

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**



Wydział Elektrotechniki, Automatyki,
Informatyki i Elektroniki

Katedra Automatyki

Praca magisterska

**SYSTEM BIOMETRYCZNY IDENTYFIKUJĄCY
OSOBY NA PODSTAWIE CECH OSOBNICZYCH
TWARZY**

Promotor:

dr Adrian Horzyk

Zespół autorski:

Małgorzata Lewicka, Krzysztof Stańczyk

Kierunek:

Automatyka i Robotyka

Kraków, maj 2008 r.

OŚWIADCZENIE:

Oświadczamy, świadomi odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą pracę dyplomową wykonaliśmy osobiście i samodzielnie (w zakresie wyszczególnionym we wstępie) i że nie korzystaliśmy ze źródeł innych niż wymienione w pracy.

Składamy serdeczne podziękowanie
Panu dr Adrianowi Horzykowi
za pomoc i cenne uwagi podczas
realizacji tej pracy

Spis treści

Wstęp.....	6
1. Wprowadzenie.....	8
1.1. Właściwości twarzy.....	9
1.2. Struktura, zasada działania i rodzaje systemów rozpoznawania twarzy.....	12
1.2.1. Ogólna struktura systemów rozpoznawania twarzy.....	12
1.2.2. Automatyczna identyfikacja twarzy.....	14
1.3. Metody rozpoznawania twarzy.....	25
1.4. Systemy rozpoznawania twarzy.....	28
1.4.1. Systemy „Visitor Identification”.....	31
1.4.2. Systemy „Access Control”.....	35
1.4.3. Systemy “Name-It”.....	38
1.5. Przykłady istniejących aplikacji i systemów.....	41
2. Metodologia.....	47
2.1. Ogólna koncepcja systemu.....	47
2.2. Budowa modelu „skin-color detection”.....	49
2.2.1. Wybór właściwego modelu przestrzeni barw.....	50
2.2.2. Reprezentacja Gaussowska rozkładu kolorów ludzkiej twarzy. Model <i>EBM</i>	55
2.2.3. Estymacja parametrów modelu.....	59
2.3. Detekcja i lokalizacja twarzy.....	59
2.3.1. Segmentacja obrazu.....	60
2.3.2. Selekcja obszarów twarzy.....	62
2.4. Ekstrakcja cech.....	67
2.4.1. Detekcja oczu.....	68
2.4.2. Normalizacja orientacji twarzy w płaszczyźnie <i>XY</i>	77
2.4.3. Detekcja ust.....	79
2.4.4. Detekcja nosa.....	83
2.4.5. Detekcja konturów twarzy. Wyznaczanie szerokości twarzy.....	87
2.4.6. Wyznaczanie środkowej części twarzy.....	91
2.4.7. Budowa wektora cech.....	93
2.5. Struktura bazy danych.....	94
2.5.1. Wprowadzenie.....	94
2.5.2. Normalizacja bazy.....	96
2.6. Identyfikacja.....	97
2.6.1. Weryfikacja rezultatu metodą korelacji fazowej.....	98
3. Implementacja.....	103
3.1. Informacje ogólne.....	103
3.2. Podstawowe funkcjonalności.....	103
3.3. Architektura systemu.....	104
3.4. Opis programu.....	107
3.4.1. Analiza obrazów twarzy. Detekcja i lokalizacja twarzy.....	107
3.4.2. Budowa wektorów cech.....	109
3.4.3. Identyfikacja.....	110
3.4.4. Obsługa bazy danych.....	111
4. Wyniki.....	112
4.1. Wprowadzenie.....	112
4.2. Baza danych twarzy.....	112
4.3. Detekcja i lokalizacja twarzy.....	113

4.4.	Ekstrakcja cech	123
4.5.	Identyfikacja twarzy	128
4.5.1.	Wprowadzenie	128
4.5.2.	Program <i>FaceRec</i>	129
4.5.3.	Pakiet <i>PRTools</i> środowiska MATLAB	138
4.5.4.	Pakiet <i>RapidMiner</i>	139
4.5.5.	Analiza wyników	141
5.	Podsumowanie	145
	Bibliografia	149
	Spis rysunków	152
	Spis tabel	156

Wstęp

Rozpoznawanie twarzy towarzyszy każdemu człowiekowi niemal od jego narodzenia. Już niemowlęta, mimo słabo jeszcze rozwiniętego zmysłu wzroku, potrafią rozpoznać wśród obcych twarzy twarz matki. Wraz z rozwojem fizycznym dziecka i jego dorastaniem umiejętność ta jest stale rozwijana. Potrafimy rozpoznać wśród wielu twarzy te, należące do członków rodziny, znajomych lub osób znanych z widzenia i fotografii.

Jednak to co jest czynnością niemal naturalną dla ludzi, staje się skomplikowanym problemem, gdy próbuje się wykorzystać do tego celu urządzenia techniczne. Okazało się, że wymaga to opracowania wielu skomplikowanych algorytmów, wdrożenia ich w praktyce, a przede wszystkim przetwarzania ogromnej ilości danych. Nauka zajmująca się pomiarami cech istot żywych, czyli biometria, zajmowała się tym od wielu lat, jednak jej gwałtowny rozwój związany był z rewolucją informatyczną. Ogromne możliwości obliczeniowe współczesnych komputerów nie tylko wyspecjalizowanych – uniwersyteckich, lecz również osobistych sprawiły, że zagadnienie to jest rozpatrywane przez coraz większą ilość badaczy na całym świecie. Powstają zupełnie nowe algorytmy, a także ulepszane są te już istniejące. Budowane na ich podstawie aplikacje pojawiają się już nie tylko w laboratoriach i na wydziałach uczelni, ale także znajdują zastosowania w życiu codziennym.

Wielkie znaczenie miały dla tej sprawy wydarzenia z 11 września 2001 roku. To w ich następstwie zaczęto zwracać większą niż wcześniej uwagę na sprawy bezpieczeństwa, także na kontrolę dostępu do strategicznych miejsc. Naturalnym okazało się poszukiwanie jak najlepszych metod i systemów wspomagających codzienną pracę ludzi na lotniskach, dworcach i innych miejscach będących potencjalnym celem ataków. W związku z tym nastąpił gwałtowny rozwój systemów, nie tylko tych pracujących od czasu do czasu w sytuacji wystąpienia zapotrzebowania przeprowadzenia rozpoznawania (jak systemy zawierające dane o liniach papilarnych będące w posiadaniu policji wszystkich krajów), ale także i tych pracujących „*on-line*”, przeprowadzających operacje w jak najkrótszym czasie.

Również autorzy niniejszej pracy zajęli się drobnym wycinkiem tego zagadnienia. Celem pracy jest zaprojektowanie i stworzenie systemu rozpoznawania twarzy, który na podstawie informacji o wybranych cechach osobniczych pozwoli na identyfikację osób. Temat narzuca w pewien sposób metodologię rozwiązania, opartą nie na globalnej analizie twarzy a na jej analitycznym rozkładzie. Oznacza to, że przed autorami zostają postawione następujące zadania:

- wybór odpowiednich cech twarzy wykorzystywanych do identyfikacji,
- wybór i budowa odpowiednich metod detekcji i lokalizacji twarzy na obrazach wejściowych ,
- ekstrakcja cech z wyznaczonych fragmentów obrazów wejściowych zawierających obrazy twarzy,
- budowa bazy danych twarzy,
- właściwa identyfikacja,
- ocena uzyskanych wyników.

Wymienione zadania odpowiadają z grubsza przebiegowi typowego procesu identyfikacji osób. Podstawowymi składnikami systemu będą: baza danych o osobach oraz aplikacja realizująca w/w algorytmy. System powinien zapewniać możliwość rozbudowywania bazy danych na podstawie przetwarzanych przez aplikację obrazów nowych osób, a także nowych zdjęć osób w bazie już istniejących. Jednym z celów pracy jest próba odpowiedzi na pytanie, czy prezentowane podejście może znaleźć praktyczne zastosowanie w dziedzinie identyfikacji osób, a jeśli tak – to przy spełnieniu jakich warunków.

W niniejszej pracy opisany został cały proces tworzenia systemu.

Rozdział pierwszy zawiera opis ogólnej struktury systemów rozpoznawania twarzy, przedstawienie kolejnych etapów automatycznej identyfikacji, krótki przegląd używanych metod a także podział systemów rozpoznawania. Na końcu opisano kilka przykładowych istniejących aplikacji (opracowanie: Małgorzata Lewicka i Krzysztof Stańczyk).

W następnym rozdziale przedstawione są kolejne etapy budowy projektowanego systemu, wybór użytych metod, zawiera on także dokładny opis znajdowania punktów charakterystycznych na twarzy i mierzenia odległości między nimi, jak również prezentuje strukturę bazy danych, w której przechowywane są wektory opisujące osoby (opracowanie: Krzysztof Stańczyk).

Kolejny rozdział to krótkie przedstawienie powstałej aplikacji (opracowanie: Krzysztof Stańczyk).

Rozdział czwarty zawiera przykładowe wyniki eksperymentów przeprowadzonych z wykorzystaniem zaimplementowanego systemu, a także te uzyskane za pomocą pakietów *PRTools* (w środowisku *MATLAB*) oraz *RapidMiner* - wykonane na podstawie uzyskanych danych. Następnie zaprezentowane zostały wyniki sumaryczne oraz ich analiza (opracowanie: Małgorzata Lewicka).

Rozdział piąty to podsumowanie pracy, wyników oraz wnioski końcowe.

1. Wprowadzenie

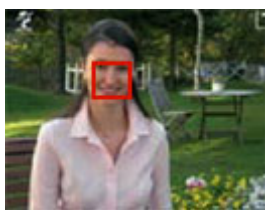
Rozpoznawanie obrazów to jedna z ciekawszych i najszybciej rozwijających się dziedzin informatyki. Z zastosowaniami tego zagadnienia mamy coraz częściej do czynienia w różnych dziedzinach naszego życia i działalności. Przykładami mogą być: codzienne wykorzystywanie kodów kreskowych na towarach, analiza obrazów z wideoradarów przez policję drogową, jak również programy typu OCR używane przez skanery do rozpoznawania tekstu.

Szczególne znaczenie mają systemy rozpoznające twarz. Znajdują one najczęściej zastosowanie w zakresie nadzoru i kontroli. Nadzór można rozumieć jako sprawdzanie czy dana osoba nie przekracza obszaru, w którym może się poruszać – przykładem może być tutaj szpital psychiatryczny i nadzór nad pacjentami, innym siedziba dużej firmy z wydzielonym terenem, po którym mogą poruszać się goście.

Ostatnio coraz częściej spotykane są również zabezpieczenia w dostępie do sprzętu komputerowego stosujące rozpoznawanie – np. coraz więcej firm ma w swojej ofercie laptopy (np. HP - COMPAQ nc6400, FUJITSU - SIEMENS LB S7110, LG Express M1 M1-5221Y), które są wyposażone w skanery linii papilarnych, dzięki czemu dostęp do zawartych w nich danych mają tylko wcześniej zarejestrowane osoby.

Bardzo ciekawym zastosowaniem w praktyce zagadnienia lokalizacji twarzy na obrazie jest wykorzystanie tej funkcji w aparatach fotograficznych.

Firma Nikon swoje aparaty z serii COOLPIX (m.in. L5, P3, P4, S5, S6, S9 [WWW-10]) wyposaża w funkcję autofokus [WWW-6]. Dzięki niej, przed wykonaniem zdjęcia można znaleźć w obszarze zainteresowania twarz (rys. 1.1) i ustawić ostrość dokładnie na daną twarz niezależnie od tego, w której części kadru się ona znajduje. Wykorzystywana jest do tego metoda porównywania wzorców (*ang. template matching*) – oprogramowanie aparatu wyposażono w podstawowy wzorec graficzny konturów ludzkiej twarzy. Funkcję tę dostarcza firma Identix.



Rys. 1.1. Autofokus w aparatach Nikon (Źródło: [WWW-9])

Jeszcze dalej poszła firma FujiFilm. W aparatach tej firmy (np. FinePix S6500fd, FinePix F31fd, FinePix S5Pro, FinePix F40fd) wykrywanie twarzy na zdjęciu nie jest wykonywane programowo lecz sprzętowo przez wbudowany procesor [WWW-7]. Dzięki temu jest ono jeszcze szybsze – w ciągu 0,05s można wykryć nawet 10 twarzy na zdjęciu, a następnie dopasować optymalną ostrość i ekspozycję tak, by uzyskać jak najlepsze ujęcie portretowe. Opcja ta może być również używana przy odtwarzaniu zdjęć do kadrowania lub przybliżania twarzy fotografowanych osób.

1.1. Właściwości twarzy

Badania wykazały, że twarz z którą się urodziliśmy pozostaje identyfikowalna przez resztę naszego życia. Oznacza to, że zależności pomiędzy poszczególnymi częściami twarzy, jej kontury, charakteryzują się względną stałością niezależnie od naszego wieku (dojrzałego). Wiedza o tym pozwala nam na wykorzystanie twarzy jako części ludzkiego ciała w zadaniu rozpoznawania.

Twarz na obrazie cyfrowym można opisać za pomocą szeregu cech biometrycznych człowieka (rys. 1.2). Można je podzielić na dwie główne grupy:

a) cechy geometryczne:

- owal twarzy,
- szerokość twarzy,
- kształt ust (szerokość, kształt),
- kształt nosa (szerokość, wysokość),
- kształt czoła,
- kształt brwi,
- kształt podbródka,
- kształt uszu,

b) antropometryczne:

- odległości między środkami oczu,
- odległości między najdalszymi punktami oczu,
- odległości pomiędzy oczami i nosem,
- odległości pomiędzy linią oczu i linią ust,
- odległości od najdalszego punktu oka do środka ust itd.



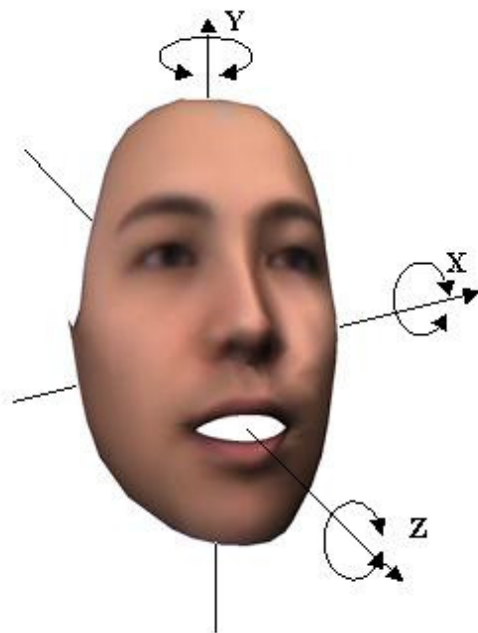
Rys. 1.2. Wybrane cechy biometryczne twarzy

Oprócz w/w podziału w literaturze funkcjonuje podział ze względu na kategorię. Wprowadza on trzy grupy cech: fizyczne, strukturalne i matematyczne. Fizyczne i strukturalne cechy obrazów wyrażane są poprzez kształty twarzy (owal, kształty podstawowych części), kolor twarzy, włosów. Najczęściej stosowaną cechą tych kategorii jest jasność barwy, wyrażona poprzez poziom wartości każdego piksela na obrazie cyfrowym twarzy. Do kategorii fizycznych i strukturalnych cech możemy również zaliczyć współrzędne punktów twarzy w miejscach odpowiadających zmianie kontrastu i będących takimi częściami twarzy jak brwi, oczy, nos, uszy, usta i owal lub kontur twarzy. Kategorie matematyczne odnoszą się do statystycznych charakterystyk obrazów (średnie, dyspersje), gradientu zmian jasności, widma spektralnego obrazów (np. z wykorzystaniem przekształceń falkowych) i innych otrzymanych w wyniku matematycznego przekształcenia obrazów.

Przez porównanie odpowiednich cech biometrycznych twarzy można decydować o mierze podobieństwa pomiędzy ludźmi. W zadaniu rozpoznawania, kluczowe znaczenie ma właściwy ich dobór oraz zapewnienie optymalnych warunków do ich pozyskania. Związanych z tym jest sporo problemów, które projektant systemu musi uwzględnić tak, aby osiągnąć dostateczną skuteczność działania systemu. Problemy te związane są przede

wszystkim z zapewnieniem w miarę porównywalnych warunków otoczenia (oświetlenie, tło) w trakcie akwizycji obrazów i właściwą orientacją twarzy w przestrzeni (rys. 1.3). Ma to szczególne znaczenie w systemach klasy „*Visitor-Identification*” gdzie nie ma możliwości ręcznej korekty pozyskanych obrazów. Twarz – jako rzut głowy człowieka na płaszczyznę XY – może charakteryzować różne parametry orientacji głowy w przestrzeni 3D. Przy porównywaniu dwóch twarzy (badanej i wzorcowej z BD) ważne jest, żeby właściwości jednej i drugiej w przybliżeniu sobie odpowiadały. W praktyce jest to trudne do osiągnięcia, dlatego konieczne staje się wykorzystanie metod i procedur normalizacji obrazów twarzy. Celem tej normalizacji jest transformacja badanego obrazu twarzy wg właściwości obrazu wzorcowego (obrót w płaszczyźnie XY, obrót wokół osi Z, obrót wokół osi Y lub X, skalowanie).

Wiąże się to niestety ze wzrostem kosztu obliczeniowego przy rozwiązywaniu zadania rozpoznawania, co w konsekwencji prowadzi do obniżenia wydajności całego systemu.



Rys. 1.3. Osie obrotów twarzy w przestrzeni 3D

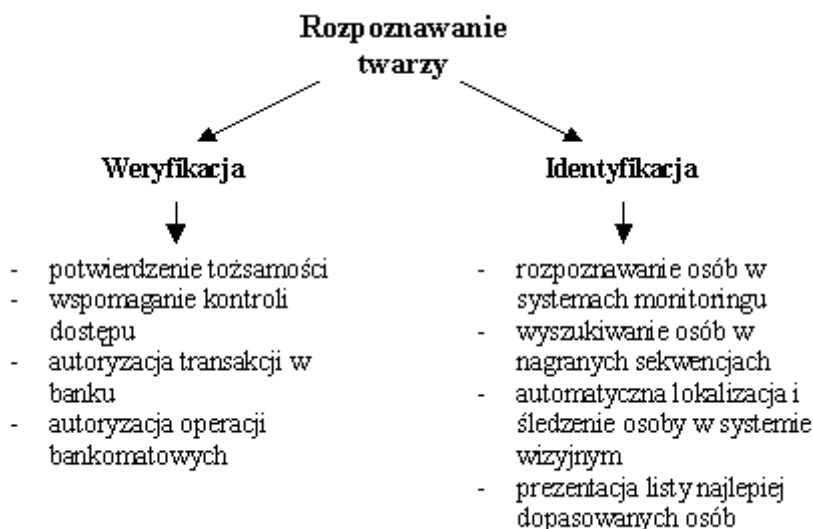
To jak bardzo właściwości jednej i drugiej twarzy muszą sobie odpowiadać, zależy jednak w dużej mierze od zastosowanej metody rozpoznawania. Istnieją techniki pozwalające na uzyskanie stosunkowo dużej skuteczności działania i dopuszczające przypadek, w którym badana twarz różni się – w ustalonych granicach – pod względem swoich właściwości w stosunku do obrazu wzorcowego. W tym przypadku problemem jest jednak skomplikowana procedura ekstrakcji cech oraz dokładność wyznaczenia współrzędnych konturów twarzy

i antropometrycznych punktów na twarzy. Metody te zostaną opisane szerzej w jednym z kolejnych rozdziałów.

1.2. Struktura, zasada działania i rodzaje systemów rozpoznawania twarzy

1.2.1. Ogólna struktura systemów rozpoznawania twarzy

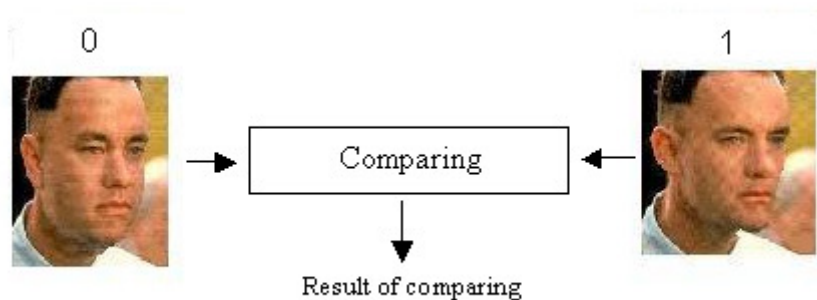
Systemy biometryczne mogą pracować w dwóch trybach: trybie weryfikacji oraz trybie identyfikacji (rys. 1.4). Tryb weryfikacji pozwala na ustalenie identyczności zadanej twarzy z pewnym wzorcem i działa na zasadzie porównania "jeden wzorzec – jeden obraz identyfikowany". Ponieważ częstym jego zastosowaniem jest sprawdzanie tożsamości osób (np. na podstawie zdjęcia z dowodu osobistego) tryb ten utożsamiany jest często z operacją autoryzacji dostępu.



Rys. 1.4. Tryby pracy systemów rozpoznawania i ich zastosowanie

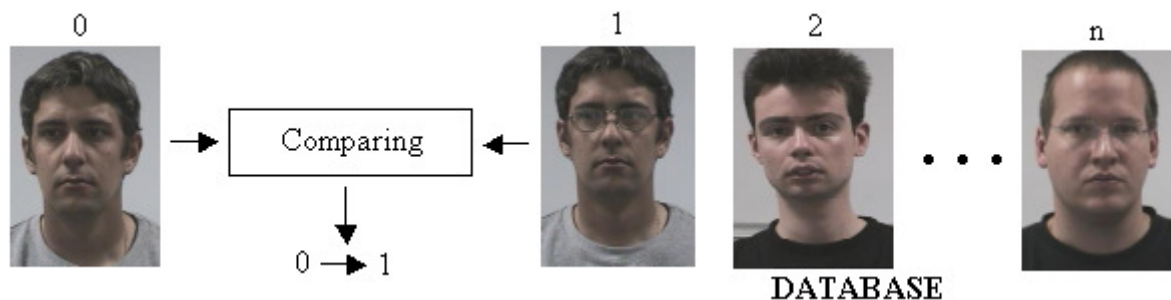
Ogólny schemat systemu rozpoznawania realizującego zadanie weryfikacji przedstawia rys. 1.5. Podstawowym blokiem w strukturze systemu jest komparator, który realizuje zadanie porównywania dwóch obrazów pomiędzy sobą i generuje decyzję na podstawie rezultatu rozpoznawania. Obrazem identyfikowanym jest twarz oznaczona cyfrą „0” a wzorcem twarz oznaczona cyfrą „1”. Rezultat weryfikacji określa się miarą podobieństwa dwóch obrazów z ustalonym progiem określającym warunki graniczne przyjęcia decyzji w procedurze weryfikacji. Wynik porównania może przyjmować dwie

wartości: „TRUE” lub „FALSE” przy czym „TRUE” oznacza, że twarze należą do tego samego człowieka a „FALSE” oznacza, że twarze należą do dwóch różnych osób.



Rys. 1.5. Ogólny schemat systemu rozpoznawania twarzy w trybie weryfikacji

Tryb identyfikacji pozwala na uzyskanie pełnej identyfikacji osoby poddawanej rozpoznaniu poprzez wyszukanie jej w bazie danych na zasadzie "jeden obraz identyfikowany – wiele wzorców" (rys. 1.6).

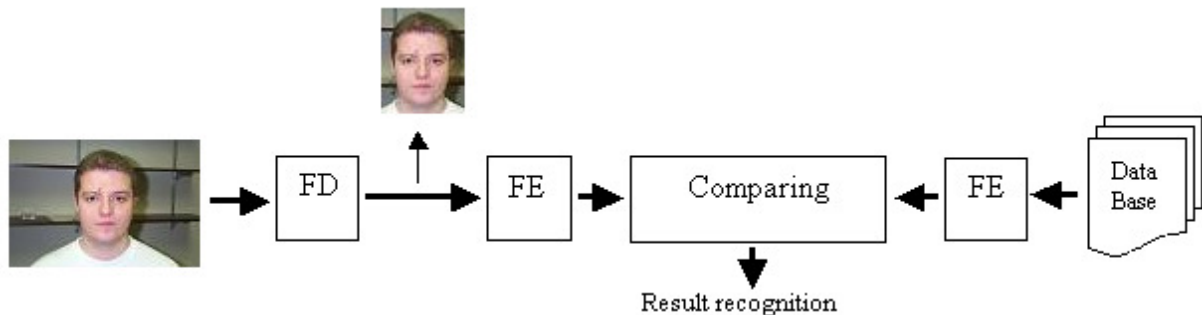


Rys. 1.6. Ogólny schemat systemu rozpoznawania twarzy w trybie identyfikacji

W stosunku do systemu realizującego zadanie weryfikacji nowym elementem jest baza danych, gdzie cyframi 1÷n oznaczono numery początkowe obrazów lub klas obrazów w bazie danych. Na prawe wejście komparatora podaje się po kolei wszystkie obrazy z bazy danych i dla każdego z tych obrazów realizowane jest zadanie weryfikacji. Przynależność badanego obrazu do pewnej klasy obrazów jest określona przez maksimum miary podobieństwa z obrazami tej klasy i przy spełnieniu warunku, że ta miara podobieństwa jest wyższa niż zadany próg.

Przy założeniu, że dane początkowe (twarze) przedstawione są w formie obrazów cyfrowych i nośnikiem informacji jest wartość jasności pojedynczego piksela, okazuje się, że komparator w trakcie pojedynczego zadania weryfikacji musi wykonać $M \times N$ porównań, gdzie M , N oznaczają odpowiednio wysokość i szerokość obrazu cyfrowego i stanowią wymiar początkowej przestrzeni cech. Jest to bardzo czasochłonna operacja proporcjonalna do rozmiarów obrazu. Na dodatek, bezpośrednie porównanie nie zawsze daje satysfakcjonujące wyniki co związane jest m.in. z niespełnieniem wymagań opisanych w jednym

z wcześniejszych rozdziałów. Dlatego stosowane są różne metody pozwalające na zmniejszenie wymiaru początkowej przestrzeni cech, przy zachowaniu wszystkich istotnych w procesie identyfikacji danych. Zmniejszenie wymiaru początkowej przestrzeni cech można osiągnąć np. usuwając zbędne w procesie identyfikacji elementy tła, ograniczając porównanie do wybranych wierszy lub kolumn obrazu badanego, czy też przedstawiając kształt twarzy nie w postaci cech jaskrawości a jako np. współrzędnych punktów twarzy w miejscach, gdzie rozmieszczone są elementy charakterystyczne twarzy lub modeli konturowych. Zadanie to realizowane jest przez blok ekstrakcji cech (FE). Ponadto, jeśli system rozpoznawania pracuje z rzeczywistymi danymi pozyskiwanymi bezpośrednio z kamery wideo, to koniecznym staje się wprowadzenie dodatkowego bloku realizującego zadanie detekcji twarzy (FD) z pola całego obrazu podanego na wejście systemu rozpoznawania (rys. 1.7).



Rys. 1.7. Ogólna struktura systemu rozpoznawania

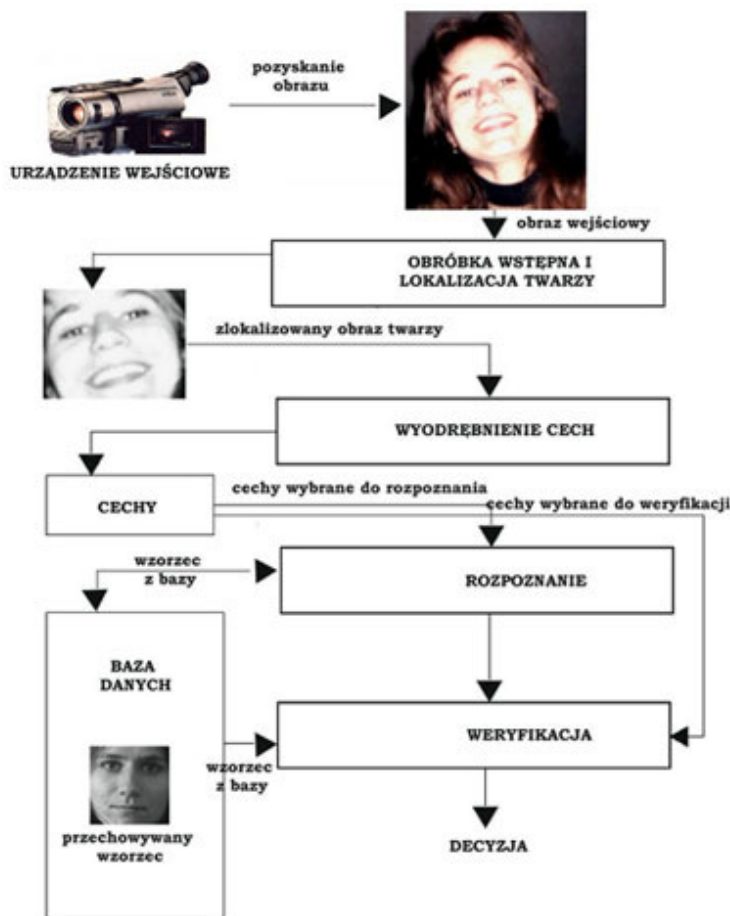
Wszystkie te bloki tworzą ogólną strukturę systemu rozpoznawania. Na jej podstawie można budować wyspecjalizowane systemy, realizujące rozpoznawanie twarzy w różnych zastosowaniach.

1.2.2. Automatyczna identyfikacja twarzy

Zadaniem każdego systemu rozpoznawania twarzy jest automatyczna identyfikacja osoby na podstawie cyfrowego obrazu jej twarzy. W oparciu o opisaną wcześniej ogólną strukturę systemu, na rysunku 1.8 przedstawiony został przykładowy schemat systemu rozpoznawania twarzy, realizującego zadanie automatycznej identyfikacji twarzy.

Proces identyfikacji składa się z pięciu podstawowych etapów:

- pozyskania obrazu z urządzenia wejściowego,
- obróbki wstępnej i lokalizacji twarzy ,
- ekstrakcji cech,
- identyfikacji (rozpoznania),
- weryfikacji i kontroli poprawności.



Rys. 1.8. Proces automatycznej identyfikacji twarzy (Źródło: [WWW-1])

Pozyskanie obrazu z urządzenia wejściowego

Na etapie pozyskania obrazu z urządzenia wejściowego zadaniem systemu jest akwizycja obrazów twarzy. W przypadku systemów „on-line” (np. „Visitor-identification”) ich źródłem jest najczęściej kamera CCD. Ponieważ cały proces odbywa się automatycznie, ważne jest, żeby warunki, w jakich pozyskuje się obrazy twarzy były jak najbardziej zbliżone do warunków, w jakich pozyskiwano obrazy przy budowie bazy danych. Problemem może być tutaj np. odpowiednie ustawienie kamery względem identyfikowanej osoby, oświetlenie itp. Niektóre z tych parametrów mogą zostać skorygowane programowo w dalszych etapach identyfikacji. W przypadku systemów „off-line”, użytkownik systemu przed rozpoczęciem rozpoznawania może dokonać ręcznej korekty parametrów obrazu niespełniających wymaganych kryteriów. Daje to większy wpływ na proces rozpoznawania, co w konsekwencji przekłada się na skuteczność jego działania.

Obróbka wstępna i lokalizacja twarzy

Etap obróbki wstępnej obrazu ma za zadanie wyeliminowanie wpływów otoczenia na

pozyskany obraz twarzy oraz dostosowanie go do pewnych ustalonych parametrów (normalizacja obrazu). Czynniki, które należy wziąć pod uwagę to:

- wpływ oświetlenia (rozkład cieni, intensywność światła, refleksy świetlne) - szczególnie duże znaczenie ma on w systemach, w których do akwizycji obrazów wykorzystuje się kamery o niezbyt wysokiej jakości (np. kamery internetowe). Problemem jest tu niestabilność parametrów (jasności, kontrastu) obrazu przez nie generowanego, które w pełni zależą od otaczającego oświetlenia. Drugi problem dotyczy przypadku, gdy wejściowe obrazy twarzy pochodzą z różnych źródeł i posiadają różne stopnie jasności, kontrastu i przeniesienia oddzielnych detali na wycinkach w obszarze twarzy i otaczającego je tła. Może to ostatecznie przełożyć się na rezultat identyfikacji, ponieważ wejściowymi cechami są wartości pikseli obrazu. Ponadto, kluczowe dla procesu identyfikacji elementy twarzy mogą wskutek niekorzystnego oświetlenia (rys. 1.9) ulec zatarciu na obrazie wejściowym i system nie będzie w stanie ich wykryć.



Rys. 1.9. Wpływ oświetlenia na obraz twarzy

W celu zminimalizowania wpływu w/w czynników ważne jest zapewnienie porównywalnych, stabilnych warunków akwizycji obrazów twarzy (i często tak się dzieje w praktyce, np. w pomieszczeniach zamkniętych oświetlanych ciągłym światłem sztucznym), zapisywanie do bazy danych obrazów należących do tej samej klasy pobranych przy różnym oświetleniu zewnętrznym (naturalnym w dzień i wieczorem, przy oświetleniu sztucznym, wada – rozmiary bazy), lokalne (dla każdego obrazu) dobieranie czułości kamery w momencie pracy z systemem rozpoznawania. Niezależnie od zastosowanego rozwiązania, system musi udostępniać metody pozwalające na korekcję parametrów pozyskanych obrazów,

- poza – twarz na obrazach może przyjmować różne pozy w stosunku do oka kamery (frontalnie, profilem, pochylona do przodu lub do tyłu).

W przypadku niektórych z nich istnieje niebezpieczeństwo, że niezbędne w procesie rozpoznawania cechy będą w całości lub częściowo niewidoczne,

- zmiany w wyglądzie zewnętrznym (okulary, zmiana uczesania, zarost),
- ekspresja twarzy – może mieć wpływ na właściwości niektórych cech twarzy (np. szerokość ust, nosa, położenie brwi),
- przeskalowanie obrazu do zadanego rozmiaru,
- usuwanie tekstury (tła), eliminacja szumów w obrazie,
- widoczność twarzy - twarz może być częściowo zakryta przez inne obiekty.

Na obrazie, który przedstawia grupę osób, niektóre twarze mogą zakrywać (częściowo) inne.

Niektóre z wymienionych czynników system jest w stanie wyeliminować, wykorzystując dostępne metody i algorytmy, jednak inne, np. te związane z widocznością twarzy czy niektórymi pozami mogą uniemożliwić przeprowadzenie procesu identyfikacji. Procedury normalizacyjne wykorzystywane na etapie obróbki wstępnej, znajdują także zastosowanie w późniejszych etapach rozpoznawania – np. do lokalnego podkreślania cech twarzy (owalu, granic ust, oczu itd.) w trakcie etapu ekstrakcji cech.

Po zakończeniu normalizacji parametrów obrazu, system może rozpocząć procedurę lokalizacji i detekcji wszystkich twarzy na zarejestrowanym przez urządzenie wejściowym obrazie. Detekcja to określenie pozycji wszystkich twarzy znajdujących się w kadrze. Jest ona realizowana przed operacją lokalizacji, tyle razy, ile twarzy znajduje się na analizowanym obrazie. Lokalizacja to określenie pozycji pojedynczej twarzy na wyodrębnionym w trakcie detekcji fragmencie obrazu wejściowego. Ponieważ w ogólnym przypadku na obrazie wejściowym mogą zostać także zarejestrowane obiekty przypominające ludzką twarz a nie będące nią, system musi umieć je eliminować, nie tracąc jednocześnie informacji o rzeczywistych obiektach twarzy. Istnieje wiele algorytmów pozwalających na określenie położenia twarzy na obrazie cyfrowym. Ogólnie można podzielić je na:

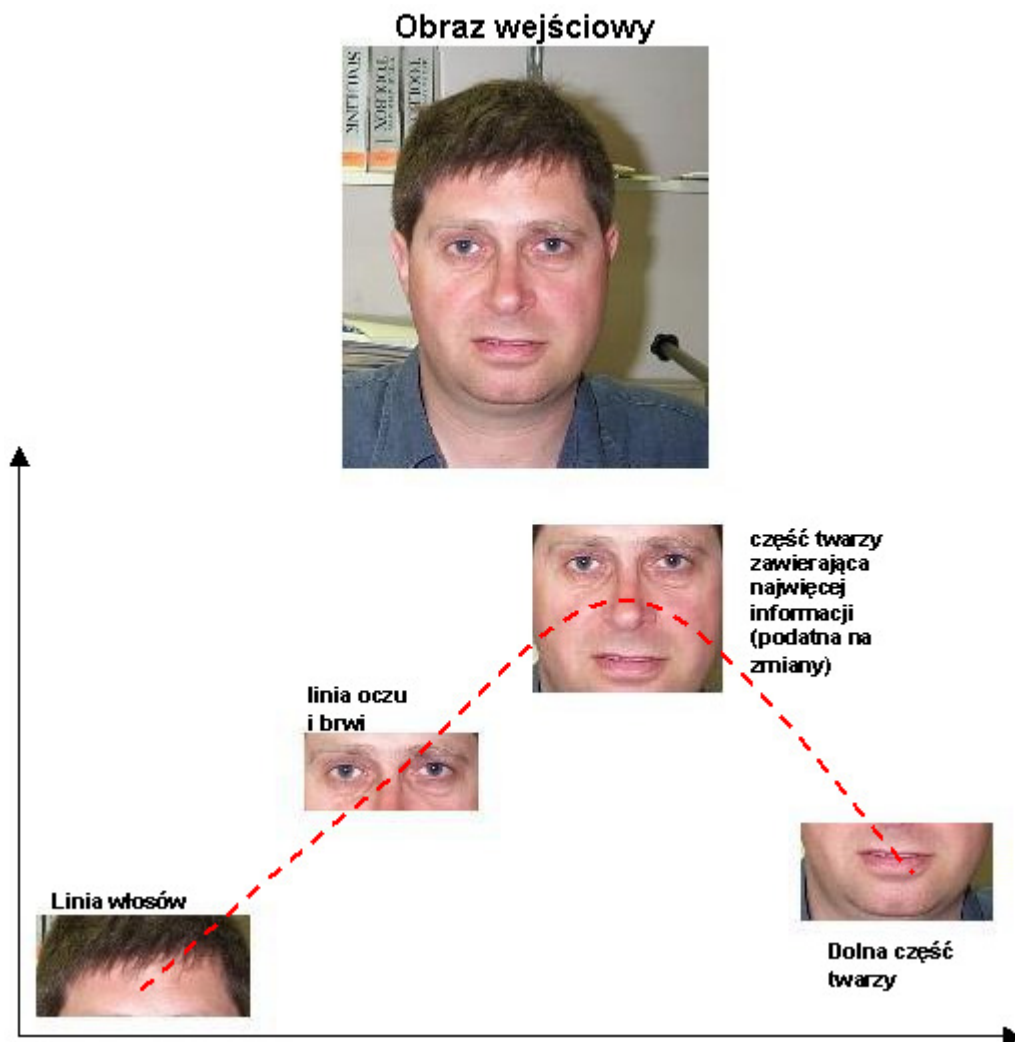
- metody bazujące na algorytmach heurystycznych i dobieraniu szablonu,
- techniki opierające się na deformacjach szablonu charakterystycznych dla twarzy ludzkiej lub jej wewnętrznych cechach (oczy, usta, nos),
- detekcja twarzy lub konturu twarzy przy użyciu bruzd lub techniki dynamicznego konturu,
- metody bazujące na eliptycznej aproksymacji i poszukiwaniu twarzy poprzez minimalizację ostatniego obszaru i zwiększaniu rejonu elipsy,

- algorytmy bazujące na transformacji Hough'a i jej adaptacjach,
- metody bazujące na poszukiwaniu znaczących grup cech (tripletów itp.),
- poszukiwanie twarzy w przestrzeni własnej przy wykorzystaniu PCA i teorii informacji,
- uczenie sieci neuronowych,
- lokalizacja twarzy na obrazach kolorowych wykorzystująca segmentację w przestrzeni kolorów,
- detekcja twarzy przy użyciu informacji o ruchu (przepływu optycznego) oraz gradientu przestrzenno-czasowego i inne.

W praktyce, w celu poprawienia wyników lokalizacji stosowane są różne kombinacje wyżej wymienionych metod.

Ekstrakcja cech

Celem ekstrakcji cech jest wyszukanie istotnych z punktu widzenia identyfikacji cech twarzy. Głównym problemem jest tu wybór cech twarzy podlegających rozpoznaniu a następnie odpowiedniej do nich metody porównania obrazów. Wiąże się to z koniecznością wyodrębnienia części ludzkiej twarzy niosących największą ilość informacji przydatnych dla danego typu rozpoznawania. Rysunek 1.10 przedstawia wybrane części twarzy i ilości dostarczanych przez nie informacji na temat indywidualnych cech osobnika. Najbardziej użyteczną z punktu widzenia rozpoznawania jest centralna część twarzy. W tym przypadku trzeba jednak uwzględnić fakt, iż jest to także część najbardziej podatna na zmiany związane np. z wyglądem danej osoby (zarost, okulary, biżuteria) lub mimiką twarzy.

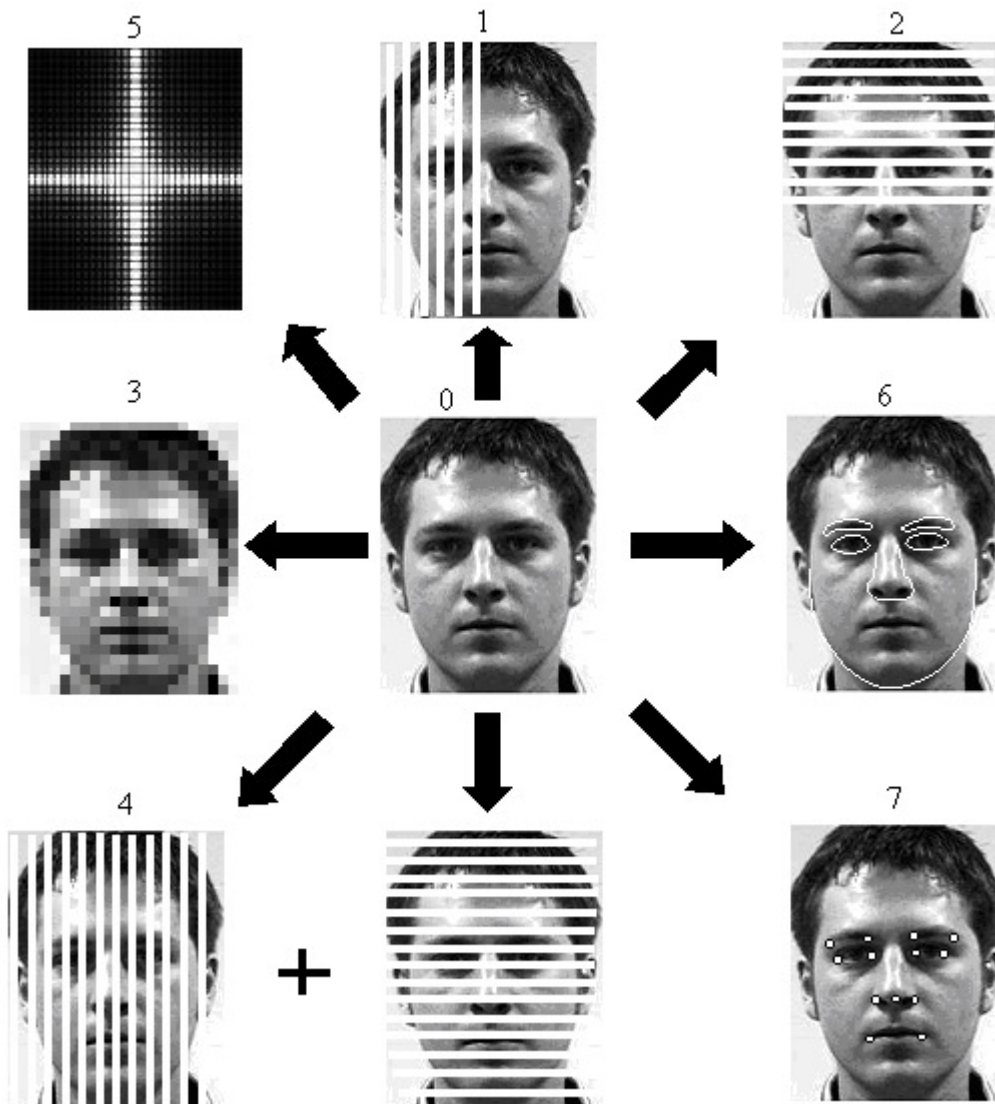


Rys. 1.10. Części twarzy a ilość dostarczanych przez nie informacji nt. cech osobniczych

Ekstrakcja cech może być zrealizowana na podstawie trzech podejść: selekcji cech, redukcji cech oraz selekcji cech wspólnie z ich redukcją. Na rysunku 1.11 przedstawiono różne sposoby przedstawienia twarzy za pomocą cech kategorii fizycznych i strukturalnych (cech jaskrawości obrazu cyfrowego) oraz cech kategorii matematycznych (widma spektralnego).

Obraz wejściowy oznaczony został cyfrą „0”. Najprostszy opis twarzy polega na zbudowaniu wektora cech drogą konkatencji kolumn (obraz 1) lub wierszy (obraz 2) obrazu wejściowego, traktowanego jako macierz danych o wymiarach $M \times N$. Tak zbudowany wektor cech będzie zawierał MN komponentów i taka też będzie jego wymiarowość. Operowanie wektorami o tak dużych rozmiarach związane jest ze wzrostem czasu pracy procedur obróbki obrazów i całego algorytmu rozpoznawania. Dlatego w celu zmniejszenia wymiarowości przestrzeni cech w praktyce obraz wejściowy przed konkatencją jest odpowiednio pomniejszany (obraz 3). W trakcie tej operacji należy się jednak liczyć z możliwością utraty części informacji dotyczących indywidualnych cech twarzy. Inne podejście polega na

rozpatrywaniu obrazu wejściowego jako zbioru kolumn i zbioru wierszy (obraz 4), będących samodzielnymi wektorami cech.



Rys. 1.11. Opis twarzy za pomocą różnych kategorii cech (rys. na podst. [12])

Wymiarowość przestrzeni cech w tym przypadku wynosi $\max\{M,N\}$, a liczba samodzielných wektorów cech jest proporcjonalna do $(M+N)$. W celu redukcji cech wykorzystuje się metodę komponentów głównych (PCA) oraz transformację Karhunen-Loeva (KLT). Redukcja stosowana jest dwa razy – oddzielnie dla kolumn i wierszy każdego obrazu. W rezultacie redukcji cech, zamiast każdego obrazu wejściowego, otrzymujemy macierz komponentów głównych posiadającą odpowiednio mniejszy wymiar. Oprócz tego sposób ten czyni system rozpoznawania bardziej stabilnym w zakresie zmian jasności obrazów, przesunięć twarzy w polu ramki wejściowego przedstawienia, a także stabilnym w stosunku do szumów nakładających się na obraz wejściowy. Ponadto metoda ta zapewnia uzyskanie wyższego współczynnika prawidłowego rozpoznawania niż tradycyjne podejście.

Kolejny sposób to przedstawienie twarzy w formie cech spektralnych. Proces ich wydzielenia polega na przeprowadzeniu dwuwymiarowego dyskretnego przekształcenia Fouriera obrazu twarzy, obliczeniu absolutnej wartości otrzymanego widma, a następnie wykonywaniu cyklicznego przesunięcia (do połowy rozmiaru wyjściowego obrazu) widma w dwóch osiach (obraz 5). Ponieważ widmo jest symetryczne, rozmiar wyjściowej przestrzeni cech może być zmniejszony podwójnie. Oprócz tego doświadczenia pokazują, że pełna informacja o obrazie twarzy zawarta jest w całości w 20 przestrzennych składowych. Z tego powodu wyjściowa przestrzeń cech dla każdego obrazu zawiera maksymalnie $(20*20)/2 = 200$ elementów.

Obrazy 6 i 7 to przedstawienie obrazów twarzy za pomocą modeli konturowych twarzy oraz współrzędnych punktów antropometrycznych. Wymiarowość wektora cech dla tych przypadków wynosi odpowiednio: $DIM=100 \div 200$ oraz $DIM=20 \div 40$. Problemem jest tutaj bardzo skomplikowana metoda wyboru odpowiednich cech oraz wymagania co do dokładności wyznaczenia współrzędnych konturów i antropometrycznych punktów na twarzy. Z tego właśnie powodu modele konturowe twarzy oraz współrzędne punktów antropometrycznych stosowane są raczej w interaktywnych systemach rozpoznawania.

Tabela 1 przedstawia rozmiary przestrzeni cech dla wszystkich omówionych wyżej sposobów wyboru cech.

Tabela 1. Rozmiary przestrzeni cech dla różnych metod (źródło [12])

Sposób Wyboru cech	0	1,2	3	4	5	6	7
<i>DIM</i>	<i>MN</i>	<i>MN</i>	$\geq (M/n*N/k)$	$\max(M,N)$	≥ 200	$100 \div 200$	$20 \div 40$

Identyfikacja twarzy

Etap identyfikacji polega na porównywaniu badanego obrazu twarzy z przechowywanymi w bazie wzorcami. W zależności od przyjętej metody rozpoznawania, ilość przechowywanych wzorców oraz forma ich zapisu jest różna. Wyróżniamy dwa podejścia: pierwsze, nazywane strukturalnym, bazuje na cechach anatomicznych twarzy oraz na statystycznych wartościach obrazu twarzy i traktuje twarz jako zespół podstawowych członów występujących w określonych relacjach względem siebie. Z kolei podejście decyzyjno-teoretyczne zakłada, że obiekt reprezentowany jest wektorem liczb otrzymanych przez odpowiednio przeprowadzone pomiary (np. próbkowanie, skanowanie). W projektowanym systemie wykorzystano podejście decyzyjno-teoretyczne – każda twarz reprezentowana jest przez wektor cech (o ustalonej wymiarowości) zbudowany na podstawie współrzędnych charakterystycznych punktów twarzy, wyznaczonych w drodze pomiarów.

Zapisane w taki sposób wzorce zgrupowane zostały następnie w klasy reprezentujące poszczególne twarze. Głównym problemem w tym przypadku jest zapewnienie odrębności obszarów poszczególnych klas i wykorzystanie metod umożliwiających skuteczne badanie różnic pomiędzy nimi. Istnieje wiele metod określania przynależności badanej twarzy do określonej klasy. Najbardziej popularne z nich to:

- metody minimalnoodległościowe,
- metody wzorców,
- metody aproksymacyjne,
- metody drzewowe i grafowe.

Omówienie ich wszystkich znacznie wykracza poza ramy niniejszej pracy, dlatego ograniczono się tylko do opisu jednej z nich, wykorzystanej w przedstawianym systemie metody αNN będącej przedstawicielem grupy metod minimalnoodległościowych. Metody te są często stosowane w praktyce ze względu na swą prostotę i intuicyjność, ale także dlatego, że dają stosunkowo dobre wyniki.

Niech I oznacza zbiór indeksów klas, D zbiór obiektów podlegających rozpoznawaniu, U jest ciągiem uczącym zdefiniowanym jako zbiór par:

$$U = \left\{ \langle \underline{x}^k, i^k \rangle, \quad k = 1, 2, \dots, N \right\} \quad (1)$$

Elementy tego zbioru to wzorce, zawierające pełną charakterystykę danego obiektu d^k i informacje na temat numeru klasy i^k do której ten obiekt powinien być zaliczony.

Ogólna zasada działania metod minimalnoodległościowych polega na wybraniu jako wyniku $i \in I$ tej klasy, do której należy obiekt $x^{i,k} \in U$ najbliższy (w sensie przyjętej metryki ρ) rozpoznawanemu obiektowi d (reprezentującemu go wektorowi cech x). Oznaczając przez C odwzorowanie opisujące stopień podobieństwa nieznanego obiektu $d \in D$ do poszczególnych klas D_i indeksowanych numerami $i \in I$ i zakładając, że w przestrzeni X zdefiniowano pewną metrykę (np. euklidesową) można odwzorowanie C zapisać w postaci:

$$C^i(\underline{x}) = \frac{1}{\rho(\underline{x}, \underline{x}^{i,k}) + \varepsilon}, \quad i = 1, 2, \dots, L \quad (2)$$

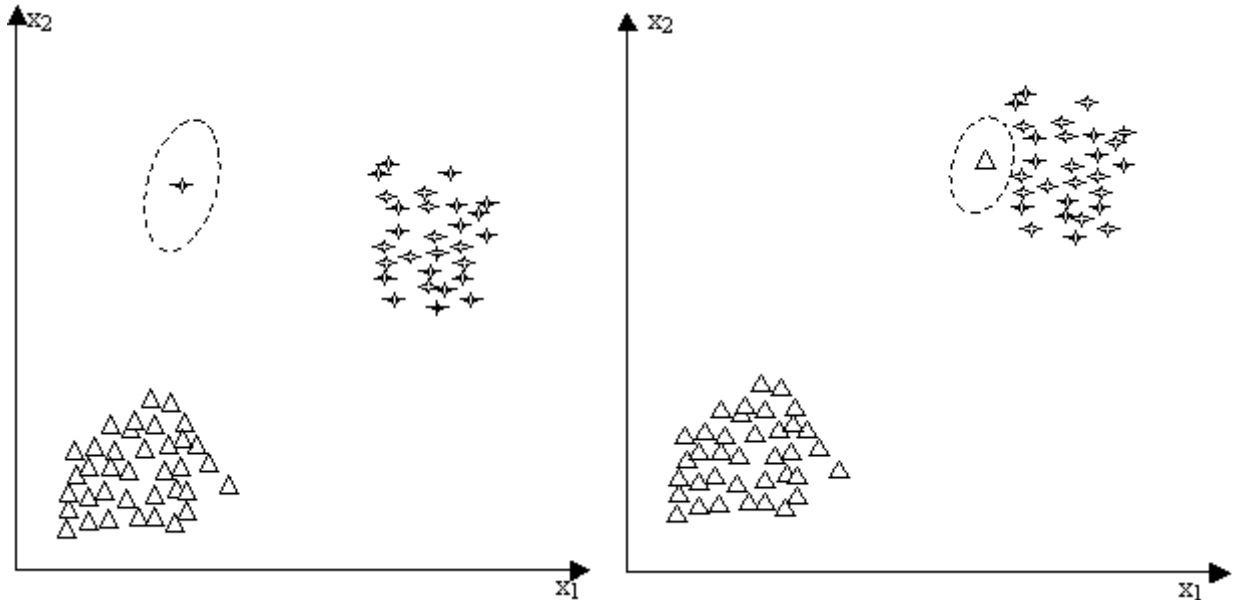
Element $\underline{x}^{i,k}$ należy do podzbioru U^i a mała stała dodatnia została wprowadzona w celu zapewnienia warunku $C^i(\underline{x}) < \infty$, gdyż teoretycznie możliwe jest uzyskanie pokrycia punktów

$$\rho(\underline{x}, \underline{x}^{i,k}) = 0 \quad (3)$$

Wybór elementu $\underline{x}^{i,k}$ w podstawowym wariacie NN (ang. *nearest neighbour*) dokonywany jest zgodnie z regułą:

$$\rho(x, x^{i,k}) = \min_{\underline{x}^\mu \in U^i} (\underline{x}, \underline{x}^\mu) \quad (4)$$

Metoda ta w tej postaci ma jednak liczne wady. Jedną z nich jest duża wrażliwość na błędy ciągu uczącego U . Jeśli błędnie określona zostanie przynależność i^k chociaż jednego elementu \underline{x}^k ciągu uczącego U (rys. 1.12), to wówczas całe jego otoczenie będzie błędnie sklasyfikowane.



