

Rozprawa doktorska  
Paweł Masicz

*Metoda lokalizacji twarzy  
na obrazach cyfrowych  
z wykorzystaniem jej owalnego kształtu  
i procedury zmiennego promienia*

Promotor rozprawy:  
Dr hab. inż. Georgy Kukharev, prof. PS

## Spis treści

|                                                                                         |     |
|-----------------------------------------------------------------------------------------|-----|
| Przedmowa .....                                                                         | 2   |
| Wykaz ważniejszych skrótów i oznaczeń .....                                             | 5   |
| 1. Wprowadzenie .....                                                                   | 6   |
| 1.1. Wprowadzenie do problemu detekcji twarzy .....                                     | 6   |
| 1.2. Analiza algorytmów detekcji .....                                                  | 13  |
| 1.3. Bazy danych zdjęć .....                                                            | 26  |
| 1.4. Podsumowanie i wnioski .....                                                       | 30  |
| 1.5. Cel i teza pracy .....                                                             | 32  |
| 2. Opracowanie metody do rozwiązania zadania lokalizacji obszarów zawierających twarz.. | 33  |
| 2.1. Wprowadzenie do projektowania detektora twarzy .....                               | 33  |
| 2.2. Omówienie użytych algorytmów obróbki obrazów i metod normalizacyjnych .....        | 35  |
| 2.3. Omówienie bazowej metody lokalizacji obiektów owalnych .....                       | 43  |
| 2.4. Omówienie zmodyfikowanego algorytmu lokalizacji obiektów owalnych .....            | 49  |
| 2.4.1. Obróbka wstępna .....                                                            | 50  |
| 2.4.2. Macierze gradientów .....                                                        | 55  |
| 2.4.3. MT (Macierz Trafień) .....                                                       | 58  |
| 2.4.4. Zmienny promień poszukiwań .....                                                 | 62  |
| 2.4.5. ZMT (Zbiorcza Macierz Trafień) .....                                             | 69  |
| 2.4.6. Wyznaczanie pięciu obszarów zawierających twarz .....                            | 71  |
| 2.4.7. Wyznaczanie wielu twarzy .....                                                   | 71  |
| 2.4.8. Wersja dla obrazów w skali szarości .....                                        | 73  |
| 2.5. Podsumowanie i wnioski .....                                                       | 74  |
| 3. Eksperymentalne badania opracowanej metody .....                                     | 75  |
| 3.1. Podstawowe założenia .....                                                         | 75  |
| 3.2. Procedura testowa .....                                                            | 77  |
| 3.3. Badania na bazie FERET i autorskiej .....                                          | 79  |
| 3.4. Porównanie z wynikami konkurencyjnych systemów .....                               | 98  |
| 3.5. Podsumowanie i wnioski .....                                                       | 101 |
| 4. Opis programu badawczego Face Localizator .....                                      | 103 |
| 4.1. Założenia wstępne i wymagania sprzętowe .....                                      | 103 |
| 4.2. Omówienie możliwości .....                                                         | 104 |
| 4.3. Podsumowanie i wnioski .....                                                       | 112 |
| Zakończenie .....                                                                       | 113 |
| Bibliografia .....                                                                      | 116 |

## Przedmowa

Rozprawa poświęcona jest problematyce detekcji twarzy na obrazach cyfrowych dla zadań rozpoznawania ludzi w systemach biometrycznych.

Detektory twarzy niezwykle rzadko występują jako niezależne systemy. Przeważnie są jednym z kroków większych systemów osadzonych w szerokiej dziedzinie zadań związanych z rozpoznawaniem ludzi i niewątpliwie należą do ich kluczowych elementów, ponieważ wynik ich działania znacząco wpływa na ogólną efektywność. Systemy używające twarzy jako cechy poddawanej rozpoznawaniu są jednymi z najbezpieczniejszych i najwygodniejszych sposobów identyfikacji tożsamości, ponieważ wydają się być najbardziej intuicyjnym i naturalnym rozwiązaniem. Niezwykle istotnym czynnikiem przemawiającym za tym rozwiązaniem jest fakt, że pobranie próbki twarzy dokonywane jest w sposób bezinwazyjny, a co ważniejsze może odbywać się przy braku interakcji z identyfikowaną osobą. Inne, często wykorzystywane cechy biometryczne (np. tęczę oka, odciski palców), nie oferują takiej elastyczności. Kolejną, niewątpliwą cechą przemawiającą za użyciem twarzy do celów rozpoznawania, jest duża ilość baz danych je zawierających.

Mimo poświęcenia ogromnej ilości czasu i energii przez wielu naukowców na całym świecie aktualnie nie ma detektora twarzy, który byłby w stanie bezbłędnie lokalizować osoby w każdych warunkach. W dalszym ciągu trwają prace nad opracowaniem jeszcze efektywniejszych i co ważne – coraz szybszych detektorów, potrafiących sprawnie pracować nawet na biurkowych komputerach. Czas działania jest istotnym czynnikiem, który bezpośrednio wpływa na klasyfikacje do jakich zadań dany detektor może być użyty. Te najszybsze są najbardziej pożądane do zadań rozpoznawania ludzi w czasie rzeczywistym, np. bezinwazyjnej kontroli ruchu na lotniskach. Dodatkowo detektor powinien dobrze radzić sobie w niekorzystnych warunkach, do których należą: zmienne, kierunkowe oświetlenie, częściowe przesłonięcie, obrót czy elementy zakłócające. Powinien również poprawnie działać z obrazami czarnobiałymi jak i kolorowymi (wtedy informacja o kolorze powinna być wykorzystana dla poprawienia efektywności, a nie pomijana).

W najnowszych rozwiązaniach cały proces detekcji jest podzielony na dwa główne etapy: generowanie kandydatów i ich weryfikację przez algorytm „twarz/nie twarz”. W literaturze poświęcono wiele miejsca na opracowanie coraz lepszych rozwiązań obejmujących drugi blok, a ten pierwszy pozostawiano przeważnie w bardzo uproszczonym kształcie. Wynikało to z prostego przesłania, że algorytmy je realizujące nie mogą być zbyt skomplikowane, a przez to kosztowne czasowo, ponieważ moc obliczeniowa musiała być

skierowana do drugiego etapu detekcji. Nie zauważono przy tym, że zawężenie ilości próbek zlokalizowanych jako obszar potencjalnie zawierający twarz, równie skutecznie potrafi zwiększyć efektywność całego systemu, zarówno w dziedzinie czasu jak i skuteczności.

Wobec przedstawionych powyżej powodów celem pracy jest „Opracowanie metody i realizacja algorytmów do zadań lokalizacji obszarów zawierających twarz na obrazach cyfrowych”, a tezą, że: „Jednoczesne uwzględnienie informacji o owalnym kształcie twarzy i procedury zmiennego promienia może być podstawą do budowy skutecznej metody lokalizacji twarzy na obrazach cyfrowych, a dodatkowo użycie informacji o kolorze skóry zwiększy jej skuteczność”.

Przedmiotem badań są metody służące do zadań detekcji twarzy. Metodyka badań opiera się na analitycznym przeglądzie istniejących podejść i rozwiązań, a w efekcie zaproponowaniu metody, dającej w rezultacie lepszy wynik niż algorytmy bazowe, od których modyfikacje wyprowadzono.

Na całość pracy składają się z cztery rozdziały, zakończenie i bibliografia.

W rozdziale 1 przedstawiono przegląd światowych osiągnięć oraz usystematyzowano wiedzę w zakresie problemu detekcji twarzy na obrazach. Opierając się na tej podstawie zdefiniowano cel i tezę pracy.

W rozdziale 2 dokonano analizy wybranych algorytmów, przedstawiono ich słabe punkty, po czym zaproponowano zmiany umożliwiające ich wyeliminowanie, wliczając uzasadnienie ich użycia. Te czynności dały przesłankę do uzyskania lepszych wyników. W rozdziale dokonano również przeglądu i krótkiego omówienia metod obróbki wstępnej.

Rozdział 3 został poświęcony opisaniu procesu badawczego. Na jego potrzeby przeprowadzono eksperymentalne badania, mające na celu uzyskanie odpowiedzi, czy zaproponowane zmiany przekładają się na polepszenie wyniku. Dodatkowo wyznaczono optymalne parametry pracy algorytmów, dla których metoda zwraca najlepsze rezultaty. Ostatecznie przeprowadzono analizę uzyskanych wyników.

W rozdziale 4 przedstawiono implementację przedstawionych algorytmów w programie „FaceLocalizator”, zrealizowanego w języku wysokiego poziomu C++, wspartego przyjaznym, graficznym środowiskiem.

Głównym rezultatem jest opracowanie metody do zadań lokalizacji twarzy na obrazach cyfrowych z wykorzystaniem jej kształtu, koloru i procedury zmiennego promienia. Przy tym można wyróżnić następujące etapy pracy nad jego uzyskaniem:

- przegląd literatury naukowej poświęconej problemowi detekcji i lokalizacji twarzy,

- sformułowanie idei wykorzystania kształtu twarzy, procedury zmiennego promienia i macierzy gradientów do zadań lokalizacji twarzy, działających zarówno z obrazami kolorowymi jak i w skali szarości;
- eksperymentalne udowodnienie, że opracowana metoda uzyskuje lepsze wyniki od metod bazowych,
- stworzenie programu w języku wysokiego poziomu (C++) realizujący opracowany algorytm.

Autor chciałby wyrazić swoje serdeczne podziękowanie dla profesora Georgy Kukharev'a za podzielenie się swoją ogromną wiedzą i doświadczeniem, jak również za zaangażowanie i poświęcony czas.

## Wykaz ważniejszych skrótów i oznaczeń

- DCT - dyskretna transformata kosinusowa (ang. Discrete Cosine Transform);  
ERR - współczynnik równowagi (ang. Equal Error Rate);  
FaReS - system rozpoznawania twarzy (ang. Face Recognition System);  
FAR - współczynnik fałszywej akceptacji (ang. False Accept Rate);  
FRR - współczynnik fałszywego odrzucenia (ang. False Reject Rate);  
FFT - szybka transformata Fouriera (ang. Fast Fourier Transform);  
ICA - analiza niezależnych komponentów (ang. Independent Component Analysis);  
MT - macierz trafień  
KLT - transformacja Karhunen-Loeve'a (ang. Karhunen-Loeve Transform);  
LDA - liniowa analiza dyskryminacyjna (ang. Linear Discriminant Analysis);  
MLP - wielowarstwowy perceptron (ang. Multi Layer Perceptron);  
PCA - analiza komponentów głównych (ang. Principal Component Analysis);  
ZMT - zbiorcza macierz trafień

$M$  - wysokość obrazu w pikselach;

$N$  - szerokość obrazu w pikselach;

## 1. Wprowadzenie

### 1.1. Wprowadzenie do problemu detekcji twarzy

Wraz z niesamowicie szybkim rozwojem branży informatycznej, rośnie zainteresowanie nowymi technikami interakcji człowieka z maszynami (HCI – Human Computer Interaction). Oczywistym jest fakt, że komunikacja między ludźmi polegająca przede wszystkim na rozmowie i obserwacji rozmówcy nie została dotąd zaadoptowana do interakcji między człowiekiem i maszyną. Tu nadal dominuje klawiatura, uzupełniana przez myszkę. Również ekrany dotykowe nie posiadają tych cech, których używa się w kontaktach między ludzkich. Niesamowity wzrost zainteresowania poszukiwaniami rozwiązań problemów z dziedziny przetwarzania obrazów twarzy bazuje na przesłaniu, że informacje o intencjach, stanie i tożsamości użytkownika mogą być wyodrębnione z obrazu, a komputer może reagować stosownie do sytuacji, np. poprzez obserwowanie jej wyrazu.

Kolejnym powodem dla którego rozwijane są systemy bazujące na przetwarzaniu cech biometrycznych twarzy są kwestie bezpieczeństwa. Na przestrzeni ostatnich lat znacząco wzrosło zagrożenie terroryzmem, a także przestępczością gospodarczą związaną z nieupoważnionym dostępem do poufnych zasobów. Niezbędnym stało się opracowanie nowych systemów dostępu i kontroli, mogących zastąpić dotychczasowe rozwiązania bazujące na kartach magnetycznych i PIN'ach.

Zauważono, że niewątpliwie najlepszą metodą identyfikacji jest użycie danych płynących z wykorzystania cech biometrycznych. Biometryka [Woo97, Kyx01] określa metody pomiarów fizycznych cech lub charakterystycznego zachowania człowieka oraz metody ich porównywania z odpowiadającymi im danymi zapisanymi w bazie danych, przeznaczona do rozpoznania lub potwierdzenia jego tożsamości. Niewątpliwą zaletą biometryki jest wykorzystanie zasady bazującej na idei „czymś, czym jesteś” a nie na „czymś, co pamiętasz, czy masz przy sobie”. Ponieważ niezwykle trudno jest podrobić lub ukraść czyjeś indywidualne dane biometryczne, systemy dostępu lub identyfikacji je wykorzystujące należą do najbezpieczniejszych i najwygodniejszych narzędzi autoryzacji.

Najczęściej wykorzystywane cechy biometryczne wykorzystywane do identyfikacji to [Rih00]: twarz, odcisk palca, tęczówka i siatkówka oka, głos, geometria dłoni i palca, podpis, sposób pisania, kod DNA, kształt ucha, zapach, a nawet sposób chodzenia. Część z nich jest słabo poznana lub niemożliwa do realizacji z powodów technicznych – trudność pobierania próbek (np. DNA, zapach), niski współczynnik różnorodności (np. sposób chodzenia, kształt

ucha, zapach). Dlatego na co dzień, w praktycznych zastosowaniach, najczęściej korzysta się z obrazu twarzy, odcisku palca i tęczówki oka.

Jeśli jako główne czynniki decydujące o jakości wziąć nieinwazyjność, brak konieczności interakcji i możliwość wykorzystania istniejących systemów, niewątpliwie najbardziej atrakcyjnymi są metody bazujące na przetwarzaniu danych płynących z obrazu twarzy. Istniejące bazy danych zdjęć twarzy (np. policyjne), mnogość systemów monitoringu (np. na lotniskach) [BTT03] jak również coraz szybsze komputery pozwalają na szybkie i niekosztowne zbudowanie efektywnych systemów identyfikacji.

Od dłuższego czasu twarz ludzka była obiektem badań wielu naukowców, chcących opracować skuteczne algorytmy pozwalające dokonywać rozpoznawania tożsamości, jak również stanu emocjonalnego, jednak dopiero w ostatnim dziesięcioleciu XX wieku zostały one zintensyfikowane. Stało się tak ponieważ powstające wtedy komputery i ich dynamiczny rozwój pozwoliły wcielić w życie część, wcześniej czysto teoretycznych pomysłów [Kir90, Wys67].

O tym jak poważnie traktuje się przetwarzanie obrazu twarzy w celach poprawy bezpieczeństwa niech świadczy dyrektywa Komisji Europejskiej, która uznała twarz ludzką za główny identyfikator biometryczny, a swoim członkom nakazała ujednolicenie formatu wiz i paszportów poprzez umieszczenie w nich cyfrowych zdjęć do 2005 roku co zostało później odroczone o kolejne 3 lata. Podobne dyrektywy wydano w stosunku do wiz i pozwoleń na pobyt dla osób spoza UE [UE03].

Jednak, aby było możliwe jakiegokolwiek przetwarzanie obrazu twarzy zachodzi potrzeba zlokalizowania jej położenia na obrazie. Dla komputera każdy obraz to tablica zawierająca piksele o różnych wartościach i to co dla człowieka jest oczywiste, dla maszyny – niekoniecznie. Lokalizacja twarzy na obrazie okazała się problemem na tyle skomplikowanym, że dotychczas nie udało się stworzyć systemu dającego gwarancje poprawnej lokalizacji, co jest powodem ciągłego poszukiwania i ulepszania istniejących rozwiązań. Niewątpliwie ważnym pytaniem przy budowie detektora twarzy jest ustalenie wymagań jakie ma spełniać na wyjściu. Nie zawsze konieczne jest poszukiwanie wszystkich twarzy (np. w systemach kontroli dostępu), albo ważny jest czas przetwarzania (np. w systemach ciągłej obserwacji na lotniskach). Istotne jest również to, że detektory twarzy sporadycznie są używane jako samodzielne systemy, gdzie wynik ich działania jest jednocześnie wyjściem, lecz stanowią zaledwie pierwszy, choć bardzo istotny etap. Dlatego często kosztem jakości ulepsza się wynik w dziedzinie czasu.

Definicja lokalizacji twarzy: *Zadaniem lokalizacji twarzy na dostarczonym obrazie nazywamy ustalenie jakie jest położenie i wielkość każdej twarzy przy założeniu, że posiada się wiedzę na temat ilości ich występowania* [Yan02].

Definicja detekcji twarzy: *Zadaniem detekcji twarzy na dostarczonym obrazie jest ustalenie czy występują na nim twarze, a jeśli tak to zwrócenie położenia i wielkości każdej z nich* [Yan02].

Istnieje szereg czynników mogących w istotny sposób wpływać na charakterystykę wizerunku twarzy. Poza tym one same, ich wzorce, nie są jednakowe. Umysł ludzki potrafi doskonale klasyfikować te drobne różnice, nie mając problemu z prawidłową klasyfikacją lecz dla maszyny nie jest to zadanie łatwe. Powstaje więc skomplikowany problem lokalizacji wzorca, czyli twarzy, który na dodatek nie jest jednakowy. Równie niekorzystnym zjawiskiem jest stosunkowo częste występowanie tych niekorzystnych dla spójności wizerunku twarzy czynników, jak również ich kumulowanie się. Na rysunku 1.1 przedstawiono przykłady obrazów zawierające zakłócenia mogące sprawić trudność przy lokalizacji. Należą do nich [Phi99, Gro01, Kuk03a]:

- kierunek obrotu twarzy – obrazy twarzy zmieniają się w zależności od relacji położenia twarz-kamera (frontalny, 45 stopni, profil, „do góry nogami”), a niektóre cechy twarzy, takie jak nos lub oczy, mogą być częściowo lub całkowicie zasłonięte;
- obecność przeszkód – cechy twarzy takie jak broda lub wąsy, jak również przedmioty, np. okulary, mogą być obecne lub nie; ich ewentualne występowanie drastycznie zmienia charakterystykę twarzy, np. kształt, kolor i wymiar;
- wyraz twarzy – wygląd twarzy jest ściśle związany z jej wyrazem; twarz ludzka posiada największe skupisko mięśni na ludzkim ciele, mogące zniekształcać ją w znaczny sposób; do różnych wyrazów wliczamy między innymi uśmiech, smutek, mimikę;
- przesłonięcie – twarz może być częściowo lub całkowicie przesłonięta przez inny obiekt; w tej grupie pojawia się również wyjątkowo trudny przypadek, gdy na obrazach zawierających grupę ludzi, część twarzy może być przesłonięta przez jedną lub wiele innych twarzy;
- jakość i charakterystyka obrazu – podczas tworzenia obrazu czynniki takie jak światło (spektrum, charakterystyka, intensywność) i charakterystyka techniczna kamery (przesłona, rozdzielczość matrycy cyfrowej, jakość zastosowanej optyki, balans bieli itp.) znacząco wpływa na wygląd twarzy;
- przynależność rasowa – niewątpliwie duża różnorodność odcieni skóry dla różnych ras ludzi jest dodatkową przeszkodą w lokalizacji szczególnie, gdy wziąć pod uwagę, że

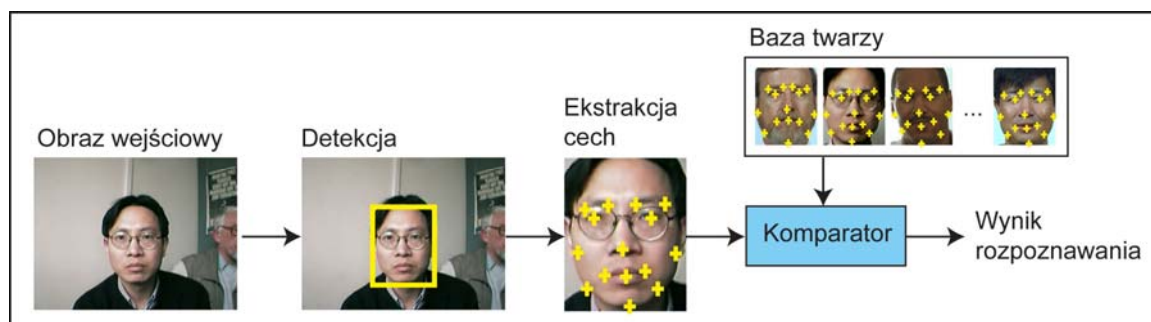
wiele systemów bazuje na segmentacji kolorów „twarzo-podobnych” uzależniając dalsze obliczenia od ich wystąpienia;



Rys. 1.1. Przykłady obrazów mogących sprawić problem przy procesie detekcji:

- a) obrót twarzy; b) obecność przeszkód – np. broda i okulary; c) wyraz twarzy;  
d) przesłonięcie; e) jakość obrazu – światło kierunkowe; f) przynależność rasowa

Istnieje bardzo wiele zagadnień z dziedziny przetwarzania obrazów zawierających twarz, które są ściśle związane z detekcją twarzy. Niemal wszystkie z nich do pełni automatycznego działania potrzebują skutecznych algorytmów zwracających położenie i wielkość twarzy. Poprawna detekcja jest warunkiem niemal niezbędnym i koniecznym ponieważ w dużym stopniu determinuje jakość wyników obliczanych w następujących po niej algorytmów. Poniżej (rys. 1.2) przedstawiono uproszczony schemat idei typowych, automatycznych systemów rozpoznawania ludzi na podstawie wizerunku twarzy.



Rys. 1.2. Schemat blokowy automatycznych systemów rozpoznawania ludzi na podstawie wizerunku twarzy

Możemy wyróżnić kilka problemów, do rozwiązania których niezbędne jest skorzystanie z metod lokalizacji twarzy [Kuk03a]:

- lokalizacja położenia twarzy – najprostszy przypadek, gdzie wynikiem systemu jest wynik lokalizatora;
- detekcja cech twarzy – celem takich systemów jest określenie istnienia i położenia cech takich jak oczy, nos, brwi, usta, uszy, itd. [Mas02a, Kas05];
- weryfikacja twarzy – twarz z obrazu wejściowego jest porównywana z jedną twarzą z bazy (problem porównania nazywany „jeden do jednego” – *ang.* „*one to one*”); często spotykana przy systemach kontroli dostępu;
- rozpoznawanie twarzy – obraz wejściowy jest porównywany z obrazami z bazy danych, po czym zwracany jest wynik określający tożsamość osoby badanej (problem porównania nazywany „jeden do wielu” – *ang.* „*one to many*”);
- tracking twarzy – celem trackingu jest obliczanie w czasie rzeczywistym położenia twarzy, jak również jej orientacji;
- rozpoznawanie ekspresji twarzy – rozwiązuje problem ustalenia stanu emocjonalnego występującego na twarzy, np. smutku, śmiechu, strachu, itp.

Podczas analizy wszelkich podejść do lokalizacji i detekcji twarzy, niezwykle istotnymi parametrami są te, dzięki którym można szybko ocenić efektywność danej metody. Zaliczamy do nich [Hje01, Yan02]:

- współczynnik detekcji (*ang.* *detection rate*) – jest to stosunek ilości twarzy zlokalizowanych przez algorytm detekcji do ilości twarzy rzeczywiście znajdujących się na obrazach;
- błędne odrzucenia (*ang.* *false negatives*) – określa, jak dużo obszarów obrazu opisujących twarz nie zostało zakwalifikowanych jako twarze;
- błędne przyjęcia (*ang.* *false positives*) – określa, jak dużo regionów obrazu zostało błędnie zakwalifikowanych jako twarz mimo, że jej nie opisuje.

Dokonując oceny systemów detekcji powinno się analizować te parametry, przy założeniu, że oceniamy wszystkie razem, a nie oddzielnie. Jest to istotne, ponieważ przeważnie występuje sytuacja, że gdy polepszamy jeden, efektywność innych maleje. Szacuje się, że istnieje około 150 różnych podejść do rozwiązania problemu detekcji twarzy [Yan02a]. Bardzo wiele z nich ma zapożyczenia z szerokiej grupy metod stosowanych do rozpoznawania tekstur czy kształtów – niekoniecznie twarzy. Często występuje również stwierdzenie, że problem detekcji twarzy można rozpatrywać jako dwuklasowe

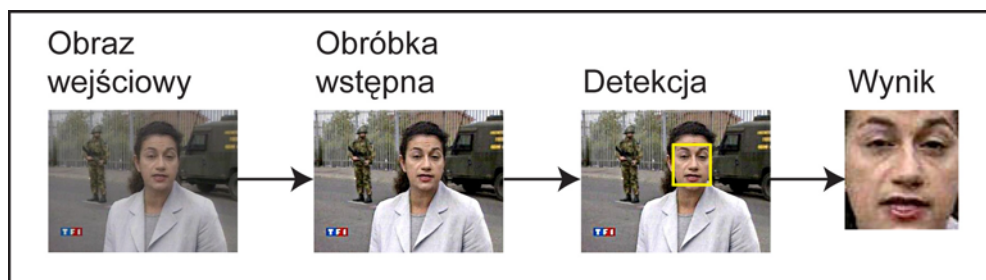
rozpoznawanie obrazu – klasy twarzy i „nie twarzy”. Oczywiście nie jest to łatwe, ponieważ istnieje wielkie zróżnicowanie wewnątrz klasowe.

Oprócz współczynników detekcji opisanych powyżej o jakości danego algorytmu detekcji decyduje również:

- szybkość działania – niezwykle istotny czynnik ponieważ zdecydowana większość zastosowań detektorów jest jedynie wstępnym procesem (np. w systemach kontroli dostępu – *ang. access control*); zaleca się pracę w czasie rzeczywistym (ponad 25 klatek na sekundę);
- wielkość twarzy możliwej do rozpoznania – określa wielkość twarzy (wyrażaną przeważnie w procentach w stosunku do wysokości obrazu wejściowego) którą system potrafi wykryć – zarówno maksymalną jak i minimalną;
- przestrzeń kolorów – przeważnie systemy budowane z dedykacją dla kolorowego obrazu są całkowicie nieprzydatne dla obrazów w odcieniach szarości, co znacząco obniża ich wartość komercyjną, gdyż kamery przemysłowe w znakomitej większości pracują właśnie w odcieniach szarości; należy wspomnieć, że dla odmiany systemy pracujące z obrazami w skali szarości doskonale radzą sobie również z kolorowym źródłem – po prostu implementowany jest dodatkowy, wstępny krok „normalizacyjny” zamieniający kolor na skalę szarości.

Niezależnie od użytej metody detekcji, każda korzysta z tej samej drogi umożliwiającej osiągnięcie sukcesu. Pierwszym krokiem jest zawsze dostarczenie obrazu w formie tablicy pikseli o określonej wielkości, gdzie wartość każdego z nich może być określona przy pomocy odcieni szarości (wartości 0-255) lub przy pomocy którejś z kolorowych przestrzeni barw – najczęściej przy pomocy przestrzeni RGB (trzy składowe – R: 0-255, G: 0-255, B: 0-255). Ten obraz jest następnie normalizowany przy pomocy narzędzi *pre-processingu* takich jak filtry, skalowanie, obracanie, redukcja cech itp. Następnie tak obrobiony materiał jest przekazywany kolejnemu blokowi funkcyjnemu – głównemu algorytmowi detekcji. Jego odpowiedni dobór w dużej mierze decyduje o efektywności całego systemu. Jednak źle przygotowane wcześniejsze kroki mogą spowodować uzyskanie znacznie gorszego rezultatu, niż był możliwy. Oczywiście ostatecznym wyjściem każdego systemu detekcji jest położenie i wielkość każdej z wykrytych twarzy zapisana przeważnie jako położenie jednego z narożników prostokąta opisującego obszar zawierający twarz i wielkości jego jednego z boków (drugi jest obliczany na zasadzie proporcjonalności). Każdy z bloków może być bardzo rozbudowany, jednak niezaprzeczalnie największe zróżnicowanie

występuje w bloku „Detekcja”, który zawiera w sobie cały – nierzadko bardzo skomplikowany – mechanizm detekcji. Rysunek 1.3 ukazuje ideę budowy detektorów.



Rys. 1.3. Schemat blokowy systemów detekcji twarzy

W najnowszych podejściach można zauważyć podział bloku detekcji na predetekcję, nazywaną wyodrębnianiem kandydatów i algorytm decyzyjny twarz/„nie twarz”. Jest to spowodowane obserwacją, że lepsze rezultaty w dziedzinie czasu otrzymuje się, gdy detektor najpierw jedynie pobieżnie zlokalizuje obszary potencjalnie zawierające twarz. Pozwala to zaoszczędzić sporo mocy obliczeniowej komputerów i skierować ją do drugiego kroku, w którym następuje decyzja czy wyodrębnione obszary są twarzą, czy też nie.

Niewątpliwie etap wyodrębniania tych potencjalnych obszarów zawierających twarz jest mniej istotny od drugiego – decyzji twarz/„nie twarz”, jednak ma swój ogromny wpływ na szybkość działania i ostateczną skuteczność detekcji. Dodatkowo jeśli pozwala zawęzić ilość próbek poddawanych dalszej analizie do kilku sztuk (gdzie ilości kombinacji w typowym obrazie 320x240 jest kilkaset tysięcy), a do tego wykazuje duże prawdopodobieństwo (np. ponad 90%), że zawiera się w nich poszukiwany wycinek zawierający poszukiwaną twarz, to warto jest poświęcić czas na opracowanie szybkich i efektywnych algorytmów go realizujących. W literaturze traktującej o detekcji twarzy proces lokalizowania tych obszarów jest obecny, jednak znacznie większą uwagę poświęca się głównemu blokowi decyzyjnemu, czyli twarz/„nie twarz”. Zachodzi więc duża szansa na opracowanie algorytmów zwiększających skuteczność detektorów, a dodatkową korzyścią jest możliwość ich użycia w wielu istniejących już systemach dodatkowo zwiększając ich efektywność.

Należy zwrócić uwagę, że starsze systemy cechowały się znacznie mniejszym skomplikowaniem od obecnych, jednak główna idea pozostaje od lat ta sama. Większość zmian to modyfikacje tzw. *tuningi*, albo użycie mieszanki różnych podejść. Mimo to, uzyskuje się coraz większe współczynniki detekcji, co rokuje nadzieje, że problem detekcji

twarzy będzie kiedyś ostatecznie rozwiązany. Pomaga przy tym nieustanny wzrost mocy obliczeniowej komputerów, nawet tych będących na wyposażeniu przeciętnego użytkownika.

## 1.2. Analiza algorytmów detekcji

W celu odpowiedniego zrozumienia mechanizmów kierujących większością podejść do problemu detekcji, a przez to i zasadność użycia skuteczniejszych lokalizatorów obszarów potencjalnie zawierających twarze, najlepiej przyjrzeć się kilku rozwiązaniom będących realnymi, działającymi systemami. Należy przy tym dodać, że mnogość rozwiązań problemu detekcji twarzy na obrazach jest ogromna. Bardzo często metody użyte nie są nowatorskie, a jedynie zaadoptowane od innych zadań, szczególnie z dziedziny rozpoznawania i klasyfikacji obiektów. Niewątpliwie, aby dokładnie i efektywnie przyjrzeć się światowym osiągnięciom z dziedziny detekcji twarzy należy dokonać pogrupowania metod. Nie jest to zadanie proste, ponieważ większość z nich nie jest przypisana do jednego typu algorytmów. Niejednokrotnie okazuje się, że najlepsze wyniki mają te, które wykorzystują najlepsze cechy z wielu podejść. Kilka lat temu pojawiła się propozycja podziału zawierająca cztery grupy, a która obowiązuje do dziś [Yan02]. Rysunek 1.4 ilustruje ten podział.



Rys. 1.4. Podział detektorów zaproponowany przez Yang'a [Yan02].

Grupy te można scharakteryzować w następujący sposób [Yan02]:

- bazujące na wiedzy – te metody wykorzystują zasady jakie ludzka wiedza posiada na temat tego, co zawiera typowa ludzka twarz; najczęściej wykorzystywane do lokalizacji.
- używające niezmienników – metody z tej grupy starają się znaleźć takie cechy twarzy, które są niezależne od pozy, kąta obrotu czy warunków oświetleniowych i użyć ich do lokalizacji twarzy; najczęściej wykorzystywane do lokalizacji.
- bazujące na poszukiwaniu wzorcem – kilka wzorców całej twarzy bądź jej komponentów są użyte do poszukiwań na obrazie; wyniki korelacji pomiędzy tymi wzorcami, a obrazem wejściowym służą do decyzji czy dany obraz zawiera twarz; wykorzystywane zarówno do lokalizacji jak i detekcji.

- bazujące na obrazie – w odróżnieniu od metod bazujących na poszukiwaniu wzorcem w tej grupie model detektora jest nauczany zestawem obrazów zawierających w sobie reprezentatywną zmienność ludzkiej twarzy w zmiennych warunkach; te nauczone modele są następnie wykorzystywane do detekcji.

Większość z tych grup ma swoje podgrupy, które są konsekwencją użycia odmiennych metod jednak spełniających warunek przynależności do grupy głównej [Yan02]:

- metody bazujące wiedzy [Kot97, Mia99].
- metody korzystające z niezmienników:
  - komponenty twarzy [Ami98, Bur95, Gra95, Han98, Leu95, Yow96, Yow97, Yow98],
  - tekstura twarzy [Aug93, Dai96],
  - kolor skóry [Alb00, Cha98, Jon02, Kuk04b, Mat02, San02, Ter98],
  - multi-cechy twarzy [Gar99, Gra00, Hot98, Sab98, Yan98].
- metody porównywania ze wzorcem:
  - przeddefiniowane wzorce twarzy [Cra92, Sca98],
  - zdeformowane wzorce [Lan95].
- bazująca na obrazie:
  - eigenface [Tur91],
  - bazujące na rozkładzie [Yan00, Dut98, Edw98, Yan00],
  - sieci neuronowe [Ana03, Ben99, Ber98, Row95, Row98a, Row98b],
  - Support Vector Machine (SVM) [Hei00, Hei01, Kim02, Osu97, Sun98],
  - Naive Bayes Classifier [Sch98, Sch00],
  - Ukryte Modele Markov'a (ang. *Hidden Markov Model* – HMM) [Das02, Nef98],
  - Podejście informacyjno-teoretyczne [Col97, Col99, Ela02, Lew96, Hua96].

Aby dokładniej przybliżyć każdą z grup poniżej przedstawione są rozszerzone charakterystyki i przykłady gotowych systemów dla każdej z grup.

### **Podejścia bazujące na wiedzy**

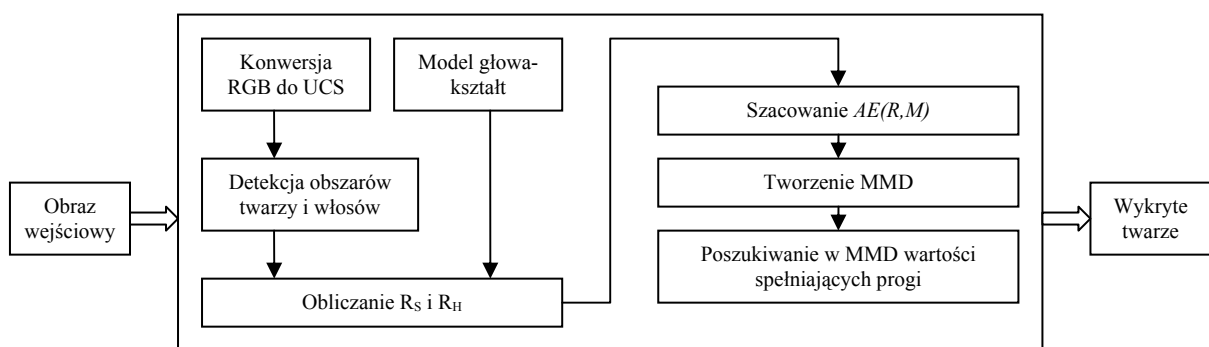
W rozwiązaniach bazujących na wiedzy wykorzystuje się bazę wiedzy człowieka o ludzkiej twarzy. Algorytmy z tej grupy wyszukują obszary, które spełniają warunki pozwalające traktować je wstępnie jako twarz (to etap generatora kandydatów), a następnie badają ich właściwości i porównują z zależnościami, które według ludzkiej wiedzy muszą wystąpić na twarzy (decyzja twarz/„nie twarz”). Dla przykładu, na twarzy najczęściej znajdują się dwoje oczu, a jeśli określimy ich położenie to nietrudno wskazać położenie ust i nosa. Dlatego cechą występującą w niemalże każdym systemie bazującym na wiedzy jest

rozpoczynanie procesu detekcji od poszukiwania cech twarzy. Następnie obszary, w których je znaleziono (nie jest wtedy jeszcze do końca pewne czy te cechy są poprawnie określone) są badane według określonych wcześniej reguł i zostają zweryfikowane, bądź odrzucane.

Niezaprzeczalną trudnością powstającą przy budowie systemów tego typu jest problem odpowiedniego zakodowania ludzkiej wiedzy do postaci zrozumiałej dla komputera. Jeśli reguły są zbyt szczegółowe, to dostarczona do systemu twarz będąca częścią obrazu może nie przejść wszystkich testów i zostać odrzucona. Z drugiej strony zbyt ogólne reguły mogą spowodować zbyt wiele błędnych przyjęć. Poza tym, niezmiernie trudnym problemem jest zaadoptowanie tej grupy metod (bazujących na wiedzy) do detekcji twarzy we wszystkich możliwych pozach, ponieważ opisanie regułami wszystkich przypadków to zadanie ogromne. Należy jednak zauważyć, że dla twarzy w widoku frontalnym jakość detekcji jest wysoka.

Przykładem obrazującym tę grupę jest podejście bazujące na użyciu teorii rozmytej w celu detekcji twarzy na obrazach kolorowych [Wu99].

Na samym początku algorytm startuje od konwersji kolorów pikseli z przestrzeni RGB do przestrzeni UCS (*perceptually Uniform Color System*) [Wys67] (patrz rys. 1.5). Przebiega to poprzez między-konwersję RGB do CIE XYZ, a następnie do UCS. W tej przestrzeni utworzono modele kolorów twarzy SCDM (*Skin Color Distribution Model*) i włosów HCDM (*Hair Color Distribution Model*) poprzez ręczne zaznaczenie kolorów i wyznaczenie dystrybucji. Model włosów jest wyznaczany nieco inaczej, ponieważ użyto mieszanki luminancji z chrominancją. Stało się tak, gdyż piksele opisujące włosy są ciemne, a przez to informacja chrominancji jest niestabilna i musi być wsparta przez luminancję.



Rys. 1.5. Schemat blokowy algorytmu detekcji bazującego na sieciach rozmytych [Wu99a].

W kolejnym kroku algorytm dokonuje detekcji obszarów twarzy i włosów. Przetestowano przy tym równoległe przestrzeń kolorów HSV, ale wyniki jednoznacznie wskazują na przewagę UCS. Do celów detekcji stworzono modele głowa-kształt w pięciu pozycjach: frontalnej, lewy i prawy pół-profil, lewy i prawy profil (wyznaczono

współczynniki  $M_F$  i  $M_H$ ). Ponieważ informacja uzyskana z detektora kolorów twarzy i włosów i modeli głowa-kształt nie dają się porównywać przy pomocy dobrze znanych metod *template matching* zaproponowano metodę nazwaną *fuzzy pattern matching*. Dla każdego kwadratowego okna pobranego z obrazu wejściowego wyznaczone są zestawy rozmyte  $R_S$  i  $R_H$  odpowiadające proporcjom pikseli „twarzo/włoso-podobnych” do ogólnej liczby pikseli (rys. 1.5). Następnie szacowane jest podobieństwo  $AE(R_{S,H}, M_{F,H})$  pomiędzy uzyskanymi zestawami rozmytymi z badanego obrazu, a tymi wyznaczonymi z modeli.

Operacja ta jest przeprowadzana dla wszystkich możliwych okienek w danym rozmiarze. Następnie rozmiar okienka jest zwiększany i operacja się powtarza. Gdy zostaną wyczerpane możliwe wielkości okienka, tworzona jest zbiorcza tablica MMD (*map of matching degree*) (patrz rys. 1.5) zawierająca maksymalne wartości  $AE$  dla każdego piksela i dodatkowo przechowująca rozmiar okna dla tej maksymalnej wartości.

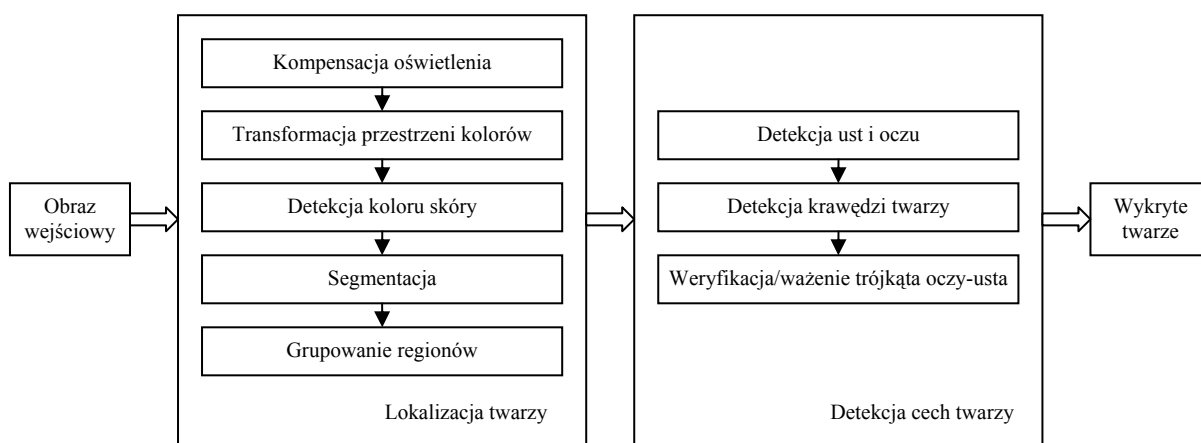
Eksperymenty przeprowadzono przy użyciu strumienia wideo pobranego ze zwykłej internetowej kamery, z parametrami ustawionymi na auto. Dla każdej rasy (azjaci, biali,...) stworzono modele SCDM i HCDM w sposób manualny. Niestety nie jest rozwiązano problemu automatycznego ich doboru. Według autorów wyniki są obiecujące, wynoszące nawet 100% dla odpowiednio dużych twarzy, a 90% dla małych. Ilość błędnych przyjęć rośnie wraz ze zmniejszającą się wielkością twarzy. Prędkość wynosiła ok. 2,5 fps na procesorze 266MHz. Metoda ma jednak swoje ograniczenia i wady, np. generuje złe wyniki przy bardzo pokolorowanych obrazach, jest nieodporna na przesłonięcia, bliskość dwóch lub więcej twarzy. Ale przede wszystkim nie potrafi rozpoznawać „łysych” i ludzi z odzieniem głowy i co ważne – konieczna jest informacja o pełnym kolorze. Dlatego nawet autorzy nazywają zlokalizowane twarze jako wykrycie kandydatów, a nie detekcję.

### **Podejścia używające niezmienników**

Algorytmy używające niezmienników, w odróżnieniu od metod bazujących na wiedzy, poszukują tych cech twarzy, które są niezmiennie i występujące na wszystkich twarzach. Wykorzystują one spostrzeżenie, że człowiek potrafi bezbłędnie rozpoznać daną twarz niezależnie od jej wyrazu, obrotu, czy warunków oświetleniowych. Dlatego muszą istnieć takie niezmiennie cechy, które to umożliwiają. Do tych cech należą: brwi, oczy, nos, usta, linia włosów. Do ich wyodrębnienia używane są najczęściej algorytmy detekcji krawędzi. Po ich zlokalizowaniu sprawdzane są zależności biometryczne występujące pomiędzy nimi. Tutaj z pomocą przychodzą wieloletnie badania kryminalistyczne. Problemami pojawiającymi się podczas używania metod dokonujących detekcji

niezmiennych cech, jest ich duża wrażliwość na zaistniałe zakłócenia, tj. zmiana oświetlenia, szum, przesłonięcie, niska jakość materiału wejściowego. Przykładowo silne światło kierunkowe padające z boku na twarz powoduje zatarcie krawędzi, które algorytm próbuje znaleźć, co z kolei uniemożliwia ich sprawne zlokalizowanie i w efekcie błędny wynik.

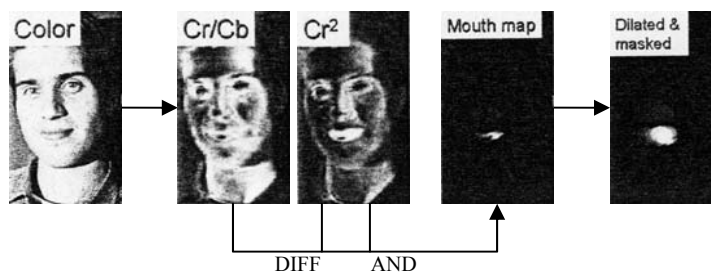
Dobrym przykładem obrazującym tę grupę jest system posługujący się segmentacją kolorów i poszukiwaniu niezmienników [Hsu02]. Zastosowane algorytmy są zorganizowane w sposób widoczny na rysunku 1.6.



