
Rozpoznawanie emocji z twarzy – podejście antropometryczne

Julia Jarkiewicz

Polsko-Japońska
Wyższa Szkoła
Technik Komputerowych
Warszawa
halofades@gmail.com

Rafał Kocielnik

Polsko-Japońska
Wyższa Szkoła
Technik Komputerowych
Warszawa
rafal.kocielnik@gmail.com

Streszczenie

Celem projektu jest detekcja, analiza oraz rozpoznawanie emocji z twarzy. System operuje na obrazach w skali szarości. Przy analizie wykorzystano kaskadowy klasyfikator cech Haar-podobnych, antropometryczny¹¹ model twarzy oraz hybrydowy sposób detekcji cech. System lokalizuje 17 punktów charakterystycznych analizowanej twarzy i bazując na ich rozkładzie jest w stanie automatycznie rozpoznać wyrażaną emocję. System został przetestowany na publicznie dostępnej bazie obrazów (Japanese Female Expression Database) JAFFE¹² uzyskując dokładność na poziomie 77% dla 7 podstawowych emocji przy użyciu różnych klasyfikatorów. Dzięki otwartej strukturze system może współpracować z dowolnym innym systemem HCI.

Słowa kluczowe

rozpoznawanie emocji, detekcja wyrazu twarzy, analiza emocji

Copyright Wydawnictwo PJWSTK Warszawa 2009
Kansei 2009
Interfejs użytkownika – Kansei w praktyce
ISBN 978-83-89244-78-9

¹¹ **Antropometria** – zespół technik i metod pomiarowych umożliwiających ścisłe badanie zróżnicowanych cech mierzalnych człowieka i ich zmienności w rozwoju osobniczym. Źródło: słownik medyczny www.biomedical.pl

¹² **The Japanese Female Facial Expression (JAFFE) Database** - <http://www.kasrl.org/jaffe.html>

1. Wstęp

Efektywne rozpoznawanie emocji zawsze stanowiło wyzwanie dla badaczy. Zadania takie jak rozpoznawanie emocji czy przedmiotów są proste dla człowieka, ale stanowią nie lada wyzwanie dla systemów komputerowych. Głównymi problemami są zmienne warunki oświetleniowe, zmiany koloru i transformacji. Efektywne wykrywanie twarzy ludzkich jest jednym z najpoważniejszych problemów analizy obrazu. Powoduje to, że skuteczna i wydajna lokalizacja cech twarzy w analizowanym obrazie oraz przełożenie ich na wyrażaną emocję jest jeszcze trudniejsze. Praca ta jest próbą utworzenia systemu komputerowego, który byłby w stanie automatycznie wykryć i rozpoznać cechy twarzy. W tym wypadku poszukiwanymi cechami są punkty charakterystyczne, umieszczone w wybranych miejscach na twarzy człowieka. Położenia tych punktów oraz odległości pomiędzy nimi zmieniają się podczas zmian wyrazu twarzy. Istnieje wiele mniej lub bardziej skutecznych rozwiązań pozwalających na detekcję i rozpoznawanie twarzy, jednakże tylko nieliczne rozwiązania łączą te cechy z systemem rozpoznawania emocji. Praca ta jest w dużej mierze oparta na rozwiązaniach zaprezentowanych w [1-3]. Do testów użyto bazy JAFFE, składającej się z 213 czarnobiałych zdjęć dziesięciu Japonek o rozmiarze 256x256 pikseli, ukazujących sześć podstawowych emocji: strach, smutek, złość, radość, zaskoczenie, odraza (każda z emocji w trzech wariantach), oraz twarz neutralną. Czarnobiałe zdjęcia o niskiej rozdzielczości stanowią wyzwanie nawet dla zaawansowanych technik.

2. Detekcja twarzy

Zanim będzie możliwe przystąpienie do wykrywania cech twarzy, konieczne jest precyzyjne określenie jej

położenia na analizowanym obrazie. Istnieje wiele różnych metod rozwiązania tego problemu. Dobry opis kilku z nich znajduje się w pracy [3]. Metoda, która znalazła zastosowanie w niniejszej pracy została opisana w kolejnych rozdziałach.