Rys. 1.12. Błędne położenie i sklasyfikowanie elementu ciągu uczącego (rys. na podst. [20])

Aby ograniczyć ten efekt wprowadza się dodatkowy parametr α , którego wybór determinuje własności metody (rys. 1.13). Parametr ten wybierany jest arbitralnie, jednak w ten sposób, aby

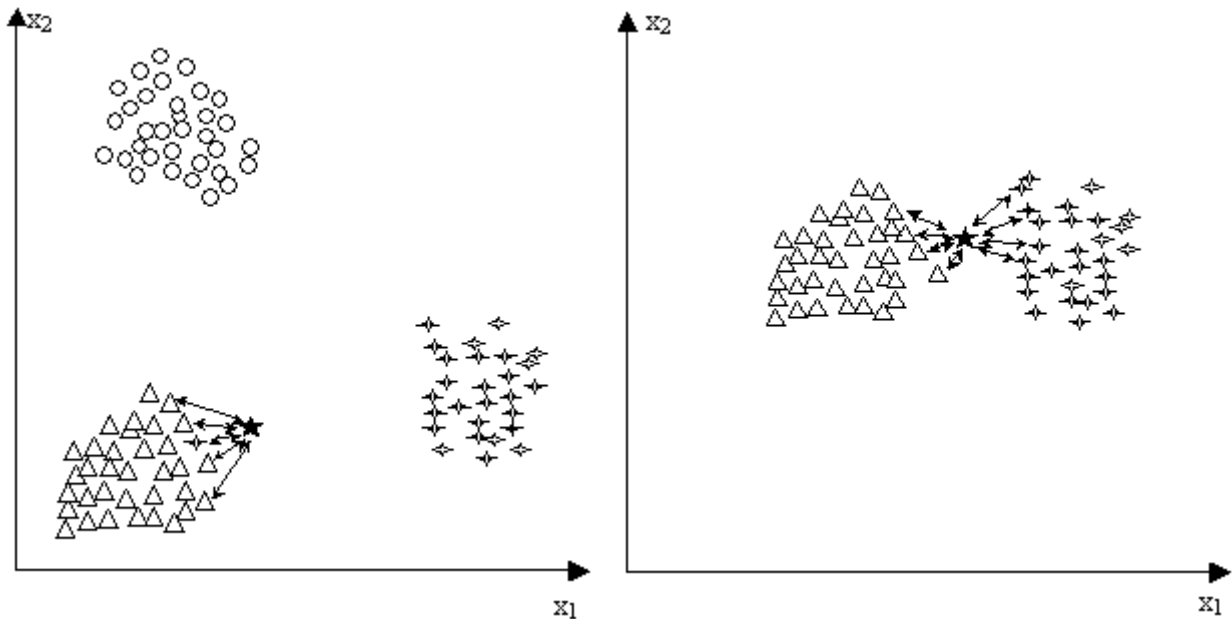
$$\alpha \ll \min_{i \in I} N^i \quad (N - \text{liczba el. ciągu uczącego}) \quad (5)$$

W praktyce α jest małą liczbą dodatnią. Po określeniu α i pojawieniu się obiektu rozpoznawania \underline{x} obliczane są wartości odległości tego obiektu od wszystkich obiektów ciągu uczącego $\rho(\underline{x}, \underline{x}^k)$, $k = 1, 2, \dots, N$. Następnie dokonuje się uporządkowania ciągu uczącego według rosnących odległości, wprowadzając numery $v = 1, 2, \dots, n$:

$$U = \left\{ \langle \underline{x}^v, i^v \rangle, \quad \varphi = 1, 2, \dots, N \right\} \quad (6)$$

przy czym

$$\forall_{v, \mu} [v > \mu \Rightarrow \rho(\underline{x}, \underline{x}^v) > \rho(\underline{x}, \underline{x}^\mu)]. \quad (7)$$



Rys. 1.13. Zasada działania metody α NN (rys. na podst. [20])

Następnie wybiera się α początkowych obiektów ciągu, tworząc podzbiór

$$U^\alpha = \left\{ \langle \underline{x}^v, i^v \rangle, \quad v = 1, 2, \dots, \alpha \right\} \quad (8)$$

rozbijany na podzbiory (mogą być puste) związane z poszczególnymi klasami

$$U^{\alpha,i} = \left\{ \langle \underline{x}^v, i^v \rangle, \quad v < \alpha \wedge i^v = i \right\} \quad (9)$$

Funkcje przynależności, których wartości określają miarę przynależności danego obiektu d do poszczególnych klas D^i ($i = 1, 2, \dots, L$) dla metody α NN można teraz wyznaczyć na podstawie liczebności podzbiorów $U^{\alpha,i}$.

$$C^i(\underline{x}) = \#U^{\alpha,i}; \quad i = 1, 2, \dots, L \quad (10)$$

Weryfikacja i kontrola poprawności

Etap weryfikacji ma na celu zminimalizowanie liczby błędnych decyzji podejmowanych przez system i ocenę wiarygodności uzyskanych rezultatów. Sama weryfikacja odbywa się praktycznie na każdym z etapów identyfikacji twarzy, począwszy od jej lokalizacji na obrazie wejściowym (np. odrzucanie obiektów nie będących twarzami). Weryfikacja rezultatu rozpoznawania pozwala na potwierdzenie otrzymania poprawnego wyniku, a w przypadku błędnej decyzji - na próbę jej wykrycia. Błędy generowane przez system rozpoznawania dzieli się na trzy główne grupy [19]:

- b) *błąd fałszywego odrzucenia* - obraz twarzy posiadający wzorec w danej bazie danych zostaje nierozpoznany i odrzucony jako obraz nie posiadający swojej reprezentacji,

- *błąd błędnej klasyfikacji* - obraz twarzy posiadający wzorec w bazie danych zostaje błędnie przypisany do innego wzorca z bazy,
- *błąd fałszywej akceptacji* - obraz twarzy nie posiadający wzorca w bazie danych zostaje błędnie przypisany do wzorca tam istniejącego.

W niektórych przypadkach ich wykrycie może być trudne, ale nawet wtedy otrzymuje się informacje o stopniu wiarygodności otrzymanych rezultatów. Aby zminimalizować prawdopodobieństwo wystąpienia błędów procedury weryfikacji w jednym systemie implementowane są na wiele sposobów:

- poprzez dodanie testu/weryfikacji z rozpoznawaniem dodatkowego obrazu lub najistotniejszych części twarzy,
- poprzez zaimplementowanie w jednym systemie dwóch różnych algorytmów rozpoznawania, pierwszy traktowany jest jako podstawowy a drugi jako dodatkowy służący w etapie weryfikacji,
- poprzez użycie kodu PIN (ang. *Personal Identification Number*) przez użytkownika w procesie weryfikacji,
- zgromadzenie i wykorzystanie różnych biometrycznych i biomedycznych informacji w jednym systemie (na przykład twarz + głos, twarz + linie papilarne itp.),
- poprzez oparcie zadań identyfikacji i rozpoznawania na cechach uzyskiwanych z różnych kategorii, ale powiązanych z jedną osobą (na przykład statystyczne i strukturalne cechy twarzy).

To jakie procedury zostaną użyte zależy od specyfiki systemu, wymagań czasowych oraz wymaganej niezawodności.

1.3. Metody rozpoznawania twarzy

Istnieje wiele sposobów klasyfikacji metod rozpoznawania twarzy. Ogólnie można je podzielić na dwie główne grupy:

- metody analityczne – w których pod uwagę bierze się anatomiczne cechy twarzy
- metody bazujące na podejściu globalnym – w których obraz twarzy traktuje się jako całość.

Tabela 2. Ogólna klasyfikacja metod rozpoznawania twarzy

Metody analityczne	Metody globalne
metody sztucznej inteligencji (sieci neuronowe, logika rozmyta)	rozpoznawanie w oparciu o jeden wzorzec
metody matematyczne	rozpoznawanie w oparciu o bazę wzorców

Metody analityczne

W podejściu analitycznym lokalizuje się charakterystyczne punkty na obrazie twarzy i mierzy się względne odległości między tymi punktami. Pod uwagę bierze się takie cechy jak:

- odległości między środkami oczu,
- odległości między najdalszymi punktami oczu,
- odległości pomiędzy oczami i nosem,
- odległości pomiędzy linią oczu i linią ust,
- odległości od najdalszego punktu oka do środka ust,
- szerokość i wysokość twarzy,
- szerokość i długość nosa, ust,
- matematyczny opis kształtu twarzy (owal twarzy).

Poszczególne cechy zostają zapisane w postaci wektora cech. Identyfikacja polega na porównaniu wektora cech reprezentującego badaną twarz z przechowywanymi w bazie danych wzorcami, poprzez badanie odległości (zgodnie z przyjętą metryką) pomiędzy nimi. Rezultat rozpoznawania określa się na podstawie minimum odległości.

Metody globalne

W metodach globalnych wykorzystywany jest cały obraz twarzy lub jego najbardziej reprezentatywna część. Identyfikacja polega na porównywaniu obrazu bieżącego z każdym z wzorców przechowywanych w bazie danych. Może być realizowana z wykorzystaniem następujących metod:

- metoda polegająca na bezpośrednim porównaniu obrazu bieżącego ze wzorcem (w przestrzeni cech wejściowych) z obliczeniem odległości pomiędzy nimi na podstawie dowolnej metryki,
- metoda polegająca na bezpośrednim porównaniu obrazu bieżącego ze wzorcem z obliczeniem miary podobieństwa pomiędzy dwoma obrazami na podstawie korelacji,

- metoda polegająca na porównaniu obrazu bieżącego ze wzorcem w przestrzeni cech zredukowanych z wykonaniem redukcji cech na podstawie transformacji Karhunenena–Loeva i koncepcji twarzy własnych (ang. „*Eigenface*”),
- metody oparte na LDA i koncepcji „*Fisherface*”,
- metody będące kombinacjami powyższych.

Jedną z prostszych jest metoda korelacyjna. Operuje ona w wejściowej przestrzeni cech. W metodzie tej bada się istnienie wzajemnego powiązania pomiędzy obrazem rozpoznawanym a wzorcem. Jeśli warunki ich pozyskiwania były zbliżone (oświetlenie, skala, pochylenie i obrót) to korelacja pomiędzy nimi jest bliska 1 a skuteczność rozpoznawania dochodzi do 90%. W przeciwnym wypadku, korelacja nie wykrywa podobieństwa dwóch obrazów i staje się nieskuteczna. Z racji tego, iż rozpoznawanie odbywa się w wejściowej przestrzeni cech, korelacja jest złożonym obliczeniowo procesem. Wyniki rozpoznawania można polepszyć rozbudowując bazę danych różnymi wariantami przechowywanych w niej wzorców (kosztem jej wielkości). Lepsze rezultaty można osiągnąć przechodząc z wejściowej przestrzeni cech do przestrzeni inwariantów Fouriera-Melina (eliminują skalę i przedstawiają obrót jako cykliczne przesunięcie) i na podstawie filtracji dopasowującej osiągnąć wysokie oceny korelacji pomiędzy dwoma obrazami. W dalszym ciągu jednak problem stanowi wielkość przestrzeni cech odpowiadająca rozmiarom obrazu wejściowego.

Metodą umożliwiającą znaczną redukcję wymiarowości wejściowej przestrzeni cech jest transformacja Karhunenena–Loeva. Do dalszego wykorzystania pozostawiane są tylko te cechy, które mają kluczowe znaczenie dla konkretnej klasy obrazów. Stopień prawidłowego rozpoznawania osiąga wartość 80% i jest stabilny, niezależnie od warunków otrzymania obrazów twarzy w stosunku do przechowywanych w bazie wzorców. Metoda dyskryminacji liniowej (LDA) jest przedstawicielem rodziny metod statystycznych. Dzieli przestrzeń cech na obszary ograniczone funkcjami liniowymi. Może być stosowana zarówno w wejściowej przestrzeni cech jak i zredukowanej, przy czym w tej drugiej osiąga się znacznie lepszą jakość klusteryzacji. Pozwala na zwiększenie współczynnika rozpoznawania do 90% przy zmiennych warunkach uzyskania obrazów.

1.4. Systemy rozpoznawania twarzy

Systemy rozpoznawania mogą realizować swoje zadania działając w dwóch trybach: trybie „off-line” lub „on-line”. Determinuje to problemy, jakie projektant systemu musi uwzględnić, aby zapewnić ich prawidłowe działanie. Systemy klasy „off-line” operują zazwyczaj na dużych bazach danych i realizują zadanie identyfikacji. Obrazy wejściowe są często poddawane ręcznemu procesowi normalizacji (interaktywny system rozpoznawania) – upraszcza to cały proces polepszając jednocześnie rezultaty rozpoznawania. Szybkość działania takich systemów nie jest czynnikiem decydującym w ocenie końcowej ich działania. W odróżnieniu od nich systemy pracujące w trybie „on-line” muszą sobie radzić z szeregiem problemów związanych z pracą w czasie rzeczywistym. Związane są one przede wszystkim z obróbką wstępną obrazów pozyskanych z urządzenia wejściowego (kamery). Problemy te dotyczą [12]:

- lokalizacji twarzy z otaczającego tła albo odróżnienia jej od innych twarzy,
- wybrania najbardziej informacyjnej części twarzy,
- dopasowania rozmiarów twarzy,
- przekształcenia twarzy zgodnie z przechowywanymi w bazie wzorcami (eliminacja obrotów),
- eliminacji wpływu mimiki, fryzury lub innych niepożądanych obiektów (okulary),
- eliminacji wpływu oświetlenia.

Ilość możliwych wariantów obrazu wejściowego twarzy jest bardzo duża - nawet w przypadku, kiedy system ma dostęp do bazy danych zawierającej dużą ilość wzorców w danej klasie, nie jesteśmy w stanie przewidzieć wszystkich, dlatego uzyskanie 100% skuteczności rozpoznawania jest praktycznie niemożliwe. Jednak zasadniczy wpływ na strukturę bazy danych mają wykorzystywane metody rozpoznawania. Ich odpowiedni dobór pozwala na uniezależnienie się w pewnym stopniu od niekompletności bazy danych. W przypadku niektórych z nich (np. PCA + LDA) dostęp do dużej ilości materiału wejściowego jest jednak warunkiem koniecznym do ich poprawnego działania. Należy także pamiętać o tym, że każda kolejna operacja porównania wydłuża proces identyfikacji – w systemach klasy „on-line” nie pozostaje to bez znaczenia.

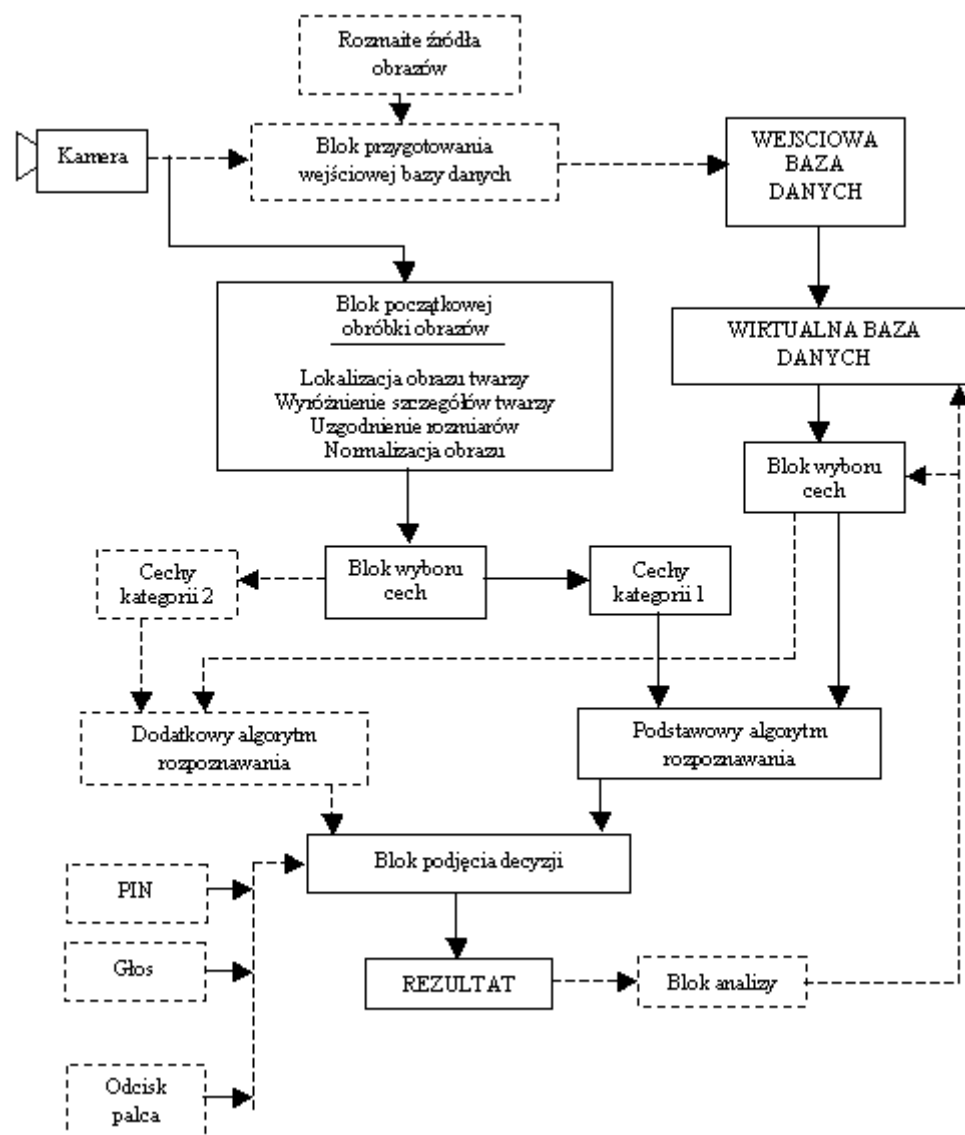
Można zatem w stosunku do tej klasy systemów sformułować następujące wymagania:

- zapewnienie wysokiej jakości rozpoznawania,
- nieskomplikowane algorytmy zapewniające dużą szybkość działania,
- łatwość modyfikacji pod kątem dopasowania do nowej bazy danych,
- eliminowanie podejmowania błędnych decyzji.

Wysoki współczynnik rozpoznawania może zostać zapewniony poprzez zapewnienie kompletności wejściowej bazy danych i wysoką jakość użytych metod, wykonywanie procedur obróbki wstępnej obrazu (*ang. preprocessing*) na nowych obrazach, takich jak: specjalna normalizacja, przekształcenie obrazów do form obrazów zgromadzonych w bazie danych itp. Oprócz tego proces rozpoznawania powinien być wspierany procedurami weryfikacyjnymi.

Ogólny algorytm działania takich systemów polega na sprawdzeniu przynależności pobranego z kamery obrazu wejściowego do zapisanych w bazie klas: albo poprzez wyliczenie miary podobieństwa pomiędzy obrazem wejściowym i każdym wzorcem, albo poprzez obliczenie miary podobieństwa między pewną charakterystyką obrazu wejściowego i wspólną charakterystyką w klasie obrazu. Obraz będzie należeć do tej klasy, dla której miara będzie największa i przy warunku, że ta miara jest większa od pewnego progu identyfikacji.

Jeden z możliwych wariantów systemu „on-line” przedstawia rysunek 1.14 [12].



Rys. 1.14. Struktura systemu rozpoznawania on-line (rys. na podst. [12])

Głównymi blokami systemu są:

1. Cyfrowa kamera – źródło obrazu wejściowego
2. Blok wstępnego obrabiania obrazów z kamery, zawierający:
 - 2.1 Blok lokalizacji twarzy z całego obrazu;
 - 2.2 Blok ekstrakcji cech twarzy;
 - 2.3 Blok normalizacji wejściowego obrazu i uzgodnienia rozmiaru;
 - 2.4 Blok eliminowania obrotów i nachyleń obrazu głowy;
3. Baza danych (BDT);
4. Wirtualna baza danych (WBT);
5. Bloki wyboru cech;
6. Blok realizujący podstawowy algorytm rozpoznawania;
7. Blok realizujący dodatkowy algorytm rozpoznawania;

8. Blok podjęcia decyzji;
9. Blok wprowadzenia PIN i bloki wprowadzenia innych kluczy;
10. Blok odzwierciedlający rezultat rozpoznawania;
11. Blok analizy rezultatu rozpoznawania.

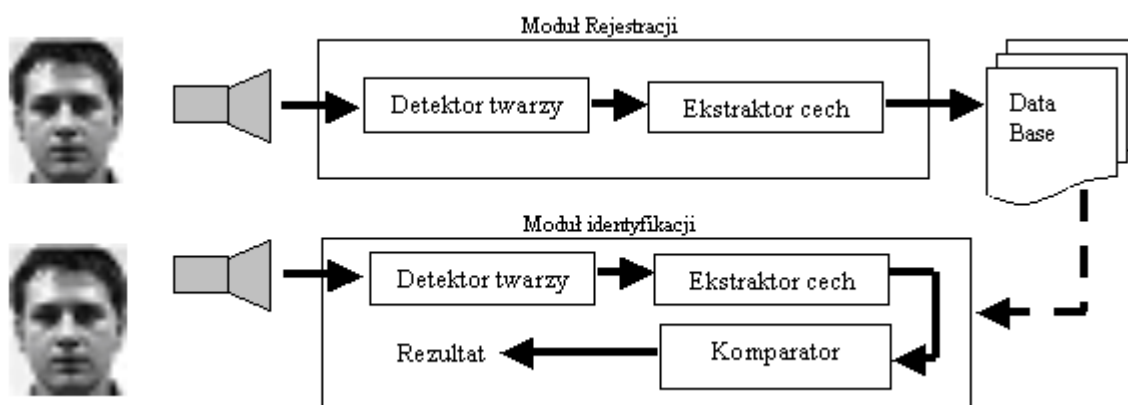
Baza danych tworzona jest z różnych źródeł (zdjęcia, inne bazy) albo bezpośrednio z obrazów napływających z kamery. Każdy obraz przechodzi etap obróbki wstępnej, w trakcie której następuje lokalizacja twarzy na obrazie, ekstrakcja cech twarzy, normalizacja rozmiarów twarzy, oświetlenia, wyznaczenie niektórych parametrów twarzy. Na podstawie wejściowej bazy danych tworzona jest tzw. wirtualna baza danych (WBT), zawierająca obrazy powstałe w wyniku przekształceń obrazów wejściowych – skalowania, wyróżnienia osobnych regionów. Określa się w niej które obrazy mają być wykorzystane do syntezy systemu a które mają pełnić funkcje kontrolne czy służyć do analizy rezultatów rozpoznawania. Ma ona szczególne znaczenie w przypadku, gdy BDT zawierają po jednym egzemplarzu danych w każdej klasie – WBT przechowuje wtedy ich różne, sztucznie wygenerowane, warianty. Pozwala ona także na uniezależnienie się w pewnym stopniu od wielkości obrazu wejściowego poprzez dynamiczne tworzenie na jego podstawie nowych obrazów – wzorców, których rozmiary odpowiadają tym przechowywanym w bazie danych. Wybór cech zdeterminowany jest realizowanym algorytmem rozpoznawania. W bloku podjęcia decyzji następuje analiza otrzymanych wyników. Tutaj podejmowane są decyzje o ewentualnym wykorzystaniu dodatkowego algorytmu rozpoznawania w celu potwierdzenia otrzymanego rezultatu. Decyzja końcowa podejmowana jest na podstawie wyniku rozpoznawania podstawowego systemu i rezultatów identyfikacji z wykorzystaniem dodatkowych metod: odcisków palców, głosu, tęczówki, PIN itp. Informacje o rezultatach rozpoznawania mogą poprzez blok analizy uzupełniać istniejącą bazę danych.

Na podstawie przedstawionej struktury budowane są wyspecjalizowane systemy, realizujące rozpoznawanie twarzy w różnych zastosowaniach np. systemy „Visitor Identification”, realizujące zadanie dostępu do pomieszczeń systemy „Access-Control” czy systemy „Name-It” służące do identyfikacji twarzy osób na planie kadru wideo.

1.4.1. Systemy „Visitor Identification”

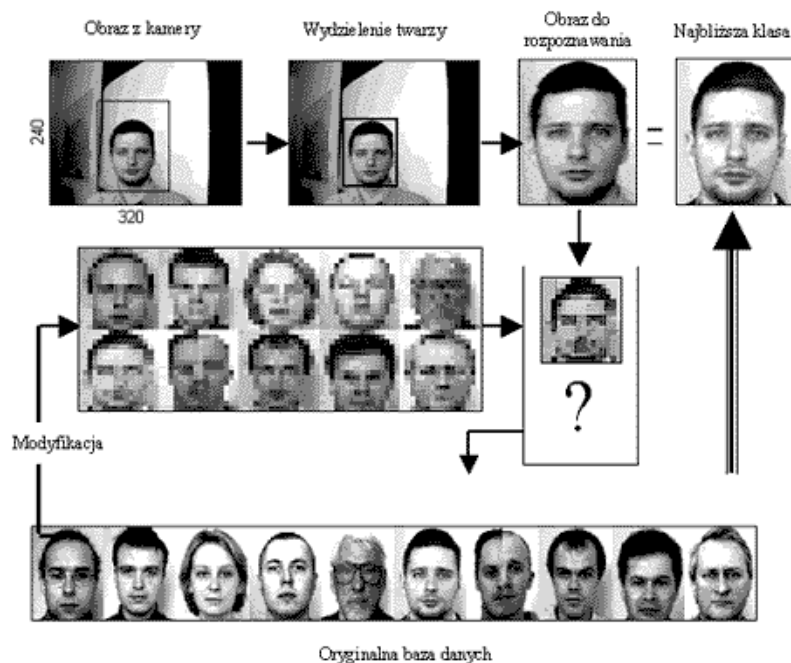
Systemy klasy “Visitor Identification” realizują rozpoznawanie związane z identyfikacją stałych klientów sklepu, gości w biurze, w domu, odwiedzających strony internetowe, obsługujących automaty z napojami, pasażerów lotnisk, kibiców na stadionach

itp. Wymaganie dotyczące pracy w czasie rzeczywistym narzuca określone reguły ich tworzenia i wyboru metod rozpoznawania. Zwykle systemy „*Visitor Identification*” są zorientowane na bazy danych zawierające od kilku do kilkudziesięciu klas. Dla większej liczby klas stosuje się omówione dalej systemy „*Access-Control*”. Rysunek 1.15. przedstawia ogólny schemat systemu „*Visitor Identification*” oparty na opisaniej wcześniej bazowej strukturze systemu rozpoznawania.



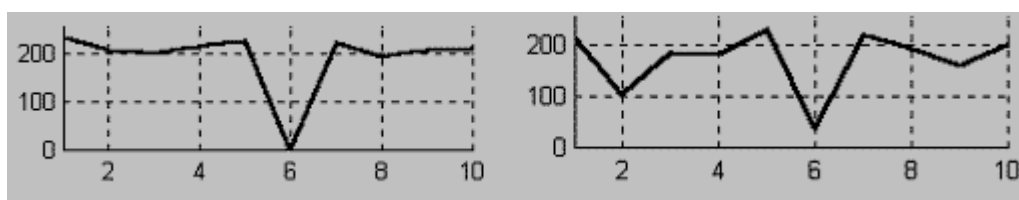
Rys. 1.15. Struktura systemu „*Visitor Identification*” (rys. na podst. [12])

System składa się z dwóch modułów – rejestracji i identyfikacji. Moduł rejestracji służy do budowy bazy danych (zarejestrowani użytkownicy). Moduł identyfikacji realizuje podstawową funkcję – rozpoznawanie. Zawiera blok komparatora, który dokonuje analizy zdjęć odwiedzającego system, wydziela zarejestrowanych użytkowników i ich identyfikuje. Na rysunku 1.16 pokazano proces rozpoznawania nowego obrazu w przykładowym systemie „*Visitor Identification*” [12]:



Rys. 1.16. Proces rozpoznawania obrazu w systemie „Visitor Identification” (Źródło: [12])

Baza danych zawiera po jednym obrazie w każdej klasie. Obraz twarzy po jej detekcji jest automatycznie wyszukiwany w polu zaznaczonej ramki. Następnie on, jak i wszystkie obrazy z bazy skalowane są do rozdzielczości 16x16. Zmniejszenie rozdzielczości powoduje, że cały algorytm rozpoznawania działa szybciej. W trakcie tej operacji eliminowany jest wpływ skali (w zakresie +/- 10%), przesunięcia twarzy (w zakresie 1-2% rozmiarów ramki), obrotu głowy (+/- 15%). Następnie rozpoznawany obraz jest porównywany z każdym przeskalowanym obrazem z bazy piksel po pikselu z wykorzystaniem ustalonej metryki. Jeśli wartość bezwzględna różnicy pikseli przewyższa pewien zadany próg, to parametr miary przekroczenia progu zwiększa się o 1. W rezultacie otrzymuje się szereg, którego elementami są parametry miar przekroczenia progu. Rozpoznawany obraz przypisywany jest do tej klasy, dla której wartość osiąga minimum.

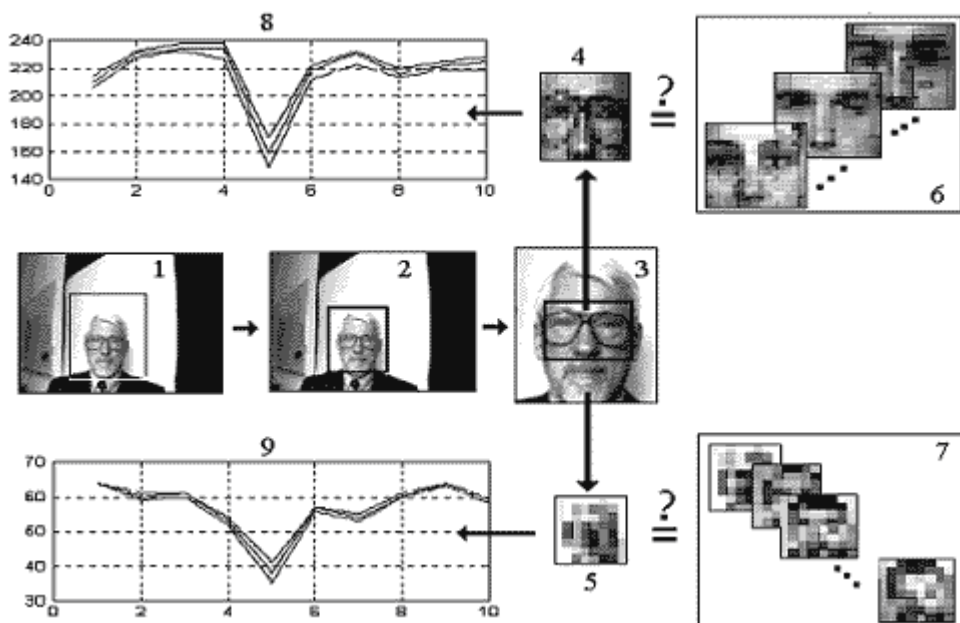


Rys. 1.17. Funkcja przekroczenia progu

Realizowany w ten sposób algorytm jest bardzo prosty. Nie wymaga skomplikowanych obliczeń, ale jego skuteczność jest bardzo ograniczona. Poprawnie wykonywana jest identyfikacja jedynie tych obrazów, które pobierane były

w porównywalnych warunkach co odpowiadające im wzorce w bazie (oświetlenie, rozmiar, warianty ustawienia twarzy w stosunku do przechowywanych w bazie wzorców). Ponadto system jest bardzo wrażliwy na zmiany w wyglądzie zewnętrznym: kształt fryzury, owal twarzy, zarost itp. System ma 100% poziom fałszywej identyfikacji dla obrazów nie należących do bazy danych (także w stosunku do obrazów nie przedstawiających w ogóle twarzy). Wyniki rozpoznawania można poprawić poprzez przechowywanie w bazie danych 3-5 wariantów obrazów dla każdej klasy, co i tak gwarantuje skuteczność rozpoznawania na poziomie co najwyżej 55%, wydłużając jednocześnie cały proces. Ponieważ nie są używane żadne metody redukcji wejściowej przestrzeni (wzorce przechowywane są w bazie jawnie), system wymaga znaczących obszarów pamięci do działania. Oprócz tego należy cały czas minimalizować albo eliminować wpływ wszystkich przeszkadzających czynników, aktualizować bazę danych bieżącymi obrazami odwiedzającego. Opisany system nie spełnia zatem podstawowych wymagań stawianych systemom rozpoznawania twarzy i dlatego jego praktyczne zastosowanie jest bardzo ograniczone.

Znaczącą poprawę wyników można osiągnąć wprowadzając dodatkowe metody weryfikacji otrzymanych rezultatów. W przypadku systemów „Visitor Identification” polega to zazwyczaj na przeprowadzeniu testu z rozpoznawaniem dodatkowego obrazu lub jego najistotniejszej części lub wykorzystaniu dwóch różnych algorytmów rozpoznawania – podstawowego i dodatkowego, służącego do weryfikacji otrzymanych wyników. Na rysunku 1.18 przedstawiono zmodyfikowany proces rozpoznawania z wykorzystaniem wymienionych metod.



Rys. 1.18. System „Visitor Identification” z dodatkową weryfikacją rezultatu (Źródło: [12])

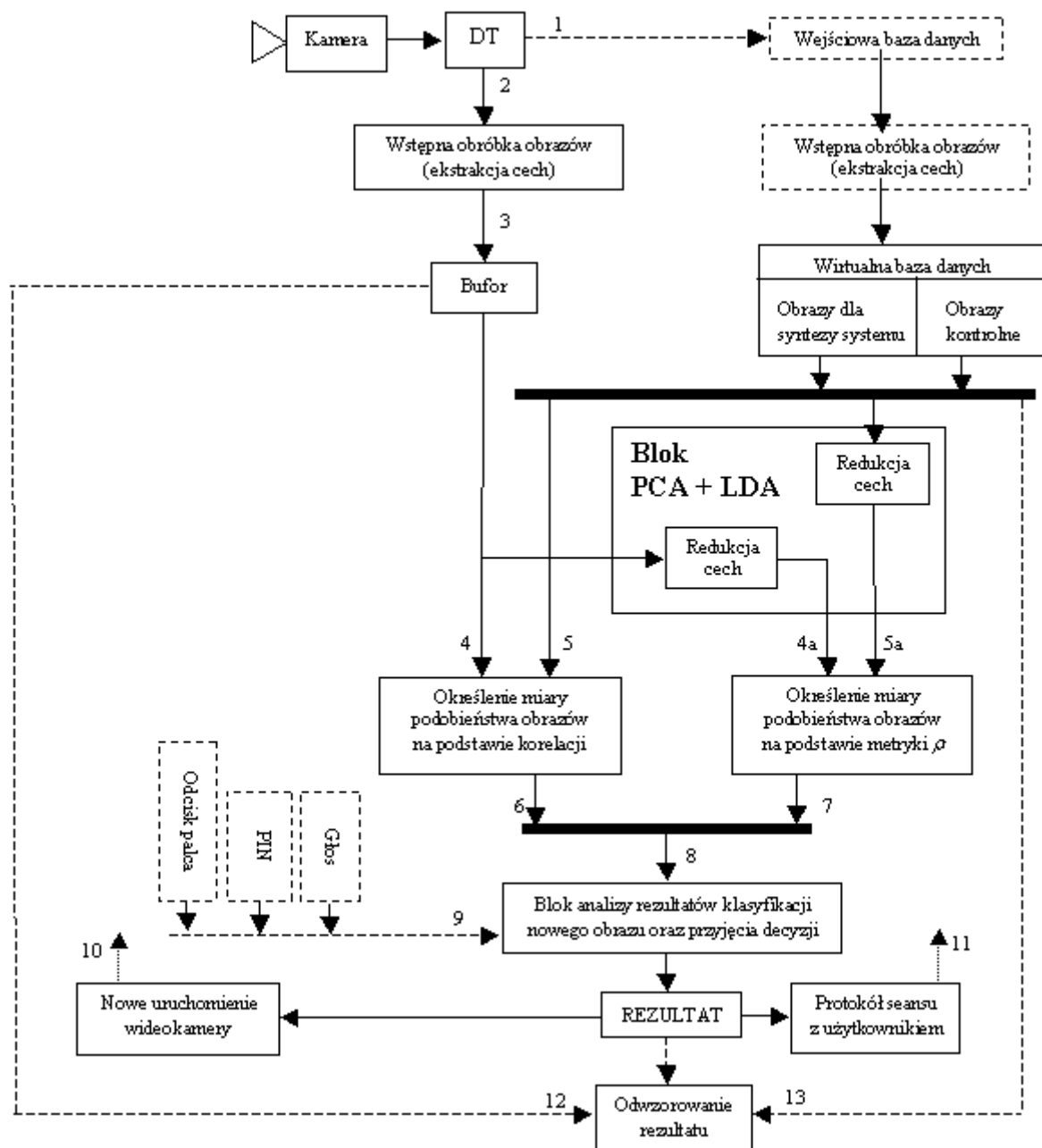
Podstawową różnicą w stosunku do opisanego wcześniej systemu jest uzupełnienie bazy danych przeskalowanymi, niosącymi najwięcej informacji o indywidualnych cechach twarzy, obrazami środkowej części twarzy uzyskanymi z obrazów wejściowych. Na rysunku numerem 1 oznaczono wynik detekcji twarzy na obrazie z kamery, 2 – zlokalizowaną twarz, 3 – wybraną ramkę obrazu do identyfikacji. Pierwszy etap rozpoznawania odbywa się podobnie jak w opisanym wyżej systemie: obraz 3 skalowany jest do rozmiarów przechowywanych w bazie wzorców a następnie porównywany z każdym wzorcem z bazy (6) piksel po pikselu z wykorzystaniem ustalonej metryki. Drugi etap to weryfikacja otrzymanego wyniku: z wejściowego obrazu wybierana jest środkowa część twarzy, a następnie z wykorzystaniem metody wzajemnej korelacji fazowej porównywana z przechowywanymi w bazie wzorcami (7).

Maksimum funkcji korelacji wskazuje do której klasy należy badany obraz. Numerami 8 i 9 oznaczono funkcje progowe opisujące rezultat rozpoznawania dla każdego z algorytmów. Jeżeli wyniki identyfikacji i weryfikacji są zbieżne, obraz uważany jest za zidentyfikowany i proces rozpoznawania zostaje zakończony.

Opisana modyfikacja pozwala osiągnąć skuteczność rozpoznawania na poziomie dochodzącym nawet do 90%.

1.4.2. Systemy „Access Control”

Zadaniem systemów kontroli dostępu („Access Control Systems”) jest dopuszczenie uprawnionych osób do określonych miejsc. Systemy te znajdują zastosowanie w kontrolowanym dostępie do budynków, pomieszczeń, autoryzacji dostępu do zasobów informatycznych, identyfikacji osób korzystających z bankomatów. Zastosowania te wymuszają niezawodność i ciągłość ich działania. Klasyczne systemy kontroli dostępu opierają się zwykle na kartach (magnetycznych, chipowych), numerach PIN czy też unikalnych identyfikatorach użytkownika. Powoduje to, że dostęp do chronionych zasobów może mieć nie tyle osoba uprawniona, co osoba posiadająca kartę, znająca numer PIN czy identyfikator. Włączenie technik biometrycznych do procesu autoryzacji powoduje faktyczną weryfikację osoby i ustalenie jej uprawnień. Systemy rozpoznawania twarzy typu „Access Control” są tego przykładem. Na rysunku 1.19 przedstawiono jedno z możliwych rozwiązań.



Rys. 1.19. Struktura systemu „Access Control” (rys. na podst. [12])

Nowymi elementami są: blok redukcji wejściowej przestrzeni cech bazujący na metodach PCA+LDA, dwa niezależne algorytmy rozpoznawania oraz wprowadzenie wirtualnej bazy danych (WBT). WBT buduje się z wykorzystaniem dwóch rodzajów obrazów: przeskalowanych obrazów bazy wejściowej oraz obrazów środkowej części twarzy – ze względu na jej najbardziej informacyjny charakter. Ponieważ system powinien potrafić rozdzielać obrazy przedstawiające twarze od tych, które jej nie zawierają, często WBT uzupełniana jest specjalną grupą obrazów – tzw. „pułapek”, określających możliwych nieupoważnionych użytkowników. W grupie tej mieszczą się np. obrazy przyrody, grupowe zdjęcia ludzi, zdjęcia głów zwierząt, przedmioty przypominające owal twarzy.

Z wejściowych obrazów, otrzymanych bezpośrednio z kamery, w bloku DT przeprowadzana jest detekcja twarzy. Jeśli realizowanym etapem jest synteza systemu, dane te uzupełniają WBT (linia 1) - po wykonaniu wstępnej obróbki danych oraz wydzieleniu środkowej części każdej z twarzy. W jej ramach dane zostają podzielone na obrazy wzorce oraz obrazy kontrolne. Używając obrazów kontrolnych (linia 3a → Bufor → linie 4, 4a) można sprawdzić skuteczność działania systemu oraz określić niezbędne progi dla prawidłowego klasyfikatora. Linia 2 → 3 rozpoczyna proces rozpoznawania: w bloku DT przeprowadzana jest detekcja twarzy oraz wybór środkowej części twarzy. Dane zapisywane są w bloku *BUFOR*. Obraz części twarzy z bufora (linia 4) oraz obrazy - wzorce części twarzy z WBT (linia 5) porównywane są w bloku „*Określenie miary podobieństwa na podstawie korelacji*”. Stopień podobieństwa obrazów określany jest za pomocą współczynnika ich wzajemnej korelacji fazowej, czułego na wszystkie zmiany indywidualnych rysów twarzy. Następnie obraz z bufora przetwarzany jest w bloku redukcji cech zgodnie z metodami PCA+LDA (linia 4a) i porównywany z obrazami - wzorcami twarzy z WBT (linia 5a) w bloku „*Określenie miary podobieństwa obrazów na podstawie metryki ρ* ”. Obrazy twarzy przechowywane są w tym przypadku w bazie w postaci niejawnej - w pamięci przechowuje się $QK(K-1)$ (Q - liczba podobrazów w klasie, K - liczba klas, $(K-1)$ - liczba cech na każdy podobraz) cech zredukowanej przestrzeni, parametry systemu i informację o odległości między środkami oczu w każdej klasie podobrazów. Ocena rezultatów porównania przeprowadzana jest w bloku „*Analiza rezultatów klasyfikacji nowego obrazu oraz przyjęcia decyzji*”. Rezultaty porównań z poprzednich etapów docierają do tego bloku wzdłuż linii 6 → 8 oraz 7 → 8. Jako ostatnie są wykorzystywane: odciski palców, głos, tęczówka oka, kształt dłoni, kody PIN, w przypadku których wykorzystuje się specjalne algorytmy i metody rozpoznawania, zespolone z podstawowym algorytmem rozpoznawania (9). Nowy obraz przyjmuje się za rozpoznany, jeśli rezultaty 6, 7 oraz 9 klasyfikują go do tej samej klasy, a miary podobieństwa w obu przypadkach są powyżej zadanego progu.

W systemach rozpoznawania „on-line” poważnym problemem jest zapewnienie porównywalnych warunków pozyskiwania obrazów twarzy. Jednym z czynników, który należy uwzględnić jest oświetlenie. W przypadku, gdy jego wpływ na obrazy wejściowe jest tak duży, że kształt i szczegóły twarzy odgrywają rolę drugorzędną, należy przejść do kategorii innych cech, np. normowanych gradientów jaskrawości. W efekcie wpływ komponentu jaskrawości w wejściowej przestrzeni cech zostaje zminimalizowany a po jej redukcji - praktycznie wyeliminowany. Pozwala to na uniezależnienie jakości rozpoznawania od warunków oświetlenia twarzy przed kamerą.

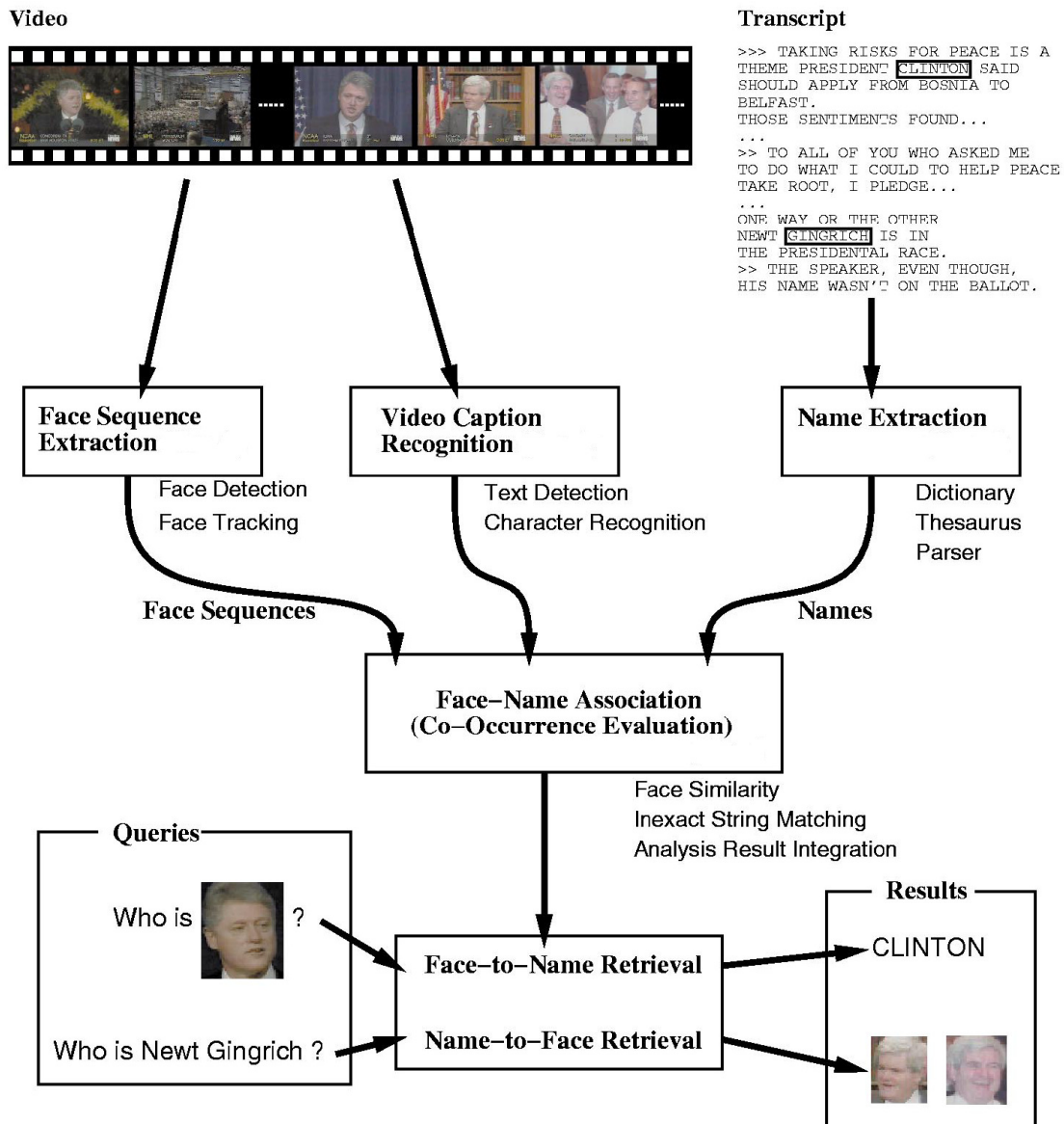
1.4.3. Systemy „Name-It”

Celem zadań klasy „Name-It” jest przypisanie identyfikowanej osobie odpowiedniej nazwy. Źródłem danych są sekwencje kadrów pochodzące m.in. z programów informacyjnych, opisy obrazów wideo, zapisy dźwiękowe głosu identyfikowanej osoby. Systemy realizujące zadania „Name-It” operują zazwyczaj na niewielkich bazach danych (10-15 osób na jeden kanał telewizyjny).

Główne zadania realizowane przez systemy klasy „Name-It” to:

- identyfikacja (nazwanie) osób znajdujących się w kadrze obrazu wideo (dziennikarza, prezentera) na podstawie twarzy,
- identyfikacja osoby na obrazie wideo na podstawie zadanej nazwy,
- identyfikacja osoby na obrazie wideo na podstawie zapisu dźwiękowego jej głosu,
- identyfikacja osoby na obrazie wideo na podstawie opisu ekranowego.

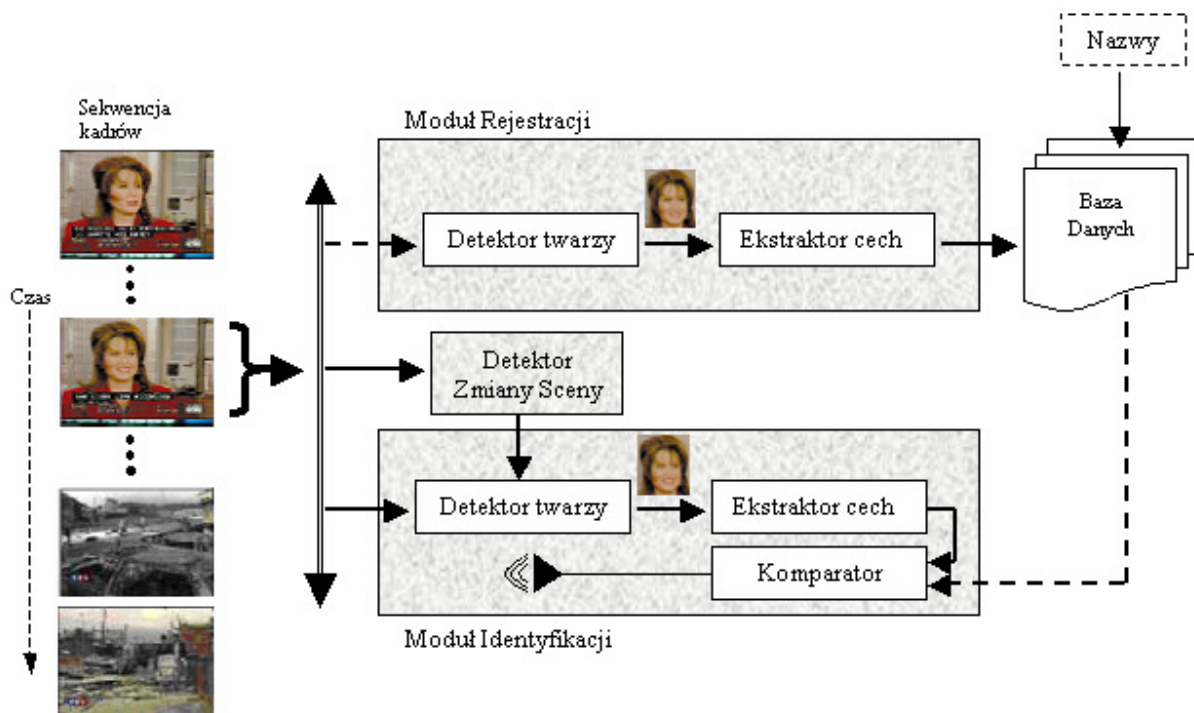
Zadania te powinny być wykonywane w czasie rzeczywistym. Potencjalne zastosowania to np. automatyczna indeksacja obrazów wideo na podstawie nazwy osoby, tworzenie baz danych informacji, w których kluczami są pary: twarz – nazwa, przeglądarki newsów, które w odpowiedzi na nazwę dobierają odpowiedni obraz twarzy, czy interaktywne odtwarzacze wideo dostarczające informacji nt. zaznaczonej twarzy. Na rysunku 1.20 przedstawiono schemat działania systemu „Name-It”.



