

Laboratorium BiNW:
LOKALIZACJA TWARZY

Jarosław Gliwiński
#74839, MiEPU

14 września 2009

1 Wstęp

Celem ćwiczenia była prezentacja algorytmów lokalizacji twarzy poprzez empiryczne ich zastosowanie w serii eksperymentów w różnych warunkach na obrazach nieruchomych oraz sekwencjach wizyjnych.

Wykorzystano wyniki działania programu *LokalizacjaTwarzy*, co zwalniało z konieczności samodzielnej implementacji wariantów algorytmów, a pozwoliło skupić się na analizie rezultatów.

2 Lokalizacja twarzy w obrazach statycznych

2.1 Metodologia

Testy przeprowadzono na zestawie 50 obrazów testowych. Obrazy wynikowe zostały podzielone na 3 kategorie wg kryterium poprawności lokalizacji twarzy (w nawiasach oznaczenia kategorii stosowane na rysunkach i w tabelach):

- lokalizacja błędna (0)
- lokalizacja niepełna (1)
- lokalizacja poprawna (2)

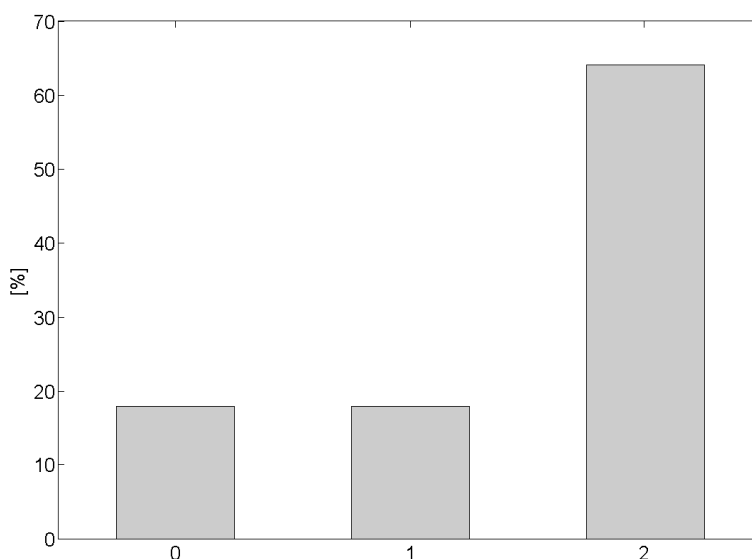
W części przypadków podział jest obciążony błędem systematycznym wynikającym z subiektywności kryterium poprawności – pewne przypadki (np. małe względne przesunięcia; niewielkie „twarze” w miejscu dłoni towarzyszące prawidłowo wykrytym twarzom; wykrycie części z licznych twarzy w obrazie) mogą być zakwalifikowane do więcej niż jednej kategorii w zależności od obserwatora.

2.2 Algorytm wykorzystujący operator Sobela

Surowe wyniki procentowe dla algorytmu wykorzystującego operator Sobela przedstawia tabela 1, dodatkowo wizualizuje je wykres 1.

kategoria	wynik [%]
0	18
1	18
2	64

Tabela 1: Wyniki działania „algorytmu Sobela”



Wykres 1: Wyniki działania „algorytmu Sobela”

W 6 przypadkach jako twarz został rozpoznany inny obiekt, w tym w jednym przypadku wynik był wyjątkowo niezadowalający (obraz 27, gdzie element krajobrazu naturalnego został zinterpretowany jako kilka nakładających się obrazów twarzy). Nie dostrzeżono zależności skuteczności lokalizacji od płci bądź koloru skóry, natomiast zaobserwowano spadek skuteczności w przypadku obrotu głowy wokół którejkolwiek osi w stosunku do pozycji *en face*.

Można podejrzewać, że wpływ na skuteczność mają słaby kontrast obrazu czy rozmycie krawędzi. W zbiorze testowym było zbyt mało obrazów o takich własnościach by wyciągnąć jednoznaczne wnioski, jednak wydaje się, że powinno to wynikać to z własności filtracji Sobela – w obydwu tych przypadkach wykrywanie krawędzi jest utrudnione.

