

## Ćwiczenie 2: Detekcja twarzy

Adrian Jałoszewski

23 października 2017

# 1 Wpływ doboru parametrów na detekcję twarzy

Ponieważ ten podpunkt ćwiczenia wymaga badania wykrywania twarzy w zależności od:

- progu binaryzacji
- elementu strukturalnego
- minimalnej wielkości obiektu

Ze względu na to, że element strukturalny składa się z jego rodzaju oraz wielkości dla ułatwienia rozważań przyjąłem, że minimalna wielkość obiektu powinna wynosić 400, a pozostałe trzy parametry (licząc element strukturalny jako dwa parametry) powinny być zmienne.

Zastosowanie takiego podejścia jest uzasadnione tym, że twarz i tak będzie prawdopodobnie większa od najmniejszego elementu. Funkcja `szukaj_twarz` zawiera już wybór największego elementu, przez co wykluczy elementy najmniejsze. Wyjątkiem od tej sytuacji jest przypadek gdy twarz nie zostanie wykryta w miejscu gdzie powinna być, a element zbyt mały będzie przypominał twarz – wtedy zostanie wykryty błędny obiekt zamiast zwrócenia braku wykrytej twarzy.

Przykład dla obrazu `testowy_0_0000.jpg`:

Próg binaryzacji	Element strukturalny	Poprawne wykrycie twarzy
0.3	diamond 2	tak
> 0.6	diamond 2	nie – wycinek szyi uznany za twarz
0.3, 0.4	diamond 3	tak
0.6	diamond 3	nie – wycinek szyi uznany za twarz
0.7	diamond 3	nie – prawe oko
0.3, 0.4, 0.5, 0.7	diamond 5	tak
0.3, 0.4, 0.5	diamond 10	tak
0.7	diamond 10	nie – prawy policzek
0.3, 0.4, 0.5, 0.6	diamond 15	tak
0.3, 0.4, 0.5, 0.6, 0.7	diamond 20	tak

Podobne wyniki były uzyskane dla pozostałych elementów strukturalnych (disk, rectangle) i ich rozmiarów (1, 3, 4, 15, 20). Rozpatrywane progi binaryzacji: 0.3, 0.4, 0.5, 0.6, 0.7. Przypadki, których nie ma w tabelce są tymi dla których algorytm `szukaj_twarz` nie znalazł twarzy.

Problematyczne były przypadki dla niektórych z obrazów, gdzie kolor skóry był podobny do tła – wymagały one większych progów binaryzacji (np. obraz `testowy_0_0001.jpg`, disk 15, próg 0.5). W przypadku zbyt dużych progów binaryzacji wykrywane były fragmenty ciała, które miały kolor skóry najbliższy średniemu i miały dużą płaską powierzchnię (zwykle fragment szyi) – fragmenty twarzy wtedy nie spełniały wtedy warunku proporcji lub miały łącznie mniej pikseli niż fragment szyi (kryteria w algorytmie `szukaj_twarz`).

W przypadku gdy algorytm miał do czynienia z twarzami na obrazach o większych rozdzielczościach należało zmienić parametry, tak aby brały pod uwagę wielkość obrazów, albo takowe przeskalować (zwiększyłyby to uniwersalność algorytmu).

## 2 Wyznaczanie modelu koloru skóry

W przypadku wyznaczenia na podstawie ROI unika się generalizacji, która musi brać pod uwagę większy zakres kolorów skóry – możliwość pomylenia skóry ze ścianą, jeżeli ta jest podobnego koloru do jednego z potencjalnych kolorów. Oprócz tego można w tym przypadku lepiej uwzględnić karnację.

### 2.1 Zaznaczenie ROI twarzy

W tym kroku wyznaczany jest obszar w którym się znajduje twarz – w ten sposób wycinane są fragmenty nie będące twarzą do następnych kroków.

### 2.2 Konwersja RGB-YCbCr i wybór składowych Cb oraz Cr

W przypadku formatu RGB wszystkie kolorory mają wpływ na jasność obrazu. Po przekształceniu w YCbCr mamy do czynienia z przypadkiem, gdzie składowa Y odpowiada za jasność – po jej usunięciu zostają tylko składowe niezależne od jasności.

### 2.3 Obliczenie wektora średniej i macierzy kowariancji

Wyznaczana jest wartość średnia oraz zależność pomiędzy składowymi Cb oraz Cr.

### 2.4 Obliczenie wielowymiarowego rozkładu normalnego

Dla obrazu wynikowego jest liczone prawdopodobieństwo, że dany piksel jest koloru skóry. Jest to rozkład normalny o średniej i macierzy kowariancji wyznaczonej w poprzednim podpunkcie.