Rys. 1.20. Struktura procesu „Name-It” (Źródło: [17])

Pierwszym etapem jest ekstrakcja danych wejściowych: detekcja i lokalizacja sekwencji twarzy w ramach jednej sceny, detekcja i rozpoznanie stopki obrazu, ekstrakcja słów kluczowych – nazw osób ze ścieżki dźwiękowej. Z pozyskanej sekwencji twarzy, wybierana jest ta położona najbardziej na wprost względem kamery. Pozwala to na bardziej niezawodną jej identyfikację. Po identyfikacji twarzy lub rozpoznaniu tekstu następuje przypisanie: nazwy do osoby w przypadku, gdy danymi wejściowymi są obrazy twarzy lub osoby do nazwy w przypadku, gdy jest to nazwa.

Praktyczne rozwiązanie takiego systemu zostało zaprezentowane w pracy [12]. Schemat systemu przedstawia rys. 1.21:



Rys. 1.21. Struktura systemu „Name-It” (Źródło: [12])

System składa się z trzech podstawowych bloków: rejestracji, identyfikacji i detektora zmiany scen. Moduł rejestracji przeznaczony jest do budowy bazy danych. Moduł identyfikacji odpowiedzialny jest za detekcję i lokalizację wszystkich twarzy w każdej wydzielonej scenie i w każdym kadrze oraz rozpoznanie osoby (osób) znajdujących się w kadrze. W przypadku gdy w bieżącym kadrze znajduje się więcej niż jedna twarz, dodatkowo wykonywane jest zadanie śledzenia twarzy (ang. *face tracking*). Istotną rolę odgrywa detektor zmiany sceny, ponieważ zmiana sceny jest sygnałem zakończenia lub początku rozwiązania zadania „Name-It”. W celu realizacji w/w zadań system wykorzystuje metody:

- do określania momentu zmiany sceny – metody opierające się porównywaniu histogramów jasności dwóch sąsiednich kadrów oraz analizę zmian gradientów średniej wartości i dyspersji komponentu H obrazu kolorowego przedstawionego w formacie HSV,
- do segmentacji kadrów i lokalizacji obszarów – kandydatów na twarz człowieka - metodę „skin-color”; w celu ich weryfikacji poszukuje się geometrycznych cech twarzy
- do identyfikacji – metody analizy komponentów głównych (PCA, PCA + LDA).

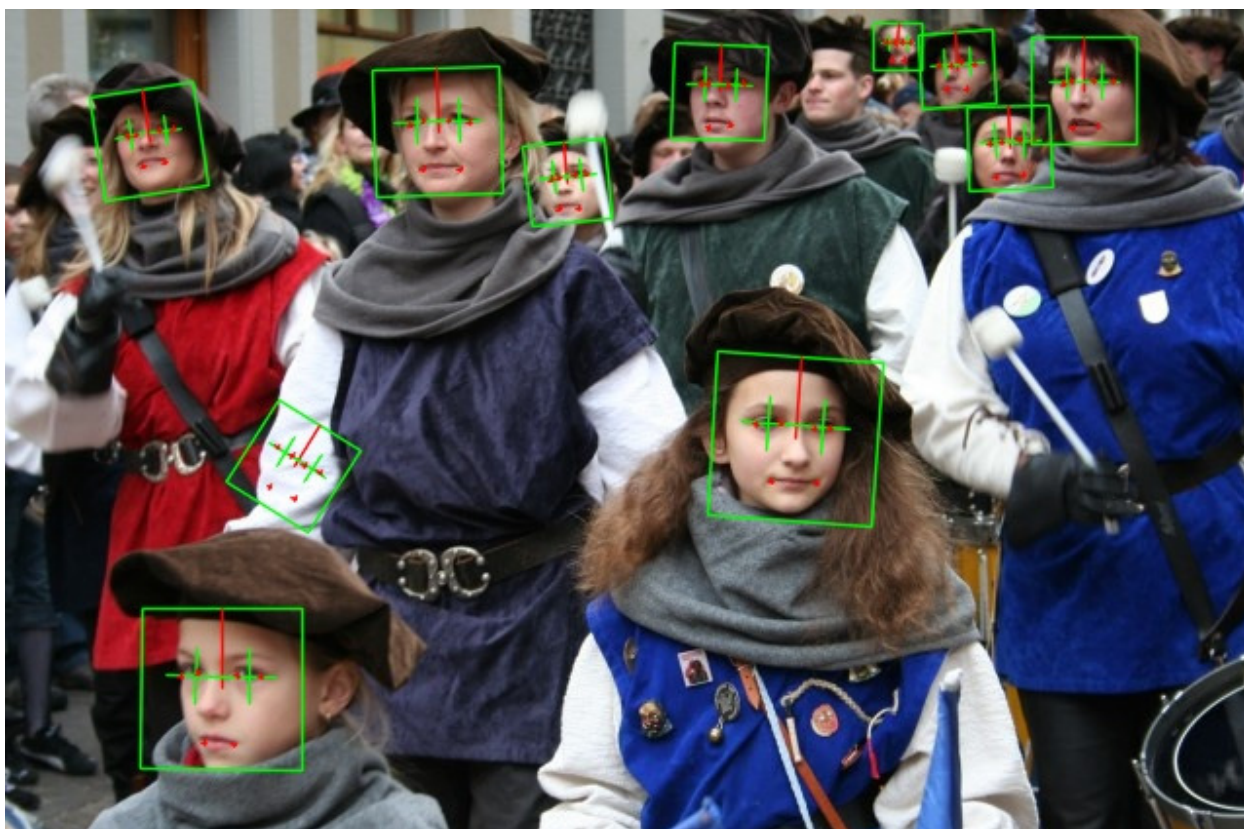
Skuteczność działania w dużej mierze uzależniona jest od efektów i jakości wyodrębnionych obszarów twarzy.

1.5. Przykłady istniejących aplikacji i systemów

Przedstawione wcześniej metody znajdują zastosowanie w wielu istniejących na rynku aplikacjach. Poniżej przedstawiono krótki przegląd systemów wykorzystujących zagadnienia biometriki. Z racji bardzo szerokiego zakresu w jakim działają te systemy, skupiono się na opisanu tych, które wykorzystują wykrywanie i rozpoznawanie twarzy, pomijając te oparte na detekcji linii papilarnych, tęczy oka czy też kształtu i ciepła dłoni.

BetaFace

System „BetaFace” [WWW-3] to produkt firmy BetaFace z Niemiec. System ten może być wykorzystywany zarówno jako aplikacja na stronach internetowych, jak i jako samodzielna aplikacja.

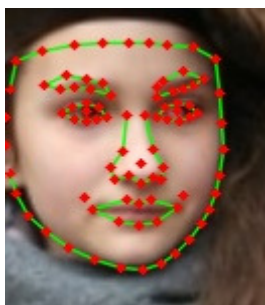


Rys. 1.22. Wykrywanie twarzy w programie BetaFace. Źródło: [WWW-3]

Algorytm „BetaFace” przeszukuje wczytany obraz (w wersji Premium także sekwencję wideo) pod kątem występowania ludzkiej twarzy. Po zlokalizowaniu twarzy wyznaczane są na niej punkty opisujące położenie charakterystycznych dla konkretnej osoby

części twarzy. W wersji podstawowej (rys.1.22) jest to 8 miejsc (oczy, kąciaki oczu oraz kąciaki ust).

Wersja „Professional” pozwala na wyznaczenie 84 punktów – opisują one kontury oczu, kontury brwi, kontur ust, kontur nosa, linię podbródka, skroni, czoła a także tęczęwkę oraz czubek nosa (rys. 1.23).



Rys. 1.23. 84 punkty opisujące twarz w systemie BetaFace. Źródło: [WWW-3]

Na stronie producenta [WWW-2] dostępna jest wersja demo „on-line”. Po wczytaniu zdjęcia następuje rozpoznanie i lokalizacja wszystkich twarzy (spełniających określone warunki – minimalną wielkość oraz ewentualny obrót wokół osi), następnie twarze te są porównywane z bazą i w wyniku otrzymuje się 5 najbardziej podobnych twarzy z bazy (rys.1.24).



Rys. 1.24. Porównanie twarzy autorki z twarzami w systemie BetaFace

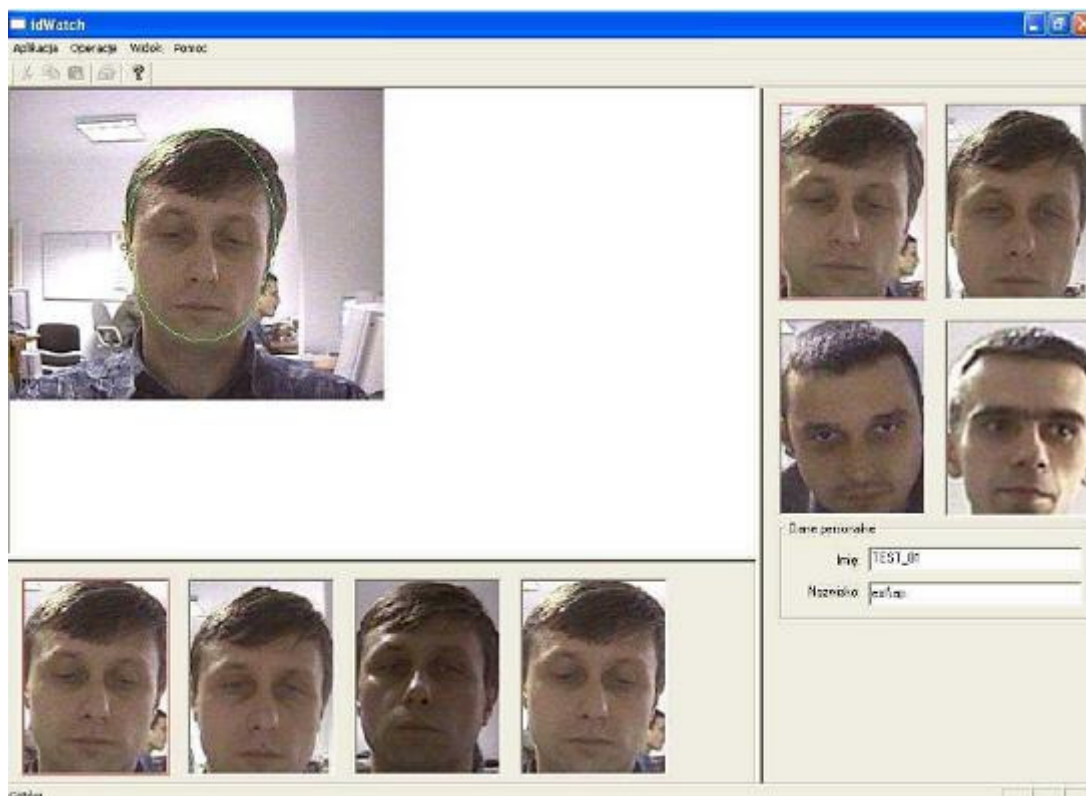
Oczywiście zdarzają się i błędy – zarówno fałszywego rozpoznania obiektu nie będącego twarzą jak i nierozpoznanie twarzy, która spełnia zadane warunki.

M3S - System Rozpoznawania Twarzy Watch ID

Jest to produkt polskiej firmy Polixel. W programie tym zastosowano algorytm analizy składowych głównych (PCA) oraz liniową analizę dyskryminacyjną Fishera (LDA).

System ten [WWW-11] wykorzystywany jest w powiązaniu z systemami monitoringu, dzięki czemu pracuje nie na pojedynczym obrazie twarzy, lecz sekwencji kilku zdjęć (rys. 1.25). Dzięki takiemu podejściu lokalizacja twarzy w systemie jest ułatwiona, gdyż porównanie kilku obrazów pozwala wyodrębnić obszar będący twarzą. Tak znaleziony obraz twarzy jest następnie poddawany normalizacji poprzez skalowanie i ewentualną rotację, aby deskryptor naliczał dane w jednakowych warunkach.

Dane takie mogą być przechowywane w bazie danych lub też np. zapisywane na chipach w kartach dostępu pracowników do wydzielonej sekcji w firmie.



Rys. 1.25. System IDWatch – okno rozpoznawania Źródło: [WWW-11]

FaceIt® Argus

Firma Identix połączona obecnie z firmą Viisage Technology jako L-1 Identity Solutions oferuje wiele rozwiązań dla firm opartych na biometryce. Są to zarówno systemy ograniczonego dostępu oparte na rozpoznawaniu i identyfikacji na podstawie odcisków palców (*FacePASS™*), jak i systemy rozpoznawania osób na podstawie odcisków palców oraz twarzy (*ABIS® System*).

Systemem opierającym się na rozpoznawaniu twarzy jest *FaceIt® Argus* [WWW-8]. System ten wyszukuje twarze w strumieniu wideo z kamer o wysokiej rozdzielczości, następnie przeszukuje zdefiniowane wcześniej listy i generuje alarm, jeśli nastąpi dokładne dopasowanie (rys. 1.26).



Rys. 1.26. Rozpoznanie twarzy w systemie FaceIt® Argus. Źródło [WWW-8]

System opiera się na połączonych technologiach rozpoznawania geometrii twarzy oraz tekstury skóry. Wykorzystywany może być szczególnie tam, gdzie należy rozpoznawać konkretne osoby (z wcześniej określonej listy) wśród tłumu innych.

Visitor Identification

System ten [WWW-13] powstał na Wydziale Informatyki Politechniki Szczecińskiej. Działa na zasadzie rozpoznawania osób na zdjęciach pochodzących z kamery cyfrowej. Może być zastosowany na przykład do identyfikacji osób czekających na wpuszczenie do posesji czy biura, weryfikacji osób wizytujących portal w Internecie, rozpoznawania częstych gości w sklepie itp.

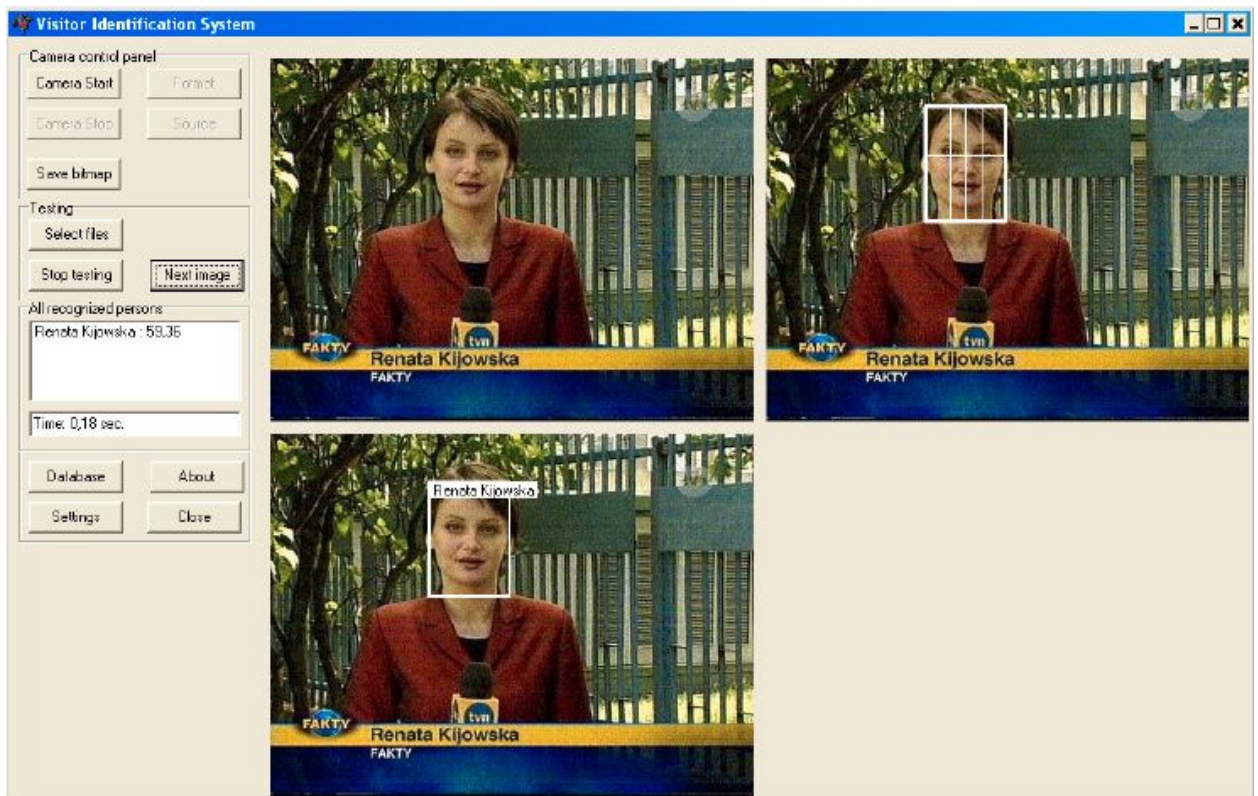
Wejściem może być plik *jpg* bądź *bmp* lub też obraz z kamery. Aplikacja wyszukuje w obrazie twarze ludzkie, następnie wybiera najważniejszą i dającą najwięcej informacji część twarzy (np. od linii brwi do podbródka), normalizuje do wymiarów obrazów w bazie danych a następnie porównuje z opisanymi w bazie osobami. Na obrazie wyjściowym wokół twarzy osoby widnieje ramka z przypisanym nazwiskiem osoby rozpoznanej (rys. 1.27).

W systemie tym do detekcji twarzy na obrazie wykorzystywane są metody oparte na rozpoznawaniu koloru skóry (ang. *skin-color detection*) w przestrzeni *RGB* oraz *YCbCr*, następnie weryfikacja następuje na podstawie cech ludzkiej twarzy (wyszukanie oczu). Do ekstrakcji cech twarzy wykorzystano transformatę kosinusową a porównanie wykonywane jest w metryce *L2*.



Rys. 1.27. Rozpoznana twarz w systemie „Visitor Identification”. Źródło [WWW-13]

Poniżej przedstawiono zrzut ekranu pokazujący działanie systemu (rys. 1.28). Po lewej stronie widnieje obraz wejściowy, z prawej wynik detekcji twarzy – ramka wokół twarzy oraz linie wskazujące oś symetrii oraz położenie oczu. Na dole znajduje się obraz z wynikiem rozpoznania – twarz otoczona ramką wraz z nazwiskiem osoby rozpoznanej.



Rys. 1.28. System „Visitor Identification”. Źródło [WWW-13]

SNAPPY Face Recogniser

Kontrolowany dostęp może dotyczyć nie tylko miejsc szczególnie ważnych z punktu bezpieczeństwa firmy, jak serwerownie czy centrale telekomunikacyjno-informatyczne. Może

on także znaleźć zastosowanie w domu, przy własnym komputerze. Często pracuje się nad ważnymi danymi, do których dostępu nie powinny mieć niepowołane osoby. Sposobem zabezpieczenia się przed takimi sytuacjami jest zastosowanie logowania do systemu na podstawie własnej twarzy. Jednym z takich systemów jest SNAPPY Face Recogniser [WWW-12] zastosowany w kamerach SNAPPY (rys. 1.29). Kamera ta może służyć jako zwykła kamera internetowa, ale również może zostać wykorzystana do ochrony komputera.



Rys. 1.29. Kamera SNAPPY. Źródło: [WWW-12]

Dane o upoważnionych użytkownikach są zapisywane w bazie. W momencie logowania system porównuje twarz widzianą w kamerze z bazą danych na podstawie analizy geometrii twarzy i zezwala na zalogowanie do systemu, jeśli nastąpi poprawna identyfikacja. System udostępnia również funkcję włączania specjalnego wygaszacza ekranu w momencie, gdy odchodzi od niego zalogowany użytkownik. Po powrocie użytkownika przed ekran monitora następuje kolejna weryfikacja, wygaszacz zostaje wyłączony i możliwa jest dalsza praca.

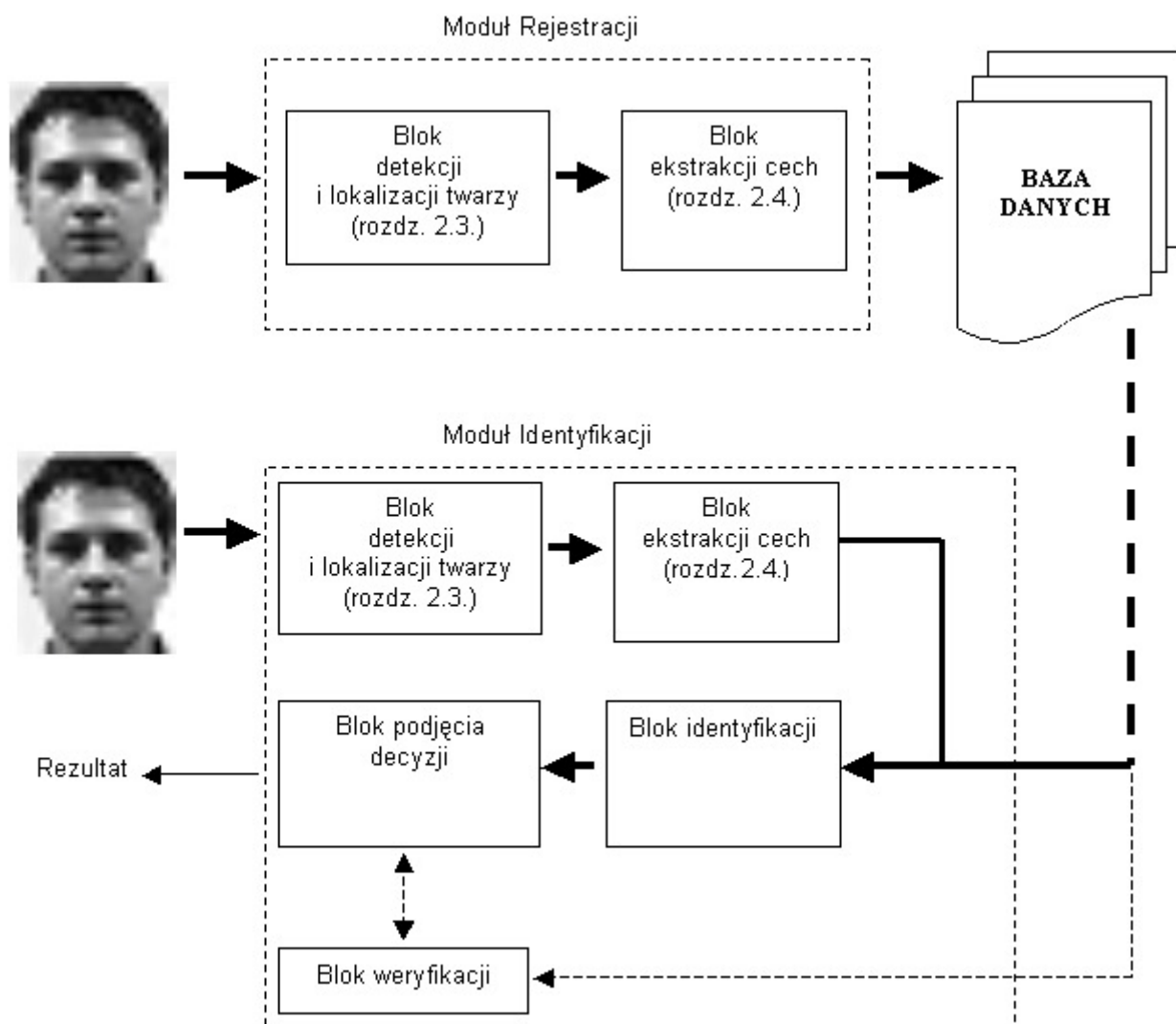
2. Metodologia

2.1. Ogólna koncepcja systemu

System realizuje zadanie identyfikacji osób w trybie „off-line”. Funkcjonalnie jego struktura oparta jest na ogólnym modelu rozpoznawania. Składa się z kilku podstawowych bloków (rys. 2.1):

- bloku detekcji i lokalizacji twarzy,
- bloku ekstrakcji cech,
- bloku identyfikacji,
- bloku podjęcia decyzji,
- bloku weryfikacji,
- bazy danych.

System może pracować w dwóch trybach: rejestracji oraz identyfikacji realizowanych w odpowiednich modułach. Podział jest umowny, gdyż oba z nich wykorzystują te same bloki funkcjonalne. Moduł rejestracji służy do budowy bazy danych. W jego skład wchodzi: blok detekcji i lokalizacji twarzy oraz blok ekstrakcji cech. Detekcja i lokalizacja twarzy opiera się na wykorzystaniu kombinacji kilku metod: metodzie detekcji koloru skóry (ang. „*skin-color detection*”) wykorzystywanej do wstępnej segmentacji obrazu wejściowego, metodzie konturowej (ang. „*active contour*”) służącej do wyznaczenia owalu twarzy oraz metodzie deformowalnych wzorców (ang. „*deformable template*”) wykorzystywanej do weryfikacji wyznaczonego obszaru. Blok ekstrakcji cech lokalizuje na wyodrębnionym obszarze twarzy: oczy, nos, usta, krawędzie twarzy (owal). Na podstawie tych informacji tworzony jest wektor cech opisujący badaną twarz ($N = 6$) dodawany do odpowiedniej klasy wzorców w bazie danych.



Rys. 2.1. Schemat funkcjonalny systemu

W module identyfikacji odbywa się właściwy proces rozpoznawania. Do identyfikacji wykorzystywane są dwa algorytmy: podstawowy opierający się na porównaniu wybranych cech anatomicznych badanej twarzy z przechowywanymi w bazie wzorcami oraz dodatkowy – służący do weryfikacji rezultatu – wykorzystujący metodę korelacji fazowej. Do określenia miary podobieństwa badanego obrazu do danej klasy wykorzystywana jest metryka euklidesowa. Algorytm dodatkowy wykorzystujący metodę korelacji fazowej wykorzystuje najbardziej informacyjną centralną część twarzy, uzyskaną w efekcie procesów obróbki wstępnej obrazów wejściowych.

Do budowy systemu wykorzystano ogólnie dostępne bazy obrazów kolorowych (*RGB*) zawierających twarze osób oraz materiały własne. Wykorzystywana przez system baza danych posiada strukturę hierarchiczną: obrazy twarzy należące do jednej osoby zgrupowane są w klasy; klasy w zależności od sposobu reprezentacji danych przyporządkowywane są do

odpowiednich grup. Aby zapewnić jak najlepsze wyniki rozpoznawania, przyjmuje się, że wejściowe obrazy twarzy spełniają następujące kryteria:

- kąt obrotu względem osi OY zawiera się w granicach +/- 15 stopni,
- kąt obrotu względem osi OX zawiera się w granicach +/- 30 stopni,
- posiadają widoczne, charakterystyczne cechy twarzy (oczy, usta, nos, ował).

Obrazy nie spełniające w/w kryteriów są odrzucane jako nieprzydatne w procesie identyfikacji.

2.2. Budowa modelu „skin-color detection”

Metoda detekcji koloru skóry (ang. „*skin-color detection*”) wykorzystywana jest do wstępnej lokalizacji obszarów-kandydatów, mogących przedstawiać twarz człowieka i opiera się na obserwacji, iż rozkład koloru ludzkiej skóry, niezależnie od rasy człowieka, skupiony jest na niewielkim obszarze w wybranej przestrzeni barw. Rozkład ten może zostać zamodelowany różnymi metodami. Najbardziej popularne z nich to:

- bezwzględne określenie granic obszaru w wybranej przestrzeni barw – zaletą metody jest jej prostota i intuicyjność reguł klasyfikacji. Jednak aby osiągnąć zadowalające wyniki działania, konieczne jest przeprowadzenie dużej liczby doświadczeń w celu budowy reguł klasyfikacji. Ponadto metoda zawodzi w przypadku, gdy obrazy wejściowe były pozyskiwane w innych niż w trakcie doświadczeń warunkach oświetleniowych;
- nieparametryczne metody modelowania koloru skóry – opierają się najczęściej na analizie histogramowej obszaru skóry. Opis rozkładu koloru skóry przedstawiany jest w postaci tablic lookup (LUT) przechowujących znormalizowane wartości prawdopodobieństw (ang. *Skin Probability Map*) opisujących w jakim stopniu dany kolor (lub przedział kolorów) reprezentuje kolor skóry. Sposób wyliczania wartości tych prawdopodobieństw uzależniony jest od przyjętej metody klasyfikacji. Najbardziej popularną jest ta wykorzystująca klasyfikator *Bayesa*. Skuteczność tych metod uzależniona jest od ilości danych treningowych dostępnych w trakcie budowy modelu. Ponadto, do przechowywania uzyskanych danych wymagają one dodatkowej ilości pamięci. Innym przykładem wykorzystania metod nieparametrycznych jest zastosowanie

sztucznych sieci neuronowych. W dziedzinie tej szczególne znaczenie mają samoorganizujące się mapy (ang. *Self Organizing Maps*, SOM, sieci Kohonena). Z siecią neuronów stowarzyszone są współrzędne na prostej, płaszczyźnie lub dowolnej n -wymiarowej przestrzeni. Uczenie tego rodzaju sieci polega na zmianach współrzędnych neuronów, tak aby dążyły one do wzorca zgodnego ze strukturą analizowanych danych. Sieć zatem „rozpiną się” wokół zbioru danych, dopasowując do nich swoją strukturę. Przeprowadzone doświadczenia wykazały, że skuteczność detektora opartego na sieci SOM jest porównywalna do wyników osiągniętych z wykorzystaniem modelu złożonego Gaussa,

- parametryczne metody modelowania koloru skóry – rozkład koloru ludzkiej skóry przedstawiany jest za pomocą sparametryzowanych modeli matematycznych. Najszerzej wykorzystywanym jest model opierający się na n -wymiarowym rozkładzie normalnym Gaussa (prostym lub złożonym). Główną zaletą tych modeli jest stosunkowo wysoka skuteczność, uniwersalność i generalizacja problemu.
- modele dynamiczne – zaprojektowane specjalnie do lokalizacji obszarów skóry w zadaniach śledzenia twarzy (ang. *face-tracking*).

Projektowany system wykorzystuje model oparty na gaussowskiej reprezentacji rozkładu kolorów ludzkiej skóry ze względu na swą uniwersalność, łatwość modyfikacji i dużą skuteczność.

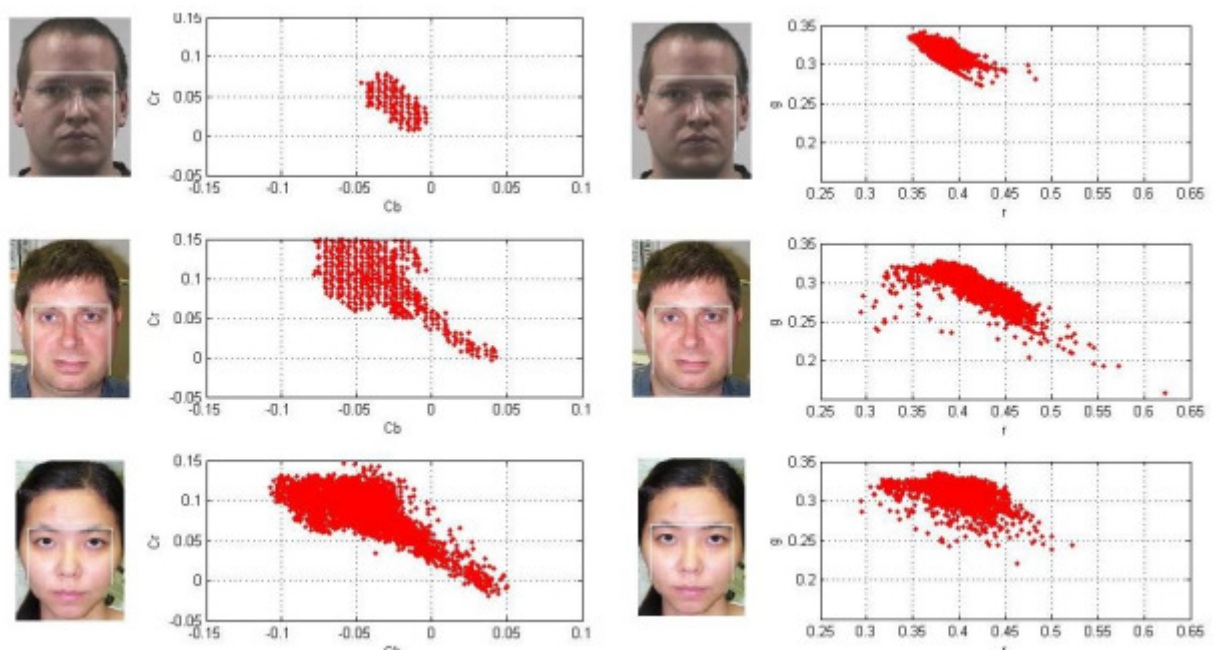
2.2.1. Wybór właściwego modelu przestrzeni barw

Model opisujący rozkład kolorów ludzkiej skóry powinien być zaadaptowany do poprawnego działania niezależnie od rasy rozpoznawanej osoby i panujących warunków zewnętrznych (np. oświetlenia). Dlatego właściwy wybór stosowanego modelu przestrzeni barw ma kluczowe znaczenie dla skuteczności jego działania.

Najbardziej popularnym jest model RGB. Jest to model wynikający z właściwości odbiorczych ludzkiego oka, w którym wrażenie widzenia dowolnej barwy można wywołać przez zmieszanie w ustalonych proporcjach trzech wiązek światła o barwie czerwonej, zielonej i niebieskiej, czyli światła o odpowiedniej częstotliwości fali elektromagnetycznej.

Jego popularność wynika z faktu, iż jako jedyny posiada on bezpośrednie odwzorowanie w realizacji sprzętowej (w postaci kineskopów pokrytych luminoforami, które

wzbudzone w odpowiednich miejscach trzema strumieniami elektronów świecą światłem w trzech barwach podstawowych). Stąd też wynika zastosowanie tego modelu w bezpośredniej prezentacji obrazu na monitorze. Najczęściej stosowany jest 24-bitowy zapis kolorów, w którym każda z barw jest zapisana przy pomocy składowych, które przyjmują wartość z zakresu 0-255. Reprezentacja RGB przestrzeni barw nie nadaje się jednak zbyt do opisu koloru ludzkiej skóry. W przestrzeni *RGB* pojedynczy komponent opisujący kolor (*r*, *g*, *b*) reprezentuje nie tylko kolor, ale także luminancję. Luminancja może powodować, że ta sama twarz postrzegana w różnych warunkach oświetlenia zewnętrznego będzie inaczej wyglądać i jej poprawna detekcja może być znacznie utrudniona. Ponadto pomiędzy poszczególnymi składowymi istnieje wysoka korelacja utrudniająca poprawną klasyfikację. Wszystko to powoduje, że model *RGB* nie jest najlepszym wyborem, jeśli chodzi o zastosowania w biometrycznych systemach rozpoznawania twarzy. Wpływ luminancji można wyeliminować wybierając chromatyczną przestrzeń kolorów. Największą popularność w zadaniach lokalizacji twarzy zdobyły przestrzenie: *RG*, *HSV*, *YCbCr* oraz *III2I3*. Ich wspólną cechą jest rozdzielenie luminancji i chrominancji.



Rys. 2.2. Rozkład wartości składowych *Cb*, *Cr* (po lewej) oraz *R*, *G* (po prawej) dla obszarów różnych twarzy

Model kolorów *RG* to przykład dwuwymiarowej chromatycznej przestrzeni barw, w której informacja na temat intensywności danego koloru nie jest przechowywana. W odróżnieniu od modelu *RGB*, w którym każdy piksel identyfikowany jest przez wartości składowych *R*, *G*, *B* w przestrzeni *RG* kolor jest reprezentowany poprzez proporcjonalny

udział każdej ze składowych w kolorze. Suma tych udziałów musi zawsze wynosić 1, więc brakującą w modelu RG znormalizowaną składową B można w razie potrzeby w łatwy sposób wyliczyć.

Mimo, iż przestrzeń RG zawiera mniej informacji niż modele RGB lub HSV, z racji opisanej wyżej własności znajduje szerokie zastosowanie w komputerowych systemach wizyjnych. Brak informacji o intensywności kolorów rozwiązuje problem związany z niestałością zewnętrznych warunków, w jakich pozyskiwany jest obraz wejściowy i powoduje, że widoczny kolor pozostaje stały. Należy tutaj zaznaczyć, że w przypadku kiedy różne części obrazu są oświetlone przez różne źródła światła, problem związany z niestałością warunków zewnętrznych nadal pozostaje.

Składowe w modelu RG (składowe chromatyczne) definiuje się na podstawie modelu RGB poprzez następujący proces normalizacji:

$$r = R/(R + G + B) \quad (11)$$

$$g = G/(R + G + B) \quad (12)$$

Znormalizowana składowa b jest redundantna, ponieważ $r + g + b = 1$.

Konwersja odwrotna jest niemożliwa, ponieważ informacja o intensywności jest w trakcie procesu tracona.



Rys. 2.3. Składowe kolorowego obrazu w przestrzeni RG

Bardzo popularnym w zastosowaniach rozpoznawania twarzy jest model *HSV* (*HSB*). Definiuje on przestrzeń kolorów składającą się z trzech składowych: *H* – opisującą kolor dominujący, *S* – nasycenie koloru i *V* – jasność koloru. Jego geometryczną interpretacją jest stożek, którego podstawą jest koło barw.

Przejście z modelu *RGB* do *HSV* odbywa się z wykorzystaniem nieliniowej transformacji zgodnie z równaniami:

$$H = \begin{cases} \text{nieustalone,} & \text{jezeli MAX} = \text{MIN} \\ 60^\circ \times \frac{G - B}{\text{MAX} - \text{MIN}} + 0^\circ, & \text{jezeli MAX} = R \text{ i } G \geq B \\ 60^\circ \times \frac{G - B}{\text{MAX} - \text{MIN}} + 360^\circ, & \text{jezeli MAX} = R \text{ i } G < B \\ 60^\circ \times \frac{B - R}{\text{MAX} - \text{MIN}} + 120^\circ, & \text{jezeli MAX} = G \\ 60^\circ \times \frac{R - G}{\text{MAX} - \text{MIN}} + 240^\circ, & \text{jezeli MAX} = B \end{cases}$$

$$S = \begin{cases} 0, & \text{jezeli MAX} = 0 \\ 1 - \frac{\text{MIN}}{\text{MAX}}, & \text{w przeciwn. przyp.} \end{cases} \quad (13)$$

$$V = \text{MAX}$$

przy czym MAX równa się $\max(R, G, B)$ a MIN równa się $\min(R, G, B)$.

Z punktu widzenia zastosowań w zadaniach segmentacji obszaru skóry, szczególnie składowa *Hue* charakteryzuje się interesującymi własnościami. Okazuje się, że jest ona niezależna od odbłyśków powodowanych białym źródłem światła, ponadto w przypadku matowych powierzchni jest również niewrażliwa na światło otaczające niezależnie od orientacji danej powierzchni względem jego źródła. Przykładowe wartości składowych używane do detekcji twarzy to:

$$\begin{aligned}
(H): & 0 \leq H \leq 25 \text{ lub } 335 \leq H \leq 360 \\
(S): & 0,4 \leq S \leq 0,6 \\
(V): & V \geq 0,4
\end{aligned} \quad (14)$$



Rys. 2.4. Składowe kolorowego obrazu w przestrzeni HSV

Model $YCbCr$ to jeden z modeli szeroko wykorzystywanych w systemach wideo, telewizji oraz kompresji obrazów. Składowa Y reprezentuje luminancję, Cb i Cr to składowe (czerwona i niebieska) chrominancji. Poszczególne składowe obrazu można otrzymać w następujący sposób (BT. 601):

$$Y = 0,2989R + 0,5866G + 0,1145B$$

$$Cb = (B - Y)$$

$$Cr = (R - Y)$$
(15)

Wartości składowych Cb i Cr używane do detekcji obszaru twarzy to:

$$(Cb): 199 < Cb + 0,6Cr < 215$$
(16)

$$(Cr): 138 < Cr < 178 \quad \text{gdzie } Cb, Cr = [0, 255].$$

Obszar twarzy najlepiej wydzielany jest w składowej Cb (rys. 2.5).



Rys. 2.5. Składowe kolorowego obrazu w przestrzeni YCbCr

Prostota transformacji i rozdzielenie komponentów luminancji i chrominancji powoduje, że również ten model znajduje szerokie zastosowanie w zadaniach detekcji twarzy.

Interesujące właściwości z punktu widzenia modelowania koloru skóry, wykazuje również przestrzeń $I1I2I3$ (rys. 2.6).

Poszczególne składowe można wyliczać w następujący sposób:

$$I1 = (R + G + B) / 3$$

$$I2 = R - B$$

$$I3 = (2G - R - B) / 2$$
(17)

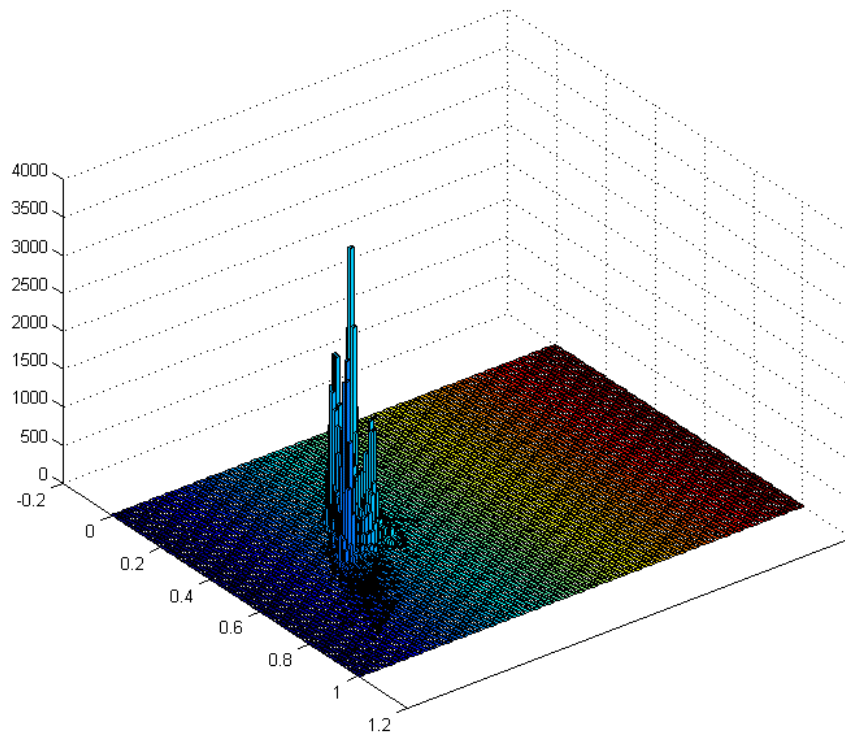
Szczególnie użyteczna jest tutaj składowa $I2$. Jej dalsza modyfikacja może okazać się przydatna w celu minimalizacji wpływu niepożądanych obiektów na twarzy (np. okularów).



Rys. 2.6. Składowe kolorowego obrazu w przestrzeni I1I2I3

2.2.2. Reprezentacja Gaussowska rozkładu kolorów ludzkiej twarzy. Model *EBM*.

Może się wydawać, że kolor skóry różnych ludzi bardzo się od siebie różni. Okazuje się jednak, że przy spełnieniu pewnych warunków, ich dystrybucja w chromatycznej przestrzeni barw jest skupiona na niewielkim obszarze (rys. 2.7) i może być reprezentowana przez klasyczny model rozkładu Gaussowskiego.



Rys. 2.7. Rozkład kolorów ludzkiej twarzy w chromatycznej przestrzeni barw

Model estymowany jest dwoma parametrami: wektorem wartości oczekiwanych m i macierzą kowariancji C :

$$\begin{aligned} m &= E\{x\} \\ C &= E\{(x - m)(x - m)^T\} \end{aligned} \quad (18)$$

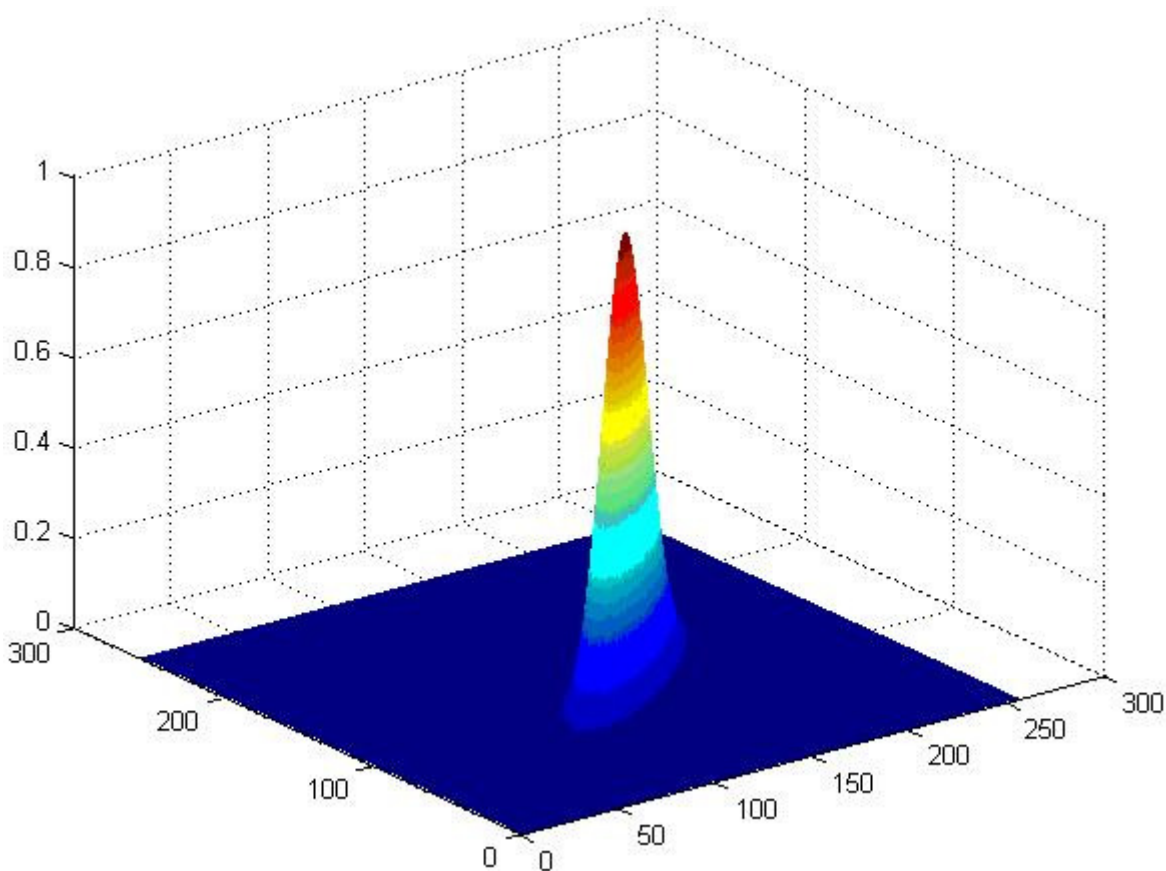
gdzie $x = (x_1, x_2, \dots, x_n)^T$ n – wymiar przestrzeni barw

Przyjmując, że rozkład kolorów ludzkiej skóry opisany jest tym modelem, prawdopodobieństwo tego, że dany wektor wejściowy chrominancji x reprezentuje obszar ludzkiej skóry wyrażone jest zależnością:

$$p(x) = \exp[-0,5(x - m)^T C^{-1}(x - m)] \quad (19)$$

gdzie: $x = (x_1, x_2, \dots, x_n)^T$, m – wektor wartości oczekiwanych, C – macierz kowariancji.

Wektor wartości oczekiwanej m i macierz kowariancji C rozkładu estymowane są na podstawie próbek serii treningowej.



Rys. 2.8. Aproksymacja rozkładu koloru ludzkiej skóry modelem Gaussa (przestrzeń RG)

W trakcie budowy modelu bazującego na reprezentacji Gaussowskiej należy wziąć pod uwagę następujące aspekty:

- charakter serii zdjęć wejściowych,
- reprezentację piksela,
- wybór typu modelu.

Charakter serii zdjęć wejściowych związany jest z warunkami, w jakich pozyskiwane były obrazy wejściowe (optymalną sytuacją jest, aby w ramach jednej serii każde ze zdjęć było pozyskiwane w podobnych warunkach) czy rasą ludzi, którą przedstawiają. Reprezentacja koloru piksela związana jest bezpośrednio z wybraną przestrzenią barw. Ma to wpływ na postać modelu Gaussa np. w przypadku przestrzeni HSV wektor wartości oczekiwanych i macierz kowariancji mają odpowiednio wymiary 3×1 i 3×3 . W przypadku modelu Gaussowskiego można mówić o dwóch jego typach: prostym i mieszanym (złożonym). Związane jest to ze sposobem, w jaki estymowane są jego parametry oraz jego formalną reprezentacją. Model prosty reprezentowany jest poprzez równanie (19). Model

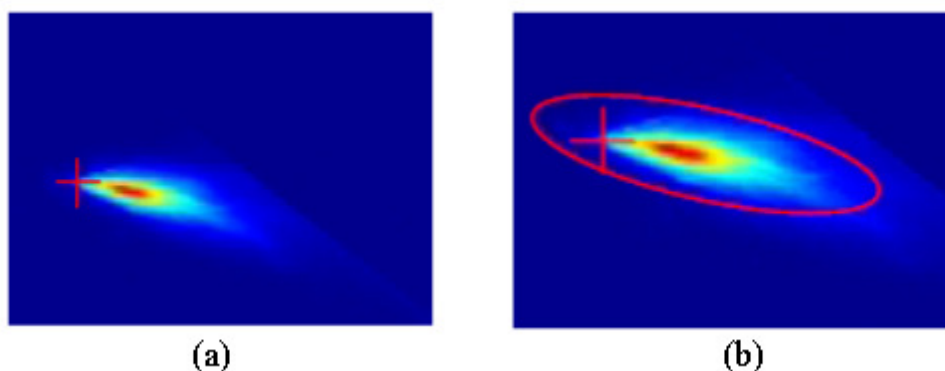
mieszany to kombinacja liniowa modeli prostych z przyporządkowanymi do nich odpowiednimi „wagami”.

$$p(x) = w_1 N_1(m, C) + w_2 N_2(m, C) + \dots + w_n N_n(m, C) \quad (20)$$

gdzie N_1, N_2, N_n to modele proste.

Każdy taki prosty model odpowiada innej pod względem charakteru serii danych wejściowych (patrz wyżej). Parametry modelu mogą być estymowane w sposób analityczny (z wykorzystaniem alg. EM) lub na drodze eksperymentalnej.

Model gaussowski prosty dobrze sprawdza się w przypadkach kontrolowanego procesu akwizycji obrazów twarzy (właściwe oświetlenie, tło). W ogólnym przypadku rozkład koloru ludzkiej skóry odbiega nieco od tego przedstawionego na rysunku 2.7. Na wykresie rozkładu chrominancji ludzkiej skóry pojawiają się zakłócenia związane m.in. z niestabilnością koloru ciemniejszych fragmentów ludzkiej skóry. Badania przeprowadzone przez Lee i Yoo [14] na losowo wybranych, odpowiednio dużych seriach obrazów, porównujące rozkłady: koloru ludzkiej skóry i innych obiektów, wykazały ponadto, że większość chrominancji obiektów nie reprezentujących ludzkiej skóry skoncentrowana jest w jednym punkcie, znajdującym się na granicy obszaru rozkładu ludzkiej skóry.



Rys. 2.9. Histogram rozkładu koloru ludzkiej skóry z zaznaczonym punktem skupienia obszarów nie reprezentujących ludzkiej skóry (a), (b) aproksymacja rozkładu modelem Gaussa (Źródło: [14])

Powoduje to, że rzeczywisty rozkład chrominancji ludzkiej skóry przestaje być symetryczny. Próba zamodelowania go z wykorzystaniem symetrycznego modelu Gaussowskiego może prowadzić do wzrostu ilości niepoprawnych decyzji. Pewne rozwiązanie stanowi wykorzystanie gaussowskiego modelu mieszanego. Jest to jednak okupione dużymi nakładami obliczeniowymi niezbędnymi do jego budowy (alg. EM) i stosunkowo wolnym działaniem, co uniemożliwia jego wykorzystanie np. w systemach „on-line”.

Fakt, iż kształt obszaru ludzkiej skóry można w chromatycznej przestrzeni barw aproksymować elipsą, został wykorzystany do stworzenia nowego modelu rozkładu kolorów

ludzkiej skóry, opartego także na rozkładzie gaussowskim, nazwanego przez autorów [14] „Elliptical Boundary Model” (*EBM*). Model ten pozwala na uwzględnienie omówionych wyżej ograniczeń, szybkość działania porównywalną do prostego modelu Gaussa i skuteczność modelu mieszanego. Model *EBM* $\Phi = (X; \psi, \Lambda)$ definiowany jest w następujący sposób:

Niech X_1, X_2, \dots, X_n oznaczają wektory chrominancji, poddane obróbce wstępnej zbioru uczącego, przy czym $X_1 \ll X_2, \dots, \ll X_n$, a $f(X_i) = f_i$ ($i=1, 2, \dots, n$) niech oznacza liczbę próbek o chrominancji równej X_i . Wtedy model $\Phi = (X; \psi, \Lambda)$ definiuje się jako

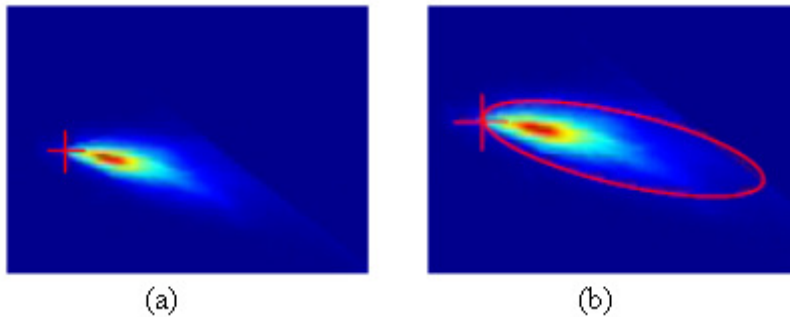
$$\Phi(X) = [X - \psi]^T \Lambda^{-1} [X - \psi] \quad (21)$$

gdzie ψ, Λ to parametry:

$$\psi = \frac{1}{n} \sum_{i=1}^n X_i \quad (22)$$

$$\Lambda = \frac{1}{N} \sum_{i=1}^n f_i (X_i - \mu)(X_i - \mu)^T$$

i $N = \sum_{i=1}^n f_i$ jest całkowitą liczbą próbek zbioru uczącego, a $\mu = \frac{1}{N} \sum_{i=1}^n f_i X_i$ - średnią z wektorów chrominancji.