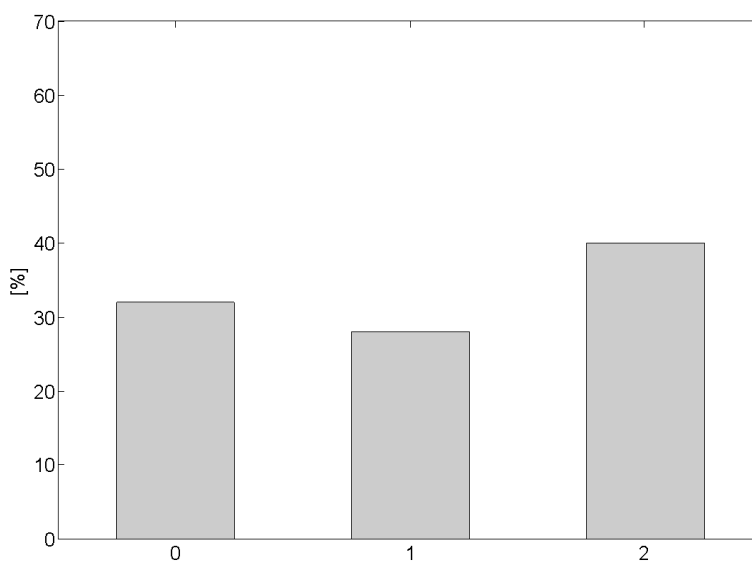
Kolejną obserwacją, którą trudno potwierdzić statystycznie (tylko jeden obraz w zbiorze testowym) jest wpływ ukrycia detali twarzy, w szczególności oczu, które w używanym wzorcu są mocno wyeksponowane.

2.3 Algorytm wykorzystujący transformację HMT

Surowe wyniki procentowe dla algorytmu wykorzystującego transformację HMT przedstawia tabela 2, dodatkowo wizualizuje je wykres 2.

kategoria	wynik [%]
0	32
1	28
2	40

Tabela 2: Wyniki działania „algorytmu HMT”



Wykres 2: Wyniki działania „algorytmu HMT”

Transformacja Hit-or-Miss korzysta z zestawu elementarnych wzorców pozwalających na wykrywanie krawędzi kształtów kolistych i owalnych – a wzorec wykorzystywany przy wyliczaniu odległości Hausdorffa wygląda inaczej niż stosowany poprzednio. Elementy te (metoda wykrywania krawędzi i wzorec twarzy) są ze sobą ściśle powiązane. Na podstawie badań empirycznych możliwie reprezentatywnego zbioru obrazów twarzy ustalany jest wzorec będący uśrednieniem wyników po całym zbiorze. Ponadto, wyliczanie metryki odbywa się z innymi parametrami, ze względu na mniejsze „nasycenie” wzorca punktami charakterystycznymi. Wydaje się to zresztą mieć wpływ na wyniki lokalizacji, o czym poniżej.

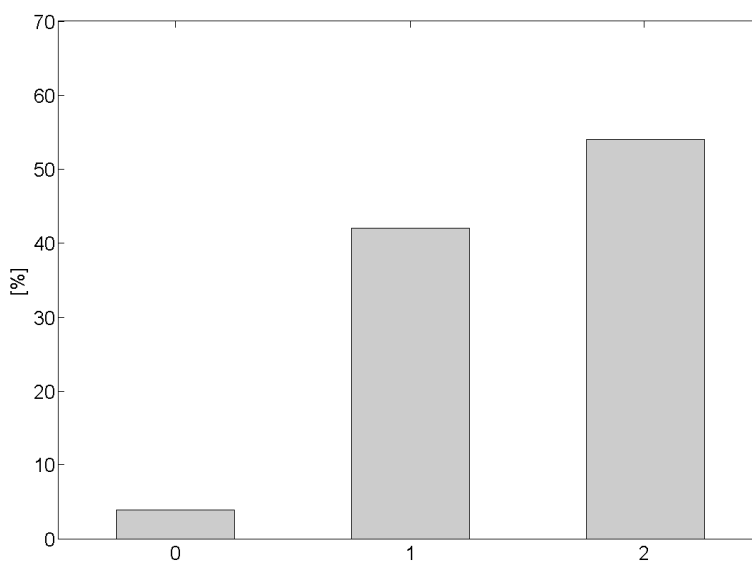
W 13 przypadkach jako twarz został rozpoznany inny obiekt, więcej niż raz element krajobrazu naturalnego został zinterpretowany jako kilka nakładających się obrazów twarzy. Nie dostrzeżono zależności skuteczności lokalizacji od płci bądź koloru skóry, ponownie trudno zaobserwować zależność skuteczności lokalizacji od obrotu głowy. Wydaje się jednocześnie, że skuteczność w pozornie prostych przypadkach (*en face*) spadła. Ogólny spadek skuteczności w stosunku do „metody Sobela” wydają się potwierdzać wyniki badań (o blisko 40% względnie mniejszy udział wyników kategorii drugiej oraz ponad 40% względny wzrost udziału wyników kategorii zero).