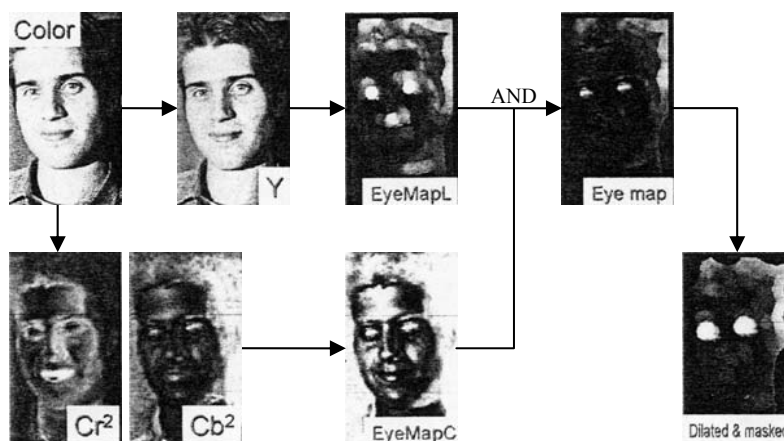
Rys. 1.6. Schemat blokowy algorytmu detekcji opartego na sieciach segmentacji kolorów i poszukiwaniu niezmienników [Hsu02a].

W uproszczeniu wszystkie użyte procedury bazują na analizie pikseli i ich przekształceniach w taki sposób, żeby uzyskać szukane elementy. Dostarczony obraz jest kompensowany pod względem światła, a następnie przeprowadzana jest segmentacja kolorów „skóro-podobnych”. Do tego celu użyto przestrzeni kolorów  $YCrCb$  [Jon02a]. Dowiedziono, że kolor skóry jest nieliniowo zależny od luminancji.

Używając odpowiednich, nieliniowych przekształceń i zakresów w poszczególnych składowych przestrzeni  $YCrCb$  tworzeni są kandydaci do bycia twarzą. Aby ich zweryfikować pod kątem „twarzowości” lokalizowane są oczy i usta. Do tego celu użyto specjalnych przekształceń widocznych na rysunkach 1.7 i 1.8.



Rys. 1.7. Zilustrowanie metody tworzenia maski umożliwiającej efektywne lokalizowanie ust [Hsu02a].



Rys. 1.8. Zilustrowanie metody tworzenia maski umożliwiającej efektywne lokalizowanie oczu [Hsu02a].

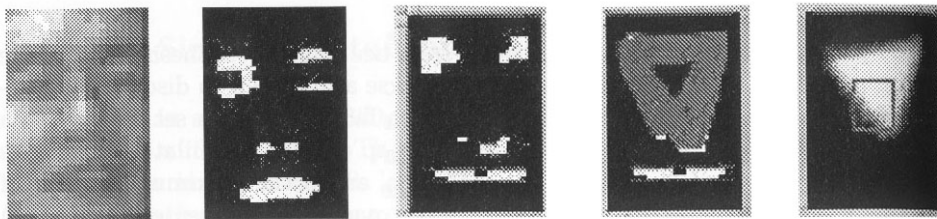
Następnie przy pomocy wariacji luminancji, geometrii trójkąta oczy-usta i predykcji, poszukiwana jest granica twarzy. Na sam koniec użyto transformacji Hough'a w celu ostatecznego obrysowania elipsą wykrytej twarzy. Metoda bazuje na wielu parametrach i progach przez co można domniemać, że jest niejako skazana na niską skuteczność. Wyniki pokazały jednak, że nie jest to zgodne z prawdą. Detekcja na poziomie 89-91% jest przyzwoitą wartością, jednak czas przetwarzania obrazu o wymiarach 640x480 pikseli wynoszący średnio 25 sek. jest zdecydowanie za długi (procesor 1,7GHz). Poza tym nie określono co uważa się za prawidłowo zlokalizowaną twarz. Warto zaznaczyć, że powodem tak długiego czasu jest – w największym stopniu – detekcja cech twarzy. Również to, że system pracuje na obrazach kolorowych eliminuje możliwość przetestowania go na typowych bazach danych do tego celu stworzonych (FERET, CMU, itd.). Dlatego wszystkie testy przeprowadzone zostały na własnych, autorskich zestawach.

### Podejścia bazujące na poszukiwaniu wzorców

W metodach przeszukiwania wzorcem, standardowy wzorec (najczęściej frontalny) jest ręcznie predefiniowany lub sparametryzowany przez funkcje. Podając obraz do przetworzenia, oblicza się niezależną korelację pomiędzy tymi wzorcami, a konturem twarzy, oczami, nosem i ustami. Decyzja o tym czy twarz występuje na obrazie czy nie, zależy od wartości korelacji. Metody te mają wielką zaletę przejawiającą się w prostocie implementacji. Mimo to dowiedziono, że nie za dobrze radzą sobie z detekcją twarzy w różnych skalach, pozach czy kształtach. W celu zwiększenia niezawodności i odporności na te zmienne warunki stosuje się „podwzorce” (w różnych rozdzielczościach i wielkościach) jak również zdeformowane wzorce. Niestety te udoskonalenia zwiększają zapotrzebowanie tych metod na

moc obliczeniową, a ponieważ szybkość działania jest jedną z nielicznych cech pozytywnie wyróżniającą tę metodę na tle innych, nie mają one wielkiego zastosowania. Nie ulega wątpliwości, że idea porównywania ze wzorcem jest znacznie częściej wykorzystywana do detekcji cech twarzy niż do samej detekcji.

W jednym z systemów zastosowano poszukiwanie wzorcem w pierwszym kroku algorytmów [Mar98]. Zaproponowano użycie trzech wzorców w trzech rozdzielczościach, każdy o połowę mniejszy od poprzedniego. Te trzy wzorce  $I_1$ ,  $I_2$ ,  $I_3$  są porównywane za pomocą korelacji z trzema odpowiadającymi im pod kątem rozdzielczości wzorcami  $T_1$ ,  $T_2$ ,  $T_3$ . Zauważono, że w słabych warunkach oświetleniowych, a na pewno przy świetle kierunkowym korelacje są mało skuteczne. Jednocześnie zwrócono uwagę, że wynik porównywania znacznie się poprawia, gdy zastosuje się podwójną korelację, dla każdej pionowej połówki twarzy oddzielnie i wybraniu maksimum do porównywania ze wzorcem. W ten sposób znacznie poprawia się jakość wyników. Ten pierwszy korelacyjny krok jest przeprowadzany w celu wyodrębnienia 10 obszarów najlepiej spełniających warunki do bycia twarzą. Następnie wyznaczone obszary są przekazywane do kolejnego kroku wyodrębniającego geometryczne cechy – oczy i usta. Zastosowane metody bazują na prostych odcięciach co ilustruje rysunek 1.9.



Rys. 1.9. Zobrazowanie kroków normalizacyjnych i sposobu działania algorytmu detekcji cech geometrycznych twarzy [Mar98]

Kolejnym krokiem algorytmu jest klasyfikacja. Do zbudowania modelu twarzy i „nie twarzy” zaproponowano użycie kompilacji algorytmów PCA i LDA. Do ostatecznej decyzji czy dany obszar jest twarzą posłużył rozkład statystyczny. Określono również wielkość minimalną i maksymalną poszukiwanej twarzy, a przedział ten jest bardzo wąski od  $30 \times 50$  pikseli do  $50 \times 80$  pikseli. Wyniki eksperymentów dokonanych na autorskiej bazie oscylujące wokół 90% nie są bardzo wysokie jednak wykazują swą przewagę nad większością algorytmów bazujących na segmentacji kolorów i poszukiwaniu kształtów szczególnie, gdy twarz jest obrócona i kierunkowo oświetlona.

Przykład opisany powyżej potwierdza, że pomieszanie trzech metod: poszukiwania wzorcem (3 wzorce i badanie połówek twarzy), używających niezmienników (cechy geometryczne) jak i bazujących na obrazie (PCA) jest często używane w celu osiągnięcia jak najlepszych rezultatów w każdym z etapów detekcji, jak również dla ograniczenia czasu wykonywania algorytmów.

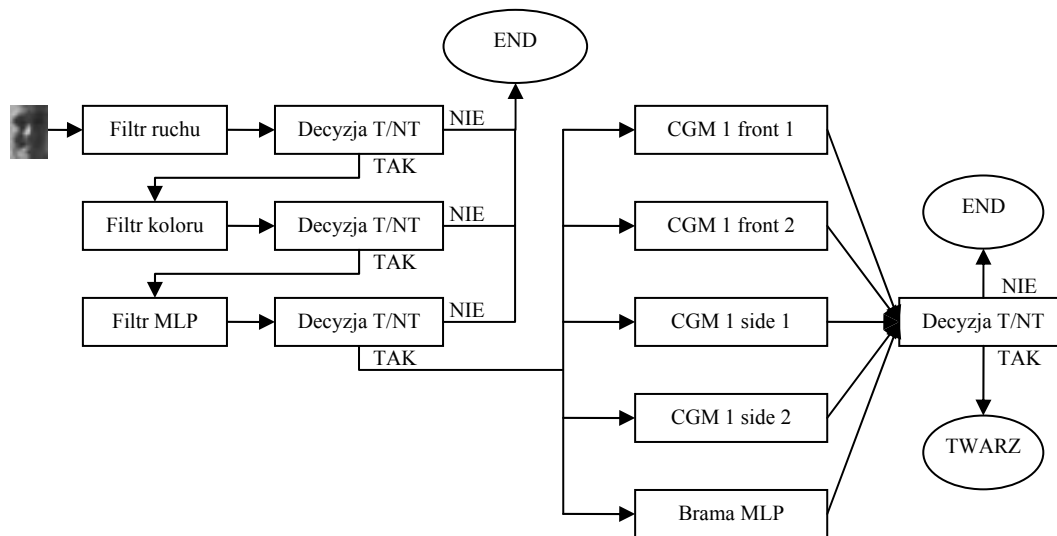
### **Podejścia bazujące na obrazie**

W odróżnieniu od metod bazujących na poszukiwaniu wzorcem, gdzie wzorce są predefiniowane przez ekspertów (statyczne), w grupie metod bazujących na obrazie, wzorce generowane są na podstawie dostarczonych zbiorów obrazów uczących [Sta02]. Uogólniając, można powiedzieć, że opierają się na analizie statystycznej i algorytmach uczących (*ang. machine learning*) w celu znalezienia zależności pozwalających określić czy badany obszar obrazu to twarz czy nie. Dynamicznie (podczas procesu uczenia) uzyskane wzorce (przeważnie zakodowane w funkcjach rozkładu bądź dyskryminacyjnych) są używane do detekcji. Bardzo ważną sprawą jest odpowiednia redukcja przestrzeni cech wpływająca na ilość obliczeń, a także powodująca zwiększenie skuteczności detekcji. Wiele metod bazujących na wyglądzie można rozumieć jako strukturę prawdopodobieństwa. Obraz lub wektor cech otrzymanych z obrazu jest uważany za losową zmienną  $x$ , a ta zmienna losowa jest scharakteryzowana jako twarz lub „nie twarz” poprzez funkcję gęstości klasowo-warunkową  $p(x|face)$  i  $p(x|nonface)$ . Do rozstrzygnięcia czy dany obszar jest twarzą czy nie najczęściej stosuje się klasyfikator Bayesian’a lub maksymalne prawdopodobieństwo.

Dużą trudnością przy konstruowaniu systemów korzystających z dziedziny metod bazujących na wyglądzie, jest budowa odpowiednich zestawów uczących. Muszą one zawierać przykłady twarzy i „nie twarzy”, przy założeniu, że zakres będzie możliwie jak najbardziej reprezentatywny. O ile zestaw uczący twarzy jest stosunkowo łatwo stworzyć, ponieważ mimo swoich subtelnych różnic twarze są stosunkowo do siebie podobne, to realizacja klasy „nie twarzy” jest znacznie trudniejsza. Dzieje się tak dlatego, że ilość kombinacji możliwych „wyglądów” tła (które nie jest twarzą), jest bliska kombinacji iluminacji wszystkich pikseli wchodzących w skład zestawu wejściowego systemu.

Jedną z metod bazujących na obrazie jest ta korzystająca z sieci neuronowych [Fer01]. Jego idea została przedstawiona na rysunku 1.10. Na każdym dostarczonym obrazie algorytm przeprowadza wycinanie okienek, które następnie są analizowane. Krokiem normalizacyjnym jest skalowanie do wymiarów 15x20 i rozciągnięcie histogramu. Założono, że najmniejszy wycinek może mieć wielkość 15x20, a największy 150x200. Dla obrazu 150x200 daje to ok.

500,000 obrazów do analizy. Jest to duża liczba, więc zastosowano algorytmy redukujące ich ilość.



Rys. 1.10. Schemat blokowy algorytmu detekcji opartego na sieciach neuronowych i wnioskowaniu [Fer01a].

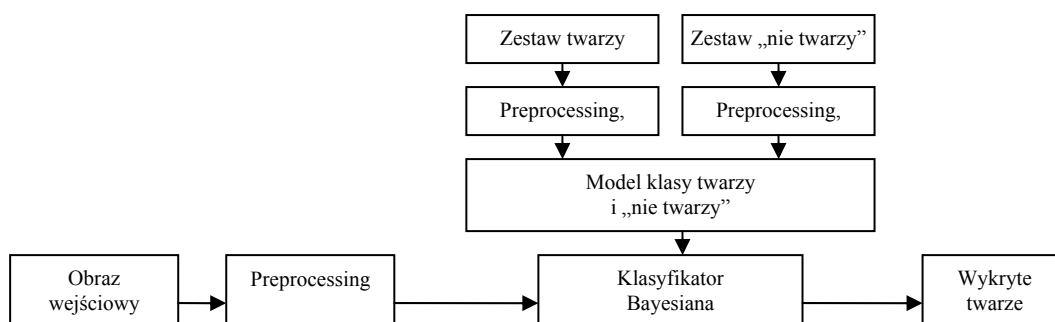
Pierwszy to filtr ruchu (patrz rys. 1.10) bazujący na prostych odcięciach, ale skutecznie eliminujący ponad 90 procent hipotez. Kolejny filtr redukuje ilość pozostałych hipotez o 60 procent, a dokonuje tego segmentując kolory skóro-podobne. Na sam koniec zaaplikowany jest prosty wielowarstwowy perceptron, posiadający 300 wejść, 20 neuronów w warstwie ukrytej, co daje razem 6041 wag. Do nauczania tej sieci użyto 8000 wizerunków twarzy i 50000 „nie twarzy”. Sieć ma niezłe wyniki w dziedzinie pozytywnej detekcji – 99%, ale negatywnych przyjęć ma też sporo – 1%, dlatego nie może być użyta jako kończący etap. Oczywiście wszystkie obrazy poddawane analizie przez tą sieć są przeskalowane do wielkości 15x20. Po tych obliczeniach pozostaje ok. 600 obrazów przekazywanych do kolejnego, głównego algorytmu.

W celu ostatecznego zdecydowania o istnieniu twarzy lub nie zdecydowano się wykorzystać *Constrained Generative Model (CGM)*. Polega on na obliczaniu prawdopodobieństwa czy dany obraz jest twarzą lub nie przy użyciu projekcji  $P$ , a następnie odległości euklidesowej do wzorcowych twarzy. Im więcej wzorców tym lepsza jakość, lecz także zwiększa się ilość operacji. W celu jej zmniejszenia użyto sieci neuronowej aproksymującej projekcję zestawu  $P(x)$ . Ma ona tyle samo wejść i wyjść (15x20), i dwie warstwy ukryte zawierające kolejno 35 i 50 neuronów. Podstawą do przyjęcia tezy, że dany wycinek przedstawia twarz jest spełnienie jeśli obliczona minimalna odległość do którejś z

twarzy wzorcowych jest poniżej pewnego progu. Aby zwiększyć skuteczność detekcji obroconych twarzy posłużono się kilkoma CGM'ami, dla każdej z pozycji (patrz rys. 1.10). Dodatkowo posłużono się wielowarstwowym perceptronem w celu dodatkowego wnioskowania czy poprawność wyników zwróconych przez CGM'y jest akceptowalna. Wszystkie te bloki funkcyjne prezentuje rysunek 1.4. Metoda została przetestowana na dwóch bazach danych. Pierwsza, *Sussex Face Database* (10 twarzy po 10 zdjęć), zawierająca twarze od lewego do prawego profilu – w celu sprawdzenia przydatności do detekcji twarzy obroconych. Zanotowano dobry wynik dla twarzy widocznych z profilu (100%), ale systematycznie malejący z rosnącym kątem obrotu (do 37%). I druga, standardowa, CMU Test Set A, zawierająca 42 obrazy z wieloma twarzami. Tutaj wyniki wahały się pomiędzy 77-85% w zależności od użytych kombinacji.

Jednym z nowszych algorytmów służącym detekcji wielu frontalnych twarzy na obrazach jest ten korzystający z BDF'a (*ang. Bayesian Discriminating Features*) [Liu03]. Nowatorstwo tej metody polega na integracji analizy cech z obrazu wejściowego, statystycznego modelowania klas twarzy/„nie twarzy”, i ostatecznej decyzji przy użyciu klasyfikatora Bayesian'a (patrz rys. 1.11). Analiza cech pozwalająca wyznaczyć wektor cech potrafiący udźwignąć ciężar detekcji, jest reprezentowana przez kombinację obrazu wejściowego, jednowymiarowej reprezentacji falkowej Harr'a i projekcji amplitudy. Udowodniono, że reprezentacja Harr'a jest efektywna do celów detekcji, a projekcja amplitudy doskonale chwyta wertykalną symetrię twarzy i horyzontalną jej charakterystykę. Statystyczne modelowanie klasy twarzy i „nie twarzy” wyznacza prawdopodobieństwo funkcji gęstości dwóch wymienionych klas. Niezaprzeczalną trudnością przy przygotowywaniu zestawów uczących jest to, że o ile klasę twarzy jest stosunkowo łatwo przygotować, o tyle klasa „nie twarzy” jest praktycznie niemożliwa do zaadoptowania. Posłużono się więc filozofią znaną z zastosowań SVM. Polega ona na tworzeniu klasy „nie twarzy” leżącej bardzo blisko klasy twarzy. Aby zamodelować obie klasy posłużono się rozkładem normalnym w którym kluczową rolę pełni przekształcenie PCA (*ang. Principal Component Analysis*), a dokładniej 10 pierwszych komponentów. Te rozkłady normalne są wyznaczane z 1200 próbek twarzy i 4500 próbek „nie twarzy” wybranych ręcznie, ponieważ muszą być bardzo zbliżonych do wyglądu twarzy. Na samym końcu użyty zostaje klasyfikator Bayesian'a, który daje możliwość decyzji czy dostarczony do systemu obraz jest twarzą czy nie. Oczywistym jest fakt, że obraz dostarczony do systemu jest analizowany okienko po okienku. Najpierw jest skalowany do rozdzielczości 16x16, następnie wyznaczana jest jednowymiarowa reprezentacja falkowa Harr'a i projekcja amplitudy. Z obrazu

wejściowego i tych policzonych, dodatkowych wektorów tworzony jest nowy wektor cech, przy pomocy którego, jak również modeli statystycznych, obliczana jest jego przynależność do klasy twarzy lub „nie twarzy”. Cały algorytm przedstawia rysunek 1.11.



Rys. 1.11. Schemat blokowy algorytmu detekcji bazującego na BDF [Liu03a].

System został przetestowany przy użyciu bazy twarzy FERET i CMU. Uzyskane rezultaty są bardzo zadowalające – 97,4% detekcji przy tylko jednej błędnej (dla CMU - 80 obrazów zawierających 227 twarzy). Jednocześnie dla bazy FERET uzyskano średni wynik wynoszący 98,7% przy żadnej błędnej. Jednak tak dobry rezultat jest spowodowany nauczaniem modelu przy pomocy próbek pochodzących właśnie z bazy FERET. Dokonano również próby detekcji obróconych twarzy przy pomocy prostego rotowania obrazu wejściowego. Wynik jest zadowalający jednak odbija się na czasie detekcji, który nie jest krótki już dla standardowego, nie obróconego obrazka wynosi średnio ok. 1 sekundy na komputerze 900MHz Sun Blade 5000.

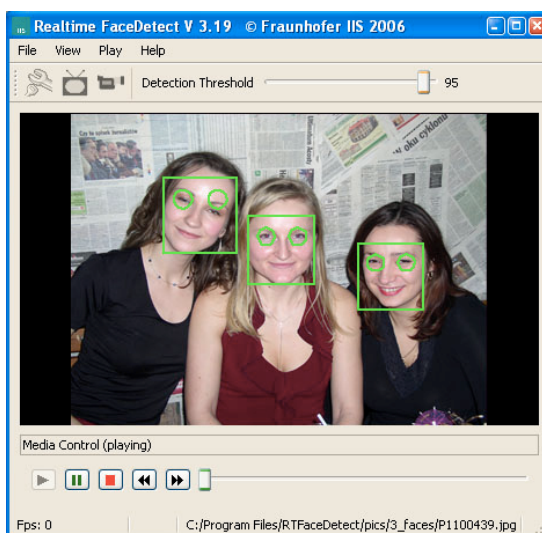
### Przegląd najlepszych systemów detekcji na rynku

Podczas prac nad rozwojem opisywanego w rozprawie systemu na świecie również trwały badania nad ulepszaniem algorytmów. Zmiany te są znaczące i zbliżają się do rozwiązania problemu lokalizacji i detekcji na obrazach cyfrowych. Do zaprezentowania wyników konkurencji najlepiej posłużyć się najlepszymi systemami dostępnymi na rynku. Uważna analiza rozwoju tej dziedziny biometryki pozwala wysnuć wniosek, że na rynku dominują dwa typy podejścia do badań. Pierwsze, typowo naukowe, sponsorowane i prowadzone pod opieką uczelni wyższych ma na celu dokładne zbadanie problemu, przez co następuje rozwój. W proces włączeni są studenci pomagający rozwijać algorytmy pod okiem doświadczonych profesorów i doktorów. W taki sposób powstaje np. RealTime FaceDetect w Fraunhofer Institut Integrierte Schaltungen [RTFD], który jest jednym z najbardziej zaawansowanych

naukowych systemów. Drugie, czysto komercyjne, nastawione na sprzedaż i rynkowy sukces, realizowane jest przez duże zespoły programistów i naukowców, wspomaganych przez ogromne budżety, mające dostęp do najnowszych zdobyczy nauki, ściśle współpracujące z wielkimi korporacjami. Wynikiem ich jest nie tyle sam algorytm, co sposób jego wykorzystania. Taki model reprezentuje bardzo popularny ostatnio system Face Tracker firmy FotoNation [FT]. Poniżej przedstawione są oba wyżej wymienione systemy.

RealTime FaceDetect opracowany przez Fraunhofer Institut Integrierte Schaltungen.

Ten system opracowany na niemieckiej uczelni jest jednym z najskuteczniejszych jakie dostępne są do przetestowania na komputerach klasy PC [Kub06, Fro04]. Prace rozwojowe ciągle trwają, a na stronie instytutu [RTFD] pojawiają się nowe aktualizacje w formie pliku do ściągnięcia. Interfejs programu, przedstawiony na rys. 1.12, w którym zaszyte są algorytmy detekcji twarzy, jest intuicyjny.

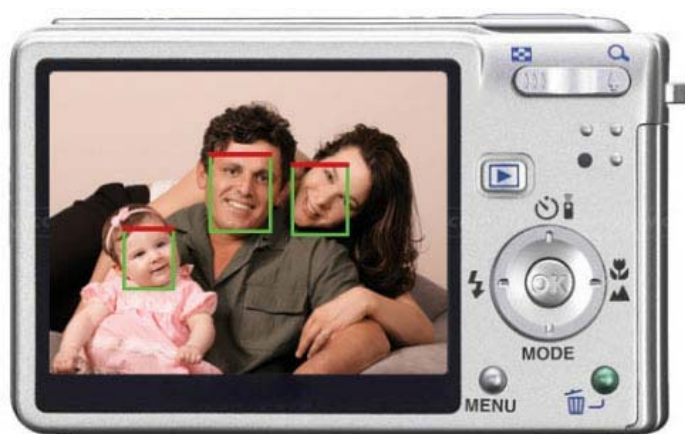


Rys. 1.12. Interfejs użytkownika programu RealTime FaceDetect wraz z przykładami jego błędnego, jak i poprawnego działania

Face Tracker opracowany przez firmę FotoNation

W odróżnieniu od poprzedniego systemu, Face Tracker [FT] jest budowany i opracowywany w konkretnym komercyjnym celu. Firma FotoNation słusznie założyła i konsekwentnie realizowała plan użycia detektora twarzy do wspomagania funkcji cyfrowych aparatów fotograficznych i aparatów telefonii komórkowych. Ich system jest tak dobrze osadzony marketingowo, że niemal każdy nowy aparat cyfrowy musi posiadać wspomaganie Auto Focusa przez moduł firmy FaceNation. Oficjalnie do użycia przyznają się Pentax, Samsung i

Nokia, choć inni producenci z wielkim prawdopodobieństwem również używają tego systemu. Nowatorskość rozwiązania polega na wykorzystaniu detekcji twarzy do wspomagania AF. Przesłanką skłaniającą przyjęcia takiego założenia jest fakt, że wg badań 90% amatorskich zdjęć zawiera twarz ludzką. Użytkownik jest w stanie zaakceptować na zdjęciu rozmyte tło, ale twarze już nie. System jest w stanie na bieżąco śledzić do dziewięciu twarzy. Informacją zwrotną dla użytkownika aparatu są ramki wyświetlone wokół wykrytych twarzy co ukazuje rys. 1.13. System wykrywa twarze zarówno w pozycjach en face, jak i z profilu.



Rys. 1.13. Wizualizacja działania Face Tracker w aparacie cyfrowym [FT]

Następnie, tuż przed wykonaniem zdjęcia FaceTracker pozwala oprogramowaniu aparatu na dobór parametrów zapewniający prawidłowe oddanie fotografowanych twarzy. W tym celu odpowiednio dobierana jest ostrość i ekspozycja w czym może, ale nie musi pomagać lampa błyskowa. Oprogramowanie FotoTracker wykrywa również obrócenie obrazu o 90° - niezależnie od tego czy aparat jest wyposażony w sensor ruchu. Zastosowanie systemu FaceTracker w oprogramowaniu aparatu cyfrowego wprowadza do menu tryb *FacePriority* - priorytet twarzy FP. (Jest to nazwa FotoNation i poszczególni producenci mogą nazywać ten tryb własnymi określeniami.) Składa się on z trzech podstawowych składowych: *FP Auto Focus* - umożliwia automatyczne ustawianie ostrości na obiekty poza centrum kadru; *FP Auto Exposure* - umożliwia tryb priorytetu ekspozycji dla pola, w którym znajduje się twarz; *FP Color Balance* - umożliwia automatyczny dobór balansu bieli zoptymalizowanego do oddania tonów skóry. Na targach PMA 2006 i 2007 (Photo Marketing Association Conference) zaprezentowano działanie systemu, które zostało bardzo pozytywnie przyjęte przez rynek.



Rys. 1.14. Demonstracja pracy Face Tracker podczas PMA 2007 [FT2]

Niestety nie ma możliwości zapoznania się z algorytmami zastosowanymi w Face Tracker, ponieważ jest to tajemnica firmowa. Niemniej z informacji, które są na rynku część algorytmów pochodzi z FaceIT, znanego systemu rozpoznawania twarzy.

### 1.3. Bazy danych zdjęć

Pomimo wielkiego wysiłku naukowców pracujących na całym świecie nad problemem detekcji twarzy na obrazach, niewiele zrobiono w celu stworzenia odpowiednich baz danych do testowania i porównywania konkurencyjnych rozwiązań. Nie należy mylić baz zdjęć twarzy służących do testowania systemów rozpoznających twarze z tymi do badania skuteczności detekcji. Problem detekcji wymaga odmiennych danych do przetwarzania, ponieważ ukazanie i ewentualne wyeliminowanie niedoskonałości algorytmów związanych z tą grupą nie pokrywa się z tymi niezbędnymi do rozpoznawania twarzy. Należy tu zwrócić uwagę, że niewiele baz danych posiada skomplikowane tło, a wielkość twarzy, jej położenie jest przeważnie podobne w całym zestawie. Często autorzy wykorzystują więc własne autorskie zestawy jednak ciężko wtedy o porównanie z innymi, istniejącymi algorytmami. Spośród najczęściej spotykanych baz danych można wymienić: ORL (Olivetti) [ORL], BioID [BioID], FERET [FER], FERET COLOR [FER], PIE [CMU].

Bazy danych najczęściej zawierają zbiór plików graficznych zgromadzonych w jednym katalogu lub też podzielone są na podkatalogi. Wtedy każdy z nich reprezentuje oddzielną klasę twarzy. Obiektem takich baz danych jest plik graficzny zawierający obraz

wraz ze znajdującą się na nim twarzą. Często dołączony jest również plik tekstowy z informacją o położeniu twarzy (np. w bazie BioID określone są położenie 20 punktów antropometrycznych na twarzy) umożliwiające szybką weryfikację pracy detektora.

Pierwsza z opisywanych baz danych to wynik pracy naukowców Olivetti Research Laboratory w Cambridge w Wielkiej Brytanii o nazwie Olivetti DB [ORL]. Powstała w latach 1992 – 1994, zawiera 400 zdjęć. Jest to zbiór obrazów 40 osób (po 10 zdjęć na każdego) wykonanych w rozdzielczości  $92 \times 112$  pikseli i w 256 odcieniach szarości. Baza zorganizowana jest w następujący sposób: każdej osobie odpowiada osobny podkatalog, którego nazwa ma format  $Sn$ , gdzie  $n$  jest numerem osoby (1-40). Kolejnymi cechami obrazów jest ciemne tło, które jest za twarzą, jak również to, że każda postać patrzy na wprost z małymi odchyleniami we wszystkich kierunkach. Należy tutaj nadmienić, że baza ta jest dowodem na to, że nie wszystkie bazy twarzy są dobre do testowania detektorów twarzy. W tym konkretnym przypadku problemem jest jednolite, ciemne tło, i zbliżona procentowo do wielkości obrazka wielkość wszystkich twarzy. W tabeli 1.1 pokazanych jest 5 przykładowych zdjęć jednej z klas.

Do zadań detekcji znacznie lepiej nadaje się baza FERET [Phi98, Phi00]. Obecnie jest to jedna z najbardziej miarodajnych baz wykorzystywanych do testowania komercyjnych systemów rozpoznawania osób (Face Recognition Vendor Test – FRVT) jak i tych czysto naukowych. Jej niezaprzeczalną zaletą jest ogromna ilość zdjęć, która jest stale rozbudowywana: w 1999 roku liczyła 14126 zdjęć 1199 osób [Phi99], a w roku 2003 test korzystał już z 121589 zdjęć, 37437 osób [Phi03]. Taką ilość danych nie może w tym momencie zaoferować żadna inna baza twarzy. Od 2003 roku zespół opracowujący bazę FERET zdecydował o przejściu na zdjęcia kolorowe i teraz rozwijana jest jedynie ta odmiana. Zdjęcia zawierają twarz w ujęciu portretowym, bądź szerokim portrecie. Są to zarówno kobiety jak i mężczyźni (każdej rasy), a kąt obrotu twarzy waha się od  $-90$  do  $90$  stopni. Postaci wyrażają różne stany emocjonalne, są nierównomiernie oświetlane, a często zawierają elementy zakłócające typu okulary czy zarost (patrz tab. 1.1). Tło za twarzami jest mało skomplikowane i przeważnie rozmyte krótką ogniskową aparatu. Parametry techniczne obrazów są następujące: rozdzielczość  $256 \times 384$  piksele w przestrzeni RGB (kiedyś w odcieniach szarości). Każdy plik zawiera unikatową nazwę, która informuje również o kilku istotnych parametrach związanych z danym obrazem.

Format nazwy pliku jest następujący: *nnnnnxxfffq\_yymmdd.tif*, gdzie [Kuz04]:

- *nnnnn* – pięciocyfrowy unikalny numer identyfikacyjny,
- *xx* – dwuznakowy kodem oznaczający sposób zobrazowania twarzy na zdjęciu,

- *fa* – zdjęcia frontalne twarzy,
- *pl* – profil lewy,
- *hr* – prawa połówka twarzy;
- *fff* – trzy znaczniki (flagi) używane do określenia:
  - czy zdjęcie może być publikowane,
  - czy histogram zdjęcia był dopasowywany,
  - czy zdjęcie było wykonane na filmie o czułości 200 czy 400 ASA;
- *q* – modyfikator, który nie zawsze występuje, informujący o tym, czy:
  - *a* – osoba nosiła okulary,
  - *e* – ubranie było cyfrowo retuszowane
  - *f* – jasność obrazu była zredukowana;
- *yymmdd* – data wykonania zdjęcia.

Baza FERET jest produktem komercyjnym. Do testowania produktów przeznaczonych do sprzedaży (detektorów lub/i systemów rozpoznawania) konieczne jest wniesienie opłat, jednak do celów naukowych udostępniono wersję uboższą, umożliwiającą sprawne przeprowadzenie badań nad efektywnością, a jedyną koniecznością jest zamieszczenie odpowiedniej adnotacji w bibliografii.

Kolejną bazą przygotowaną z obrazów w odcieniach szarości jest baza BioID opracowana przez firmę HumanScan [HS]. Posiada niewątpliwą zaletę w postaci plików dodanych do każdego obrazu, zawierających informacje o położeniu 20 punktów antropometrycznych występujących na danej twarzy (m.in. oczu, ust, nosa, brody). W bazie BioID znajduje się 1521 zdjęć 23 osób o rozdzielczości 384×286 pikseli w 256 odcieniach szarości. Są to obrazy wykonane w typowym biurze, zwykłą kamerą internetową o niewyszukanej jakości, co w założeniu miało oddać jak najbardziej docelowe warunki pracy danego detektora czy systemu rozpoznawania osób. Twarz przeważnie zajmuje centralną część obrazu, a jej wielkość nie waha się w szczególnie dużym zakresie i wynosi średnio ok. 50-70% wysokości obrazu (czyli 140-196 pikseli) (patrz tab. 1.1). Format nazewnictwa plików BioID to: *BioID\_xxxx.pgm*, gdzie *xxxx* określa unikalny numer zdjęcia. Cała baza jest dostępna pod adresem [BioID].

Niewątpliwą zaletą ostatniej z prezentowanych baz danych twarzy jest posiadanie w swych zasobach obrazów w 24-bitowej głębi kolorów. Zbiór ten o nazwie CMU Pose Illumination Expression zawiera 41368 obrazów twarzy 68 osób [Sim02]. Kolejną zaletą bazy CMU jest to, że zdjęcia były wykonywane w specjalnym pomieszczeniu o nazwie CMU 3D Room (13 zsynchronizowanych aparatów fotograficznych wysokiej jakości i 21 lamp

błyskowych). Użyte kamery i lampy zostały umieszczone z przodu fotografowanego obiektu na hemisferze [Kan98]. Parametry techniczne obrazów przedstawiają się następująco: rozdzielczość 640x480 pikseli oraz 24-bitową głębię kolorów. Cała baza jest dostępna pod adresem [CMU].

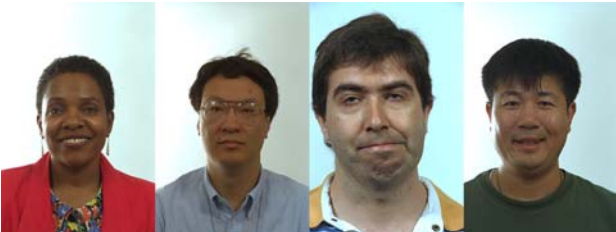
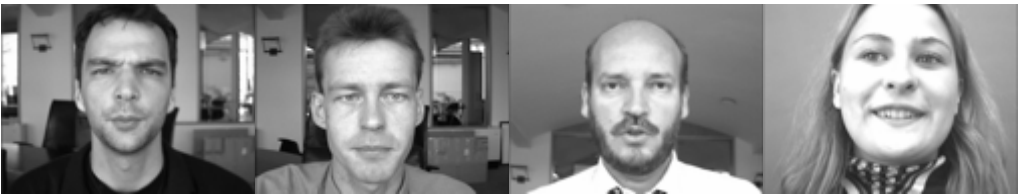
|             |                                                                                      |
|-------------|--------------------------------------------------------------------------------------|
| Olivetti    |    |
| FERET       |    |
| FERET COLOR |   |
| BioID       |  |
| CMU PIE     |  |

Tabela 1.1. Przykładowe zdjęcia znajdujące się w standardowych bazach obrazów twarzy (Źródło: [Kuz04])

Podsumowując, niezaprzeczalną zaletą gotowych baz twarzy jest ich ogólnodostępność i przez to możliwość zapoznania się z wynikami osiągnięć innych zespołów naukowych i komercyjnych. Nie występuje wtedy niebezpieczeństwo, że do testów użyta zostanie autorska baza danych, a prezentowane rewelacyjne wyniki będą rezultatem

odpowiednio spreparowanych danych wejściowych. Wadami gotowych baz jest ich „liniowość” (ta sama rozdzielczość, twarze centralnie położone, ich wielkość nie różniąca się w dużym zakresie) i przygotowanie zdecydowanie do zadań rozpoznawania, a nie detekcji. Niestety, żadna z powyższych pięciu baz nie jest miarodajnym narzędziem do testowania detektorów, a jedynie wstępnym krokiem.

## 1.4. Podsumowanie i wnioski

### Podsumowanie

Niewątpliwie problem detekcji twarzy na obrazach jest zagadnieniem poważnym i nierozwiązanym. Bardzo duże zainteresowanie jego rozwiązaniem przez naukowców jak i firmy komercyjne jest dostatecznym dowodem, że istnieje potrzeba udoskonalania wyników działania detektorów w dziedzinie jakości i szybkości. Szczególnie ostatnia wymieniona cecha jest istotna (przy zachowaniu jakości na wysokim poziomie), ponieważ detektory rzadko występują jako niezależne systemy, a są jedynie pierwszym etapem innych takich jak *visitor identification* czy *access control*. Należy przy tym pamiętać, że detektory odgrywają zasadniczą rolę w wyżej wymienionych systemach, gdyż bez nich nie uda się stworzyć skutecznych i w pełni automatycznych systemów bazujących na biometrii twarzy. Niezaprzeczalnie ważną częścią dzisiejszych detektorów jest część odpowiedzialna za lokalizowanie obszarów potencjalnie zawierających twarz, następnie weryfikowanych przez kolejny blok decyzyjny twarz/„nie twarz”. Dobrze opracowany, istotnie wpływa na wynik szczególnie w dziedzinie czasu, drastycznie ograniczając ilość analizowanych dalej próbek. W skrajnym przypadku taki generator może posłużyć jako samodzielny lokalizator twarzy.

Na podstawie przeprowadzonej analizy prac naukowych należy stwierdzić, że różnorodność podejść do zagadnienia detekcji twarzy na obrazach jest bardzo duża. W większości przypadków razem z nowym systemem detekcji opracowywany jest dedykowany lokalizator obszarów potencjalnie zawierających twarz, zwanym w uproszczeniu generatorem kandydatów. Założenie to jest słuszne jedynie w sytuacji, gdy rzeczywiście wynik jego działania znacząco odbiega od konkurencji, oczywiście w kierunku doskonałości. W trakcie analizy istniejących rozwiązań takiej tendencji nie zauważono. Nasuwa się wręcz stwierdzenie, że naukowcy rzadko, porównują tę część swoich prac, a przez to rezultaty osiągnięte ostatecznie przez cały system potrafią być słabe, mimo bardzo dobrego algorytmu twarz/„nie twarz”.

Dodatkowym problemem, który utrudnia porównywanie generatorów kandydatów, jak i samych detektorów jest brak bardzo dobrych, efektywnych, dedykowanych i

ogólnodostępnych baz danych pozwalających na rzetelne ich porównanie. Należy spodziewać się, że bazy takie w niedalekiej przyszłości zostaną opracowane i zyskają przychyłność naukowców.

Mimo, że na przestrzeni ostatnich dwóch dekad postęp w dziedzinie detekcji twarzy na obrazach jest ogromny, nie ma wątpliwości, że zostało jeszcze dużo do zrobienia, aby ostateczny system był kompletny, czyli odporny na [Kuk03a, Yan02]:

- warunki oświetleniowe,
- orientację, pozę i częściowe przesłonięcia,
- ekspresję twarzy,
- obecność okularów, zarostu i zróżnicowania w fryzurze.

Równie istotnym czynnikiem wpływającym na ocenę detektora, a tym bardziej generatora kandydatów jest szybkość jego działania, co w prostej linii przekłada się na jego przydatność do wykorzystania w systemach automatycznego rozpoznawania twarzy lub innych systemach bezpieczeństwa.

Istnieją cztery grupy podejść do detekcji [Yan02]: bazujące na wiedzy, używające niezmienników, porównywania ze wzorcem i bazujących na wyglądzie. Warto zaznaczyć, że w ostatnich latach najczęściej wykorzystywane są metody z ostatniej z wymienionych grup jednak nie występujące samodzielnie, lecz wspierane przez metody z trzech pierwszych. To w nich przeważnie stosuje się podział na dwa etapy: lokalizacji obszarów potencjalnie zawierających twarz i decyzje twarz/„nie twarz”. Celem takich działań jest redukcja czasu potrzebnego na wykonanie detekcji, jak również na nie analizowanie obszarów, gdzie wystąpienie twarzy jest minimalnie prawdopodobne.

## Wnioski

Podsumowując, można stwierdzić, że problem detekcji twarzy na obrazach nie został w pełni rozwiązany mimo niewątpliwego postępu poczynionego na przestrzeni ostatnich dwudziestu lat. Szczególnie domena czasu jest ciągle bardzo zaniedbana. Systemy pracujące z prędkością większą niż 2-5 obrazów na sekundę (*ang. fps – frame per second*) należą do rzadkości, a ich wyniki w dziedzinie efektywności nie są zadowalające. W znacznej części przypadków sytuacja ta spowodowana jest analizą zbyt dużej ilości danych, gdzie nie ma szans na znalezienie twarzy. Rozwiązaniem jest opracowanie skutecznego i szybkiego lokalizatora obszarów potencjalnie zawierających twarz, dostarczającego minimalną ilość próbek do analizy, przy założeniu wysokiego prawdopodobieństwa, że znajdzie się wśród

nich ta zawierająca poszukiwaną twarz. Jak widać, istnieje zapotrzebowanie na szybki i skuteczny lokalizator obszarów zawierających twarz spełniający następujące założenia:

1. efektywność – ponad 95% współczynnik zawierania w swoim zbiorze próbek opisujących twarze przy obrazach kolorowych, cechujących się trudnymi warunkami takimi jak:
  - kierunkowe oświetleniowe,
  - zmienną orientację i kierunek obrotu
  - pozę
  - częściowe przesłonięcia,
  - ekspresję twarzy,
  - obecność dodatkowych elementów na twarzy: okularów, zarostu
  - zróżnicowania w fryzurze.
2. szybkość działania – czas potrzebny do wykonania operacji na jednym obrazie mniejszy od 1/20 sekundy na zwykłym komputerze biurkowym

### 1.5. Cel i teza pracy

Opierając się na powyższym **celem** niniejszej pracy jest „**Opracowanie metody i realizacja algorytmów do zadań lokalizacji obszarów zawierających twarz na obrazach cyfrowych**”.

Osiągnięcie postawionego celu opierało się na:

- przeglądzie literatury naukowej poświęconej problemowi detekcji i lokalizacji twarzy;
- sformułowaniu idei wykorzystania owalnego kształtu twarzy i procedury zmiennego promienia do zadań lokalizacji oraz opracowaniu bazowej metody realizującej tę ideę;
- stworzeniu lokalizatora twarzy (w języku C++) realizującego opracowaną metodę, umożliwiające działanie w dwóch trybach: z wiersza poleceń (do integracji z systemami rozpoznawania) i graficznej prezentacji (do nauczania i badania problemów lokalizacji i detekcji twarzy);
- eksperymentach i doświadczalnemu udoskonaleniu bazowej metody w zadaniach lokalizacji twarzy.

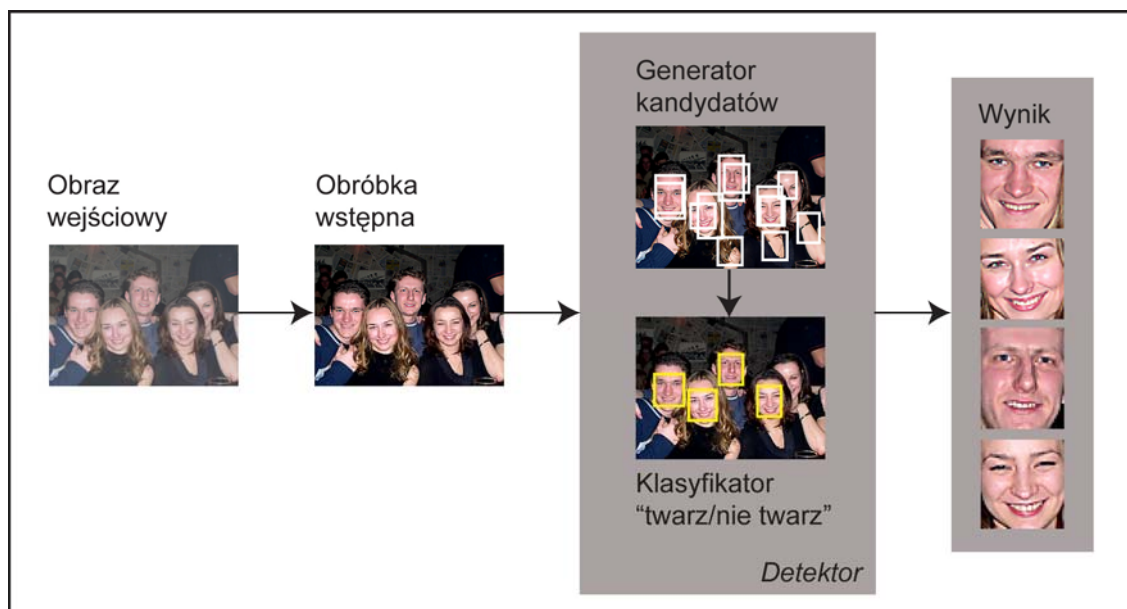
Ponadto, w pracy zostanie udowodniona teza, że: „**Jednoczesne uwzględnienie informacji o owalnym kształcie twarzy i procedury zmiennego promienia, może być podstawą do budowy skutecznej metody lokalizacji twarzy na obrazach cyfrowych, a dodatkowo użycie informacji o kolorze skóry zwiększy jej skuteczność**”.