2.1 Kaskadowy klasyfikator cech Haar-podobnych

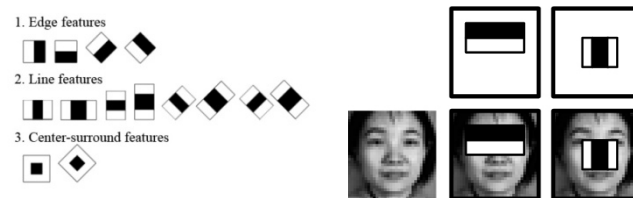
Zastosowane rozwiązanie wykorzystuje cechy Haar-podobne bazujące na bardzo prostej funkcji (Wzór 1).

$$H(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } 0 \leq t \leq 0,5 \\ -1 & \text{for } 0,5 \leq t < 1 \\ 0 & \text{for } t \geq 1 \end{cases}$$

wzór 1 – funkcja Haara

Metoda działa w różnych skalach. Dla zwiększenia wydajności, analiza rozpoczyna się od największej (najmniej dokładnej) skali, a w momencie odnalezienia wzorca, otaczający region jest przeszukiwany ponownie - tym razem w dokładniejszej skali. Podejście to można więc określić, jako odgórne. Cały proces jest powtarzany, aż do osiągnięcia najdokładniejszej skali. Metoda jest jedną z najpopularniejszych i najczęściej wykorzystywanych dzięki swojej efektywności i wydajności [2]. Kluczowymi czynnikami są trzy główne charakterystyki tego rozwiązania:

- **Klasyfikacja kaskadowa** – cały proces klasyfikacji jest podzielony na etapy, które następują po sobie, tworząc ciąg. Odrzucenie obrazu na jednym z etapów, powoduje, że sprawdzenie w dalszych etapach nie jest już konieczne. Podnosi to znacznie wydajność i zmniejsza wykorzystanie mocy obliczeniowej.



rysunek 1. Maski cech Haar-podobnych i ich ułożenie na przykładowym analizowanym obrazie [7]

- **Maski obrazu (słabe klasyfikatory Haar-podobne) oraz boosting** – każdy z etapów w opisanej wcześniej kaskadzie jest odrębnym, skomplikowanym klasyfikatorem bazującym na technice boostingu¹³ (Adaboost) [9]. Klasyfikator ten ma formę drzewa decyzyjnego z co najmniej dwoma liśćmi na każdym etapie. Końcowa ocena dokonywana jest jednak z wykorzystaniem cech Haar-podobnych. Są to w zasadzie proste maski obrazu, zaprojektowane do wykrycia różnych cech takich jak linie, rogi, itd. Rozszerzony zestaw takich masek, wykorzystywany w niniejszej pracy prezentuje rysunek (Rys. 1).
- **Wykorzystanie obrazu całkowego** – w celu detekcji wzorca w różnych położeniach, wykorzystywana jest technika przesuwanego okna. Problem odnajdywania obiektów w różnych rozmiarach, został rozwiązany dzięki użyciu obrazu

¹³ **Boosting** – jest ogólna metoda poprawy dokładności dowolnego algorytmu uczącego oparta o głosowanie z wagami. Źródło: [15]

całkowego¹⁴. Pozwala to na modyfikacje rozmiarów klasyfikatora, a nie całego obrazu [2].

Implementacja opisanego algorytmu wykorzystana w pracy jest częścią darmowej biblioteki przetwarzania obrazu - OpenCV[5]. Rezultaty lokalizacji prezentuje rysunek (Rys. 2)



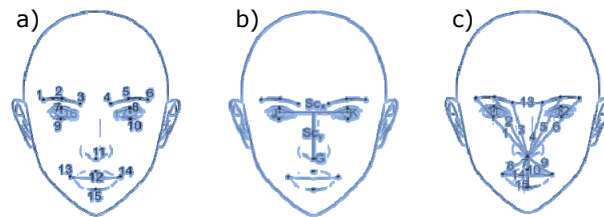
rysunek 2. Przykłady lokalizacji twarzy (baza JAFFE oraz Internet)

2.2 Normalizacja twarzy

Aktualnie system opiera się na dwóch typach normalizacji: oświetlenia oraz geometrycznej. Normalizacja oświetlenia znajduje swoje zastosowanie na różnych etapach analizy, od normalizacji dla całego obrazu do normalizacji dla konkretnych zidentyfikowanych obszarów zawierających elementy twarzy. Wykorzystywane są metody wyrównania, dopasowania histogramu oraz elementy metod Retinex[6]. Normalizacja geometryczna ma na celu skalowanie, przycinanie, bądź obracanie obrazu twarzy. W niniejszej pracy przeprowadzane są jednak tylko operacje przycinania oraz obrotu twarzy. Skalowanie nie jest zastosowane z uwagi na groźbę utraty detali, które są szczególnie istotne przy wyszukiwaniu cech twarzy. Jak można zaobserwować na rysunku (Rys. 2) twarze są lokalizowane dość niedokładnie oraz bez jakiegokolwiek informacji na temat ich rotacji. Konieczne