2.4 Algorytm wykorzystujący segmentację koloru

Surowe wyniki procentowe dla algorytmu wykorzystującego segmentację koloru przedstawia tabela 3, dodatkowo wizualizuje je wykres 3.

kategoria	wynik [%]
0	4
1	42
2	54

Tabela 3: Wyniki działania segmentacji koloru



Wykres 3: Wyniki działania segmentacji koloru

Algorytm wydaje się nie być bardziej wrażliwy na kolor skóry twarzy niż poprzednie – wynika to z faktu zastosowania transformacji przestrzeni koloru. Skuteczność detekcji dla twarzy o różnorodnej lokalizacji jest natomiast lepsza. Wobec powyższych głównym problemem pozostaje klasyfikacji innych części ciała o odsłoniętej skórze (dłoni, szyi, etc.) jako twarzy. Nie dziwi to wobec faktu znacznego udziału w algorytmie składowej detekcji koloru skóry.

Powyższe spostrzeżenia korespondują z wynikami eksperymentalnymi. Ogólna skuteczność jest wysoka (kategorie 1 i 2 to w sumie 96% przypadków), co wynika z faktu podobnej zależności w kwestii koloru skóry jak przy poprzednich algorytmach, przy jednocześnie zmniejszonej wrażliwości na położenie twarzy względem obiektywu. Względnie wysoki jest także procent obrazów kategorii 1, co z kolei jest związane z trudnościami przy „odsiewaniu” dłoni, etc.

3 Lokalizacja twarzy w sekwencjach wizyjnych

3.1 Wyniki eksperymentów

W przypadku sekwencji wizyjnych wyniki zebrano w tabeli 4, zaś krótką dyskusję wyników przeprowadzono osobno. Ponownie użyte zostały te same algorytmy, jednak w dwóch wersjach: zwykłej oraz korzystającej z wektorów ruchu („+mv”) Kategoryzowano wg skali zdefiniowanej dla obrazów statycznych, uszczegółowiając ją, gdy było to konieczne (uzyskano to wprowadzając klasyfikację ułamkową – np. kategoria $1^{1/2}$ oznacza lokalizację nie do końca poprawną, jednak wystarczająco dobrą, by nie sposób było ją uznać za „niepełną” (kategorii 1).

algorytm	wynik lokalizacji (kategoria)				
	akiyo	claire	container	missa	silent
sobel	2	1	2	2	0
sobel+mv	2	$1^{1/2}$	2	2	0
HMT	1	1	0	1	1
HMT+mv	1	$1^{1/2}$	0	1	1
kolor	1	1	2	$1/4$	$1^{3/4}$
kolor+mv	2	1	2	0	2

Tabela 4: Wyniki lokalizacji w sekwencjach wizyjnych

3.2 Komentarz do wyników

Algorytm korzystający z operatora Sobela okazał się niewystarczająco dobry do sceny z językiem migowym (*silent*). Z kolei w metodzie korzystającej z transformacji HMT kwadrat oznaczający lokalizację w pozornie losowy sposób pojawiał się i znikał, z klatki na klatkę.

Algorytm oparty o segmentację koloru lepiej poradził sobie z sekwencją „migową”, nie był jednak w stanie prawidłowo oznaczyć pozornie prostej (pozostałe algorytmy spisywały się co najmniej przeciętnie) sekwencji *missa* – z mówiącą prezenterką. Ciężko określić, czym może być to spowodowane.

4 Podsumowanie

Algorytmy działają dość dobrze dla pewnych klas obrazów. Dla badanych obrazów jak i sekwencji można zaobserwować nieproporcjonalne różnice w wynikach między algorytmami – nie ma warunków jednoznacznie „trudnych” czy „łatwych” jednocześnie dla każdej metody. Na ocenę wpływa także subiektywny poziom akceptacji błędów – oznaczania elementów tła, niedokładności granic twarzy, dodatkowych oznaczeń wewnątrz prawidłowego oznaczenia.

Można wyciągnąć wniosek, że możliwe byłoby poprawienie wyników przez zastosowanie metod hybrydowych, korzystających w zależności od warunków z więcej niż jednej z badanych metod jednocześnie i na tej podstawie syntezujących wynik. Możliwe byłoby także uzyskanie skutecznego zestawu wzorców. Na możliwość poprawy wskazuje też lepsza skuteczność dostępnych komercyjnych implementacji – jak chociażby w cyfrowych aparatach fotograficznych.