## 2. Opracowanie metody do rozwiązania zadania lokalizacji obszarów zawierających twarz

W niniejszym rozdziale zostanie przedstawiona metoda wykorzystująca ideę użycia kształtu twarzy, jej koloru, procedury zmiennego promienia i macierzy gradientów do rozwiązania zadania lokalizacji obszarów zawierających twarz. Aby opis metody był bardziej zrozumiały na początku rozdziału opisane zostaną podstawowe algorytmy normalizacyjne, bez których nie może obejść się żaden nowoczesny system operujący obrazem jako źródłem danych.

### 2.1. Wprowadzenie do projektowania detektora twarzy

Każdy detektor, a tak naprawdę każdy system musi być wstępnie zaprojektowany. Bardzo istotny jest dobór odpowiednich metod na każdym etapie projektowania, a one są bardzo zależne od przyjętych wstępnych założeń. Na przestrzeni ostatnich dwóch dekad wykształciło się parę ogólnych wzorców budowania detektorów twarzy, jednak aktualnie nowoczesne systemy w przytłaczającej większości bazują na jednym z nich. Idee jego działania ukazuje poniższy schemat.



Rys. 2.1. Schemat blokowy detektora z blokiem rozpoznawania „twarz/nie twarz” (Źródło: opracowanie własne)

Obraz wejściowy może być pojedynczą klatką (pozyskaną np. z aparatu cyfrowego) bądź sekwencją wideo. Może zawierać pełną informację o kolorze, bądź być przedstawiony w odcieniach szarości. W bardzo specjalizowanych rozwiązaniach stosuje się też obraz

pozyskany z kamer działających w podczerwieni (tam detekcja jest dużo łatwiejsza z powodu specyficznego obrazu ciepła wydzielanego przez ludzką twarz). Obraz wejściowy jest przeważnie w wielkości nie mniejszej niż  $240 \times 320$  pikseli i nie większej niż  $480 \times 640$  pikseli (orientacja występuje zarówno w wersji poziomej jak i pionowej), czyli proporcje są stałe i wynoszą 4:3.

Kolejny krok – normalizacja – jest istotnym blokiem funkcyjnym detektorów. Projektujący system detekcji twarzy musi posiadać odpowiednią wiedzę pozwalającą mu dobrać odpowiednie metody normalizacyjne, ponieważ efektem ich działania jest maksymalne ujednolicenie danych wejściowych będących często bardzo różnej jakości. Pozwala to na budowanie bardziej spójnej klasy twarzy, co jest nie bez znaczenia dla efektywności działania całego systemu i to niezależnie czy klasa ta jest opisywana zależnościami między elementami twarzy (w metodach bazujących na wiedzy czy używających niezmienników), porównywana ze wzorcem, czy wspomagająca się algorytmami uczącymi (w podejściach bazujących na wyglądzie).

Najczęstszymi metodami normalizacyjnymi stosowanymi do obrazów są [Kuk03]:

- skalowanie,
- obrót,
- filtracja,
- operacje na histogramie: wyrównanie, obcięcie, itd.
- redukcja przestrzeni cech,
- konwersja do innej przestrzeni barw.

Niewątpliwie sercem każdego detektora jest blok zawierający algorytmy odpowiadające za ostateczną detekcję. W nawiązaniu do przeglądu literatury przeprowadzonej w poprzednim rozdziale należy stwierdzić, że istnieje obecnie trend do budowania systemów dwuetapowych. Pierwszy z nich – lokalizator obszarów potencjalnie zawierających twarz, a nazywany często generatorem kandydatów – jest krokiem wstępnym, ale niezwykle istotnym, gdyż pozwala wyeliminować znaczne obszary obrazu, gdzie główny, przeważnie znacznie bardziej skomplikowany obliczeniowo algorytm, nie musi być wykonywany z racji małego prawdopodobieństwa zaistnienia tam twarzy. Drugi z etapów realizowany jest modułem nazywanym blokiem decyzyjnym twarz/„nie twarz”. Warto zauważyć, że jego parametry są tak dobrane, że uzyskiwane wyniki są trafne jedynie w przypadku dostarczenia mu odpowiednich danych wejściowych (generatora kandydatów). Natomiast zastosowanie go z pominięciem pierwszego etapu daje w rezultacie słaby wynik, spowodowany zbyt dużym współczynnikiem błędnych przyjęć. Należy również nadmienić, że

często oba z wyżej wymienionych podbloków systemu są realizowane algorytmami pochodzącymi z całkowicie różnych grup podejść do detekcji opisanych w rozdziale 1.2.

Wynikiem działania detektora jest zestaw współrzędnych określających położenie wszystkich twarzy. Najczęściej są to współrzędne prostokąta, w obszarze którego znajduje się poszukiwana twarz, jednak zdarza się też określenie środka twarzy i promienia. Przy pomocy tych danych systemy przeważnie dokonują wycięcia odpowiednich obszarów obrazu wejściowego i przekazują je do kolejnych bloków większego systemu, np. modułu rozpoznawania twarzy.

Biorąc pod uwagę powyższe rozważania jak i temat niniejszej rozprawy dotyczący opracowania skutecznego lokalizatora obszarów zawierających twarz należy dokładnie przeanalizować dostępne metody zaproponowane przez inne zespoły badawcze i wybrać takie, które wstępnie oferują szanse spełnienia warunków wejściowych. Przy przyjętych założeniach, gdzie ważny jest czas wykonywania algorytmu nie mogący być większy niż 1/10 sekundy na typowym komputerze biurowym (stan na rok 2006) pod uwagę mogą być brane metody oferujące:

- wysoką efektywność czyli wysoki współczynnik poprawnej detekcji przy minimalnych współczynnikach błędnych przyjęć i błędnych odrzuceń,
- odporność na zmienne warunki oświetleniowe, przesłonięcia, obroty i dodatkowe elementy na twarzy (okulary, zarost, blizny),
- możliwość detekcji więcej niż jednej twarzy,
- niskie zapotrzebowanie na moc obliczeniową maszyny pozwalające na pracę określaną czasem rzeczywistym (*ang. real-time*).

## 2.2. Omówienie użytych algorytmów obróbki obrazów i metod normalizacyjnych

Systemy operujące na danych płynących z obrazu dwuwymiarowego, czyli również i detektory używają algorytmy normalizacyjne będące niezbędnym elementem wpływającym na efektywność końcową całego procesu. Należy zaznaczyć, że istnieje bardzo wiele różnych algorytmów służących jednemu celowi. Dobrym przykładem jest skalowanie, które można zaimplementować przy użyciu kilku metod: najbliższy sąsiad (*ang. nearest naiborhood*), dwuliniowe (*ang. bilinear*), bicubic, Lanczos, itd. Jakość użytych algorytmów, również sposób implementacji, bezpośrednio wpływa na szybkość całego systemu szczególnie, gdy są często, cyklicznie wykorzystywane (np. operacja skalowania wycinków obrazu do wielkości 20×20 pikseli jest przeprowadzana setki tysięcy razy dla każdego badanego obrazu w większości algorytmów bazujących na wyglądzie). Poniżej przedstawione zostaną algorytmy

obróbki obrazów i metod normalizacyjnych użytych w budowie metody lokalizacji obszarów zawierających twarz, jednak ich opis będzie ograniczony do minimum, ponieważ są one powszechnie znane.

### Konwersja przestrzeni barw

W detektorach twarzy jak i innych systemach pracujących przy użyciu obrazów często zachodzi potrzeba użycia innej przestrzeni barw niż ta, którą posiada obraz wejściowy. Najczęściej używaną konwersją kolorów jest ta zamieniająca przestrzeń *RGB* (ang. *RedGreenBlue*) na odcienie szarości (ang. *grayscale*). Wynikowa przestrzeń jest odwzorowaniem jasności pikseli nazywaną luminancją. Operację tą realizuje następujący wzór:

$$Y = 0.30 \cdot R + 0.59 \cdot G + 0.11 \cdot B, \quad (2.1)$$

gdzie:

- $Y$  – wartość sygnału luminancji;
- $R$  – składowa czerwona koloru;
- $G$  – składowa zielona koloru;
- $B$  – składowa niebieska koloru.

Na odmiennosc współczynników przy każdym z kolorów wejściowych wpływa różna wrażliwość ludzkiego oka na każdą z tych składowych.

Przy detekcji twarzy na obrazach kolorowych bardzo często używa się informacji o kolorze skóry do wspomagania skuteczności algorytmów. Najczęściej używane przestrzenie barw to [Kuk03, Yan02]:

- *RGB*,
- znormalizowane *RGB*,
- *HSV* (albo *HSI*),
- *YCrCb*,
- *CIE XYZ*,
- *CIE LUV*.

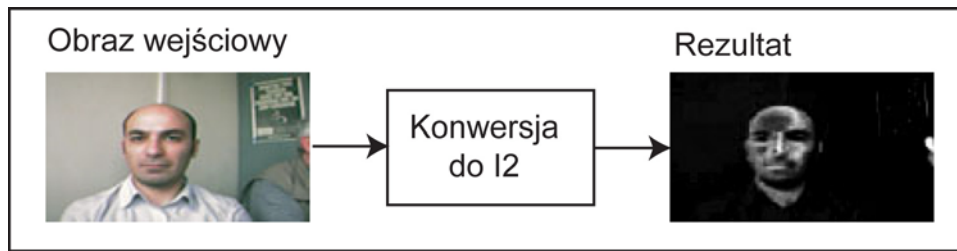
W niniejszej rozprawie zdecydowano się na użycie nietypowego modelu barw, ponieważ badania nad nim dowiodły jego skuteczności przy wybranej metodzie generowania kandydatów. Model ten, określany *III2I3* jest szeroko opisywany w [Kuk03]. Bardzo interesująco z punktu widzenia widoczności koloru skóry wygląda komponent *I2*. Konwersja obrazu w przestrzeni kolorów *RGB* do *I2* jest bardzo prosta co jest dodatkowym atutem w optymalizacji kodu algorytmu, a przedstawia się następująco:

$$I = R - G, \quad (2.2)$$

gdzie:

- $I$  – wartość składowej I2 w przestrzeni I1I2I3;
- $R$  – składowa czerwona koloru;
- $G$  – składowa zielona koloru.

Rysunek 2.2 przedstawia przykładowy rezultat działania konwersji kolorów z przestrzeni RGB do komponentu I2 z przestrzeni I1I2I3.



Rys. 2.2. Zobrazowanie wyniku konwersji do komponentu I2 z przestrzeni I1I2I3

### Skalowanie – wersja dwuliniowa

Skalowanie jest operacją powodującą zmianę rozdzielczości obrazu. Może być przeprowadzona zarówno w celu zwiększenia jak i zmniejszenia obrazu, jak również do zmiany proporcji obrazu, czyli skalowania nieproporcjonalnego. Do powszechnie znanych rozwiązań tego problemu należą trzy metody: „najbliższego sąsiada”, dwuliniowe i bicubic. Pierwsza z nich, choć najszybsza i najprostsza, nie jest szeroko wykorzystywana ze względu na niską jakość generowanego obrazu. Dwie pozostałe mają zaszyte w sobie algorytmy filtrujące, powodujące zdecydowanie lepsze odwzorowanie pikseli w wynikowym obrazie. Na potrzeby niniejszej rozprawy zdecydowano się użyć skalowania w wersji dwuliniowej.

Algorytm ten operuje na liczbach zmiennoprzecinkowych. Dla każdego punktu obrazu wynikowego przeprowadzana zostaje seria operacji. Dokonywane jest rzutowanie punktu wynikowego do obrazu wyjściowego, a następnie posługując się informacją zebraną z czterech najbliższych punktów, obliczana jest wartość wynikowego punktu. Realizacja wygląda następująco:

$$P_n(x_n, y_n) = f_{y2} * (f_{x2} * P_p(x_1, y_1) + f_{x1} * P_p(x_2, y_1)) + f_{y1} * (f_{x2} * P_p(x_1, y_2) + f_{x1} * P_p(x_2, y_2)), \quad (2.3)$$

$$\left\{ \begin{array}{l} y_1 = \text{floor}(y_n * s_y) \\ y_2 = \text{ceil}(y_n * s_y) \\ x_1 = \text{floor}(x_n * s_x) \\ x_2 = \text{ceil}(x_n * s_x) \\ f_{y1} = y_n * s_y - y_1 \\ f_{y2} = 1 - f_{y1} \\ f_{x1} = x_n * s_x - x_1 \\ f_{x2} = 1 - f_{x1} \\ s_y = \frac{Y_p}{Y_n} \\ s_x = \frac{X_p}{X_n} \end{array} \right. , \quad (2.4)$$

gdzie:

- $Y_p$  – wysokość obrazu wejściowego;
- $Y_n$  – wysokość obrazu wyjściowego;
- $X_p$  – szerokość obrazu wejściowego;
- $X_n$  – szerokość obrazu wyjściowego;
- $P_n(y,x)$  – nowa wartość punktu o współrzędnej pionowej  $m$  i poziomej  $n$ ;
- $P_p(y,x)$  – wartość punktu z obrazu wejściowego.

Zaletą opisanego powyżej algorytmu jest umiarkowanie duża szybkość, a wynik jest bliski ideału. Metoda bicubic, będąca jeszcze bardziej precyzyjna, nie oferuje dobrego współczynnika jakość/szybkość, więc została pominięta. Rysunek 2.3 obrazuje efekt działania skalowania w wersji dwuliniowej.



Rys. 2.3. Skalowanie obrazu przy użyciu wersji dwuliniowej

### Rozciągnięcie histogramu

Jednym z głównych czynników wpływających na jakość działania detektorów jest jednolitość danych wejściowych. Analiza literatury jednoznacznie wskazuje, że wpływ oświetlenia znacząco tę jednolitość zaburza, a przez to jest jednym z głównych czynników

wpływającym na spadek efektywności. Stwierdzono również, że te niekorzystne warunki można zlikwidować operacjami normalizacyjnymi na histogramie. Jego rozciągnięcie lub wyrównanie powoduje, że charakterystyka danych poddawanych dalszym obliczeniom mających na celu wykrycie twarzy jest znacznie korzystniejsza, przez co szanse na sukces są znacznie wyższe.

Histogram to statystyczny rozkład ilości pikseli dla poszczególnych wartości jasności. Te wartości zawierają się w przedziale od 0 do  $(2^n - 1)$  ( $n$ -bitowy zakres kolorów) i dla każdego z nich określana jest ilość wystąpień na obrazie. Analiza histogramu jest źródłem wiedzy na temat jakości oświetlenia, optyki urządzenia przechwytyjącego i obrazu samego w sobie. Dobre zdjęcie charakteryzuje się rozpiętością tonalną w całym zakresie. Na podstawie histogramu łatwo stwierdzić czy zdjęcie jest niedoświetlone lub prześwietlone, ponieważ znacznie maleje ilość używanych poziomów jasności, gdy reszta jest całkowicie pominięta.

Operacja rozciągnięcia histogramu ma na celu rozszerzenie dostępnej skali poziomów jasności ( $Z_1 \div Z_2$ ) na całą dostępną skalę ( $Z_{min} \div Z_{max}$ ). Osiąga się dzięki temu polepszenie kontrastu obrazu. Nową wartość piksela oblicza się według wzoru [Kuc98]:

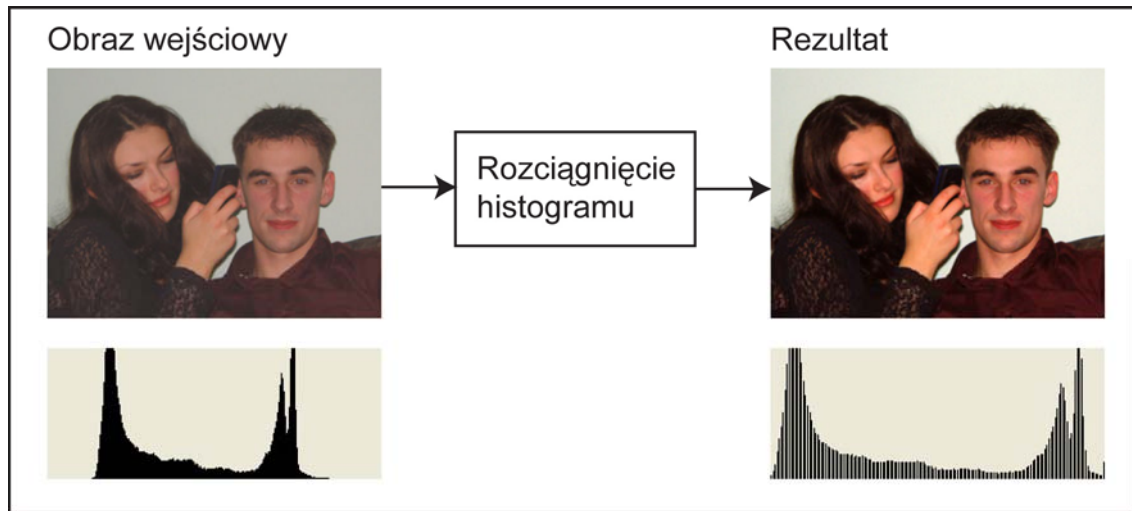
$$Z_n = (Z_{max} - Z_{min}) \cdot \frac{Z_p - Z_1}{Z_2 - Z_1}, \quad (2.5)$$

gdzie:

- $Z_p$  – oryginalny poziom jasności piksela;
- $Z_n$  – poziom jasności piksela po rozciągnięciu histogramu.

Graniczne poziomy jasności  $Z_1$  i  $Z_2$  mogą być wyznaczone ręcznie, ale bezobsługowe systemy detekcji pracujące przeważnie w tandemie z systemami rozpoznawania FaReS muszą dokonywać wyznaczania tych wartości w sposób automatyczny. Najczęściej wykorzystuje się do tego prosty algorytm poszukujący pierwszej wartości histogramu przekraczającą pewną zadaną wcześniej wartość progową. Gdy to nastąpi  $Z_1$  otrzymuje wartość aktualnej pozycji w histogramie. Podobna operacja przeprowadzana jest od drugiej strony histogramu.

Na rysunku 2.4 przedstawiony jest przykład rozciągnięcia histogramu. Łatwo zauważyć dużą, niezagospodarowaną przestrzeń histogramu w obrazie wejściowym, zarówno na początku skali jak i na jej końcu. Po dokonaniu opisywanej operacji obraz zdecydowanie zyskał na kontraście. Warto dodać, że operacje na histogramie charakteryzuje wyjątkowa prostota implementacji jak również szybkość działania, a efekt jest w zdecydowanej większości przypadków korzystny dla jakości danych wejściowych poddawanych dalszej obróbce.



Rys. 2.4. Przykład ilustrujący działanie algorytmu rozciągania histogramu wraz z zobrazowaniem wartości histogramu.

### Filtracja obrazu

Bardzo często obrazy wejściowe są zaszumione, nieostre, co utrudnia pracę detektorów jak i kompleksowych systemów FaReS. Czasem zachodzi też potrzeba rozmycia obrazu (czyli pozbycia się wysokich częstotliwości) lub wyszukania krawędzi obiektów. Do uzyskania odpowiedniego efektu w wyżej wymienionych dziedzinach często używa się operacji filtracji obrazu wejściowego.

Najczęściej wykorzystywaną metodą filtracyjną jest splot. Charakteryzuje go wielka prostota i łatwość implementacji (również sprzętowej). Do jego stosowania w systemach przetwarzania obrazów zachęca jednak przede wszystkim duża szybkość wykonywania, przy wielkiej mnogości efektów jaką można uzyskać, zmieniając jedynie wartości w macierzy filtru. Wartości poszczególnych komórek tej macierzy mogą być dodatnie jak i ujemne. Ich charakter determinuje uzyskany efekt. Do obliczenia wartości piksela obrazu wyjściowego użyta jest suma iloczynów odpowiadającego mu piksela z obrazu wejściowego i jego sąsiadów, z odpowiadającymi im elementami macierzy filtra. Ostateczna wartość jest dzielona przez współczynnik normalizujący (*norm*). Najczęściej jest to suma wszystkich elementów macierzy filtra splotu (wtedy moc całego obrazu nie zostaje zachwiana). Aby obliczyć nową wartość piksela, przy założeniu, że macierz filtra ma rozmiar  $I \times J$ , należy posłużyć się wzorem [Kuc98]:

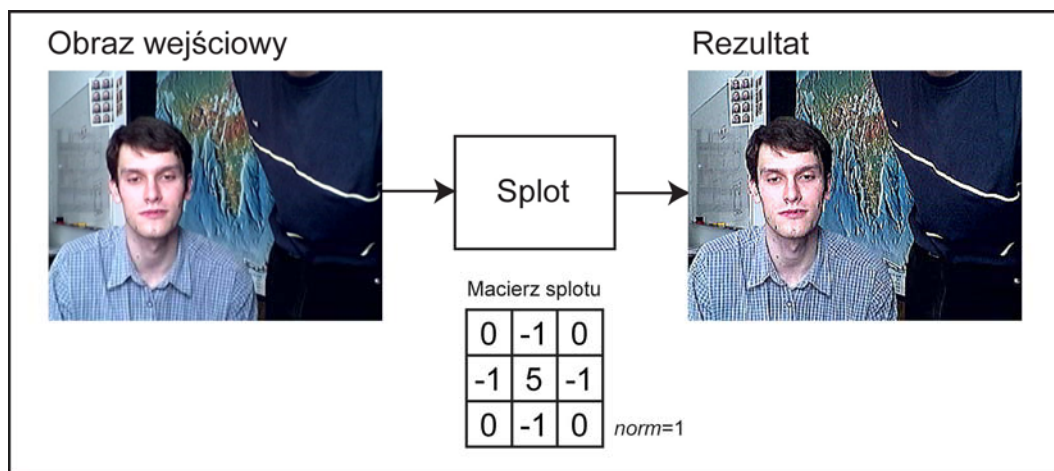
$$P_n(m, n) = \frac{1}{norm} \cdot \sum_{i=1}^I \sum_{j=1}^J f(i, j) \cdot P_p(m, n), \quad (2.6)$$

gdzie:

- $P_n(m, n)$  – nowa wartość punktu o współrzędnych  $m$  i  $n$ ;

- $P_p(m,n)$  – wartość punktu z obrazu wejściowego;
- $I, J$  – wysokość i szerokość maski;
- $f(i,j)$  – współczynnik splotu;
- $norm$  – współczynnik normalizujący.

Rysunek 2.5 obrazuje działanie splotu, w którym parametry są dobrane w taki sposób, by uzyskać efekt wyostrażania krawędzi. Technika ta jest używana często do detekcji obiektów przemysłowych np. statków, samolotów lub elementów elektroniki, jak również w zastosowaniach poligraficznych, gdzie ostrość obrazu jest istotnym elementem wpływającym na odbiór fotografii.

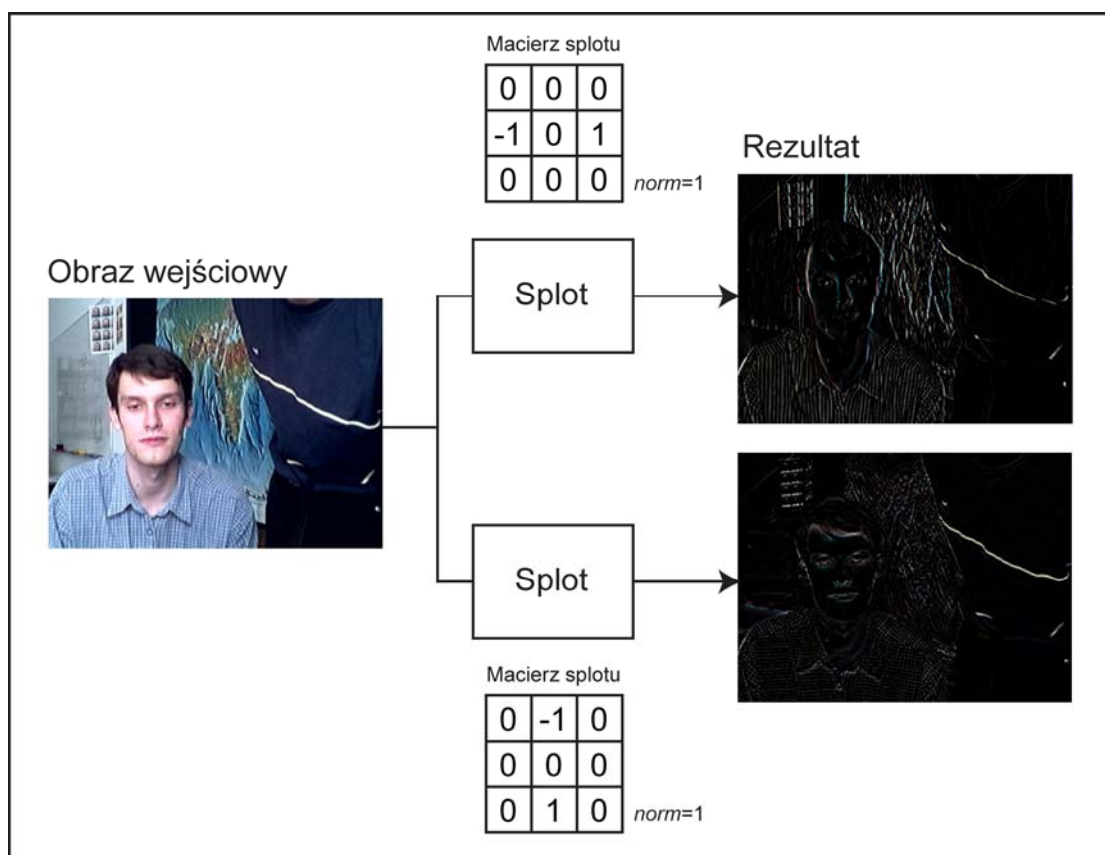


Rys. 2.5. Zilustrowanie działania przykładowego filtra splotu – macierz dająca efekt wyostrażający

W niniejszej rozprawie do budowy lokalizatora generatora kandydatów wykorzystywane są macierze gradientów w płaszczyźnie pionowej i poziomej. Są one jak kolejną wariacją splotu. Odpowiadające im macierze splotu jak i zobrazowanie wyniku przedstawia rysunek 2.6.

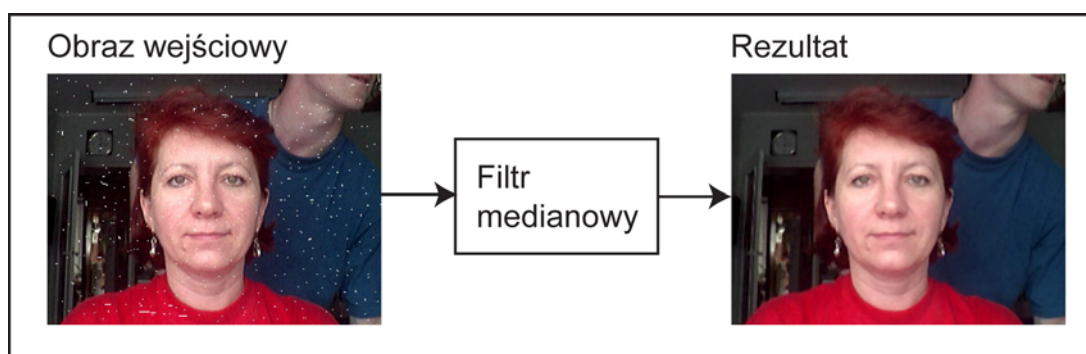
Filtr splotu jest dobrym narzędziem do uzyskiwania efektów wyostrażania, wykrywania krawędzi lub liniowego rozmywania obrazu, jednak słabo sprawuje się przy usuwaniu losowego szumu pojawiającego się często w przypadku słabych urządzeń dostarczających sygnały cyfrowe. Do tego celu znacznie lepiej nadaje się filtr medianowy. Zasada jego działania jest następująca: otoczenie każdego piksela (np. kwadrat 3x3 piksele) jest przepisywane do wektora, sortowane, a następnie wybierany jest środkowy element i wpisywany w miejsce wynikowego piksela. Zabieg ten doskonale usuwa szum z obrazów, a jego minimalna złożoność obliczeniowa (tu warunkiem jest użycie efektywnego algorytmu

sortowania) zachęca do korzystania szczególnie, gdy opracowywany system ma współpracować z niskiej jakości sprzętem np. tanimi kamerami internetowymi.



Rys. 2.6. Wartości macierzy splotu do obliczania rozkładu gradientów i zobrazowanie rezultatu ich działania

Rysunek 2.7 ilustruje przykład działania filtru medianowego dla macierzy 3x3 piksele. Uwidocznione jest całkowite wyeliminowanie szumu (białych pikseli) jednak obraz stracił na ostrości. Powiększanie macierzy filtru medianowego powoduje dalszą degradację w dziedzinie ostrości.



Rys. 2.7. Zobrazowanie usunięcia szumu przy użyciu filtru medianowego

### 2.3. Omówienie bazowej metody lokalizacji obiektów owalnych

Podczas analizy istniejących rozwiązań, a biorąc pod uwagę cel postawiony w niniejszej rozprawie, szczególną uwagę zwrócono na algorytm detekcji obiektów owalnych [Mai98, Bry00]. Metoda ta spełnia większość podstawowych założeń niniejszej rozprawy, a główną jej zaletą jest bardzo szybka praca. Zalicza się ją do rozwiązań bazujących na niezmiennikach. W uproszczeniu idea działania polega na poszukiwaniu obiektów owalnych (a taki kształt niewątpliwie posiadają głowy ludzkie) przy wykorzystaniu macierzy gradientów, które pozwalają badać lokalne zmiany luminancji w obrazie. Przeważnie do liczenia macierzy rozkładu gradientów stosowane są obrazy w skali szarości, a informacja o kolorze jest pomijana. Dane uzyskane z tych obliczeń mogą mieć wiele zastosowań, np. wykrywanie krawędzi, kształtów – w tym obiektów owalnych [Mai98].

Niewątpliwie obiektami owalnymi są również inne przedmioty, np. piłki, balony, żarówki, itd., dlatego algorytm lokalizacji obiektów owalnych potrzebuje dobrego systemu weryfikacji otrzymanego wyniku. Niemniej – co warto zauważyć – otrzymane wyniki wykazują bardzo dobry współczynnik bezpośrednich trafnych lokalizacji, jeśli twarz rzeczywiście jest obecna na obrazie, dlatego metoda ta wzbudziła zainteresowanie, szczególnie, że została porzucona przez naukowców, a zauważono sporo miejsc, gdzie można poprawić ostateczny wynik działania. Jej szybkość i trafność pozwala założyć, że dokonana zostanie znaczna redukcja ilości danych dostarczanych do bloku decyzyjnego twarz/„nie twarz”.

Wyodrębnianie twarzy za pomocą metody lokalizacji obiektów okrągłych z wykorzystaniem macierzy gradientów można podzielić na następujące etapy [Bry00]:

- wprowadzenie obrazu do pamięci,
- rozciągnięcie histogramu,
- nieproporcjonalne rozszerzenie obrazu,
- obliczenie macierzy gradientów  $G_x$  i  $G_y$ ,
- analiza poszczególnych punktów obrazu,
- tworzenie MT (macierzy trafień),
- analiza i obróbka MT,
- poszukiwanie współrzędnych środka twarzy,
- obliczenie rzeczywistych współrzędnych środków twarzy.

Algorytm zaczyna swoją pracę od kroku normalizacyjnego polegającego na zamianie przestrzeni kolorów obrazu na skalę szarości (o ile jest w przestrzeni RGB bądź CMYK), po czym wykonuje operacje rozciągnięcia histogramu i filtrów splotu. Następnie skaluje obraz

wejściowy do rozmiarów  $90 \times 120$  pikseli. Kolejny krok ma za zadanie doprowadzenie kształtu twarzy do maksymalnie bliskiego okręgu z wyjściowej elipsy. W tym celu zastosowane jest nieproporcjonalne skalowanie. Operacja ta zasadniczo zmniejsza złożoność obliczeniową późniejszych obliczeń ponieważ okrąg jest szczególnym przypadkiem elipsy, a przez to część wzorów ulega znacznemu uproszczeniu. Eksperymenty dowiodły, że szerokość twarzy wynosi około 75% jej wysokości.

W celu wyznaczenia rozkładu gradientów oblicza się macierze  $Gx$  i  $Gy$  zawierające lokalne zmiany luminancji w dwu osiach – poziomej i pionowej [9]:

$$Gx_{y,x} = \frac{1}{2} (P_{y,x+1} - P_{y,x-1}), \quad (2.7)$$

$$Gy_{y,x} = \frac{1}{2} (P_{y+1,x} - P_{y-1,x}), \quad (2.8)$$

gdzie:

$P_{y,x}$  - wartość koloru w punkcie o współrzędnych  $[y,x]$ ,

$Gx_{y,x}$  - wartość gradientu poziomego w punkcie  $P_{y,x}$ ,

$Gy_{y,x}$  - wartość gradientu pionowego w punkcie  $P_{y,x}$ .

Długość wektora gradientu  $G_{y,x}$  dla każdego piksela obrazu obliczamy z prostego wzoru opartego na twierdzeniu Pitagorasa:

$$G_{y,x} = \sqrt{(Gx_{y,x})^2 + (Gy_{y,x})^2}, \quad (2.9)$$

gdzie:

$G_{y,x}$  - długość wektora gradientu w punkcie o współrzędnych  $[y,x]$ ,

$Gx_{y,x}$  - wartość gradientu poziomego w punkcie o współrzędnych  $[y,x]$ ,

$Gy_{y,x}$  - wartość gradientu pionowego w punkcie o współrzędnych  $[y,x]$ .

Rysunek 2.8 ilustruje macierz  $Gx$  (2.8a),  $Gy$  (2.8b) i ich suma (2.8c). Widać na nim, że macierze gradientów dobrze podkreślają krawędzie w dostarczonym obrazie, w tym również krawędzie twarzy.



Rys. 2.8. Zobrazowanie macierzy gradientów  $Gx$  (a) i  $Gy$  (b) i ich suma (c) (Źródło: opracowanie własne)

Posiadając długość wektora  $G$  z dalszych obliczeń wyeliminowane zostają te punkty, w których gradienty przyjmują wartości zbyt małe, a ich analiza spowodowałaby jedynie wydłużenie czasu wyodrębniania. Ponadto mogłyby zniekształcić wyniki i niekorzystnie wpłynąć na efekt końcowy.

Właściwa procedura poszukiwania środka twarzy polega na analizie trzech parametrów określających gradienty: składowej poziomej  $G_x$ , składowej pionowej  $G_y$ , oraz promienia  $R$ , a ideę działania przedstawia rysunek 2.9. Mając wartości macierzy gradientów  $G_x$  i  $G_y$ , łatwo można wyznaczyć wektor  $G$ , którego kierunek jest pomocny przy wyznaczeniu punktu  $S$  (rys. 2.9). Obliczeń dokonuje się dla wszystkich punktów, których wartości nie zostały wyeliminowane w poprzednim procesie. Poniższy rysunek przedstawia zobrazowanie analizy jednego z takich punktów.

Oznaczenia:

- $P$  - analizowany punkt obrazu,
- $S_{y,x}$  - prawdopodobny środek twarzy,
- $R$  - spodziewany promień twarzy,
- $G_{y,x}$  - wektor gradientu dla punktu  $P$  (otrzymany z  $G_{x_{y,x}}$  i  $G_{y_{y,x}}$ ).
- $Gx'_{y,x}$  - składowa pozioma wektora  $R$ ,
- $Gy'_{y,x}$  - składowa pionowa wektora  $R$ .



Rys. 2.9. Wyznaczanie  $S$  poprzez obliczanie przesunięcia  $Gx'_{y,x}$  i  $Gy'_{y,x}$  w oparciu o gradienty i promień  $R$

Punkty środka twarzy wyznaczamy posługując się wzorami, które łatwo wyprowadzić w oparciu o powyższy rysunek:

$$R^2 = (Gx'_{y,x})^2 + (Gy'_{y,x})^2, \quad (2.10)$$

$$G_{y,x}^2 = (Gx_{y,x})^2 + (Gy_{y,x})^2, \quad (2.11)$$

$$\frac{Gx_{y,x}}{G_{y,x}} = \frac{Gx'_{y,x}}{R}, \quad (2.12)$$

$$\frac{Gy_{y,x}}{G_{y,x}} = \frac{Gy'_{y,x}}{R}. \quad (2.13)$$

Po przekształceniach otrzymujemy:

$$Gx'_{y,x} = \frac{R \cdot Gx_{y,x}}{\sqrt{Gx_{y,x}^2 + Gy_{y,x}^2}}, \quad (2.14)$$

$$Gy'_{y,x} = \frac{R \cdot Gy_{y,x}}{\sqrt{Gx_{y,x}^2 + Gy_{y,x}^2}}. \quad (2.15)$$

Poszukiwane współrzędne środków  $S$  wyniosą:

$$S_{y,x} = (y + Gy'_{y,x}, x + Gx'_{y,x}), \quad (2.16)$$

gdzie:

$x$  - współrzędna  $x$  badanego punktu,

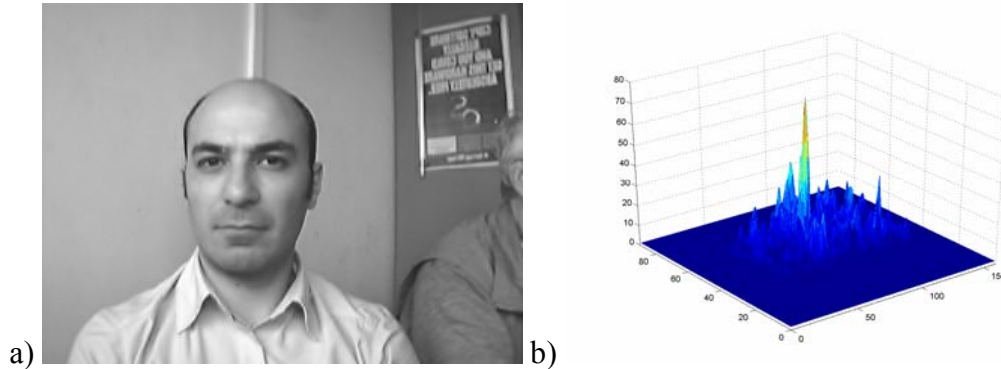
$y$  - współrzędna  $y$  badanego punktu.

Wyżej omówione obliczenia dla pojedynczego punktu dokonuje się następnie dla wszystkich punktów spełniających warunki początkowe. Aby przechowywać wyniki tych obliczeń tworzona jest MT (macierz trafień) o wartościach początkowych równych 0. Każde wyliczone  $S$  inkrementuje wartości MT według wartości ukazanych w rysunku 2.10, gdzie współrzędne punktu  $z$  wartością +3 są identyczne z współrzędnymi punktu  $S$ . Ten zmodyfikowany sposób w jaki inkrementowane są wartości MT ma zaletę polegającą na zwiększeniu efektywności celności trafień, ponieważ powoduje zmniejszenie czułości całego systemu na drobne zakłócenia.

|  |    |    |    |  |
|--|----|----|----|--|
|  |    |    |    |  |
|  | +1 | +2 | +1 |  |
|  | +2 | +3 | +2 |  |
|  | +1 | +2 | +1 |  |
|  |    |    |    |  |

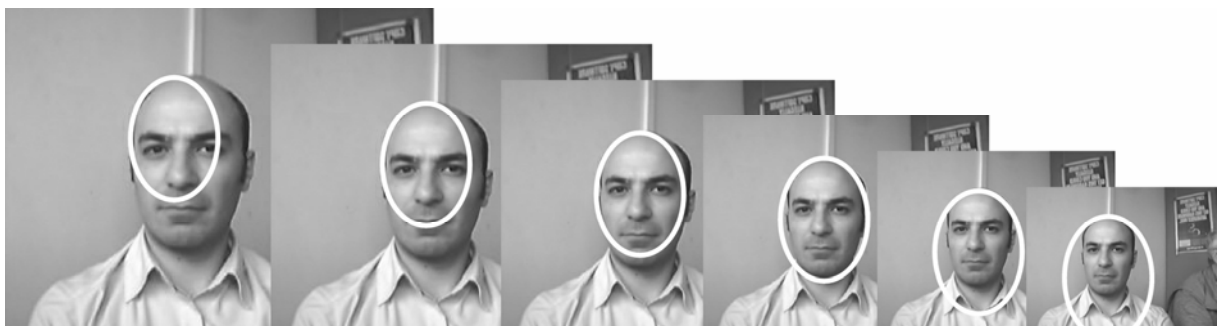
Rys. 2.10. Sposób uzupełniania MT (macierzy trafień)

Po analizie wszystkich punktów otrzymujemy obraz umownie nazwany „macierzą trafień”, co obrazuje rysunek 2.11b (powstała ona po analizie obrazu przedstawionego w rysunku 2.11a).



Rys. 2.11. Obraz badany – 240×320 pikseli (a) i „macierz trafień” (b) dla promienia 22 piksele

Na rys. 2.11b widać wyraźnie, że w okolicach centrum twarzy powstało największe skupisko pikseli o dużych wartościach. Znajduje się tam również maksymalna wartość z całej MT. Z dużym prawdopodobieństwem jest ona środkiem szukanej twarzy. Należy tutaj dodać, że twarz musi posiadać taką samą wielkość jak promień poszukujący, ponieważ tylko wtedy wartość punktu środka twarzy będzie inkrementowana przez trafiające w niego promienie  $R$  zgodne z  $G_x$  i  $G_y$ . Aby przeciwdziałać temu ograniczeniu przeważnie stosuje się piramidę obrazów (rys. 2.12). Jest to rozwiązanie kosztowne czasowo, ale pozwala przeszukiwać obraz w celu znalezienia twarzy tylko jednym ustalonym promieniem  $R$ . Obraz wejściowy jest skalowany do różnych rozdzielczości, więc twarz będzie miała za każdym razem inną wielkość przedstawioną w pikselach. Logicznym jest więc założenie, że trafi się w końcu taka wielkość twarzy (na którymś z przeskalowanych obrazów), która odpowiada promieniowi poszukującemu  $R$  i wynik poszukiwania będzie tam najwyższy. Zobrazowanie całej metody lokalizacji obiektów owalnych przy pomocy macierzy gradientów opisanej powyżej przedstawione zostało na poniższym rysunku.

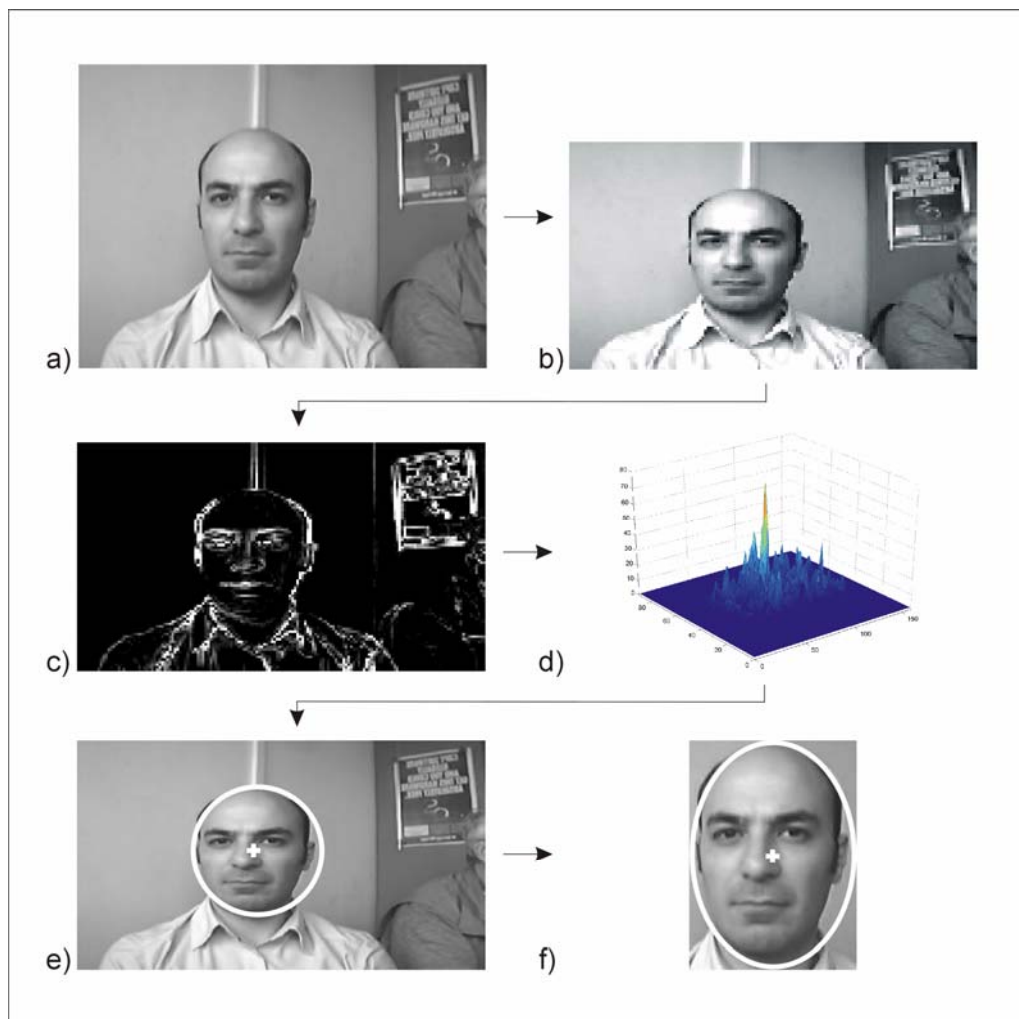


Rys. 2.12. Piramida zdjęć z wynikami poszukiwań dla zadanego  $R$  (Źródło: opracowanie własne)

Należy tutaj dodać, że poszukiwanie owalów na wielu przeskalowanych obrazach niesie za sobą wygenerowanie olbrzymiej ilości potencjalnych twarzy, gdyż dla każdego z nich tworzona jest MT.