¹⁴ **Obraz całkowity** – jest zmodyfikowaną wersją obrazu pierwotnego, gdzie wartość w punkcie (x,y) jest sumą jasności wszystkich pikseli powyżej i na lewo od punktu (x,y) z punktem (x,y) włącznie. Źródło: [2]

jest zlokalizowanie twarzy w sposób jak najdokładniejszy, gdyż umożliwia to poprawne umiejscowienie elementów twarzy. W celu poprawienia niedokładności oraz uzyskania informacji dotyczącej rotacji twarzy wykorzystywana jest normalizacja geometryczna oparta o położenia środków oczu. Pozioma linia łącząca środki oczu, pomimo jej zmiennej długości dla różnych twarzy, praktycznie zawsze wyznacza oś symetrii twarzy (Rys. 3b). Kolejną ważną informacją jest rotacja twarzy, która jest wyliczana na podstawie pomiaru kąta pomiędzy środkami oczu (Rys. 3b).



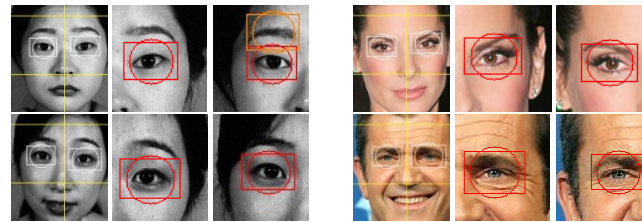
rysunek 3. Normalizacja odległości w modelu antropometrycznym a) lokalizacja punktów, b) normalizacja rotacji, c) przeliczenie odległości

3. Lokalizacja cech twarzy

Antropometria jest działem biologii zajmującym się pomiarami ludzkiego ciała i jego elementów. Autorzy pracy [1] stworzyli antropometryczny model ludzkiej twarzy dokonując pomiarów na 300 obrazach twarzy en face dla ponad 150 osób pochodzących z różnych części świata. Model może być wykorzystany dla lokalizacji elementów twarzy w oparciu o pozycje środków oczu. Na rysunku (Rys. 3a) pokazana została lokalizacja podstawowych punktów. Dobrze opisują one cechy twarzy, które zmieniają się wraz z wyrażaniem emocji.

3.1 Lokalizacja oczu

Środki oczu muszą być zlokalizowane bardzo dokładnie, jeśli mają służyć za podstawę modelu antropometrycznego. Niestety, większość metod lokalizacji oczu określa położenie źrenicy, a nie środka oka. Prawdziwa trudność pojawia się jednak w momencie, gdy oko jest przymrużone lub całkowicie zamknięte. Aby spełnić sformułowane wymagania oraz poradzić sobie ze wspomnianymi problemami zdecydowano się ponownie zastosować klasyfikator oparty na cechach Haar-podobnych. Tym razem jednak, biorąc za podstawę rozwiązanie z pracy [8], został on specjalnie wytrenowany w celu wykrycia oddzielnie lewego i prawego oka. Lokalizacja oczu oparta na cechach Haar-podobnych ma wiele zalet (zdolność działania na obrazach w odcieniach szarości, możliwość wykrycia przymrużonych lub zamkniętych oczu, odporność na zmienne warunki oświetleniowe, łatwość użycia oraz wydajność). Wstępne próby wykorzystania metody ujawniły jednak pewne problemy. W większości przypadków proces detekcji dostarcza wielu prawdopodobnych lokalizacji dla jednego oka; właściwa lokalizacja znajduje się jednak zawsze wśród tej listy. Zauważono również, że właściwa lokalizacja samej twarzy jest ograniczona do obrotów o wartości mniej więcej $\pm 30^\circ$. W związku z tym nie ma potrzeby wyszukiwania oczu w całym obszarze twarzy.