Rys. 2.10. Histogram rozkładu koloru ludzkiej skóry z zaznaczonym punktem skupienia obszarów nie reprezentujących ludzkiej skóry (a), (b) aproksymacja rozkładu modelem *EBM* (Źródło: [14])

Piksel klasyfikowany jest jako reprezentujący obszar skóry w przypadku, gdy odpowiadający mu wektor chrominancji X spełnia nierówność $\Phi(X) < \theta$, gdzie θ jest ustalonym progiem. Im większa jego wartość tym rośnie liczba poprawnych detekcji, ale także zwiększa się ryzyko pojawienia się decyzji błędnych. Równość $\Phi(X) = \theta$ definiuje eliptyczny obszar oddzielający przestrzeń chrominancji skóry od pozostałej, którego centrum dane jest przez ψ a główne osie definiowane są poprzez Λ . Jak to przedstawiono na rysunku 2.10 model ten dobrze aproksymuje rozkład chrominancji ludzkiej skóry i jest niewrażliwy na zakłócenia związane z niestabilnością koloru ciemniejszych fragmentów ludzkiej skóry.

2.2.3. Estymacja parametrów modelu

Proces budowy modelu składa się z trzech etapów: przygotowania jak najbardziej reprezentatywnego zbioru próbek stanowiących elementy zbioru uczącego, ich obróbki i właściwej estymacji parametrów. Do budowy modelu wykorzystano dwie różne bazy obrazów przedstawiających twarze osób należących do różnych ras. Z każdego obrazu zostały pobrane kwadratowe próbki, będące niewielkim wycinkiem obszaru skóry człowieka. Wszystkie próbki zostały przekonwertowane z przestrzeni *RGB* do chromatycznej przestrzeni *YCbCr*. W celu zmniejszenia niekorzystnego wpływu zakłóceń, wszystkie próbki zostały przefiltrowane filtrem dolnoprzepustowym. Następnie na podstawie (22) wyliczono końcową postać modelu:

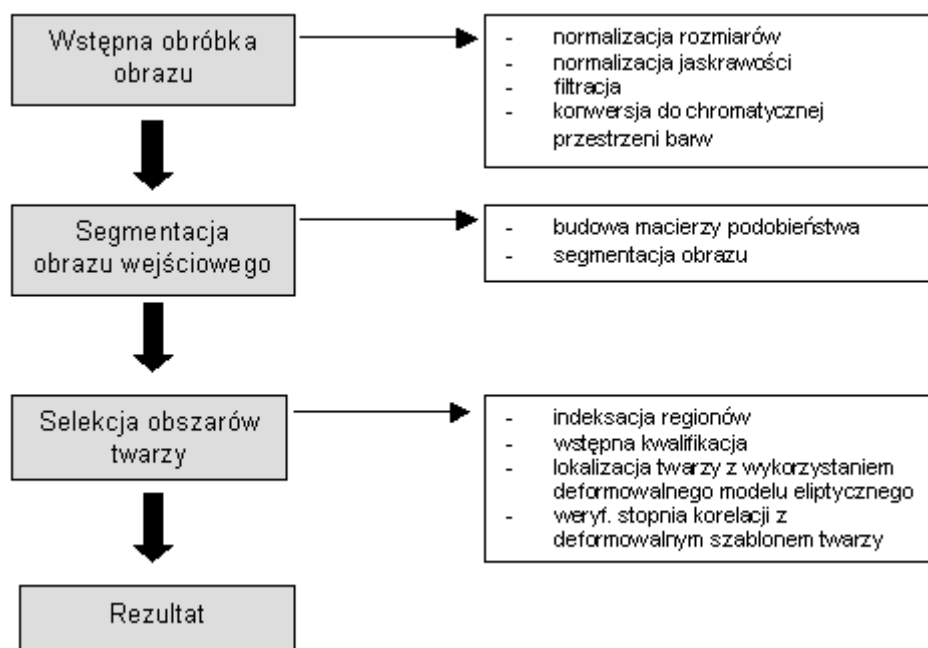
$$\Phi(X) = [X - \psi]^T \Lambda^{-1} [X - \psi] \quad (23)$$

gdzie:

$$\psi = \begin{bmatrix} -0,0481 \\ 0,0889 \end{bmatrix} \text{ i } \Lambda = \begin{bmatrix} 0,7939 & -0,4679 \\ -0,4679 & 0,6444 \end{bmatrix} \times 10^{-3} \quad (24)$$

2.3. Detekcja i lokalizacja twarzy

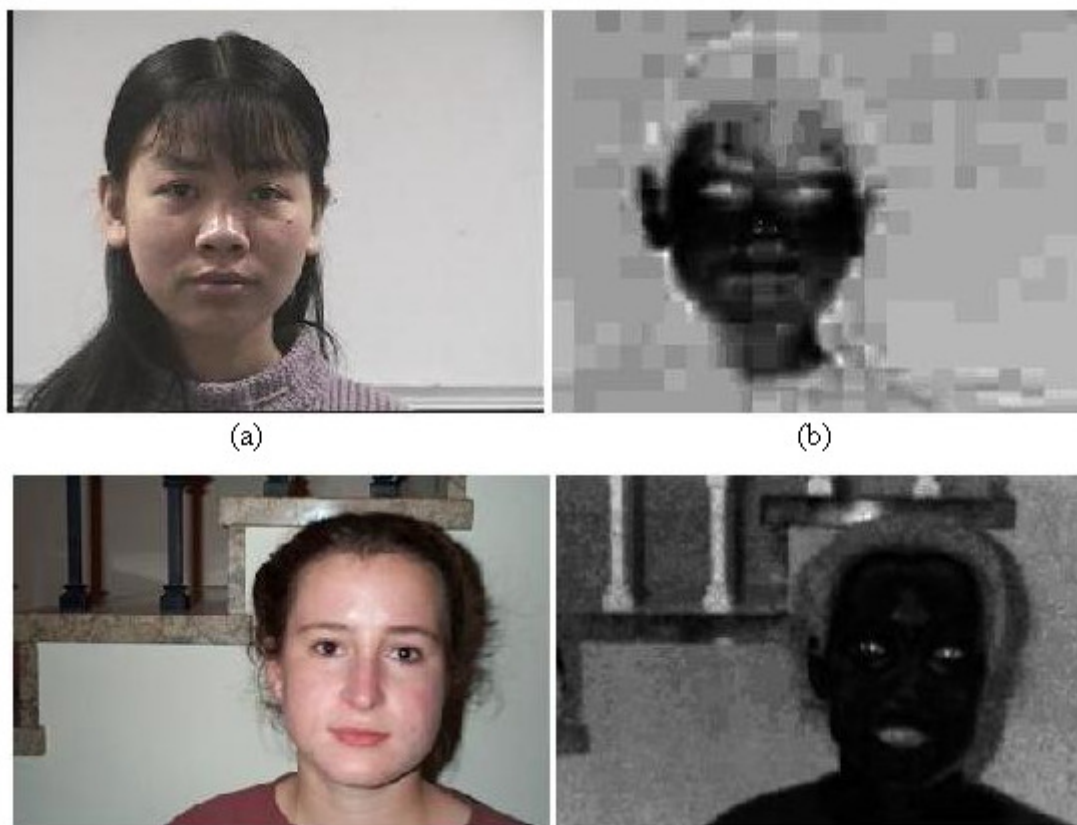
Etap lokalizacji (ogólnie – detekcji) twarzy na obrazie wejściowym ma kluczowe znaczenie dla dalszego procesu rozpoznawania twarzy. Prawidłowa lokalizacja twarzy przekłada się na wynik działania algorytmów ekstrakcji cech – w konsekwencji na wynik rozpoznawania. Rysunek 2.11. przedstawia schemat postępowania:



Rys. 2.11. Schemat postępowania algorytmu detekcji i lokalizacji twarzy

2.3.1. Segmentacja obrazu

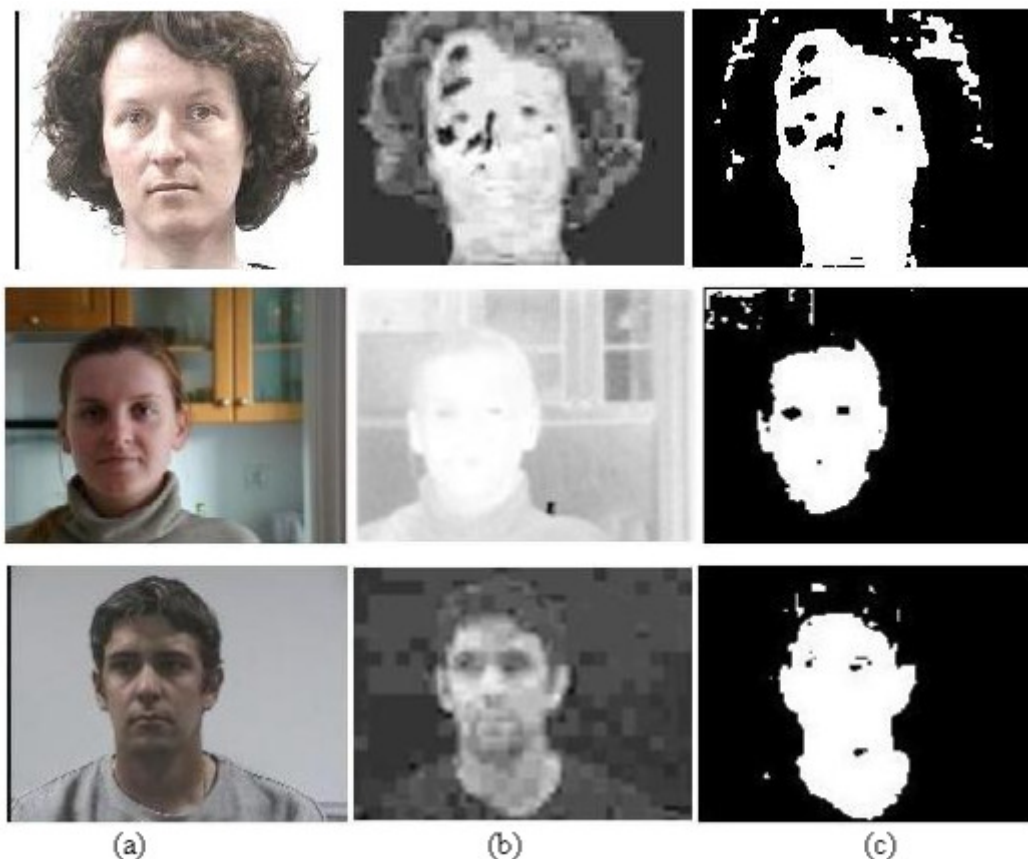
Pierwszym etapem w procesie detekcji twarzy jest analiza obrazu wejściowego pod kątem wstępnego wyodrębnienia obszarów – kandydatów mogących reprezentować twarz człowieka z wykorzystaniem metody detekcji koloru skóry (ang. „*skin-color detection*”). Obraz wejściowy transformowany jest do chromatycznej przestrzeni barw $YCbCr$. Następnie, zgodnie z (21) dla każdego piksela obliczany jest stopień podobieństwa reprezentowanej przez niego barwy do koloru skóry z wykorzystaniem zbudowanego modelu. W rezultacie tego otrzymuje się macierz o takich samych wymiarach co obraz wejściowy, której wartościami są obliczone prawdopodobieństwa. Wartości te można łatwo transformować do skali szarości tworząc obraz podobieństwa, będący wizualną reprezentacją macierzy.



Rys. 2.12. Obraz wejściowy (a) i obraz podobieństw (b). Obszary ciemniejsze reprezentują regiony skóry

Obszary obrazu, w których wartości te są najmniejsze (ciemniejsze od pozostałych) z większym prawdopodobieństwem reprezentują regiony skóry niż pozostałe. Podstawowym celem tej operacji jest wyeliminowanie z dalszego procesu lokalizacji tych obszarów, które nie reprezentują koloru skóry. Eliminacja odbywa się poprzez binaryzację obrazu podobieństw z ustalonym (na stałe lub liczonym adaptacyjnie) progiem. Jeżeli wartość podobieństwa jest mniejsza lub równa ustalonej wartości progowej przyjmuje się, że odpowiadający piksel reprezentuje kolor skóry, w przeciwnym wypadku przyjmuje się, że nie. Piksele odpowiadające kolorowi skóry przyjmują wartość „1”, pozostałe wartość „0”. Podstawowym problemem jest tutaj estymacja wartości progu binaryzacji. Możliwe są dwa podejścia: przyjęcie stałej wartości progu lub adaptacyjne wyliczanie jego wartości dla każdego z obrazów. Metoda adaptacyjna bazuje na obserwacji, iż krokowe zwiększanie wartości progu powoduje zwiększanie się powierzchni obszaru segmentacji. Wraz ze zwiększaniem wartości progu wzrost ten stopniowo maleje. Przy pewnej wartości progu (zmianie jego wartości), powierzchnia obszaru segmentacji gwałtownie rośnie, ponieważ w jej skład zaczynają wchodzić także obszary o niskim stopniu podobieństwa do koloru skóry. Jeśli dla pewnej zmiany wartości progu (wartość jest zwiększana) obserwuje się najmniejszy ze wszystkich wzrost powierzchni obszaru segmentacji, wtedy przyjmuje się, że

wartość optymalna mieści się w tym przedziale. Przykładowo, jeśli wartość progu jest zwiększana od $0,05$ do $0,65$ ze skokiem $0,1$ i najmniejszy wzrost obszaru segmentacji następuje przy zmianie z $0,35$ na $0,45$ wtedy jako wartość optymalną przyjmuje się $0,4$. Otrzymany w ten sposób obraz binarny zawiera te regiony obrazu wejściowego, które z największym prawdopodobieństwem reprezentują kolor ludzkiej skóry.



Rys. 2.13. Obraz wejściowy (a), obraz podobieństw (dopełnienie) (b), obraz po segmentacji (c) (przestrzeń YCbCr)

W związku z powyższym może on zawierać nie tylko ludzką twarz (twarze) ale także inne odkryte części ciała (ręce, ramiona, szyję) lub, co często ma miejsce w praktyce, zupełnie inne obiekty. Kolejne etapy detekcji pozwolą na wybranie tych obszarów, które z największym prawdopodobieństwem reprezentują ludzką twarz.

2.3.2. Selekcja obszarów twarzy

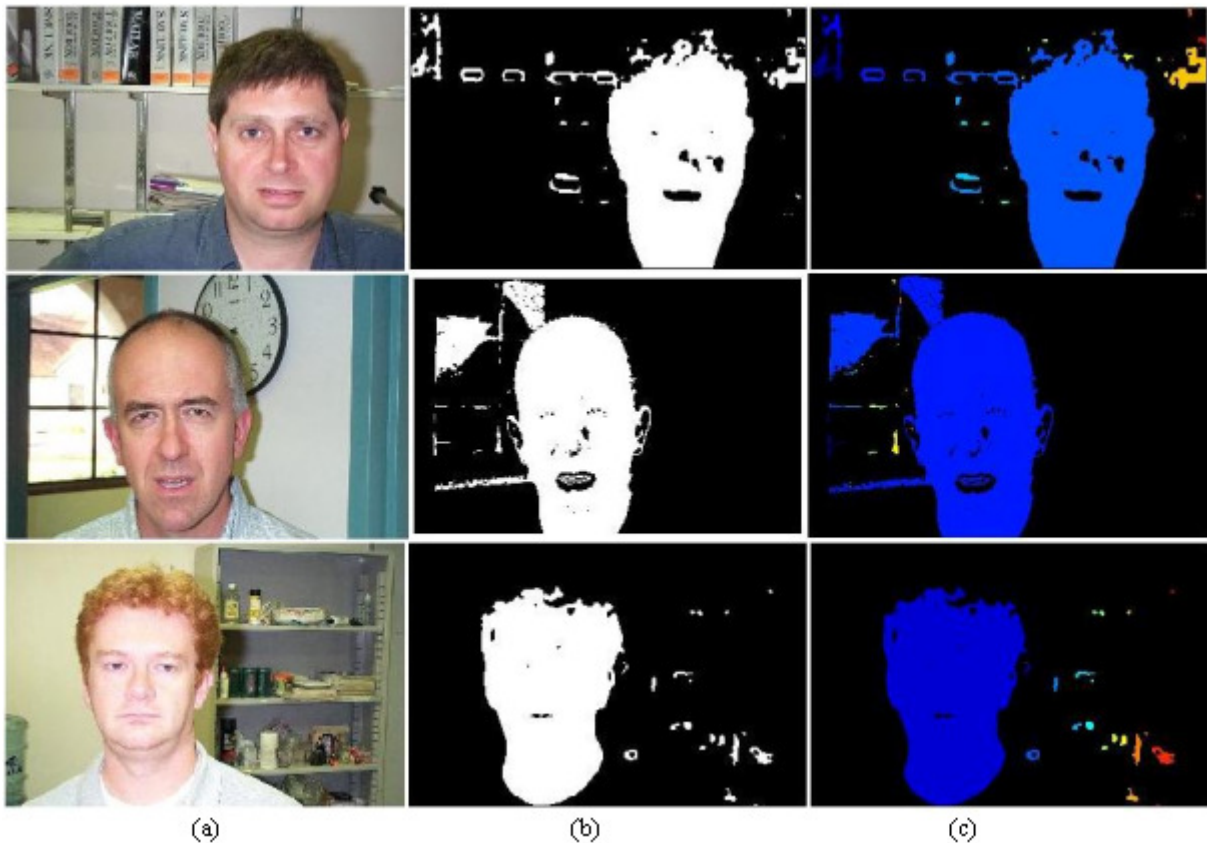
Otrzymany w wyniku segmentacji obraz binarny rozpatruje się jako zbiór niezależnych regionów i każdy z nich analizuje osobno.

Przyjęto następujące założenia:

- a) ludzką twarz definiuje się jako zamknięty, eliptyczny obszar na obrazie, zawierający jedną lub więcej dziur (oczy, usta),
- b) pojedynczy region może zawierać co najwyżej jedną twarz.
- c) stosunek wysokości do szerokości obszaru twarzy jest nie większy niż 3,
- d) obszar twarzy stanowi co najmniej 1/5 powierzchni największego regionu obrazu segmentacji i jest niemniejszy niż 1/10 wielkości całego obrazu,
- e) współczynnik korelacji pomiędzy obszarem twarzy a wzorcem jest większy lub równy 0,6.

Warunkiem pozytywnej kwalifikacji danego regionu jest spełnienie wszystkich wymienionych warunków.

Po indeksacji obrazu binarnego każdemu regionowi zostaje przypisana unikalna etykieta wykorzystywana do jego identyfikacji w dalszym przetwarzaniu.



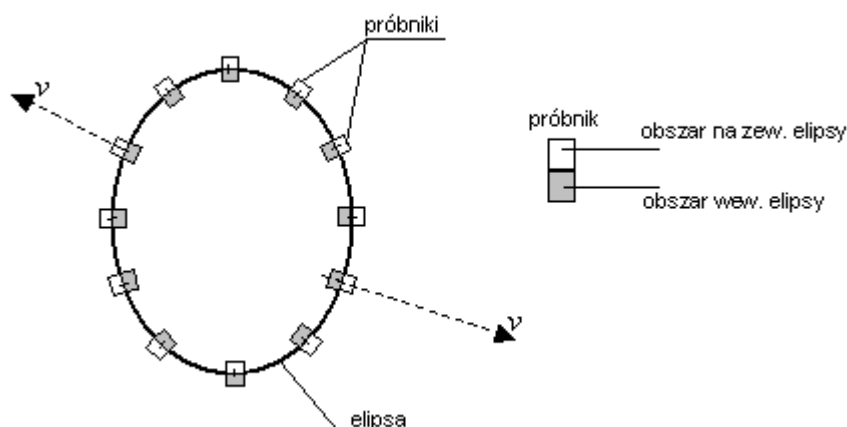
Rys. 2.14. Obraz wejściowy (a), obraz po segmentacji (b), obraz po indeksacji (c)

Następnie dla każdego ze zidentyfikowanych w ten sposób regionów wykonywane jest:

- sprawdzenie warunków dotyczących jego powierzchni i proporcji,
- obliczenie liczby dziur wewnątrz regionu,
- ekstrakcja eliptycznego obszaru twarzy,

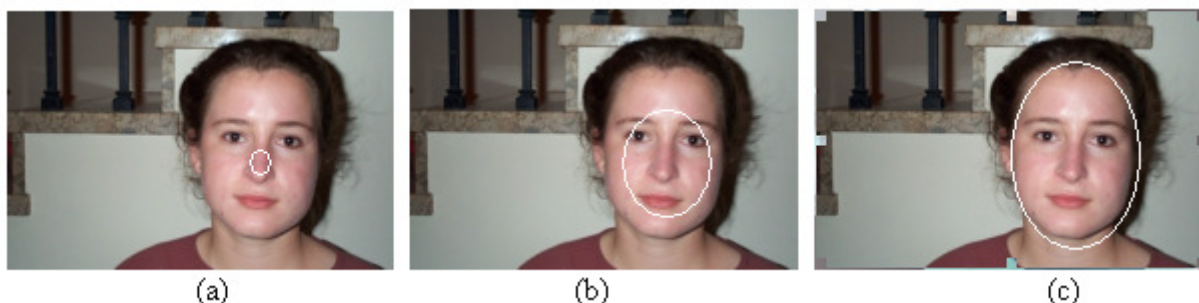
- sprawdzenie stosunku wysokości do szerokości obszaru,
- określenie orientacji obszaru twarzy,
- sprawdzenie stopnia korelacji z przechowywanym uniwersalnym wzorcem środkowej części twarzy.

Wstępna selekcja pozwala na wyeliminowanie tych obszarów, które nie spełniają kryteriów powierzchniowych i proporcji oraz nie zawierają wewnątrz odpowiedniej liczby dziur. Regiony te mogą reprezentować zarówno inne części ciała człowieka (np. odkryte ręce, nogi itd.) jak i inne obiekty. Po wstępnej kwalifikacji regionu kolejnym etapem jest ekstrakcja eliptycznego obszaru obejmującego twarz. Do tego celu wykorzystywany jest deformowalny model eliptyczny (rys. 2.15).



Rys. 2.15. Deformowalny model eliptyczny

Inicjacja modelu następuje w miejscu najbardziej prawdopodobnego położenia środka twarzy. Może to być np. środek masy rozpatrywanego regionu (rys. 2.16a). Deformacja modelu odbywa się z wykorzystaniem zbioru prostokątnych próbników zmieniających swe położenie zgodnie z ustalonym algorytmem (rys. 2.16b i rys. 2.16c).



Rys. 2.16. Zasada działania modelu deformowalnego

Niech N_{in} i N_{out} oznaczają odpowiednio liczbę pikseli próbniaka odpowiadających kolorowi skóry wewnątrz i na zewnątrz elipsy twarzy a S_p - pole powierzchni pojedynczego

próbnika. Gęstość rozkładu pikseli odpowiadających kolorowi skóry próbnika wewnątrz i na zewnątrz elipsy twarzy kontroluje wartość wektora przemieszczenia i -tego próbnika \vec{v}_i :

$$\vec{v}_i = \begin{cases} -k_{in} \times \vec{n}_i, & \text{gdy } \frac{N_{in}}{S_p} < T_1, \\ k_{out} \times \vec{n}_i, & \text{gdy } \frac{N_{out}}{S_p} > T_2, \\ 0, & \text{w przeciwn. przyp} \end{cases} \quad (25)$$

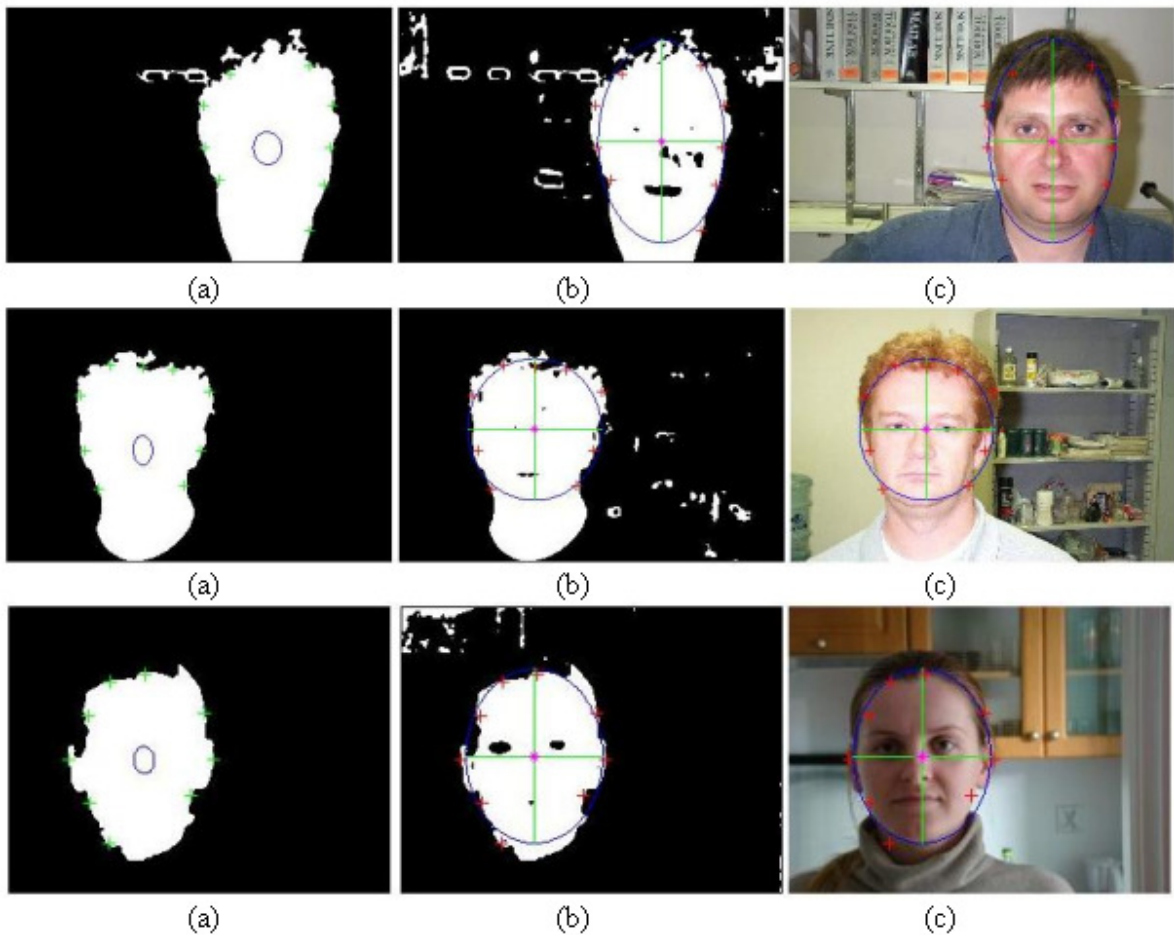
gdzie: i – indeks próbnika,
 \vec{n}_i – wektor kierunku ekspansji i -tego próbnika (o kierunku zgodnym z kierunkiem normalnej do krawędzi elipsy i zwrocie na zewnątrz)
 T_1, T_2 – wartości progowe (dobre doświadczalnie),
 k_{in}, k_{out} – współczynniki szybkości (≥ 0) deformacji modelu.

Po ustaleniu się wartości wszystkich wektorów przemieszczenia, kształt elipsy dopasowywany jest do środków przemieszczonych próbników wyznaczając tym samym granicę owalu twarzy. Wykorzystanie eliptycznego szablonu pozwala na uniezależnienie się w pewnym stopniu od błędów segmentacji powstałych np. w sytuacji, gdy tło obrazu tworzy z obszarem twarzy spójny region. Rezultaty działania algorytmu przedstawiono na rys. 2.17.

Po wyznaczeniu owalu twarzy, obliczane są przybliżone parametry obszaru: szerokość, wysokość, kąt nachylenia względem poziomu (orientacja owalu). Jeśli stosunek wysokości do szerokości mieści się w ustalonym przedziale, kolejnym etapem jest badanie stopnia korelacji wyznaczonego obszaru z uniwersalnym deformowalnym wzorcem środkowej części twarzy. Jeśli stosunek ten przekracza połowę górnej granicy dopuszczalnego przedziału ($r > 1,5$) przy spełnieniu pozostałych warunków, system próbuje skorygować rozmiar obszaru (wysokość) tak, aby proporcje obszaru mieściły się poniżej tej wartości. Wyznaczone parametry owalu pozwalają na lepsze dopasowanie wzorca do badanego obszaru.

Wzorcem jest uśredniona środkowa część twarzy, bez okularów, pozbawiona zarostu i innych ozdób, ustawiona frontem względem oka kamery. Proces dopasowywania wzorca do wyznaczonego owalu polega na takiej jego modyfikacji, aby jego rozmiary i orientacja były zgodne z wyznaczonymi wcześniej parametrami owalu. Następnie wyliczana jest wartość korelacji pomiędzy dopasowanym wzorcem a wyznaczonym owalem twarzy. Jeżeli jest ona większa niż 0,6 przyjmuje się, że wyznaczony owal rzeczywiście reprezentuje ludzką twarz. Przeprowadzone doświadczenia wykazały, że nie ma ścisłego wymagania co do orientacji

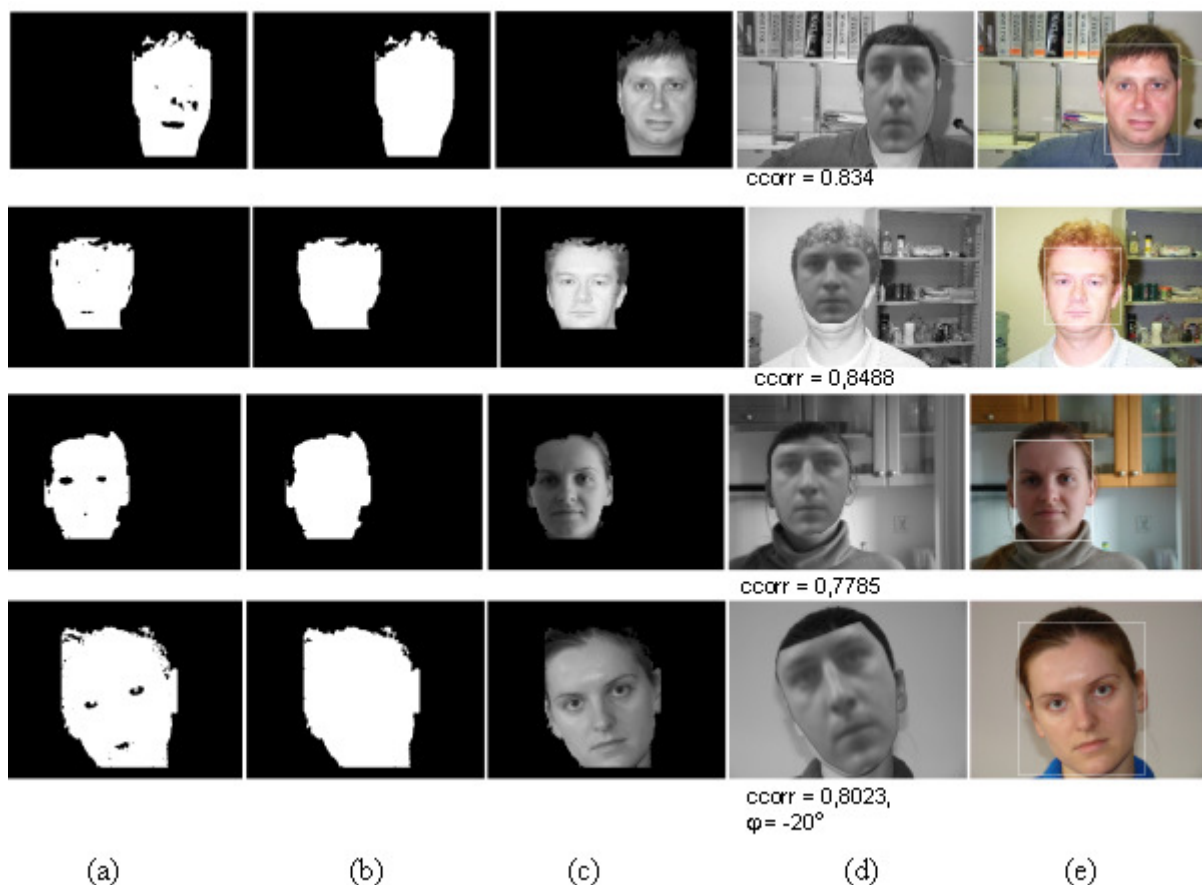
twarzy na obrazie wejściowym – ważne, żeby kąt obrotu twarzy względem osi Y zawierał się w przedziale ± 15 stopni względem pozycji frontalnej.



Rys. 2.17. Wstępne wyznaczenie owalu twarzy; (a) zakwalifikowany region obrazu binarnego z zaznaczoną pozycją startową modelu i wyznaczonymi punktami końcowymi; wyznaczony owal na obrazie binarnym (b) i wejściowym (c)

Po pozytywnej klasyfikacji regionu wyznaczony na jego podstawie prostokątny fragment obrazu wejściowego poddawany jest dalszemu procesowaniu. Na rysunku 2.18 pokazano kolejne etapy wyznaczania regionu twarzy dla przykładowych obrazów.

Obraz (a) przedstawia rezultat wyznaczania owalu twarzy, obraz (b) to owal twarzy z wypełnionymi wewnętrznymi otworami; (c) to rezultat przemnożenia obrazu (b) przez obraz wejściowy; (d) – rezultat dopasowania wzorca do wyznaczonego obszaru twarzy wraz z wyliczonymi współczynnikami korelacji; obraz (e) przedstawia rezultat końcowy – wyznaczony region twarzy.



Rys. 2.18. Wyznaczanie regionu twarzy

2.4. Ekstrakcja cech

Prawidłowa ekstrakcja cech ma kluczowe znaczenie dla całego procesu identyfikacji. Rezultat jej działania uzależniony jest od wielu czynników m.in. jakości obrazu wejściowego, orientacji twarzy względem obserwatora, poprawnej lokalizacji twarzy na obrazie, obecności na twarzy innych obiektów (okularów, kolczyków) czy zarostu identyfikowanej osoby uniemożliwiającego lokalizację niektórych cech twarzy. W związku z tym, konieczne staje się określanie wymagań dotyczących identyfikowanych twarzy. Dopiero gdy wymagania te są spełnione można przystąpić do dalszego procesowania.

Celem ekstrakcji jest wyszukanie i matematyczny opis wybranych cech twarzy. W kolejnym kroku na tej podstawie budowany jest wektor cech wykorzystywany do identyfikacji osoby. W przypadku opisywanego systemu proces składa się z następujących etapów:

- określenia położenia i rozstawu oczu,
- sprawdzenia orientacji twarzy na obrazie,

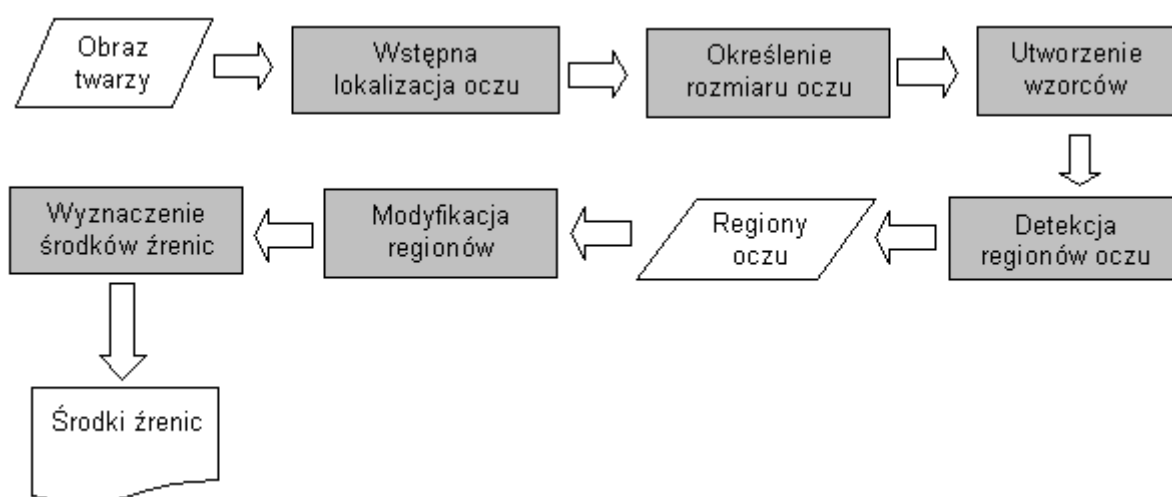
- określenia szerokości ust,
- określenia szerokości nosa,
- określenia szerokości twarzy – granic owalu na poziomie linii nosa.

Na podstawie wyznaczonych parametrów budowany jest wektor cech opisujący twarz oraz wyznaczana jest środkowa część twarzy, wykorzystywana do weryfikacji rezultatów rozpoznawania algorytmu podstawowego metodą korelacji fazowej.

2.4.1. Detekcja oczu

Oczy to jedne z najmniej narażonych na zniekształcenia miejsc na twarzy człowieka i jednocześnie bardzo charakterystyczne. Znalezienie ich jest podstawą do obliczenia pozostałych parametrów twarzy. Na podstawie wiedzy o położeniu oczu można określić m.in. oś symetrii twarzy, orientację twarzy, rozmiar itd. Wiedza na temat ich położenia pozwala również na normalizację obszaru twarzy (rozmiar, orientacja) zgodnie z wymaganiami. Skuteczność metod detekcji ograniczona jest jedynie poprzez możliwe zamknięcia oczu lub wpływ okularów.

Proponowana metoda opiera się na wykorzystaniu kilku technik: funkcji projekcji horyzontalnej i wertykalnej obrazu gradientowego do wstępnego wyznaczenia położenia oczu, metody „*template matching*” do lokalizacji regionów oczu oraz metody *Daugmana* [5] do lokalizacji środków źrenic. Schemat postępowania przedstawiono na rysunku 2.19.



Rys. 2.19. Schemat postępowania w algorytmie lokalizacji środków źrenic

Wstępna lokalizacja oczu

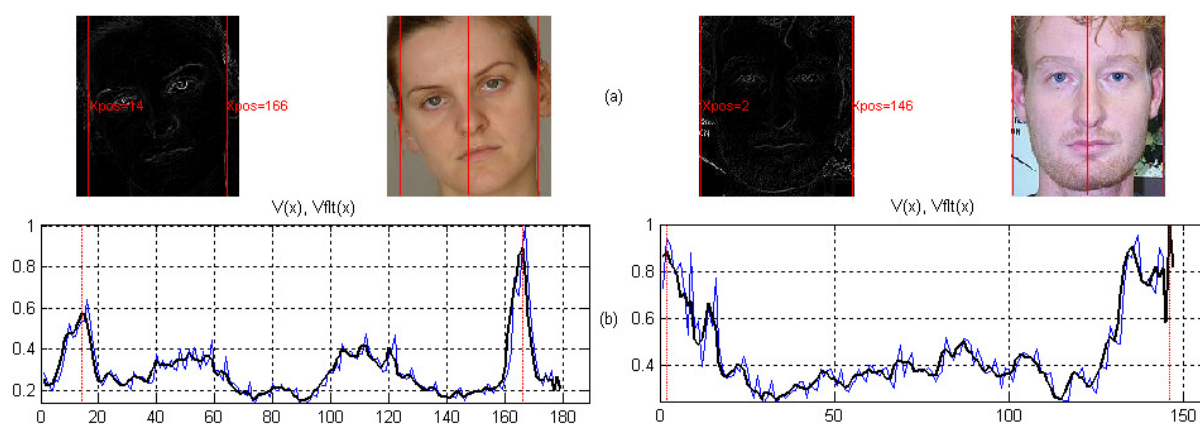
Podstawowym zadaniem jest poprawne wyznaczenie prostokątnych obszarów bezpośredniego sąsiedztwa oczu. Do tego celu wykorzystana zostanie para wzorców dla lewego i prawego oka. Ponieważ przeszukanie całego obszaru twarzy metodą „*template matching*” byłoby bardzo kosztowną i nieefektywną z punktu widzenia złożoności obliczeniowej i czasowej metodą, konieczne staje się wstępne wyznaczenie regionu poszukiwań. Do tego celu wykorzystano metody projekcji horyzontalnej i wertykalnej obrazu gradientowego twarzy oraz wiedzę na temat budowy ludzkiej twarzy. Rysunek 2.20 przedstawia etapy procedury wstępnej lokalizacji oczu.

Obraz wejściowy twarzy w formacie *RGB* transformowany jest do przestrzeni *III2I3*, która dobrze eksponuje ciemniejsze fragmenty twarzy. Następnie dla tak zmodyfikowanego obrazu $I(y, x)$ ($\forall y = 1, 2, \dots, M-1$ oraz $\forall x = 1, 2, \dots, N-1$) wyliczane są gradienty jasności pikseli Dy i Dx odpowiednio wzdłuż kolumn i wierszy obrazu. Na tej podstawie obliczany jest gradient obrazu $D(y, x)$ zgodnie z:

$$D(y, x) = \sqrt{DyDy + Dx Dx} \quad (26)$$

Dla obrazu gradientowego wyliczana jest projekcja wertykalna $V(x)$. Funkcja ta określa sumę wszystkich wierszy obrazu gradientowego:

$$V(x) = \sum_{y=1}^M D(y, x), \quad \forall x = 1, 2, \dots, N \quad (27)$$



Rys. 2.20. Projekcja wertykalna obrazu gradientowego twarzy (b) i rezultat określenia szerokości twarzy (a)

Na wykresie funkcji widoczne są dwa wyraźne piki w okolicach granic obrazu twarzy odpowiadające prawemu i lewemu konturowi twarzy. Ustalenie ich położenia pozwala na wyliczenie szerokości twarzy.

Kolejnym krokiem jest ustalenie położenia linii oczu. W zależności od wyników otrzymanych w trakcie lokalizacji twarzy wybierana jest inna metoda wyliczania ich pozycji. W przypadku, gdy kąt obrotu twarzy względem osi Z (na płaszczyźnie XY) nie przekracza 15 stopni linia oczu wyliczana jest na podstawie analizy projekcji horyzontalnej całego obszaru twarzy, w przeciwnym wypadku – dla każdej połowy oddzielnie. W obu przypadkach metodologia jest podobna, omówiony więc zostanie pierwszy przypadek z zaznaczeniem ewentualnych różnic.

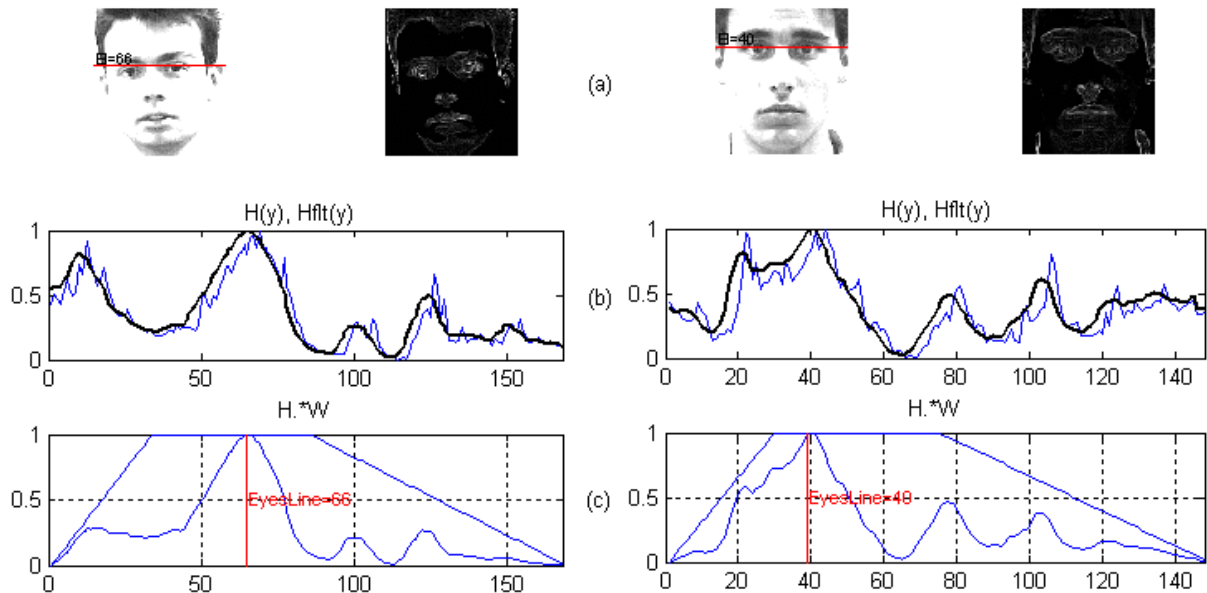
Funkcja projekcji horyzontalnej $H(y)$ określa sumy wszystkich kolumn obrazu gradientowego $D(y,x)$ tzn.:

$$H(y) = \sum_{x=1}^N D(y,x), \quad \forall y = 1,2,\dots,M \quad (28)$$

Maksima lokalne funkcji $H(y)$ odpowiadają położeniu charakterystycznych cech twarzy: oczu, nosa, ust i podbródka. Aby wyznaczyć te, które aktualnie są potrzebne wprowadza się pomocniczą funkcję okna $W(y)$ – określa ona interwał, w którym mogą znajdować się żądane cechy. Kształt funkcji okna $W(y)$ (rys. 2.21b) ustalany jest w oparciu o posiadaną wiedzę o prawdopodobnej lokalizacji szukanych cech twarzy. W przypadku oczu wartości i parametry funkcji okna $W(y)$ obliczane są na podstawie założeń [12]:

$$0.2M \leq m_1 \leq 0.32M; m_2 \leq 0.42M; m_3 = M - (m_1 + m_2 + 2),$$

gdzie wartości parametrów m_1 oraz m_2 są zaokrąglone do najbliższej liczby całkowitej;

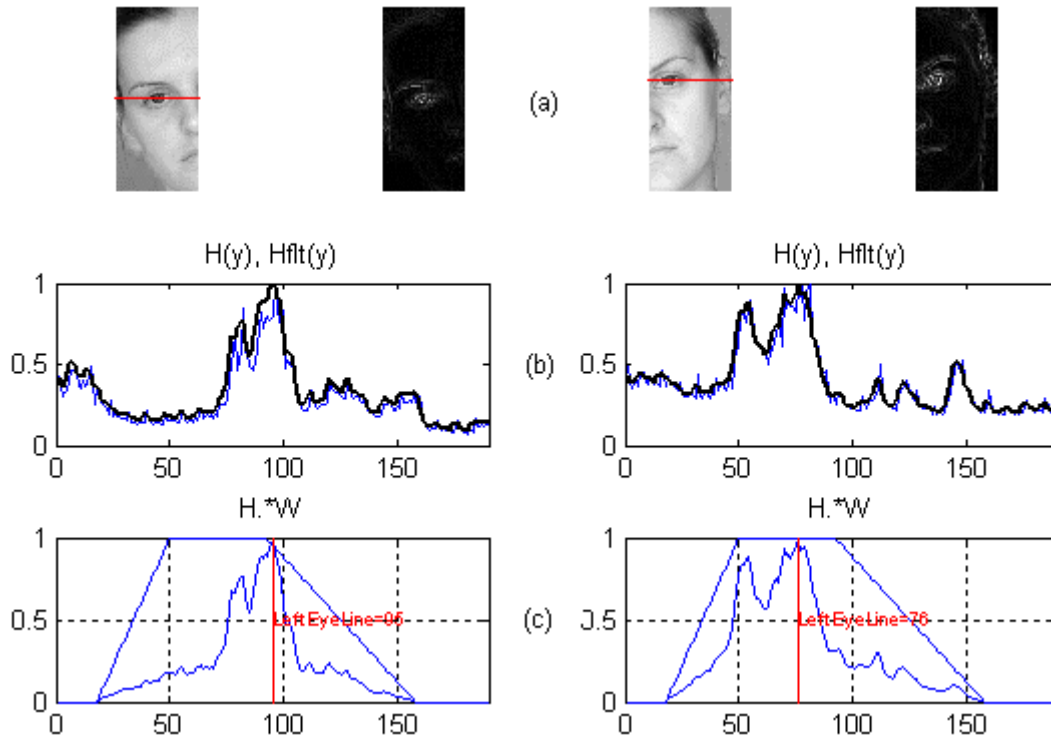


Rys. 2.21. Obliczanie położenia linii oczu na obrazie twarzy; (a) obraz wejściowy z linią oczu i obraz gradientowy; (b) - projekcja horyzontalna obszaru twarzy $H(y)$ i $Hflt(y)$ – funkcja po filtracji dolnoprzepustowej; (c) – rezultat pokropkowego przemnożenia funkcji $H(y)$ i $W(y)$

Funkcja $W(y)$ przyjmuje postać:

$$W(y) = \begin{cases} (y-1)/m_1, & \text{dla } y=1, 2, \dots, (m_1+1); \\ 1, & \text{dla } y=(m_1+2), (m_1+3), \dots, (m_1+m_2+1); \\ (M-y)/m_3, & \text{dla } y=(m_1+m_2+2), (m_1+m_2+3), \dots, M. \end{cases} \quad (29)$$

W przypadku wyznaczania linii oczu dla każdej połowy twarzy niezależnie, konieczna staje się modyfikacja kształtu funkcji (rys. 2.22) ponieważ wpływ oka na obrazie połowy twarzy na funkcję $H(y)$ może być mniejszy niż na przykład ust lub granicy włosy-czoło.



Rys. 2.22. Obliczanie położenia linii oczu dla każdej połowy twarzy; (a) obrazy wejściowe z linią oczu i ich odpowiadające im obrazy gradientowe; (b) – projekcja horyzontalna lewej i prawej połowy twarzy $H(y)$ i $Hflt(y)$ – funkcja projekcji po filtracji dolnoprzepustowej; (c) – rezultat pokropkowego przemnożenia funkcji $H(y)$ i $W(y)$.

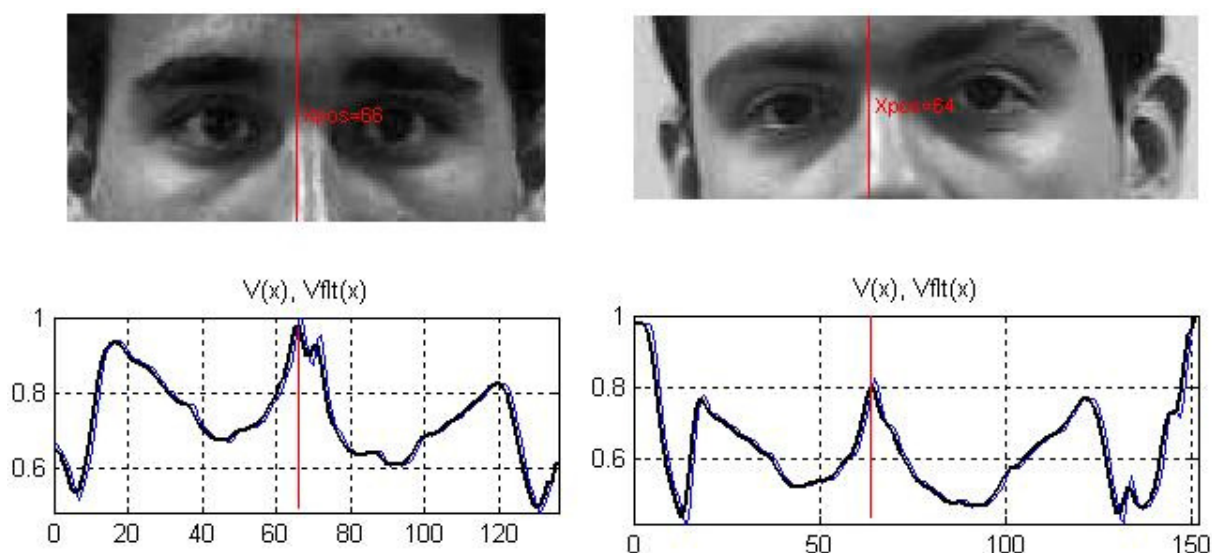
Wpływ ten zależy przede wszystkim od stopnia pochylenia twarzy: przy dużych pochyleniach twarzy granica połowy obrazu twarzy może trafić bezpośrednio na obszar oka.

Po ustaleniu kształtu funkcji okna $W(y)$, funkcja projekcji jest filtrowana dolnoprzepustowo a następnie przemnażana pokropkowo przez funkcję okna. W rezultacie otrzymuje się nową funkcję (rys. 2.21c):

$$F(y) = H(y) * W(y) \quad (30)$$

której maksimum wskazuje na położenie linii oczu.

Po wyznaczeniu linii oczu definiowany jest nowy obszar, rozciągający się symetrycznie wzdłuż niej, którego wysokość uzależniona jest od wysokości twarzy. Dla tak określonego obszaru wyliczana jest projekcja wertykalna (rys. 2.23).



Rys. 2.23. Projektcja wertykalna obszaru oczu

Środkowe maksimum odpowiada w przybliżeniu wertykalnemu centrum twarzy w ramach analizowanego obszaru. Jego obecność związana jest z faktem, że obszar pomiędzy oczami jest znacznie jaśniejszy niż jego bezpośrednie sąsiedztwo w ramach analizowanego regionu.

Położenie maksimum wyznacza jednocześnie granice obszarów oczu dla każdej połowy twarzy.