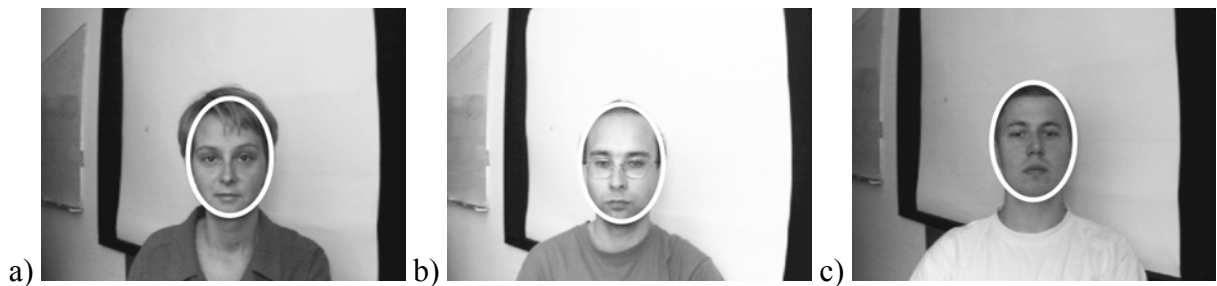
W dużym skrócie powyższą metodę można przedstawić jako zestaw algorytmów wykonywanych w kolejności:

- wprowadzenie obrazu do analizy (rys. 2.13a),
- obróbka wstępna (rys. 2.13b),
- obliczenie macierzy gradientów (rys. 2.13c),
- kalkulacja MT (rys. 2.13d),
- wybór najwyższej wartości z MT (rys. 2.13e),
- wyodrębnienie twarzy (rys. 2.13f).

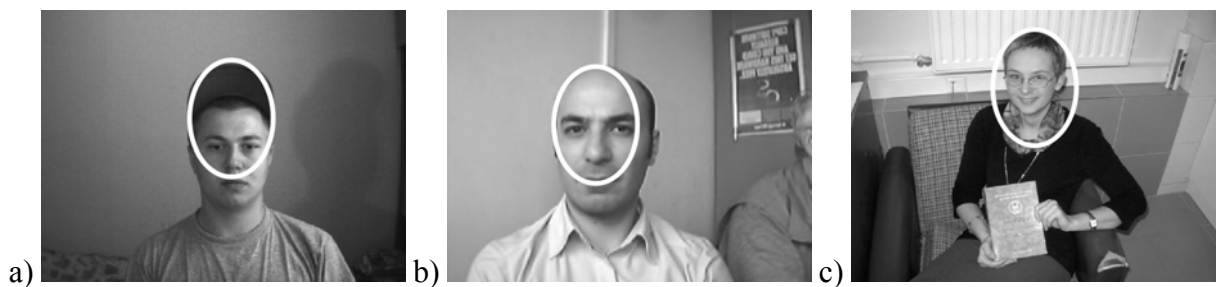


Rys. 2.13. Schemat działania metody lokalizacji obiektów owalnych z wykorzystaniem macierzy gradientów: obraz wejściowy (a), znormalizowane nieproporcjonalne skalowanie (b), mapa rozkładu gradientów (c), MT (d), wynik poszukiwania (e), wyodrębniona twarz (f)

Aby bliżej zobrazować wyniki działania algorytmu i jego ograniczenia występujące w postaci znanej z publikacji [Mai98] poniżej przedstawione są rysunki obrazujące poprawne wyniki działania algorytmu (rys. 2.14) jak również te błędne (rys. 2.15). Uwidaczniają one niezłą odporność na zmienne oświetlenie, co było jednym z kolejnych argumentów za bliższym przyjrzeniem się metodzie poszukiwania owalów. Niestety w przypadku, gdy promień twarzy nie odpowiada długości promienia  $R$  generowane są błędne wyniki. Również inne obiekty owalne zakłócające „owalność” twarzy mogą powodować błędy (w tym przypadku jest to czapka).



Rys. 2.14. Przykładowe wyniki poprawnego działania metody wykrywania obiektów owalnych



Rys. 2.15. Przykłady błędnego działania metody wykrywania obiektów owalnych bazującej na stałym promieniu  $R$

## 2.4. Omówienie zmodyfikowanego algorytmu lokalizacji obiektów owalnych

Podczas analizy opisanego powyżej algorytmu zauważono, że nie został on dostatecznie przebadany, a prace nad nim zarzucono w 1998 roku. Jego cechy, w tym szczególnie niezwykle duża szybkość wykonywania, spowodowały, że algorytm ten został poszerzony i wnikliwie przebadany, a rezultaty działania przewyższają te oryginalne i to w sposób znaczący. Zachowano przy tym zaletę szybkości, a nawet ją poprawiono, jednocześnie wyeliminowano niektóre wady. Przede wszystkim uniezależniono cały system od stałego promienia poszukiwań, co pozwoliło lokalizować twarze o różnych wielkościach, jako opcję umożliwiono poszukiwanie wielu twarzy, wykorzystano informację o kolorze (co poprawia

efektywność), jak również dodano dodatkowe macierze poszukiwań. Wszystkie te, jak i inne usprawnienia zostały dogłębnie przemyślane i przeanalizowane pod kątem sensowności ich użycia, a działanie przetestowane na bazach danych, zarówno komercyjnych jak i autorskich. Metodę zaimplementowano w języku C++, co umożliwiło porównanie szybkości działania z komercyjnymi detektorami.

Usprawniony algorytm wykrywania obiektów owalnych z wykorzystaniem macierzy rozkładu gradientów zawiera następujące kroki (dla obrazów kolorowych):

- konwersja z przestrzeni kolorów *RGB* do przestrzeni *III2I3*,
- rozciągnięcie histogramu,
- obliczanie macierzy gradientów  $G_x$  i  $G_y$ ,
- analiza poszczególnych punktów obrazu z wykorzystaniem  $G_x$ ,  $G_y$  i zmiennego promienia poszukiwań z jednoczesnym tworzeniem ZMT (zbiorczej macierzy trafień),
- obliczanie macierzy gradientów  $G_{45}$  i  $G_{135}$ ,
- analiza poszczególnych punktów obrazu z wykorzystaniem  $G_{45}$ ,  $G_{135}$  i zmiennego promienia poszukiwań z jednoczesną aktualizacją ZMT,
- analiza i obróbka ZMT,
- poszukiwanie współrzędnych środka jednej lub wielu twarzy,
- obliczenie rzeczywistych współrzędnych środków twarzy.

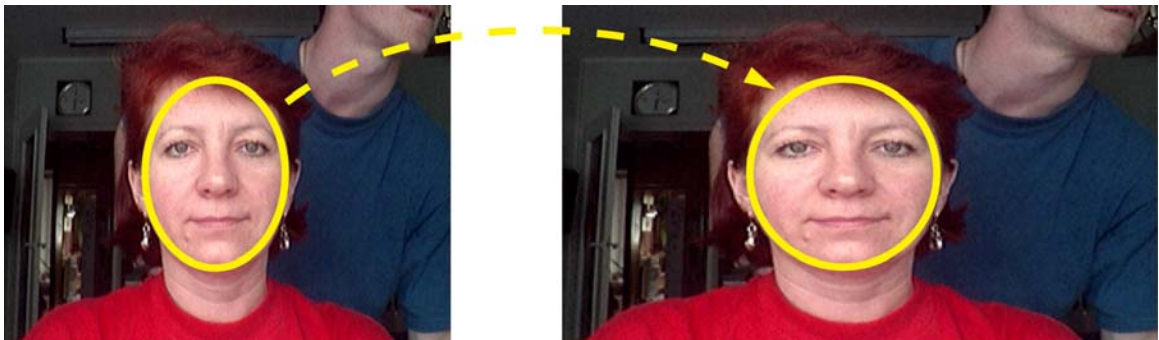
Cały rozdział 2.4 podzielono na podrozdziały w celu lepszej orientacji, a należą do nich: Obróbka wstępna; Macierze gradientów; MT (Macierz Trafień); Zmienny promień poszukiwań; ZMT (Zbiorcza Macierz Trafień); Wyznaczanie pięciu obszarów zawierających twarz; Wyznaczanie wielu twarzy; Wersja dla obrazów w skali szarości.

Poniżej zostanie przedstawione szczegółowe działanie algorytmów składających się na działanie metody. Szczególną uwagę poświęcono analizie nowych rozwiązań zwiększających efektywność działania.

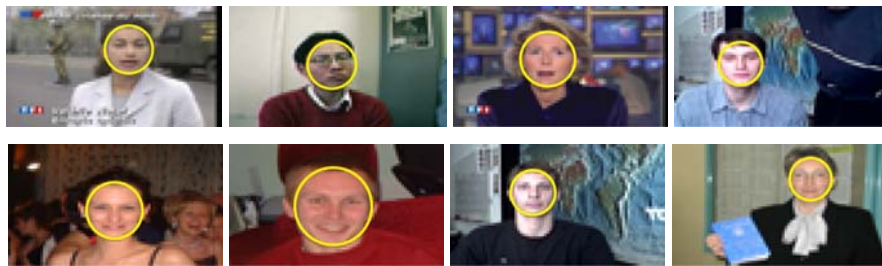
#### 2.4.1. Obróbka wstępna

Pierwszym krokiem algorytmu, a zarazem pierwszym usprawnieniem działania metody jest nieproporcjonalne przeskalowanie (opisane w rozdziale 2.2) obrazu wejściowego do rozmiarów  $45 \times 80$  pikseli (rozmiar ten jest obowiązujący dla obrazów wejściowych o proporcjach 4:3). Łatwo zauważyć, że powyższy rozmiar nie spełnia proporcji 4:3. Jest to celowy zabieg mający na celu zbliżenie kształtu twarzy do okręgu, ponieważ standardowo jest ona owalem (elipsą). Powód jest oczywisty – obliczenia konieczne do przeprowadzenia dla kształtów okrągłych w porównaniu z owalnymi (eliptycznymi) są dużo prostsze i znacznie

mniej kosztowne czasowo dla jednostki obliczeniowej (okrąg to szczególny przypadek elipsy). Badania przeprowadzone na obrazach twarzy wykazały, że szerokość twarzy wynosi w przybliżeniu 75% wysokości [Kuk03]. Dlatego rozmiar docelowego obrazu przeznaczonego do dalszej analizy jest, tak jak napisano na początku rozdziału, równy  $45 \times 80$  pikseli. Należy dodać, że skalowanie do tych wartości jest przeprowadzane obligatoryjnie dla wszystkich obrazów wejściowych, niezależnie od posiadanej rozdzielczości. Dowiedziono, że zmniejszenie wymiarów obrazu jest korzystne w dwóch aspektach: czasowym i jakościowym. O ile pierwszy jest oczywisty ponieważ mniej pikseli, to mniej obliczeń, tak drugi jest już wynikiem wnikliwych badań i obserwacji wyników, które wykazały, że zmniejszenie rozdzielczości jest dodatkowym, skutecznym krokiem normalizującym i eliminującym szum z obrazu. Poniższe rysunki prezentują efekt działania nieproporcjonalnego skalowania.



Rys. 2.16. Przedstawienie koncepcji nieproporcjonalnego skalowania



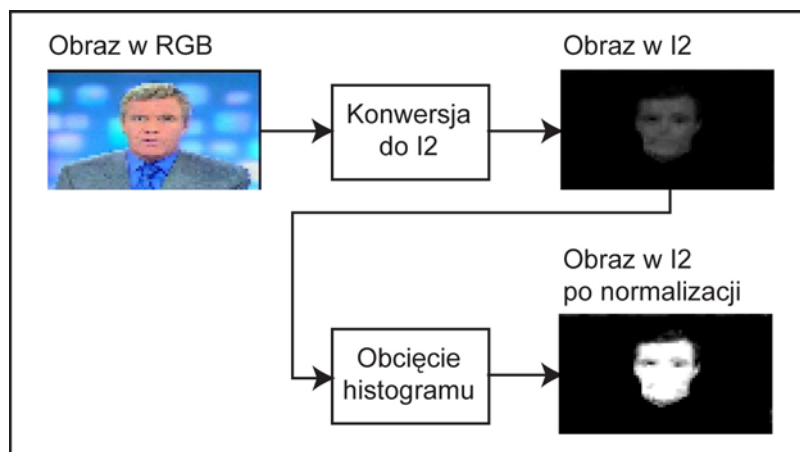
Rys. 2.17. Zobrazowanie „okrągłości” twarzy po przeskalowaniu do rozdzielczości  $80 \times 45$  pikseli

Kolejnym krokiem algorytmu jest zbadanie w jakiej przestrzeni barw są dane wejściowe. Jeśli wykryty zostanie tryb *RGB* to algorytm dokonuje konwersji tej przestrzeni do komponentu *I2* z przestrzeni *III2I3* [Kuk03], a opisaną w rozdz. 2.2. Wybór komponentu *I2* poprzedzony został badaniami, które wykazały, że doskonale nadaje się do poszukiwania twarzy, gdyż twarz jest na nim widoczna, a pozostałe obszary obrazu nie będące w odcieniu zbliżonym do twarzy są ukryte. Jest to niewątpliwa zaleta, która udowodniła swoją wartość podczas testów.

Jednak obraz w komponencie  $I2$  bez wstępnej obróbki nie jest wartościowym materiałem do analizy. Dlatego zastosowano operację rozciągnięcia histogramu, która mimo swej prostoty działania powoduje polepszenie jakości materiału pod względem widoczności twarzy, a przez to znacząco poprawia efektywność całego systemu. Badania dowiodły, że opłaca się obciąć znaczną część histogramu, w sposób stały – pełny zakres od 0 do 255 zawężony jest do przedziału od 0 do 70 i rozciągnięty ponownie do 256 wartości.

Operacja rozciągnięcia histogramu opisana została w rozdziale 2.2. To drastyczne obcięcie znaczącej części histogramu wynika ze spostrzeżenia, że wartości odpowiedzialne za kolor skóry znajdują się w zakresie 70-200, więc obcięcie wartości powyżej 70 i rozciągnięcie na cały zakres, spowoduje przypisanie pikselom twarzy wartości bliskie 255, co oznacza kolor biały. Powoduje to natomiast powstanie dużego kontrastu na granicy twarz/tło. Ta cecha jest bardzo pożądana dla działania algorytmu lokalizacji obiektów owalnych z wykorzystaniem macierzy gradientów. Dodatkowo użycie komponentu  $I2$  z krokiem normalizacyjnym powoduje, że znaczna część pikseli obrazu przyjmuje wartości bliskie 0 lub 255 co oznacza, że gradienty będą bardzo małe, a więc poniżej progu, który jest założony. Dlatego te punkty są pomijane przez główny algorytm (jest to krok optymalizacyjny i jednocześnie eliminujący wpływ szumu cyfrowego) co dodatkowo zwiększa szybkość jak i efektywność całego systemu.

Poniższy rysunek obrazuje przekształcenie obrazu z przestrzeni  $RGB$  do komponentu  $I2$  zarówno przed i po kroku normalizacyjnym.



Rys. 2.18. Zobrazowanie konwersji do  $I2$  z krokiem normalizacyjnym

Tabela 2.1 jest próbą zobrazowania zasadności użycia komponentu  $I2$  do obliczania macierzy gradientów w zastępstwie zwykłego obrazu w skali szarości.

W każdym wierszu widocznych jest pięć różnych obrazów:

- oryginalny w *RGB*,
- konwersja do skali szarości,
- zobrazowanie macierzy gradientów dla obrazu w skali szarości,
- konwersję do komponentu *I2* z normalizacją (obcięcie i rozciągnięcie histogramu),
- zobrazowanie macierzy gradientów dla obrazu w komponentcie *I2* z przestrzeni *III2I3*.

| Obraz w przestrzeni barw RGB                                                        | Obraz w odcieniach szarości i komponentcie I2                                        | Ilustracje macierzy gradientów                                                        |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|    |    |    |
|                                                                                     |   |   |
|  |  |  |
|                                                                                     |  |  |
|  |  |  |
|                                                                                     |  |  |

Tabela 2.1. Przykładowe konwersje obrazów z przestrzeni kolorów RGB do komponentu *I2* z przestrzeni *III2I3* wraz z odpowiadającymi im ilustracjami macierzy gradientów

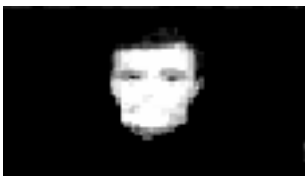
| Obraz w przestrzeni barw RGB                                                        | Obraz w odcieniach szarości i komponencie I2                                         | Ilustracje macierzy gradientów                                                        |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|    |    |    |
|                                                                                     |    |    |
|    |    |    |
|                                                                                     |   |   |
|  |  |  |
|                                                                                     |  |  |
|  |  |  |
|                                                                                     |  |  |

Tabela 2.1 cd.

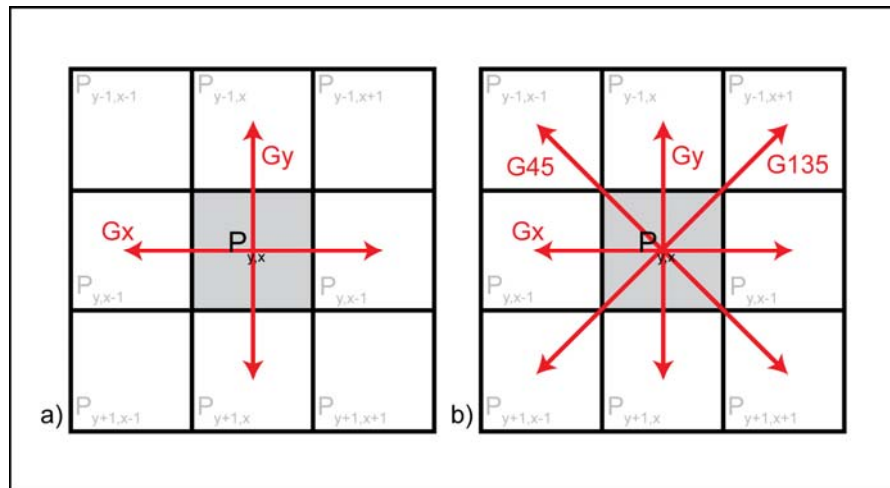
Tabela 2.1 ukazuje niewątpliwą zaletę użycia komponentu *I2*. Większość obrazów poddana tej konwersji daje w wyniku znacznie korzystniejsze dane do wykorzystania w lokalizatorze obszarów zawierających twarz używającym macierze gradientów. Są one dobrze wyizolowane z tła, zaznaczony jest zdecydowany kontrast na granicy twarz-tło i – co warto uwagi – także na granicy twarz-włosy (jedynie włosy koloru rudego nie gwarantują tego zjawiska). Dodatkowo ilość obszarów, gdzie występują dodatnie, wysokie wartości jest niewielka, co skraca czas wykonywania algorytmu. Oczywiście może się zdarzyć przypadek, gdy tło będzie w kolorze zbliżonym do koloru twarzy – to najbardziej pesymistyczny wariant, ale nawet wtedy powstaje kontrast na granicy twarz-tło, co daje możliwość poprawnej pracy systemu.

Warto uwagi jest to, że obrazy użyte w tabeli 2.1 pochodzą z różnych źródeł: TV, aparatów cyfrowych (różnych producentów), i kamer internetowych (bardzo słabej jakości). Ten zabieg jest wprowadzony celowo, ponieważ pokazuje jak bardzo zróżnicowane mogą być dane wejściowe i jak ten czynnik wpływa na jakość obrazu, która jest nierozdzielnie związana z efektywnością całego systemu. Niewątpliwie najgorsze dane generują proste kamery internetowe posiadające najniższej jakości matryce CCD, jak również słabe, ciemne obiektywy. Nie należy jednak pomijać ich przy testowaniu algorytmów lokalizacji, jak również detekcji, gdyż to one są głównym źródłem danych wejściowych, ze względu na powszechność ich zastosowania w systemach czasu rzeczywistego, gdzie np. nadzór jest prowadzony w sposób ciągły – tam źródłem jest właśnie kamera cyfrowa. Różnorodność zdjęć pomaga również zobrazować jak warunki oświetleniowe wpływają na macierze gradientów, a właśnie odporność na te, jak i inne zakłócenia świadczy o jakości generatora kandydatów.

#### 2.4.2. Macierze gradientów

Po obcięciu i rozciągnięciu histogramu algorytm przechodzi do wyznaczania macierzy gradientów. Operacja ta jest rozszerzona w stosunku do bazowej metody lokalizacji obiektów owalnych. Poza wyznaczaniem macierzy gradientów w płaszczyźnie poziomej i pionowej (czyli dla kątów  $0^\circ$  i  $90^\circ$ ) obliczane są dwie dodatkowe macierze dla płaszczyzn ukośnych (czyli dla kątów  $45^\circ$  i  $135^\circ$ ). Te dodatkowe macierze są wykorzystywane do wydobywania dodatkowych informacji z obrazu, które są następnie użyte do zwiększenia efektywności całego systemu. Idea użycia tych dodatkowych, ukośnych macierzy jest uzasadniona spostrzeżeniem, że gradienty, czyli kontrasty powstają zarówno w płaszczyznach „prostokątnych” jak i każdych innych. Można wyznaczyć inne kąty i dla nich obliczać

gradienty, jednak jedynie dla kątów  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  i  $135^\circ$  są one stworzone z wartości pozyskanych realnych pikseli co pokazuje rysunek 2.19. Ukazuje on również idee zasadności użycia wyżej wymienionych czterech macierzy. Widać na nim wyraźnie, że niewykorzystanie macierzy ukośnych powoduje pominięcie istotnej informacji z otoczenia piksela  $P_{y,x}$ , a przez to prawdopodobnie uzyskanie słabszego wyniku. Te czysto teoretyczne rozumowanie, dotyczące wykorzystania dodatkowych dwóch macierzy ukośnych, zostało dowiedzione eksperymentalnie. Należy zauważyć, że wszystkie obliczenia dla kątów  $45^\circ$  i  $135^\circ$  przeprowadzane są niezależnie od tych dla  $0^\circ$  i  $90^\circ$ . Oznacza to, że algorytm jest w tym momencie rozdzielony na dwie ścieżki obliczeń, które zbiegają się dopiero w MT, która jest inkrementowana symultanicznie, zarówno przez wyniki dla pary  $0^\circ/90^\circ$  jak i  $45^\circ/135^\circ$ .



Rys. 2.19. Ilustracja obrazująca zasadność obliczania gradientów w czterech płaszczyznach:

a) metoda bazowa, b) metoda rozszerzona (Źródło: opracowanie własne)

W celu zilustrowania zasadności użycia macierzy gradientów w płaszczyznach ukośnych, poniżej (rys. 2.20) przedstawione są obrazy: wejściowy, w komponencie  $I2$  z przestrzeni  $III2I3$ , macierze gradientów dla  $0^\circ$ ,  $90^\circ$ ,  $45^\circ$ ,  $135^\circ$ , suma  $0^\circ/90^\circ$ , suma  $45^\circ/135^\circ$  i suma wszystkich –  $0^\circ/90^\circ/45^\circ/135^\circ$ . Wartym podkreślenia jest fakt, że suma wszystkich czterech macierzy nie jest w rzeczywistości analizowana ponieważ algorytm jest rozdzielony na dwie oddzielne ścieżki dla obu par macierzy, jednak doskonale ukazuje, że dodatkowe macierze ukośne dodają użytecznych informacji do przetwarzania. Natomiast suma macierzy  $45^\circ/135^\circ$  ilustruje, że są one tak samo dobrym źródłem do przetwarzania jak macierze z bazowej metody i przez to są pełnowartościowym materiałem. Przeprowadzono testy, które wykazały, że można użyć jedynie tych ukośnych macierzy, a wynik będzie zbliżony do tego uzyskanego z analizy macierzy  $0^\circ/90^\circ$ .

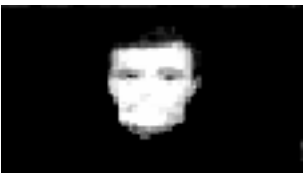
|                                                    |                                                                                     |                                                                                      |                                                                                       |
|----------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Obraz w RGB                                        |    |    |    |
| Obraz w komponentcie I2 z etapem norm.             |    |    |    |
| Macierz gradientów dla kąta 0°                     |    |    |    |
| Macierz gradientów dla kąta 90°                    |    |    |    |
| Macierz gradientów dla kąta 45°                    |   |   |   |
| Macierz gradientów dla kąta 135°                   |  |  |  |
| Suma macierzy gradientów dla kątów 0°/90°          |  |  |  |
| Suma macierzy gradientów dla kątów 0°/90°          |  |  |  |
| Suma macierzy gradientów dla kątów 0°/90°/45°/135° |  |  |  |

Tabela 2.2. Ilustracja macierzy gradientów dla wszystkich kątów (0°, 90°, 45°, 135°) i ich sumy dla odpowiadającym im obrazów wejściowych.

Macierze gradientów wyznacza się w następujący sposób:

$$Gx_{y,x} = \frac{1}{2}(P_{y,x+1} - P_{y,x-1}), \quad (2.17)$$

$$Gy_{y,x} = \frac{1}{2}(P_{y+1,x} - P_{y-1,x}), \quad (2.18)$$

$$G45_{y,x} = \frac{1}{2}(P_{y+1,x+1} - P_{y-1,x-1}), \quad (2.19)$$

$$G135_{y,x} = \frac{1}{2}(P_{y+1,x-1} - P_{y-1,x+1}), \quad (2.20)$$

gdzie:

$P_{y,x}$  - wartość koloru w punkcie o współrzędnych  $[y,x]$ ,

$Gx_{y,x}$  - wartość gradientu poziomego w punkcie  $P_{y,x}$ ,

$Gy_{y,x}$  - wartość gradientu pionowego w punkcie  $P_{y,x}$ .

$G45_{y,x}$  - wartość gradientu pod kątem  $45^\circ$  w punkcie  $P_{y,x}$ ,

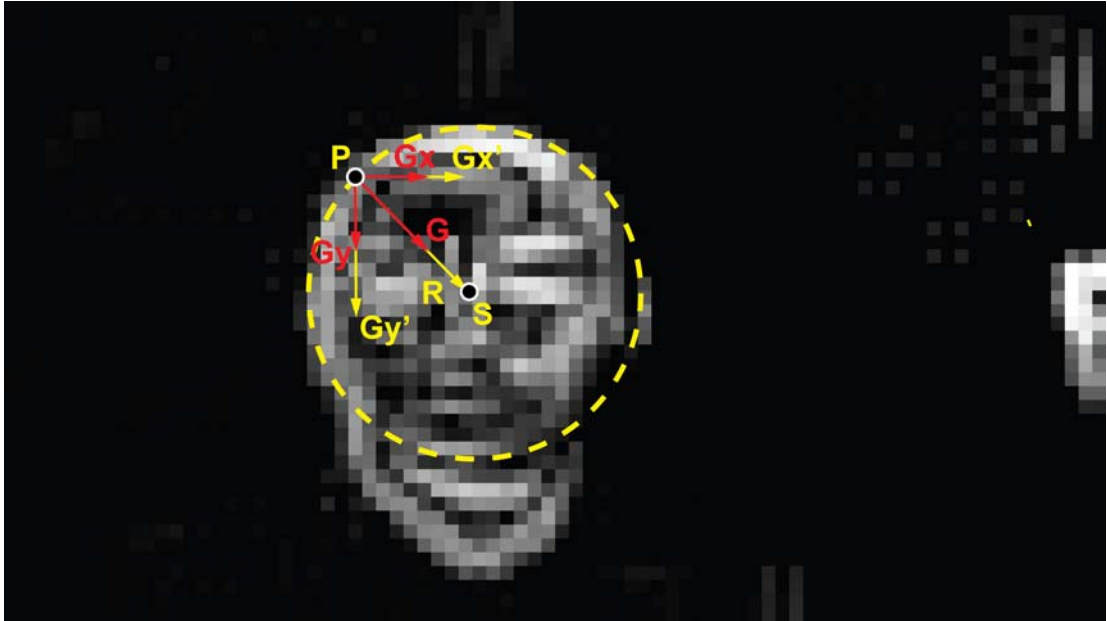
$G135_{y,x}$  - wartość gradientu pod kątem  $135^\circ$  w punkcie  $P_{y,x}$ .

Następnie, w odróżnieniu od bazowej metody wykrywania owalów, nie jest obliczana obligatoryjnie długość wektora gradientów dla każdego piksela (ze wzoru 2.9), ponieważ jest on kosztowny czasowo. Badana jest natomiast wartość gradientu w każdym kierunku dla każdej z par ( $0^\circ/90^\circ$  i  $45^\circ/135^\circ$ ). Służy to do oszacowania, czy dokonywać obliczeń dla badanego piksela, czy też go pominąć. Założenie jest proste: jeśli wartości gradientu są niskie, to z dużym prawdopodobieństwem jest to jedynie szum, a nie granica obiektu (w tym np. twarzy). Dlatego, jeśli żadna z bezwzględnej wartości gradientu nie przekracza założonego progu, algorytm pomija bieżący punkt i przechodzi do analizy kolejnego.

#### 2.4.3. MT (Macierz Trafień)

Po wyznaczeniu macierzy gradientów algorytm przechodzi do wyliczania MT (macierz trafień) przy użyciu dwóch par macierzy gradientów: pionowej/poziomej i ukośnych. Etap ten przebiega dwutorowo, całkowicie niezależnie. Dopiero wyniki „trafień” są wpisywane do jednej ZMT (zbiorczej macierzy trafień), która stanowi nowość w stosunku do metody bazowej, a jej założenia, sposób inkrementacji i analizy zostały opracowane na potrzeby zmodyfikowanej metody lokalizacji obiektów owalnych. Aby zobrazować ten proces, poniżej zaprezentowane zostaną oddzielnie dwie ścieżki.

Poniższy rysunek (2.24) przedstawia analizę jednego punktu z obszaru pogranicza twarzy dla macierzy gradientów pionowej i poziomej.



Rys. 2.20. Wyznaczanie przesunięcia  $Gx'$  i  $Gy'$  w oparciu o gradienty i promień  $R$  (Źródło: opracowanie własne)

Punkty środka twarzy wyznaczamy posługując się wzorami, które łatwo wyprowadzić w oparciu o powyższy rysunek:

$$R^2 = (Gx')^2 + (Gy')^2, \quad (2.21)$$

$$G^2 = (Gx)^2 + (Gy)^2, \quad (2.22)$$

$$\frac{Gx}{G} = \frac{Gx'}{R}, \quad (2.23)$$

$$\frac{Gy}{G} = \frac{Gy'}{R}. \quad (2.24)$$

Po przekształceniach otrzymujemy:

$$Gx' = \frac{R \cdot Gx}{\sqrt{Gx^2 + Gy^2}}, \quad (2.25)$$

$$Gy' = \frac{R \cdot Gy}{\sqrt{Gx^2 + Gy^2}}. \quad (2.26)$$

Poszukiwane współrzędne środków  $S$  wyniosą:

$$S = (y + Gy', x + Gx'), \quad (2.27)$$

gdzie:

$x$  - współrzędna  $x$  badanego punktu,

$y$  - współrzędna  $y$  badanego punktu.

Analogicznie wyznaczamy punkt „trafień” dla ukośnych macierzy gradientów o kierunkach  $45^\circ$  i  $135^\circ$ . Poniższy rysunek (2.25) ukazuje w jaki sposób analizowany jest pojedynczy punkt obrazu.



Rys. 2.21. Wyznaczanie przesunięcia  $G45'$  i  $G135'$  w oparciu o gradienty i promień  $R$  (Źr.: opracowanie własne)

Z powyższego rysunku można zbudować zależności, przy pomocy których wyprowadzono wartości przesunięć, które w wyniku dadzą współrzędne punktu  $S2$ :

$$R2^2 = (G45')^2 + (G135')^2, \quad (2.28)$$

$$G2^2 = (G45)^2 + (G135)^2, \quad (2.29)$$

$$\frac{G45}{G2} = \frac{G45'}{R2}, \quad (2.30)$$

$$\frac{G135}{G2} = \frac{G135'}{R2}. \quad (2.31)$$

Po przekształceniach otrzymujemy:

$$G45' = \frac{R2 \cdot G45}{\sqrt{G45^2 + G135^2}}, \quad (2.32)$$

$$G135' = \frac{R2 \cdot G135}{\sqrt{G45^2 + G135^2}}. \quad (2.33)$$

Poszukiwane współrzędne środków  $S2$  wyniosą:

$$S2 = (y + \frac{G45'}{\sqrt{2}} + \frac{G135'}{\sqrt{2}}, x + \frac{G45'}{\sqrt{2}} - \frac{G135'}{\sqrt{2}}). \quad (2.34)$$

gdzie:

$x$  - współrzędna  $x$  badanego punktu,

$y$  - współrzędna  $y$  badanego punktu.

Aby przechowywać wyniki obliczeń zarówno dla pary macierzy gradientów w orientacji pionowej/poziomej, jak również dla ukośnych, tworzona jest MT o wartościach początkowych równych 0. Każde wyliczone  $S$  i  $S2$  inkrementuje wartości tej macierzy. Sposób inkrementacji macierzy stał się powodem badań, ponieważ proste założenie, że gradienty nieomylnie wskażą środek twarzy wydał się zbyt optymistyczny, szczególnie po etapach obróbki wstępnej. Dlatego zaproponowano dwie metody inkrementacji (aktualizacji) MT przedstawione na rysunku 2.22.

|  |    |    |    |  |
|--|----|----|----|--|
|  |    |    |    |  |
|  | +1 | +2 | +1 |  |
|  | +2 | +3 | +2 |  |
|  | +1 | +2 | +1 |  |
|  |    |    |    |  |

Rys. 2.22. Wagi przypisane odpowiednim pikselom w otoczeniu wyliczonego  $S$  (lub  $S2$ ) – wersja 1.

|  |      |      |   |      |      |  |
|--|------|------|---|------|------|--|
|  |      |      |   |      |      |  |
|  | 0,16 | 0,76 | 1 | 0,76 | 0,16 |  |
|  | 0,76 | 1,58 | 2 | 1,58 | 0,76 |  |
|  | 1    | 2    | 3 | 2    | 1    |  |
|  | 0,76 | 1,58 | 2 | 1,58 | 0,76 |  |
|  | 0,16 | 0,76 | 1 | 0,76 | 0,16 |  |
|  |      |      |   |      |      |  |

Rys. 2.23. Wagi przypisane odpowiednim pikselom w otoczeniu wyliczonego  $S$  (lub  $S2$ ) – wersja 2.

Bazują one na założeniu, że algorytm wyznaczania obiektów owalnych będzie generował wyniki obarczone pewnym błędem i MT należy aktualizować w sposób bardziej rozmyty. Polega on na inkrementacji punktu  $S$  i  $S2$ , jak również punktów znajdujących się w jego pobliżu, o wartość przypisanej im wagi. Daje to gwarancje, że mimo minimalnie

błędne wyliczenia środka twarzy, wartość poprawnego punktu opisującego środek twarzy będzie inkrementowana o pewną, choć minimalną, wartość.

#### 2.4.4. Zmienny promień poszukiwań

Poszukiwanie twarzy w metodzie bazowej korzystało z założenia, że posiadając wartości gradientu w płaszczyźnie pionowej i poziomej, jak również założony wcześniej promień poszukiwań, wyznaczone były trafienia  $S$  (patrz rys. 2.9) inkrementujące następnie MT. Rozwiązanie to miało swoją jedną niewątpliwą wadę. Ograniczało możliwości detektora do poszukiwań twarzy o wiadomej od początku wielkości. Oznacza to, że gdy promień twarzy był mniejszy lub większy od założonego promienia poszukiwań, wynik był błędny. Dlatego zaproponowano rozwiązanie, które ten problem wyeliminowało. Bazuje ono na przesłaniu, że poszukiwań można dokonywać wszystkimi dostępnymi promieniami, a wyniki wpisywać do równoległe założonych MT – czyli dla każdego promienia jest tworzona oddzielna macierz. Założono, że promień poszukiwań będzie zawierał się w przedziale od  $M/2$  do  $M/5$ , gdzie  $M$  to wysokość obrazu wejściowego. Ostatecznie przeszukuje się wszystkie obliczone MT w celu wykrycia najwyższej wartości, która połączona z wiadomym promieniem (każda MT ma swój kod mówiący o tym jaki promień poszukiwań ją tworzył) daje w rezultacie wynik będący z dużym prawdopodobieństwem poszukiwaną twarzą. Wynik tej analizy tworzy ZMT (zbiorną macierz trafień) przechowującą dwie liczby dla każdej współrzędnej piksela – wartość gradientu i przypadający jej promień poszukiwań.

Na rysunku 2.24 przedstawiono (w sposób uproszczony) przygotowanie obrazu z postaci  $RGB$  do komponentu  $I2$  z krokiem normalizacyjnym (dodatkowo ilustracja macierzy gradientów – 2.24c), który na kolejnych rysunkach – 2.25, 2.26, 2.27 – jest poddawany poszukiwaniom promieniami o trzech różnych długościach: za krótkim, idealnym i za długim.



Rys. 2.24. Obraz w  $RGB$  (a), w komponentcie  $I2$  (b) i zilustrowanie macierzy gradientów (c)

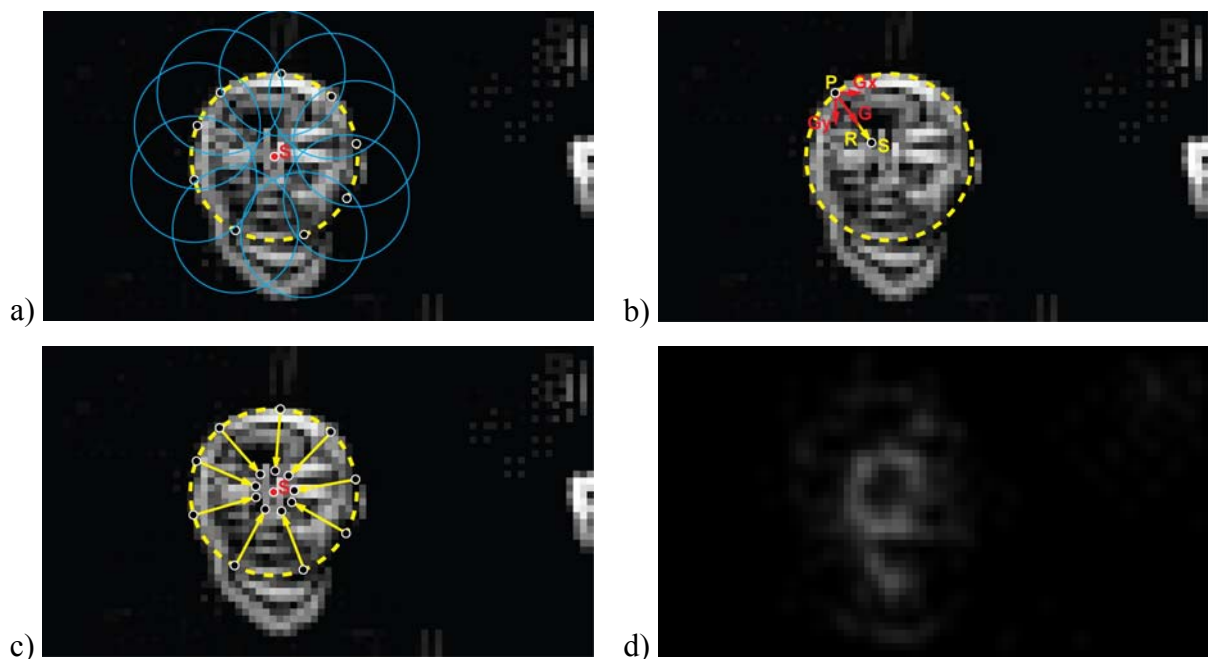
Pierwsze obrazy w rysunkach 2.25, 2.26 i 2.27 (oznaczone literką „a”) przedstawiają hipotetyczne okręgi zataczone przez promienie poszukiwań dla wybranych punktów znajdujących się na granicy twarzy. Kolejne (oznaczone „b”) ukazują w jaki sposób lokalne gradienty wskazują w którym kierunku skierowany ma zostać promień poszukiwań  $R$ .

Następnie (rysunki „c”) pokazują kilka zanalizowanych punktów z obszaru granicy twarzy i wyznaczone przez nie punkty  $S$ . Ostatnie obrazy (z literką „d”) to zobrazowanie MT, gdzie obowiązuje zasada – im jaśniejsze piksele tym wyższa wartość w macierzy, co znaczy, że tam najczęściej „trafiały” promienie poszukiwań.

Rysunek 2.25 ukazuje w bardzo widoczny sposób, dlaczego zbyt krótki promień poszukiwań nie ma szans na poprawne zlokalizowanie położenia twarzy. Wyznaczone przez niego punkty  $S$  rozciągają się z realnym położeniem środka twarzy, przez co nie mają szans na poprawną inkrementację MT, co bardzo trafnie przedstawia rys. 2.25d.

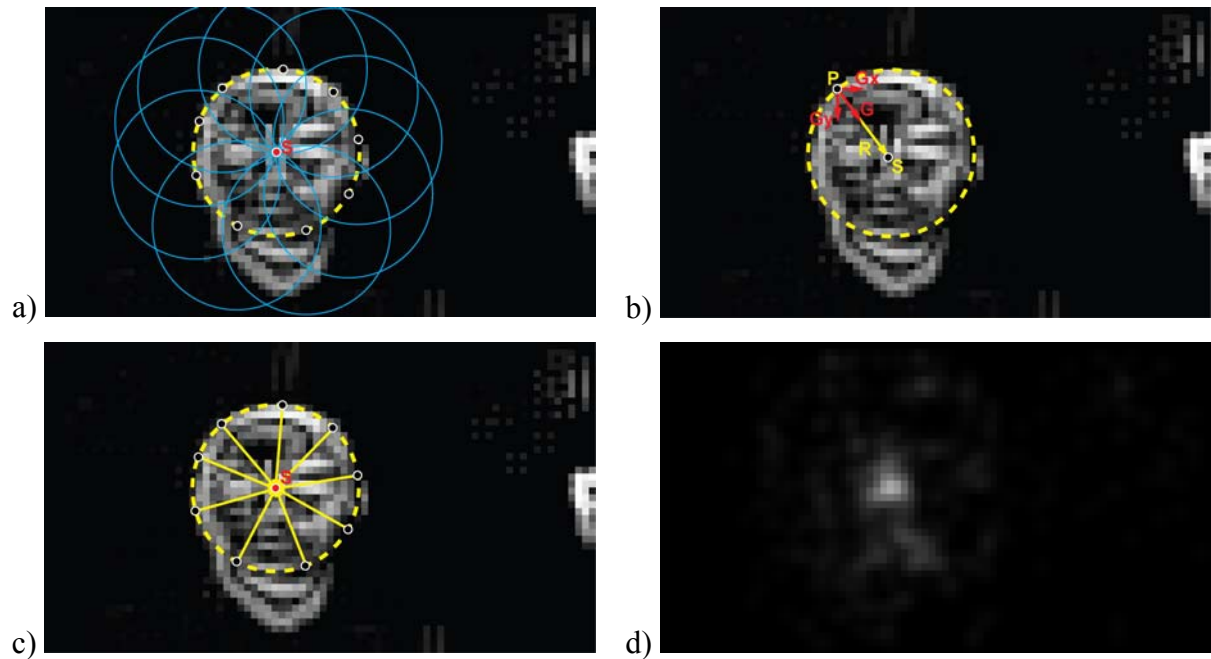
Oznaczenia (dla rysunków 2.25, 2.26, 2.27):

- $S$  - prawdopodobny środek twarzy,
- $R$  - spodziewany promień twarzy,
- $G_x$  - gradient poziomy,
- $G_y$  - gradient pionowy,
- $G$  - wektor gradientu.