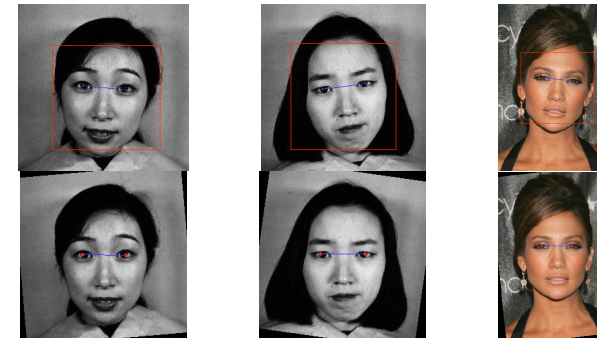


rysunek 4. Lokalizacja oczu w obszarach ograniczających (baza JAFFE i Internet)

Na biało – końcowe pozycje oczu (białe koło i prostokąt), obszary ograniczające (żółte linie)

Na czerwono – kandydaci na lokalizacje oczu wewnątrz obszarów ograniczających

Jak pokazano na obrazie (Rys. 4), w większości przypadków, użycie regionów ograniczających eliminuje problem niejednoznacznej lokalizacji. Regiony ograniczające zostały wybrane tak, aby uwzględnić wszystkie możliwe położenia oczu dla każdej wykrywanej rotacji twarzy. Dodatkowe doprecyzowanie zostało oparte na odpowiednio skonstruowanej funkcji kosztu, która preferuje kandydatów wykazujących najwyższy stopień dopasowania oraz cechujących się najmniejszym kątem pomiędzy środkami oczu. Na tym etapie środki oczu są już zlokalizowane - możliwe jest, więc przeprowadzenie normalizacji geometrycznej. W tym celu obliczony musi zostać kąt między środkami oczu, który jest następnie wykorzystywany przy korekcyjnej rotacji twarzy (Rys. 5).

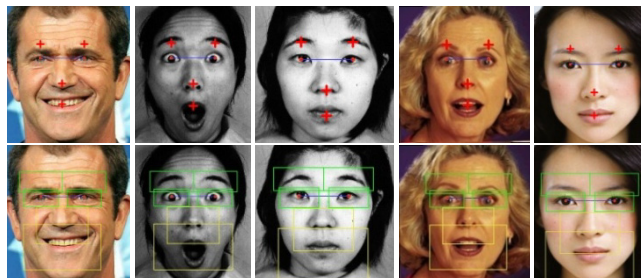


rysunek 5. Obrazy przed (górny rząd) oraz po (dolny rząd) korekcyjnej rotacji bazującej na kącie pomiędzy środkami oczu (baza JAFFE i Internet)

3.2 Lokalizacja obszarów twarzy

Położenia twarzy oraz oczu są znane na tym etapie. Co więcej, twarz jest już po korekcyjnej rotacji. Kolejnym etapem jest wykorzystanie modelu antropometrycznego w celu lokalizacji pozostałych elementów twarzy (ust, brwi, nosa). Zgodnie z modelem [1], statystyczne proporcje pomiędzy konkretnymi odległościami na przeciętnej ludzkiej twarzy pozostają mniej więcej stałe. W związku z tym, położenia punktów P2, P5, P11 oraz P12 (Rys. 3a) mogą być określone w oparciu o odległość pomiędzy środkami obu oczu P16 oraz P17 jako przesunięcie w relacji do geometrycznego środka pomiędzy punktami środkowymi oczu SC_x (Rys. 3b). Po zlokalizowaniu środków obszarów zawierających elementy twarzy, konieczne jest jeszcze określenie rozmiarów tych obszarów. W tym celu wykorzystana została odległość pomiędzy środkami oczu P16 oraz P17. Wielkości obszarów zostały dobrane w taki sposób, aby obejmowały dany element dla każdej możliwej jego deformacji. Prostokąt obejmujący usta jest w tym przypadku dość duży i może, w przypadku zamkniętych ust, zawierać duże obszary twarzy lub nawet jej

otoczenia. Regiony obejmujące elementy twarzy dla różnych jej wyrazów prezentuje rysunek (Rys. 6).



rysunek 6. Lokalizacja elementów twarzy w oparciu o model antropometryczny (baza JAFFE oraz Internet)

Z uwagi na wspomniane zróżnicowanie wyrazów twarzy, zdecydowano o dodaniu etapu uściślającego lokalizacje dla ust i nosa. W obu tych przypadkach większa dokładność uzyskiwana jest dzięki wykorzystaniu specjalnie wytrenowanego klasyfikatora cech Haar-podobnych opisanego w pracy [8].