Detekcja regionów oczu

Detekcja regionów oczu odbywa się z wykorzystaniem metody „*template matching*” w ramach obszarów wyznaczonych dla każdej połowy twarzy. Metoda polega na przeglądzie obszaru lokalną maską i porównaniu obrazu w polu tej maski z pewnym wzorcem. W przypadku detekcji oczu wzorcami są obrazy oczu (lewego i prawego) pobrane z rzeczywistych twarzy. Rozmiar i orientacja wzorca nie musi dokładnie odpowiadać wydzielanemu obszarowi obrazu wejściowego jednak najlepsze rezultaty osiągnęte są wtedy, gdy różnica ta mieści się w przedziale 5-10%. Dysponując informacjami na temat wielkości obszarów sąsiedztwa oczu, szerokości twarzy i jej geometrycznej strukturze można określić rozmiar wzorców wystarczająco dokładnie. Niewielkie odchyłki nie mają większego wpływu na wynik porównania.

Najczęściej używaną miarą różnicy pomiędzy wzorcem a obszarem w polu maski jest wartość błędu średniokwadratowego wyrażonego jako:

$$MSE = \sum_{(i,j) \in R} (f - g)^2 \quad (31)$$

gdzie f – obszar w polu maski, g – wzorec i R – obszar wzorca.

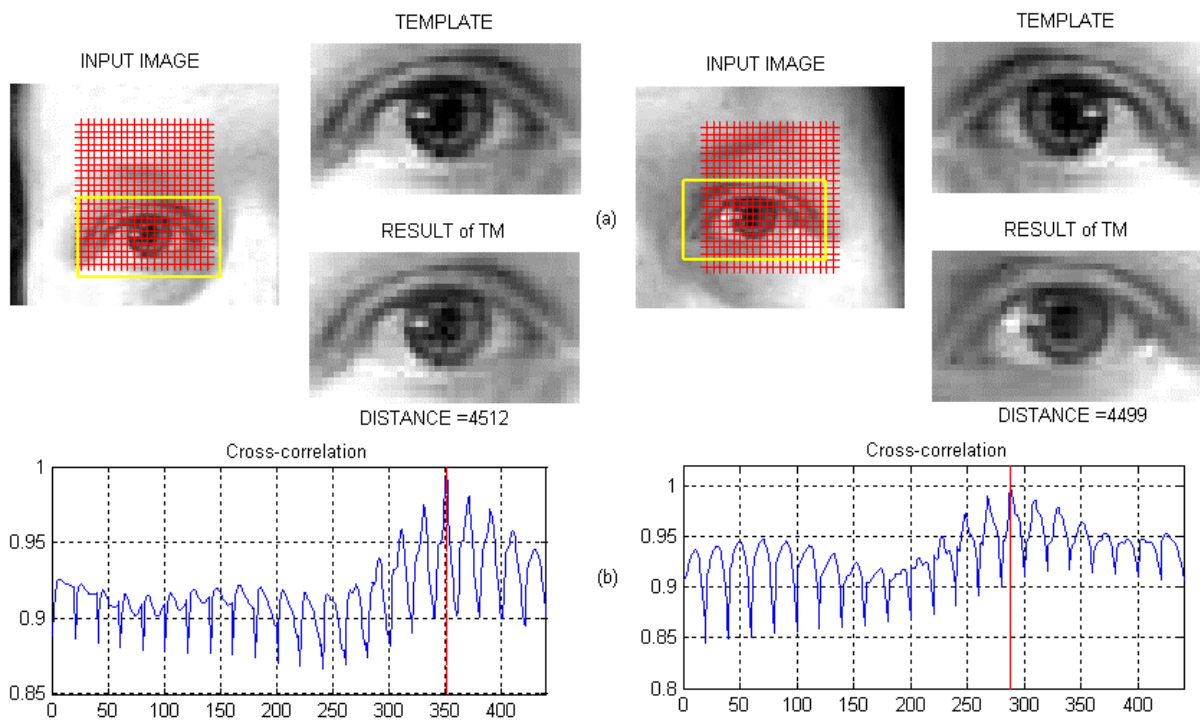
Minimum średniokwadratowego błędu otrzymane w rezultacie całego przeglądu, świadczy o najlepszym podobieństwie między obszarem w polu maski a wzorcem.

W przypadku metody „*template matching*” miara (31) może zostać określona pośrednio redukując koszt obliczeniowy całej procedury. Równanie (31) zapisać można jako:

$$\sum_{[i,j] \in R} (f - g)^2 = \sum_{[i,j] \in R} f^2 + \sum_{[i,j] \in R} g^2 - 2 \sum_{[i,j] \in R} fg \quad (32)$$

Przyjmując teraz, że f i g są niezmiennie, miarą różnicy staje się $\sum fg$ (wzajemna korelacja).

Obraz przeszukiwany jest poczynając od lewego górnego rogu a kończąc na przeciwległym krańcu pomniejszonym o wymiary wzorca. W celu przyspieszenia skanowania obrazu przejście do nowej pozycji maski odbywa się co pewien krok większy niż jeden piksel.



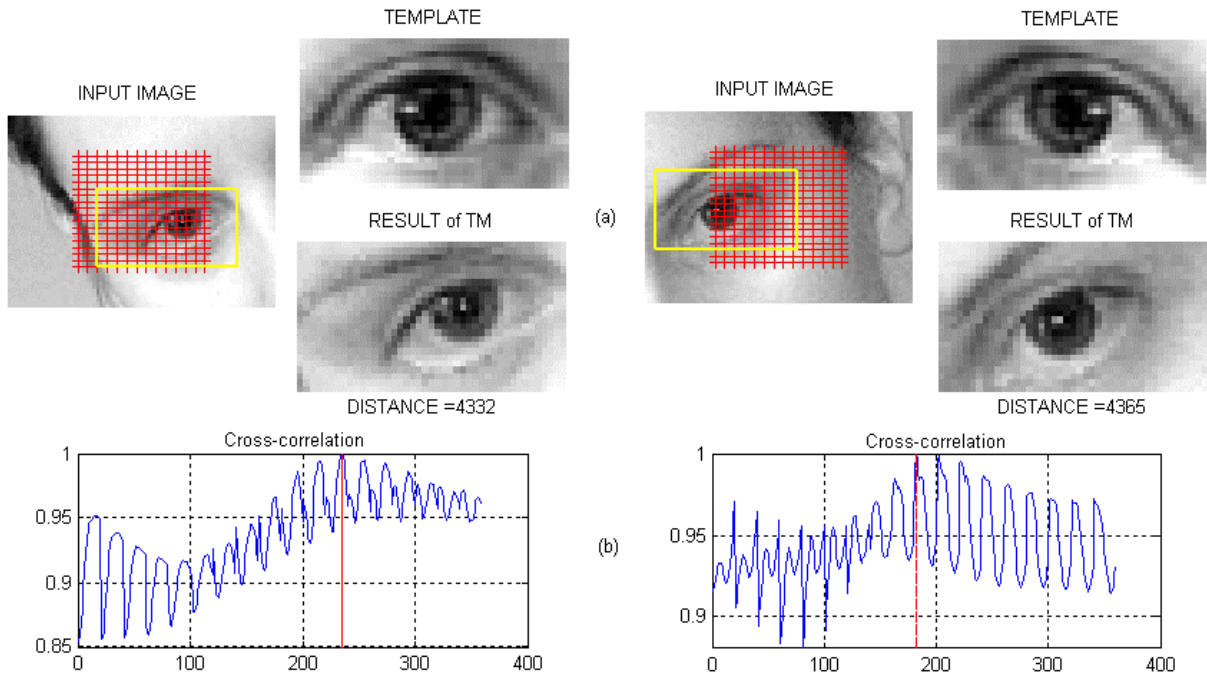
Rys. 2.24. Detekcja regionów oczu dla każdej połowy twarzy metodą „*template matching*”; (a) – wydzielony obszar poszukiwań z obrazem wzorca i rezultatem; (b) – wykres znormalizowanej wzajemnej korelacji wzorca z obszarem wewnątrz pola maski

Zakładając, że rozmiar wzorca wynosi $m \times n$ a rozmiar przeszukiwanego obrazu $M \times N$ dla każdego punktu obrazu obliczana jest wartość znormalizowanej (ze względu na założenie o niezmienności f i g) wzajemnej korelacji:

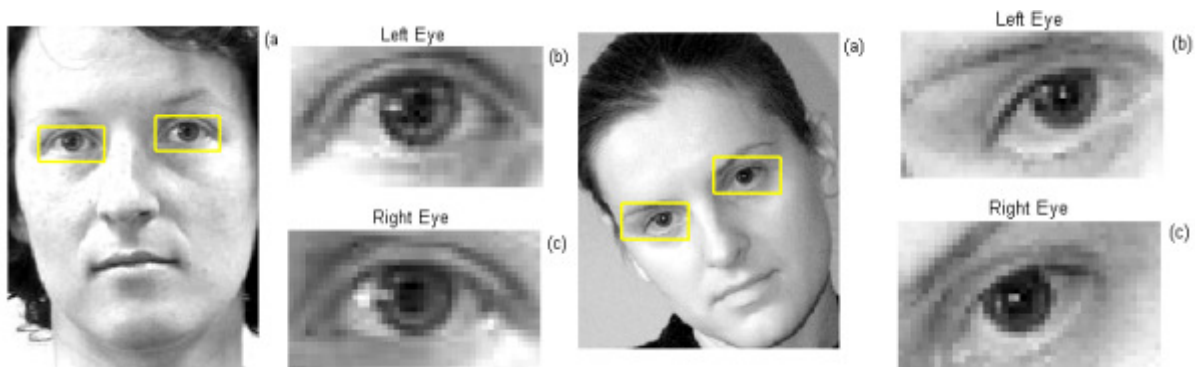
$$M(i, j) = \frac{C_{fg}(i, j)}{\left\{ \sum_{k=1}^m \sum_{l=1}^n f^2(i+k, j+l) \right\}^{1/2}} \quad (33)$$

gdzie: $C_{fg}(i, j) = \sum_{k=1}^m \sum_{l=1}^n g(k, l) f(i + k, j + l)$

Największa wartość $M(i, j)$ wskazuje na punkt, w otoczeniu którego najprawdopodobniej znajduje się obszar podobny do wzorca. Obszar ten jest poszukiwanym regionem oczu.



Rys. 2.25. Detekcja regionów oczu dla każdej połowy twarzy metodą „*template matching*” w przypadku znacznego nachylenia twarzy i niedokładnego dopasowania wzorców ; (a) – wydzielony obszar poszukiwań z obrazem wzorca i rezultatem; (b) – wykres znormalizowanej wzajemnej korelacji wzorca z obszarem wewnątrz pola maski



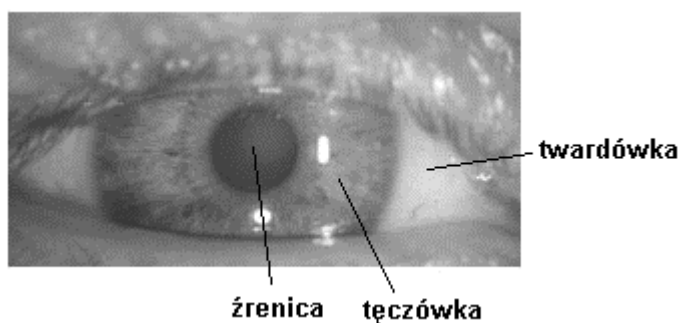
Rys. 2.26. Rezultaty detekcji oczu

Wyznaczanie środków źrenic

Ostatnim etapem procedury detekcji oczu jest wyznaczenie środków źrenic. Danymi wejściowymi są w tym przypadku wyznaczone regiony oczu. Obliczone współrzędne

środków źrenic zostaną wykorzystane do normalizacji skali i orientacji twarzy, określenia niektórych parametrów twarzy oraz budowy wektora cech.

Kształt i wzajemne położenie tęczówki i źrenicy można w przybliżeniu opisać za pomocą dwóch współśrodkowych okręgów (rys. 2.27). Obliczając parametry okręgu opisującego kształt tęczówki można określić współrzędne środka źrenicy. Ze względu na swe właściwości tęczówka stanowi łatwiejszy w lokalizacji element ludzkiego oka w stosunku do źrenicy. Wynika to z faktu, iż granica tworzona przez tęczówkę z twardówką jest dużo bardziej kontrastowa niż ta tworzona ze źrenicą – w konsekwencji łatwiejsza do zlokalizowania.



Rys. 2.27. Ludzkie oko

Na algorytm lokalizacji środków źrenic składa się kilka etapów:

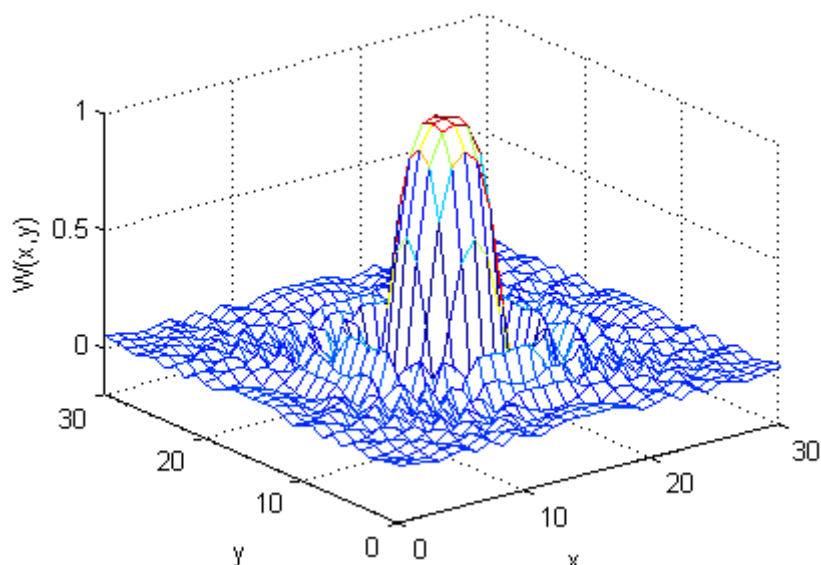
- normalizacja skali wyznaczonych regionów oczu,
- modyfikacja regionów w celu wzmocnienia kolistych regionów ciemnych pikseli,
- detekcja tęczówki z wykorzystaniem operatora *Daugmana*,
- wyznaczenie środków źrenic.

Operacje związane z lokalizacją środków źrenic odbywają się na wydzielonym z obrazu wejściowego kanale R , wzmacniającym dodatkowo granice tęczówki. Związane jest to z tym, iż tęczówka reprezentowana jest zazwyczaj niskimi wartościami czerwieni (niezależnie od koloru oczu) w odróżnieniu od otaczających ją pikseli (twardówki i skóry) mających znacznie większe wartości składowej R .

Pierwszym krokiem jest normalizacja skali wyznaczonego regionu oczu. Dzięki temu wykorzystywane algorytmy detekcji będą mogły operować na bezwzględnych wartościach parametrów i progów. Następnie wyznaczony region oczu poddawany jest modyfikacji mającej na celu wzmocnienie kolistych regionów ciemnych pikseli odpowiadających obszarowi źrenicy. Modyfikacja składa się z dwóch etapów: najpierw w celu eliminacji często występujących na źrenicy błysków światła („flash” aparatu, odbicie sztucznego oświetlenia

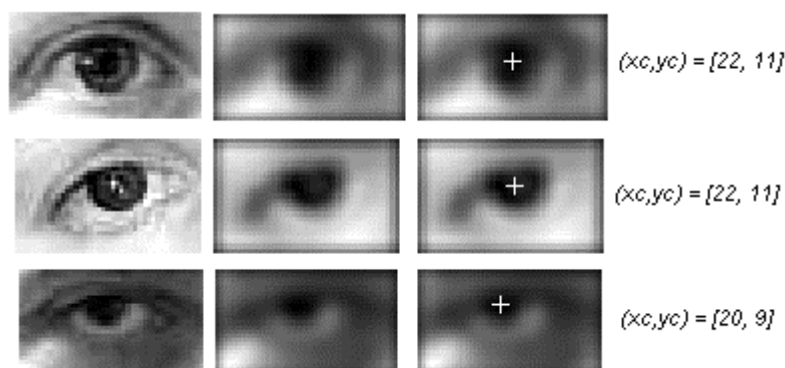
itp.) mogących mieć niekorzystny wpływ na wynik detekcji tęczy, obraz jest filtrowany filtrem minimum o wielkości maski 5x5. Następnie wykonywana jest konwolucja obrazu oka z funkcją wagową $W(x,y)$ postaci [22]:

$$W(x, y) = \frac{\sin((x^2 + y^2) / c)}{(x^2 + y^2) / c} \quad (34)$$



Rys. 2.28. Funkcja wagowa $W(x,y)$ wzmacniająca obszary koliste w centrum maski konwolucji

Funkcja wzmacnia najbardziej koliste obszary w centrum maski konwolucji – w przypadku oczu jest to obszar źrenicy i tęczy charakteryzujący się niskimi wartościami luminancji, pozostałej części procesowanego obszaru przyporządkowując ujemne wagi. Promień obszaru o największych wagach regulowany jest parametrem c i jego wartość powinna w przybliżeniu odpowiadać przewidywanej wielkości tęczy. Po konwolucji, uśrednione położenie grupy najciemniejszych pikseli wskazuje na przybliżone położenie środka źrenicy (rys. 2.29)



Rys. 2.29. Modyfikacja regionów oczu; (a) – obraz wejściowy, składowa R obrazu oka; (b) – obraz oka po filtracji i konwolucji; (c) – przybliżony położenie środka źrenicy, wyliczone jako środek masy obszaru najciemniejszych pikseli obrazu (b)

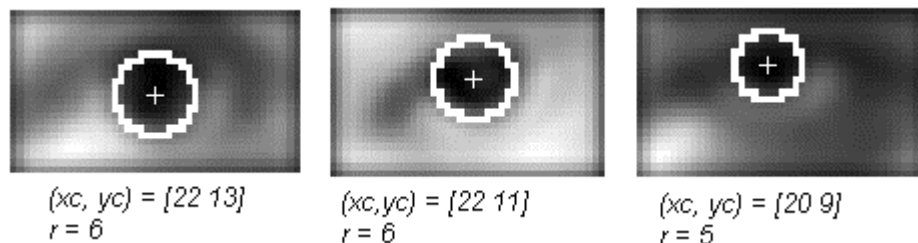
Zmodyfikowane obrazy oczu stanowią dane wejściowe dla procedury detekcji tęczy opartej na algorytmie zaproponowanym przez *Daugmana* [5]. Do lokalizacji kolistych regionów tęczy i źrenicy używany jest różniczkowo-całkowy operator definiowany jako:

$$\max_{(r,x_0,y_0)} \left| G_{\sigma}(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \quad (35)$$

gdzie $I(x,y)$ jest obrazem oka, r jest szukanym promieniem, $G_{\sigma}(r)$ jest gaussowską funkcją wygładzającą o odchyleniu standardowym σ , s jest konturem okręgu danym przez r, x_0, y_0 ,

* - oznacza konwolucję.

Operator przeszukuje obraz pod kątem kolistych obszarów, dla których występuje największa zmiana wartości pikseli, zmieniając wartość promienia r i współrzędne środka x i y kolistego konturu. Przeszukiwanie odbywa się iteracyjnie, redukując, w miarę osiągnięcia coraz dokładniejszego wyniku, stopień wygładzania. Ze względu na to, iż detekcja odbywa się w oparciu o informację o pierwszej pochodnej obrazu i geometrycznej charakterystyce szukanego obiektu, działanie operatora można porównać do działania detektora krawędzi kolistych obiektów opartego np. na transformacie *Hough'a*.



Rys. 2.30. Rezultaty lokalizacji środków źrenic

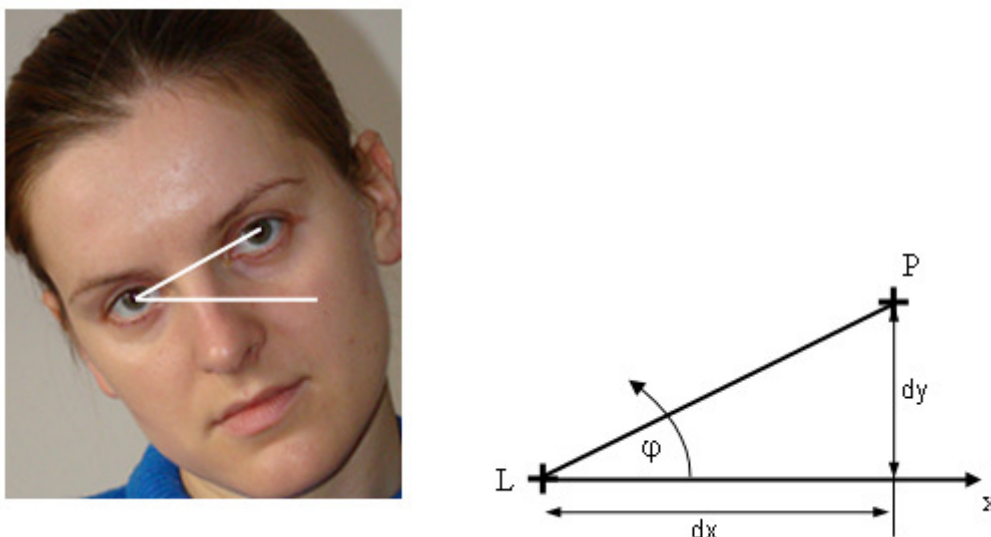
Wartość promienia r zmienia się w ustalonych granicach wyznaczonych na podstawie wielkości wyznaczonego regionu oka. Przykładowe wyniki działania algorytmu przedstawiono na rysunku 2.30. Na wynik detekcji składają się parametry okręgu (środek i promień) opisującego w najlepszym stopniu kształt tęczy zgodnie z opisanymi wyżej kryteriami. Współrzędne środka tak wyznaczonego okręgu są jednocześnie szukanymi współrzędnymi środka źrenicy.

2.4.2. Normalizacja orientacji twarzy w płaszczyźnie XY

Wstępne określenie nachylenia twarzy względem osi Z zostało już przeprowadzone na etapie lokalizacji twarzy (rozd. 2.3). Wtedy wykorzystywanymi informacjami były obliczone

cechy obszaru twarzy. Na obecnym etapie możliwe jest dokładniejsze wyliczenie kąta nachylenia twarzy w oparciu o wiedzę na temat położenia środków źrenic. Pozwoli to na ewentualną korektę jej położenia, polepszając wyniki identyfikacji. Eliminowanie nachylenia twarzy w płaszczyźnie XY nie zawsze jest niezbędne. Małe kąty nachyleń (+/- 5 stopni) nie powinny zaszkodzić procesowi identyfikacji i zmniejszyć jakości rozpoznania. Przy kątach nachylenia większych niż 30 stopni można mówić o nieprzydatności obrazu do jego rozpoznania.

Zasadę wyliczania kąta nachylenia twarzy w oparciu o wiedzę o położeniu oczu przedstawiono na rysunku 2.31



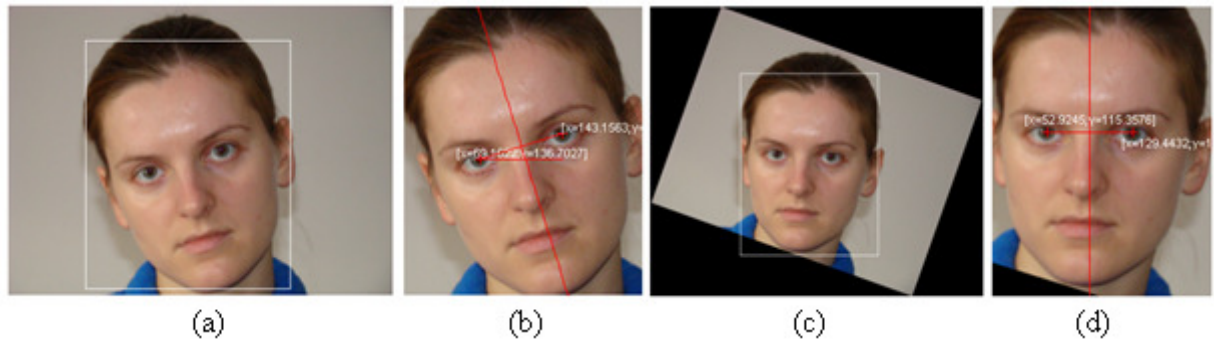
Rys. 2.31. Zasada określania kąta nachylenia twarzy w płaszczyźnie XY

Kąt nachylenia twarzy φ wylicza się według wzoru:

$$\pm \varphi = \arctg \frac{dy}{dx} \quad (36)$$

Jeżeli lewe oko na obrazie twarzy znajduje się poniżej prawego, to twarz jest przechylona w lewo (kąt φ ze znakiem +), w przeciwnym wypadku twarz jest pochylona w prawo (kąt φ ze znakiem -).

Cały proces normalizacji orientacji twarzy pokazany jest na rysunku 2.32.



Rys. 2.32. Normalizacja orientacji twarzy w płaszczyźnie XY; (a) – wejściowy obraz z zaznaczoną ramką twarzy; (b) – wydzielony obszar twarzy z wyliczonymi współrzędnymi oczu i osią symetrii; (c) – kompensacja nachylenia poprzez obrót obrazu wejściowego o kąt $-\varphi$ wyznaczona nowa ramka obszaru twarzy; (d) – znormalizowany obszar twarzy z wyliczonymi nowymi współrzędnymi oczu

Na obrazie wejściowym (rys. 2.32a) ramką zaznaczono zlokalizowany rejon twarzy. Po wyliczeniu współrzędnych lewego i prawego oka, określono kąt nachylenia linii oczu w odniesieniu do poziomu φ . Kompensacja nachylenia polega na obrocie obrazu twarzy o kąt $-\varphi$. Nowa ramka obszaru twarzy wyznaczana jest na podstawie pozycji obróconego segmentu twarzy z wejściowego obrazu twarzy. Wyliczona macierz transformacji pozwala na wyznaczenie współrzędnych oczu w nowym układzie współrzędnych (O, x_r, y_r) obróconym względem układu wejściowego (O, x, y) o kąt $-\varphi$. Ponieważ przy wykonaniu obrotu obrazu w płaszczyźnie XY jego rozmiar powiększa się (rys. 2.32c), konieczne jest uwzględnienie tego faktu w trakcie transformacji.

2.4.3. Detekcja ust

Ze względu na mimikę ludzkiej twarzy, usta są jednym z trudniejszych obiektów do lokalizacji. Ich kształt ulega zmianie - gdy mówimy, śmiejemy się, jesteśmy zdenerwowani czy smutni. W przypadku posiadania zarostu usta przestają być widoczne i tym samym ich lokalizacja staje się niemożliwa. Choć zdarzają się sytuacje, że zmianie ulega także ich kolor, w większości przypadków wyraźnie odcina się on od pozostałej części twarzy. Podobnie jak w przypadku oczu, fakt ten można wykorzystać do ich lokalizacji. Na obrazach w skali szarości intensywny czerwony kolor ust skutkuje tym, iż zajmowany przez nie obszar posiada wyraźnie mniejszą jasność niż sąsiednia część twarzy. Jest to szczególnie widoczne w przypadku, gdy są one otwarte. Poprzez odpowiednią modyfikację obrazu można tę cechę jeszcze bardziej uwypuklić. Analiza zmian kształtu ust w zależności od stanu emocjonalnego wykazała, że ich szerokość zmienia się w niewielkim stopniu. Własność ta zostanie

wykorzystana w proponowanej procedurze detekcji, której zadaniem jest lokalizacja a następnie wyznaczenie szerokości ust (współrzędnych kąćków ust).

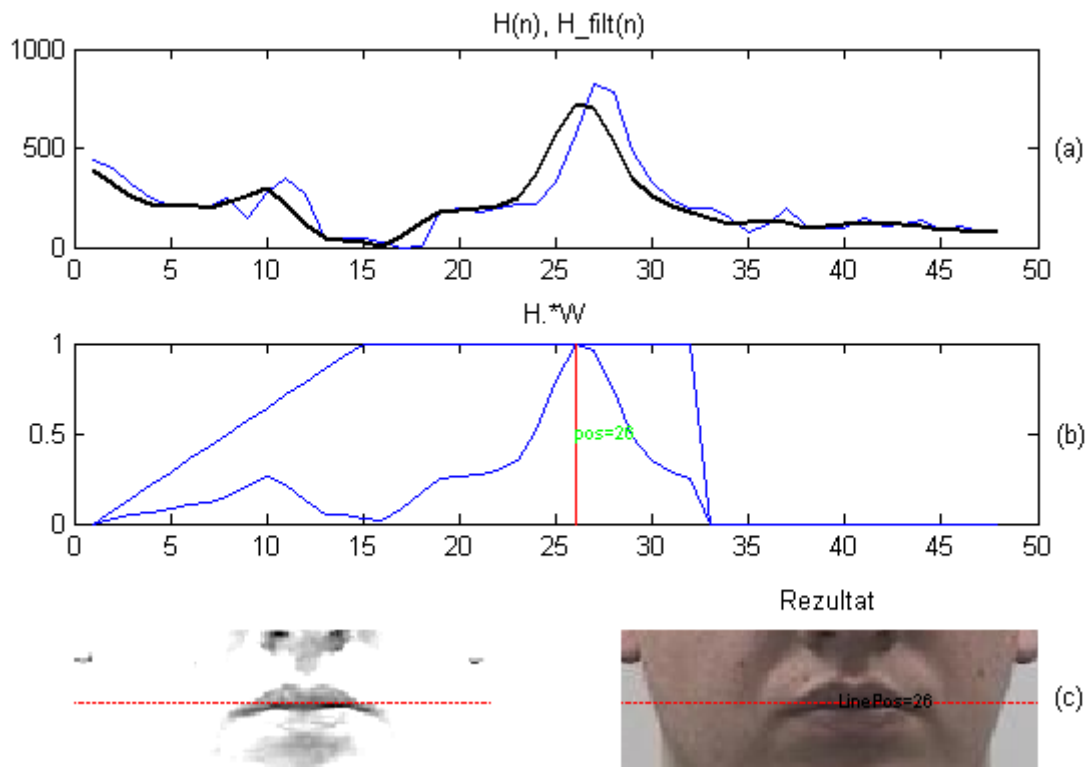
Procedura detekcji ust składa się z kilku etapów:

- wyznaczenia poziomej linii ust,
- określenia regionu poszukiwań ust,
- modyfikacji obrazu w celu wzmocnienia charakterystycznych cech ust,
- wyznaczenia współrzędnych kąćków ust (szerokości ust).

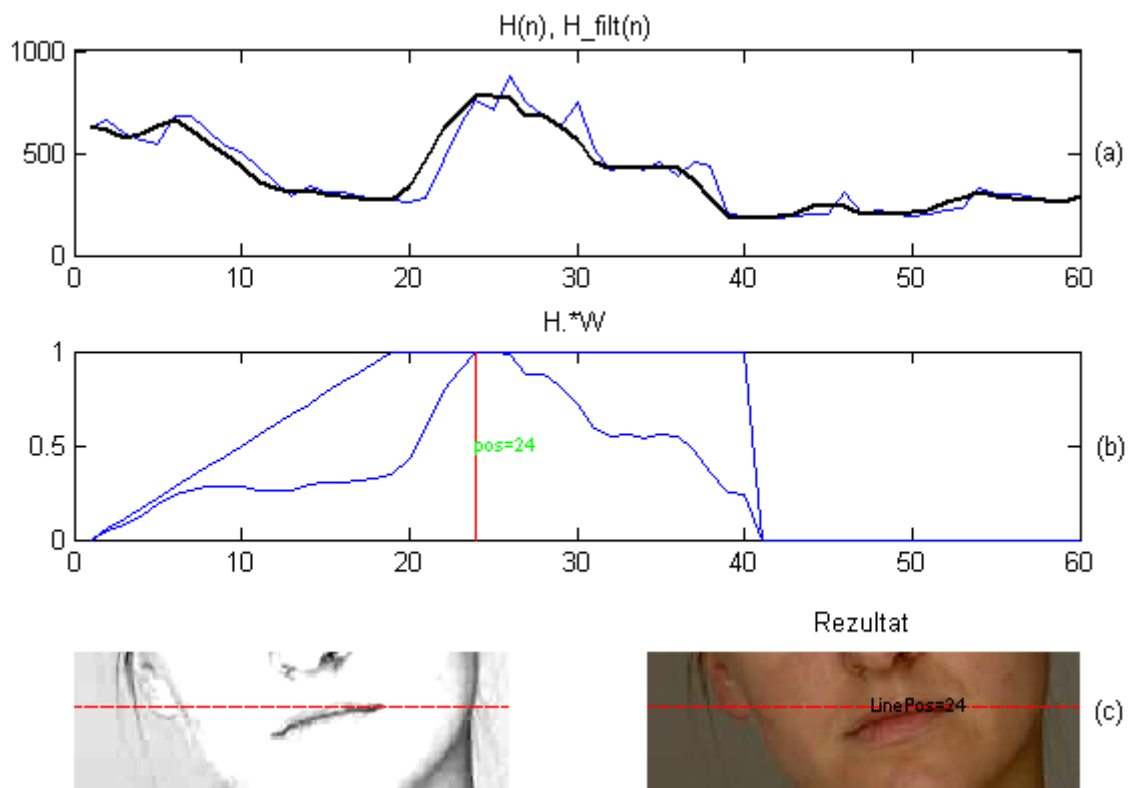
Wymagania dotyczące tego czy analizowany obszar można uznać za usta czy też nie sformułowano na podstawie posiadanej wiedzy o położeniu oczu oraz analizie przebiegu funkcji projekcji wertykalnej obszaru ust. Przyjmuje się, że muszą być spełnione następujące warunki:

- a) usta położone są poniżej zlokalizowanej linii nozdrzy,
- b) na wykresie funkcji projekcji wertykalnej obszaru ust, zaobserwować można wyraźne lokalne minimum odpowiadające środkowej części ust,
- c) stosunek szerokości ust do szerokości twarzy mieści się w ustalonych granicach.

Na podstawie wyznaczonych współrzędnych środków źrenic oczu definiowany jest region twarzy, w ramach którego poszukiwana będzie pozioma linia ust. Granice regionu wyznaczone są od góry poziomą linią nosa, od dołu linią podbródka. Do wyznaczenia poziomej linii ust, podobnie jak miało to miejsce w przypadku lokalizacji oczu, wykorzystano metodę projekcji horyzontalnej (rys. 2.33). Zmianie uległ kształt funkcji okna W stosownie do charakteru analizowanego obszaru.



Rys. 2.33. Projekcja horyzontalna obszaru ust. Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W , wynik końcowy (c)

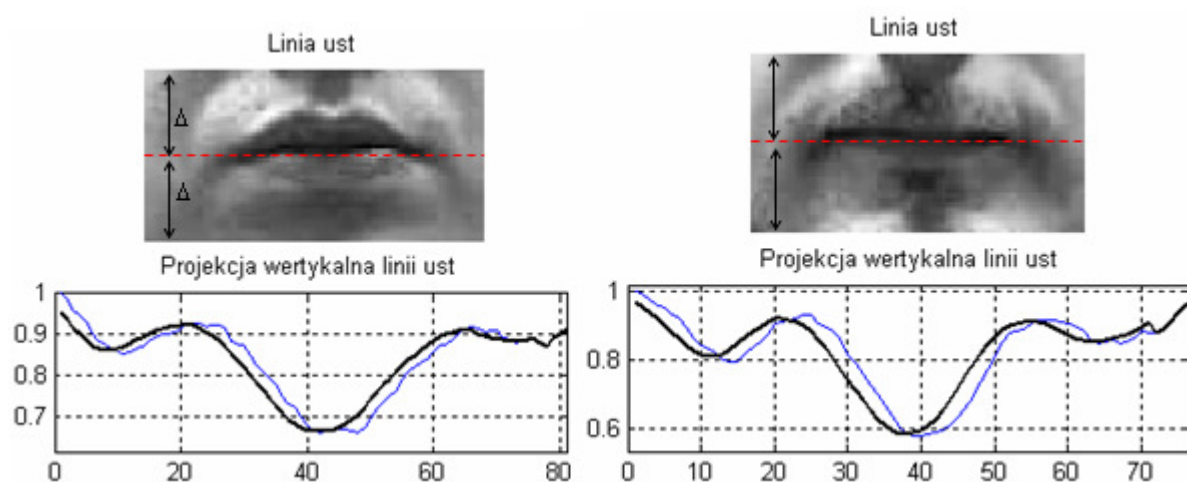


Rys. 2.34. Projekcja horyzontalna obszaru ust twarzy pochylonej (15st). Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W , wynik końcowy (c)

Wynik lokalizacji nie zależy od nachylenia twarzy na płaszczyźnie XY (rys. 2.34) pod warunkiem, że mieści się ono w ustalonych granicach (+/- 30 stopni).

Po ustaleniu pozycji poziomej linii ust, definiowany jest właściwy obszar poszukiwań ust rozciągający się symetrycznie wzdłuż wyznaczonej linii o wysokości równej 2Δ i szerokości zależnej od rozstawu oczu (rys. 2.35). Dla tak wyznaczonego obszaru wyliczana jest projekcja wertykalna. Funkcja ta określa sumę wszystkich wierszy obszaru:

$$V(x) = \sum_{y=y_1}^{y_2} I(x, y), \quad \forall x = x_1, x_1+1, \dots, x_2 \quad (37)$$

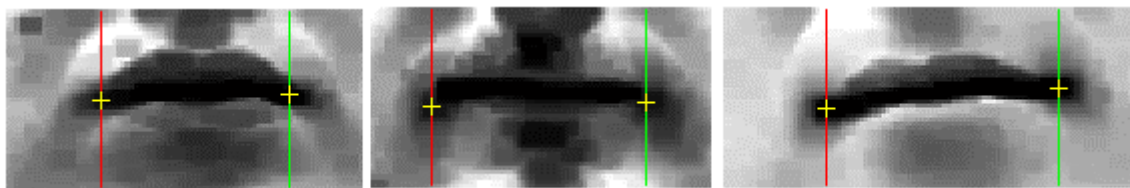


Rys. 2.35. Wyznaczone regiony ust i funkcje projekcji wertykalnej dla wybranych twarzy

Jeśli warunek b) jest spełniony, wyliczana jest przybliżona pozycja środka ust. Następnie wyznaczony obszar ust modyfikowany jest w celu wzmocnienia ciemnych regionów odpowiadających kolorowi ust. W pierwszym kroku obraz jest przekształcany morfologicznie z wykorzystaniem operacji erozji. Elementem strukturalnym jest prostokąt (5x3) zorientowany liniowo w tym samym kierunku co linia ust, co pozwala uwypuklić obszar ust. Następnie wzmocniany jest kontrast poprzez zastosowanie następującego przekształcenia:

$$p(x, y) = \begin{cases} \min & \text{gdy } I(x, y) - \min < \max - I(x, y) \\ \max & \text{w przeciwnym wypadku} \end{cases} \quad (38)$$

gdzie: \min i \max oznaczają odpowiednio minimalną i maksymalną wartość poziomu szarości sąsiedztwa analizowanego piksela $I(x, y)$. Granice sąsiedztwa wyznaczane są prostokątem o wymiarach (5x3).



Rys. 2.36. Rezultat modyfikacji regionu ust i wyznaczona szerokość ust

W wyniku tych modyfikacji powstaje obraz, na którym obszar ust wyraźnie odcina się od pozostałej jego części (rys. 2.36). Na jego podstawie tworzony jest drogą segmentacji nowy obraz, zawierający tylko obszar ust. Dla tego obszaru wyliczane są współrzędne lewego i prawego rogu odpowiadające kąćkom ust. Po transformacji współrzędnych lokalnych do współrzędnych całego obszaru twarzy procedura kończy działanie. Przeprowadzone doświadczenia wykazały dużą skuteczność opisaną metodą, niezależnie od warunków w jakich pozyskiwano analizowane zdjęcia, rasy przedstawianej osoby czy wieku.

2.4.4. Detekcja nosa

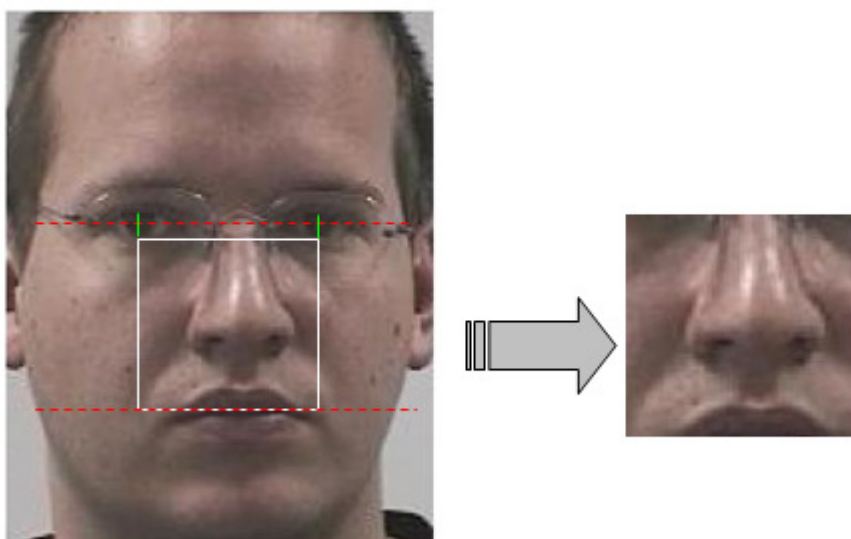
W zadaniu rozpoznawania twarzy nos nie ma tak istotnego znaczenia jak np. oczy, ale ustalenie wartości pewnych jego parametrów może stanowić cenną informację w całym procesie, zwiększając jego skuteczność. Parametrami tymi mogą być: szerokość nosa, współrzędne czubka nosa, położenie otworów nosowych lub opis konturowy nosa. Wybór parametrów uzależniony jest od specyfiki systemu i przewidywanej jakości analizowanych obrazów. Do wyznaczenia w/w parametrów stosuje się techniki oparte na analizie wartości pikseli poddanego odpowiednim przekształceniom obrazu wejściowego (metody geometryczne, „*template matching*”) lub analizie gradientowej wykorzystując metody konturowe, wykrywania narożników, projekcji obszarów etc. W przypadku projektowanego systemu nos opisywany jest za pomocą dwóch parametrów: szerokości i współrzędnych czubka nosa. Wybór szerokości jako parametru pozwala na uniezależnienie się od widoczności otworów nosowych, których lokalizacja w zależności od ustawienia twarzy może być utrudniona lub niemożliwa. Położenie czubka nosa charakteryzuje się dużą stabilnością i w związku z tym może zostać wykorzystane np. do normalizacji obrazów.

Algorytm detekcji nosa składa się z kilku etapów:

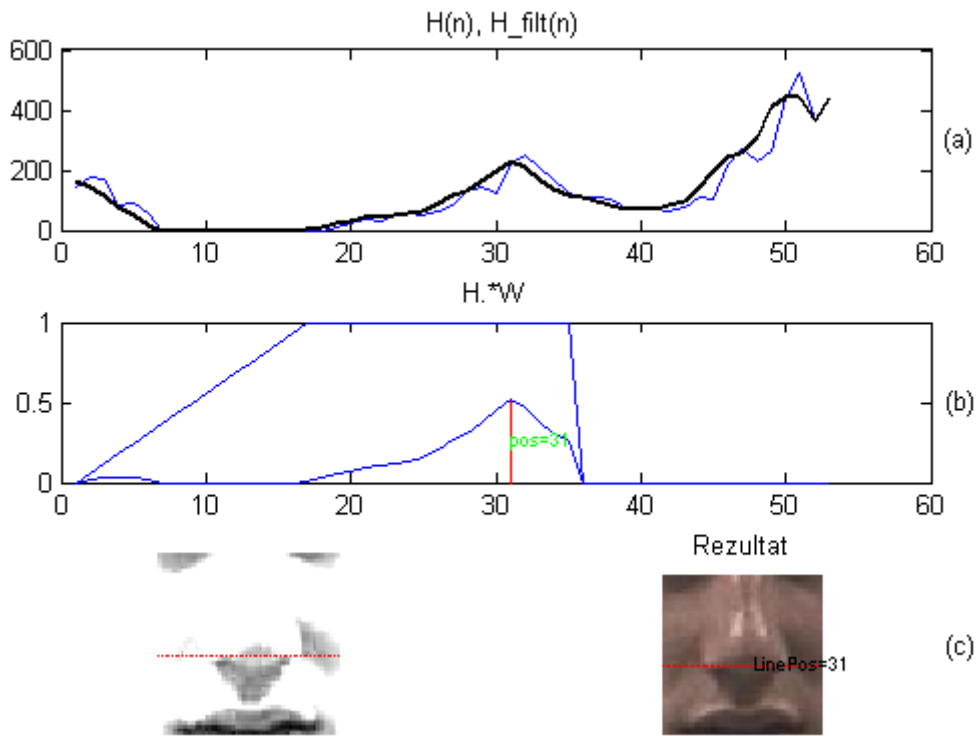
- wyznaczenia poziomej linii nozdrzy,
- określenia regionu sąsiedztwa nosa,
- detekcji krawędzi nosa,
- wyznaczenia szerokości nosa z wykorzystaniem projekcji wertykalnej,

- wyznaczenia współrzędnych czubka nosa.

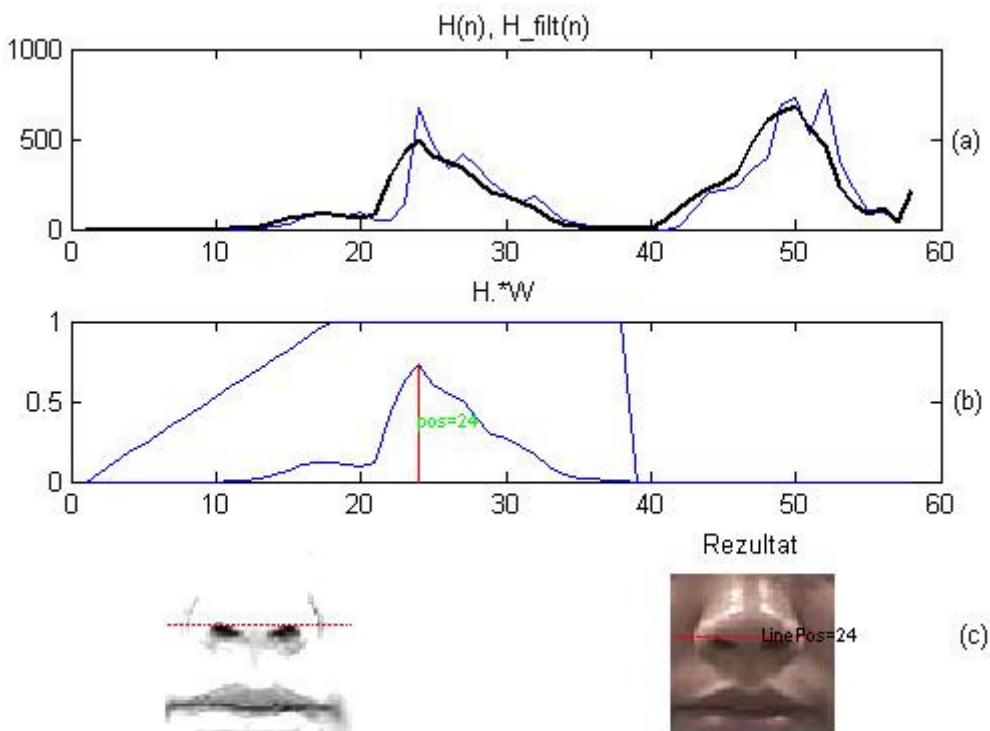
Dysponując informacjami na temat położenia oczu i położeniu ust, wyznaczono region poszukiwań linii nozdrzy. Obszar ten ograniczony jest od góry linią oczu i współrzędnymi środków źrenic oczu a od dołu linią ust (rys. 2.37). Do wyznaczenia linii nozdrzy wykorzystano metodę projekcji horyzontalnej. Kształt funkcji okna W pozostał niezmienny w stosunku do detekcji ust. Następnie wyliczane jest przybliżone położenie środka nosa jako punktu przecięcia się poziomej linii nozdrzy i symetrycznej odcinka łączącego środki źrenic oczu.



Rys. 2.37. Region poszukiwań linii nozdrzy; od góry ograniczony linią oczu i współrzędnymi środków źrenic, od dołu linią ust



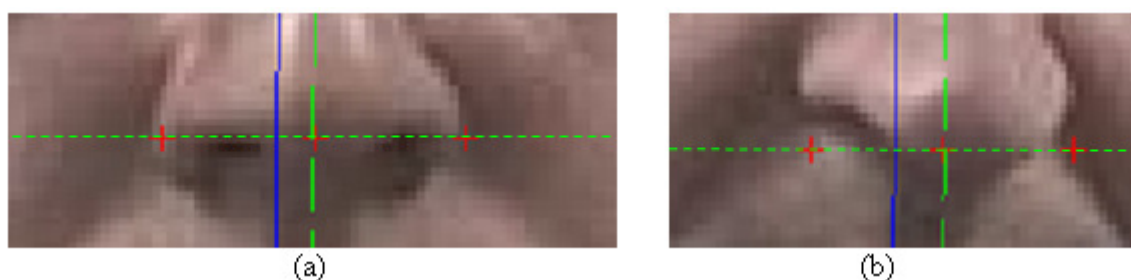
Rys. 2.38. Wyznaczanie linii nozdrzy. Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W , wynik końcowy (c)



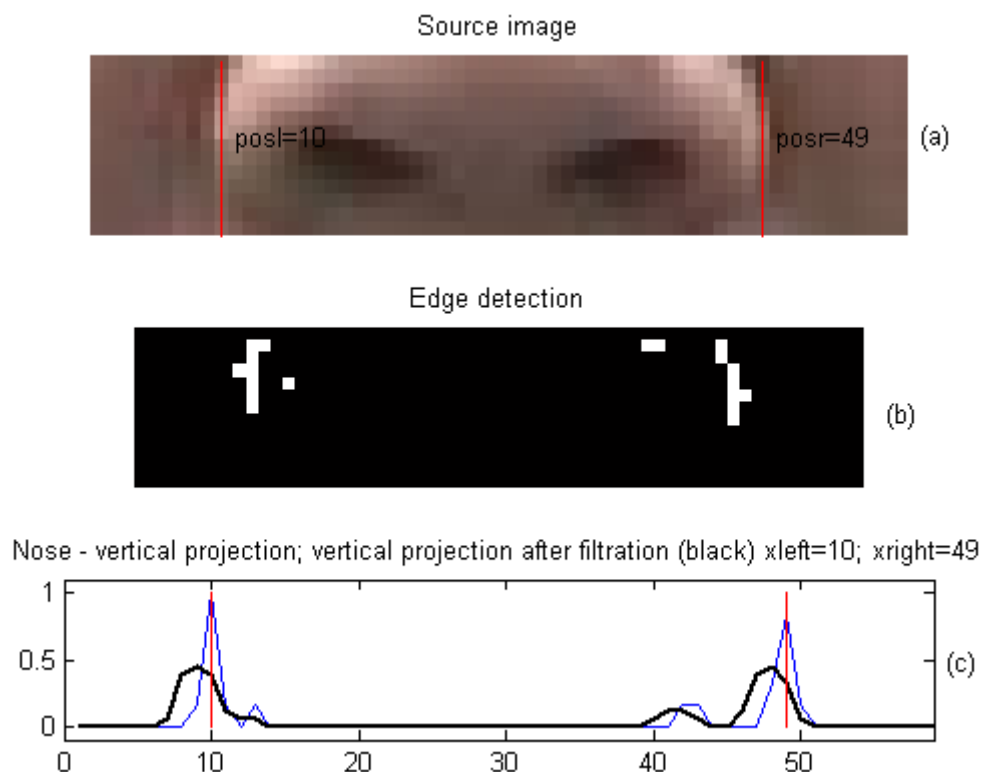
Rys. 2.39. Wyznaczanie linii nozdrzy dla innej twarzy

Po wyznaczeniu położenia linii nozdrzy definiowany jest nowy obszar, rozciągający się symetrycznie wzdłuż niej obejmujący koniec nosa – nozdrza oraz czubek nosa. Granice z lewej i prawej strony wyznacza rozstaw źrenic oczu. Po normalizacji obszaru obejmującej

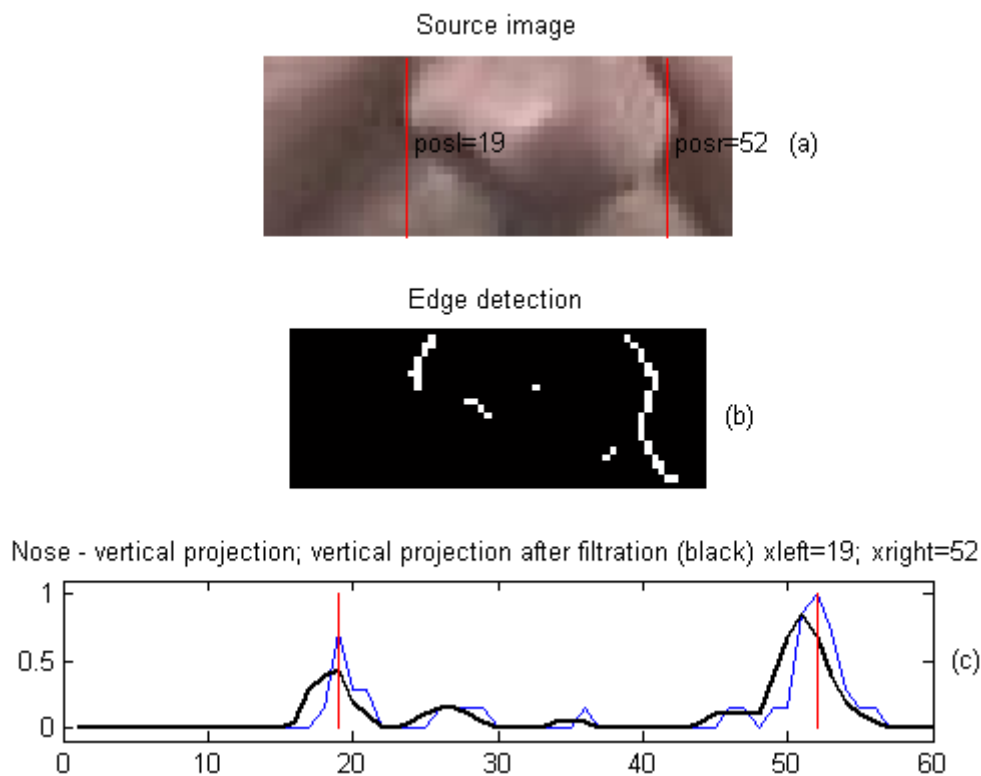
wyrównanie histogramu, filtrację zakłóceń przeprowadzana jest wertykalna detekcja krawędzi. Dla tak zmodyfikowanego obszaru wyliczana jest funkcja projekcji wertykalnej (rys. 2.38c). Położenie maksimum funkcji po lewej i prawej stronie względem wyznaczonej wcześniej przybliżonej współrzędnej czubka nosa wyznacza granice nosa na linii nozdrzy. Na tej podstawie wyliczane są nowe współrzędne czubka nosa – jako punktu będącego środkiem odcinka utworzonego z punktów stanowiących lewą i prawą granicę nosa (rys. 2.40).