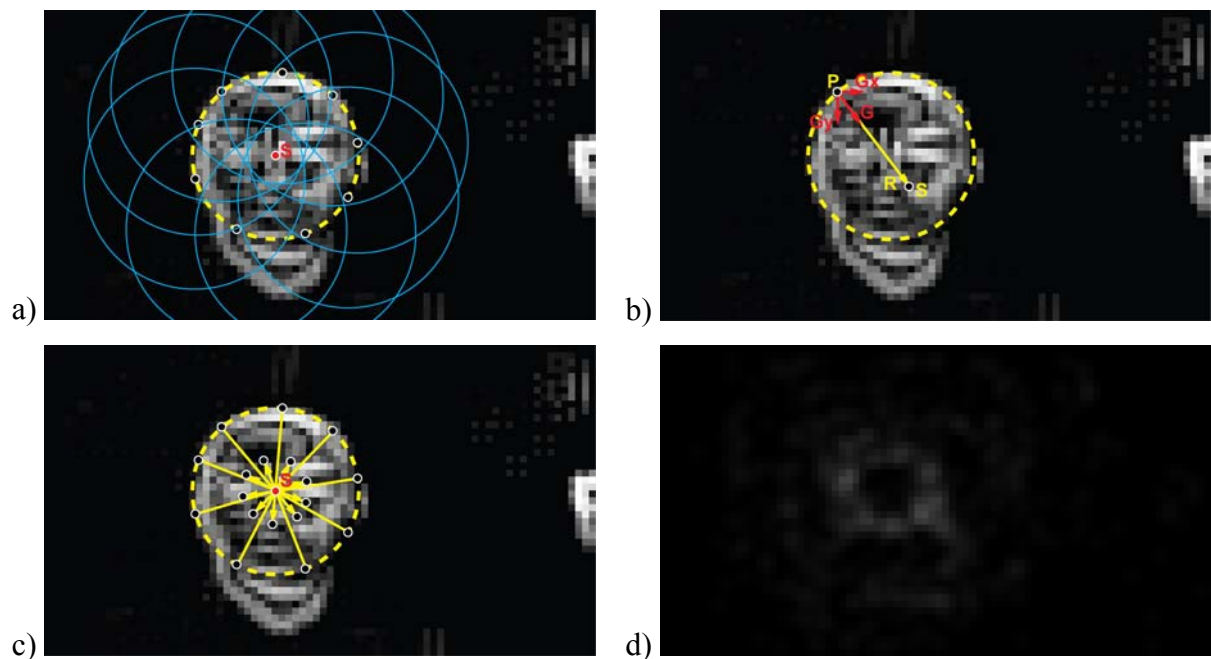
Rys. 2.25. Poszukiwanie promieniem o wielkości 7 pikseli: a) hipotetyczne okręgi poszukiwań, b) ilustracja wyznaczania poprawnego kierunku poszukiwań, c) zobrazowanie analizy kilku punktów, d) MT o kodzie 7

Na rys. 2.26 promień poszukiwań jest bardzo zbliżony do promienia twarzy przez co punkty  $S$  wyznaczone przez promienie i wartości gradientów są w okolicach środka twarzy. Powoduje to wzrost ilości inkrementacji wartości MT dla punktu środka twarzy. Rys. 2.25d ukazuje jak wygląda wynikowa MT o kodzie 12 (czyli dla promienia poszukiwań o wielkości 12 pikseli).



Rys. 2.26. Poszukiwanie promieniem o wielkości 12 pikseli: a) hipotetyczne okręgi poszukiwań, b) ilustracja wyznaczania poprawnego kierunku poszukiwań, c) zobrazowanie analizy kilku punktów, d) MT o kodzie 12

Przy zbyt długim promieniu poszukiwań (rys. 2.27) sytuacja przypomina tą, gdy promień jest za krótki. Znow inkrementowane są niepoprawne wartości w MT o kodzie 17, a przez to wynik jest znow rozbieżny z oczekiwanym (rys. 2.27d).



Rys. 2.27. Poszukiwanie promieniem o wielkości 17 pikseli: a) hipotetyczne okręgi poszukiwań, b) ilustracja wyznaczania poprawnego kierunku poszukiwań, c) zobrazowanie analizy kilku punktów, d) MT o kodzie 17

W celu dokładniejszego przyjrzenia się MT dla poszczególnych promieni poszukiwań poniżej przedstawiona jest tabela (2.3) obrazująca uzyskane wyniki.

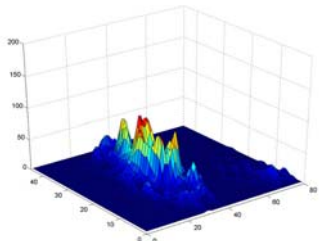
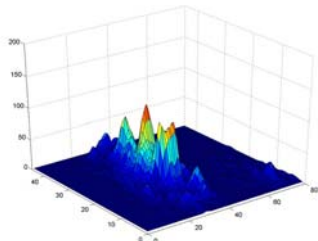
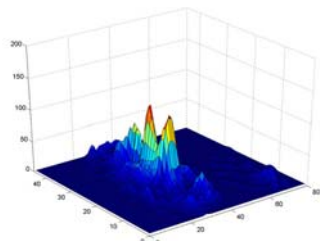
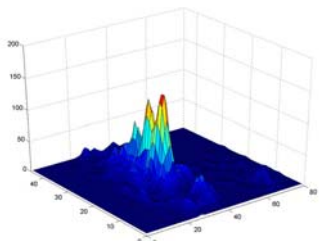
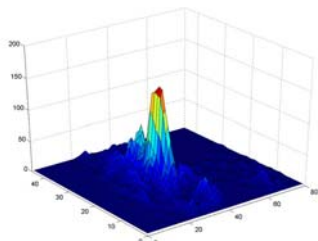
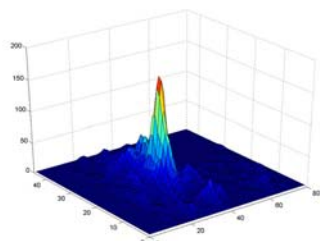
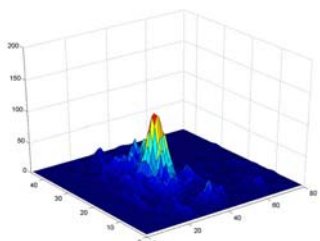
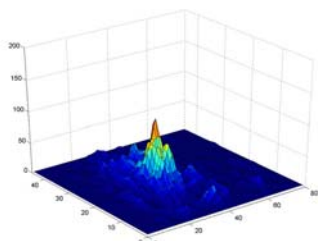
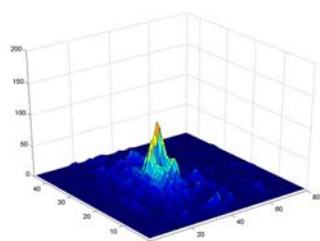
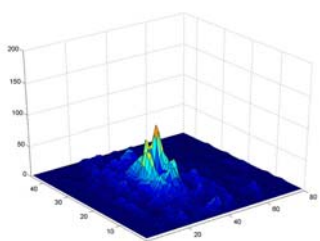
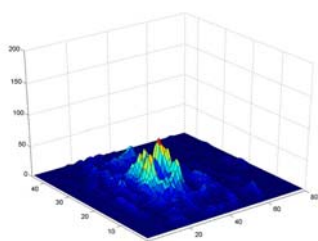
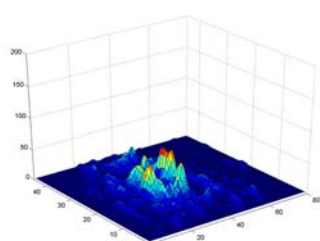
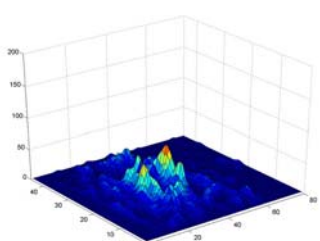
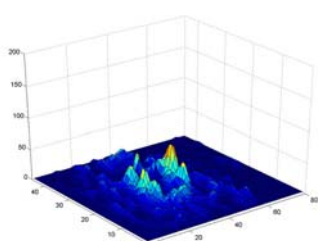
|                                                                                     |                                                                                     |                                                                                       |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|    |    |    |
| Obraz wejściowy                                                                     | MT 7                                                                                | MT 8                                                                                  |
|    |    |    |
| MT 9                                                                                | MT 10                                                                               | MT 11                                                                                 |
|  |  |  |
| MT 12                                                                               | MT 13                                                                               | MT 14                                                                                 |
|  |  |  |
| MT 15                                                                               | MT 16                                                                               | MT 17                                                                                 |
|  |  |  |
| MT 18                                                                               | MT 19                                                                               | MT 20                                                                                 |

Tabela 2.3. Ilustracja MT dla różnych promieni R i zaznaczonym najlepszym wynikiem

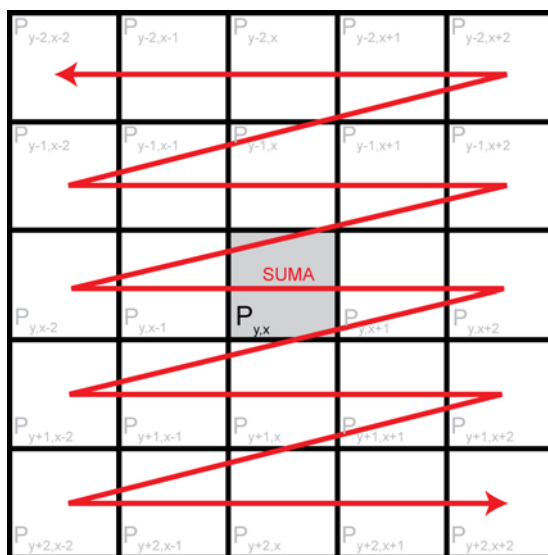
Wszystkie wykresy MT przedstawione w tabeli 2.3 posiadają tę samą skalę, co poprawia ich czytelność. Analizując te wyniki łatwo zauważyć, że w MT 12 (czyli dla tej uzyskanej przy użyciu promienia  $R=12$ ) znajduje się zdecydowanie najwyższa wartość ze wszystkich MT, a współrzędne punktu ją określającego są automatycznie wynikiem całego algorytmu. Określa bowiem środek lokalizowanej twarzy, a promień  $R$  (w tym przypadku o rozmiarze 12 pikseli) to promień twarzy.

Pomimo uzyskiwania podobnych, równie dobrych i precyzyjnych wyników jak w opisywanym powyżej przykładzie, podczas badań zdarzały się przypadki błędnych. Po dogłębnej analizie zauważono, że pomimo użycia zmodyfikowanych metod inkrementacji zilustrowanych rysunkami 2.22 i 2.23, a mających zapobiegać nieprecyzyjności danych wejściowych, uzyskiwane wyniki nie zawsze są zadowalające. Stwierdzono, że odchylenia te są wynikiem działania zewnętrznych czynników takich jak:

- szum obrazów wejściowych,
- oświetlenie kierunkowe,
- przesłonięcie części twarzy,
- możliwość wystąpienia podobnego koloru włosów i tła do koloru skóry
- zniekształcenie informacji poprzez nieproporcjonalne skalowanie,
- kształt głowy nie będący idealnym owalem (a po nieproporcjonalnym przeskalowaniu – nie będący idealnym okręgiem),
- założenie algorytmu, będące teoretycznym, ale zbyt idealnym rozważaniem na temat owalu twarzy,
- elementy zakłócające, w tym szczególnie: kształt włosów, okrycie głowy.

Aby przeciwdziałać powyższym czynnikom przetestowano możliwości dodatkowego rozmycia wyników MT (macierzy trafień), które zniwelują niedoskonałości obliczeń na realnych obrazach. Do tego celu posłużono się metodą w założeniu podobną do tej zastosowanej w zaawansowanym inkrementowaniu MT. Każdy piksel z MT jest modyfikowany w sposób przedstawiony na rysunku 2.28. Jest to prosta suma pikseli z otoczenia tego badanego (zastosowano obszar sumowania o wielkości 5x5 pikseli, choć przebadano również 3x3) wpisana następnie w jego miejsce. Stosowność użycia tej metody bazuje na przesłaniu, że w otoczeniu piksela z najwyższą wartością powinny znajdować się również inne piksele o wysokich wartościach. Istnieje bowiem prawdopodobieństwo, że jakiś przypadkowy piksel zostanie inkrementowany do najwyższej wartości w całej MT. Zastosowanie sumy okalających pikseli taką ewentualność w dużej mierze wyeliminuje.

Powyższe rozważania zostały przetestowane i udowodnione podczas badań efektywności całego systemu.



Rys. 2.28. Zobrazowanie idei aktualizowania wyników  
(macierzą o wymiarach 5x5 pikseli)

Rysunek 2.28 można przedstawić za pomocą wzoru::

$$P_{y,x} = \sum \left( \begin{aligned} &P_{y-2,x-2} + P_{y-2,x-1} + P_{y-2,x} + P_{y-2,x+1} + P_{y-2,x+2} + \\ &P_{y-1,x-2} + P_{y-1,x-1} + P_{y-1,x} + P_{y-1,x+1} + P_{y-1,x+2} + \\ &P_{y,x-2} + P_{y,x-1} + P_{y,x} + P_{y,x+1} + P_{y,x+2} + \\ &P_{y+1,x-2} + P_{y+1,x-1} + P_{y+1,x} + P_{y+1,x+1} + P_{y+1,x+2} + \\ &P_{y+2,x-2} + P_{y+2,x-1} + P_{y+2,x} + P_{y+2,x+1} + P_{y+2,x+2} \end{aligned} \right), \quad (2.35)$$

gdzie:

$P_{v,x}$  - aktualizowany punkt z „macierzy trafień”.

Powyższe rozwiązanie wpływa na charakter wartości przechowywanych w MT. Stają się one znacznie bardziej rozmyte w stosunku do „ostrych”, uzyskanych bezpośrednio z algorytmu poszukiwania obiektów owalnych z wykorzystaniem macierzy gradientów.

W celu zobrazowania wyników uzyskanych po opisywanym etapie aktualizacyjnym poniżej przedstawiona jest tabela zawierająca trzy kolumny z MT:

- bez aktualizacji,
- aktualizowaną przez macierz o rozmiarach  $3 \times 3$  piksele,
- aktualizowaną przez macierz o rozmiarach  $5 \times 5$  pikseli.

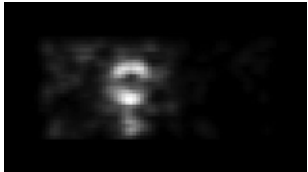
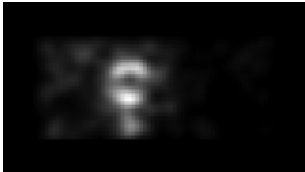
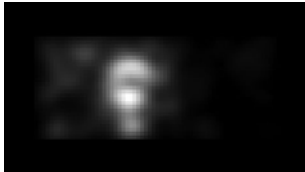
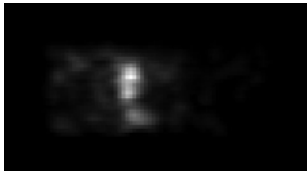
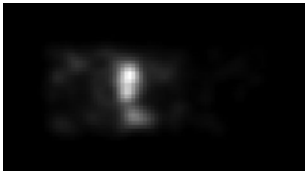
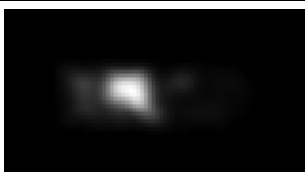
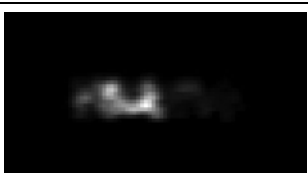
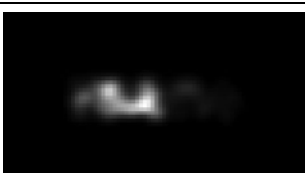
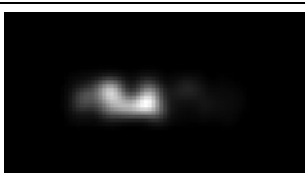
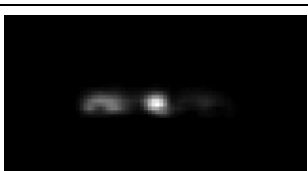
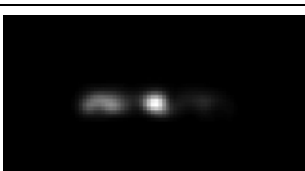
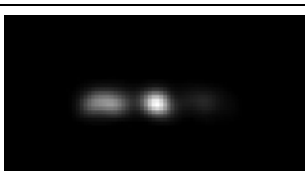
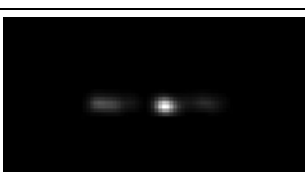
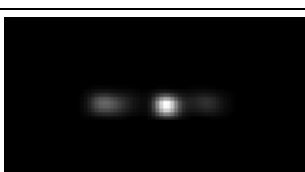
|                 |                                                                                     |                                                                                      |                                                                                       |
|-----------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Obraz wejściowy |   |                                                                                      |                                                                                       |
|                 | bez aktualizacji                                                                    | aktualizacja poprzez<br>sczytywanie macierzą<br>o rozmiarach $3 \times 3$ piks.      | aktualizacja poprzez<br>sczytywanie macierzą<br>o rozmiarach $5 \times 5$ piks.       |
| R=8             |    |    |    |
| R=10            |    |    |    |
| R=12            |   |   |   |
| R=14            |  |  |  |
| R=16            |  |  |  |
| R=18            |  |  |  |
| R=20            |  |  |  |

Tabela 2.4. Ilustracja macierzy trafień przed i po kroku aktualizacyjnym  
(ciemniejszym kolorem zaznaczono ostateczny wynik)

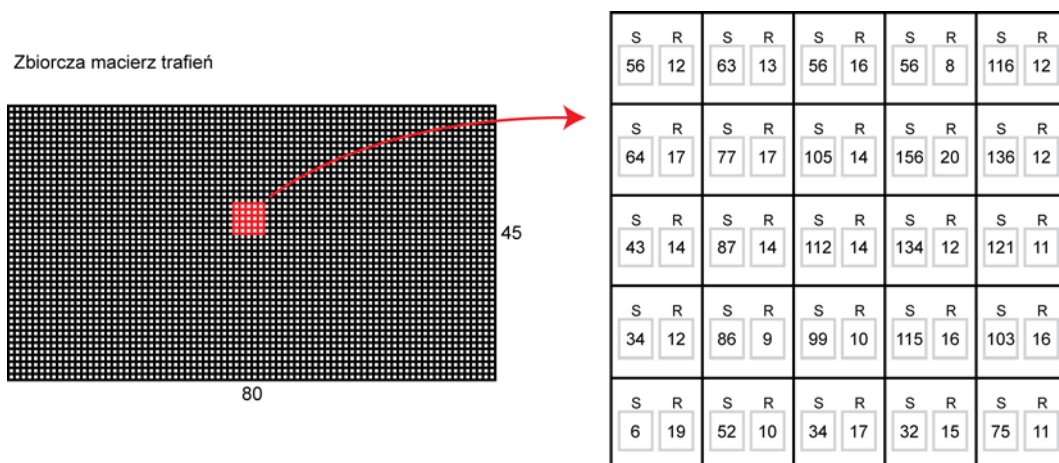
### 2.4.5. ZMT (Zbiorcza Macierz Trafień)

Całość rozważań przedstawionych w rozdziale 2.4 daje w wyniku położenie jednej twarzy, a w rzeczywistości pierwszego kandydata do bycia twarzą. W celu zlokalizowania kolejnych obszarów potencjalnie zawierających twarz zaimplementowano dodatkowy algorytm będący kontynuacją poprzednich rozwiązań. Bazuje on na przesłaniu, że warto zaprojektować wygodne narzędzie do poszukiwań kolejnych kandydatów, wykorzystując obliczone wcześniej MT.

Aby to osiągnąć stworzono ZMT (zbiorczą macierz trafień) w której każde pole przechowuje dwie zmienne:

- najwyższa wartość  $S$  ze wszystkich MT dla tych współrzędnych,
- identyfikator, z której MT jest powyższa liczba (czyli  $R$ ).

Taki zabieg eliminuje konieczność przeszukiwania wszystkich MT dla kolejnych twarzy (w implementacji w języku C++ wszystkie twarze, łącznie z pierwszą, poszukiwane są w ZMT). Wystarczy analizować po raz kolejny ZMT. To rozwiązanie zdecydowanie przyspiesza działanie programu, ponieważ kolejnego kandydata poszukuje się poprzez wyznaczenie najwyższej wartości w ZMT, co jest operacją niezwykle prostą. W ten sposób można wygenerować w szybkim czasie bardzo wiele kolejnych obszarów potencjalnie zawierające twarz. Rysunek 2.29 prezentuje przykład (wycinek) zaproponowanej powyżej ZMT.



Rys. 2.29. Ilustracja „zbiorczej macierzy trafień” przechowującej w każdej komórce wartości  $S$  i  $R$

Aby zaprezentować ideę ZMT, posłużono się tabelą 2.5 przedstawiającą analizę realnych obrazów, w skład której wchodzi:

- oryginalny obraz (po etapie nieproporcjonalnego skalowania),

- obraz w komponencie  $I2$  ze przestrzeni  $III2I3$  wraz z krokiem normalizacyjnym,
- ZMT (komponent  $S$ ) z uwagą, że dla poszczególnych MT z których była wyznaczana przeprowadzony był krok aktualizacyjny sczytywania macierzą o rozmiarach  $5 \times 5$  pikseli.

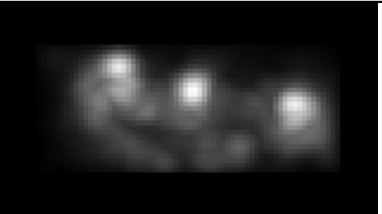
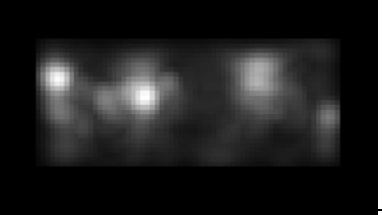
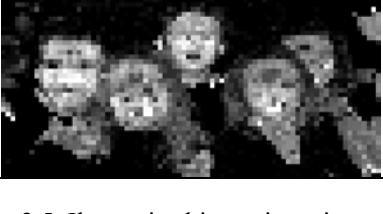
| Obraz oryginalny                                                                    | Obraz w komponencie I2                                                               | „zbiorcza macierz trafień”                                                            |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|    |    |    |
|    |    |    |
|   |   |   |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Tabela 2.5. Ilustracja zbiorczej macierzy trafień

W bardzo konkretny sposób można określić ilość iteracji występujących podczas wyznaczania obszarów twarzy. Ta cecha jest bardzo istotna przy projektowaniu systemów dla określania zasobów niezbędnych dla prawidłowego funkcjonowania.

Ilość iteracji wyznaczamy przy pomocy wzoru:

$$\text{Metoda opracowana: } Q = G_N \sum_{R=M/5}^{M/2} (M - 2R)(N - 2R), \quad (2)$$

gdzie:

$G_N$  – ilość zestawów macierzy gradientów ( $G_N = 1$  lub  $G_N = 2$ );

$M, N, R$  – wysokość, szerokość analizowanego obrazu i promień poszukiwań.

Przykładowo dla  $M=45, N=80, L=2$  wartość  $Q=21028$ .

#### 2.4.6. Wyznaczanie pięciu obszarów zawierających twarz

Mając do dyspozycji obliczoną ZMT wyznacza się kolejno następne obszary, które są wyjściem całego systemu. W niniejszej rozprawie zdecydowano się wyznaczać jedynie pięć propozycji obszarów potencjalnie zawierających twarz. Takie założenie, przy jednoczesnym wysokim współczynniku zawierania się w tej grupie twarzy (powyżej 95%) i bardzo krótkim czasie wykonania (poniżej 20ms) pozwala użyć przedstawiony system jako pierwszy krok w detektorach już opracowanych, jak również posłużyć do budowy nowych.

Realizacja wyznaczenia pierwszych pięciu obszarów potencjalnie zawierających twarz jest operacją niezwykle prostą. Po wyszukaniu pierwszego, opisanego w całym punkcie 2.4, algorytm wyklucza z dalszej analizy piksel o współrzędnych równym środku wyznaczonego (bo on jest wynikiem działania – obszar jest wynikiem przeliczenia  $S$  i  $R$ ). Do dyspozycji jest już wyliczona ZMT, a przez to odpada konieczność wykonywania tych samych, skomplikowanych obliczeń mających na celu ponowne wyliczenie MT. Operacja wyznaczenia kolejnego obszaru polega więc na prostym wyszukaniu najwyższej wartości w ZMT (z wyłączeniem tej wyznaczonej poprzednio). Przy jej pomocy określany jest obszar będący drugim kandydatem. Znowu wyłącza się ten piksel z dalszych obliczeń (więc są już dwa wyłączone) i operację powtarza się 5-krotnie. Uzyskane w ten sposób położenie pięciu obszarów potencjalnie zawierających twarz jest przekazywane na wyjście systemu

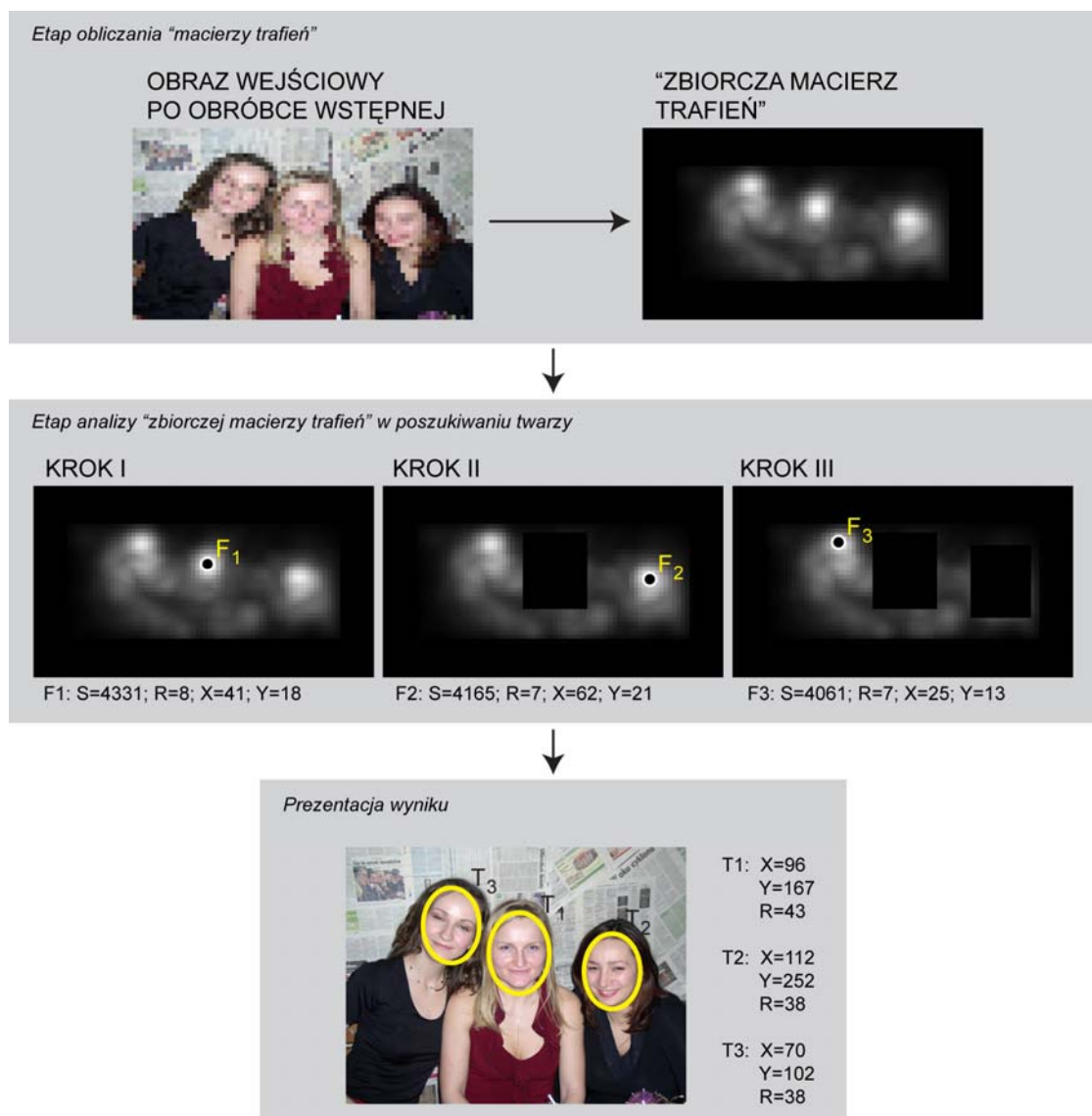
#### 2.4.7. Wyznaczanie wielu twarzy

Całe powyższe rozważania opisane w rozdziale 2.4 są poświęcone poszukiwaniu pięciu obszarów potencjalnie zawierających twarz. Nic nie stoi jednak na przeszkodzie, aby

po zweryfikowaniu jednego z nich jako twarz (lub wykluczeniu jej występowania) wyznaczyć kolejnych pięciu kandydatów do poszukiwania kolejnej twarzy.

Po zweryfikowaniu obszaru zawierającego pierwszą twarz najbezpieczniej jest go całkowicie wykluczyć z dalszych obliczeń. Takie rozwiązanie, poza oczywistym nie dublowaniem tego samego obszaru (jedynie minimalnie przesuniętego), daje możliwość poszukiwania środka kolejnej twarzy już na granicy poprzednio znalezionej co powoduje, że możliwe jest znalezienie twarzy przesłoniętej inną twarzą. Ta cecha jest bardzo unikalna szczególnie w metodach bazujących na segregacji kolorów.

Rysunek 2.30 przedstawia schemat postępowania przy poszukiwaniu trzech twarzy.



Rys. 2.30. Schemat postępowania w celu zlokalizowania twarzy na obrazie

Pierwszy etap nazwany „Etapem obliczania MT” jest bardzo skrótowym pokazaniem wszystkich etapów opisanych w rozdziale 2.4. W kolejnym etapie – „Etap analizy ZMT w

celu zlokalizowania obszaru twarzy” – uwidocznione zostało w jaki sposób poszukiwana jest pierwsza jak i kolejne twarze. Po wykryciu najwyższej wartości w ZMT zapamiętywana jest wartość  $X$ ,  $Y$  oraz  $R$ , a obszar, który pokrywa zlokalizowana twarz jest maskowany, czyli zerowane są wszystkie wartości w jego obrębie co gwarantuje, że w kolejnym kroku, następna twarz będzie poszukiwana w innym miejscu.

Ostatnim etapem jest „Prezentacja wyników”. Niezbędnym jest przeliczenie wartości uzyskanych z algorytmów do skali obrazu wejściowego, gdyż na samym początku działania jest przeprowadzone nieproporcjonalne skalowanie, a przez to uzyskane wyniki są w innej skali.

Proces ten realizują następujące wzory:

$$\begin{cases} y_w = y_i * \frac{Y_w}{Y_i} \\ x_w = x_i * \frac{X_w}{X_i} \end{cases} \quad (2.36)$$

gdzie:

- $y_i, x_i$  - współrzędne  $y$  i  $x$  środka twarzy uzyskane bezpośrednio z algorytmu (w proporcjach nieproporcjonalnie przeskalowanego obrazu),
- $y_w, x_w$  - współrzędne  $y$  i  $x$  środka twarzy przeliczone do proporcji obrazu wejściowego,
- $Y_i, X_i$  - rozdzielczość obrazu po przeskalowaniu (45×80 pikseli),
- $Y_w, X_w$  - rozdzielczość obrazu wejściowego.

W tym miejscu algorytm wykrywania obiektów owalnych z wykorzystaniem macierzy gradientów w celu lokalizacji wielu twarzy kończy swoją pracę. Dla obrazu wejściowego zwracana jest macierz zawierająca położenie i wielkość twarzy.

#### 2.4.8. Wersja dla obrazów w skali szarości

Opracowano również wersję algorytmu dla obrazów w skali szarości. Potrafi ona automatycznie wykrywać czy obraz jest pozbawiony informacji o kolorze i przełączać swój tryb pracy na odpowiednią wersję. Zdecydowana część algorytmu pozostaje bez zmian. Nie można zastosować komponentu  $I2$  z przestrzeni  $III2I3$ , więc do obliczenia macierzy gradientów, podobnie jak w bazowej metodzie, wykorzystany jest bezpośrednio obraz czarnobiały. Przykładowe wizualizacje wyników obliczania tych macierzy zawiera tabela 2.1.

W przypadku obrazów czarnobiałych konieczne jest podwójne obliczanie trafień algorytmu wykrywania obiektów w owalnych. Dualizm bierze się z faktu, że granica twarzy może być zarówno przy przejściu z jasnego tła do ciemnego koloru skóry twarzy (np.

opalonej), bądź odwrotnie – z ciemnego tła do jasnej skóry. Nie można zawyrokować, który wariant wystąpi, więc oblicza się oba, a wyniki wpisywane są do tych samych MT. Uwidacznia się tu kolejna zaleta stosowania komponentu  $I_2$ , który pozwala założyć, że kolor skóry twarzy po konwersji jest jaśniejszy od tła (co ukazuje tabela 2.1) – ta cecha pozwala zaoszczędzić znaczną ilość czasu, ponieważ obliczanie MT jest jednym z głównych zadań obciążających procesor w całej metodzie wykrywania owalów.

## 2.5. Podsumowanie i wnioski

W rozdziale 2 opisano metodę lokalizacji i realizujące ją algorytmy. Na wstępie omówiono algorytmy obróbki wstępnej, jak również sposób projektowania detektorów. Następnie opisano główny algorytm lokalizacji w którym posłużono się ideą według której jednocześnie uwzględnienie informacji o owalnym kształcie twarzy i procedury zmiennego promienia, może być podstawą do budowy skutecznej metody lokalizacji twarzy na obrazach cyfrowych, a dodatkowo użycie informacji o kolorze skóry zwiększy jej skuteczność.

Aby powyższą ideę zrealizować sięgnięto po metodę spełniającą minimalne postawione kryteria, którą następnie zmodyfikowano w celu poprawienia wyników działania. Każde proponowane usprawnienie zostało poprzedzone uważną analizą jego potencjalnego wpływu na efektywność, a wnioski zawarto w opisie.

W opracowanej metodzie można wyróżnić następujące elementy nowości w stosunku do metody bazowej:

1. Integracja w jednej metodzie informacji o owalnym kształcie twarzy, jej nieprzewidywalnej wielkości i przedstawieniu w polu gradientów w czterech niezależnych kierunkach (dwie pary: pion/poziom i ukośne);
2. Zastosowanie komponentu  $I_2$  z przestrzeni  $I1I2I3$  wraz z procedurami normalizacji, jako głównego elementu obróbki obrazów kolorowych dla różnych ras ludzkich;
3. Propozycja zbiorczej macierzy trafień do zadania poszukiwania kolejnych twarzy (do 10), bez powtórnego wyliczania poszczególnych macierzy trafień;
4. Modyfikacja wyznaczania centrum twarzy za pomocą procedury wygładzania, eliminująca niedokładność formy owalów w macierzy gradientów;

Opracowana metoda jest przygotowana do przeprowadzenia na niej eksperymentalnych badań, mających na celu udowodnienie poprawności użytych algorytmów i ich modyfikacji, jak również potwierdzić zasadność postawionej tezy.

### 3. Eksperymentalne badania opracowanej metody

W niniejszym rozdziale zostały przedstawione eksperymentalne badania opracowanej metody lokalizacji twarzy. Całą metodę poddano ściśle opisanej procedurze testowej mającej na celu kompleksowe jej zbadanie. Jednocześnie każda użyta modyfikacja została zweryfikowana pod kątem pozytywnego wpływu na ostateczny wynik. Dodatkowym efektem eksperymentalnych badań metody jest wyznaczenie optymalnych parametrów pracy dla każdego z algorytmów tego wymagających.

#### 3.1. Podstawowe założenia

Każdy system pretendujący do bycia dobrym detektorem, bądź lokalizatorem musi, poza uzasadnieniem trafności użycia akurat tych, a nie innych algorytmów, przejść procedurę testową ukazującą jego skuteczność, bądź jej brak. Wskazane jest też, by uzyskiwane wyniki były wyprzedzające rozwiązania już dostępne na rynku detektorów. Należy jednak zauważyć, że nie zawsze jedynym i decydującym czynnikiem czy badany system jest innowacyjny musi być ostateczna efektywność w dziedzinie współczynnika detekcji. Często ważniejszy jest współczynnik błędnych przyjęć, szybkość lub współczynnik jakość/szybkość.

Dlatego zdefiniowano następujące założenia dotyczące badań:

- użyte zostaną dwie bazy danych: komercyjna (FERET) i autorska,
- współczynnik poprawnego wskazania obszaru na którym występuje twarz w pięciu próbkach będzie większy od 95% (dla każdej bazy danych),
- każdy parametr będzie przetestowany w zakresie umożliwiającym ustalenie tendencji wzrostu bądź spadku efektywności całego systemu,
- stworzona zostanie dokumentacja całego procesu.

W celu określenia skuteczności algorytmów użytych w budowie systemu lokalizatora obszarów potencjalnie zawierających twarz skorzystano z dwóch baz zdjęć. Pierwsza z nich to komercyjna baza FERET opracowana w Stanach Zjednoczonych, a druga – autorska. Baza FERET jest zdecydowanie najbardziej promowaną bazą zdjęć do testowania systemów rozpoznawania ludzi, ulega ciągłej rozbudowie, a najnowsze rozwiązania detekcji i rozpoznawania w skali światowej nie mogą się obejść bez testów na niej przeprowadzonych. Niemniej zauważono, że jest ona przewidywalna i stanowi stosunkowo łatwy materiał do poszukiwania twarzy ponieważ tło jest przeważnie jednolite, rozmydlone, twarze dobrze doświetlone, a ich wielkość znormalizowana. Można więc stwierdzić, że bardziej nadaje się do celów rozpoznawania niż detekcji. Dlatego mimo testów na FERET zdecydowano się

stworzyć drugą bazę, będącą lepszą odpowiedzią na założenia dotyczące wymagań materiału testowego, ponieważ zróżnicowanie w oświetleniu, tle, wielkości twarzy jest znacznie bardziej kompleksowe.

Podczas analizy literatury nie stwierdzono występowania jako przedmiotu badania problemu, który został sprecyzowany w tej pracy. Dlatego ciężko porównywać wyniki płynące z analizy opracowanego systemu z innymi rozwiązaniami. Niemniej najczęściej ilość próbek dostarczana przez lokalizator obszarów potencjalnie zawierających twarz do klasyfikatora twarz/„nie twarz” wynosiła kilkaset sztuk, a i tak potrafiła nie zawierać poszukiwanej twarzy. W niniejszej pracy założono, że już w pierwszych pięciu obrazach wydzielonych przez algorytm będzie twarz, a współczynnik poprawności spełnienia tego założenia wyniesie więcej niż 95%.

W założeniach każdy algorytm opracowany w niniejszej rozprawie powinien być dostatecznie przetestowany pod kątem zasadności jego użycia. Dodatkowo niemal wszystkie algorytmy użyte ostatecznie w badanym lokalizatorze zawierają parametry, których regulacja pozwala znacznie zwiększyć lub ograniczyć efektywność całego systemu. Dlatego podczas testowania duży nacisk położono na kompleksowe zbadanie efektywności przed i po wprowadzeniu usprawnień, jak i parametrów już wybranych metod.

Po dogłębnej analizie algorytmów pod względem doboru optymalnej ścieżki testowej zdecydowano się na przeprowadzenie testu obejmującego generowanie wyników przy zmianie wszystkich parametrów względem siebie. Aby tego dokonać powstała specjalna wersja programu Face Localizator.

Wszelkie testy zostały przeprowadzone na komputerze przenośnym klasy PC o parametrach:

- procesor: Athlon Mobile XP 2800+ 1450MHz
- pamięć RAM: 512MB
- karta grafiki: ATI Radeon Mobile 9600
- system operacyjny: Microsoft Windows XP
- platforma programistyczna: Borland Builder 5

Na potrzeby dokumentacji jak i archiwizacji danych stworzono procedurę umożliwiającą przeanalizowanie całego procesu i możliwość odtworzenia przeprowadzonych badań. Dla obu testowych baz procedura testowa wygląda w ten sam sposób, co dodatkowo pozwala na opracowanie wniosków dotyczących wpływu charakteru testowej bazy danych na zmianę zachowania się algorytmów przy różnych parametrach.

### 3.2. Procedura testowa

Poniżej przedstawiona jest procedura według której przeprowadzony został proces testu algorytmu.

#### 1. CEL DOKUMENTU

Stworzenie bazy danych będącej materiałem do badań algorytmów wchodzących w skład lokalizatora obszarów zawierających twarz opracowanych na potrzeby rozprawy doktorskiej.

#### 2. WEJŚCIE PROCESU

Baza zdjęć testowych zawierających twarze o określonych parametrach technicznych i zestaw parametrów przy pomocy których algorytm ma być badany.

#### 3. WYJŚCIE PROCESU – CEL PROCESU

Zestaw plików tekstowych zawierających:

- wyniki działania algorytmów,
- wyniki przeprowadzonych przy ich pomocy obliczeń.

#### 4. DEFINICJE

- Program testowy – specjalna wersja programu *Face Localizator* dostosowana do testowania algorytmu w sposób zautomatyzowany. Na jego wyjściu generowane są pliki tekstowe zawierające informacje dla jakich parametrów wejściowych aktualny test jest przeprowadzany, jak również wynik działania algorytmu dla każdej twarzy wg schematu:

```
0 1 1 120 0 100 0
```

```
99 151 27 91 147 35 94 147 35 88 149 38 91 149 38 70
```

```
102 153 35 99 153 35 99 155 32 102 155 32 104 153 35 70
```

```
99 184 35 99 182 35 91 182 27 96 184 32 88 182 24 60
```

...

gdzie:

linia pierwsza (parametry):

|     |                                                                                                                 |
|-----|-----------------------------------------------------------------------------------------------------------------|
| 0   | macierze gradientów 45/135 włączone/wyłączone                                                                   |
| 1   | wartość odcięcia gradientu                                                                                      |
| 1   | wielkość obszaru czytywanego dla każdego piksela                                                                |
| 120 | rozdzielczość obrazka (jego wartość w płaszczyźnie poziomej po przeskalowaniu proporcjonalnym) poddana analizie |
| 0   | wartość odcięcia dolnego histogramu                                                                             |
| 100 | wartość odcięcia górnego histogramu                                                                             |
| 0   | typ uaktualniania wyników                                                                                       |

Linie od drugiej do ilości zdjęć poddanych analizie (poniżej przykładu dla linii nr 2):

|     |                                                        |
|-----|--------------------------------------------------------|
| 99  | środek kandydata nr 1 – wartość w pł. Y (w pikselach)  |
| 151 | środek kandydata nr 1 – wartość w pł. X (w pikselach)  |
| 27  | promień kandydata nr 1 (w pikselach)                   |
| 91  | środek kandydata nr 2 – wartość w pł. Y (w pikselach)  |
| 147 | środek kandydata nr 2 – wartość w pł. X (w pikselach)  |
| 35  | promień kandydata nr 2 (w pikselach)                   |
| 94  | środek kandydata nr 3 – wartość w pł. Y (w pikselach)  |
| 147 | środek kandydata nr 3 – wartość w pł. X (w pikselach)  |
| 35  | promień kandydata nr 3 (w pikselach)                   |
| 88  | środek kandydata nr 4 – wartość w pł. Y (w pikselach)  |
| 149 | środek kandydata nr 4 – wartość w pł. X (w pikselach)  |
| 38  | promień kandydata nr 4 (w pikselach)                   |
| 91  | środek kandydata nr 5 – wartość w pł. Y (w pikselach)  |
| 149 | środek kandydata nr 5 – wartość w pł. X (w pikselach)  |
| 38  | promień kandydata nr 5 (w pikselach)                   |
| 70  | czas wygenerowania pięciu powyższych kandydatów (w ms) |

- Baza danych zdjęciowych – nazywamy zestaw zdjęć zawierających twarze, których wielkość jest znormalizowana i wynosi 240×320 pikseli (dla bazy autorskiej) lub 256×384 pikseli (dla bazy FERET), przestrzeń kolorów – 8 lub 24 bity. Dodatkowo nazewnictwo plików graficznych jest uporządkowane w następujący sposób: *xxxx.jpg*, gdzie *xxxx* – kolejny numer porządkowy.
- Zestaw wzorcowych wyników – plik tekstowy zawierający wzorcowe wyniki lokalizacji dla wejściowej bazy zdjęć, uporządkowane w sposób:

99 151 27  
 102 153 35  
 99 184 35  
 ...

gdzie wszystkie linie poniższej tabeli zawierają informacje (przykład dla linii nr 1):

|     |                                                      |
|-----|------------------------------------------------------|
| 99  | realny środek twarzy – wartość w pł. Y (w pikselach) |
| 151 | realny środek twarzy – wartość w pł. X (w pikselach) |
| 27  | realny promień twarzy (w pikselach)                  |

## 5. OPIS PROCESU

- Przygotowanie bazy danych zdjęciowych – normalizacja parametrów technicznych (przebieg kolorów, rozmiar w pikselach, rodzaj formatu zapisu) i zadanie odpowiedniego nazewnictwa.
- Opracowanie zestawu wzorcowych wyników – dla baz komercyjnych jest on przeważnie dostarczony wraz z całą bazą. Bazy autorskie wymagają ręcznego

określenia każdej twarzy za pomocą trzech liczb opisanych w wyjaśnieniu „Zestaw wzorcowych wyników”.