3.3 Lokalizacja punktów charakterystycznych twarzy

Posiadając już wyizolowane regiony zawierające elementy twarzy takie jak oczy, nos, usta oraz brwi, możliwe jest zlokalizowanie 17 charakterystycznych punktów twarzy. Punkty te zostały dobrane tak, aby reprezentować mimikę twarzy najlepiej i najoszczędniej jak to możliwe (Rys. 3). Dla maksymalnej efektywności zastosowano podejście hybrydowe – każdy element twarzy analizowany jest w inny, indywidualnie dobrany sposób zapewniający maksymalne wykorzystanie jego charakterystyk.

3.3.1 Analiza nosa

Zgodnie z modelem konieczne jest odnalezienie dwóch

punktów zlokalizowanych w środkach otworów nosowych. Otwory nosowe są na obrazie widoczne, jako dwa wyraźnie kontrastujące, ciemne, okrągłe, bądź eliptyczne obszary. Detekcja środków otworów nosowych odbywa się dzięki filtracji, która jest w stanie oddzielić ciemne, okrągłe obszary od jaśniejszego tła. W procesie tym wykorzystywany jest tzw. filtr LoG¹⁵ (Laplacian of Gaussian) [3]. Aby uczynić proces odpornym na zróżnicowane rozmiary otworów, wielkość filtra jest zmieniana i cała operacja ulega wielokrotnemu powtarzaniu. Punkty środkowe otworów nosowych są odnajdywane, jako lokalne maksima obszarów po filtracji. Następnym etapem jest przeprowadzenie weryfikacji bazującej głównie na ograniczeniach antropometrycznych. Aby dalej poprawić uzyskany rezultat wykorzystywany jest poziomy filtr Sobela¹⁶ mający uwidocznić linie końcówki nosa. Ostateczna decyzja, co do lokalizacji podejmowana jest dzięki wykorzystaniu specjalnie skonstruowanej funkcji celu bazującej na opisanych wcześniej informacjach.

3.3.2 Analiza ust

Usta są prawdopodobnie najbardziej zmiennym spośród elementów twarzy. Pierwszym etapem analizy jest dopasowanie kontrastu obrazu tak, aby uczynić usta ciemniejszymi, a otaczający obszar skóry jaśniejszym. Po tym etapie poszukiwane są dwa punkty

¹⁵ **Filtr LoG** - filtr Laplace'a oparty na funkcji Gaussa. Wartości tej funkcji w przestrzeni filtra są odzwierciedleniem odchylenia standardowego. Filtr Laplace'a jest stosowany do wykrywania krawędzi. Filtr Gaussa reprezentuje rozkład normalny zmiennej losowej, Źródło: dokumentacja OpenCV

¹⁶ **Filtr Sobela** – filtr krawędziowy, którego zadaniem jest dawanie bardzo silnych odpowiedzi w miejscach występowania silnych zmian kontrastu. Obszary te najczęściej reprezentują krawędzie. Filtry Sobela są kierunkowe, co oznacza, że mogą wykrywać krawędzie poziome, pionowe, bądź ukośne. Źródło: [16]

umiejscowione w kąciach ust. W procesie tym znajdują zastosowanie dwa spostrzeżenia. Po pierwsze, punkty zwykle znajdują się w silnych rogach obrazu. Po drugie, pomiędzy kącikami ust, niezależnie od mimiki, widoczny jest ciemny, podłużny obszar, który może zostać wyostrzony przy użyciu poziomego filtru Sobela. Kandydackie lokalizacje kącików ust są uzyskiwane w wyniku poszukiwania silnych rogów obrazu, a następnie weryfikowane w oparciu o własności opisane powyżej. Po zlokalizowaniu kącików ust możliwe jest określenie położenia środkowej części ust i poszukiwanie środkowych punktów na wargach. Obraz jest odwracany i binaryzowany z użyciem progu obliczanego za pomocą iteracyjnej procedury. Wybierany jest największy kontur spełniający ograniczenia, który podlega dalszej analizie. Lokalizowane są na nim dwa punkty mające koordynaty na osi x dokładnie w połowie odległości pomiędzy wcześniej zlokalizowanymi kącikami ust. Punkt posiadający większą wartość na osi y wybierany jest na punkt środkowy dolnej wargi, zaś drugi punkt, mający mniejszą wartość wybierany jest na punkt środkowy górnej wargi.

