na możliwość wykorzystania danych o położeniu oczu z annotacji zdecydowano przesunąć poszukiwanie bardziej efektywnej metody do następnego etapu badań.

## 4. Wyniki

Do testów wykorzystano bazę danych BioID [3, 6] ze względu towarzyszącą jej annotację położenia oczu, co pozwoliło na skupienie się na opisywaniu twarzy za pomocą deskryptorów bez konieczności dokładnej analizy położenia punktów charakterystycznych. Ponadto baza ta zawiera sekwencje obrazów z wieloma wystąpieniami tych samych osób, co pozwoliło lepiej zweryfikować postawioną tezę.

Baza zawiera 1521 obrazów twarzy *en face* (w odcieniach szarości) 23 osób. Przykładowy obraz z tej bazy został pokazany na rysunku 1a. Jednakże nie wszystkie obrazy mogły być wykorzystane w testach. Część z nich nie zawierała pełnego wizerunku (zazwyczaj „obcięta" była górna część twarzy). W efekcie twarze na takich obrazach albo nie były w ogóle wykrywane przez kaskadę Haara, albo wykryty fragment nie zawierał dostatecznego otoczenia punktów charakterystycznych umożliwiającego deskrypcję (wykryta twarz była „obcięta"). Z tego powodu do analizy wyników wykorzystano tylko 1416 obrazów.

W deskryptorze HOG przyjęto rozmiar bloku 32×32. Deskryptor SIFT ograniczono tak, oby obejmował podobny obszar. Deskryptory „wzorcowe" były wyliczane dla 3 punktów charakterystycznych, natomiast przy rozpoznawaniu wyznaczano deskryptory dla 3 punktów charakterystycznych oraz ich 24 najbliższych sąsiadów (aby jeszcze bardziej uniezależnić się od wpływu przesunięcia). Przy klasyfikacji kierowano się kryterium najmniejszej odległości euklidesowej pomiędzy wzorcem a deskryptorem punktu. Jako obrazy wzorcowe wybrano jeden lub dwa (w zależności od liczebności zdjęć danej osoby) z obrazów każdej z reprezentowanych w bazie osób. Jeżeli daną osobę reprezentowały dwa zdjęcia, to były one wybierane z różnych sekwencji (przy różnym oświetleniu). Jeżeli dana osoba była fotografowana w okularach i bez, to jako wzorce brano oba takie przypadki.

Jak widać w tabeli 1 użycie jednego punktu charakterystycznego nie daje zadowalających wyników (poprawne rozpoznawania stanowią około 79%), przy czym nieco lepsza okazała się reprezentacja HOG (o jeden procent). Uwzględniając przypadki, w których oba deskryptory jednocześnie wskazały tę samą klasę, „rozpoznanie zgodne" uzyskuje się poprawne rozpoznanie tylko w 69,7% przypadków. Powstaje pytanie jak rozstrzygać „sporne" przypadki. Jako rozwiązanie zaproponowano dodatkowy opis dotychczasowego punktu charakterystycznego, ale obejmujący jego bliższe otoczenie (okolice nasady nosa) oraz dołożenie dwóch dodatkowych punktów charakterystycznych obejmujących okolice oczu (por. rys. 2).

Dodanie punktów charakterystycznych pozwoliło na generowanie odpowiedzi na podstawie rozpoznań cząstkowych z poszczególnych punktów. Rozpoznaniom cząstkowym przypisano różne wagi – punkt centralny 10, „powiększony" punkt centralny – 7, okolice oczu – 4. Dodanie drugiego deskryptora punktu centralnego pozwoliło na poprawienie rozpoznań do nieco ponad 89%, podobnie w przypadku dwóch dodatkowych punktów z okolicą oczu. Połączenie wszystkich czterech deskryptorów zwiększyło skuteczność rozpoznawania do ponad 90%, co można uznać za wynik satysfakcjonujący. Przeprowa-

dzono także eksperyment, w którym wagi HOG zostały zwiększone o 1 w stosunku do SIFT, aby oddać fakt, że deskryptor HOG dawał nieco lepsze rezultaty. Przyniosło to dalszą, aczkolwiek niewielką, poprawę rozpoznań (z 90,68% do 90,75%).

**Tabela 1**
Zestawienie wyników (szczegółowy opis w tekście)

| Wyniki | Liczba poprawnych rozpoznań | Procentowo |
|---|---|---|
| Tylko SIFT, jeden punkt | 1115 | 78,74% |
| Tylko HOG, jeden punkt | 1129 | 79,73% |
| Wyniki „zgodne” | 987 | 69,70% |
| Wyniki ważone (nasada nosa), 3 punkty | 1261 | 89,05% |
| Wyniki ważone (okolice oczu), 3 punkty | 1262 | 89,12% |
| Wyniki ważone (łącznie), 4 punkty | 1284 | 90,68% |
| Wyniki ważone ze zwiększeniem wag HOG | 1285 | 90,75% |

## 5. Podsumowanie

Wyniki w dużej mierze potwierdzają wynik eksperymentu psychologicznego, w którym pokazano, że do rozpoznania twarzy wystarcza skoncentrowanie się na jednym punkcie charakterystycznym. Ponadto wykazały, że do opisu takiego punktu można wykorzystać deskryptory HOG i SIFT (przy czym deskryptor HOG dał w tym wypadku nieco lepsze wyniki niż deskryptor SIFT). Analiza przypadków błędnie rozpoznanych wykazała, że gradientowe deskryptory punktów charakterystycznych są dość wrażliwe na oświetlenie. Jednakże nie chodzi tu o poziom oświetlenia (który raczej nie wpływa na poziom rozpoznawania, dając podobne opisy gradientowe), a o nierównomierność oświetlenia, odblaski i prześwietlenia (które są wzmacniane w wyniku gradientowania). Z tego powodu w praktycznych zastosowaniach należałoby zapewnić stabilne warunki oświetlenia lub wprowadzić dodatkowe przetwarzanie wstępne zmniejszające negatywne zjawiska prześwietlenia.

## Literatura

[1] van Belle G., Ramon M., Lefèvre P and Rossion B., *Fixation patterns during recognition of personally familiar and unfamiliar faces*. Front. Psychology, 1:20, 2010.

[2] Bindemann M., Scheepers C., Burton A. M., *Viewpoint and center of gravity affect eye movements to human faces.* Journal of Vision, 9(2):7, 2009, 1–16.

[3] *BioID-Technology Research. The BioID Face Database*. http://support.bioid.com/downloads/facedb/index.php. Internet.

[4] Dalal N., Triggs B., *Histograms of oriented gradients for human detection.* Proc. Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005, 886–893.
[5] Hsiao J. H.-W., Cottrell G., *Two fixations suffice in face recognition.* Psychological Science, 19, 2008, 998–1006.
[6] Jesorsky O., Kirchberg K.J., Frischholz R.W., *Robust Face Detection Using the Hausdorff Distance*. Audio- and Video-Based Biometric Person Authentication, 3rd International Conference, Springer, Lecture Notes in Computer Science, LNCS-2091, 2001, 90–95.
[7] Lowe D.G., *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision, 60, 2, 2004, 91–110.
[8] Paisitkriangkrai S., Shen C., Zhang J., *Face Detection with Effective Feature Extraction.* Proc. the 10th Asian conference on Computer vision, Springer, Lecture Notes in Computer Science, vol. 6494/2011, 2011, 460–470.
[9] Viola P., Jones M., *Rapid Object Detection Using a Boosted Cascade of Simple Features.* Proc. Int. Conf. Computer Vision and Pattern Recognition, vol. 1, 2001, 511–518.