- Opracowanie zestawu parametrów – odpowiedni ich dobór ma decydujący wpływ na kompletność całego procesu testowego.
- Uruchomienie programu testowego – do programu należy dostarczyć (wskazać) katalog zawierający bazę danych zdjęciowych i zestaw parametrów dla których będzie przeprowadzone badanie.
- Analiza wyników – dane otrzymane z programu testowego (pliki tekstowe, których zawartość jest opisana w definicjach przy okazji omawiania „programu testowego”) zostają poddane obróbce i analizie przy pomocy zestawu wzorcowych wyników.
- Wypełnienie sprawozdania z przebiegu badania – do tego celu przygotowane są odpowiednie załączniki
- Archiwizacja wyników – wszystkie otrzymane wyniki (pliki tekstowe) zostają zapisane na płytach CD (z odpowiednim opisem) w celu umożliwienia powrotu do nich bez konieczności ponownego uruchamiania procesu testowego.

## 6. ZAŁĄCZNIKI

- Sprawozdanie z badań - załącznik ten służy ewidencji i wraz z archiwum jest podstawowym dowodem potwierdzającym przeprowadzenie testów i badań algorytmów opracowywanych na potrzeby rozprawy doktorskiej

### 3.3. Badania na bazie FERET i autorskiej

Każdy opracowywany system działający w szerokiej dziedzinie systemów biometrycznych, tym również ten operujący obrazem twarzy wymaga przetestowania pod kątem jego efektywności. Należy również wskazać jego zalety na tle już istniejących rozwiązań. Dlatego opracowano bazy danych zdjęć oferujące zestaw obrazów zawierających twarz, przy pomocy których każdy zespół badawczy ma możliwość podzielenia się rezultatami swej pracy z całym środowiskiem. Bazy danych zdjęć zostały omówione w rozdziale 1.3. Najatrakcyjniejszą spośród nich pod względem przydatności do testów detektorów, czyli również generatorów kandydatów jest FERET. Zdecydowano się również na testy przy użyciu bazy autorskiej.

Krótką charakterystykę i kilka przykładów z bazy FERET zawarto w rozdziale 1.3. Jednak jest to zbyt ograniczony opis. Dlatego w tabeli 3.2 przedstawiono reprezentatywną grupę obrazów z badanej kolorowej bazy FERET, pozwalającą zrozumieć jej charakter.



Tabela 3.1. Reprezentatywna podgrupa z bazy FERET wykorzystanej do badania algorytmu wyodrębniania twarzy.

Baza kolorowa FERET ma następującą charakterystykę:

- obrazy wchodzące w jej skład mają orientację pionową,
- występują wszystkie znaczące rasy ludzkie – biali, czarnoskórzy, azjaci, itd.
- twarz w centralnej części obrazu,
- postaci fotografowane w kadrze portretowym lub popiersie (czyli wielkość ich średnicy waha się od 1/4 do 2/3 wysokości obrazu),
- każda z postaci reprezentowana jest przy pomocy 2 do 6 obrazów,
- występują niewielkie odchylenia twarzy w różnych płaszczyznach (do 10%),
- zmienna mimika – śmiech, smutek, zdziwienie itd.,
- oświetlenie kierunkowe lub rozproszone,
- występują zakłócenia na twarzy – okulary, zarost itd.,
- tło nieskomplikowane, jasne, bez znaczących detali.

Ponieważ w trakcie konstruowania metody bardzo często korzystano ze zdjęć, które nie wchodziły w skład żadnej komercyjnej, bądź badawczej bazy danych zdjęciowych (zabieg ten miał na celu badanie reakcji algorytmów na przypadkowo dostarczone zdjęcia z różnych źródeł) zdecydowano się przetestować metodę przy użyciu autorskiej bazy danych. Stwierdzono, że wyniki tych obserwacji warto przeanalizować w sposób podobny jak dla bazy FERET. Dodatkowym argumentem przemawiającym za tym rozwiązaniem jest pewnego rodzaju schematyczność bazy FERET i jej ukierunkowanie na badanie całych systemów rozpoznawania twarzy, a nie tylko detektorów. Dlatego większość twarzy z jej zasobów ma znormalizowaną wielkość, tło nie jest specjalnie skomplikowane, a jako źródło pozyskiwania danych służy aparat cyfrowy o dobrej jakości. Jednocześnie stwierdzono, że warto zweryfikować zaskakująco dobre wyniki dla bazy FERET (być może wynikające właśnie z jej przychylnych warunków).

W celu osiągnięcia maksymalnie zróżnicowanych danych badawczych zdecydowano się zgromadzić serię 207 zdjęć pozyskanych (reprezentatywny obraz ich charakteru zawiera tabela 3.2) z przypadkowych źródeł charakteryzujących się bardzo różną jakością techniczną, jak i występującymi warunkami fotograficznymi. Zmienność zawartą w tej bazie można podzielić ze względu na:

- źródło pochodzenia: aparat cyfrowy (różne modele), kamera internetowa (różne modele), obraz z telewizji poddany digitalizacji,
- jakość techniczną: balans bieli, ostrość, zaszumienie, ekspozycja,
- oświetlenie: dzienne, sztuczne, lampą błyskową, rozproszone, kierunkowe,
- wielkość i umiejscowienie twarzy.

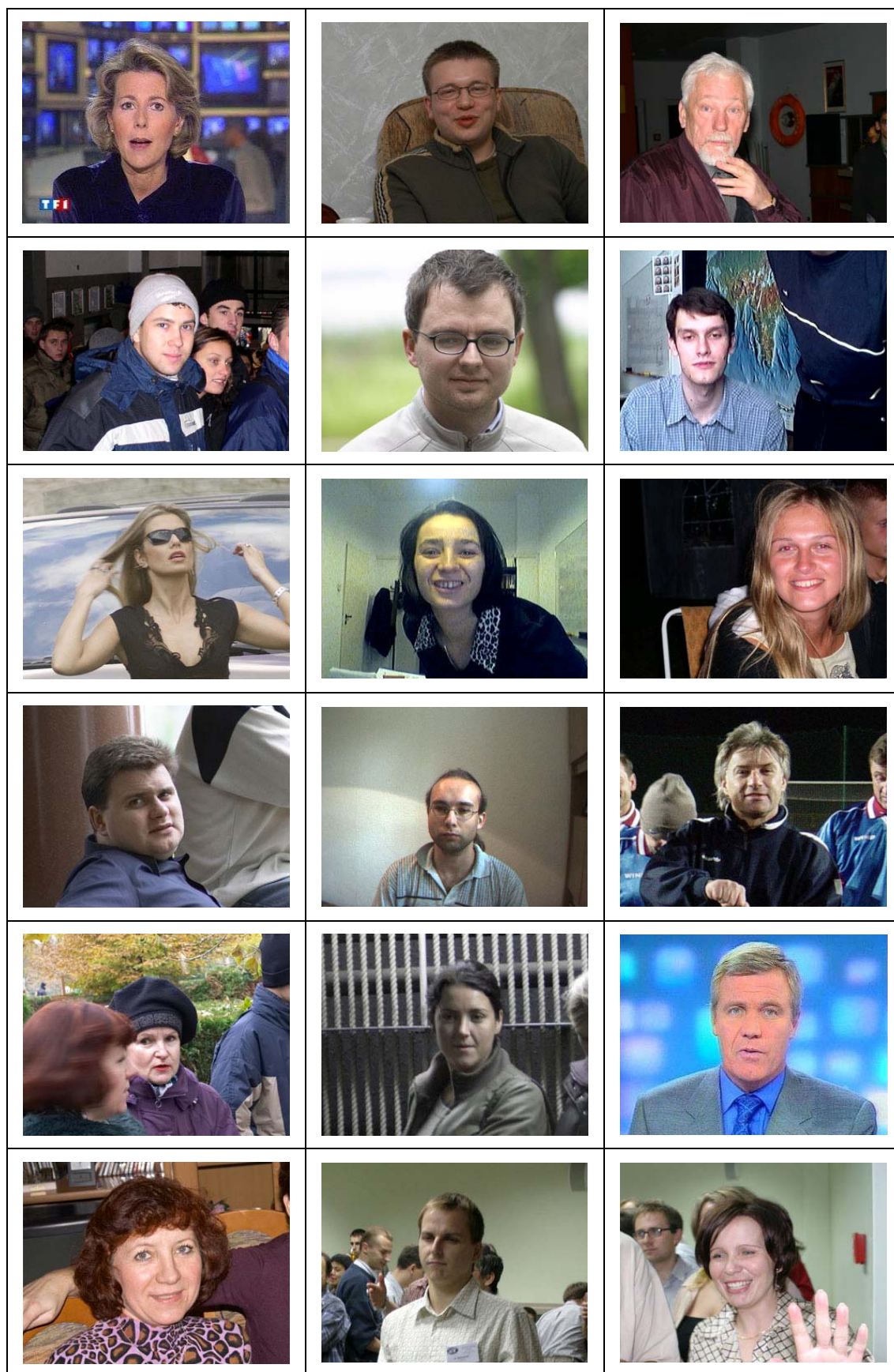


Tabela 3.2. Reprezentatywna podgrupa z bazy autorskiej wykorzystanej do badania algorytmów wyodrębniania twarzy.

Największy wpływ na jakość zdjęć ma źródło pochodzenia. Zdecydowanie najgorsze dane pozyskiwane są z kamer internetowych, lepsze oferują digitizery sygnału telewizyjnego, a zdecydowanie najlepsze pochodzą z aparatów cyfrowych (w szczególności z tych nowej generacji). Należy jednak założyć, że detektory, jak i lokalizatory działające na ich potrzeby, muszą wykazywać dużą efektywność już dla najsłabszego źródła, czyli kamer internetowych, ponieważ to one są głównym dostarczycielem danych do przetwarzania, szczególnie w systemach kontroli dostępu, czy innych działających w trybie online.

Używając wiedzy o działaniu algorytmów opisywanych w tej pracy ustalono zestaw parametrów przy pomocy których generowano wyniki służące dalszej analizie. Poniższa tabela (3.3) ukazuje parametry badania bazy FERET i autorskiej. Należy zaznaczyć, że są to jedynie parametry – wyniki, analiza i przemyślenia zawarte są w dalszej części rozdziału.

| Nazwa bazy zdjęć                   | FERET / AUTORSKA          |                        |
|------------------------------------|---------------------------|------------------------|
| Ilość zdjęć                        | 215 / 207                 |                        |
| Rozdzielczość                      | 256×384 / 240×320 pikseli |                        |
| Przestrzeń barw                    | 24 bity (RGB)             | 8 bitów (Grayscale)    |
| Macierze gradientów 45/135         | 0 1                       | 0 1                    |
| Wartość odcięcia gradientu         | 1 3 5 7                   | 1 3 5 7                |
| Wielkość obszaru sczytywanego      | 1 3 5                     | 1 3 5                  |
| Rozdzielczość badawcza             | 40 80 120 / 80 120 160    | 40 80 120 / 80 120 160 |
| Wartość odcięcia dolnego histogr.  | 0                         | 0 45                   |
| Wartość odcięcia górnego histogr.  | 50 70 90 255              | 165 210 255            |
| Typ uaktualniania wyników          | 0 1 2                     | 0 1 2                  |
| Ilość kombinacji powyższych param. | 864                       | 1296                   |
| Ilość wyodrębnień kandydatów       | 185760 / 178848           | 268272 / 278640        |

Tabela 3.3. Parametry badania bazy FERET i AUTORSKIEJ.

W tabeli zawarto pozycje i liczby do nich przypisane, które wymagają bliższego omówienia (pełna informacja zawarta jest w rozdziale 2.4):

- Przestrzeń barw – testy przeprowadzono dla obrazów kolorowych (RGB) i dla skali szarości (Grayscale) z założeniem, że są to te same zdjęcia,
- Macierze gradientów 45/135 – możliwość włączenia (1) lub wyłączenia (0) przetwarzania w dwóch dodatkowych ukośnych płaszczyznach,
- Wartość odcięcia gradientu – określa dla jakich minimalnych wartości gradientu bieżący piksel będzie przez algorytm analizowany,
- Wielkość obszaru sczytywanego – wielkość otoczenia sczytywanego dla każdego badanego piksela, np. 3 → maska 3×3 pikseli,

- Rozdzielczość badawcza – określa wielkość (jej składową poziomą) do której jest skalowany obraz wejściowy, np. dla  $80 \rightarrow 45 \times 80$  pikseli (dla autorskiej),  $90 \times 80$  pikseli (dla FERET)
- Wartość odcięcia dolnego i górnego histogramu – ta operacja jest przeprowadzana podczas kroków normalizacyjnych na początku działania, ale z powodu jej istotnego wpływu na wynik jest ujęta w procesie badawczym,
- Typ uaktualniania wyników – określa sposób w jaki uaktualniane są wyniki w macierzy trafień: 0 – pojedyncza inkrementacja, 1 – inkrementacja maską  $3 \times 3$  pikseli, 2 – inkrementacja maską  $5 \times 5$  pikseli.

W celu miarodajnego przebadania wpływu poszczególnych parametrów zdecydowano się przeprowadzić testy dla wszystkich możliwych kombinacji parametrów zawartych w tabeli 3.3 (daje to dokładnie 2160 możliwych kombinacji), a w wyniku uzyskano prawie 0,5mln pojedynczych badań. Taka ilość daje podstawy by zaobserwować tendencje w jaki sposób poszczególne parametry i ich zmiany wpływają na ostateczny wynik.

Niezbędnym stało się skonstruowanie programu automatyzującego proces zmieniania parametrów i testowania całej bazy przy ich pomocy. Uzyskane za jego pomocą wyniki zapisane zostały w plikach tekstowych. Ich format jest zgodny z tym użytym przy tworzeniu „Zestawu wzorcowych wyników” opisanych w rozdziale 3.2.

Tak przygotowane dane użyto do kolejnego kroku badawczego – analizy. Zdecydowano się zautomatyzować również ten proces. Ustalono wartości odchylen od wzorca jakie może wystąpić w badanym przypadku. Wynoszą one odpowiednio:

$$\begin{aligned} war_1 : \sqrt{(x_b - x_w)^2 + (y_b - y_w)^2} &< 0,15 * r_w \\ war_2 : |r_b - r_w| &< 0,15 * r_w \end{aligned} \quad (3.1)$$

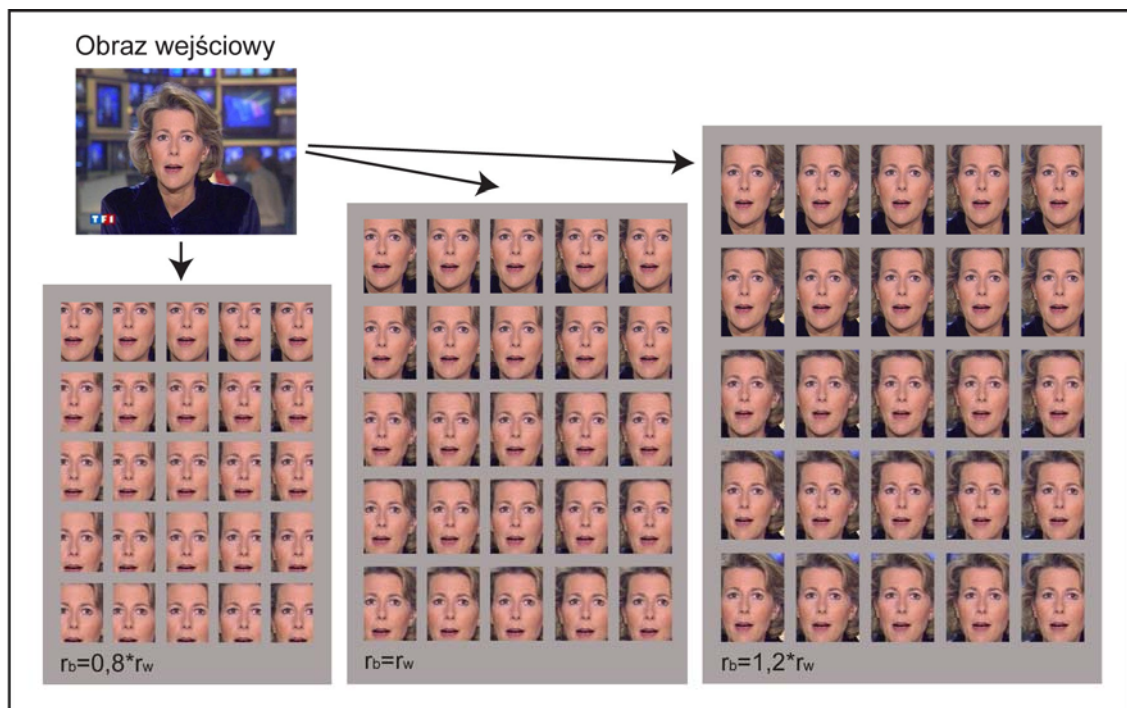
gdzie:

$x_b, y_b$  - współrzędna x i y środka wygenerowanego kandydata do bycia twarzą,

$x_w, y_w$  - współrzędna x i y środka twarzy wzorcowej,

$r_b, r_w$  - promień twarzy zaproponowanej przez generator i wzorcowej.

Zasadność użycia powyższych warunków łatwiej zrozumieć przyglądając się i analizując rysunek 3.1. Zawiera on przykłady wycinków obrazów, które mogą zostać zaproponowane przez algorytmy wyodrębniania kandydatów, a zakwalifikowe jako twarz. Należy zaznaczyć, że ukazują one w znacznej większości graniczne akceptacje, natomiast realny wynik działania algorytmu to przeważnie wycinki zbliżone do wzorca, lub całkowicie inne obszary obrazu, nie zawierające twarzy.



Rys. 3.1. Ilustracja ukazująca jakie twarze zaproponowane przez algorytmy lokalizacji obszarów zawierających twarz mogą zostać zakwalifikowane jako „trafne”.

Etap weryfikacji przedstawiony przy pomocy wzorów 3.1 przeprowadzony jest dla wszystkich pięciu wyodrębnionych kandydatów do bycia twarzą. Jeśli choć jeden mieści się w zaproponowanych warunkach, automatycznie cała seria próbek jest klasyfikowana jako „trafiony”.

Podczas badania algorytmu zauważono, że ma on wysoki wskaźnik trafień dla pierwszego wyodrębnionego kandydata do bycia twarzą, często wynoszący grubo ponad 90%. Zdecydowano się więc na dodatkowe zapamiętywanie efektywności już dla pierwszego kandydata (czyli jego „trafności”). Wyniki okazały się zaskakująco dobre. Pozwoliło to wysnuć wniosek, że metoda może służyć jako pełnoprawny lokalizator twarzy – czyli zgodnie z definicją wskazywać położenie i wielkość twarzy na obrazie, gdy jest pewność, że ona się tam znajduje.

Duża ilość przeprowadzonych testów umożliwiła wskazanie ustawień przy których metoda wykazuje najlepszy wynik. Jednak nie można założyć, że ustawienie to jest na pewno najlepszym ze wszystkich, ponieważ drugi i kolejne w kolejności wykazują jedynie minimalnie mniejszą efektywność. Niemniej zaobserwowano pewne zależności w doborze wartości parametrów, które wpływają pozytywnie na wynik.

Tabela 3.3 zawiera najlepsze 30 wyników dla bazy FERET (biorąc pod uwagę ogólny współczynnik poprawnego wyodrębnienia w pierwszych 5 próbach).

| L.p. | Kolor/skala szarości | Macierze gradientów 45/135 | Wartość odcięcia gradientu | Wielkość obszaru szczytowanego | Rozdzielczość badawcza | Wartość odcięcia dolnego histogramu | Wartość odcięcia górnego histogramu | Typ uaktualniania wyników | Odległość od środka twarzy w odniesieniu do promienia dla I kandydata | Różnica w długości promienia dla I kandydata | Współczynnik lokalizacji dla I kandydata | Odległość od środka twarzy w odniesieniu do promienia | Różnica w długości promienia | Współczynnik poprawnego wygenerowania kandydatów w 5 próbach | Średnia ilość poprawnych trafień w 5 zaproponowanych próbach | Średni czas wykonania dla obrazu |
|------|----------------------|----------------------------|----------------------------|--------------------------------|------------------------|-------------------------------------|-------------------------------------|---------------------------|-----------------------------------------------------------------------|----------------------------------------------|------------------------------------------|-------------------------------------------------------|------------------------------|--------------------------------------------------------------|--------------------------------------------------------------|----------------------------------|
|      | A                    | B                          | C                          | D                              | E                      | F                                   | G                                   | H                         | I                                                                     | J                                            | K                                        | L                                                     | M                            | N                                                            | O                                                            | P                                |
| 1    | 1                    | 1                          | 1                          | 5                              | 40                     | 0                                   | 90                                  | 2                         | 6,5                                                                   | 4,0                                          | <b>93,4</b>                              | 8,5                                                   | 4,0                          | <b>99,5</b>                                                  | 4,38                                                         | 14                               |
| 2    | 1                    | 1                          | 7                          | 5                              | 40                     | 0                                   | 50                                  | 0                         | 7,4                                                                   | 4,1                                          | <b>92,5</b>                              | 8,5                                                   | 4,2                          | <b>99,0</b>                                                  | 4,28                                                         | 10                               |
| 3    | 1                    | 0                          | 7                          | 5                              | 40                     | 0                                   | 50                                  | 1                         | 6,0                                                                   | 3,7                                          | <b>93,0</b>                              | 8,5                                                   | 3,9                          | <b>99,0</b>                                                  | 4,40                                                         | 10                               |
| 4    | 1                    | 0                          | 5                          | 5                              | 40                     | 0                                   | 90                                  | 1                         | 5,8                                                                   | 3,7                                          | <b>90,6</b>                              | 8,4                                                   | 4,0                          | <b>99,0</b>                                                  | 4,40                                                         | 10                               |
| 5    | 1                    | 1                          | 1                          | 5                              | 40                     | 0                                   | 255                                 | 0                         | 7,2                                                                   | 4,1                                          | <b>91,1</b>                              | 8,6                                                   | 4,3                          | <b>98,6</b>                                                  | 4,35                                                         | 9                                |
| 6    | 1                    | 1                          | 1                          | 5                              | 40                     | 0                                   | 255                                 | 2                         | 6,5                                                                   | 3,6                                          | <b>93,0</b>                              | 8,4                                                   | 3,9                          | <b>98,6</b>                                                  | 4,41                                                         | 12                               |
| 7    | 1                    | 1                          | 3                          | 5                              | 40                     | 0                                   | 90                                  | 2                         | 6,5                                                                   | 4,0                                          | <b>92,5</b>                              | 8,5                                                   | 4,0                          | <b>98,6</b>                                                  | 4,40                                                         | 13                               |
| 8    | 1                    | 1                          | 5                          | 5                              | 40                     | 0                                   | 90                                  | 2                         | 6,5                                                                   | 3,9                                          | <b>92,5</b>                              | 8,5                                                   | 4,0                          | <b>98,6</b>                                                  | 4,41                                                         | 13                               |
| 9    | 1                    | 1                          | 5                          | 5                              | 40                     | 0                                   | 50                                  | 2                         | 6,6                                                                   | 3,6                                          | <b>92,0</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,37                                                         | 14                               |
| 10   | 1                    | 1                          | 7                          | 5                              | 40                     | 0                                   | 50                                  | 2                         | 6,3                                                                   | 3,6                                          | <b>92,0</b>                              | 8,5                                                   | 3,8                          | <b>98,6</b>                                                  | 4,42                                                         | 13                               |
| 11   | 1                    | 1                          | 7                          | 5                              | 40                     | 0                                   | 70                                  | 2                         | 6,3                                                                   | 3,6                                          | <b>92,0</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,45                                                         | 13                               |
| 12   | 1                    | 1                          | 3                          | 5                              | 40                     | 0                                   | 255                                 | 2                         | 6,1                                                                   | 3,7                                          | <b>91,6</b>                              | 8,4                                                   | 3,9                          | <b>98,6</b>                                                  | 4,42                                                         | 11                               |
| 13   | 1                    | 1                          | 5                          | 5                              | 40                     | 0                                   | 70                                  | 0                         | 7,4                                                                   | 4,1                                          | <b>91,6</b>                              | 8,5                                                   | 4,2                          | <b>98,6</b>                                                  | 4,30                                                         | 10                               |
| 14   | 1                    | 1                          | 7                          | 5                              | 40                     | 0                                   | 90                                  | 0                         | 6,8                                                                   | 4,1                                          | <b>91,6</b>                              | 8,6                                                   | 4,2                          | <b>98,6</b>                                                  | 4,33                                                         | 10                               |
| 15   | 1                    | 1                          | 1                          | 5                              | 40                     | 0                                   | 255                                 | 1                         | 6,2                                                                   | 3,9                                          | <b>91,1</b>                              | 8,4                                                   | 3,9                          | <b>98,6</b>                                                  | 4,40                                                         | 11                               |
| 16   | 1                    | 1                          | 7                          | 5                              | 40                     | 0                                   | 50                                  | 1                         | 6,2                                                                   | 3,7                                          | <b>91,1</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,40                                                         | 11                               |
| 17   | 1                    | 1                          | 3                          | 5                              | 40                     | 0                                   | 255                                 | 0                         | 6,8                                                                   | 4,4                                          | <b>89,7</b>                              | 8,5                                                   | 4,4                          | <b>98,6</b>                                                  | 4,30                                                         | 9                                |
| 18   | 1                    | 1                          | 7                          | 3                              | 40                     | 0                                   | 90                                  | 2                         | 6,0                                                                   | 3,6                                          | <b>87,9</b>                              | 8,4                                                   | 3,9                          | <b>98,6</b>                                                  | 4,34                                                         | 11                               |
| 19   | 1                    | 0                          | 3                          | 5                              | 40                     | 0                                   | 70                                  | 1                         | 6,3                                                                   | 3,8                                          | <b>94,4</b>                              | 8,5                                                   | 4,0                          | <b>98,6</b>                                                  | 4,38                                                         | 10                               |
| 20   | 1                    | 0                          | 5                          | 5                              | 40                     | 0                                   | 50                                  | 2                         | 6,5                                                                   | 3,5                                          | <b>94,4</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,41                                                         | 10                               |
| 21   | 1                    | 0                          | 3                          | 5                              | 40                     | 0                                   | 70                                  | 2                         | 6,3                                                                   | 3,8                                          | <b>93,9</b>                              | 8,6                                                   | 4,1                          | <b>98,6</b>                                                  | 4,40                                                         | 10                               |
| 22   | 1                    | 0                          | 1                          | 5                              | 40                     | 0                                   | 70                                  | 2                         | 6,4                                                                   | 3,7                                          | <b>93,4</b>                              | 8,6                                                   | 4,0                          | <b>98,6</b>                                                  | 4,36                                                         | 10                               |
| 23   | 1                    | 0                          | 3                          | 5                              | 40                     | 0                                   | 50                                  | 1                         | 6,4                                                                   | 3,6                                          | <b>93,4</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,39                                                         | 10                               |
| 24   | 1                    | 0                          | 7                          | 5                              | 40                     | 0                                   | 50                                  | 2                         | 6,4                                                                   | 3,7                                          | <b>93,4</b>                              | 8,5                                                   | 3,9                          | <b>98,6</b>                                                  | 4,40                                                         | 10                               |
| 25   | 1                    | 0                          | 7                          | 5                              | 40                     | 0                                   | 70                                  | 2                         | 6,3                                                                   | 3,6                                          | <b>93,0</b>                              | 8,4                                                   | 4,0                          | <b>98,6</b>                                                  | 4,39                                                         | 10                               |
| 26   | 1                    | 0                          | 1                          | 5                              | 40                     | 0                                   | 255                                 | 1                         | 6,2                                                                   | 4,0                                          | <b>92,5</b>                              | 8,4                                                   | 4,1                          | <b>98,6</b>                                                  | 4,39                                                         | 9                                |
| 27   | 1                    | 0                          | 3                          | 5                              | 40                     | 0                                   | 90                                  | 2                         | 6,7                                                                   | 3,9                                          | <b>92,0</b>                              | 8,5                                                   | 4,0                          | <b>98,6</b>                                                  | 4,35                                                         | 10                               |
| 28   | 1                    | 0                          | 5                          | 5                              | 40                     | 0                                   | 70                                  | 1                         | 6,0                                                                   | 3,6                                          | <b>92,0</b>                              | 8,4                                                   | 4,0                          | <b>98,6</b>                                                  | 4,38                                                         | 10                               |
| 29   | 1                    | 0                          | 7                          | 5                              | 40                     | 0                                   | 70                                  | 1                         | 6,0                                                                   | 3,7                                          | <b>92,0</b>                              | 8,4                                                   | 3,9                          | <b>98,6</b>                                                  | 4,39                                                         | 10                               |
| 30   | 1                    | 0                          | 5                          | 5                              | 40                     | 0                                   | 90                                  | 0                         | 6,9                                                                   | 4,5                                          | <b>90,6</b>                              | 8,6                                                   | 4,4                          | <b>98,6</b>                                                  | 4,26                                                         | 9                                |

Tabela 3.4. Pierwsze trzydzieści wyników pochodzących z badań na bazie FERET

Analizując powyższą tabelę wyciągnięto następujące wnioski:

- zaobserwowano wysoki współczynnik trafności wytypowania kandydatów twarzy w pierwszych pięciu próbach (kolumna N), sięgający 99,5%, co daje jedną błędną na 215 obrazów poddanych analizie,
- wysoki współczynnik utrzymuje się dla wszystkich 30 wyników, więc można stwierdzić, że algorytmy są dobrze opracowane i nie wymagają dalszych usprawnień,
- w kolumnie K ukazującej wyniki lokalizacji dla pierwszego kandydata, a oscylujące w najlepszych przypadkach w okolicach 94% umożliwiają uznać algorytmy za dobrej klasy lokalizator, zgodnie z definicją wskazującą jedną twarz na obrazie gdy jest pewność, że ona się tam znajduje – niekonieczny jest wtedy moduł twarz/„nie twarz”.
- średni czas wykonania obliczeń dla jednego obrazu oscyluje w granicach  $9\div 15$ ms (kolumna P), co stawia opracowane algorytmy bardzo wysoko w porównaniu z konkurencją na rynku lokalizatorów,
- wszystkie 30 najlepszych wyników osiągnięto przy użyciu obrazów kolorowych (kolumna A) – dla skali szarości najlepszy wynik wyniósł 84,1%,
- włączenie macierzy ukośnych jest korzystne – najlepszy wynik jak i 16 z 30 przypadków w udziale tej włączonej opcji (kolumna B),
- wszystkie wyniki w tabeli 3.3 uzyskano, gdy użyto powiększonego obszaru sczytywania przy wpisywaniu do ZMT co udowadnia słuszność jego użycia,
- skalowanie do wielkości  $45\times 40$  pikseli (rozdzielczość ta jest różna od opisywanej w rozdziale 2.4  $45\times 80$  pikseli ponieważ obraz jest w orientacji pionowej) występuje we wszystkich 30 najlepszych ustawieniach (kolumna E),
- górna granica histogramu obcięta do wartości poniżej 100 (dla wersji z wykorzystaniem komponentu  $I_2$ ) nieznacznie zwiększyła efektywność w stosunku do nie obciętego (kolumna G),
- uaktualnianie wyników trafień algorytmu wyszukiwania obiektów owalnych przy pomocy masek  $3\times 3$  i  $5\times 5$  pikseli występuje w 24 na 30 przypadków (kolumna H),
- średnia różnica w długości promienia twarzy wykrytej do wzorcowej to tylko 4,0% (kolumny L i M), a średnia odległość pomiędzy wykrytym środkiem twarzy, a wzorcowym wynosi jedynie 8,5% (w stosunku do długości promienia twarzy)
- stwierdzono wysoki współczynnik poprawnych trafień dla wszystkich wygenerowanych próbek (ponad 4 poprawne na 5 zaproponowanych), czyli algorytm ma małą szansę nie wykryć twarzy, ponieważ większość najlepszych wyników skupiona jest wokół punktu określającego środek szukanej twarzy, a ten punkt jest celem poszukiwań (kolumna O).

Poniżej przedstawiono wyniki dla bazy autorskiej. Tabela (3.5) zawiera pierwszych 30 ustawień parametrów z najwyższą skutecznością.

| L.p. | Kolor/skala szarości | Macierze gradientów 45/135 | Wartość odcięcia gradientu | Wielkość obszaru szczytowanego | Rozdzielczość badawcza | Wartość odcięcia dolnego histogramu | Wartość odcięcia górnego histogramu | Typ uaktualniania wyników | Odległość od środka twarzy w odniesieniu do promienia dla I kandydata | Różnica w długości promienia dla I kandydata | Współczynnik lokalizacji dla I kandydata | Odległość od środka twarzy w odniesieniu do promienia | Różnica w długości promienia | Współczynnik poprawnego wygenerowania kandydatów w 5 próbach | Średnia ilość poprawnych trafień w 5 zaproponowanych próbach | Średni czas wykonania dla obrazu |
|------|----------------------|----------------------------|----------------------------|--------------------------------|------------------------|-------------------------------------|-------------------------------------|---------------------------|-----------------------------------------------------------------------|----------------------------------------------|------------------------------------------|-------------------------------------------------------|------------------------------|--------------------------------------------------------------|--------------------------------------------------------------|----------------------------------|
|      | A                    | B                          | C                          | D                              | E                      | F                                   | G                                   | H                         | I                                                                     | J                                            | K                                        | L                                                     | M                            | N                                                            | O                                                            | P                                |
| 1    | 1                    | 1                          | 1                          | 5                              | 80                     | 0                                   | 90                                  | 2                         | 6,1                                                                   | 1,5                                          | 91,7                                     | 7,7                                                   | 1,7                          | 99,5                                                         | 4,16                                                         | 28                               |
| 2    | 1                    | 1                          | 1                          | 5                              | 80                     | 0                                   | 70                                  | 2                         | 5,7                                                                   | 1,6                                          | 91,3                                     | 7,6                                                   | 1,6                          | 99,5                                                         | 4,14                                                         | 28                               |
| 3    | 1                    | 1                          | 1                          | 5                              | 80                     | 0                                   | 90                                  | 1                         | 5,9                                                                   | 1,8                                          | 90,8                                     | 7,7                                                   | 1,9                          | 99,5                                                         | 4,12                                                         | 25                               |
| 4    | 1                    | 1                          | 1                          | 5                              | 80                     | 0                                   | 70                                  | 1                         | 6,1                                                                   | 1,8                                          | 90,3                                     | 7,8                                                   | 1,8                          | 99,5                                                         | 4,12                                                         | 25                               |
| 5    | 1                    | 1                          | 1                          | 3                              | 80                     | 0                                   | 90                                  | 2                         | 5,8                                                                   | 1,7                                          | 89,8                                     | 7,9                                                   | 2,1                          | 99,5                                                         | 4,02                                                         | 24                               |
| 6    | 1                    | 1                          | 1                          | 3                              | 80                     | 0                                   | 70                                  | 2                         | 6,0                                                                   | 1,9                                          | 91,3                                     | 7,7                                                   | 2,1                          | 99,0                                                         | 4,05                                                         | 25                               |
| 7    | 1                    | 1                          | 1                          | 3                              | 80                     | 0                                   | 70                                  | 1                         | 5,9                                                                   | 2,5                                          | 87,4                                     | 7,8                                                   | 2,6                          | 99,0                                                         | 3,98                                                         | 21                               |
| 8    | 1                    | 0                          | 5                          | 3                              | 80                     | 0                                   | 70                                  | 2                         | 5,8                                                                   | 2,2                                          | 85,9                                     | 7,8                                                   | 2,6                          | 99,0                                                         | 3,94                                                         | 16                               |
| 9    | 1                    | 0                          | 5                          | 3                              | 80                     | 0                                   | 70                                  | 1                         | 6,4                                                                   | 2,5                                          | 84,5                                     | 7,8                                                   | 2,8                          | 99,0                                                         | 3,84                                                         | 16                               |
| 10   | 1                    | 1                          | 1                          | 3                              | 80                     | 0                                   | 70                                  | 0                         | 6,6                                                                   | 3,0                                          | 82,6                                     | 8,0                                                   | 3,1                          | 99,0                                                         | 3,70                                                         | 16                               |
| 11   | 1                    | 0                          | 3                          | 1                              | 80                     | 0                                   | 70                                  | 2                         | 6,6                                                                   | 2,6                                          | 82,1                                     | 7,9                                                   | 3,2                          | 99,0                                                         | 3,77                                                         | 14                               |
| 12   | 1                    | 0                          | 1                          | 1                              | 80                     | 0                                   | 70                                  | 2                         | 6,5                                                                   | 2,4                                          | 81,6                                     | 8,0                                                   | 3,0                          | 99,0                                                         | 3,74                                                         | 15                               |
| 13   | 1                    | 1                          | 1                          | 1                              | 80                     | 0                                   | 70                                  | 1                         | 6,4                                                                   | 3,2                                          | 80,1                                     | 8,0                                                   | 3,4                          | 99,0                                                         | 3,52                                                         | 18                               |
| 14   | 1                    | 1                          | 5                          | 5                              | 80                     | 0                                   | 90                                  | 2                         | 6,0                                                                   | 2,3                                          | 93,7                                     | 7,7                                                   | 2,1                          | 98,5                                                         | 4,12                                                         | 25                               |
| 15   | 1                    | 1                          | 7                          | 5                              | 80                     | 0                                   | 90                                  | 2                         | 5,7                                                                   | 2,2                                          | 92,7                                     | 7,7                                                   | 2,2                          | 98,5                                                         | 4,12                                                         | 25                               |
| 16   | 1                    | 1                          | 3                          | 5                              | 80                     | 0                                   | 90                                  | 2                         | 6,1                                                                   | 1,7                                          | 92,7                                     | 7,7                                                   | 1,8                          | 98,5                                                         | 4,13                                                         | 27                               |
| 17   | 1                    | 1                          | 7                          | 5                              | 80                     | 0                                   | 70                                  | 2                         | 5,7                                                                   | 2,2                                          | 92,2                                     | 7,7                                                   | 2,0                          | 98,5                                                         | 4,16                                                         | 25                               |
| 18   | 1                    | 1                          | 5                          | 5                              | 80                     | 0                                   | 70                                  | 2                         | 5,7                                                                   | 1,8                                          | 91,7                                     | 7,7                                                   | 1,9                          | 98,5                                                         | 4,13                                                         | 26                               |
| 19   | 1                    | 1                          | 3                          | 5                              | 80                     | 0                                   | 70                                  | 2                         | 5,5                                                                   | 1,7                                          | 91,7                                     | 7,6                                                   | 1,7                          | 98,5                                                         | 4,14                                                         | 27                               |
| 20   | 1                    | 1                          | 5                          | 3                              | 80                     | 0                                   | 90                                  | 2                         | 5,7                                                                   | 2,0                                          | 91,3                                     | 7,8                                                   | 2,2                          | 98,5                                                         | 4,04                                                         | 22                               |
| 21   | 1                    | 1                          | 7                          | 5                              | 80                     | 0                                   | 90                                  | 1                         | 5,9                                                                   | 2,4                                          | 90,8                                     | 7,7                                                   | 2,2                          | 98,5                                                         | 4,12                                                         | 22                               |
| 22   | 1                    | 1                          | 5                          | 5                              | 80                     | 0                                   | 90                                  | 1                         | 6,1                                                                   | 2,3                                          | 90,8                                     | 7,8                                                   | 2,1                          | 98,5                                                         | 4,15                                                         | 23                               |
| 23   | 1                    | 1                          | 3                          | 5                              | 80                     | 0                                   | 90                                  | 1                         | 6,0                                                                   | 1,9                                          | 90,8                                     | 7,8                                                   | 1,9                          | 98,5                                                         | 4,13                                                         | 23                               |
| 24   | 1                    | 1                          | 1                          | 5                              | 120                    | 0                                   | 70                                  | 2                         | 6,5                                                                   | 4,2                                          | 90,8                                     | 7,3                                                   | 4,2                          | 98,5                                                         | 4,37                                                         | 136                              |
| 25   | 1                    | 0                          | 1                          | 5                              | 80                     | 0                                   | 70                                  | 1                         | 6,2                                                                   | 2,2                                          | 90,3                                     | 7,8                                                   | 2,4                          | 98,5                                                         | 4,08                                                         | 20                               |
| 26   | 1                    | 1                          | 3                          | 3                              | 80                     | 0                                   | 90                                  | 2                         | 59                                                                    | 17                                           | 90,3                                     | 7,8                                                   | 2,1                          | 98,5                                                         | 4,03                                                         | 23                               |
| 27   | 1                    | 0                          | 3                          | 5                              | 80                     | 0                                   | 70                                  | 2                         | 57                                                                    | 21                                           | 88,8                                     | 7,8                                                   | 2,4                          | 98,5                                                         | 4,12                                                         | 20                               |
| 28   | 1                    | 1                          | 1                          | 5                              | 80                     | 0                                   | 70                                  | 0                         | 72                                                                    | 18                                           | 88,8                                     | 7,8                                                   | 2,5                          | 98,5                                                         | 3,93                                                         | 22                               |
| 29   | 1                    | 0                          | 3                          | 5                              | 80                     | 0                                   | 70                                  | 1                         | 58                                                                    | 23                                           | 88,4                                     | 7,8                                                   | 2,5                          | 98,5                                                         | 4,06                                                         | 19                               |
| 30   | 1                    | 1                          | 5                          | 3                              | 80                     | 0                                   | 90                                  | 1                         | 60                                                                    | 23                                           | 88,4                                     | 7,9                                                   | 2,7                          | 98,5                                                         | 3,95                                                         | 20                               |

Tabela 3.5. Najlepsze trzydzieści wyników uzyskane z badań na bazie autorskiej

W powyższej tabeli można zaobserwować między innymi następujące zależności:

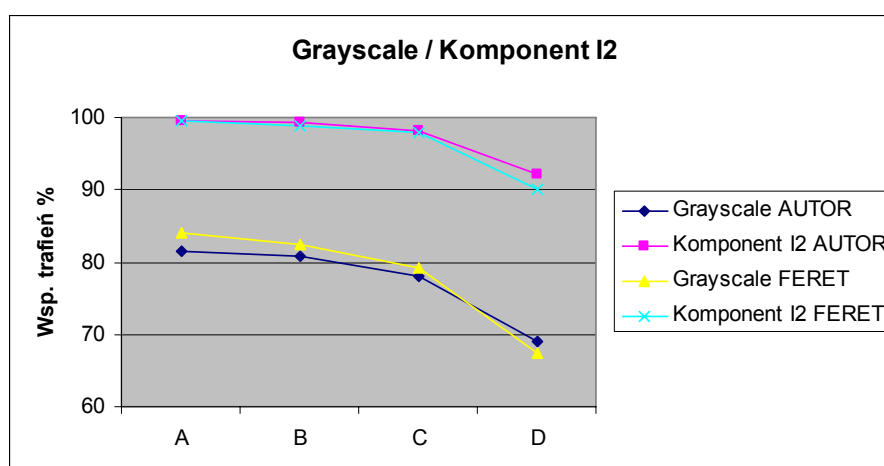
- istnieje kilkanaście ustawień parametrów dla których algorytm wykazuje wysoki, sięgający 99,5% współczynnik poprawnego wytypowania w pierwszych pięciu próbach obszaru zawierającego twarz (kolumna N),
- nie występuje szczególny przypadek ustawień parametrów, zdecydowanie przewyższający inne, co w połączeniu z wysokim współczynnikiem wspomnianym powyżej, daje możliwość stwierdzenia, że system ma na tyle przemyślane i zaprojektowane algorytmy, że nie istnieje konieczność dalszego badania,
- kolumna ukazująca wysoki wynik lokalizacji dla pierwszego kandydata (szczególnie ustawienia w linii 14 – 93,7%) dodaje algorytmowi waloru w postaci możliwości zastosowania go w systemach detekcji bez modułu twarz/„nie twarz” – oczywiście przy założeniu, że twarz na obrazie występuje (mamy wtedy do czynienia z klasycznym lokalizatorem) (kolumna K); dodatkowo sprawdzono wynik, gdy ograniczenia warunków 3.1 zwiększono z 0,15 do 0,2 – wyniósł 97,4%,
- średni czas wykonania obliczeń dla jednego obrazu oscyluje w granicach 20÷30ms, co jest wynikiem bardzo dobrym w porównaniu z innymi detektorami (kolumna P),
- jedynie obrazy kolorowe dają możliwość pracy z tak wysokim współczynnikiem poprawnych generowań (najwyższy wynik dla odcieni szarości to 81,6%) (kolumna A),
- włączenie macierzy ukośnych jest korzystne – 23 na 30 najlepszych wyników osiągnięto przy ich użyciu (kolumna B),
- użycie powiększonego obszaru sczytywania przy wpisywaniu do „zbiorczej macierzy trafień” pozwala na konstruowanie systemów o wysokim współczynniku trafień przy pierwszej próbie (kolumna D),
- skalowanie do wielkości 45×80 pikseli zdominowało przedstawione wyniki (kolumna E),
- obcięcie górnej granicy histogramu (dla wersji z wykorzystaniem komponentu  $I_2$ ) do wartości poniżej 100 zwiększyło efektywność w stosunku do nie obciętego (kolumna G),
- uaktualnianie wyników trafień algorytmu wyszukiwania obiektów owalnych przy pomocy zaproponowanych masek 3×3 i 5×5 występuje w 28 na 30 przypadków (kolumna H),
- średnia odległość wykrytego i zweryfikowanego środka twarzy od środka twarzy wzorcowej w stosunku do promienia wzorcowej twarzy wyniósł średnio 7,8%, a różnice w długości promienia to tylko 2,4% (kolumny L i M),
- dla ustawień oferujących najlepsze wyniki stwierdzono wysoki globalny współczynnik trafień, co oznacza, że algorytm na 5 zaproponowanych próbek dla danego obrazu generował ponad 4 poprawnie zakwalifikowane (kolumna O).