Rys. 2.40. Wyznaczone parametry nosa. Linia niebieska to symetryczna odcinka łączącego środki źrenic, która w punkcie przecięcia się z linią nozdrzy wyznacza przybliżone położenie czubka nosa; (b) – przykład działania procedury w przypadku niewidocznych otworów nosowych i nietypowego kształtu nosa



Rys. 2.41. Wyznaczanie szerokości nosa. Rezultat końcowy (a), wynik wertykalnej detekcji krawędzi (b) i przebieg funkcji projekcji wertykalnej obszaru z zaznaczonymi dwoma maksimum wyznaczającymi granice nosa.

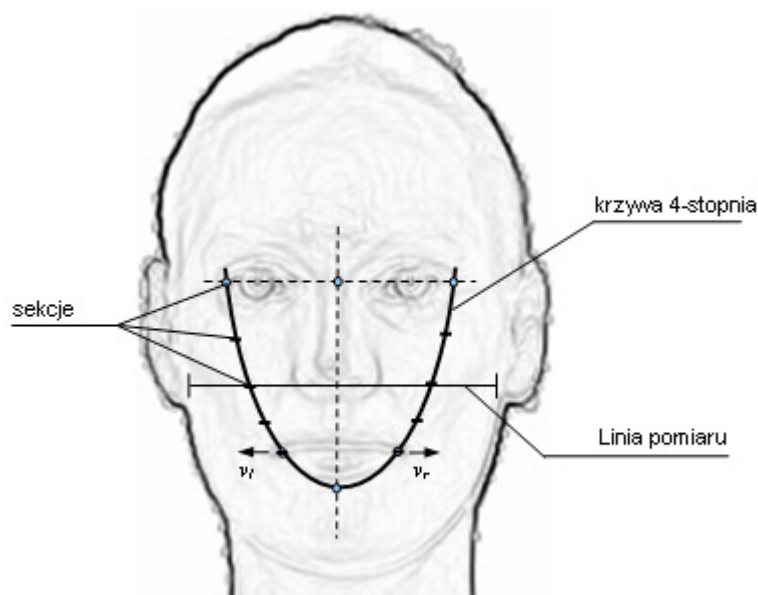


Rys. 2.42. Rezultat działania procedury w przypadku niewidocznych otworów nosowych

2.4.5. Detekcja konturów twarzy. Wyznaczanie szerokości twarzy

Ze względu na trójwymiarową naturę ludzkiej twarzy, poprawne wyznaczenie jej konturu jest złożonym procesem. Do tego celu wykorzystuje się posiadane informacje na temat lokalizacji pozostałych cech: oczu, nosa, ust oraz wiedzy na temat budowy anatomicznej twarzy. W najprostszym przypadku owal twarzy można opisać w przybliżeniu parabolą. W celu lepszego odwzorowania kształtu konieczne jednak staje się wykorzystanie krzywych wyższego stopnia. Detekcja owalu opiera się zazwyczaj na analizie mapy krawędzi obszaru twarzy. Procedura detekcji musi być zatem tak zaprojektowana, aby radzić sobie z często występującymi na niej zakłóceniami, jak i niekompletnym obrazem konturu twarzy, co ma często miejsce w praktyce. Ponadto powinna ona także działać w pewnym zakresie zmian orientacji twarzy.

Proponowana metoda opiera się na wykorzystaniu deformowalnego szablonu reprezentowanego przez krzywą czwartego stopnia (rys. 2.43). Proces deformacji modelu został tak zaprojektowany, aby detekcja owalu była możliwa także w przypadku obecności zakłóceń na mapie krawędzi obszaru, jak i niekompletności konturu twarzy. Brakujące fragmenty są uzupełniane na drodze estymacji parametrów krzywej zbudowanej na bazie dostępnych danych.



Rys. 2.43. Szablon defomowalny do wyznaczenia szerokości twarzy

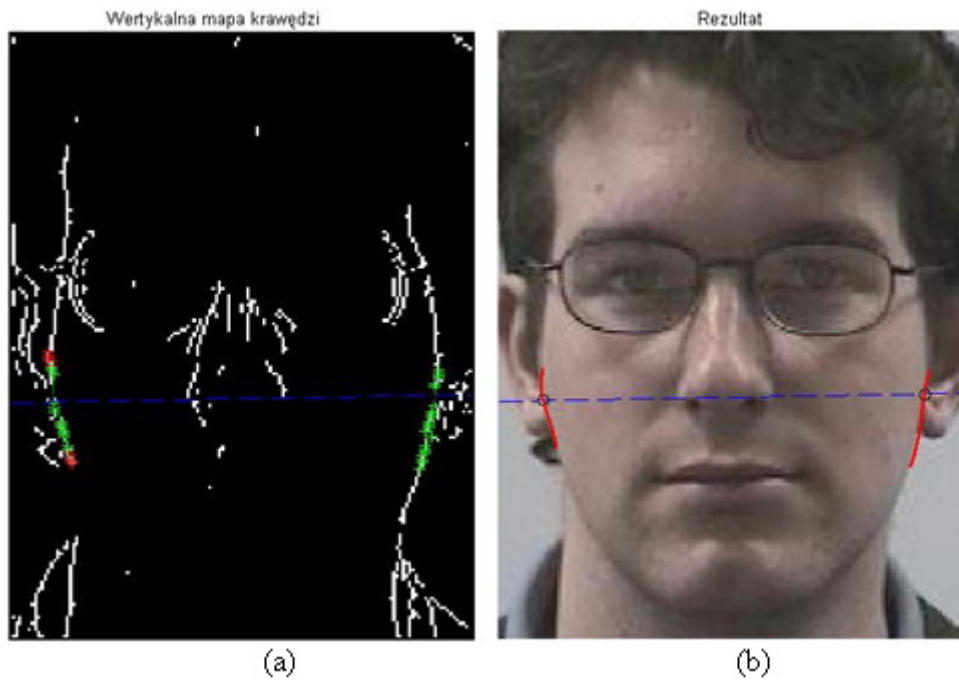
Szukany parametrem jest szerokość twarzy. Wybór ten podyktowany został rezultatami przeprowadzonych doświadczeń. Analiza map krawędzi badanych twarzy wykazała, że w przeważającej większości przypadków krawędzie dolnej części twarzy (szczególnie podbródka) były zupełnie niewidoczne. W tej sytuacji wyznaczenie całej linii owalu stałoby się utrudnione - wymagałoby opierania się wyłącznie na proporcjach ludzkiej twarzy.

Parametry startowe modelu wyliczane są na podstawie położenia zlokalizowanych do tej pory cech. Po inicjalizacji rozpoczyna się ekspansja modelu w kierunku granic twarzy zgodnie z wyznaczonymi wcześniej wektorami kierunkowymi. Ponieważ celem jest wyznaczenie szerokości twarzy, ekspansja odbywa się tylko w kierunku granic policzków (rys. 2.43), niezależnie od siebie. Pomiar szerokości twarzy wykonywany jest na linii nozdrzy. Proces trwa do momentu, gdy wykryta zostanie obecność krawędzi kolinearnych z krzywą modelu. Krzywa modelu podzielona jest na kilka sekcji. Przemieszczenie pojedynczej sekcji odbywa się do momentu, kiedy zostanie wykryta wystarczająca liczba pikseli kolinearnych z aktualną sekcją krzywej, przy czym kolinearność w tym wypadku oznacza, że dany piksel należy do krzywej lub znajduje się w bliskim jej sąsiedztwie. Po ustaleniu przemieszczeń wszystkich sekcji, wyznaczane są parametry dwóch nowych krzywych: pierwszej opisującej lewy kontur twarzy i drugiej opisującej prawą część.

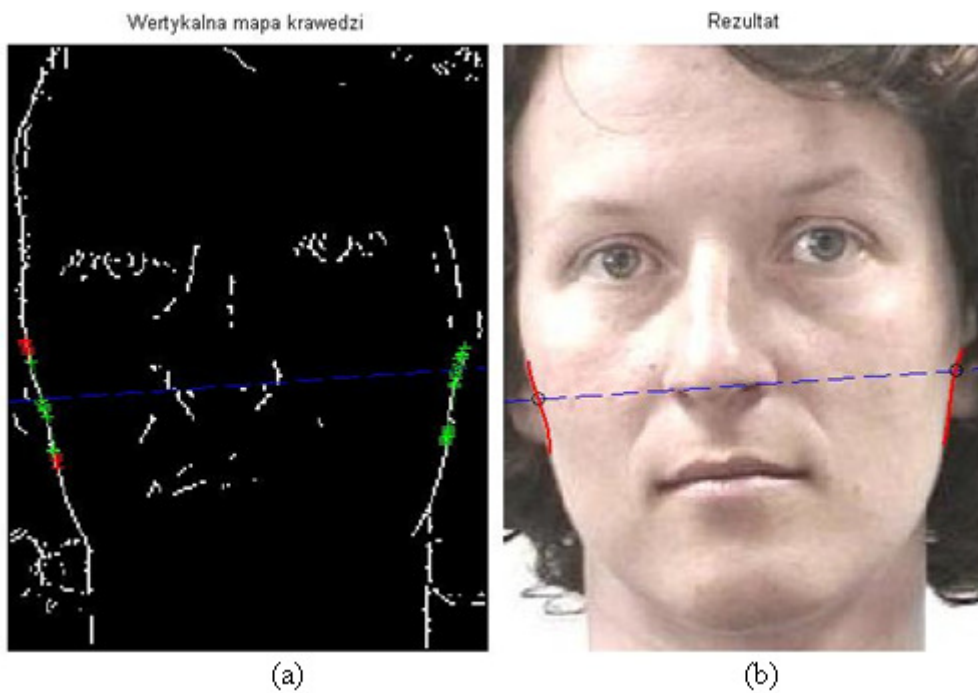


Rys. 2.44. Twarze z naniesionymi pozycjami startowymi modelu

Parametry krzywych wyliczane są z wykorzystaniem metody najmniejszych kwadratów. Przed ich wyliczeniem wśród punktów tworzących krawędzie twarzy przeprowadzana jest selekcja, mająca na celu eliminację tych, które mogą stanowić źródło potencjalnych zakłóceń. Dla całego zbioru punktów wyliczane jest odchylenie standardowe i średnia wartość względem osi OX . Punkty, których położenie odbiega od średniej więcej niż wyliczone odchylenie, są eliminowane. Po ustaleniu równań krzywych wykonywany jest pomiar szerokości twarzy na poziomie linii nozdrzy (długość wektora zbudowanego z punktów przecięcia się wyznaczonych krzywych z linią pomiaru).



Rys. 2.45. Wertykalna mapa krawędzi twarzy (a) z naniesionymi punktami zlokalizowanej krawędzi i rezultat pomiaru szerokości (b)



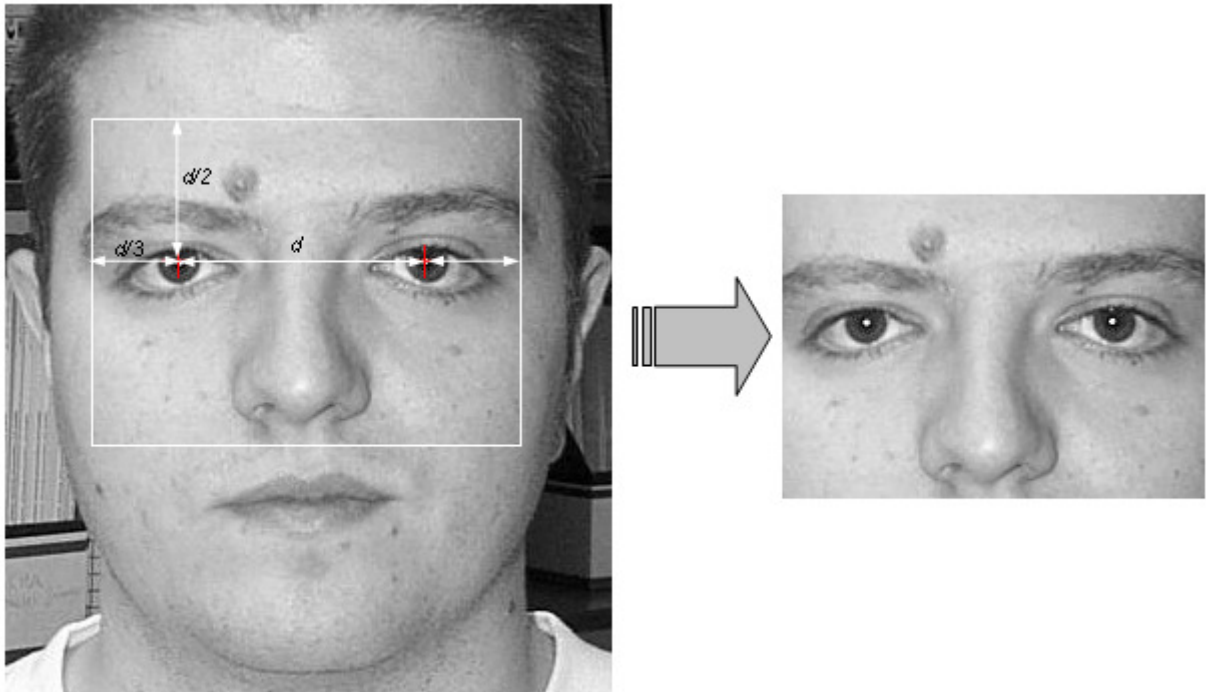
Rys. 2.46. Wertykalna mapa krawędzi twarzy (a) z naniesionymi punktami zlokalizowanej krawędzi i rezultat pomiaru szerokości (b) dla innej twarzy

W przypadku, gdy wyznaczenie parametrów krzywej staje się niemożliwe (np. w przypadku zupełnego braku konturu) szerokość twarzy nie jest włączana do wektora cech.

2.4.6. Wyznaczanie środkowej części twarzy

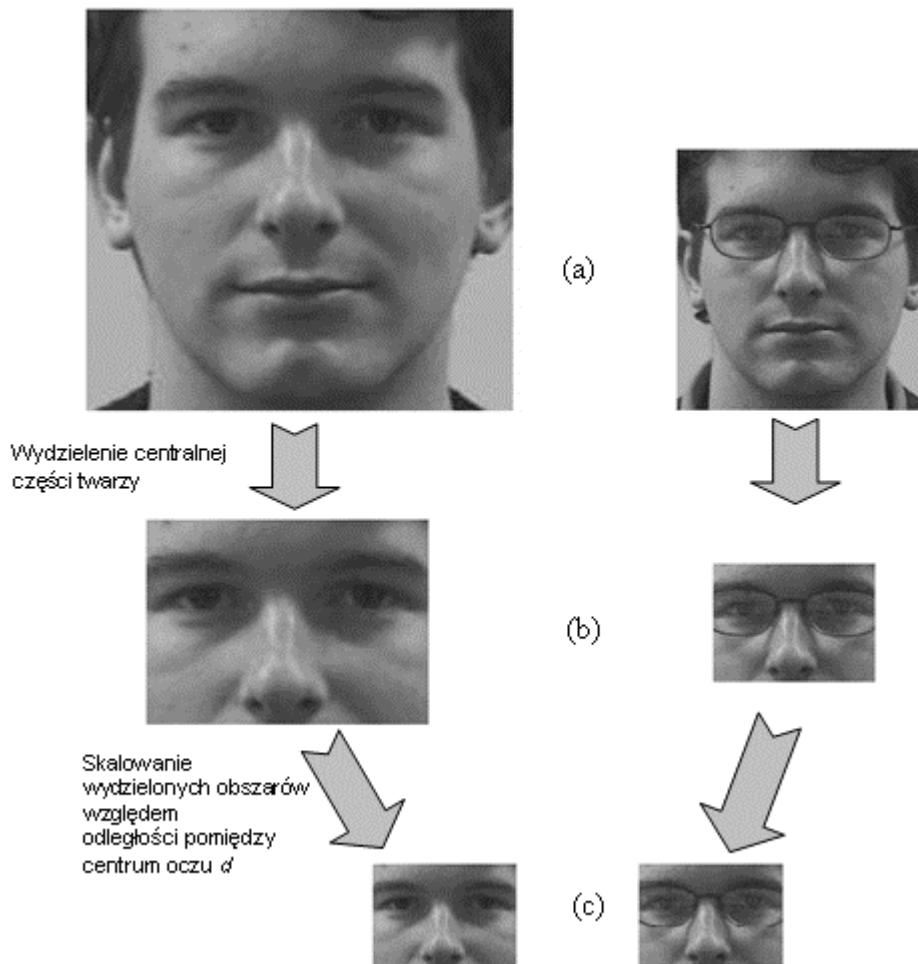
W proponowanym systemie weryfikacja rezultatów identyfikacji odbywa się poprzez badanie stopnia podobieństwa obrazów z wykorzystaniem fazowej wzajemnej korelacji. Aby rezultaty weryfikacji były jak najbardziej wiarygodne, należy zadbać o to, aby porównywane obrazy charakteryzowały się podobnymi właściwościami jeśli chodzi o rozmiar, kąt obrotu i zakres niesionych informacji. Jak zostało to już opisane, najbardziej informacyjnym i niezmiennym obszarem twarzy jest jej środkowa część (brwi, oczy, nos). Wyłączone z tego obszaru są usta, ze względu na możliwe zmiany ich kształtu (usta otwarte, zamknięte, uśmiech itd.) i obecność zarostu, co mogłoby mieć niekorzystny wpływ na przebieg procesu. Warunkiem skutecznie przeprowadzonej weryfikacji jest zatem poprawne wyznaczenie granic tego obszaru i normalizacja jego parametrów.

Granice obszaru środkowej części twarzy można wyznaczyć w oparciu o posiadaną wiedzę na temat odległości pomiędzy środkami oczu. Znając odległość d pomiędzy centrum oczu na twarzy, obszar centralnej części twarzy (brwi, oczy, nos) ma rozmiar $\frac{5d}{4} \times \frac{5d}{3}$ pikseli. Granica obszaru centralnej części twarzy z prawej oraz lewej strony jest przesunięta na odległość $d/3$ od centrum oczu. Linia oczu jest rozmieszczona poniżej górnej granicy obszaru centralnej części twarzy na odległość $d/2$ pikseli.



Rys. 2.47. Zasada wyznaczania centralnej części twarzy

Następnie wydzielony obszar jest normalizowany. Podstawowym zadaniem tej operacji jest dopasowanie rozmiarów wydzielonego obszaru do rozmiarów przechowywanych w bazie wzorców. Samo zadanie zmiany rozmiarów nie stwarza żadnych trudności. Problem polega na tym, czy obraz należy pomniejszyć czy powiększyć, a także jak bardzo trzeba zmienić jego rozmiar. Proponowane rozwiązanie opiera się na wykorzystaniu jako parametru, względem którego obraz będzie skalowany, odległości między środkami oczu d . Początkowa wartość d dana jest na podstawie przeprowadzonej lokalizacji oczu. Wydzielony obszar skalowany jest na podstawie stosunku bieżącej odległości d do ustalonej wartości wzorcowej. Ponieważ sposób wyznaczania centralnej części twarzy jest zawsze taki sam, niezależnie od wielkości twarzy, zmiana skali obrazu wg opisaney metody zapewnia dostateczne dopasowanie obrazu do przechowywanych w bazie wzorców.



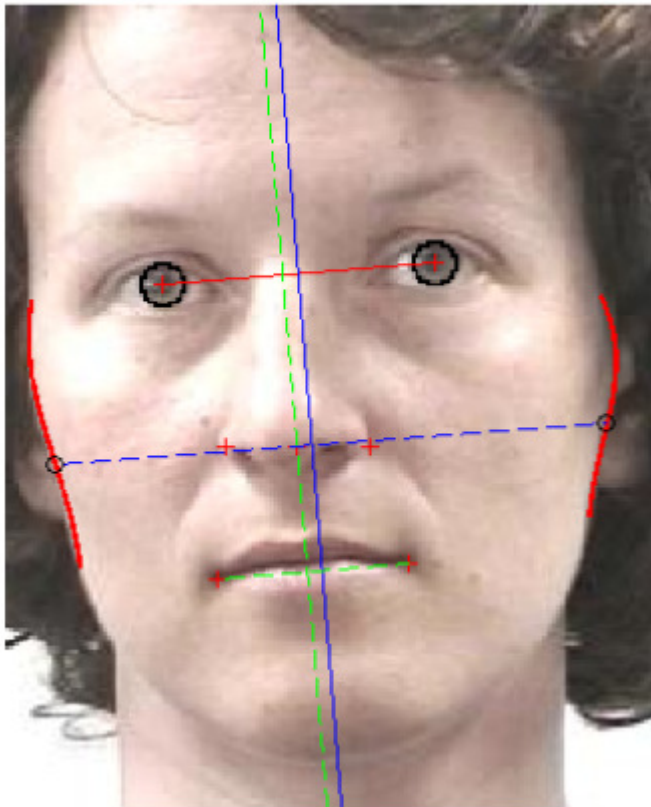
Rys. 2.48. Normalizacja skali wydzielonego obszaru centralnej części twarzy; (a) różnej wielkości wejściowe obrazy twarzy; (b) wydzielone centralne obszary twarzy; (c) obszary wydzielone po normalizacji skali

2.4.7. Budowa wektora cech

W oparciu o współrzędne zlokalizowanych cech twarzy (rys. 2.49) budowany jest wektor cech opisujący analizowaną twarz. Elementy wektora stanowią znormalizowane wartości stosunków geometrycznych i antropometrycznych cech twarzy. Cechy te to (rys. 2.50):

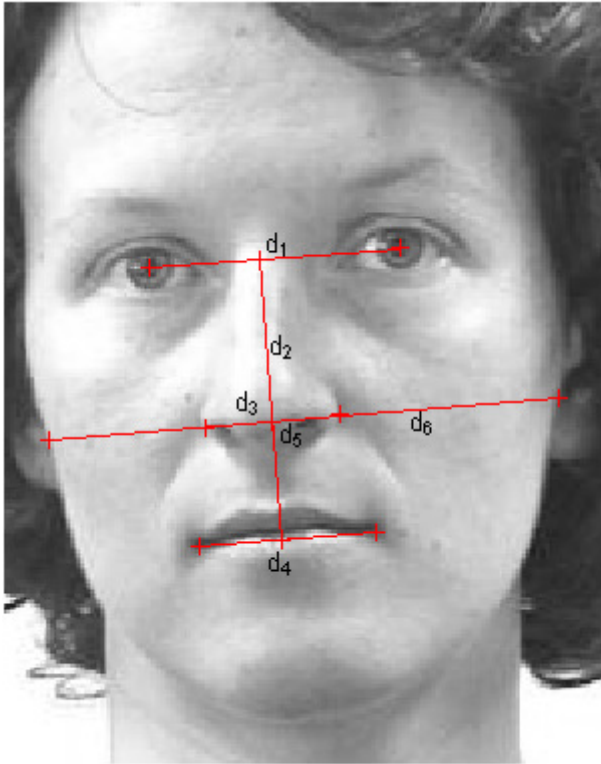
- odległość między środkami oczu d_1 ,
- odległość pomiędzy oczami i nosem d_2 ,
- odległość pomiędzy linią oczu i linią ust d_3 ,
- szerokość ust d_4 ,
- szerokość nosa d_5 ,
- szerokość twarzy d_6 .

Ekstrakcja cech - rezultat



1.	LeftEyeCenter	[44; 80]
2.	RightEyeCenter	[119; 74]
3.	CenterBEyes	[81.5; 77]
4.	CenterEyesNose	[77.2; 77.34]
5.	LeftNoseCorner	[61.06; 126]
6.	RightNoseCorner	[101; 126]
7.	CenterNose	[81.05; 126]
8.	LeftMouthCorner	[59.06; 163.5]
9.	RightMouthCorner	[112.06; 159]
10.	CenterMouth	[85.56; 161.25]
11.	CenterMouthNose	[83.89; 161.38]
12.	LeftCheekEdge	[14.38; 131.33]
13.	RightCheekEdge	[166.57; 119.15]

Rys. 2.49. Rezultat lokalizacji cech twarzy



Rys. 2.50. Składowe wynikowego wektora cech

Wektor cech c przyjmuje postać:

$$c = [c_1 \quad c_2 \quad c_3 \quad c_4 \quad c_5 \quad c_6] \quad (39)$$

$$\text{gdzie: } c_1 = d_1 / d_2; \quad c_2 = d_1 / d_3; \quad c_3 = d_1 / d_4; \quad c_4 = d_1 / d_6; \quad c_5 = d_4 / d_5; \\ c_6 = d_1 / d_5;$$

2.5. Struktura bazy danych

2.5.1. Wprowadzenie

Proces budowy bazy danych jest ważnym etapem budowy każdego systemu rozpoznawania twarzy. Dobrze zaprojektowana struktura umożliwia zwiększenie skuteczności działania systemu rozpoznawania. Istotną rzeczą jest zapewnienie odpowiedniej jakości przechowywanych w bazie danych wzorców. Należy zadbać o to, aby w trakcie uczenia systemu, zapisywane dane charakteryzowały się odpowiednią wiarygodnością i dokładnością. Na poprawę skuteczności ma także wpływ ilość zgromadzonych danych – systemy, które w ramach jednej klasy (opisującej jedną osobę) zawierają wiele wzorców pobranych w różnych warunkach (różna skala, oświetlenie, rotacja) wykazują zazwyczaj większą skuteczność.

Bazę danych każdego systemu rozpoznawania możemy scharakteryzować poprzez:

- reprezentatywność przechowywanych danych,
- sposób strukturyzacji danych,
- format przechowywanych danych, w tym:
 - format graficznego przedstawienia obrazu;
 - rozmiar każdego obrazu;
 - format koloru;
 - wymiar i postać wektora cech;
- przeznaczenie.

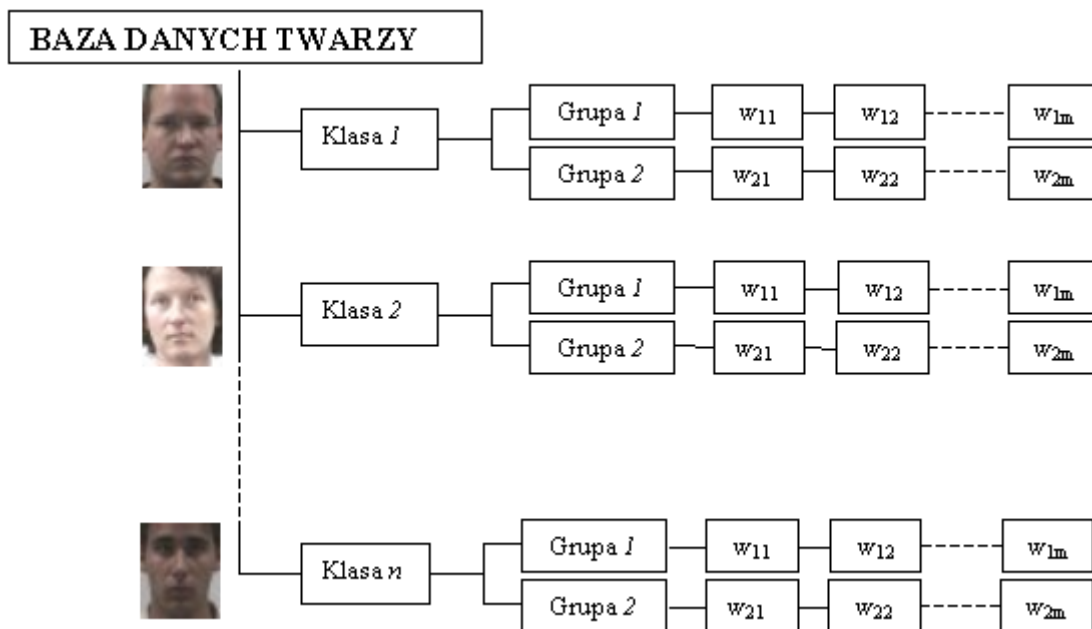
Reprezentatywność bazy danych określa pożądaną ilość obrazów (ogólnie: wzorców) odpowiadających różnym orientacjom twarzy, obecności lub nieobecności dodatkowych przedmiotów (okulary, kolczyki itp.), a także mimiki twarzy. Ponadto, istotnym jest, aby warunki otrzymania przechowywanych wzorców były zbliżone (oświetlenie, nieznaczne wiekowe zmiany twarzy przedstawiane w każdej klasie). Z kolei sposób strukturyzacji bazy wyrażony jest liczbą klas, liczbą wzorców w każdej klasie oraz organizacją bazy ze względu na format i rodzaj wykonywanych operacji [12]. Każda baza danych ma indywidualne przeznaczenie. Jako przykład może posłużyć baza danych zbudowana w laboratorium *LIM Ecole Centrale de Lyon*. Obrazy wchodzące w jej skład przedstawiają twarze, których rozmiar i orientacja nie są kontrolowane a twarze nie są wycentrowane. Baza przeznaczona jest głównie do testowania procedur lokalizacji twarzy na obrazach kolorowych, detekcji położenia oczu oraz testowania systemów klasy „*Visitor Identification*”. Baza zawiera obrazy osób różnych ras i płci: Europejczyków, Chińczyków, Arabów, Afrykańczyków. W zastosowaniach identyfikacji twarzy dużą popularność zdobyła baza *ORL Database of Faces*. W jej skład wchodzi 400 obrazów twarzy zgrupowanych w 40 klas po 10 obrazów w każdej klasie. Obrazy przedstawione są w 256 poziomach szarości, na ciemnym tle. Fotografie obrazów twarzy wykonywane są w różnym czasie (rozpiętość sięga 2 lat). Obrazy w ramach klasy różnią się skalą, kątem patrzenia, warunkami oświetlenia, jasności tła – nadają się więc doskonale do testowania systemów rozpoznawania twarzy.

Przystępując do projektowania struktury bazy danych systemu, należy zdefiniować format przechowywanych danych – wzorców. Związany jest on ściśle z wybraną metodologią pozyskiwania cech i zastosowanym algorytmem identyfikacji. Odnosząc to do projektowanego systemu dane w ramach pojedynczej klasy podzielone zostały na dwie grupy:

- grupa I - w której skład wchodzi wektory cech (wymiar $N = 6$) powstałe w wyniku ekstrakcji charakterystycznych punktów twarzy;

- grupa II – zawierająca wektory utworzone poprzez konkatencję wierszy znormalizowanego obrazu środkowej części twarzy (N = 256)

Każda grupa obrazów poddawana jest innej ścieżce przetwarzania w zależności od wybranego algorytmu identyfikacji (rys. 2.51).



Rys. 2.51. Struktura bazy danych systemu

2.5.2. Normalizacja bazy

Elementy przechowywanych w bazie danych wektorów cech charakteryzować mogą się dużą rozpiętością wartości zarówno w ramach pojedynczego wektora jak i względem swych odpowiedników w pozostałych wektorach. Aby zapobiec sytuacjom, kiedy wartość jednej cechy ma decydujący wpływ na wynik porównania, konieczna staje się normalizacja wartości elementów wektorów cech. Polega ona na odpowiednim przeskalowaniu wartości poszczególnych składowych wektora cech, tak aby każda z nich miała równy udział w procesie identyfikacji, sterowany ewentualnie przydzielonymi przez operatora systemu wagami. Wyliczenie wartości odpowiednich współczynników skalujących odbywa się dwuetapowo:

1. Wyznaczenie wartości średnich każdej z cech dla wszystkich obiektów bazy danych:

$$\bar{c}_i = \left(\sum_{j=1}^n db(j,i) \right) / n \quad \forall i = 1, 2, \dots, m \quad (40)$$

2. Wyznaczenie średnich odchyłeń każdej z cech określonych jako:

$$\bar{\sigma}_i = \sqrt{\left(\sum_{j=1}^n (1 - db(j,i)/\bar{c}_i)^2 \right) \setminus n} \quad \forall i = 1, 2, \dots, m \quad (41)$$

przy czym: n – ilość obiektów bazy danych, m – wymiar wektora cech

Wtedy współczynniki skalujące dla i -tej cechy wyliczane są jako:

$$s_i = 10 \cdot \bar{\sigma}_i / \bar{c}_i \quad (42)$$

2.6. Identyfikacja

W zależności od aktualnej ścieżki przetwarzania identyfikacja odbywa się z wykorzystaniem dwóch algorytmów: podstawowego, który polega na porównywaniu wyznaczonego wektora cech opisującego badany obraz z przechowywanymi w bazie wzorcami oraz dodatkowego, wykorzystującego metodę korelacji fazowej do weryfikacji wyniku rozpoznania.

W trybie podstawowym rozpoznawania, wejście stanowi wyznaczony wektor cech c (wymiar $N = 6$). Każdej składowej wektora cech zostaje przyporządkowana waga w_i będąca liczbą z przedziału $[0; 1]$. Określa ona w jakim stopniu poszczególne cechy wpływają na wynik rozpoznania. Identyfikacja odbywa się poprzez wyliczanie odległości pomiędzy badanym wektorem a przechowywanymi w bazie wzorcami z wykorzystaniem metryki euklidesowej zgodnie z:

$$d_j = \sqrt{\sum_{i=1}^N w_i (db(j,i) - c_i s_i)^2}, \quad \forall j = 1, 2, \dots, M \quad (43)$$

gdzie: $i = 1, 2, \dots, N$ – składowe wektora cech, $j = 1, 2, \dots, M$ – j -ty obiekt bazy danych,

w_i – waga przypisana i – tej składowej,

c_i – i -ta składowa wektora cech,

s_i – współczynnik skalujący,

$db(j,i)$ – i -ta składowa j -tego obiektu bazy danych.

Klasyfikacja odbywa się na podstawie zdefiniowanych wcześniej kryteriów:

- a) najbliższy obraz bazowy wskazuje klasę, do której należy badany obraz (1 of 1);
- b) obliczane są średnie obrazy w klasach, najbliższa średnia określa numer klasy, do której należy badany obraz (ang. *average*);
- c) badany obraz należy do klasy, do której należą dwa z trzech najbliższych obrazów w bazie (2 of 3);

- d) badany obraz należy do klasy, do której należą trzy z pięciu najbliższych obrazów w bazie (3 of 5);

Dodatkowe kryteria stanowią wartości progowe:

- T_1 - minimalna wartość podobieństwa, poniżej której identyfikacja kończy się z wynikiem negatywnym,
- T_2 – wartość progowa (procentowo) określająca odległość średniego obrazu najbliższej klasy do średniego obrazu następnej najbliższej klasy.

Możliwe są kombinacje powyższych warunków. Zwiększanie ilości warunków klasyfikacji powoduje zmniejszanie się współczynnika prawdopodobieństwa błędnego zakwalifikowania badanego obrazu do klasy (FAR – ang. *False Acceptance Ratio*) lecz zazwyczaj powoduje także zmniejszenie się wartości współczynnika rozpoznawania ($FArec$). Na przykład jeśli do klasyfikacji wykorzystane zostaną warunki (a) i (b) a wartość progu zostanie określona $T_2 = 1,25$ to badany obraz zostanie przyporządkowany do pewnej klasy, jeżeli najbliższy obraz bazowy i średni obraz najbliższej klasy należą do tej samej klasy oraz odległość do średniego obrazu następnej najbliższej klasy jest większa o 25% od odległości do średniego obrazu najbliższej klasy.

2.6.1. Weryfikacja rezultatu metodą korelacji fazowej

Weryfikację oparto na porównywaniu wyników rozpoznawania uzyskanych za pomocą algorytmu podstawowego z odpowiednimi rezultatami rozpoznawania uzyskanymi w wyniku zastosowania korelacji fazowej do badania stopnia podobieństwa centralnych części twarzy. Procedura weryfikacji uruchamiana jest, kiedy decyzja nie może być podjęta w oparciu o wyniki uzyskane za pomocą algorytmu podstawowego. W przypadku, gdy obie metody dają sprzeczne wyniki identyfikacja uważana jest za negatywną. Do badania stopnia podobieństwa obrazów wykorzystuje się metodę korelacji fazowej. W przetwarzaniu obrazów, korelacja fazowa jest szybką, realizowaną w dziedzinie częstotliwości metodą, mogącą służyć do badania stopnia podobieństwa pomiędzy obrazami a także estymacji względnego przemieszczenia pomiędzy dwoma obrazami. W zastosowaniach rozpoznawania twarzy mamy do czynienia ze specyficznymi obrazami – obrazami twarzy człowieka. Każda twarz składa się z tych samych części – brwi, oczu, ust, nosa, owalu. Indywidualne rysy twarzy (brwi, oczy, usta, nos, owal) a także tło na obrazach w ogólnym przypadku mogą się zmieniać nie na tyle znacznie, aby odczuć je w zmianach zwykłej dwuwymiarowej wzajemnej

korelacji, która jest bardziej czuła dla zmian w niskoczęstotliwościowej części i jest mniej więcej jednakowa dla wszystkich obrazów twarzy. Wady tej nie posiada korelacja fazowa, czuła dla wszystkich zmian indywidualnych rysów twarzy.

Metoda

Oznaczmy porównywane obrazy jako $f(m,n)$ i $g(m,n)$, każdy o rozmiarach $M \times N$. Zakładamy, że indeksy zmieniają się w zakresach odpowiednio $m = -M_2 \dots M_2$ ($M_2 > 0$) i $n = -N_2 \dots N_2$ ($N_2 > 0$), stąd $M = 2M_2 + 1$ i $N = 2N_2 + 1$. Niech $F(k_1, k_2)$ i $G(k_1, k_2)$ oznaczają dwuwymiarowe dyskretne transformacje Fouriera (2D DFT) badanych obrazów i

$$F(k_1, k_2) = \sum_{m,n} f(m,n) W_M^{k_1 m} W_N^{k_2 n} = A_F(k_1, k_2) e^{j\theta_F(k_1, k_2)} \quad (44)$$

oraz

$$G(k_1, k_2) = \sum_{m,n} g(m,n) W_M^{k_1 m} W_N^{k_2 n} = A_G(k_1, k_2) e^{j\theta_G(k_1, k_2)} \quad (45)$$

gdzie: $k_1 = -M_2 \dots M_2$, $k_2 = -N_2 \dots N_2$, $W_M = e^{-j\frac{2\pi}{M}}$, $W_N = e^{-j\frac{2\pi}{N}}$,

$A_F(k_1, k_2)$ i $A_G(k_1, k_2)$ – składowe amplitudowe,

$e^{j\theta_F(k_1, k_2)}$ i $e^{j\theta_G(k_1, k_2)}$ – składowe fazowe.

Wzajemne widmo fazowe $R_{FG}(k_1, k_2)$ pomiędzy $F(k_1, k_2)$ i $G(k_1, k_2)$ definiujemy jako:

$$R_{FG}(k_1, k_2) = \frac{F(k_1, k_2) G^*(k_1, k_2)}{|F(k_1, k_2) G^*(k_1, k_2)|} = e^{j\theta(k_1, k_2)} \quad (46)$$

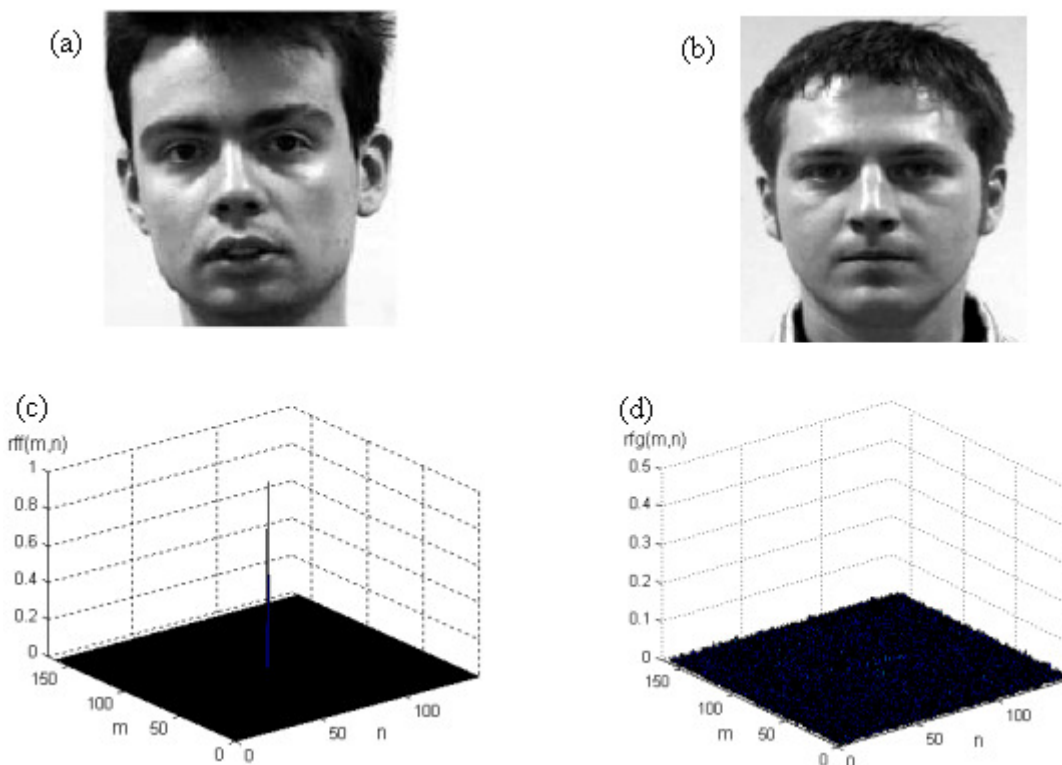
Wtedy funkcja korelacji fazowej r_{fg} jest dwuwymiarową odwrotną dyskretną transformacją Fouriera (2D IDFT) $R_{FG}(k_1, k_2)$ i dana jest jako:

$$r_{fg}(m, n) = \frac{1}{MN} \sum_{k_1, k_2} R_{FG}(k_1, k_2) W_M^{-k_1 m} W_N^{-k_2 n} \quad (47)$$

W przypadku gdy $f(m,n)$ i $g(m,n)$ są tym samym obrazem, tzn. $f(m,n) = g(m,n)$, funkcja korelacji fazowej przyjmuje postać:

$$r_{ff}(m, n) = \frac{1}{MN} \sum_{k_1, k_2} W_M^{-k_1 m} W_N^{-k_2 n} = \delta(m, n) = \begin{cases} 1 & \text{gdy } m = n = 0 \\ 0 & \text{w przeciwnym wypadku} \end{cases} \quad (48)$$

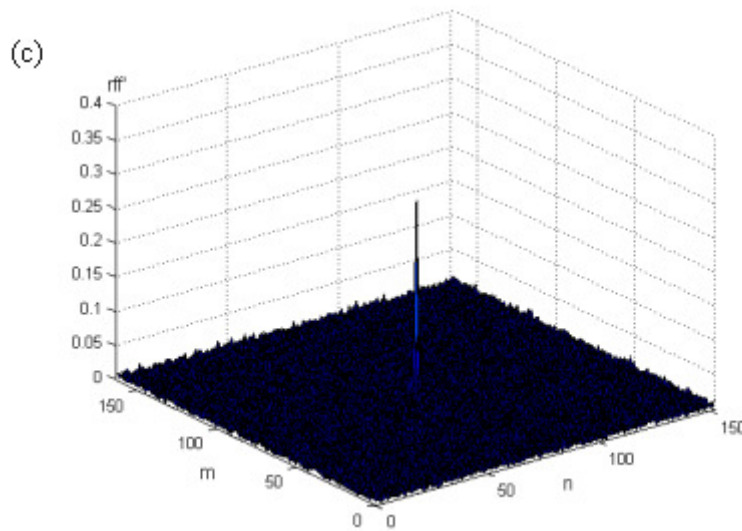
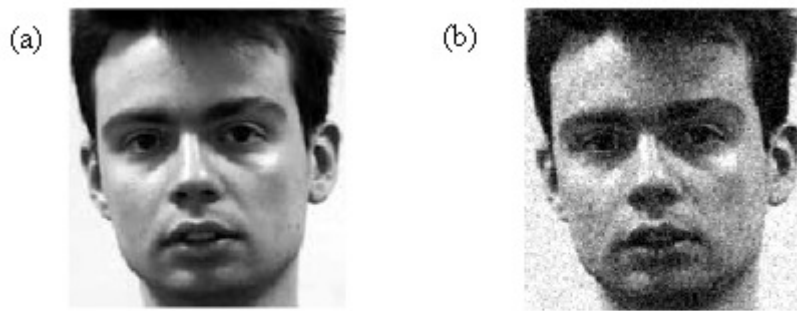
Oznacza to, że korelacja fazowa pomiędzy dwoma identycznymi obrazami jest równa wartości funkcji delty Kroneckera $\delta(m,n)$ (rys. 2.52)



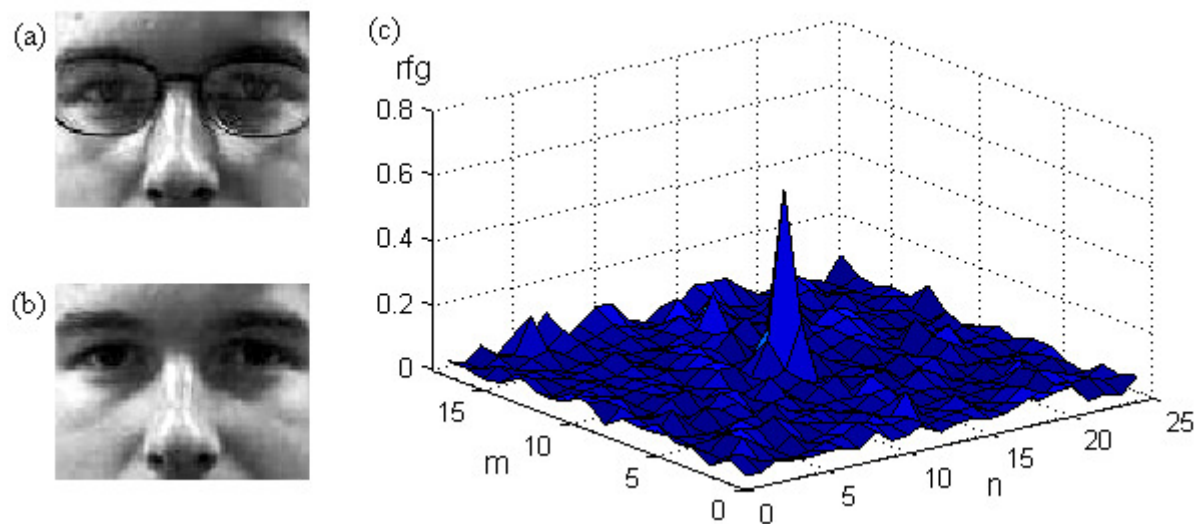
Rys. 2.52. Rezultat zastosowania funkcji korelacji fazowej $rfg(m, n)$: (a) – obraz $f(m,n)$; (b) – obraz $g(m,n)$; (c) – funkcja korelacji fazowej pomiędzy dwoma identycznymi obrazami (obraz $f(m,n)$); (d) – funkcja korelacji fazowej pomiędzy obrazami $f(m, n)$ i $g(m, n)$

Gdy dwa obrazy są do siebie podobne, funkcja ich wzajemnej korelacji fazowej daje w odpowiedzi wyraźny ostry pik. W przeciwnym wypadku, pik ten zanika. Cecha ta powoduje, że w stosunku do klasycznej wzajemnej korelacji, funkcja korelacji fazowej posiada lepsze właściwości dyskryminacyjne. Wysokość (amplituda) piku może być wykorzystana jako miara stopnia podobieństwa dwóch obrazów. Innymi, ważnymi z punktu widzenia porównywania obrazów właściwościami funkcji korelacji fazowej są jej niewrażliwość na przesunięcia obrazów, zmiany jasności i zaszumienie obrazów (rys. 2.53).

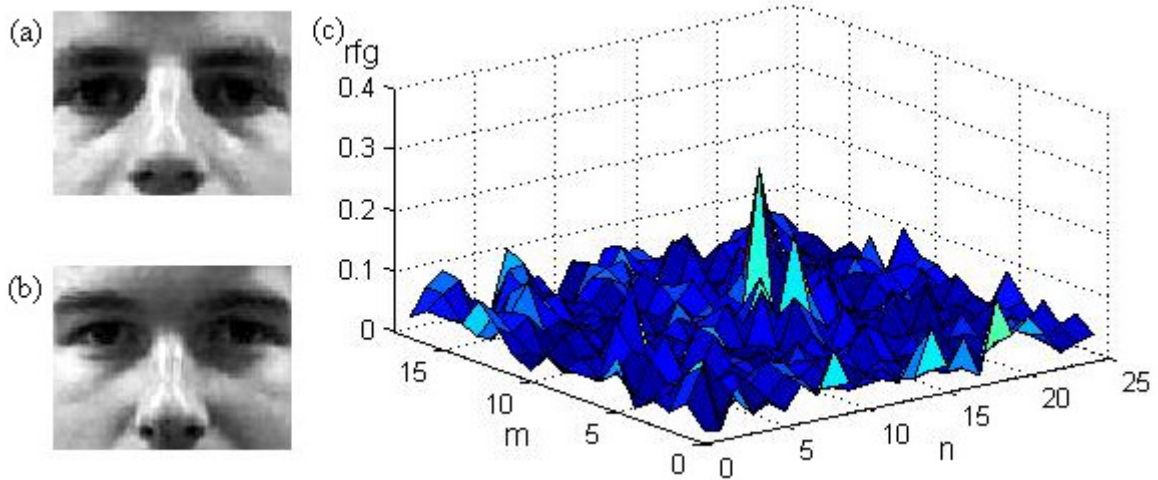
Przykłady wykorzystania funkcji korelacji fazowej do badania stopnia podobieństwa obrazów przedstawiono na rys. 2.54 – 2.56. Dla porównania zostały również wyznaczone współczynniki korelacji wzajemnej. Wyniki potwierdzają, że zastosowanie korelacji fazowej do oceny podobieństwa obrazów twarzy (centralnych części twarzy) pozwala na uzyskanie wiarygodnej informacji o tym, czy mamy do czynienia z obrazem tej samej osoby czy też nie. Analogiczne rezultaty otrzymane w wyniku zastosowania metody korelacji wzajemnej takiej informacji już nie dostarczają. We wszystkich przypadkach otrzymano wysokie wartości korelacji wzajemnej, niezależnie od konfiguracji porównywanych obrazów (rys. 2.54 – 2.56). Stwierdzony został więc fakt, że porównywane obrazy należą do tej samej klasy – klasy obrazów twarzy, nie został jednak dokonany ich właściwy podział.



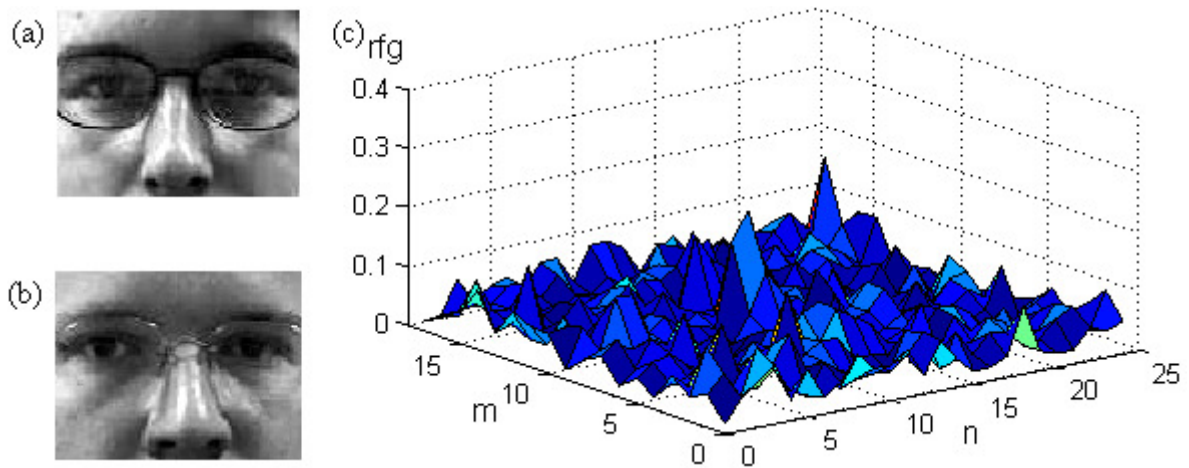
Rys. 2.53. Rezultat zastosowania korelacji fazowej (c) do badania stopnia podobieństwa obrazu oryginalnego (a) do zaszumionego odpowiednika (b). Wartości współczynników korelacji: fazowego – 0,2868; zwyczajny – 0,9598



Rys. 2.54. Funkcja korelacji fazowej r_{fg} centralnych części twarzy tej samej osoby z okularami (a) i bez nich (b). Współczynniki korelacji: fazowy = 0,61; zwyczajny = 0,9231



Rys. 2.55. Funkcja korelacji fazowej rfg centralnych części twarzy dwóch różnych osób. Współczynniki korelacji: fazowy = 0,2622; zwyczajny = 0,7423



Rys. 2.56. Funkcja korelacji fazowej rfg centralnych części twarzy dwóch różnych osób – twarze z okularami. Współczynniki korelacji: fazowy = 0,2089; zwyczajny = 0,7439

3. Implementacja

3.1. Informacje ogólne

Wykonany program „FaceRec” zapewnia realizację wszystkich omówionych do tej pory zadań. Daje użytkownikowi pełną kontrolę nad procesem rozpoznawania, udostępniając ponadto wiele dodatkowych funkcji. Proces jego realizacji przebiegał w dwóch etapach: w pierwszym zbudowany został z wykorzystaniem środowiska MATLAB prototypowy model systemu, w kolejnym – właściwa implementacja, wykorzystująca platformę .NET i język C#.