3.3.3 Analiza oka

Wynikami analizy oka powinny być położenia czterech punktów. Najpierw lokalizowane są dwa punkty w kącikach oka, następnie dwa punkty pośrodku górnej i dolnej powieki. W pierwszej fazie następuje dopasowanie intensywności obrazu tak, aby zwiększyć kontrastowość brwi. Obraz jest następnie konwertowany do postaci binarnej przy wykorzystaniu iteracyjnie obliczonego progu [3]. Obszary uzyskane na tym obrazie są otaczane konturem przy wykorzystaniu procedury 8-connected contour following algorithm [11]. Największy uzyskany kontur podlega dalszej analizie w celu wyboru zewnętrznego i wewnętrznego

kącika oka. Aby poradzić sobie z mylącymi cieniami, które czasem pojawiają się w kącikach oka, położenia są sprawdzane przy użyciu ograniczeń antropometrycznych. Jeśli ograniczenia te nie są spełnione - wykorzystywany jest dodatkowy mechanizm korygujący oparty na lokalizacji silnych rogów w obrazie. Po uzyskaniu położenia kącików oka, rozpoczyna się poszukiwanie punktów środkowych brwi. Kontur oka jest dzielony na trzy równe części. Na podstawie analizy środkowego obszaru wybierana jest para punktów charakteryzujących się maksymalną pionową odległością między sobą. Punkty te są następnie sprawdzane w oparciu o ograniczenia antropometryczne i dodatkowo korygowane, jeśli zajdzie taka potrzeba, przy pomocy mechanizmu opisanego w [12].

3.3.4 Analiza brwi

Dla każdej brwi lokalizowane są trzy punkty charakterystyczne. W pierwszym etapie następuje inwersja kolorów w celu wyróżnienia brwi, jako białych obszarów na czarnym tle (w większości przypadków). Owłosienie głowy mogące znajdować się w wyznaczonym obszarze brwi jest eliminowane poprzez odjęcie obrazu średniej jasności tła od obrazu oryginalnego. W celu uzyskania średniej jasności tła przeprowadzana jest operacja morfologiczna otwarcia z elipsą o promieniu 10 px jako elementem strukturującym. Następnie, kontrast obrazu jest zwiększany w oparciu o sumacyjny rozkład jasności pikseli, po czym uzyskiwany jest obraz binarny z progiem ustalonym przy użyciu metody Otsu [13]. Jednolite obszary w obrazie otaczane są konturami w wyniku użycia metody 8-connected contour following algorithm [11]. Największy kontur o szerokości większej od wysokości jest wybierany jako kontur brwi i podlega dalszej analizie. Skrajne punkty brwi

lokalizowane są jako pierwsze. Następnie wyliczana jest wartość punktu środkowego brwi na osi x jako średnia z wartości x punktów skrajnych. Wybierany jest punkt na konturze o najmniejszej wartości na osi y. Zidentyfikowane punkty są weryfikowane i ewentualnie poprawiane w oparciu o warunki antropometryczne takie jak kąty i wzajemne odległości.

3.4 Rezultaty testów

Podsumowując, średnia precyzja lokalizacji 17 punktów dla bazy JAFFE wynosi 2.63 pixela (wyliczona w relacji do ręcznego zaznaczania punktów) oraz 95,58% (stosunek odległości od ręcznego zaznaczenia podzielony przez odległość między oczami).

4. Normalizacja danych dla wykrywania emocji

Ponieważ na percepcję emocji wpływ mają różnice międzyosobnicze w proporcjach twarzy, uzyskane odległości pomiędzy punktami na twarzy poddane zostały normalizacji w celu dokonania klasyfikacji wyrazów twarzy niezależnej od osoby jak i od rozmiaru oraz położenia twarzy na zdjęciu.

Normalizacja cech [3] jest konieczna, gdyż na rozpoznawanie emocji wpływ mają proporcje twarzy danej osoby. Normalizacja ta jest przeprowadzona poprzez wygenerowanie średniej twarzy neutralnej ze wszystkich twarzy neutralnych w zbiorze treningowym. Następnie wektor cech tej twarzy jest odejmowany od każdej z twarzy w zbiorze danych.