Obie tabele (3.4 i 3.5) przedstawiają najlepsze wyniki i mogłyby stanowić koniec badań. Jednak, aby poznać dlaczego tak się stało, że to właśnie one są najlepsze, trzeba dokonać głębszej analizy. Zdecydowano się przedstawić szereg zestawień zawierających zmienność parametrów i ich wpływ na ostateczną efektywność systemu. Kolejność ich prezentacji uzależniona jest od występowania w algorytmach.

Pierwszym badanym czynnikiem jest zasadność wykorzystania informacji o pełnym kolorze. Z punktu widzenia logiki taki zabieg powinien dawać korzyść. Należy jednak zdecydować w jaki sposób użyć tej informacji, żeby maksymalnie ją wykorzystać. Tabela 3.4 i odpowiadający jej rysunek 3.2 ukazują wyniki uzyskane przy obliczeniach z wykorzystaniem bazy w postaci kolorowej, jak i przekonwertowanej do odcieni szarości.

|                       | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|-----------------------|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|                       | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|                       | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| <b>Skala szarości</b> | 81,6                                     | 84,1  | 80,7                                   | 82,5  | 78,0                                     | 79,2  | 69,0                                | 67,5  | 22,0                                                               | 11,2  |
| <b>Komponent I2</b>   | 99,5                                     | 99,5  | 99,2                                   | 98,8  | 98,2                                     | 98,0  | 92,2                                | 90,1  | 17,0                                                               | 8,7   |

Tabela 3.6. Wyniki uzyskane dla włączonej i wyłączonej opcji analizowania koloru



Rys. 3.2. Graficzna interpretacja wyników przedstawionych w tabeli 3.6

(A, B, C, D – oznaczenia kolumn z tabeli 3.6)

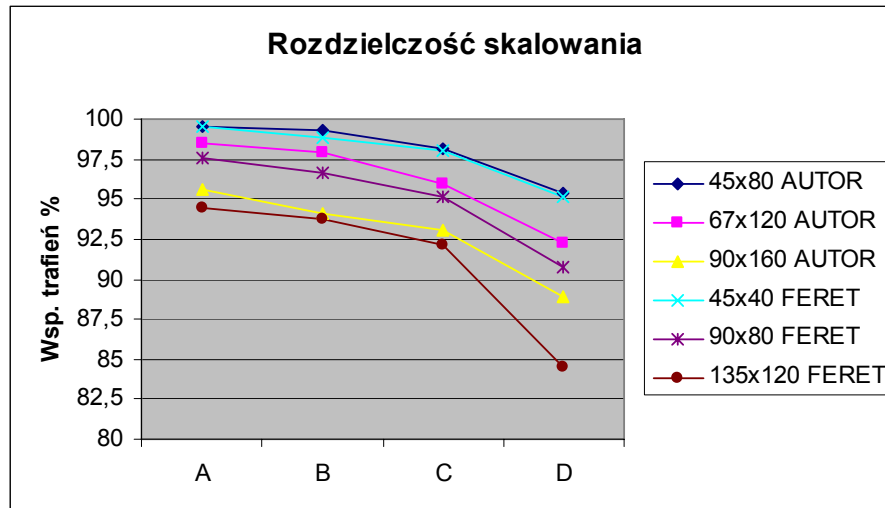
Komponent *I2* z przestrzeni *III2I3* ukazał swoją przydatność. Efektywność wzrosła, a dodatkowo spadło zapotrzebowanie na moc obliczeniową. Spowodowane jest to specyfiką obliczeń operujących na komponencie *I2* pomijających spore obszary obrazu (więcej o tej zależności w rozdziale 2.4). Tabela 3.6 ukazuje, że najlepszy uzyskany wynik w przestrzeni Grayscale jest gorszy nawet od średniej ze wszystkich zbadanych dla kolorowej. Również czas wykonania algorytmu wynosi średnio prawie cztery razy więcej. Stąd wniosek, że odpowiednie użycie informacji o kolorze jest jednym z głównych czynników wpływających na drastyczną poprawę działania całego systemu. Na uwagę zasługuje również niewielki spadek efektywności (trzymający się na poziomie powyżej 97%) nawet dla 100 pierwszych konfiguracji parametrów. Taka informacja pozwala założyć, że system jest dobrze zaprojektowany i jedynie radykalne zmiany klasy testowanych parametrów mogą wpłynąć bardzo niekorzystnie na efektywność. Ciekawym jest też zjawisko podobieństwa efektywności dla obu baz danych testowych i to zarówno w wersji kolorowej jak i w skali szarości. Można więc założyć, że wyniki z jednej bazy potwierdzają te uzyskane z drugiej, eliminując podejrzenie przypadkowości.

Ze względu na zbyt duże wypaczenie wyników w dalszych rozważaniach testy przeprowadzone na bazie obrazów w skali szarości nie będą brane pod uwagę. Opcja pracy w tym trybie została zostawiona, jednak nie jest rekomendowana.

Kolejnym badanym aspektem algorytmów jest niezwykle istotny parametr wpływającym na szybkość działania jak i na efektywność. Jest nim rozdzielczość do jakiej skalowany jest obraz wejściowy. Tabela 3.7 przedstawia wyniki uzyskane przez algorytm dla różnych rozdzielczości pracy algorytmu.

|                                 | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45×80 (AUTOR)<br>R=45×40 (FERET) |       |
|---------------------------------|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|                                 | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|                                 | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| <b>45×80 (45×40) pikseli</b>    | 99,5                                     | 99,5  | 98,8                                   | 99,3  | 98                                       | 98,2  | 95,1                                | 95,4  | 8,7                                                                | 17    |
| <b>67×120 (90×80) pikseli</b>   | 97,6                                     | 98,5  | 96,6                                   | 97,9  | 95,1                                     | 95,9  | 90,7                                | 92,3  | 49,7                                                               | 59,7  |
| <b>90×160 (135×120) pikseli</b> | 94,4                                     | 95,6  | 93,7                                   | 94,1  | 92,1                                     | 93,1  | 84,5                                | 88,9  | 150,4                                                              | 110,8 |

Tabela 3.7. Wyniki uzyskane dla różnych wielkości skalowania.



Rys. 3.3. Graficzna interpretacja wyników przedstawionych w tabeli 3.7

(A, B, C, D – oznaczenia kolumn z tabeli 3.7)

Już na początku rozważań skonstruowano tezę, że nie jest zasadnym analizowanie obrazu w rozdzielczości wejściowej ponieważ niepotrzebnie wydłuża cały proces, a dodatkowo szum zawarty w niemal każdym obrazie cyfrowym byłby analizowany w identyczny sposób jak piksele przechowujące rzeczywistą, użyteczną wartość. Dlatego skalowanie uważa się również za etap normalizacyjny. Widać wyraźnie przewagę używania niskiej rozdzielczości do której przeprowadzane jest skalowanie. Wyniki są zdecydowanie lepsze zarówno w dziedzinie efektywności, ale bardzo istotne jest też znikome obciążenie komputera. Czas wykonywania wynoszący ok. 20ms na zwykłym komputerze biurowym (stan na rok 2006) to wynik rekomendujący ten system do pracy w czasie rzeczywistym. Wraz ze wzrostem rozdzielczości do której skalowany jest obraz wejściowy spada wydajność i efektywność dla obu badanych baz zdjęć. Jednak nawet wtedy najlepsze wyniki znacznie przekraczające 90% są bardzo dobre.

Kolejnym badanym parametrem będzie zestaw dwóch liczb określający odciecie histogramu po konwersji z przestrzeni RGB do komponentu  $I_2$  (pozyskanego z przestrzeni  $I_1I_2I_3$ ). Zdecydowano się na użycie czterech par tych parametrów: 0/50, 0/70, 0/90, 0/255, gdzie pierwsza liczba określa dolną, a druga górną wartość dla tej operacji. Pozwoliło to na obserwację zachowania efektywności pod wpływem eliminowania coraz większej ilości informacji z histogramu jak i w przypadku pominięcia tego kroku (czyli dla 255). Sposób w jaki odciecie histogramu wpływa na obraz w przestrzeni  $I_2$  jest szerzej w rozdziale 2.4. Poniższa tabela 3.8 przedstawia wyniki uzyskane z analizy wyników.

|                      | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|----------------------|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|                      | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|                      | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| Obcięcie hist. 0/50  | 99                                       | 97,5  | 98,6                                   | 97,3  | 95,9                                     | 95,4  | 93,7                                | 92,5  | 9,1                                                                | 17,5  |
| Obcięcie hist. 0/70  | 98,6                                     | 99,5  | 98,5                                   | 99,1  | 95,8                                     | 97,5  | 93,3                                | 94,1  | 9,1                                                                | 17,6  |
| Obcięcie hist. 0/90  | 99,5                                     | 99,5  | 98,7                                   | 98,8  | 96,1                                     | 97,0  | 93                                  | 93,6  | 8,9                                                                | 17,5  |
| Obcięcie hist. 0/255 | 98,6                                     | 98,0  | 98,4                                   | 97,9  | 93,6                                     | 94,4  | 80,4                                | 88,6  | 7,7                                                                | 15,1  |

Tabela 3.8. Wyniki uzyskane dla różnych konfiguracji obcięcia histogramu.

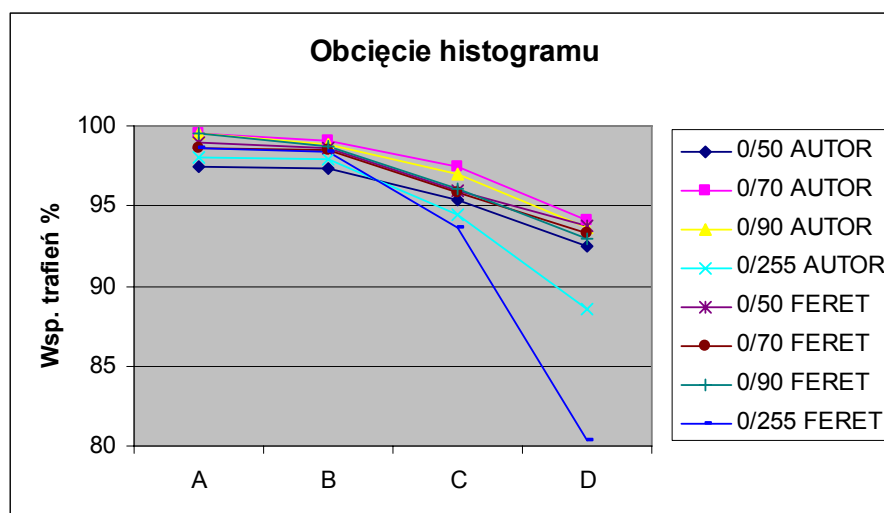
Rys. 3.4. Graficzna interpretacja wyników przedstawionych w tabeli 3.8  
(A, B, C, D – oznaczenia kolumn z tabeli 3.8)

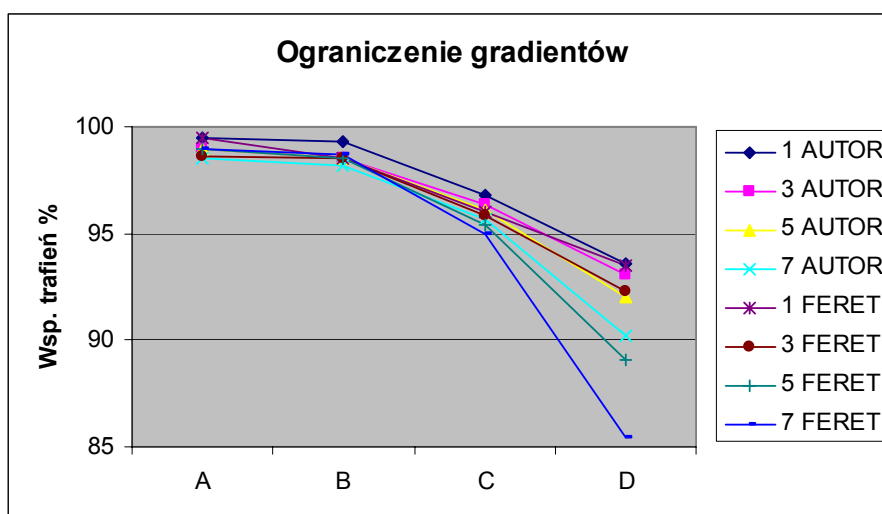
Tabela 3.8 i wykres ją obrazujący ukazują zależność efektywności systemu i czasu wykonywania od parametrów obcięcia histogramu. Wynika z nich prawidłowość, którą zauważono już na etapie projektowania, a która znalazła swoje potwierdzenie w tych testach, czyli o bezzasadności analizowania całego obrazu w komponencie *I2* skoro twarz zawiera się w zakresie luminancji 0–70. Dlatego nie dziwi fakt, że najlepszą efektywność uzyskano właśnie dla takiego obcięcia histogramu. Niemal identyczne wyniki zanotowało obcięcie na poziomie 0–90, a 0–50 również znacząco nie odstawało – takie wyniki utwierdzają w przekonaniu, że zabieg obcięcia histogramu jest słuszny. Trzeba jednak również zauważyć, że założenie o zwiększeniu szybkości wykonywania algorytmu poprzez wyeliminowanie analizy niepotrzebnych pikseli (obcięcie niweluje różnicę w jasności pikseli, czyli wyklucza

powstawanie drobnych wartości gradientów) nie sprawdziło się. Czas obliczenia obciążenia histogramu okazał się bardziej kosztowny czasowo niż zmniejszenie ilości obliczeń (sumarycznie wydłużył się o około 15–20%). Wyniki jednak jednoznacznie wskazują, że obciążenie histogramu jest czynnikiem wpływającym korzystnie na efektywność.

Niewątpliwie parametrem mogącym mieć znaczący wpływ na ostateczny wynik jest wartość odcięcia gradientu dla którego algorytm wykonuje obliczenia. Jego wielkość determinuje w jaki sposób traktowana będzie tablica rozkładu gradientów. Dla małych wartości ilość analizowanych punktów rośnie, i choć na wyjściu jest więcej danych to mogą one zawierać szum. Poniższa tab. 3.9 i rys. 3.5 przedstawia wyniki analizy ilustrując również w jaki sposób wielkości odcięcia wartości analizowanych gradientów zmieniają wynik.

|                              | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|------------------------------|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|                              | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|                              | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| <b><math>G \geq 1</math></b> | 99,5                                     | 99,5  | 98,5                                   | 99,3  | 96                                       | 96,8  | 93,5                                | 93,6  | 8,9                                                                | 16,2  |
| <b><math>G \geq 3</math></b> | 98,6                                     | 99    | 98,5                                   | 98,5  | 95,8                                     | 96,4  | 92,3                                | 93,1  | 8,8                                                                | 16,6  |
| <b><math>G \geq 5</math></b> | 99                                       | 99    | 98,5                                   | 98,5  | 95,4                                     | 96,1  | 89,1                                | 92    | 8,6                                                                | 17,3  |
| <b><math>G \geq 7</math></b> | 99                                       | 98,5  | 98,7                                   | 98,2  | 95                                       | 95,7  | 85,4                                | 90,2  | 8,5                                                                | 17,7  |

Tabela 3.9. Wyniki uzyskane dla analizie ograniczonej ilości gradientów.



Rys. 3.5. Graficzna interpretacja wyników przedstawionych w tabeli 3.9

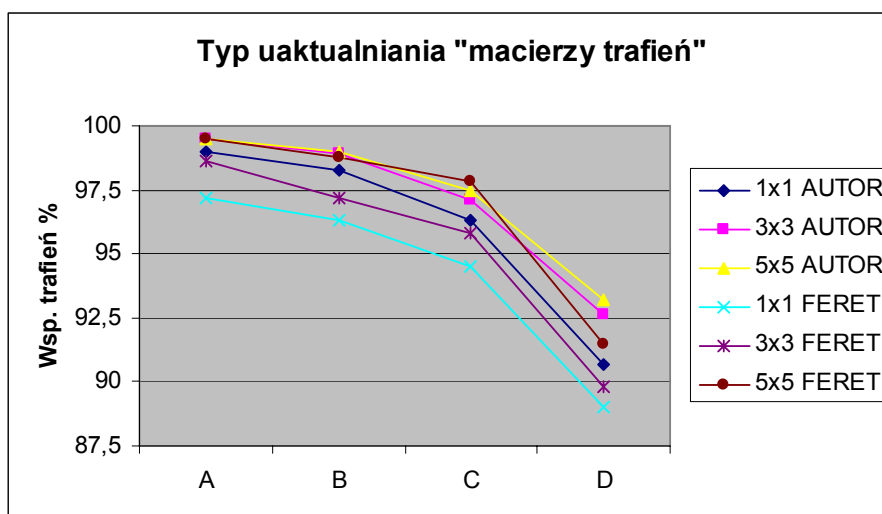
(A, B, C, D – oznaczenia kolumn z tabeli 3.9)

Wyniki są bardzo zbliżone do siebie. Nie można jednoznacznie stwierdzić, że wprowadzenie dodatkowego warunku obciążenia histogramu wpływa negatywnie bądź pozytywnie na efektywność. Jednak we wszystkich zestawieniach najlepiej (choć nieznacznie) wypadły te, dla których warunek działania ustawiony był na  $G \geq 1$ . Oznacza to, że można całkowicie pominąć ten etap usprawniania systemu.

Parametr określający sposób aktualizowania MT (macierzy trafień) podczas analizy macierzy gradientów w założeniach miał mieć znaczny wpływ na efektywność całego systemu. Tą tezę oparto na obserwacji, że standardowe rozwiązanie polegające na uaktualnianiu jednym punktem budziło dużo kontrowersji związanych z niedokładnością obliczeń spowodowaną szumem i charakterem danych wejściowych, często nienajlepszej jakości. Tabela 3.10 i rysunek 3.6 przedstawia wyniki badań nad tym parametrem.

|     | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|-----|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|     | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|     | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| 1x1 | 97,2                                     | 99    | 96,3                                   | 98,3  | 94,5                                     | 96,3  | 89                                  | 90,7  | 7,3                                                                | 14,5  |
| 3x3 | 98,6                                     | 99,5  | 97,2                                   | 98,9  | 95,8                                     | 97,1  | 89,8                                | 92,6  | 8,7                                                                | 17,3  |
| 5x5 | 99,5                                     | 99,5  | 98,8                                   | 99    | 97,8                                     | 97,5  | 91,5                                | 93,2  | 10,2                                                               | 19,2  |

Tabela 3.10. Wyniki uzyskane dla różnych typów uaktualnień MT.



Rys. 3.6. Graficzna interpretacja wyników przedstawionych w tabeli 3.10

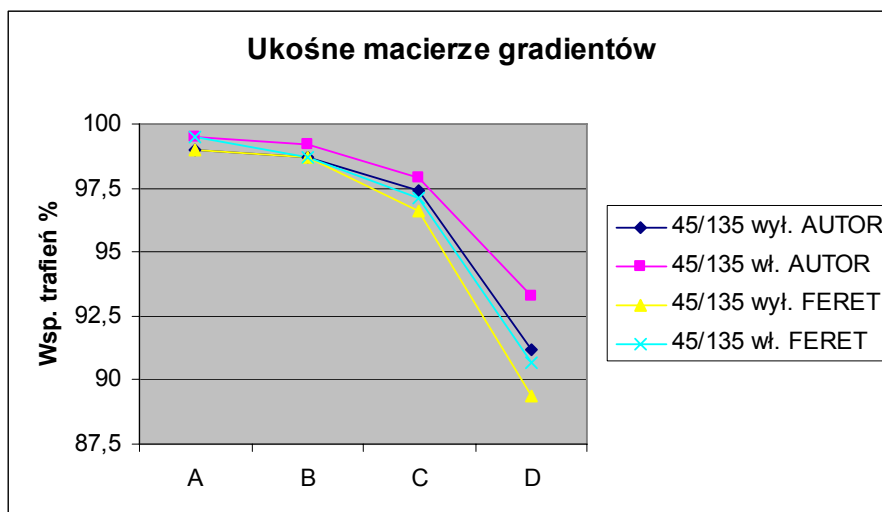
(A, B, C, D – oznaczenia kolumn z tabeli 3.10)

Typy uaktualnień są opisane w rozdziale 2.4, a tutaj dla ułatwienia nazwano je  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ . Wyniki jednoznacznie wskazują, że użycie zmodyfikowanego sposobu uaktualniania macierzy trafień generuje lepsze wyniki, przy czym sposób  $5 \times 5$  wykazuje minimalnie lepsze rezultaty niż  $3 \times 3$ . Wykonywanie tych dodatkowych, powierzchniowych uaktualnień ma swoje odbicie w czasie wykonania. Jest różnica ta jest akceptowalna szczególnie w obliczu zyskania kilku dodatkowych punktów procentowych efektywności.

Jednym z nowatorskich rozwiązań w udoskonalaniu bazowej metody jest zastosowanie ukośnych macierzy gradientów. Umożliwiają analizę dodatkowych danych płynących z gradientów w kierunkach  $45^\circ$  i  $135^\circ$ , a nie jedynie w  $0^\circ$  i  $90^\circ$ . Jest to szczególnie istotne przy skalowaniu do najniższej rozdzielczości, gdzie ilość danych do analizy jest drastycznie obniżona. Każda więc dodatkowa ich ilość jest wskazana. Tabela 3.11 i rysunek 3.7 ukazują wpływ ukośnych macierzy gradientów na wynik działania algorytmów.

|                                  | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|----------------------------------|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|                                  | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|                                  | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| <b>Macierze ukośne wyłączone</b> | 99                                       | 99    | 98,7                                   | 98,7  | 96,6                                     | 97,4  | 89,4                                | 91,2  | 7,7                                                                | 14,6  |
| <b>Macierze ukośne włączone</b>  | 99,5                                     | 99,5  | 98,7                                   | 99,2  | 97,1                                     | 97,9  | 90,7                                | 93,3  | 9,7                                                                | 19,4  |

Tabela 3.11. Wyniki uzyskane dla włączonej i wyłączonej ukośnej macierzy gradientów.



Rys. 3.7. Graficzna interpretacja wyników przedstawionych w tabeli 3.10

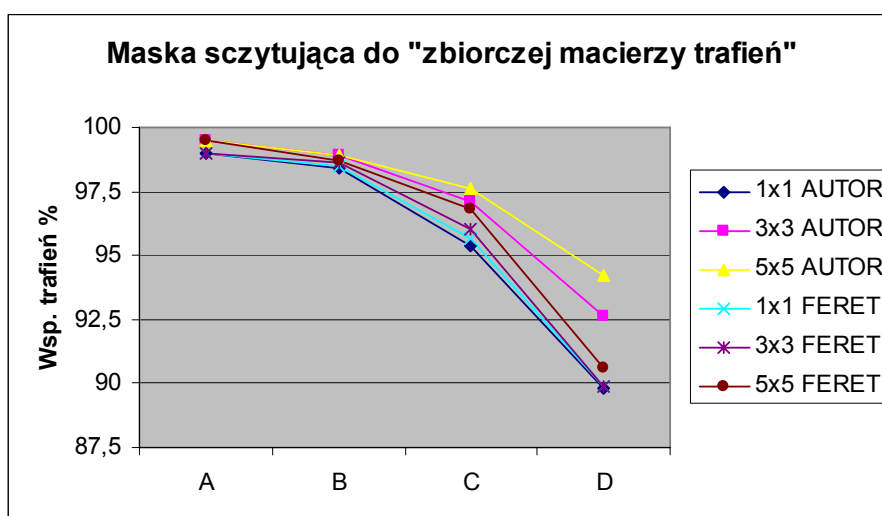
(A, B, C, D – oznaczenia kolumn z tabeli 3.10)

Analizując powyższe wyniki wpływu ukośnych macierzy gradientów można jednoznacznie stwierdzić, że dają one dodatkową użyteczną informację, więc sens jej wykorzystania jest uzasadniony. Czas wykonywania wzrósł o ok. 30%, lecz i tak ciągle jest sporo poniżej 25 klatek na sekundę.

Ostatnim parametrem dla którego przeprowadzone zostały badania jest wielkość maski szczytującej do ostatecznej ZMT (zbiorczej macierzy trafień). Sposób działania opisano w rozdziale 2.4, warto jednak przypomnieć, że jej zastosowanie wiąże się z dodatkowym uniewrażliwieniem na szum obrazów cyfrowych. Przypomina w założeniu filtr medianowy, więc ma niekorzystny wpływ na szybkość działania. Tabela 3.12 i rysunek 3.8 ukazują wyniki dla włączonej jak i wyłączanej maski szczytującej.

|     | Najlepszy wynik uzyskany w testach (w %) |       | Średnia z 10 najlepszych wyników (w %) |       | Średnia ze 100 najlepszych wyników (w %) |       | Średnia ze wszystkich wyników (w %) |       | Średni czas wykonania (w ms)<br>R=45x80 (AUTOR)<br>R=45x40 (FERET) |       |
|-----|------------------------------------------|-------|----------------------------------------|-------|------------------------------------------|-------|-------------------------------------|-------|--------------------------------------------------------------------|-------|
|     | A                                        |       | B                                      |       | C                                        |       | D                                   |       | E                                                                  |       |
|     | FERET                                    | AUTOR | FERET                                  | AUTOR | FERET                                    | AUTOR | FERET                               | AUTOR | FERET                                                              | AUTOR |
| 1x1 | 99                                       | 99    | 98,5                                   | 98,4  | 95,7                                     | 95,4  | 89,9                                | 89,8  | 7,8                                                                | 14    |
| 3x3 | 99                                       | 99,5  | 98,6                                   | 98,9  | 96                                       | 97,1  | 89,9                                | 92,6  | 8,7                                                                | 16,9  |
| 5x5 | 99,5                                     | 99,5  | 98,7                                   | 98,9  | 96,8                                     | 97,6  | 90,6                                | 94,2  | 9,7                                                                | 20,1  |

Tabela 3.12. Wyniki uzyskane dla różnych masek szczytujących do ZMT



Rys. 3.8. Graficzna interpretacja wyników przedstawionych w tabeli 3.12

(A, B, C, D – oznaczenia kolumn z tabeli 3.12)

Teoretyczne rozważania potwierdziły przeprowadzone testy. Maski 5×5 okazała się zdecydowanym liderem wśród wszystkich zaproponowanych. Szczególną cechą wartą zauważenia jest jej dobry wynik dla wszystkich ustawień (dla bazy autorskiej). Daje to niejako gwarancję, że ten parametr jest dobrze wyregulowany, a poprawiać można inne. Czas wykonywania jest zauważalnie dłuższy dla maski 5×5, ale ciągle poniżej granicy 25 klatek na sekundę.

### 3.4. Porównanie z wynikami konkurencyjnych systemów.

W celu przybliżenia jak prezentują się wyniki opisywanego w tej rozprawie systemu na tle konkurencyjnych systemów, poniżej zaprezentowane osiągnięcia dwa najlepszych obecnie systemów detekcji twarzy na obrazach. Celowo zostaną rozdzielone, ponieważ w przypadku Face Tracker nie ma możliwości uruchomienia demo na komputerze klasy PC, a jedynie w aparatach cyfrowych, przez co została opracowana inna metoda porównania.

#### **RealTime Face Detector opracowany przez Fraunhofer Institut Integrierte Schaltungen**

Ten system opracowany na niemieckiej uczelni jest jednym z najskuteczniejszych jakie dostępne są do przetestowania na komputerach klasy PC. Prace rozwojowe ciągle trwają, a na stronie instytutu [RTFD] ciągle pojawiają się nowe aktualizacje w formie pliku do ściągnięcia. Interfejs programu, przedstawiony na rys. 3.9, w którym zaszyte są algorytmy detekcji twarzy, jest intuicyjny, jednak nie umożliwia jakiegokolwiek obserwacji etapów pośrednich, przez co nie jest ciekawym narzędziem do analizy kroków pośrednich. Przeprowadzając badania na bazach przygotowanych do testowania algorytmu opracowywanego na potrzeby tej rozprawy, czyli autorskiej i FERET, stwierdzono, że niemiecki algorytm jest skuteczny, jednak ma kilka dyskwalifikujących wad. Wynik tych ostatnich widoczny jest na rys. 3.9. Pojawiają się zarówno błędne detekcje, jak i nie wykrycie twarzy.

Do zalet systemu należy zaliczyć:

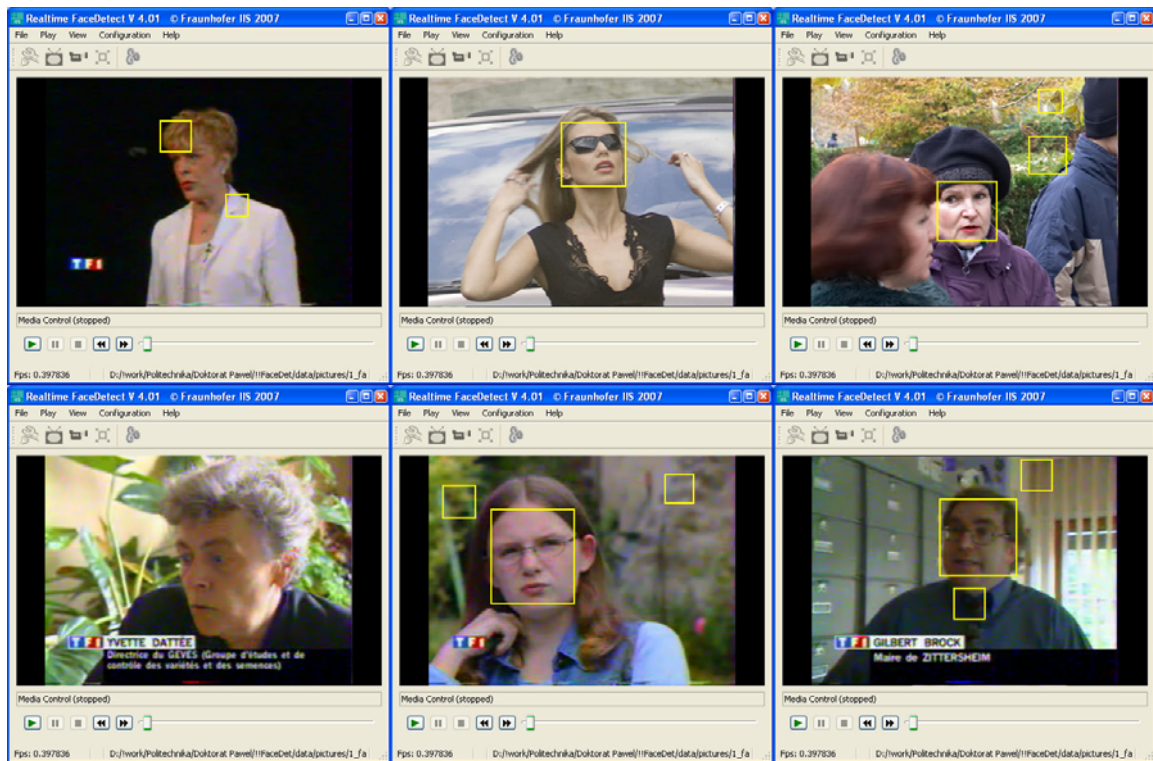
- stosunkowo wysoki współczynnik detekcji,
- detekcja wielu twarzy
- mała wielkość minimalnej twarzy,
- możliwość pracy w trybie tracking,
- dobra implementacja programowa w środowisku Windows.

Podczas badań wykryto następujące wady:

- długi czas detekcji (ok. 4 sek)

- błędne weryfikacje „face/no face”.

Należy jednak dodać, że system ma naprawdę wysoki współczynnik detekcji, sięgający 97-98% przy obu testowanych bazach. Daje to przesłanki do kontynuowania badań nad doskonaleniem szybkości, która jest jednym z najpoważniejszych wad, w tym momencie wręcz dyskwalifikującą.



Rys. 3.9. Interfejs użytkownika programu RealTime FaceDetektor wraz z przykładami jego błędnego, jak i poprawnego działania

### Face Tracker opracowany przez firmę FotoNation

Testy systemu Face Tracker dokonano przy użyciu cyfrowego aparatu fotograficznego Nikon Coolpix S7 wyposażonego w matrycę 7mln pikseli. Użytkownik ma możliwość włączenia funkcji Face Tracker w celu wspomoczenia uzyskiwania lepszej jakości zdjęć. Po jej aktywowaniu następuje przejęcie przez aparat wszystkich funkcji związanych z wyborem balansu bieli, ostrości, użycia lampy błyskowej jak również jej siły. Wszystko odbywa się automatycznie. Producent bardzo chwali możliwości systemu, i choć jako ciekawy gadżet spełnia swoją rolę budząc entuzjazm przeciętnego użytkownika, jego praca w rzeczywistości nie jest tak efektywna jak zakłada producent.

Do zalet Face Tracker możemy zaliczyć:

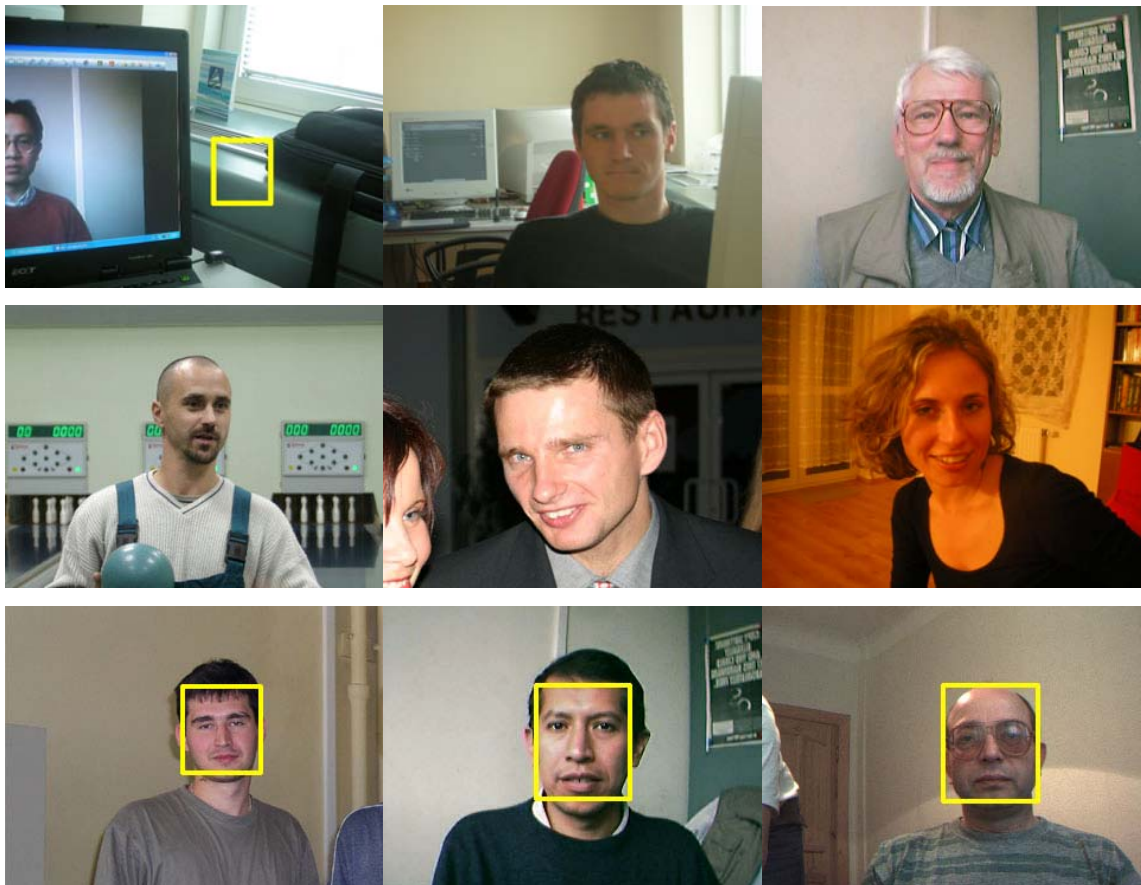
- skalarność i integracja w hardware aparatu cyfrowego,

- praca w orientacji poziomej i pionowej,
- możliwość detekcji twarzy w dużym zakresie obrotu (jednak jedynie w trybie tracking, gdyż pierwsze wykrycie twarzy musi nastąpić w pozycji zbliżonej do centralnej)

Wykrytymi wadami Face Tracker są:

- konieczność wstępnego wykrycia twarzy w jej centralnej pozycji (bez obrotu),
- duża wrażliwość na niedoświetlenie twarzy i światło kierunkowe,
- szybkość ok. 0,5 klatki na sekundę (przy deklarowanych przez producenta 30 kl./sek.),
- ciągła praca mechanizmu ostrzenia powodująca duże zużycie baterii,
- okazjonalne błędne przyjęcia,
- słaba ogólna użyteczność tej opcji w aparacie, ponieważ wydłużenie do kilku sekund wykonania pojedynczego zdjęcia jest dyskwalifikujące.

Poniżej, na rysunku 3.10, przedstawiono błędy działania systemu. Jeśli nie ma żółtej ramki to znaczy, że system nie wykrył żadnej twarzy.



Rys. 3.10. Przykład poprawnego i błędnego działania systemu Face Tracker

### 3.5. Podsumowanie i wnioski

Celem przeprowadzonych eksperymentalnych badań było przetestowanie opracowanej metody i wchodzących w jej skład algorytmów pod kątem efektywności. Przeprowadzono dogłębną analizę uzyskanych wyników, ze zwróceniem szczególnej uwagi na zasadność zastosowania dodatkowych ulepszeń algorytmów bazowych. Wyznaczono również zestaw parametrów dla których metoda zwraca najlepsze rezultaty. Wyniki udowodniły, że idea jednoczesnego uwzględnienia informacji o owalnym kształcie twarzy i procedury zmiennego promienia, może być podstawą do budowy skutecznej metody lokalizacji twarzy na obrazach cyfrowych, a dodatkowo użycie informacji o kolorze skóry zwiększy jej skuteczność.

Do najważniejszych wniosków należą:

- Opracowane algorytmy oferują bardzo wysoki współczynnik wyodrębniania kandydatów do bycia twarzą, zbliżający się do granicy 100%,
- Algorytmy doskonale potrafiły analizować profesjonalną bazę FERET (USA), jak i autorską, przygotowaną specjalnie na potrzeby tej rozprawy,
- Obie bazy testowych obrazów dostarczyły wyników, które w większości są zbieżne ze sobą – fakt ten, pozwala stwierdzić, że są one wiarygodne,
- Szybkość działania algorytmów wynosząca średnio ok. 15ms jest jednym z głównych atutów całego systemu, ponieważ rekomenduje go do zastosowań w systemach czasu rzeczywistego,
- Zaskakujący wysoki okazał się wynik efektywności już dla pierwszego generowanego kandydata, a wynoszący ponad 94% dla obu baz – pozwala stwierdzić, że system ten można użyć jako lokalizator twarzy bez dedykowanego modułu twarz/„nie twarz”; dodatkowo, gdy zwiększy się tolerancję przyjmowania jako „trafiony” położenia twarzy i jej średnicy do wzorca z wartości 15% do 20%, wynik dla pierwszego kandydata rośnie do ponad 97%,
- Głębsza analiza ustawień parametrów nie wskazała szczególnego ich zestawu dla których wynik znacząco by się wyróżniał, co w połączeniu z bardzo wysokim średnim współczynnikiem efektywności pracy algorytmów pozwala stwierdzić, że został on dobrze przemyślany, dostatecznie przetestowany i nie zachodzi konieczność dalszego jego badania,
- Algorytmy mają możliwość pracy z obrazem w skali szarości, jednak tylko informacja o pełnym kolorze wykazuje wysoki współczynnik lokalizacji,
- Szereg wprowadzonych usprawnień wykazało swoją przewagę nad bazowymi rozwiązaniami – należą do nich:

- zmienny promień poszukiwań,
- użycie komponentu  $I_2$  ze zmodyfikowanej przestrzeni  $III_2I_3$ ,
- ukośne macierze gradientów,
- powiększony obszar sczytywania przy wpisywaniu do ZMT,
- skalowanie do wielkości  $45 \times 80$  pikseli ( $45 \times 40$  dla bazy FERET),
- obcięcie histogramu,
- uaktualnianie wyników trafień przy pomocy masek  $3 \times 3$  i  $5 \times 5$  pikseli.

Należy również dodać, że w implementacji w języku C++ wprowadzono szereg optymalizacji kodu źródłowego niwelującego nadmiarowe obliczenia i powodujących znaczne przyspieszenie wykonywania całego programu. Olbrzymia ilość przeprowadzonych badań sięgająca miliona pojedynczych wyodrębnień, daje podstawę do stwierdzenia, że metoda została rzetelnie przetestowana, a zaproponowane nowe rozwiązania i usprawnienia poprawiają dotychczasowe wyniki. Całą procedurę testową przeprowadzono z zachowaniem dbałości o sporządzenie odpowiedniej dokumentacji potwierdzającej jej prawdziwość.

## 4. Opis programu badawczego Face Localizator

Podczas rozwoju prac nad algorytmami opisanymi w poprzednich dwóch rozdziałach zaistniała potrzeba użycia platformy programistycznej umożliwiającej sprawne testowanie nowych rozwiązań. W tym celu rozważono wykorzystanie dwóch środowisk: Matlab i C++. Niewątpliwie oba posiadają swoje wady i zalety. Zasadniczą różnicą jest stopień ułatwień w pracy z grafiką oferowany przez oba pakiety. Ogromną przewagę posiada Matlab, za wykorzystaniem którego przemawia dodatkowo duże wsparcie w akademickich środowiskach badawczych. Jednak jego prędkość pozostawia bardzo wiele do życzenia, a w stosunku do języka C++ jest to różnica bardzo znacząca. Jednocześnie wymagania współczesnego, komercyjnego rynku są nastawione na porównywanie wyników płynących z implementacji w tych językach programowania, w których docelowo algorytmy będą pracować, a które są przenoszalne pomiędzy popularnymi systemami operacyjnymi typu Windows, MacOS czy Linux. Dlatego po krótkim okresie pracy w Matlabie zdecydowano się na środowisko C++. Wszystkie wyniki uzyskane podczas eksperymentalnych badań opisanych w rozdziale 3 zostały otrzymane przy użyciu programu napisanego w programie Borland C++ Builder 5.

### 4.1. Założenia wstępne i wymagania sprzętowe

Lokalizator obszarów potencjalnie zawierających twarz jest pierwszym z kroków niemal każdego nowoczesnego detektora, a przez to również jednym z trybów większych systemów rozpoznawania ludzi. Głównym więc czynnikiem decydującym o jego jakości jest w równej mierze skuteczność, jak i szybkość działania. Dlatego poczyniono szereg zabiegów natury programistycznej mających na celu optymalizację algorytmów pod kątem czasu wykonywania (integracja, optymalizacja cyklicznych operacji np. skalowania). Dodatkowo nowe pomysły opisane w rozdziale 2.4 przyczyniły się do poprawy szybkości (wykorzystanie komponentu *I2*, skalowanie do wielkości  $45 \times 80$  pikseli, zmienny promień poszukiwań, MT i ZMT) w stosunku do bazowej metody poszukiwania obiektów owalnych.

Wyniki uzyskane podczas testów opracowanych algorytmów, opisane w rozdziale 3, zachęciły do stworzenia dwóch wersji programu:

- Pełnoprawnego lokalizatora twarzy – wynikiem są współrzędne i promień pierwszego otrzymanego wyniku,
- Lokalizatora zgrubnego – zwraca współrzędne i promień pierwszych pięciu otrzymanych wyników lokalizacji obszarów potencjalnie zawierających twarz.

Pierwszy z nich to nic innego jak system lokalizacji twarzy, który mimo braku algorytmów decyzyjnych „twarz/nie twarz” może być zastosowany w prostych systemach rozpoznawania jako etap lokalizacji. Takie śmiałe posunięcie podparte jest bardzo dobrymi współczynnikami trafień w pierwszej próbie i oczywiście bazuje na przesłaniu, że na obrazie twarz się znajduje.

Drugi tryb, czyli lokalizator obszarów potencjalnie zawierających twarz zwraca pięć wyników dla każdego obrazu. Dodatkową opcją jest możliwość przesłania macierzy zawierającej obszary, gdzie twarz nie ma być poszukiwana, co umożliwia poszukiwanie kolejnych twarzy.

Obie, powyższe wersje programu są przekompilowane przy użyciu pakietu Borland Builder C++ 5, a ich wynikiem jest wykonywalny program, uruchamiany z linii poleceń, który z powodzeniem może być użyty przez inne systemy np. w środowisku Matlab. Spełnione jest też założenie, że wykonanie programu w obu wersjach powinno być całkowicie „przezroczyste” dla zwykłego użytkownika, a dostęp do opcji dostępny jedynie dla wyszkolonego administratora systemu i temu służy opcja przesłania dodatkowych parametrów opisanych w następnym podrozdziale.

Zauważono też, że warto poświęcić czas na przygotowanie wersji z pełnym interfejsem użytkownika, który może posłużyć zarówno do badania efektywności dla różnych parametrów, jak również dać możliwość nauczania studentów tajników działania algorytmów opisanych w niniejszej rozprawie.

Określono następujące wymagania sprzętowe systemu:

- system Microsoft Windows 95 lub nowszy,
- karta grafiki: VGA (akcelerator nie jest wymagany),
- pamięć RAM: 128 MB,
- dysk twardy: 10 MB,
- opcjonalnie: kamera internetowa (dla wersji z pełnym GUI – ang. *Graphics User Interface*).

## 4.2. Omówienie możliwości

W pierwszej kolejności zostanie omówiony program w wersji wykonywalnej z linii poleceń dla obu wersji o których wspomniano w powyższym podrozdziale, czyli jako pełnoprawnego lokalizatora i lokalizatora obszarów potencjalnie zawierających twarz, a następnie wersję z pełnym interfejsem użytkownika.