Możliwości jakie oferuje MATLAB w zakresie dostępności gotowych algorytmów (w dziedzinie analizy obrazów, transformacji, operacji macierzowych), szybkości tworzenia własnych programów i możliwość wizualizacji rezultatów ich działania znacznie ułatwiły proces tworzenia całego systemu i umożliwiły przetestowanie poprawności działania własnych algorytmów.

3.2. Podstawowe funkcjonalności

System realizuje zadanie identyfikacji osób w oparciu o ich wybrane cechy osobnicze twarzy. Twarze wprowadzane są do systemu w formie kolorowych obrazów zapisanych w różnych formatach (*.bmp, *.jpg). Istnieje możliwość jego rozbudowy o inne źródła: skaner, aparat fotograficzny, kamera internetowa.

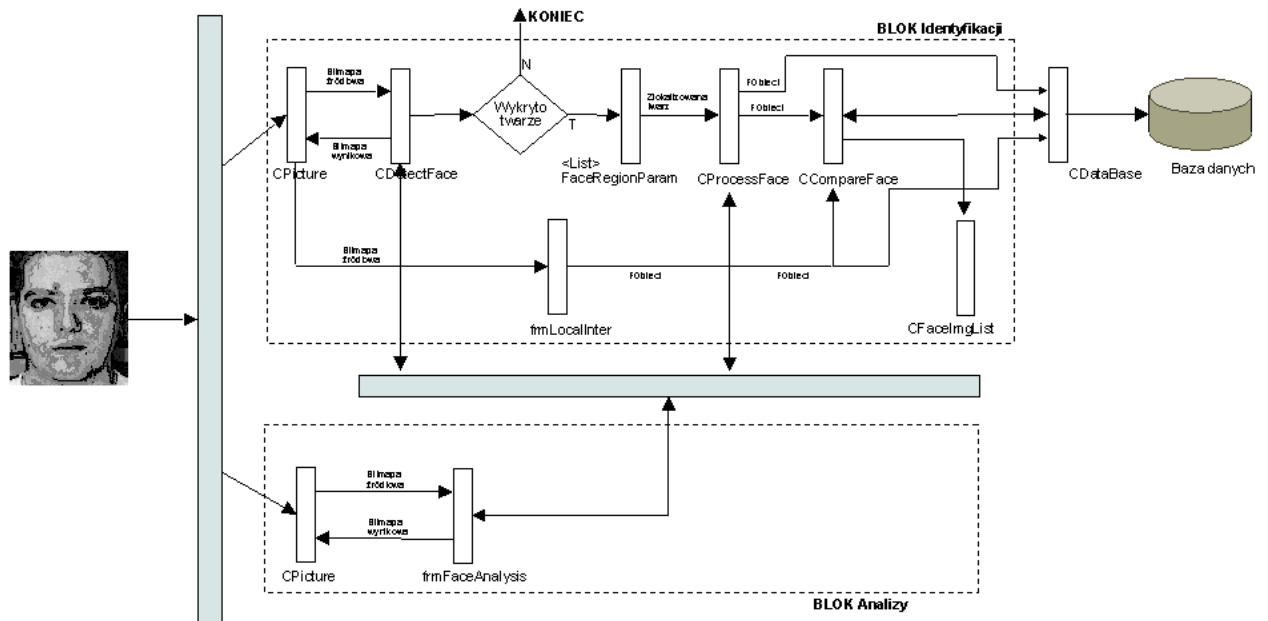
Podstawowe funkcjonalności programu to:

- obsługa obrazów zapisanych w formatach bmp, JPG,
- operacje na obrazach - w programie jest dostęp do wybranych algorytmów obróbki statycznych obrazów wykorzystywanych w identyfikacji. Ze względu na ich uniwersalność można z nich również korzystać w odniesieniu do obrazów nie powiązanych bezpośrednio z samą identyfikacją. Operacje te to m.in. binaryzacja, ekstrakcja składowych obrazów, indeksacja, filtry kontekstowe, inwersja itp.,
- bezpośredni dostęp do wybranych algorytmów procesu identyfikacji,
- detekcja i lokalizacja twarzy na obrazach źródłowych – nie ma ograniczenia co do liczby twarzy na obrazie źródłowym. Natomiast zostało wprowadzone ograniczenie dotyczące minimalnej powierzchni obszaru twarzy,

- lokalizacja charakterystycznych punktów twarzy człowieka - program umożliwia wyznaczenie położenia oczu, nosa, ust oraz punktów opisujących szerokość twarzy,
- ekstrakcja cech twarzy – na podstawie położenia charakterystycznych punktów twarzy,
- interaktywna ekstrakcja cech – funkcja umożliwia ręczną lokalizację punktów twarzy wykonywaną przez użytkownika systemu, na podstawie których wyliczane są odpowiednie wartości wektora cech,
- identyfikacja twarzy – wektory cech opisujące badaną twarz porównywane są z przechowywanymi w bazie danych wzorcami twarzy. Klasyfikacja odbywa się zgodnie z przyjętymi kryteriami identyfikacji,
- analiza krokowa poszczególnych etapów całego procesu identyfikacji od segmentacji po ekstrakcję cech twarzy,
- budowa i administracja bazy danych twarzy.

3.3. Architektura systemu

Architektura programu odpowiada logicznemu podziałowi całego procesu identyfikacji osoby. Poszczególne operacje procesu zostały zgrupowane w klasy udostępniające realizujące je metody. Na rysunku 3.1. przedstawiono ogólny schemat systemu.



Rys. 3.1. Ogólny schemat systemu

W systemie można wyróżnić dwa główne bloki przetwarzania: blok identyfikacji i blok analizy. Blok identyfikacji agreguje wszystkie najważniejsze klasy związane z procesem identyfikacji i udostępnia mechanizmy służące do komunikacji pomiędzy nimi. Obraz źródłowy zawierający twarze umieszczany jest w obiekcie klasy *CPicture*. Wszystkie kolejne rezultaty przetwarzania tego obrazu zostają umieszczone także w tej klasie. Pierwszym krokiem jest detekcja i lokalizacja wszystkich twarzy na obrazie wejściowym. Zadanie to realizowane jest przez klasę *CDetectFace*. Udostępnia ona metody służące do:

- budowy obrazu podobieństwa,
- segmentacji obrazu wejściowego,
- wyznaczenia owalu twarzy,
- procesowania poszczególnych segmentów obrazu: detekcji i lokalizacji twarzy.

W przypadku pomyślnej detekcji (zlokalizowana została minimum 1 twarz) każdy zlokalizowany fragment obrazu źródłowego (ang. *blob*) reprezentujący ludzką twarz umieszczany jest w odpowiadającej mu strukturze *FaceRegionParam*. Struktura ta, oprócz samego obrazu twarzy, przechowuje ponadto informacje na temat jego położenia na obrazie źródłowym, oszacowany na podstawie orientacji segmentu kąt nachylenia twarzy w płaszczyźnie *XY* oraz nowy obraz twarzy powstały w wyniku kompensacji wyliczonego kąta nachylenia. Kolejnym krokiem jest budowa wektorów cech opisujących zlokalizowane twarze poprzedzona ekstrakcją położenia charakterystycznych punktów opisujących twarz

danej osoby. Odpowiednie metody udostępnia klasa *CProcessFace*. Realizowane przez nią zadania to m.in.:

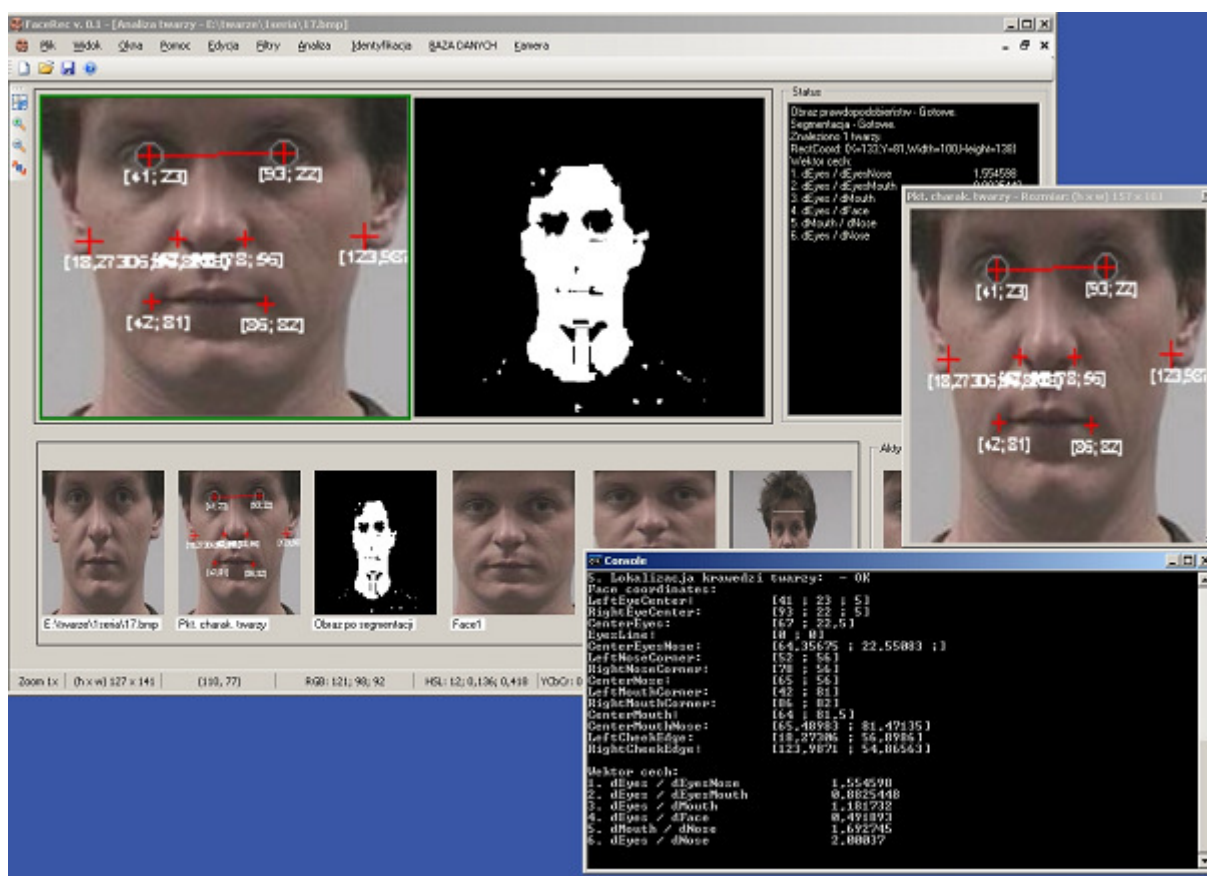
- lokalizacja centrów oczu,
- lokalizacja nosa,
- lokalizacja ust,
- wyznaczenie szerokości twarzy,
- wyznaczenie środkowej (centralnej) części twarzy,
- budowa wektora cech.

Rezultat przetwarzania zostaje umieszczony w obiekcie klasy *FObject*. W takiej postaci przechowywane są także wzorce w bazie danych twarzy. Identyfikacja odbywa się z wykorzystaniem obiektu klasy *CCompareFace*. Najlepsze ostatnie pięć wyników porównań przechowywanych jest przez obiekt klasy *CFaceImgList*. Opisany algorytm przedstawia cały proces identyfikacji. Użytkownik ma możliwość pełnego wpływu na jego przebieg. Oznacza to, że dysponując na przykład gotowymi zdjęciami przedstawiającymi obraz twarzy, może on pominąć etap związany z jej lokalizacją. W sytuacjach, gdy automatyczna detekcja lub ekstrakcja cech zakończy się niepowodzeniem lub gdy użytkownik uzna, że uzyskane rezultaty nie są w pełni satysfakcjonujące, program udostępnia możliwość interaktywnej obsługi procesu budowy wektora cech. Stosowne funkcjonalności udostępnia klasa *frmLocalInter*. Pozwala ona na manualne wydzielenie obszaru twarzy oraz lokalizację położenia punktów charakterystycznych twarzy. Na tej podstawie budowany jest nowy obiekt klasy *FObject* opisujący analizowany obraz twarzy, który może być wykorzystany do przeprowadzenia identyfikacji danej osoby. Niezależnie od sposobu wyznaczenia tego obiektu (automatycznie lub w sposób interakcyjny) operator systemu może zdecydować o tym, czy obiekt ma być włączony do bazy danych czy też ma być przeprowadzana wyłącznie identyfikacja osoby, którą opisuje. Komunikacja z bazą danych odbywa się poprzez obiekt klasy *CDataBase*.

Podstawową funkcją realizowaną przez blok analizy jest natomiast udostępnienie użytkownikowi interfejsu umożliwiającego dostęp do metod bloku identyfikacji. Pozwala to na śledzenie wyników działania poszczególnych algorytmów programu, korektę parametrów przetwarzania, dostęp do funkcji przetwarzających obrazy. Dostęp do nich realizowany jest z wykorzystaniem metod klasy *frmFaceAnalysis*.

3.4. Opis programu

Podstawowy widok interfejsu użytkownika programu przedstawia rys. 3.2. Został on tak zaprojektowany, aby obsługa programu była prosta i intuicyjna. Umożliwia on dostęp do wszystkich funkcji systemu związanych z analizą obrazów twarzy i identyfikacją. W zależności od trybu pracy zmienia się rola poszczególnych pól, natomiast ich układ pozostaje niezmienny. Szczegółowy opis poszczególnych funkcji programu znajduje się poniżej.



Rys. 3.2. Podstawowy widok GUI programu w trybie analizy obrazów twarzy

3.4.1. Analiza obrazów twarzy. Detekcja i lokalizacja twarzy

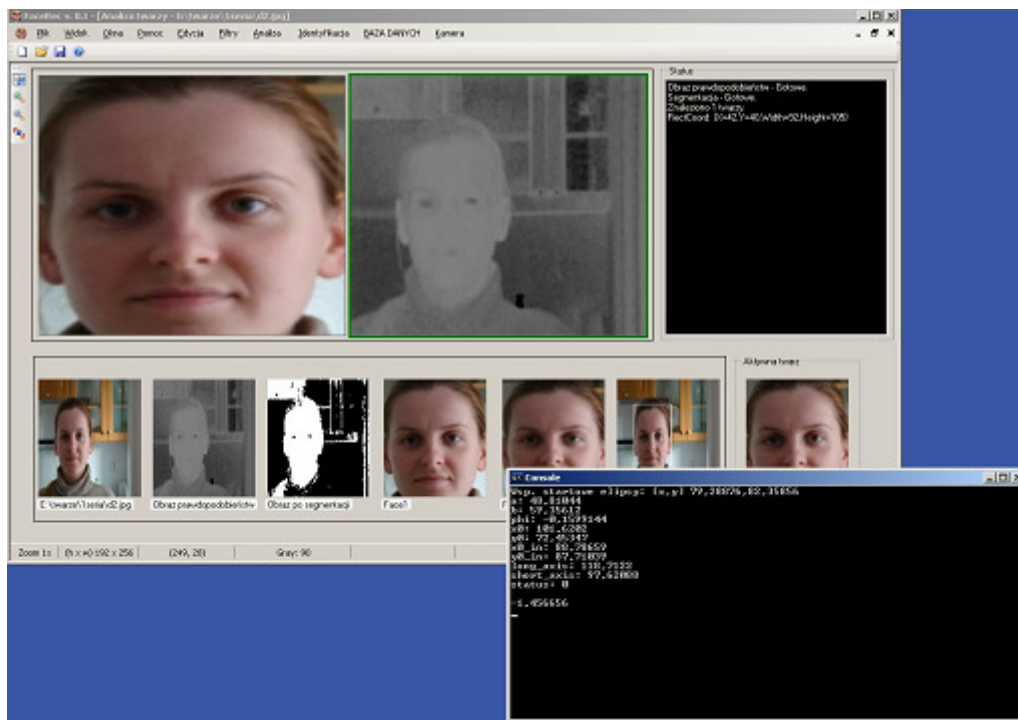
Funkcje z tej grupy pozwalają użytkownikowi na modyfikacje obrazów źródłowych oraz przeprowadzenie detekcji i lokalizacji twarzy. Dostęp do wybranych funkcji modyfikujących obrazy możliwy jest poprzez menu **Filtry**. Znaczenie poszczególnych funkcji jest następujące:

- **Filtry / Kolor** – funkcje konwersji pomiędzy różnymi przestrzeniami barw,
- **Filtry / Filtracja 2D** – filtracja obrazów (mediana, f. dolnoprzepustowa)

- **Filtry / Binaryzacja** – binaryzacja obrazów w skali szarości (z progiem, adaptacyjna).
- **Filtry / Korekcja kontrastu**,
- **Filtry / Min2D** – filtr 2D minimum,
- **Filtry / Operacje morfologiczne**,
- **Filtry / Inne** – indeksacja obrazów binarnych, wyznaczenie liczby dziur w segmencie
- **Filtry / Obrót**,
- **Filtry / Powiększ + , Pomniejsz -**.

W celu przeprowadzenia operacji lokalizacji twarzy (rys. 3.3) należy wykorzystać funkcje z menu *Analiza*.

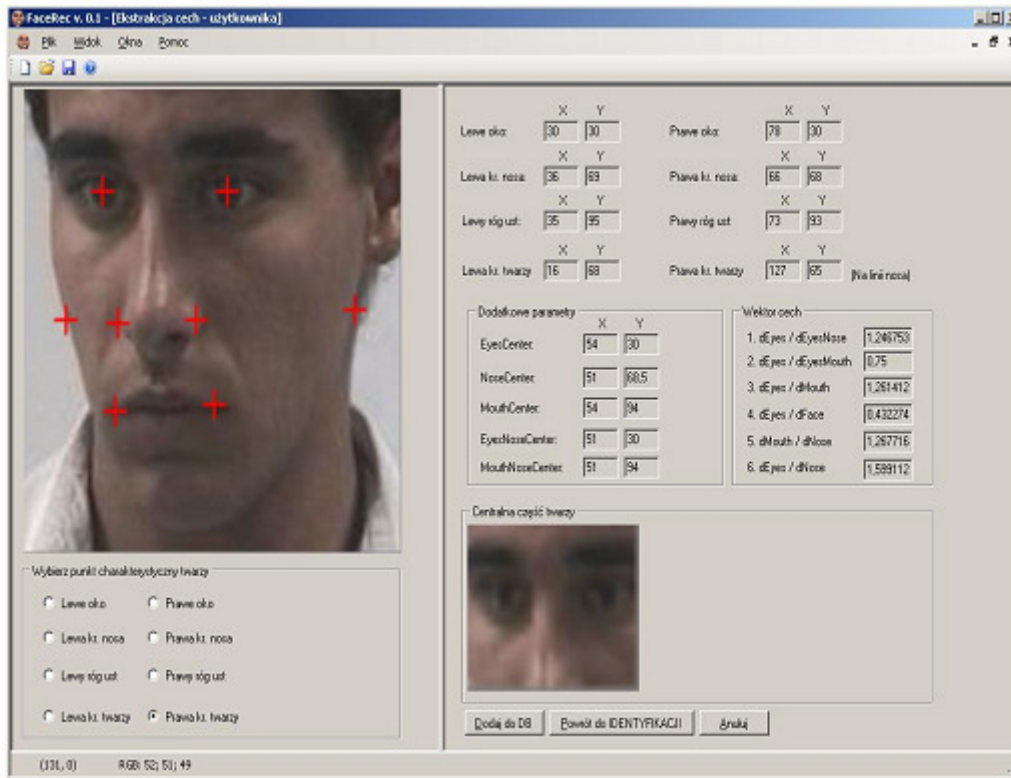
- **Analiza / START Lokalizacja** – uruchomienie automatycznej procedury lokalizacji twarzy,
- **Analiza / Obraz źródłowy** – wyświetlenie w wybranym panelu obrazu źródłowego,
- **Analiza / Obraz podobieństwa** – wygenerowanie i wyświetlenie w wybranym panelu obrazu podobieństw,
- **Analiza / Segmentacja** – segmentacja obrazu podobieństw z ustalonym lub wyliczonym adaptacyjnie progiem,
- **Analiza / Lokalizacja twarzy** – lokalizacja twarzy na zsegmentowanym obrazie źródłowym,
- **Analiza / Opcje** – ustawienie parametrów procesu lokalizacji obrazów twarzy



Rys. 3.3. Lokalizacja twarzy na obrazie źródłowym

3.4.2. Budowa wektorów cech

Informacjami wejściowymi do procedury budowy wektora cech opisującego analizowany obraz są współrzędne charakterystycznych punktów twarzy. Zatem proces budowy wektora cech można podzielić na dwa etapy: lokalizację na wyodrębnionym obszarze twarzy ich położenia oraz właściwą procedurę wyliczającą współrzędne wektora. Lokalizacja może odbywać się na dwa sposoby: z wykorzystaniem wbudowanych procedur automatycznej lokalizacji poszczególnych cech twarzy lub w sposób interakcyjny przeprowadzany przez użytkownika. Funkcje służące do realizacji wymienionych zadań zlokalizowane zostały w menu **Analiza / Zlokalizuj** oraz **Analiza / Ekstrakcja cech**. Wybór funkcji **Analiza / Lokalizuj interaktywnie** daje możliwość manualnej lokalizacji położenia charakterystycznych punktów twarzy (rys. 3.4.).



Rys. 3.4. Interaktywna budowa wektora cech

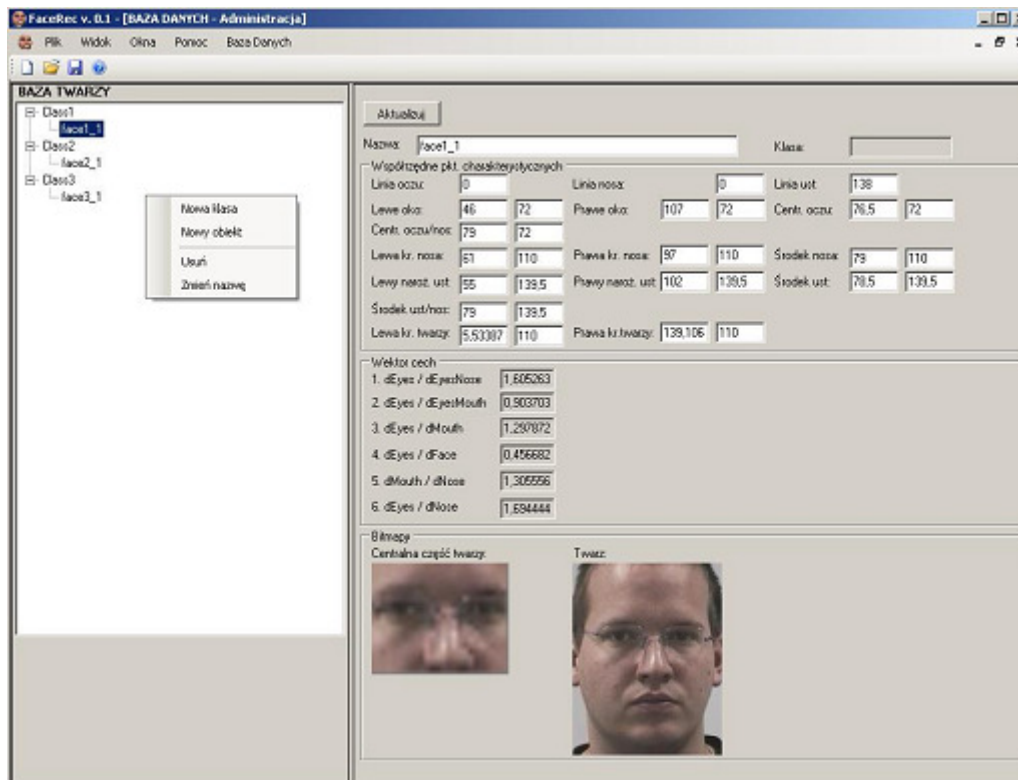
Na podstawie wybranych współrzędnych wyliczany jest wektor cech oraz wyznaczana jest centralna część twarzy. Niezależnie od wybranej metody ekstrakcji cech, rezultatem jej działania jest zawsze obiekt klasy *FObject* (rozd. 3.3.) opisujący analizowaną twarz. Użytkownik ma możliwość uzupełnienia istniejącej bazy danych o ten obiekt (**Dodaj do DB**) lub przeprowadzenia jego identyfikacji.

3.4.3. Identyfikacja

Wybranie polecenia **Identyfikacja / START** z menu głównego rozpoczyna procedurę identyfikacji bieżącego obiektu twarzy. Parametry (rozd. 2.6.) procedury można zmodyfikować poprzez pole menu **Identyfikacja / Opcje**. Przez cały okres trwania procedury można obserwować na bieżąco rezultaty jej działania: aktualnie porównywany obiekt twarzy oraz aktualną listę najlepiej dopasowanych dotychczas obiektów bazy, wraz z informacją o stopniu podobieństwa.

3.4.4. Obsługa bazy danych

Dostęp do panelu obsługi bazy danych możliwy jest bezpośrednio poprzez wybranie polecenia **Baza Danych / Administracja** z menu głównego (rys.3.5) lub pośrednio – podczas dodawania nowego obiektu twarzy.



Rys. 3.5. Panel obsługi bazy danych twarzy

Panel umożliwia użytkownikowi:

- dodawanie, usuwanie, modyfikowanie klas obiektów,
- dodawanie, usuwanie obiektów w ramach klas,
- modyfikacje danych obiektów.

Struktura bazy danych odpowiada wcześniej opisanemu schematowi.

4. Wyniki

4.1. Wprowadzenie

W rozdziale zostaną przedstawione wyniki przeprowadzonych doświadczeń, sprawdzające skuteczność działania systemu pod kątem: detekcji i lokalizacji twarzy, ekstrakcji z wydzielonych obszarów twarzy współrzędnych charakterystycznych punktów twarzy (z pominięciem lokalizacji interaktywnej) oraz identyfikacji. Duży wpływ na skuteczność działania systemu mają wartości kilku podstawowych parametrów. Należą do nich: parametry modelu „*skin-color detection*” ψ i Λ , próg segmentacji t , parametry procedury identyfikacji (T_1 , T_2 , wybór kryteriów identyfikacji). Doświadczenia przeprowadzono dla kilku kombinacji parametrów. Parametry modelu „*skin-color detection*” wyliczono zgodnie z opisaną w rozdz. 2.2.3. procedurą.

4.2. Baza danych twarzy

Do eksperymentów wykorzystano obrazy twarzy pochodzące z baz danych: *PRIMA*, *CALTECH* oraz źródeł własnych.

Baza danych *PRIMA* jest rezultatem pracy wykonywanej w ramach projektu badawczego *PRIMA* realizowanego przez Centrum Badawcze *INRIA Grenoble* a dotyczącego określania orientacji twarzy. Baza zawiera 15 różnych zbiorów danych przedstawiających obrazy twarzy. Każdy zbiór zawiera dwie serie danych, każda po 93 zdjęcia tej samej osoby w różnych pozach. W bazie umieszczono zdjęcia osób o różnych kolorach skóry, noszących lub nie okulary.

Ze względu na cel projektu, w ramach którego baza powstała, zawarte w niej zdjęcia przedstawiają obrazy twarzy, których orientacja jest ściśle określona i determinowana dwoma kątami (h , v), z których pierwszy oznacza obrót twarzy w poziomie a drugi – w pionie. Wartości kątów zmieniają się w granicach od -90 stopni do +90 stopni. Wszystkie obrazy zapisane są w formacie *JPEG*, w pełnej palecie kolorów (24bit *RGB*) o rozmiarach 388 x 284.



Rys. 4.1. Próbkę zdjęć przedstawiających tę samą osobę. Baza danych *PRIMA*.

Baza danych *CALTECH* powstała w ramach prowadzonych prac badawczych z dziedziny cyfrowego przetwarzania obrazów na *California Institute of Technology*. Zawiera 450 kolorowych zdjęć, przedstawiających 27 różnych osób o różnej płci, rasie, noszących okulary lub nie. Warunki powstawania zdjęć były bardzo zróżnicowane pod kątem oświetlenia, miejsca wykonania zdjęcia (wewnątrz i na zewnątrz budynków), mimiki twarzy, tła. Stała pozostawała orientacja twarzy – frontem do oka kamery.



Rys. 4.2. Próbkę zdjęć przedstawiających tę samą osobę. Baza danych *CALTECH*.

4.3. Detekcja i lokalizacja twarzy

Doświadczenia przeprowadzono dla następujących parametrów modelu:

$$\psi = \begin{bmatrix} -0,0481 \\ 0,0889 \end{bmatrix} \text{ i } \Lambda = \begin{bmatrix} 0,7939 & -0,4679 \\ -0,4679 & 0,6444 \end{bmatrix} \times 10^{-3}.$$




Model zbudowany został na podstawie próbek pobranych z obszarów twarzy zdjęć należących do bazy *PRIMA*. Segmentacja przeprowadzona została dla różnych wartości progów ustalanych manualnie i dla progów wyliczanych adaptacyjnie przez program. Dany segment klasyfikowany jest jako zawierający twarz, gdy spełnia zadane kryteria (rozdz. 2.3.2).

Poniżej przedstawiono otrzymane rezultaty dla różnych kombinacji wartości progów segmentacji i obrazów zawierających twarze pochodzących z kilku różnych źródeł. Należy zaznaczyć, że zdjęcia zostały wykonane w warunkach niekontrolowanych (różne tło, oświetlenie, wielkość zdjęć) – w niektórych przypadkach będzie miało to zasadniczy wpływ na wyniki detekcji.

Przypadek 1 – próg segmentacji 0,1





Obraz 1:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	jednolite tło, rasa: biała, twarz na pierwszym planie frontem, światło z prawej strony
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 2:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, światło z lewej strony
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 3:

źródło:	CALTECH Database
wymiary:	212 x 320
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem,
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			




Obraz 4:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, zdjęcie wykonane na zewnątrz, niejednoznaczny rezultat segmentacji
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			




Obraz 5:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: biała (Latynos), twarz na pierwszym planie frontem,
rezultat:	negatywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 6:





źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: żółta, twarz na pierwszym planie obrócona pod kątem 30 stopni w prawo
rezultat:	negatywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			BRAK

Przypadek 2 – próg segmentacji 0,35




Obraz 1:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	jednolite tło, rasa: biała, twarz na pierwszym planie frontem, światło z prawej strony
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			



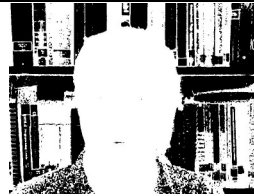

Obraz 2:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, światło z lewej strony
rezultat:	negatywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			BRAK





Obraz 3:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem,
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

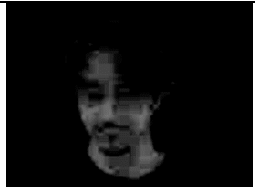
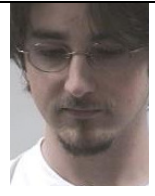
Obraz 4:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, zdjęcie wykonane na zewnątrz,
rezultat:	negatywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 5:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: biała, broda, twarz na pierwszym planie pochylona 15 stopni w dół
rezultat:	negatywny , włosy zniekształciły owal twarzy

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

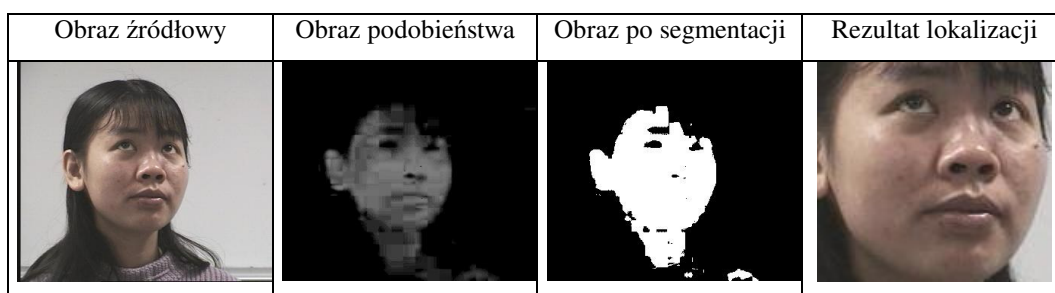
Obraz 6:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: biała (Latynos), twarz na pierwszym planie obrócona pod kątem 15 stopni w lewo
rezultat:	pozytywny




Obraz 7:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: żółta, twarz na pierwszym planie obrócona pod kątem 30 stopni w lewo i 15 stopni do góry
rezultat:	pozytywny



Obraz 8:





źródło:	PRIMA DATABASE
wymiary:	587 x 749
charakterystyka:	jednolite tło, rasa: biała, dwie twarze: pierwsza 15 stopni w dół i 15 stopni w lewo; druga: 15 stopni w górę
rezultat:	pozytywny

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Przypadek 3 – próg segmentacji liczony adaptacyjnie





Obraz 1:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	jednolite tło, rasa: biała, twarz na pierwszym planie frontem, światło z prawej strony
rezultat:	pozytywny , thr. = 0.5

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			





Obraz 2:

źródło:	własne
wymiary:	192 x 256
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, światło z lewej strony
rezultat:	pozytywny , thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			





Obraz 12:

źródło:	własne
wymiary:	256 x 192
charakterystyka:	jednolite tło, rasa: biała, twarz na pierwszym planie frontem, światło z prawej strony
rezultat:	pozytywny , thr. = 0.3

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 3:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem,
rezultat:	pozytywny , thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			



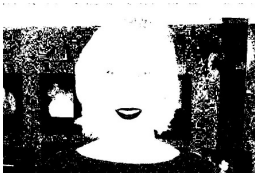
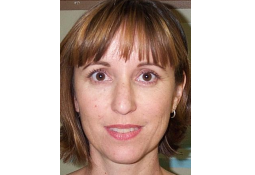
Obraz 4:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, zdjęcie wykonane na zewnątrz, niejednoznaczny rezultat segmentacji
rezultat:	pozytywny , thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			


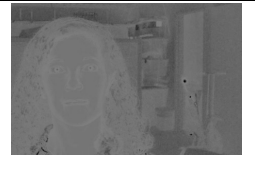


Obraz 9:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, niejednoznaczny rezultat segmentacji
rezultat:	pozytywny , thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			


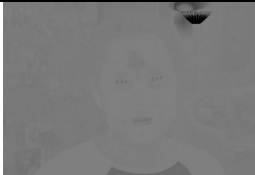


Obraz 10:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem, niejednoznaczny rezultat segmentacji, obszar twarzy przeświecony
rezultat:	negatywny , thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

Obraz 11:

źródło:	CALTECH Database
wymiary:	592 x 896
charakterystyka:	złożone tło, rasa: biała, twarz na pierwszym planie frontem
rezultat:	negatywny ; thr. = 0.1

Obraz źródłowy	Obraz podobieństwa	Obraz po segmentacji	Rezultat lokalizacji
			

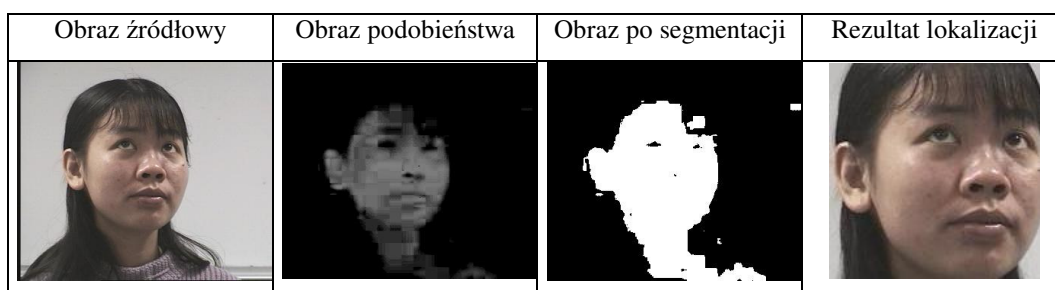
Obraz 5:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: biała (Latynos), twarz na pierwszym planie obrócona pod kątem 15 stopni w lewo
rezultat:	pozytywny , thr. = 0.5



Obraz 6:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: żółta, twarz na pierwszym planie obrócona pod kątem 30 stopni w lewo i 15 stopni do góry
rezultat:	pozytywny (możliwe do skorygowania w dalszym procesowaniu), thr. = 0.5



Obraz 13:

źródło:	PRIMA DATABASE
wymiary:	288 x 384
charakterystyka:	jednolite tło, rasa: biała, twarz na pierwszym planie pochylona 15 stopni w dół
rezultat:	negatywny , thr. = 0.2





4.4. Ekstrakcja cech

Doświadczenia przeprowadzono na wybranych twarzach z każdego dostępnego katalogu. Wynik każdego z eksperymentów jest subiektywną oceną dokonaną przez przeprowadzającego doświadczenie.

Baza PRIMA

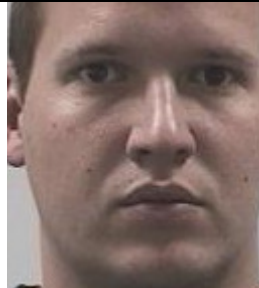

Obraz Class 1, ID: 1:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, okulary, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,605263
		2. dEyes / dEyesMouth	0,9037037
		3. dEyes / dMouth	1,297872
		4. dEyes / dFace	0,4668803
		5. dMouth / dNose	1,382353
		6. dEyes / dNose	1,794118


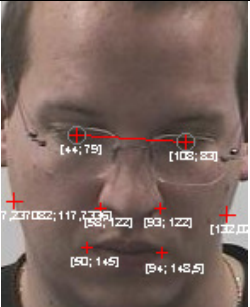
Obraz Class 1, ID: 2:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, bez okularów, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,575566
		2. dEyes / dEyesMouth	0,9012553
		3. dEyes / dMouth	1,371919
		4. dEyes / dFace	0,4633152
		5. dMouth / dNose	1,344113
		6. dEyes / dNose	1,844015

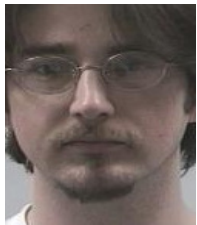

Obraz Class 1, ID: 5:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, pochylenie 30 stopni w dół, okulary, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,56588
		2. dEyes / dEyesMouth	0,9734848
		3. dEyes / dMouth	1,452795
		4. dEyes / dFace	0,5128559
		5. dMouth / dNose	1,261114
		6. dEyes / dNose	1,832139



Obraz Class 4, ID: 1:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, obrót 5 stopni w prawo, okulary, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,59515
		2. dEyes / dEyesMouth	0,897723
		3. dEyes / dMouth	1,222349
		4. dEyes / dFace	0,486467
		5. dMouth / dNose	1,730876
		6. dEyes / dNose	2,115734


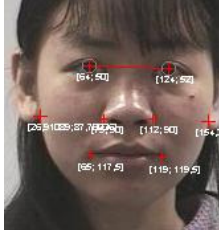
Obraz Class 4, ID: 2:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, bez okularów, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,579557
		2. dEyes / dEyesMouth	0,9075655
		3. dEyes / dMouth	1,055598
		4. dEyes / dFace	0,5058088
		5. dMouth / dNose	1,894962
		6. dEyes / dNose	2,000319


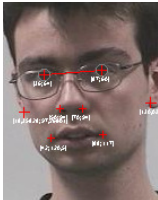
Obraz Class 3, ID: 1:

źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, bez okularów, rasa: żółta
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,579557
		2. dEyes / dEyesMouth	0,9075655
		3. dEyes / dMouth	1,055598
		4. dEyes / dFace	0,5058088
		5. dMouth / dNose	1,894962
		6. dEyes / dNose	2,000319

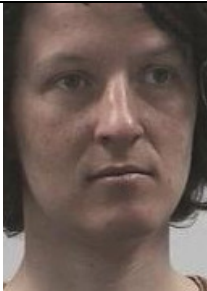

Obraz Class 5, ID: 1:

źródło:	PRIMA DATABASE
charakterystyka:	obrót 10 stopni względem osi Y w prawo, okulary, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,609471
		2. dEyes / dEyesMouth	0,8998539
		3. dEyes / dMouth	1,108895
		4. dEyes / dFace	0,4566784
		5. dMouth / dNose	2,306648
		6. dEyes / dNose	2,557831



Obraz Class 2, ID: 3:

źródło:	PRIMA DATABASE
charakterystyka:	obrót 15 stopni w lewo, rasa: biała
rezultat:	niepełna ekstrakcja – brak detekcji lewej krawędzi twarzy z powodu niepełnego obr. krawędzi

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,595163
		2. dEyes / dEyesMouth	0,8990296
		3. dEyes / dMouth	1,340967
		4. dEyes / dFace	-1
		5. dMouth / dNose	1,467005
		6. dEyes / dNose	1,967206



Obraz Class 7, ID: 2:

źródło:	PRIMA DATABASE
charakterystyka:	pochylenie 30 stopni do przodu + obrót 15 stopni w lewo, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,598624
		2. dEyes / dEyesMouth	1,019051
		3. dEyes / dMouth	1,073078
		4. dEyes / dFace	0,4909343
		5. dMouth / dNose	1,833787
		6. dEyes / dNose	1,967796

Obraz Class 15, ID: 5:



źródło:	PRIMA DATABASE
charakterystyka:	orientacja frontem, pochylenie 15 stopni w dół, twarz obrócona 15 stopni w lewo, rasa: biała
rezultat:	negatywny , niepoprawna lokalizacja centrum prawego oka w wyniku błędnej detekcji obszaru oka

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,287129
		2. dEyes / dEyesMouth	0,9318996
		3. dEyes / dMouth	1,341093
		4. dEyes / dFace	0,5932347
		5. dMouth / dNose	1,603122
		6. dEyes / dNose	2,149935

Baza CALTECH



Obraz Class 20, ID: 1:

źródło:	CALTECH DATABASE
charakterystyka:	orientacja frontem, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,700443
		2. dEyes / dEyesMouth	0,9815127
		3. dEyes / dMouth	1,407042
		4. dEyes / dFace	0,4916651
		5. dMouth / dNose	1,315229
		6. dEyes / dNose	1,850583


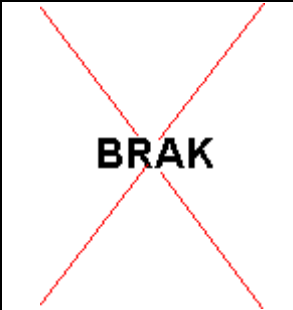
Obraz Class 21, ID: 1:

źródło:	CALTECH DATABASE
charakterystyka:	orientacja frontem, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,855916
		2. dEyes / dEyesMouth	1,077135
		3. dEyes / dMouth	1,372362
		4. dEyes / dFace	0,5198217
		5. dMouth / dNose	1,270288
		6. dEyes / dNose	1,743296

Obraz Class 1, ID: 1:

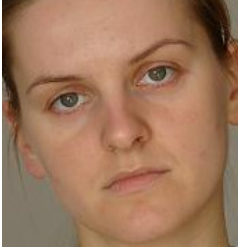

źródło:	CALTECH DATABASE
charakterystyka:	orientacja frontem, nachylenie 10 stopni w prawo, rasa: biała
rezultat:	negatywny , główna przyczyna – niedoświetlenie obiektu

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	
		2. dEyes / dEyesMouth	
		3. dEyes / dMouth	
		4. dEyes / dFace	
		5. dMouth / dNose	
		6. dEyes / dNose	

Baza własna


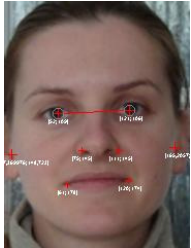
Obraz Class 100, ID: 1:

źródło:	własne
charakterystyka:	orientacja frontem, nachylenie 30 stopni w prawo, rasa: biała
rezultat:	niepełna ekstrakcja - brak detekcji lewej krawędzi twarzy z powodu niepełnego obrazu krawędzi

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,69626
		2. dEyes / dEyesMouth	0,9636229
		3. dEyes / dMouth	1,367215
		4. dEyes / dFace	-1
		5. dMouth / dNose	1,361395
		6. dEyes / dNose	1,861318

Obraz Class 100, ID: 2:

źródło:	własne
charakterystyka:	orientacja frontem, rasa: biała
rezultat:	pozytywny

Źródło	Rezultat det. pkt. charakt.		
		1. dEyes / dEyesNose	1,829689
		2. dEyes / dEyesMouth	1,006648
		3. dEyes / dMouth	1,167915
		4. dEyes / dFace	0,4371634
		5. dMouth / dNose	1,642651
		6. dEyes / dNose	1,918477

4.5. Identyfikacja twarzy

4.5.1. Wprowadzenie



Doświadczenia przeprowadzono z wykorzystaniem obrazów twarzy z bazy PRIMA. Z całego zbioru dostępnych obrazów dokonano selekcji obrazów spełniających wymagane kryteria. Następnie wydzielone obrazy podzielono na dwa zbiory: zbiór uczący, stanowiący w przypadku programu *FaceRec* wbudowaną bazę twarzy oraz zbiór testowy wykorzystany


do przeprowadzenia doświadczeń. W celu oceny skuteczności wbudowanego algorytmu identyfikacji przeprowadzone zostały dodatkowe doświadczenia z wykorzystaniem pakietów *RapidMiner* oraz *PRTools* udostępniających popularne metody klasyfikacji danych (sieci neuronowe, LDA, SVC).

4.5.2. Program *FaceRec*

Doświadczenia przeprowadzono dla różnych wartości parametrów identyfikacji. Bazę testową stanowił zbiór 44 obiektów, będących obrazami twarzy spełniającymi wymagane kryteria. Warunkiem udziału w identyfikacji była poprawna ekstrakcja cech z obrazu wejściowego.


Przypadek 1	
Parametry identyfikacji	Kryterium klasyfikacji: (a) Warunki progowe: brak Weryfikacja: brak

Obraz testowy:		
	Przynależność:	Class1
	Wynik identyfikacji	pozytywny
	Uwagi	3 z 5 twarzy zgodnych na pierwszych pozycjach
		

Obraz testowy:		
	Przynależność:	Class1
	Wynik identyfikacji	pozytywny
	Uwagi	Twarz znacznie pochylona do przodu, 1 z 5 twarzy zgodnych, niski poziom ufności




Obraz testowy:



	Przynależność:	Class1
	Wynik identyfikacji	pozytywny
	Uwagi	twarz obrócona w prawo pod kątem 30stopni, 2 z 5 twarzy zgodnych






Obraz testowy:

	Przynależność:	Class2
	Wynik identyfikacji	pozytywny
	Uwagi	2 z 5 twarzy zgodnych



Obraz testowy:		
	Przynależność:	Class3
	Wynik identyfikacji	pozytywny
	Uwagi	2 z 5 twarzy zgodnych, twarz obrócona w prawo pod kątem 15 stopni
		

Obraz testowy:		
	Przynależność:	Class4
	Wynik identyfikacji	Pozytywny
	Uwagi	2 z 5 twarzy zgodnych, twarz obrócona w lewo pod kątem 15 stopni
		

Obraz testowy:		
	Przynależność:	Class6
	Wynik identyfikacji	pozytywny
	Uwagi	3 z 5 twarzy zgodnych



Obraz testowy:

	Przynależność:	Class8
	Wynik identyfikacji	pozytywny
	Uwagi	3 z 5 twarzy zgodnych




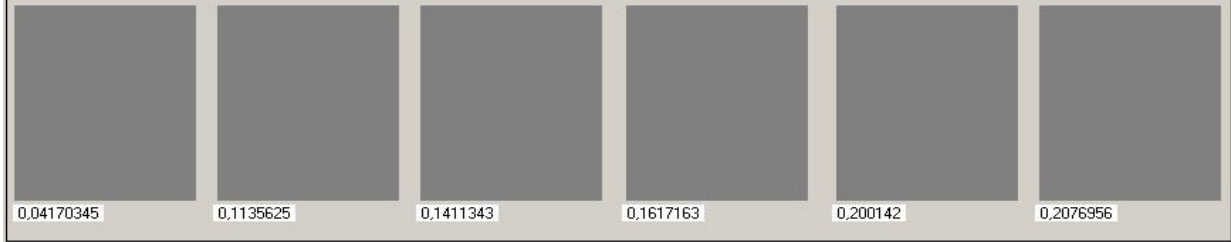
Obraz testowy:



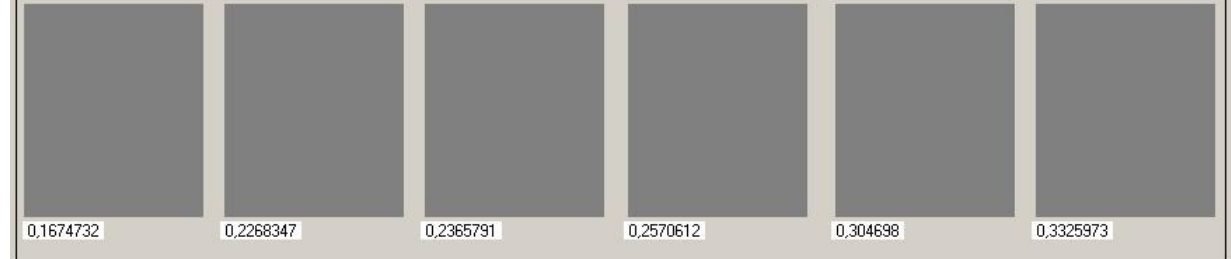
	Przynależność:	Class8
	Wynik identyfikacji	pozytywny
	Uwagi	5 z 5 twarzy zgodnych

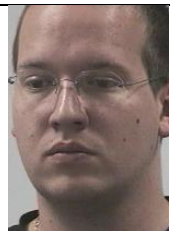


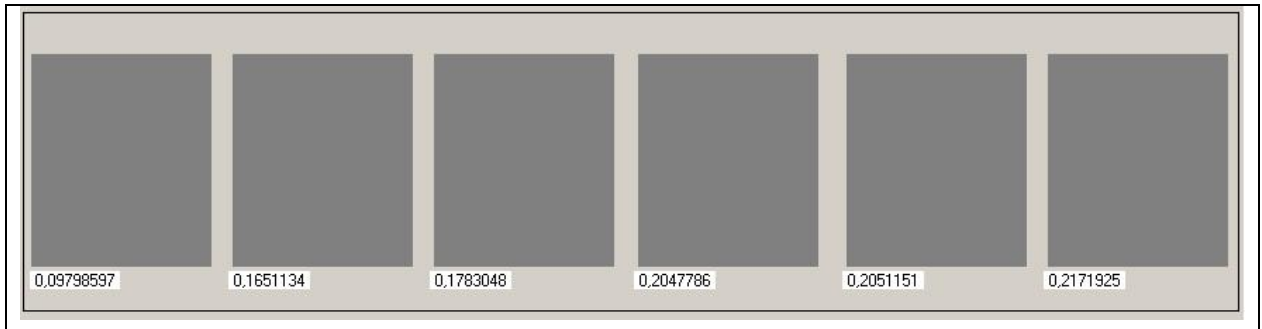
Przypadek 2


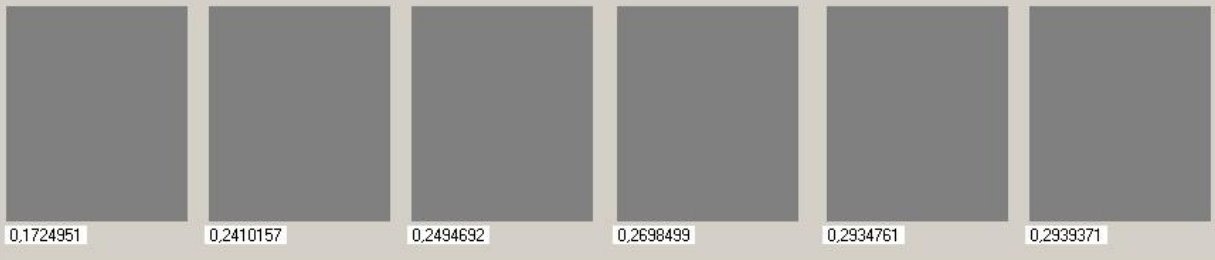
Parametry identyfikacji	Kryterium klasyfikacji: (b)
	Warunki progowe: brak
	Weryfikacja: brak


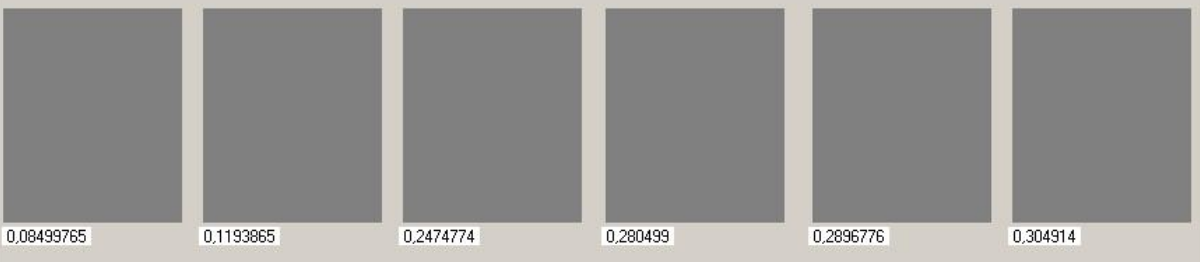
Obraz testowy:		
	Przynależność:	Class1
	Wynik identyfikacji	pozytywny
	Uwagi	
		




Obraz testowy:		
	Przynależność:	Class1
	Wynik identyfikacji	negatywny - wynik identyfikacji: Class20
	Uwagi	twarz pochylona do przodu pod kątem 30 stopni 
		

Obraz testowy:		
	Przynależność:	Class1
	Wynik identyfikacji	pozytywny
	Uwagi	twarz pochylona do przodu pod kątem 15 stopni i obrócona w prawo o 15



Obraz testowy:		
	Przynależność:	Class3
	Wynik identyfikacji	pozytywny
	Uwagi	twarz obrócona w prawo pod kątem 15 stopni
		

Obraz testowy:		
	Przynależność:	Class6
	Wynik identyfikacji	pozytywny
	Uwagi	
		

Obraz testowy:		
	Przynależność:	Class2
	Wynik identyfikacji	negatywny – wynik identyfikacji Class13
	Uwagi	Rezultat: 
		
0,05050994 0,09217958 0,1337259 0,1407618 0,1901239 0,2072434		

Przypadek 3	
Parametry identyfikacji	Kryterium klasyfikacji: (a) Warunki progowe: brak Weryfikacja: brak




Obraz testowy:			
		Przynależność:	Class1
		Wynik identyfikacji	pozytywny
		Wynik weryfikacji	pozytywny
		Uwagi	Korelacja fazowa: 0,37
			
0,06715569 0,07701537 0,1020382 0,1095709 0,1145189 0,1160262			



Obraz testowy:									
		Przynależność:	Class5						
		Wynik identyfikacji	pozytywny						
		Wynik weryfikacji	pozytywny						
		Uwagi	Korelacja fazowa: 1,0						
									
<table border="0"> <tr> <td>0,01892025</td> <td>0,1616338</td> <td>0,1956574</td> <td>0,2060726</td> <td>0,2368832</td> <td>0,2609309</td> </tr> </table>				0,01892025	0,1616338	0,1956574	0,2060726	0,2368832	0,2609309
0,01892025	0,1616338	0,1956574	0,2060726	0,2368832	0,2609309				

Obraz testowy:									
		Przynależność:	Class6						
		Wynik identyfikacji	pozytywny						
		Wynik weryfikacji	pozytywny						
		Uwagi	Twarz pochylona do góry 15 stopni, w lewo 15 stopni Korelacja fazowa: 0,481						
									
<table border="0"> <tr> <td>0,1362079</td> <td>0,1686369</td> <td>0,3397105</td> <td>0,3474183</td> <td>0,3593293</td> <td>0,4834496</td> </tr> </table>				0,1362079	0,1686369	0,3397105	0,3474183	0,3593293	0,4834496
0,1362079	0,1686369	0,3397105	0,3474183	0,3593293	0,4834496				

Obraz testowy:									
		Przynależność:	Class7						
		Wynik identyfikacji	pozytywny						
		Wynik weryfikacji	pozytywny						
		Uwagi	Korelacja fazowa: 0,3572						
									
<table border="0"> <tr> <td>0,07025709</td> <td>0,1102379</td> <td>0,1296119</td> <td>0,1421812</td> <td>0,1438371</td> <td>0,1473266</td> </tr> </table>				0,07025709	0,1102379	0,1296119	0,1421812	0,1438371	0,1473266
0,07025709	0,1102379	0,1296119	0,1421812	0,1438371	0,1473266				

Obraz testowy:		Przynależność:	Class9
		Wynik identyfikacji	pozytywny
		Wynik weryfikacji	negatywny
		Uwagi	Korelacja fazowa: 0,1859 Mimo poprawnej identyfikacji (2 z 5), niska wartość korelacji fazowej związana przede wszystkim ze zniekształconym obrazem oczu
			

Obraz testowy:		Przynależność:	Class10
		Wynik identyfikacji	negatywny
		Wynik weryfikacji	negatywny
		Uwagi	Korelacja fazowa: 0,147 Poprawna klasa na 2 pozycji
			

Obraz testowy:		Przynależność:	Class12
		Wynik identyfikacji	negatywny
		Wynik weryfikacji	negatywny
		Uwagi	Korelacja fazowa: 0,17293 Poprawna klasa na 3 pozycji



Wyniki eksperymentów zestawiono w tabeli 3.