## 3 Operacja zamknięcia w procesie detekcji twarzy

Jest to dokonanie operacji erozji na wynik operacji dyatacji. Operacja dyatacji zwiększa powierzchnię pokrywana przez białe komórki (jeżeli jedna z komórek w elemencie strukturalnym jest biała, to wynik też jest biały) operacja erozji zwiększa powierzchnię pokrywana przez czarne komórki (jeżeli jedna komórka elementu strukturalnego czarna – wynik czarny). Przez takie połączenie tych operacji dwa oddzielne obiekty mogą być scalone (mogły być rozdzielone w wyniku binaryzacji), przez co przy usuwaniu zbyt małych elementów nie są usuwane elementy istotne.

## 4 Inne metody lokalizacji twarzy na obrazie

- Głębokie Konwolucyjne Sieci Neuronowe (CNN) – wykorzystywane głównie w celach klasyfikacyjnych, czy obraz jest twarzą
- Sieci Neuronowe Fast R-CNN – wykrywanie położenia obrazu przy pomocy sieci neuronowych
- Histogram of Oriented Gradients – obraz jest dzielony na obszary dla których jest liczony histogram gradientów, wykrywanie odbywa się na podstawie wyniku łączenia histogramów

- Edge-Orientation Matching – rozpoznawanie na podstawie orientacji krawędzi
- Wykorzystanie Metryki Hausdorffa – wyznaczenie odległości między obrazami
- Cascading Classifiers – łączenie klasyfikatorów, podając wyjście z poprzedniego klasyfikatora jako dodatkowa informacja do następnych klasyfikatorów
- Wykrycie twarzy na podstawie charakterystycznych ruchów.

## 5 Rezultaty detekcji twarzy przy pomocy kaskad Hara

### 5.1 Bez ustawiania dodatkowych parametrów

Poprawnie wykryte zostały wszystkie twarze ustawione frontalnie. Niewykryte zostały twarze ustawione profilem, zdarzały się przy nich pomyłki. Twarze ustawione skośnie do kamery były wykryte w wielu przypadkach.

### 5.2 Profile face

Z tym ustawieniem wykryte zostało wiele twarzy, które zostały ustawione profilem do kamery. Nie wszystkie jednak zostały z nich wykryte. Bardzo często jedno ucho w połączeniu z okiem zdawało się być brane za twarz z profilu, przez co wykrywane były połówki twarzy, a czasami również całe twarze.

### 5.3 Eye pair big / small

Bardzo często wykrywało parę oczu, zdarzały się przypadki, gdzie jednak nie były wykrywane. Czasami z oczami mylone były nozdrza. Często niewykrywane były oczy na obrazach, gdzie osoba patrzy pod kątem na kamerę, a obrazy z profilu nie były wykrywane nigdy (brak drugiego oka). Zdarzały się przypadki, gdzie obok jednego oka brane było coś w czym było rozpoznawane oko (np. kosmyk włosów), albo dwa takie obiekty.

### 5.4 Upper body

Wykrywało też twarz w przypadku jak miało do czynienia z popiersiem, w innych przypadkach często doszukiwało się jej tam gdzie jej nie ma – wadą tego jest to, że obszar wykryty jako popiersie jest bardzo duży, podobnie jak liczba błędnych wykryć.

### 5.5 Max/Min size

W zależności ustawienia tych parametrów były ignorowane obiekty, które były mniejsze lub większe od określonego rozmiaru. Wyrzucało to przypadki gdzie były wykrywane fragmenty ubioru lub włosów mniejsze od spodziewanej twarzy oraz rzeczy od niej większe – wadą tego rozwiązania jest konieczna znajomość tego jaki fragment obrazu będzie zajmowany przez twarz i obliczenie rozmiaru twarzy w zależności od rozdzielczości obrazu. Ustawienia wybrane dla obrazów o małej rozdzielczości nie sprawdzały się dla obrazów o dużej rozdzielczości jeżeli chodzi o maksymalny rozmiar oraz odwrotnie jeżeli chodzi o minimalny.

## 6 Metoda detekcji twarzy przy pomocy kaskad Haara

Metoda kaskad Haara polega na uczeniu klasyfikatora przy pomocy wielu obrazów zarówno takich zawierających twarz jak i takich niezawierających twarzy. Uczenie odbywa się poprzez wykrywanie cech Haara – są to ramki czarno-białe. Po nałożeniu takiej ramki na obraz od sumy pikseli, które się pokrywają z polami białymi jest odejmowana suma pól znajdujących się pod polami czarnymi.

Następnie jest liczona jest ważona suma słabych klasyfikatorów (nazwa ponieważ jako same w sobie nie są w stanie wykryć twarzy).

Metoda kaskad Haara polega na tym, że się stosuje poszczególne klasyfikatory, podając wynik poprzedniego jako wejście następnego. Jeżeli któryś z klasyfikatorów stwierdzi, że na obrazie nie znajduje się twarz należy przypadek ten odrzucić, jeżeli ostatni z klasyfikatorów stwierdzi, że mamy do czynienia z twarzą – algorytm wykrył twarz.

Po nauczeniu klasyfikatora kaskad Haara podobne operacje są wykonywane na obrazie na którym chcemy dokonać detekcji, a wynik jest porównywany do tego uzyskanego na zbiorze uczącym.