### Wersja wykonywalna z linii poleceń

Idea wykorzystania linii poleceń jest efektem przemyślenia, że opracowany lokalizator można z łatwością wykorzystać w innych systemach, niekoniecznie napisanych w języku C++. Wiersz poleceń ma również dodatkową zaletę w postaci szybkiego czasu wykonywania. Istnieje oczywiście również możliwość pełnej integracji kodu źródłowego opisywanych w niniejszej rozprawie algorytmów bezpośrednio w większym systemie rozpoznawania ludzi bądź innym, gdzie lokalizacja twarzy jest ważnym elementem funkcjonowania całości.

Zasada użycia obu wersji algorytmu w wersji dla linii poleceń jest bardzo prosta i ogranicza się do napisania formuły według zasad opisanych poniżej:

- dla pełnoprawnego lokalizatora

faceloc\_1.exe obraz\_wej.jpg wynik.txt <F> <is45> <G> <M> <R> <HL> <HH> (4.1)

- dla lokalizatora pięciu obszarów potencjalnie zawierających twarz

faceloc\_5.exe obraz\_wej.jpg wynik.txt maska.txt <is45> <G> <M> <R> <HL> <HH> (4.2)

gdzie:

- obraz\_wej.jpg – obraz wejściowy w formacie JPG,
- wynik.txt – plik tekstowy zawierający wynik działania programu,
- maska.txt – plik tekstowy zawierający odpowiednio opisane obszary, gdzie nie należy poszukiwać twarzy.
- <is45> <G> <M> <R> <HL> <HH> - parametry opisane w tabeli 4.1.

Tabela 4.1. Parametry programu *facedet\_lok.exe* i *facedet\_gen.exe* wykonywanego z linii poleceń.

| Nazwa parametru | Możliwe stany        | Opis                                                                            |
|-----------------|----------------------|---------------------------------------------------------------------------------|
| <F>             | 1,2,3,4,5,6,7,8,9,10 | ilość twarzy do poszukiwania (jedynie w wersji <i>faceloc_1.exe</i> )           |
| <is45>          | 0,1                  | użycie macierzy G45 i G135                                                      |
| <G>             | 0,1,2,3,4,5          | odcięcie gradientu, np. 2 → gradienty poniżej wartości 2 nie są brane pod uwagę |
| <M>             | 3,5,7                | obszar sczytywania dla danego piksela, np. 5 → maska 5×5 pikseli                |
| <R>             | 80,120,160           | rozdzielczość do której obraz jest skalowany, np. 80 → 45×80 pikseli            |
| <HL>            | 0÷70                 | odcięcie histogramu – dolne                                                     |
| <HH>            | 70÷255               | odcięcie histogramu – górne                                                     |

Obrazem wejściowym może być dowolny plik graficzny zapisany w formacie JPG. Obie wersje posiadają możliwość użycia parametrów opisanych w tabeli 4.1. Należy

podkreślić, że parametry są jedynie opcją i jeśli nie zostaną podane, przypisane im zostaną wartości domyślne, uznane w niniejszej rozprawie jako optymalne, ale niekoniecznie trafione dla zastanych warunków.

Dodatkowo w wersji *faceloc\_5.exe* istnieje możliwość dostarczenia listy obszarów, które algorytm ma pomijać przy poszukiwaniu – służy do tego plik *maska.txt*. Zabieg ten ma za zadanie umożliwienie poszukiwania kolejnych twarzy po wykryciu poprzedniej. Zachodzi wtedy potrzeba wyłączenia dostępu do obszaru na którym twarz już jest zlokalizowana. Dodatkowym atutem jest możliwość zamaskowania dowolnego obszaru obrazu, na którym na pewno nie ma twarzy (np. dla statecznie umocowanej kamery w części obrazu może występować ściana).

Plik *maska.txt* zawiera linie tekstu sformatowane w następujący sposób:

$$Y \ X \ Y_s \ X_s \quad (4.3)$$

gdzie:

- $Y$  – współrzędna „y” górnego lewego wierzchołka opisującego wykluczany obszar,
- $X$  – współrzędna „x” górnego lewego wierzchołka opisującego wykluczany obszar,
- $Y_s$  – wielkość w płaszczyźnie pionowej wykluczanego obszaru,
- $X_s$  – wielkość w płaszczyźnie poziomej wykluczanego obszaru.

Każda z wyżej wymienionych wartości jest rozdzielona znakiem spacji, co ułatwia dostęp do nich z programów zewnętrznych.

Przykładowy plik tekstowy zawierający maskę wygląda następująco:

*maska.txt*

|                |
|----------------|
| 80 100 80 60   |
| 180 240 120 90 |
| 88 189 100 75  |

W celu przekazania wyników działania wyżej opisywanych programów tworzony jest plik *wynik.txt*. W związku z tym, że obie wersje programu posiadają odmienne wytyczne odnośnie wyniku lista ta różni się w specyfice odbioru, choć zasada wpisywania kolejnych linii jest identyczna i wygląda następująco:

$$Y \ X \ R \quad (4.4)$$

gdzie:

- $Y$  – współrzędna „y” środka twarzy/obszaru zawierającego twarz,
- $X$  – współrzędna „x” środka twarzy/obszaru zawierającego twarz,
- $R$  – wielkość promienia wykrytej twarzy/obszaru zawierającego twarz.

Każda z wyżej wymienionych wartości jest rozdzielona znakiem spacji co ułatwia szybki i nieomylny do nich dostęp. Dodatkowo w ostatniej linii, tuż za wszystkimi wpisanymi twarzami umieszczony jest czas wykonania programu – podany w milisekundach. Informacja ta jest często przydatna do określenia sprawności użytych parametrów, jak i do szybkiego testu danego komputera.

Poniżej przedstawiony został przykładowy wynikowy plik tekstowy:

*wynik.txt*

```
96 150 38 40
107 154 38 30
102 183 38 30
96 167 38 40
160 227 43 40
48
```

### Wersja z pełnym interfejsem użytkownika

Podczas prac nad rozwojem algorytmu zauważono, że warto poświęcić czas na przygotowanie dodatkowej wersji programu realizującego algorytmy opisywane w niniejszej rozprawie, a posiadającego pełny interfejs użytkownika. Rozwiązanie to pomagało obserwować wyniki, umożliwiając szybkie reagowanie na powstające problemy, czym przyczyniło się do poprawy jakości ostatecznych algorytmów. Dodatkowo powstało narzędzie pozwalające prezentować poszczególne kroki algorytmu w bardzo przystępny sposób, jak również dające możliwość efektywniej dobierać optymalne parametry pracy systemu.

Okno programu Face Localizator bezpośrednio po wczytaniu wygląda w sposób przedstawiony na rysunku 4.1.



Rys. 4.1. Okno główne programu Face Localizator

W lewej części okna dostępnych jest siedem przycisków, dających dostęp do wszystkich opcji programu:

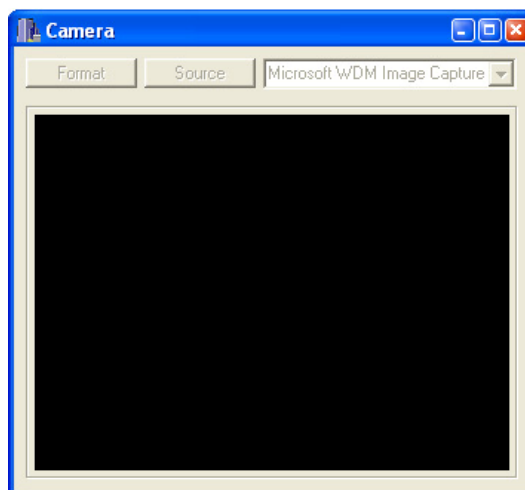
- File... – okno wczytania plików,
- Avi File... – okno wczytania plików multimedialnych,
- Camera Start – okno obsługujące kamerę internetową,
- Grab Frame – inicjuje przechwytywanie klatek z kamery internetowej,
- Options – okno opcji,
- Show Face(s) – okno prezentacji wyodrębnionej/nych twarzy,
- Results – okno prezentujące tabelę wyników.

Pierwsze dwa przyciski nie wymagają opisu, gdyż uruchamiają standardowe okna do wyboru plików. Należy nadmienić, że akceptowalne pliki to:

- dla okna „File...” – JPG i BMP,
- dla okna „AVI File...” – AVI (nieskompresowany).

Po wczytaniu obrazu następuje automatyczne uruchomienie algorytmów lokalizacji. Gdy zaznaczona zostanie większa ilość obrazów (np. w sytuacji, gdy sekwencja wideo jest podzielona na klatki) możliwe jest uruchomienie automatycznego lokalizowania przy pomocy klawisza „Auto skip”. Natomiast animacja jest automatycznie przetwarzana, klatka po klatce.

Kolejny przycisk – „Camera Start” – uruchamia okno obsługujące kamerę internetową w którym można wybrać urządzenie (jeśli jest ich kilka podłączonych), jak również dostosować parametry jej pracy do optymalnych. Okno to wygląda następująco:



Rys. 4.2. Okno obsługi kamery internetowej

Przycisk „Grab Frame” służy do ściągania pojedynczej klatki, bądź sekwencji obrazów pozyskanych z kamery internetowej. Lokalizacja przebiega w sposób automatyczny.

Jest wartym podkreślenia, że na szybkich komputerach (procesor główny taktowany powyżej 2 GHz) osiągana jest prędkość 25 klatek na sekundę, co jest wynikiem wyjątkowym.

Kolejny przycisk – „Options” – to najważniejsza opcja z punktu widzenia zasadności stworzenia wersji z interfejsem użytkownika. Ukryte są pod nim możliwości obserwacji kolejnych kroków algorytmu, jak również zmiany parametrów z możliwością błyskawicznej obserwacji wpływu na efektywność systemu. Jego wygląd przedstawia rysunek 4.3.



Rys. 4.3. Okno dialogowe „Options”

W oknie „Options” istnieje możliwość ustawiania parametrów algorytmów opisywanych w niniejszej rozprawie. Realizują je suwaki, pola zaznaczeń i przyciski opisane w tabeli 4.2.

W tabeli tej zaznaczono jakim narzędziem jest realizowana możliwość zmiany danego parametru:

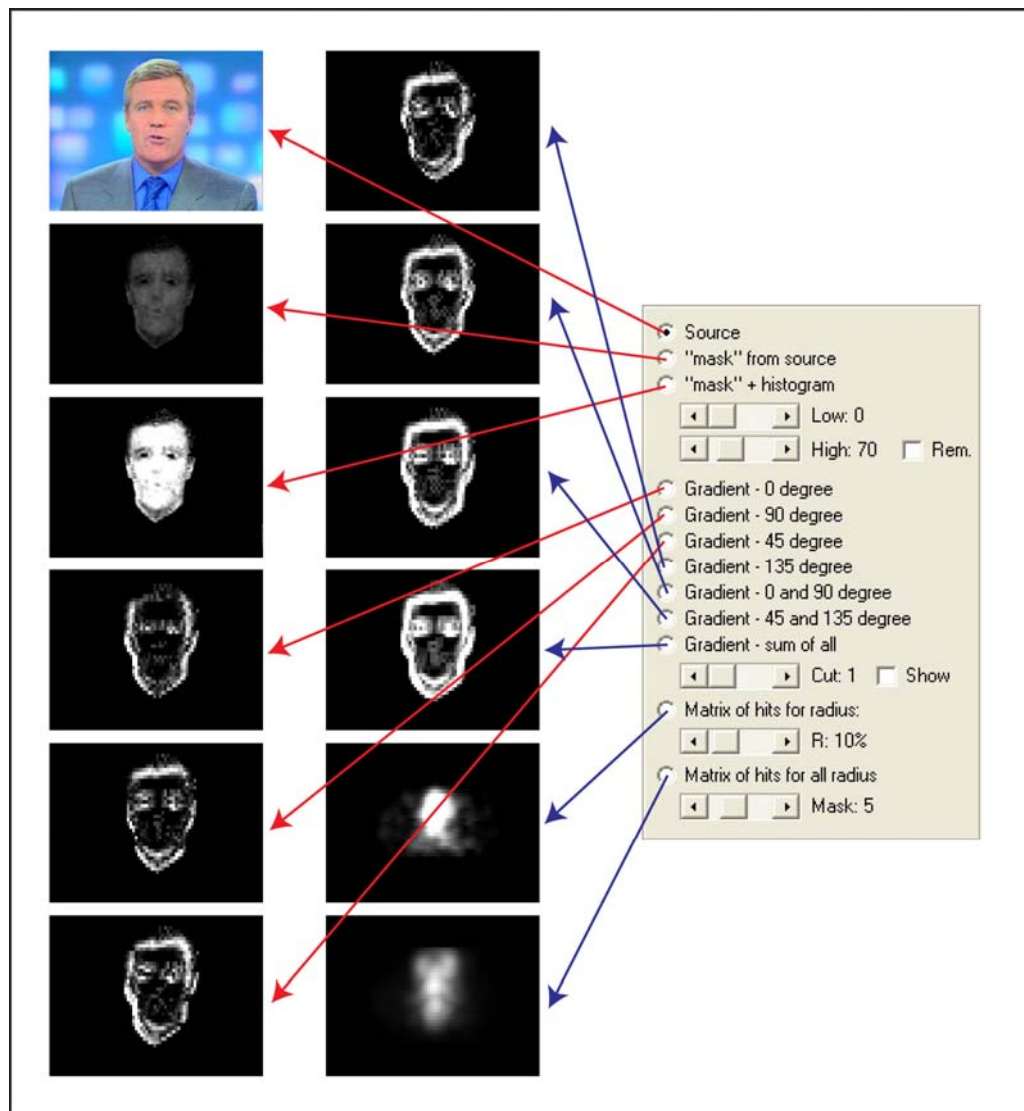
- S – ScrollBar – suwak z określonymi stanami
- C – CheckBox – pole zaznaczenia
- R – RadioButton – przycisk zmieniający widok w głównym oknie

Przycisk „Execute” służy wykonaniu pełnego algorytmu lokalizacji obiektów owalnych z aktualnie ustawionymi parametrami. Opcja ta jest przydatna przy testowaniu nowych wartości i całkowicie uniezależnia okno „Options” od okna głównego programu Face Localizator. Dodatkowo umieszczono przydatne koordynaty w płaszczyźnie pionowej i poziomej w celu ułatwienia określenia położenia piksela obrazu, nad którym aktualnie znajduje się kursor myszy.

Tabela 4.2. Opis suwaków, pól zaznaczeń i przycisków z okna „Options”

| Nazwa parametru                | Typ | Opis                                                                        |
|--------------------------------|-----|-----------------------------------------------------------------------------|
| Resol:                         | S   | Rozdzielczość do której ma być skalowany obraz                              |
| Face(s):                       | S   | Ilość twarzy do poszukiwania: 1÷10                                          |
| MoH up.                        | S   | Typ uaktualniania macierzy trafień:<br>simply, simply+                      |
| 45/135 degree                  | C   | Użycie macierzy ukośnych: TAK/NIE                                           |
| Save report                    | C   | Zapis raportu do pliku tekstowego: TAK/NIE                                  |
| Source                         | R   | Obraz wejściowy                                                             |
| „mask” from source             | R   | Obraz po przekształceniu do komponentu $I_2$<br>z przestrzeni $III_2I_3$    |
| „mask” + histogram             | R   | Obraz po obcięciu histogramu                                                |
| Low:                           | S   | Dolna wartość obcięcia histogramu: 0÷69                                     |
| High:                          | S   | Górna wartość obcięcia histogramu: 70÷255                                   |
| Rem                            | C   | Zapamiętanie nowych ustawień histogramu                                     |
| Gradient - 0 degree            | R   | Zobrazowanie macierzy gradientów dla kąta 0 stopni                          |
| Gradient - 90 degree           | R   | Zobrazowanie macierzy gradientów dla kąta 90 stopni                         |
| Gradient – 45 degree           | R   | Zobrazowanie macierzy gradientów dla kąta 45 stopni                         |
| Gradient - 135 degree          | R   | Zobrazowanie macierzy gradientów dla kąta 135 stopni                        |
| Gradient – 0/90 degree         | R   | Zobrazowanie sumy wartości macierzy gradientów<br>dla kątów 0 i 90 stopni   |
| Gradient – 45/135 degree       | R   | Zobrazowanie sumy wartości macierzy gradientów<br>dla kątów 45 i 135 stopni |
| Gradient – sum of all          | R   | Zobrazowanie sumy wartości wszystkich macierzy gradientów                   |
| Cut:                           | S   | Maska wartości gradientów do pominięcia przez algorytm                      |
| Show                           | C   | Pokazywanie ustawień „Cut”                                                  |
| Matrix of hits for radius:     | R   | Zobrazowanie MT dla wartości ustawionych w „R”                              |
| R:                             | S   | Dla jakiego promienia pokazywać „macierz trafień”: 7÷50%                    |
| Matrix of hits for all radius: | R   | Zobrazowanie ZMT                                                            |
| Mask:                          | S   | Maska, którą aktualizuje macierz trafień: 3, 5, 7                           |

W celu dokładniejszego przyjrzenia się efektowi działania przycisków typu *RadioButton* oznaczonych w tabeli symbolem R stworzono rysunek 4.4. Należy zwrócić uwagę, że okno „Options” działa symultanicznie z głównym i w związku z tym może być widoczne cały czas, dając możliwość podglądu zaznaczonej opcji nawet dla analizowanej animacji przez główne okno.



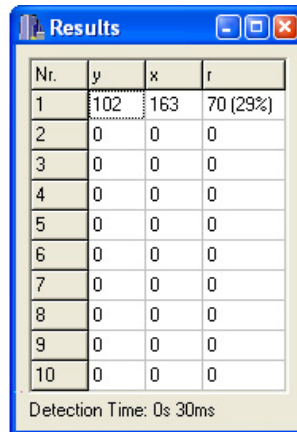
Rys. 4.4. Przykład działania podglądów w oknie „Options”

Kolejnym przyciskiem umieszczonym w oknie głównym jest „Show Face/s”. Jego wcisnięcie powoduje wyświetlenie okna przedstawionego na rysunku 4.5 przedstawiającego podgląd wyznaczonych twarzy, których ilość (1 lub 5) jest zależna od ustawionych parametrów i trybu pracy.



Rys. 4.5. Okno „Show Faces” w wersji dla pełnoprawnego lokalizatora twarzy (a)  
jak i lokalizatora 5 obszarów potencjalnie zawierających twarz (b)

Ostatnim przyciskiem jest „Results”. Ukryta pod nim tabela zawiera numeryczne zobrazowanie wyniku. Pozwala na przyjrzenie się wynikom, ich weryfikację lub ewentualne porównanie z oczekiwaniami. Jego wygląd ukazuje rysunek 4.6.



| Nr. | y   | x   | r        |
|-----|-----|-----|----------|
| 1   | 102 | 163 | 70 (29%) |
| 2   | 0   | 0   | 0        |
| 3   | 0   | 0   | 0        |
| 4   | 0   | 0   | 0        |
| 5   | 0   | 0   | 0        |
| 6   | 0   | 0   | 0        |
| 7   | 0   | 0   | 0        |
| 8   | 0   | 0   | 0        |
| 9   | 0   | 0   | 0        |
| 10  | 0   | 0   | 0        |

Detection Time: 0s 30ms

Rys. 4.6. Okno „Results”

### 4.3. Podsumowanie i wnioski

Przedstawione w tym rozdziale wersje programu „Face Localizator” ukazały sposób implementacji algorytmów opracowanych na potrzeby tej rozprawy. Wszystkie metody wykorzystane w implementacji zostały napisane w postaci funkcji, umożliwiając łatwą przenoszalność do innych systemów biometrycznych.

Przygotowano dwie wersje programu:

- dla trybu z linii poleceń
- z pełnym interfejsem użytkownika.

Obie zostały przemyślane pod kątem optymalności i funkcjonalności. Wersja z linii poleceń jest dedykowana użyciu w innych systemach, m. in. w Matlabie. Szybkość działania, łatwość przekazywania własnych parametrów, generowanie prostego w przetwarzaniu pliku wynikowego to podstawowe zalety tego rozwiązania. Druga, posiadająca własne GUI, może zostać z powodzeniem wykorzystana do nauczania studentów sposobu działania opracowanych algorytmów. Nic nie stoi na przeszkodzie, by z jego pomocą w jeszcze głębszym stopniu przetestować algorytmy. Dla przykładu można użyć inne bazy czy dobrać inne parametry.

Cały system przygotowany w środowisku Borland C++ Builder 5 może być niewielkim nakładem pracy przeniesiony na każdą platformę systemową, np. Linux, MacOS.

## Zakończenie

W pracy poruszony został problem opracowania metody lokalizacji twarzy na obrazach cyfrowych. Na łamach czterech rozdziałów przedstawiono zagadnienie automatycznej lokalizacji, zdefiniowano cel i tezę, wybrano metody bazowe, stworzono ich udoskonalone wersje, przetestowano przy użyciu komercyjnej i autorskiej bazy danych, jak również opisano implementację w języku wysokiego poziomu.

W rozdziale 1 dokonano przeglądu i usystematyzowania światowych osiągnięć w dziedzinie detekcji twarzy na obrazach. Zwrócono uwagę na ich wielkiemu znaczeniu w systemach rozpoznawania ludzi. Przy czym, rozdział 1.1 poświęcono wprowadzeniu do problemu detekcji, omówiono dlaczego ciągle jest to aktualny temat, zaznaczono najistotniejsze cechy jakie charakteryzują system detekcji i z jakimi przeszkodami musi się zmierzyć projektant. Rozdział 1.2 zawiera przegląd, analizę, usystematyzowanie i krótką charakterystykę istniejących na świecie podejść do detekcji. Przedstawiono przy tym najnowsze trendy w tym podejście bazujące na tworzeniu dwuetapowych detektorów, zawierających generator kandydatów i weryfikator tychże. Zaznaczono przy tym, że mimo ogromnej ilości pracy naukowców na całym świecie, opracowujących bardzo wiele różnych podejść, problem detekcji jest ciągle nierozwiązany, a przez to ciągle aktualny. Z tego powodu w rozdziale 1.5 za cel pracy przyjęto opracowanie metody i realizacja algorytmów do zadań lokalizacji obszarów zawierających twarz na obrazach cyfrowych. Postawiono tezę, iż jednoczesne uwzględnienie informacji o owalnym kształcie twarzy i procedury zmiennego promienia może być podstawą do budowy skutecznej metody lokalizacji twarzy na obrazach cyfrowych, a dodatkowo użycie informacji o kolorze skóry zwiększy jej skuteczność.

W rozdziale 2 przedstawiono metodę wykorzystującą ideę użycia kształtu twarzy, jej koloru, procedury zmiennego promienia i macierzy gradientów do rozwiązania zadania lokalizacji obszarów zawierających twarz. Przy czym, w rozdziale 2.1 wprowadzono założenia niezbędne do poprawnego projektowania lokalizatora. W rozdziale 2.2 omówiono metody obróbki obrazu i metod normalizacyjnych będących bardzo dobrze znanymi i często wykorzystywanymi. Wśród nich znalazły się też te niezbędne do poprawnego działania metody wykrywania obiektów owalnych przy użyciu macierzy gradientów, opisaną dokładnie w rozdziale 2.3. Posłużyła ona za bazę do zbudowania znacznie efektywniejszego i szybszego lokalizatora obiektów owalnych szczegółowo opisaną w rozdziale 2.4. Objętość i ważność materiału zawartego w tym rozdziale dała podstawę do podziału na mniejsze podrozdziały ułatwiające jego ogarnięcie. Podczas omawiania każdego kroku metody i

usprawnień starano się udowodnić dlaczego modyfikacje są wprowadzane i jakie płyną z tego korzyści, zarówno w dziedzinie efektywności jak i czasu wykonania.

W rozdziale 3 przedstawiono eksperymentalne badania opracowanej metody lokalizacji twarzy. W rozdziale 3.1 określono główne założenia testów oraz zdefiniowano jakimi bazami danych i w jakim zakresie dokonane zostaną badania. Do ich przeprowadzenia stworzono precyzyjną procedurę testową opisaną w rozdziale 3.2, pozwalającą odtworzyć cały proces badawczy, a przez to go uwiarygodniającą. Rozdział 3.3 zawiera przedstawienie głównego procesu badawczego oraz szczegółową charakterystykę dwóch baz danych: FERET i autorskiej, wraz z wyjaśnieniem dlaczego akurat one zostały użyte. Następnie przetestowano każdą modyfikację dającą się bezpośrednio porównać z bazową. Wyniki poddano analizie po czym sprecyzowano wnioski.

W rozdziale 4 omówiono program badawczy Face Localizator będący implementacją opisywanych algorytmów w języku C++. Ze względu na bardzo dobre wyniki lokalizacji już przy pierwszym proponowanym przez algorytm obszarze zdecydowano się zaimplementować dwie wersje programu: pełnoprawnego lokalizatora twarzy i lokalizatora obszarów potencjalnie zawierających twarz generującego pięć propozycji. Omówiono ich możliwości i ograniczenia. Dodatkowo dokonano analizy możliwości programu Face Localizator w wersji z pełnym interfejsem użytkownika. Ze względu na prostotę obsługi i bogactwo opcji pozwalających podglądać etapy pracy algorytmu, zaproponowano jego użycie do celów badawczych, jak i czysto dydaktycznych.

Metoda przedstawiona w niniejszej dysertacji jest wynikiem trzyletniej pracy. Kolejne etapy badań i eksperymentów zostały opisane w publikacjach [Mas02b, Mas03, Mas06, Kuk03b, Kuk04a, Kuk05a, Kuk05b, Kuk07].

Głównymi elementami nowości w opracowanym podejściu są:

1. Integracja w jednej metodzie informacji o owalnym kształcie twarzy, jej nieprzewidywalnej wielkości i przedstawieniu w polu gradientów w czterech niezależnych kierunkach (dwie pary: pion/poziom i ukośne);
2. Zastosowanie komponentu I2 z przestrzeni I1I2I3 wraz z procedurami normalizacji, jako głównego elementu obróbki obrazów kolorowych dla różnych ras ludzkich;
3. Propozycja zbiorczej macierzy trafień do zadania poszukiwania kolejnych twarzy (do 10), bez powtórnego wyliczania poszczególnych macierzy trafień;
4. Modyfikacja wyznaczania centrum twarzy za pomocą procedury wygładzania, eliminująca niedokładność formy owalów w macierzy gradientów;

Na podstawie rezultatów przedstawionych w rozdziale 3, będących wynikiem działania algorytmów omówionych w rozdziale 2, można stwierdzić, że cel pracy został osiągnięty. Przy tym:

- 1) dokonano wyboru metody lokalizacji twarzy, która posłużyła jako baza do dalszego rozwoju w dziedzinie efektywności i czasu wykonania;
- 2) opracowano modyfikację powyższej metody i zrealizowano algorytmy przeznaczone do zadań lokalizacji obszarów zawierających twarz na obrazach cyfrowych;
- 3) na przykładzie wielu eksperymentów udowodniono, że wybrane modyfikacje, jak i dobór metody jest wystarczający do realizacji skutecznego lokalizatora obszarów zawierających twarz, przez co jest użyteczny w systemach rozpoznawania ludzi;
- 4) udowodniono przez to, że wspólne wykorzystanie informacji o kolorze, owalnym kształcie i rozmiarach twarzy jest podstawą do budowy skutecznych algorytmów lokalizacji obszarów zawierających twarz na obrazach cyfrowych.

Powyższe rezultaty są dowodem na postawioną w pracy tezę.

## Bibliografia

- [Ana03] Anagnostopoulos I., Anagnostopoulos C., Vergados D., Loumos V., Kayafas E., *A Probabilistic Neural Network for Human Face Identification based on Fuzzy Logic Chromatic Rules*, paper submitted to IEEE MED03, June 2003, Rhodes
- [Alb00] Albiol A. *A simple and efficient face detection algorithm for video database applications*. In Proceedings of the 2000 International Conference on Image Processing, 2000
- [Ami98] Amit Y., Geman D., Jedynak B. *Efficient focusing and face detection*. In H. Wechsler and J. Phillips, editors, *Face Recognition: From Theory to Applications*, NATO ASI Series F. Springer-Verlag, Berlin, 1998
- [Aug93] Augusteijn M.F., Skujca T.L. *Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology* Proc. IEEE Conf. Neural Networks, pp. 392-398, 1993
- [Ben99] Ben-Yacoub S., Fasel B., Luttin J. *Fast face detection using MLP and FFT*. In Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999
- [Ber98] Bernier O., Collobert M., Ferau R., Lemaire V., Viallet J.E., Collobert D. *MULTRAK: A System for Automatic Multiperson Localization and Tracking in Real-Time*. vol. 1, pp. 136-140, Chicago, Illinois, 1998
- [Bry00] Bryłka T. *Wyodrębnianie twarzy człowieka ze zdjęć*, Praca magisterska, Politechnika Szczecińska, 2000
- [BTT03] Biometric Technology Today. Facial Recognition – Part 1. Nov/Dec 2003
- [Bur95] Burl M.C., Leung T.K., Perona P. *Face Localization via Shape Statistics*. In Intl. W. on Automatic Face and Gesture Recognition, 1995
- [Cha98] Chai D., Ngan K.N. *Locating facial Region of a Head-and-Shoulders Color Image*. Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, pp. 124-129, 1998
- [Col97] Colmenarez A.J., Huang T.S. *Face Detection with Information-Based Maximum Discrimination*. In Computer Vision and Pattern Recognition, pp. 782-787, 1997
- [Col99] Colmenarez A., Frey B., Huang T. *Detection and tracking of faces and facial features*. In ICIP, 1999
- [Cra92] Craw I., Tock D., Bennett A. *Finding face features*. In Proceedings of the First European Conference on Computer Vision, pp. 92-96, 1992

- 
- [Dai96] Dai Y., Nakano Y. *Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene*, Pattern Recognition, vol. 29, no. 6, pp. 1007-1017, 1996
  - [Das02] Dass S. *Face Detection and Synthesis Using Markov Random Field Models*.
  - [Dut98] Duta N., Jain A.K. *Learning the Human Face Concept From Black and White Pictures*. Proc. Int'l Conf. Pattern Recognition, pp. 1365-1367, 1998
  - [Edw98] Edwards G.J., Taylor C.J., Cootes T.F. *Learning to identify and track faces in image sequences*. In IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998
  - [Ela02] Elad M., Hel-Or Y., Keshet R. *Rejection based classifier for face detection*. Pattern Recognition Letters, vol. 23, pp. 1459-1471, 2002
  - [Fer01] Feraud R., Bernier O.J., Villet J.-E., Collobert M. *A Fast and Accurate Face Detector Based on Neural Network*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 42-53, 2001
  - [Fri00] Frischolz R.W., Dieckmann U. *BioID: A multimodal biometric identification system*. IEEE Comput., vol. 33, 2000
  - [Fro04] Fröba B, Ernst A., *Face Detection with the Modified Census-Transform*; International Conference on Automatic Face and Gesture Recognition (FG '2004), Seoul, Korea, Mai 2004
  - [Gar99] Garcia C., Tziritas G. *Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis*. IEEE Transactions on Multimedia, p.264-277, 1999
  - [Gra95] Graf H.P., Chen T., Petajan E., Cosatto E. *Locating Faces and Facial Parts*. In Int. Workshop on Automatic Face and Gesture Recognition, pp. 41-46, 1995
  - [Gra00] Graf H., Cosatto E., Ezzat T. *Face analysis for the synthesis of photo-realistic talking heads*. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000
  - [Gro01] Gross R., Shi J., Cohn J. *Quo vadis Face Recognition?* Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania, June 2001
  - [Han98] Han C.-C., Liao H.-Y.M., Yu K.-C., Chen L.-H. *Fast Face Detection via Morphology-based Pre-processing*. Proc. Ninth Int'l Conf. Image Analysis and Processing, pp. 469-476, 1998
  - [Hei00] Heisele B., Poggio T., Pontil M. *Face Detection in Still Gray Images*. Artificial Intelligence Laboratory, MIT, 2000

- 
- [Hei01] Heisele B., Serre T., Pontil M., Poggio T. *Component-based face detection*. In CVPR01, pp. 657-662, 2001
  - [Hje01] Hjelmås E., Low B.K., Low B.K. *Face Detection: A Survey*. Computer Vision and Image Understanding, pp. 236-274, 2001
  - [Hot98] Hotta K., Kurita T., Mishima T. *Scale Invariant Face Detection Method using Higher-Order Local Autocorreltaion features of Log-Polar Image*. Proc. Third Int'l Conf. Automatic Face and Gesture Recognition, pp. 70-75, 1998
  - [Hsu02] Hsu R.-L., Abdel-Mottaleb M., Jain A. K. *Face Detection in Color Images*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 24, no. 5, pp. 696-706, 2002
  - [Hua96] Huang J., Gutta S., Wechsler H. *Detection of human faces using decision trees*. In Proceedings, Second International Conference on Automatic Face and Gesture Recognition, pp. 248-252. IEEE Computer Society Press, 1996
  - [Jon02] Jones J. M., Rehg J. M. *Statistical Color Models with Application to Skin Detection*. International Journal of Computer Vision, vol. 46, no. 1, pp. 81-96, 2002
  - [Kas05] Kasiński A., Król M. *Application of Active Shapes to the Structural Face Model*. CORES'05 - Fourth International Conference on Computer Recognition Systems, Leszno, 2005
  - [Kim02] Kim T.K., Lee S.U., Lee H.J., Kee S.C., Kim S.R. *Multiple Face Detection for Video Surveillance*. ICPR 2002, Quebec, Canada, 2002
  - [Kir90] Kirby M., Sirovich L. *Application of the Karhunen-Loeve Procedure for Characterization of Human Face*. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, pp. 103-108, 1990
  - [Kot97] Kotropoulos C., Pitas I. *Rule-based face detection in frontal views*. In Proc. of International Conference on Acoustics, Speech, and Signal Processing, pp. 21-24, 1997
  - [Kub06] Küblbeck C., Ernst A., *Face Detection and Tracking in Video Sequences using the Modified Census Transformation*; Image and Vision Computing 24 (2006) , Pages 564-572, 2006
  - [Kuc98] Kuchariew G. *Przetwarzanie i analiza obrazów cyfrowych*. Politechnika Szczecińska Wydział Informatyki, Szczecin 1998
  - [Kuk03a] Kukharev G., Kuźmiński A. *Techniki Biometryczne Część 1 – Metody Rozpoznawania Twarzy*. Politechnika Szczecińska Wydział Informatyki, 2003

- 
- [Kuk03b] Kukharev G., Masicz P., Masicz Pi. *Real-time face detection using modified gradient method*, Proceedings of the Seventh International Conference PRIP'2003, 195-199, Minsk, Belarus, 21-23.V.2003
  - [Kuk04a] Kukharev G., Masicz P., Masicz Pi. *A Fast and Accurate Faces Localization using Gradient Method*, WSCG'2004 Short Communications. 157-164, Plzen, Czech Republic, 2-6.II.2004
  - [Kuk04b] Kukharev G., Nowosielski A. *Fast and efficient algorithm for face detection in colour images*. Machine GRAPHICS & VISION 13(4), 377-399, 2004
  - [Kuk05a] Kukharev G., Masicz P., Masicz Pi. *Modified Gradient Method for Face Localization*, Enhanced Methods in Computer Security, Biometric and Artificial Intelligence Systems. Kluwer Academic Publisher, 2005
  - [Kuk05b] Kukharev G., Masicz P., Masicz P. *Faces Localization using Modified Gradient Method*, Computing, Multimedia and Intelligent Techniques. Special issue on Live Biometrics and Security 1(1), 2005
  - [Kuk07] Kukharev G., Masicz P. *Scientific research of the "Face Locator" environment which uses the method of oval object location and the variable ray procedure*, Sejmik Młodych Informatyków, Świnoujście 2007
  - [Kuz04] Kuźmiński A. *Opracowanie środowisko projektowo-badawczego do modelowania systemów rozpoznawania twarzy opartych na redukcji cech metodami komponentów głównych*. Rozprawa doktorska, Politechnika Szczecińska, 2004
  - [Kyx01] Кухарев Г.А. *Методы и средства идентификации личности человека*. Политехника Издательство, Санкт-Петербург, 2001
  - [Lan95] Lanitis A., Taylor C. J., Cootes T. F. *Automatic Face Identification System Using Flexible Appearance Models*. Image and Vision Computing, pp. 392-401, vol. 13, no. 5, 1995
  - [Leu95] Leung T.K., Burl M.C., Perona P. *Finding faces in cluttered scenes using random labeled graph matching*. In Fifth International Conference on Computer Vision, pp. 637-644, Cambridge, Massachusetts, June 1995
  - [Lew96] Lew M.S. *Information theoretic view-based and modular face detection*. 2nd. International Conference on Automatic Face and Gesture Recognition, Killington, VT, pp. 198-203, 1996
  - [Liu03] Liu Ch. *A Bayesian Discriminating Features Method for Face Detection*. IEEE Transactions On Pattern Analysis And Machine Intelligence. vol. 25, no. 6, pp. 725-740, 2003

- 
- [Mai98] Maio D., Maltoni D. *Fast Face Location in Complex Backgrounds*, in H. Wechsler et. al., *Face Recognition from Theory to Applications*, Nato Asi Series, Springer Verlag, 1998.
  - [Mar98] Mariani R. *A Face Location Algorithm Robust to Complex Lighting Conditions*,
  - [Mat02] Mateos G.G, Chicote C.V. *Face Detection on Still Images Using HIT Maps*. Lecture Notes in Computer Science
  - [Mas02a] Masicz P. *Określanie biometrycznych cech w obrazie twarzy*. Praca magisterska, Politechnika Szczecińska, Szczecin 2005
  - [Mas02] Masicz P. *System automatycznego określania cech biometrycznych w obrazie twarzy*, Materiały VII Sesji Naukowej Informatyki, Szczecin, 2002
  - [Mas03] Masicz Pi., Masicz P. *Rozpoznawanie twarzy za pomocą dwuwymiarowej transformaty cosinusowej*, Materiały VIII Sesji Naukowej Informatyki, Szczecin, 2003
  - [Mas06] Masicz P. Lokalizacja twarzy na obrazach kolorowych w słabych warunkach oświetleniowych, *Metody Informatyki Stosowanej w Technice i Technologii*. Roczniki Informatyki Stosowanej WI PS nr 8, Szczecin, 2006
  - [Mia99] Miao J., Yin B., Wang K., Shen L., Chen X. *A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template*. Pattern Recognition, vol. 32, no. 7, pp. 1237-1248, 1999
  - [Nef98] Nefian A.V., Hayes M. H. *Face detection and recognition using Hidden Markov Models*. in International Conference on Image Processing, 1998
  - [Osu97] Osuna E., Freund R., Girosi F. *Training support vector machines: an application to face detection*. CVPR'97, 1997
  - [Phi98] Phillips P. J., Wechsler H., Huang J., Rauss P. *The FERET database and evaluation procedure for face recognition algorithms*, Image and Vision Computing J, vol.16, no.5, pp. 295-306, 1998
  - [Phi99] Phillips P.J., Moon H., Rizvi S.A., Rauss P.J. *The FERET Evaluation Methodology for Face-Recognition Algorithms*. Technical Report NISTIR 6264, January 1999
  - [Phi00] Phillips P.J., Moon H., Rizvi S.A., Rauss P. J. *The FERET Evaluation Methodology for Face Recognition Algorithms*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 1090-1104, 2000.
  - [Rih00] Riha Z., Matyas V. *Biometric Authentication Systems*. Faculty of Informatics Masaryk University, November 2000

- 
- [Row95] Rowley H.A., Baluja S., Kanade T. *Human face detection in visual scenes*. CMU-CS-95-158R, Carnegie Mellon University, November 1995
- [Row98a] Rowley H.A., Baluja S., Kanade T. *Neural Network-Based Face Detection*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 20, no. 1, pp. 23-38, 1998
- [Row98b] Rowley H.A., Baluja S., Kanade T. *Rotation Invariant Neural Network-Based Face Detection*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 20, no. 6, pp. 38-44, 1998
- [Sab98] Saber E., Tekalp A.M. *Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions*. Pattern Recognition Letters, vol. 19, no. 8, pp. 669-680, 1998
- [San02] Sandeep K., Rajagopalan A.N. *Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information*, ICVGIP, 2002
- [Sca98] Scassellati B.M. *Eye Finding via Face Detection for a Foveated Active Vision System*. In Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, WI, July 1998
- [Sch98] Schneiderman H., Kanade T. *Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Santa Barbara, CA, pp. 45-51, 1998
- [Sch00] Schneiderman H., Kanade T. *A Statistical Method for 3D Object Detection Applied to Faces and Cars*. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 746-751, 2000
- [Sta02] Stateczny A., Praczyk T. *Sztuczne sieci neuronowe w rozpoznawaniu obiektów morskich*. Gdańskie Towarzystwo Naukowe, Gdynia 2002
- [Sun98] Sung K.K., Poggio T. *Example-Based Learning for View-Based Human Face Detection*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 20, no. 1, pp. 39-51, 1998
- [Ter98] Terrillon J.-C., David M., Akamatsu S. *Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments*. In Proc. of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 112-117, 1998
- [Tur91] Turk M., Pentland A. *Eigen Faces for Recognition*. Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991

- [UE03] COM(2003)558 final, Commission of the European Communities, Brussels, 24.09.2003
- [Woo97] Woodward J.D. *Biometrics: Privacy's Foe or Privacy's Friend*. Proceedings of The IEEE, vol. 85, no. 9, pp. 1480 – 1491, September 1997
- [Wu99] Wu H., Chen Q., Yachida M. *Face Detection From Color Images Using a Fuzzy Pattern Matching Method*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 21, no. 6, pp. 557-563, 1999
- [Wys67] Wyszecki G., Stiles W. S. *Color Science*. Wiley, New York, 1967
- [Yan98] Yang M.-H., Ahuja N. *Detecting Human Faces in Color Images*. Proc. IEEE Int'l Conf. Image Processing, vol. 1, pp. 127-130, 1998
- [Yan00] Yang M.-H., Ahuja N., Kriegman D.J. *Face Detection Using Mixtures of Linear Subspaces*. Proc. Fifth Int'l Conf. Automatic Face and Gesture Recognition, pp. 70-76, 2000
- [Yan00] Yang M.-H., Ahuja N., Kriegman D.J. *A SNoW-Based Face Detector*. Advances in Neural Information Processing System 12, S.A. Solla, T.K. Leen, and K.-R. Muller, eds., pp. 855-861, MIT Press, 2000
- [Yan02] Yang M.-H., Kriegman D.J., Ahuja N. *Detecting Faces in Images: A Survey*. IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002
- [Yow96] Yow K.C., Cipolla R. *A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection*. Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, pp. 16-21, 1996
- [Yow97] Yow K.C., Cipolla R. *Feature-based Human Face Detection*. Image and Vision Computing, vol. 15, no. 9, pp. 713-735, 1997
- [Yow98] Yow K.C., Cipolla R. *Enhancing Human Face Detection Using Motion and Active Contours*, Proc. Third Asian Conf. Computer Vision, pp. 515-522, 1998

### Strony internetowe

- [BioID] Baza twarzy: BioID. <ftp://ftp.uni-erlangen.de/pub/facedb/>
- [Bor] Strona internetowa producenta środowiska programistycznego C++Builder. <http://www.borland.pl/cbuilder/>
- [CMU] Baza twarzy: CMU Pose, Illumination, and Expression (PIE) Database. [http://www.ri.cmu.edu/projects/project\\_418.html](http://www.ri.cmu.edu/projects/project_418.html)

- 
- [FER] Baza twarzy: The Facial Recognition Technology (FERET) Database.  
[http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html)
- [FT] Strona główna firmy FotoNation, <http://www.fotonation.com>
- [FT2] Artykuł dotyczący projektu Face Detector firmy FotoNation:  
<http://www.fotopolis.pl/index.php?n=5646>
- [HS] HumanScan – Producent BioID. <http://www.humanscan.de/company/index.php>
- [ORL] Baza twarzy Olivetti. [http://www.uk.research.att.com/pub/data/att\\_faces.zip](http://www.uk.research.att.com/pub/data/att_faces.zip)
- [RTFD] Strona główna RealTime Face Detector:  
<http://www.iis.fraunhofer.de/fhg/iis/EN/bf/bv/kognitiv/biom/dd.jsp>
- [SinIn] Singular Inversions - Producent oprogramowania FaceGen Modeller.  
<http://www.facegen.com>
- [Vis] Viisage – Producent systemów biometrycznych.  
[http://www.viisage.com/ww/en/pub/viisage\\_products.htm](http://www.viisage.com/ww/en/pub/viisage_products.htm)

**Załącznik nr 1**  
**SPRAWOZDANIE Z BADAŃ**

1. Numer badania: .....

2. Termin badania:

|                        |  |
|------------------------|--|
| Data wykonania badania |  |
| Czas rozpoczęcia       |  |
| Czas zakończenia       |  |

3. Specyfikacja techniczna:

|                  |  |
|------------------|--|
| Nazwa bazy zdjęć |  |
| Ilość zdjęć      |  |
| Rozdzielczość    |  |
| Przestrzeń barw  |  |

4. Parametry:

|                                                   |  |
|---------------------------------------------------|--|
| macierze gradientów 45/135 włączone/wyłączone     |  |
| wartość odcięcia gradientu                        |  |
| wielkość obszaru sczytywanego dla każdego piksela |  |
| rozdzielczość obrazka poddana analizie            |  |
| wartość odcięcia dolnego histogramu               |  |
| wartość odcięcia górnego histogramu               |  |
| typ uaktualniania wyników                         |  |