Tabela 3. Wyniki eksperymentów przeprowadzonych w programie *FaceRec*

Lp.	Próba	Parametry	Błąd klasyfikacji
1	Przypadek 1	najbliższy sąsiad, bez weryfikacji	27%
2	Przypadek 2	najbliższy średni obraz, bez weryfikacji	38%
3	Przypadek 3	najbliższy sąsiad, z weryfikacją	27%, (weryfikacja odrzuciła 15% poprawnie sklasyfikowanych twarzy)

4.5.3. Pakiet *PRTools* środowiska *MATLAB*

Pakiet *PRTools* jest zbiorem procedur służących do klasyfikacji obiektów zaimplementowanych w środowisku *MATLAB*. Umożliwia on zastosowanie większości znanych obecnie metod klasyfikacji m.in. liniowej analizy dyskryminacyjnej (*LDA*), kwadratowej analizy dyskryminacyjnej (*QDA*), metody *k-NN* (*k*–najbliższych sąsiadów), drzew decyzyjnych czy różnego rodzaju sieci neuronowych. Możliwość wizualizacji wyników pozwala na łatwiejszą analizę otrzymanych wyników. W celu oceny skuteczności wbudowanego w program *FaceRec* algorytmu identyfikacji przeprowadzono serie doświadczeń z wykorzystaniem następujących algorytmów klasyfikacji:

- a) liniowej analizy dyskryminacyjnej (*LDA*),
- b) metody *k-NN* (*k*–najbliższych sąsiadów) dla $k = 1$,
- c) metody *Nearest Mean*,
- d) metody analizy głównych składowych *PCA*,
- e) jednokierunkowej sieci neuronowej, algorytm wstecznej propagacji błędów, 5 warstw ukrytych.

Doświadczenia przeprowadzono z wykorzystaniem dwóch zbiorów danych:

- uczącego – składającego się z 69 obiektów zgrupowanych w 15 klasach, każdy z nich opisany poprzez 6-wymiarowy wektor cech,

- testowego – składającego się z 44 obiektów zgrupowanych w 15 klasach, opisanych 6-wymiarowymi wektorami cech.

Dla każdej wybranej metody klasyfikacji wyliczony został średni błąd klasyfikacji. Rezultaty przeprowadzonych doświadczeń przedstawiono w tabeli 4.

Tabela 4. Wyniki eksperymentów przeprowadzonych w pakiecie *PRTools*

Lp.	Metoda	Błąd klasyfikacji
1	Liniowa analiza dyskryminacyjna (LDA)	38%
2	Metoda $k - NN$	27%
3	<i>Nearest Mean</i>	38%
5	<i>PCA</i>	38%
6	<i>SSN</i>	88%

4.5.4. Pakiet *RapidMiner*

RapidMiner jest złożonym środowiskiem służącym do prowadzenia doświadczeń i analiz z dziedziny uczenia maszynowego (ang. *Machine learning*) oraz eksploracji danych (ang. *Data Mining*). Program powstał na Uniwersytecie w Dortmundzie. Do projektu stopniowo przystępowały inne organizacje, w wyniku czego możliwy był jego dalszy rozwój. Od roku 2004 program udostępniany jest na licencji GPL, a koordynacją prac nad jego rozwojem zajmuje się firma Rapid-I.

Program oferuje ponad 400 operatorów (funkcji) pozwalających na wyszukiwanie różnych powiązań między danymi, obsługuje popularne formaty używane przez inne programy (pliki tekstowe, arkusze Excela, pliki XML, formaty bazodanowe *PostgreSQL*, *Oracle* i in.), blokowa struktura pozwala na łatwą rozbudowę i konkretne implementacje klienckie. Udostępnia szerokie możliwości analizy danych, wizualizację zbiorów wejściowych i wyników.

Modelowanie procesów w *RapidMiner* odbywa się poprzez budowę drzewa, którego poszczególne węzły stanowią wybrane operatory. Odbywać się to może na dwa sposoby: za pomocą odpowiednich plików XML opisujących strukturę procesu lub z wykorzystaniem graficznego interfejsu udostępnianego przez program. Doświadczenia przeprowadzono z wykorzystaniem następujących algorytmów klasyfikacji:

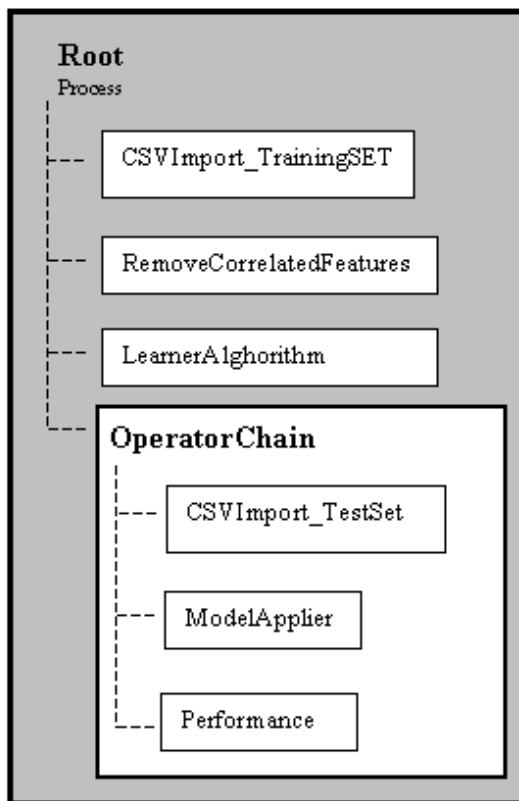
- metody $k - NN$ (k – najbliższych sąsiadów) dla $k = 1$,
- metody wektorów wspierających *SVM* (ang. *Support Vector Machine*),
- algorytmu drzewa decyzyjnego J48,
- naiwnego klasyfikatora bayesowskiego (ang. *NaiveBayes*),

e) sieci Bayesa.

Doświadczenia, podobnie jak wcześniej, przeprowadzono z wykorzystaniem dwóch zbiorów danych:

- uczącego – składającego się z 69 obiektów zgrupowanych w 15 klasach, każdy z nich opisany poprzez 6-wymiarowy wektor cech,
- testowego – składającego się z 44 obiektów zgrupowanych w 15 klasach, opisanych 6-wymiarowymi wektorami cech.

Model procesu dla każdego z przypadków przedstawiony został na rys. 4.3.



Rys. 4.3. Model procesu dla przeprowadzonych doświadczeń

W tabeli 5 zestawiono otrzymane rezultaty.

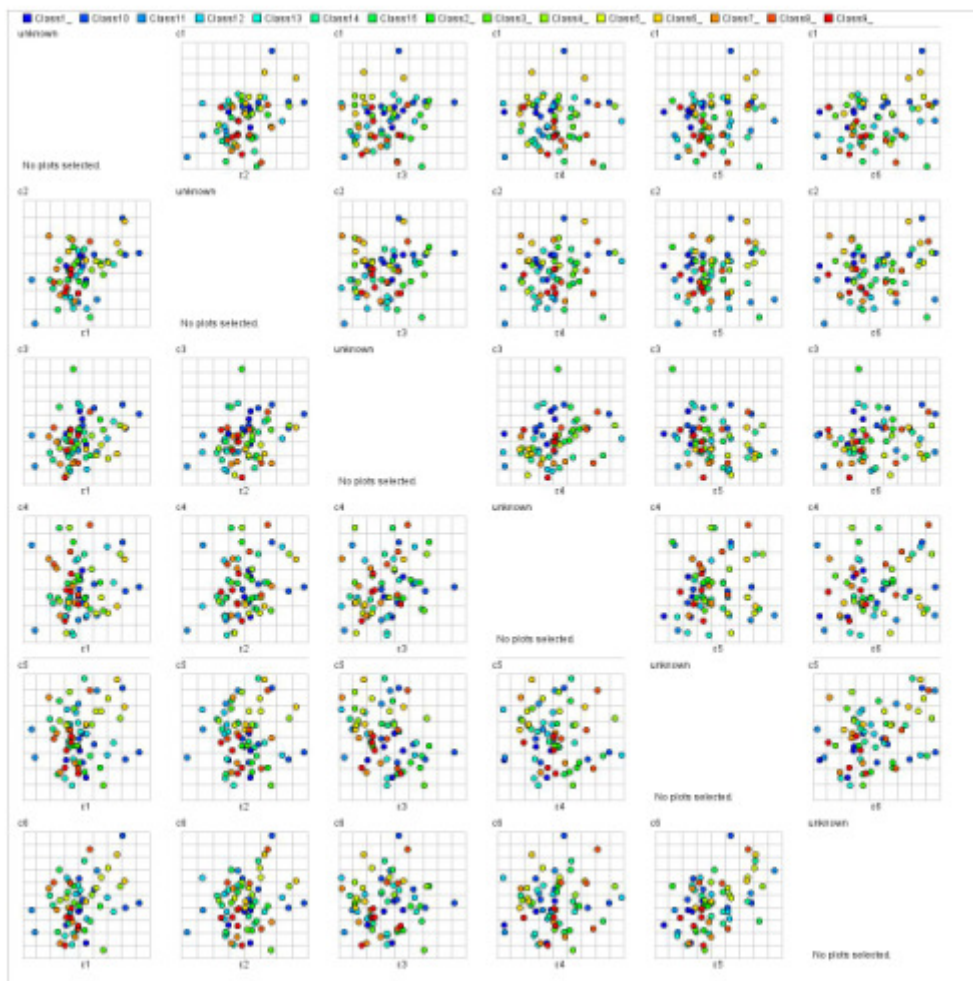
Tabela 5. Wyniki eksperymentów przeprowadzonych w programie *RapidMiner*

Lp.	Metoda	Błąd klasyfikacji
1	Metoda $k - NN$	27,27%
2	SVM	38,64%
3	Drzewo J48	52,27%
5	<i>NaiveBayes</i>	43,18%
6	Sieć Bayesa	63,64%

4.5.5. Analiza wyników

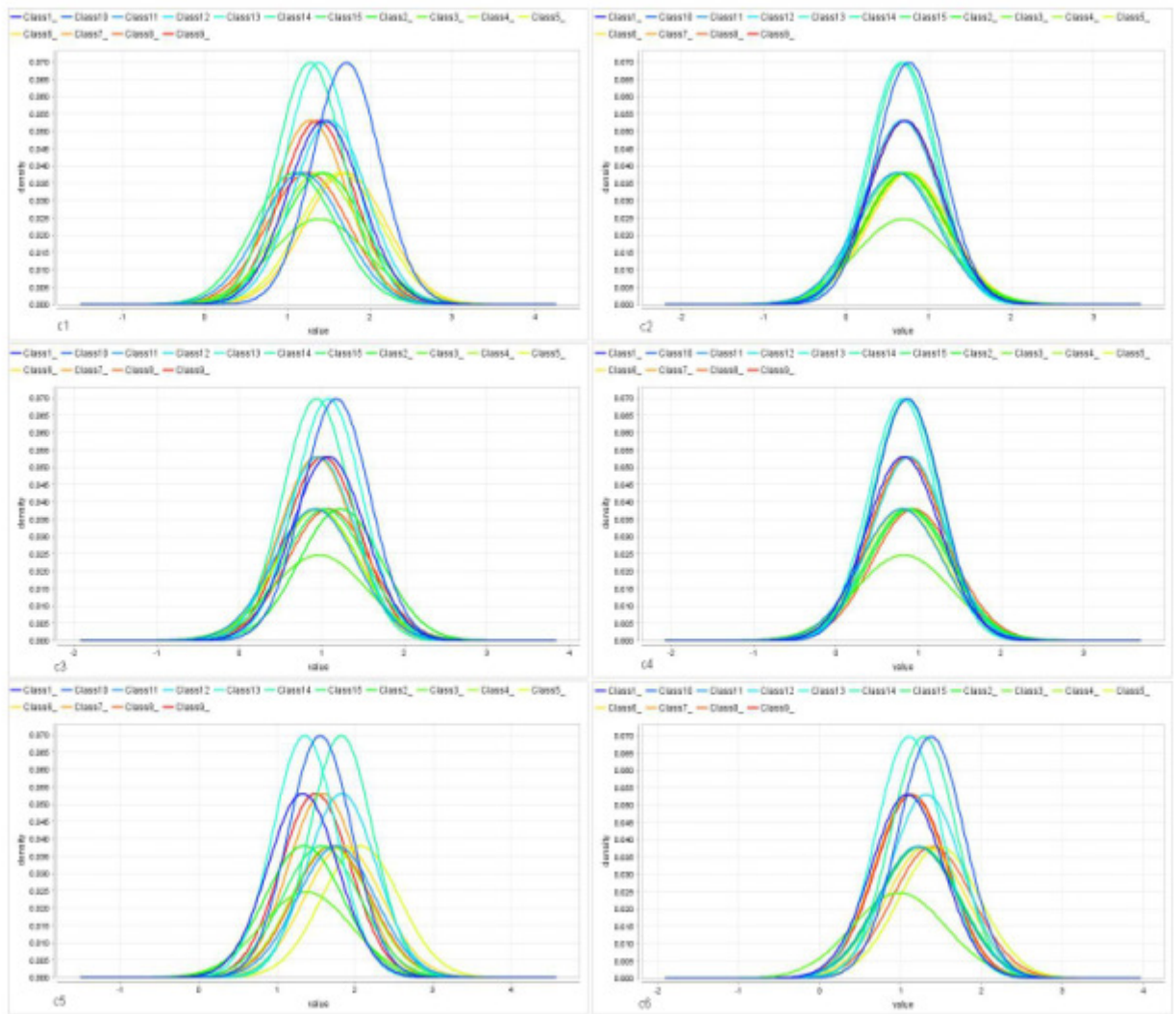
Analiza otrzymanych rezultatów potwierdza słuszność wybranej metody klasyfikacji. We wszystkich badanych przypadkach najlepsze wyniki osiągnięto korzystając z metody najbliższych sąsiadów. Błąd klasyfikacji wynosił w tym przypadku ok. 27% co potwierdziły wszystkie przeprowadzone doświadczenia. Próby uogólnienia problemu za pomocą pozostałych metod zakończyły się niepowodzeniem – uzyskane wyniki odbiegały od oczekiwanych. Najlepsze rezultaty uzyskano w tym przypadku z wykorzystaniem liniowych klasyfikatorów *LDA* i *PCA* oraz algorytmu *SVM*. Na taki rezultat złożyło się kilka czynników. Na rysunku 4.2. przedstawiono wykresy rozrzutu wybranych par zmiennych wykorzystywanych zbiorów danych.

Analiza wykresów rozrzutu pokazuje, iż wybrane zmienne niezbyt dobrze różnicują badane obiekty. Wynikać to może z wykorzystania zbyt małej liczby obiektów zbioru uczącego lub jego niereprezentatywności. Należy wziąć pod uwagę iż wykorzystane obrazy przedstawiały twarze w różnych, często znacznie się różniących pozach (obroty wokół osi, dodatkowe obiekty na twarzy, różne warunki oświetlenia). Rozbudowanie zbioru uczącego i wprowadzenie bardziej rygorystycznych warunków co do orientacji twarzy pozwoliłoby na uzyskanie znacznie lepszej separacji zmiennych. Analiza wykresów dystrybucji (rys. 4.3.) pozwala na wskazanie tych cech które najlepiej różnicują poszczególne klasy obiektów. Są to zmienne: *c1*, *c3*, *c5* oraz *c6*.



Rys. 4.4. Wykresy rozrzutu wybranych par zmiennych zbioru uczącego

Ocenę jakości identyfikacji dla poszczególnych klas można przeprowadzić w oparciu o macierz kontyngencji (ang. *confusion matrix*).



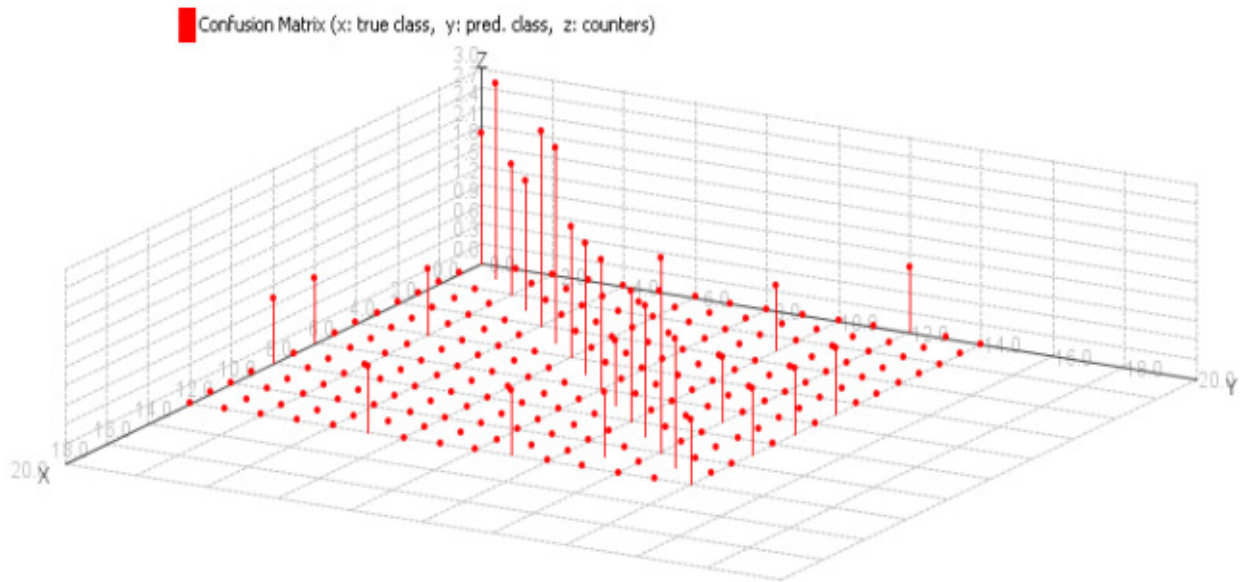
Rys. 4.5. Dystrybucje poszczególnych zmiennych zbioru uczącego

classification_error: 27.27%

	true Class1_	true Class10	true Class11	true Class12	true Class13	true Class14	true Class15	true Class2_	true Class3_	true Class4_	true Class5_	true Class6_	true Class7_	true Class8_	true Class9_	class precision
pred. Class1_	2	0	0	0	0	0	0	0	1	0	1	0	0	0	0	50.00%
pred. Class10	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Class11	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	66.67%
pred. Class12	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Class13	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	100.00%
pred. Class14	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	75.00%
pred. Class15	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	100.00%
pred. Class2_	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	100.00%
pred. Class3_	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	100.00%
pred. Class4_	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	50.00%
pred. Class5_	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	66.67%
pred. Class6_	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	66.67%
pred. Class7_	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	60.00%
pred. Class8_	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	100.00%
pred. Class9_	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	25.00%
class recall	66.67%	100.00%	100.00%	66.67%	100.00%	100.00%	66.67%	66.67%	66.67%	33.33%	66.67%	66.67%	100.00%	66.67%	33.33%	

Rys. 4.6. Macierz kontyngencji dla alg. k – NN. Błąd klasyfikacji 27,27%.

W zależności od klasy czułość metody waha się w granicach od 33% (*Class4*, *Class9*) do 100% (*Class10*, *Class13*). Najgorszą specyficznością wykazała się klasa 9 – tylko 25%, osiągając największą liczbę trafień.



Rys. 4.7. Macierz kontyngencji dla alg. k – NN – reprezentacja graficzna

5. Podsumowanie

W niniejszej pracy opisano proces tworzenia systemu identyfikacji osób w oparciu o wybrane cechy osobnicze twarzy na podstawie zaproponowanego przez autorów rozwiązania. Na początku przedstawiono podstawowe pojęcia z dziedziny rozpoznawania twarzy: opisano możliwości biometrycznego opisu kształtu twarzy, zdefiniowano pojęcie identyfikacji, omówiono podstawowe struktury systemów rozpoznawania twarzy, przedstawiono kolejne etapy automatycznej identyfikacji twarzy, dokonano przeglądu istniejących metod, opisano istniejące implementacje. W dalszej części przedstawiona została wybrana metodologia: opis realizacji poszczególnych etapów projektu, poczynając od budowy modelu „*skin-color detection*” poprzez lokalizację twarzy, ekstrakcję cech, kończąc na właściwej identyfikacji. Opisano strukturę zaimplementowanej bazy danych, sposób standaryzacji zawartych w niej danych. Na końcu przedstawiono wyniki przeprowadzonych eksperymentów i dokonano ich analizy.

Założenia jakie przyjęto przy rozpoczęciu pracy nad systemem zostały w dużej mierze zrealizowane. Zaprojektowany system rozwiązuje zadania związane z detekcją i lokalizacją twarzy na kolorowych obrazach, umożliwia ekstrakcję wybranych cech twarzy, identyfikację na tej podstawie osób, oferuje funkcjonalności związane z budową i zarządzaniem bazą danych. Należy zaznaczyć, że przyjęte w trakcie jego tworzenia analityczne podejście do tematu identyfikacji nie jest zbyt rozpowszechnione w komercyjnych rozwiązaniach oferujących automatyczną identyfikację osób ze względu na trudności związane przede wszystkim z precyzyjną lokalizacją wybranych charakterystycznych punktów twarzy. Dlatego rozwiązanie to znajduje zastosowanie przede wszystkim w interaktywnych systemach identyfikacji osób. W trakcie pracy nad systemem, autorzy mieli okazję zapoznać się z istniejącymi metodami realizującymi w lepszym lub gorszym stopniu poszczególne etapy całego procesu identyfikacji, dokonując w niektórych przypadkach próby ich usprawnienia.

Otrzymane wyniki potwierdziły skuteczność stosowania modelu „*skin-color detection*” do wstępnej detekcji obszarów ludzkiej skóry. Okazało się, że mimo ograniczonej ilości wykorzystanych próbek skóry, zbudowany model dobrze radził sobie z postawionym mu zadaniem. W przypadku testowej bazy *PRIMA* (na podstawie której został on zbudowany) jego skuteczność była bardzo wysoka, niezależnie od koloru skóry badanej osoby, choć w tym wypadku całe zadanie było nieco ułatwione ze względu na nieskomplikowaną fakturę tła i porównywalne warunki akwizycji poszczególnych zdjęć. Próba wykorzystania modelu do

lokalizacji obszarów ludzkiej skóry na obrazach z bazy *CALTECH* również dała pozytywne rezultaty, mimo tego, iż model zbudowany został na podstawie próbek pochodzących z zupełnie innej bazy. Ze względu jednak na bardzo zróżnicowane warunki ich akwizycji, zdarzały się przypadki w których metoda zawodziła. Najczęściej związane to było z bardzo silnym oświetleniem lub niedoświetleniem obszaru twarzy. Doświadczenia pokazały, że kluczową sprawą dla dalszego procesowania było właściwe dobranie progu segmentacji. W programie przewidziano dwie metody jego wyznaczenia: poprzez ręczne ustalenie wartości progu przez użytkownika lub jego adaptacyjne wyliczenie na podstawie różnicy powierzchni obszarów skóry i pozostałych. W przypadku bazy *PRIMA* obie metody dały pozytywne rezultaty, w przypadku bazy *CALTECH* najlepsze rezultaty osiągnięto ustalając jego wartość na stałym poziomie. Optymalnym rozwiązaniem wydaje się ustalanie stałej wartości progu w zależności od procesowanej bazy zdjęć – daje to gwarancje otrzymywania najlepszych rezultatów. Rezultat segmentacji w dużej mierze przekłada się na kolejny etap lokalizacji twarzy – wyznaczanie jej owalu z wykorzystaniem deformowalnego modelu eliptycznego. Poprawnie wyznaczona ramka twarzy znacznie zwiększa skuteczność procedur ekstrakcji cech. Do wyznaczenia poziomego położenia linii oczu, nosa oraz ust wykorzystano metodę projekcji horyzontalnej obrazu gradientowego twarzy. Otrzymane wyniki potwierdziły jej skuteczność – modyfikując odpowiednio kształt funkcji okna można było uzyskać bardzo dokładne wyznaczenie ich położenia. Zaobserwowane przypadki pomyłek związane były najczęściej z niezupełnością obrazu gradientowego twarzy oraz niewłaściwymi proporcjami wyznaczonego obszaru twarzy. W przypadku lokalizacji ramek oczu nie miało to najczęściej większego wpływu na dalszy przebieg procesu, pod warunkiem, że różnica nie była zbyt duża.

Wybranie metody wzorców do lokalizacji ramek oczu pozwoliło na uproszczenie całego procesu wyznaczenia ich środków, nie wpływając na jego skuteczność. W przeważającej większości przypadków otrzymane rezultaty były zgodne z oczekiwanymi. Należy jednak pamiętać, że możliwości wykorzystania metody (nie tylko do lokalizacji regionów oczu) ograniczone są koniecznością posiadania wiedzy na temat przybliżonej wielkości poszukiwanego obiektu, w mniejszym stopniu jego orientacji. Pozwala to na dopasowywanie obiektu do wielkości wzorca lub na odwrót. W przypadku twarzy wystarcza wiedza o jej całkowitych rozmiarach (wymiary owalu) – na tej podstawie możliwe jest dopasowanie wielkości wzorców według przyjętej metody. Błędy na tym etapie to wynik przede wszystkim nieprecyzyjnego wyznaczenia ramki twarzy. Możliwe są dwa przypadki: owal nie wypełnia w całości ramki lub ramka obejmuje jedynie fragment twarzy. W obu

efektem jest zaburzenie poprawnych proporcji twarzy – w rezultacie nieprecyzyjne wyznaczenie wielkości wzorca. Pewnym rozwiązaniem jest wykorzystanie tzw. piramidy obrazów czyli zbioru odpowiednio przeskalowanych (np. +/- 5% względem wyznaczonego rozmiaru bazowego) wzorców i sekwencyjne przeszukiwanie obszaru z ich wykorzystaniem. Oprócz kryterium wielkościowego na wynik lokalizacji mogą mieć wpływ takie czynniki jak obecność okularów (w przypadku testowej bazy nie miało to jednak miejsca), słaba jakość zdjęć twarzy (niedoświetlenie obszaru oka) lub inne (np. zamknięte powieki, duże odchylenia twarzy) mogące wpłynąć na wartość wzajemnej korelacji pomiędzy wzorcem a wybranym fragmentem przeszukiwanego obszaru. Wiedza na temat przybliżonej wielkości oczu wymagana jest także w przypadku algorytmu *Daugmana*, wykorzystywanego do wyznaczenia ich środków, w którym parametrami wejściowymi są m.in. minimalna i maksymalna wartość promienia źrenicy. Szczególnie wrażliwym na jakość wejściowego obrazu twarzy okazał się algorytm detekcji kształtu (szerokości) nosa. Wynika to z faktu, iż podstawową informacją, na której bazuje jest gradientowy obraz twarzy. Przełożyło się to na rezultaty jego działania. W przypadku mało kontrastowych i słabych jakościowo obrazów twarzy nie było możliwości odtworzenia krawędzi nosa – w rezultacie wyznaczenia jego kształtu. Miało to miejsce w kilku testowanych przypadkach. Innym algorytmem, którego działanie oparte było w dużej mierze na analizie obrazu gradientowego była procedura detekcji konturów twarzy. W tym jednak przypadku ewentualne niedoskonałości mapy krawędzi obszaru czy obecność zakłóceń mogły zostać skompensowane na drodze estymacji parametrów krzywej opisującej kontur twarzy bazując na dostępnych danych.

Osiągnięty wynik identyfikacji (błąd klasyfikacji 27%) można uznać za dostateczny. Jak wykazały to przeprowadzone doświadczenia, na taki wynik wpływ miało kilka czynników. Jednym z nich była metoda budowy wektora cech opisującego poszczególne twarze. Poszczególne składowe wyliczane były na podstawie wyznaczonych w trakcie etapu ekstrakcji cech współrzędnych charakterystycznych punktów twarzy. Ich wybór i ilość determinowały sposób tworzenia zależności, będących poszczególnymi składowymi wektora cech. Wszelkie powstałe na tym etapie błędy lokalizacji przekładały się na końcową postać wektora. Nieudane próby uogólnienia problemu za pomocą innych metod klasyfikacji można, oprócz powyższego, wytłumaczyć zbyt łagodnymi w stosunku do zastosowanej metody kryteriami, które musiały spełniać analizowane twarze, zbyt małą liczbą obiektów zbioru uczącego lub jego niereprezentatywnością. Doświadczenia przeprowadzone na testowej bazie zawierającej dane 15 różnych osób wykazały dobrą skuteczność metody w przypadku rozpoznawania osób, których orientacja pozostawała ustalona, np. frontem do oka kamery.

Możliwości rozwoju systemu, to przede wszystkim eliminacja opisanych wyżej problemów. Pod względem koncepcyjnym rozszerzenie listy punktów pomiarowych np. o krzywe opisujące całość owalu twarzy, kształt brwi a w konsekwencji zastosowanie innych metod ich ekstrakcji wykorzystujących w większym stopniu metody oparte na deformowalnych krzywych. Uzyskane w ten sposób dane pozwolą być może na lepsze różnicowanie poszczególnych obiektów i zastosowanie innych metod ich klasyfikacji. Na etapie lokalizacji twarzy warto rozważenia wydaje się zastosowanie dodatkowego wzorca centralnej części twarzy umożliwiającego lokalizację tej części twarzy. Umożliwi to bardziej precyzyjne określenie punktu startowego dla modelu deformowalnego wyznaczającego owal twarzy. Pod względem funkcjonalnym to rozszerzenie zakresu działania samej aplikacji – rozbudowa istniejącego szkieletu o nowe możliwości np. związane z pracą w trybie *on-line*.

Bibliografia

1. Chang T. C., Huang T. S., Novak C., Facial feature extraction from color images, in Int. Conf. Pattern Recognition, (Israel), Oct. 1994, pp. 39-43.
2. Chang H., Robles U., Face Detection, EE368 Final Project Report, 2000.
3. Colmenarez A. J, Huang T. S., Frontal view face detection, in SPIE, vol. 2501, 1995, pp. 90-98.
4. Craw I, Ellis H., Lishman J., Automatic extraction of face features, Pattern Recognition Letters, vol. 5, 1987, pp. 183-187.
5. Daugman J., How iris recognition works, Proceedings of 2002 International Conference on Image Processing, Vol. 1, 2002.
6. Fuksa S., Systemy wizyjne – wykłady, Państwowa Wyższa Szkoła Zawodowa w Tarnowie, 2002.
7. Gu H., Su G., Du Cheng., Feature Points Extraction from Faces, Image and Vision Computing New Zealand 2003.
8. Huang C. L., Chen C. W., Human facial feature extraction for face interpretation and recognition, Pattern Recognition, vol. 25, no. 12, 1992, pp. 1435-1444.
9. Jones M. J., Rehg J. M., Skin Color Modeling and Detection, Hewlett-Packard Company, June 2002.
10. Kampmann M., Estimation of the Chin and Cheek Contours for Precise Face Model Adaption, International Conference on Image Processing (ICIP--97), Santa Barbara, CA, 1997.
11. Kirby M., Sirovich I., Application of the Karhunen-Loeve procedure for the characterization of human faces, IEEE Trans. Patt. Anal. Mach. Intel., vol. 12, 1990, pp. 103-108.
12. Kukharev G., Kuzminski A., Techniki Biometryczne Część I: Metody rozpoznawania twarzy, Pracownia Poligraficzna WI Politechnika Szczecińska, Szczecin, 2003
13. Kumar R.T., Raja S.K., Ramakrishnan A.G., Eye detection using color cues and projection functions, Proceedings 2002 International Conference on Image Processing ICIP, Rochester, 2002, pages Vol.3 337-340.
14. Lee J.Y., Yoo S.I., An Elliptical Boundary Model for Skin Color Detection, In CISST'02, 2002.
15. Nasiri J.A., Khanchi S., Pourreza H.R., Eye Detection Algorithm on Facial Color Images, ams, 2008 Second Asia International Conference on Modelling & Simulation (AMS), 2008, pp. 344-349.
16. Peng K., Chen L., Ruan S., Kukharev G., A Robust Alghorithm for Eye Detection on Gray Intensity Face without Spectacles, Journal of Computer Science & Technology, 2005.

17. Satoh S., Nakamura Y., Kanade T., Name-It: Naming and Detecting Faces in News Videos, IEEE MultiMedia, Vol. 6, No. 1, January-March, 1999, pp. 22-35.
18. Sedlacek M., Evaluation of RGB and HSV model in Human Faces Detection, Central European Seminar on Computer Graphics, 2004.
19. Sim T., Sukthar R., Mullin M., Baluja S., High Performance Memory-based Face Recognition for Visitor Identification, Just Research Technical Report, 1999
20. Tadeusiewicz R., Flasiński M., Rozpoznawanie obrazów, PWN, 1991.
21. Tadeusiewicz R., Korohoda P., Komputerowa analiza i przetwarzanie obrazów, Wydawnictwo Fundacji Postępu Telekomunikacji, 1997.
22. Tsekeridou S., Pitas I., Facial feature extraction in frontal views using biometric analogies, in Proc. of the IX European Signal Processing Conference, vol. I, 1998, pp. 315-318.
23. Utkin M., Identyfikacja biometryczna, Młody Technik, nr 10/2005, (<http://rozrywka.onet.pl/prasa/3920,1257701,1,druk.html>).
24. Vezhnevets V., Degtiareva A., Robust and Accurate Eye Contour Extraction, Faculty of Computational Mathematics and Cybernetics Moscow State University, 2003.
25. Vezhnevets V., Sazonov V., Andreeva A., A Survey on Pixel-Based Skin Color Detection Techniques, Proc. Graphicon-2003, Moscow, Russia, September 2003, pp. 85-92.
26. Yang G., Huang T. S., Human face detection in a complex background, Pattern Recognition, vol. 27, no. 1, 1994, pp. 53-63.
27. Yang M.-H., Ahuja N., Gaussian Mixture Model for Human Skin Color and its Applications in Image and Video Databases, In the 1999 SPIE/EI&T Storage and Retrieval for Image and Video Databases, San Jose, January 1999, pp. 458-466.
28. Yang M.-H., Kriegman D., Ahuja N., Detecting Faces in Images: A Survey, In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, 1, 2002, pp. 34-58
29. Yow K. C., Cipolla R., Feature-based human face detection, Image and Vision Computing, vol. 15, 1997, pp. 713-735.
30. Yuille A. L., Hallinan P. W., Cohen D. S., Feature extraction from faces using deformable templates, Int. Jour. Computer Vision, vol. 8, no. 2, 1992, pp. 99-111.

Źródła internetowe:

1. <http://4safe.pl>
2. <http://www.betaface.com/demo.aspx?lang=pl>
3. <http://www.betaface.com/products.aspx?lang=pl>
4. <http://code.google.com/p/aforge>
5. <http://www.compas.com.pl/biometria.html>

6. http://www.europe-nikon.com/press_room/news.html?locale=pl_PL&bandwidth=broad&id=236&type_index=2
7. http://www.fujifilm.pl/3295_3309.htm
8. http://www.l1id.com/index.php?option=com_content&task=view&id=30&Itemid=133
(FaceIt Argus)
9. www.nikon.biz.pl
10. <http://www.nikoncoolpix.info/main.php?lang=PL&country=PL>
11. <http://www.polixel.com.pl/p29-watch-id.htm>
12. <http://www.snappy.pl/info.html>
13. <http://zpiro.wi.ps.pl/index.php?id=121> (Visitor Identification)
14. <http://zpiro.wi.ps.pl/index.php?id=31>

Spis rysunków

Rys. 1.1. Autofokus w aparatach Nikon (Źródło: [WWW-9]).....	8
Rys. 1.2. Wybrane cechy biometryczne twarzy	10
Rys. 1.3. Osie obrotów twarzy w przestrzeni 3D	11
Rys. 1.4. Tryby pracy systemów rozpoznawania i ich zastosowanie.....	12
Rys. 1.5. Ogólny schemat systemu rozpoznawania twarzy w trybie weryfikacji	13
Rys. 1.6. Ogólny schemat systemu rozpoznawania twarzy w trybie identyfikacji	13
Rys. 1.7. Ogólna struktura systemu rozpoznawania.....	14
Rys. 1.8. Proces automatycznej identyfikacji twarzy (Źródło: [WWW-1]).....	15
Rys. 1.9. Wpływ oświetlenia na obraz twarzy	16
Rys. 1.10. Części twarzy a ilość dostarczanych przez nie informacji nt. cech osobniczych....	19
Rys. 1.11. Opis twarzy za pomocą różnych kategorii cech (rys. na podst. [12])	20
Rys. 1.12. Błędne położenie i sklasyfikowanie elementu ciągu uczącego (rys. na podst. [20])	23
Rys. 1.13. Zasada działania metody α NN (rys. na podst. [20]).....	24
Rys. 1.14. Struktura systemu rozpoznawania on-line (rys. na podst. [12]).....	30
Rys. 1.15. Struktura systemu „Visitor Identification” (rys. na podst. [12])	32
Rys. 1.16. Proces rozpoznawania obrazu w systemie „Visitor Identification” (Źródło: [12]).	33
Rys. 1.17. Funkcja przekroczenia progu	33
Rys. 1.18. System „Visitor Identification” z dodatkową weryfikacją rezultatu (Źródło: [12])	34
Rys. 1.19. Struktura systemu „Access Control” (rys. na podst. [12])	36
Rys. 1.20. Struktura procesu „Name-It” (Źródło: [17])	39
Rys. 1.21. Struktura systemu „Name-It” (Źródło: [12]).....	40
Rys. 1.22. Wykrywanie twarzy w programie BetaFace. Źródło: [WWW-3].....	41
Rys. 1.23. 84 punkty opisujące twarz w systemie BetaFace. Źródło: [WWW-3].....	42
Rys. 1.24. Porównanie twarzy autorki z twarzami w systemie BetaFace	42
Rys. 1.25. System IDWatch – okno rozpoznawania Źródło: [WWW-11].....	43
Rys. 1.26. Rozpoznanie twarzy w systemie FaceIt® Argus. Źródło [WWW-8]	44
Rys. 1.27. Rozpoznana twarz w systemie „Visitor Identification”. Źródło [WWW-13].....	45
Rys. 1.28. System „Visitor Identification”. Źródło [WWW-13].....	45
Rys. 1.29. Kamera SNAPPY. Źródło: [WWW-12].....	46
Rys. 2.1. Schemat funkcjonalny systemu	48
Rys. 2.2. Rozkład wartości składowych Cb, Cr (po lewej) oraz R, G (po prawej) dla obszarów różnych twarzy	51
Rys. 2.3. Składowe kolorowego obrazu w przestrzeni RG	52
Rys. 2.4. Składowe kolorowego obrazu w przestrzeni HSV	53
Rys. 2.5. Składowe kolorowego obrazu w przestrzeni YCbCr	54
Rys. 2.6. Składowe kolorowego obrazu w przestrzeni I1I2I3	54
Rys. 2.7. Rozkład kolorów ludzkiej twarzy w chromatycznej przestrzeni barw.....	55
Rys. 2.8. Aproksymacja rozkładu koloru ludzkiej skóry modelem Gaussa (przestrzeń RG) ..	56
Rys. 2.9. Histogram rozkładu koloru ludzkiej skóry z zaznaczonym punktem skupienia obszarów nie reprezentujących ludzkiej skóry (a), (b) aproksymacja rozkładu modelem Gaussa (Źródło: [14])	57
Rys. 2.10. Histogram rozkładu koloru ludzkiej skóry z zaznaczonym punktem skupienia obszarów nie reprezentujących ludzkiej skóry (a), (b) aproksymacja rozkładu modelem EBM (Źródło: [14])	58

Rys. 2.11. Schemat postępowania algorytmu detekcji i lokalizacji twarzy	60
Rys. 2.12. Obraz wejściowy (a) i obraz podobieństw (b). Obszary ciemniejsze reprezentują regiony skóry	61
Rys. 2.13. Obraz wejściowy (a), obraz podobieństw (dopełnienie) (b), obraz po segmentacji (c) (przestrzeń YCbCr)	62
Rys. 2.14. Obraz wejściowy (a), obraz po segmentacji (b), obraz po indeksacji (c)	63
Rys. 2.15. Deformowalny model eliptyczny	64
Rys. 2.16. Zasada działania modelu deformowalnego	64
Rys. 2.17. Wstępne wyznaczanie owalu twarzy; (a) zakwalifikowany region obrazu binarnego z zaznaczoną pozycją startową modelu i wyznaczonymi punktami końcowymi; wyznaczony owal na obrazie binarnym (b) i wejściowym (c)	66
Rys. 2.18. Wyznaczanie regionu twarzy	67
Rys. 2.19. Schemat postępowania w algorytmie lokalizacji środków źrenic	68
Rys. 2.20. Projekcja wertykalna obrazu gradientowego twarzy (b) i rezultat określenia szerokości twarzy (a)	69
Rys. 2.21. Obliczanie położenia linii oczu na obrazie twarzy; (a) obraz wejściowy z linią oczu i obraz gradientowy; (b) - projekcja horyzontalna obszaru twarzy $H(y)$ i $Hflt(y)$ – funkcja po filtracji dolnoprzepustowej; (c) – rezultat pokropkowego przemnożenia funkcji $H(y)$ i $W(y)$	70
Rys. 2.22. Obliczanie położenia linii oczu dla każdej połowy twarzy; (a) obrazy wejściowe z linią oczu i ich odpowiadające im obrazy gradientowe; (b) – projekcja horyzontalna lewej i prawej połowy twarzy $H(y)$ i $Hflt(y)$ – funkcja projekcji po filtracji dolnoprzepustowej; (c) – rezultat pokropkowego przemnożenia funkcji $H(y)$ i $W(y)$	71
Rys. 2.23. Projekcja wertykalna obszaru oczu	72
Rys. 2.24. Detekcja regionów oczu dla każdej połowy twarzy metodą „ <i>template matching</i> ”; (a) – wydzielony obszar poszukiwań z obrazem wzorca i rezultatem; (b) – wykres znormalizowanej wzajemnej korelacji wzorca z obszarem wewnątrz pola maski.....	73
Rys. 2.25. Detekcja regionów oczu dla każdej połowy twarzy metodą „ <i>template matching</i> ” w przypadku znacznego nachylenia twarzy i niedokładnego dopasowania wzorców ; (a) – wydzielony obszar poszukiwań z obrazem wzorca i rezultatem; (b) – wykres znormalizowanej wzajemnej korelacji wzorca z obszarem wewnątrz pola maski.....	74
Rys. 2.26. Rezultaty detekcji oczu	74
Rys. 2.27. Ludzkie oko	75
Rys. 2.28. Funkcja wagowa $W(x,y)$ wzmacniająca obszary koliste w centrum maski konwolucji	76
Rys. 2.29. Modyfikacja regionów oczu; (a) – obraz wejściowy, składowa R obrazu oka; (b) – obraz oka po filtracji i konwolucji; (c) – przybliżony położenie środka źrenicy, wyliczone jako środek masy obszaru najciemniejszych pikseli obrazu (b)	76
Rys. 2.30. Rezultaty lokalizacji środków źrenic	77
Rys. 2.31. Zasada określania kąta nachylenia twarzy w płaszczyźnie XY	78
Rys. 2.32. Normalizacja orientacji twarzy w płaszczyźnie XY; (a) – wejściowy obraz z zaznaczoną ramką twarzy; (b) – wydzielony obszar twarzy z wyliczonymi współrzędnymi oczu i osią symetrii; (c) – kompensacja nachylenia poprzez obrót obrazu wejściowego o kąt $-\varphi$ wyznaczona nowa ramka obszaru twarzy; (d) – znormalizowany obszar twarzy z wyliczonymi nowymi współrzędnymi oczu.....	79
Rys. 2.33. Projekcja horyzontalna obszaru ust. Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W, wynik końcowy (c)	81

Rys. 2.34. Projekcja horyzontalna obszaru ust twarzy pochylonej (15st). Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W, wynik końcowy (c)	81
Rys. 2.35. Wyznaczone regiony ust i funkcje projekcji wertykalnej dla wybranych twarzy...	82
Rys. 2.36. Rezultat modyfikacji regionu ust i wyznaczona szerokość ust	83
Rys. 2.37. Region poszukiwań linii nozdrzy; od góry ograniczony linią oczu i współzrędnymi środków źrenic, od dołu linią ust.....	84
Rys. 2.38. Wyznaczanie linii nozdrzy. Przebiegi funkcji projekcji horyzontalnej H (oryginalny i przefiltrowany) (a), (b) rezultat przemnożenia funkcji projekcji przez funkcję okna W, wynik końcowy (c)	85
Rys. 2.39. Wyznaczanie linii nozdrzy dla innej twarzy	85
Rys. 2.40. Wyznaczone parametry nosa. Linia niebieska to symetryczna odcinka łączącego środki źrenic, która w punkcie przecięcia się z linią nozdrzy wyznacza przybliżone położenie czubka nosa; (b) – przykład działania procedury w przypadku niewidocznych otworów nosowych i nietypowego kształtu nosa	86
Rys. 2.41. Wyznaczanie szerokości nosa. Rezultat końcowy (a), wynik wertykalnej detekcji krawędzi (b) i przebieg funkcji projekcji wertykalnej obszaru z zaznaczonymi dwoma maksimumami wyznaczającymi granice nosa.....	86
Rys. 2.42. Rezultat działania procedury w przypadku niewidocznych otworów nosowych....	87
Rys. 2.43. Szablon defomowalny do wyznaczenia szerokości twarzy.....	88
Rys. 2.44. Twarze z naniesionymi pozycjami startowymi modelu	89
Rys. 2.45. Wertykalna mapa krawędzi twarzy (a) z naniesionymi punktami zlokalizowanej krawędzi i rezultat pomiaru szerokości (b).....	90
Rys. 2.46. Wertykalna mapa krawędzi twarzy (a) z naniesionymi punktami zlokalizowanej krawędzi i rezultat pomiaru szerokości (b) dla innej twarzy.....	90
Rys. 2.47. Zasada wyznaczania centralnej części twarzy.....	91
Rys. 2.48. Normalizacja skali wydzielonego obszaru centralnej części twarzy; (a) różnej wielkości wejściowe obrazy twarzy; (b) wydzielone centralne obszary twarzy; (c) obszary wydzielone po normalizacji skali.....	92
Rys. 2.49. Rezultat lokalizacji cech twarzy	93
Rys. 2.50. Składowe wynikowego wektora cech	94
Rys. 2.51. Struktura bazy danych systemu	96
Rys. 2.52. Rezultat zastosowania funkcji korelacji fazowej $r_{fg}(m, n)$: (a) – obraz $f(m, n)$; (b) – obraz $g(m, n)$; (c) – funkcja korelacji fazowej pomiędzy dwoma identycznymi obrazami (obraz $f(m, n)$); (d) – funkcja korelacji fazowej pomiędzy obrazami $f(m, n)$ i $g(m, n)$..	100
Rys. 2.53. Rezultat zastosowania korelacji fazowej (c) do badania stopnia podobieństwa obrazu oryginalnego (a) do zaszumionego odpowiednika (b). Wartości współczynników korelacji: fazowego – 0,2868; zwyczajny – 0,9598	101
Rys. 2.54. Funkcja korelacji fazowej r_{fg} centralnych części twarzy tej samej osoby z okularami (a) i bez nich (b). Współczynniki korelacji: fazowy = 0,61; zwyczajny = 0,9231	101
Rys. 2.55. Funkcja korelacji fazowej r_{fg} centralnych części twarzy dwóch różnych osób. Współczynniki korelacji: fazowy = 0,2622; zwyczajny = 0,7423	102
Rys. 2.56. Funkcja korelacji fazowej r_{fg} centralnych części twarzy dwóch różnych osób – twarze z okularami. Współczynniki korelacji: fazowy = 0,2089; zwyczajny = 0,7439. 102	
Rys. 3.1. Ogólny schemat systemu.....	105
Rys. 3.2. Podstawowy widok GUI programu w trybie analizy obrazów twarzy	107
Rys. 3.3. Lokalizacja twarzy na obrazie źródłowym.....	109
Rys. 3.4. Interaktywna budowa wektora cech	110
Rys. 3.5. Panel obsługi bazy danych twarzy	111

Rys. 4.1. Próbkę zdjęć przedstawiających tę samą osobę. Baza danych <i>PRIMA</i>	113
Rys. 4.2. Próbkę zdjęć przedstawiających tę samą osobę. Baza danych <i>CALTECH</i>	113
Rys. 4.3. Model procesu dla przeprowadzonych doświadczeń.....	140
Rys. 4.4. Wykresy rozrzutu wybranych par zmiennych zbioru uczącego.....	142
Rys. 4.5. Dystrybucje poszczególnych zmiennych zbioru uczącego	143
Rys. 4.6. Macierz kontyngencji dla alg. k – NN. Błąd klasyfikacji 27,27%.....	143
Rys. 4.7. Macierz kontyngencji dla alg. k – NN – reprezentacja graficzna	144

Spis tabel

Tabela 1. Rozmiary przestrzeni cech dla różnych metod (źródło [12])	21
Tabela 2. Ogólna klasyfikacja metod rozpoznawania twarzy	26
Tabela 3. Wyniki eksperymentów przeprowadzonych w programie <i>FaceRec</i>	138
Tabela 4. Wyniki eksperymentów przeprowadzonych w pakiecie <i>PRTools</i>	139
Tabela 5. Wyniki eksperymentów przeprowadzonych w programie <i>RapidMiner</i>	140