Normalizacja przestrzeni danych jest podobna do skalowania i nakładania na siebie najważniejszych regionów dwóch zdjęć (nos, oczy, usta, brwi) z jak największą dokładnością. Pełne dopasowanie dwóch różnych zdjęć nie jest możliwe, jednakże możliwe jest takie ułożenie i przeskalowanie zdjęć by rejonu oczu,

brwi i nosa znajdowały się mniej więcej w tym samym miejscu. Tego typu normalizacja sprawia, że twarze w zestawie nie muszą być znormalizowane ze względu na ich rozmiary i pozycje na zdjęciach. Wykonanie normalizacji przestrzeni danych polega na uśrednieniu dwóch odległości, które następnie służą jako skale - pionowa i pozioma - dla każdego zestawu punktów twarzy. Odległość pozioma to odległość między środkami oczu, natomiast odległość pionowa łączy uśrednioną pozycję środków oczu z punktem znajdującym się pomiędzy dziurkami w nosie.

5. Klasyfikacja emocji

Testowano rozpoznawanie siedmiu podstawowych emocji, tj. strachu, smutku, złości, radości, zaskoczenia, odrazy, a także neutralnego wyrazu twarzy. Do klasyfikacji użyto trzech typów klasyfikatorów oraz dwie metody uśredniania wyników - głosowanie i średnia ważona. Pierwszym z użytych klasyfikatorów jest sieć neuronowa oparta na funkcjach radialnych gaussowskich [10], wybrana ze względu na łatwość trenowania a także dobre dopasowanie do danych. Składa się ona z 15 neuronów wejściowych, 7 neuronów wyjściowych oraz z pewnej liczby neuronów w warstwie ukrytej - ich liczba zależy od danych użytych do trenowania sieci. Warstwa ukryta sieci to tak naprawdę siedem niezależnych podsieci z których każda odpowiada za inną z rozpoznawanych emocji. Taka izolacja podsieci ułatwia dobór wag i sprawia, że rozpoznawanie każdej z emocji jest osobnym zadaniem, niepowiązanym z innymi. Funkcję aktywacji neuronów warstwy ukrytej prezentuje (Wzór 2)

$$N = \exp\left(-\sum_{i=1}^M \frac{(x_i - c_i)^2 w_i}{2r^2}\right)$$

wzór 2 – funkcja aktywacji neuronów

gdzie M jest liczbą cech, c jest wektorem centrów neuronów, w jest wektorem wag, r jest promieniem danego neuronu. Sieć jest trenowana poprzez dodanie nowego neuronu dla każdej uśrednionej emocji dla każdej twarzy w zbiorze treningowym. Dla każdej błędnie rozpoznanej próbki dodawany jest nowy neuron aż wszystkie są rozpoznawane poprawnie. Każde z centrów neuronów jest trenowane poprzez uśrednienie wszystkich cech we wszystkich próbkach, które powinny być przez ten neuron rozpoznawane, czyli wszystkich próbek dla danej emocji należącej do danej twarzy dla której ten neuron został utworzony. Promień każdego neuronu jest trenowany poprzez uśrednienie odległości od wszystkich innych neuronów w podsieci do której ten neuron należy. Zabieg ten rozszerza wpływ danej podsieci na cały obszar na którym emocja rozpoznawana przez tę podsieć mogłaby się teoretycznie pojawić, wg danych treningowych. Ponieważ sieci radialne są liniowe w wagach, możliwym było wytrenowanie jej poprzez użycie algorytmu SVD bez potrzeby uciekania się do czasochłonnego trenowania on-line. Poza siecią RBF użyto dwóch innych klasyfikatorów: naiwnego klasyfikatora Bayesowskiego oraz KNN. Zauważono, że różne klasyfikatory dokonywały błędnej klasyfikacji dla innego typu danych. Połączenie ich pozwoliło na poprawienie wyników klasyfikacji, co można szczególnie zauważyć dla metody średniej ważonej.

5.1 Wyniki i dyskusja

W ogólności system działa dobrze nawet dla danych uzyskanych automatycznie (Tabelki 1 i 2). Należy zauważyć, że najlepsze wyniki są uzyskane dla ważonej kombinacji KNN i sieci RBF. Potwierdza to dobrą zdolność sieci radialnych do generalizacji danych, koniecznej do rozpoznawania twarzy nieznanymi. Natomiast dla twarzy znanych, gdzie dokładność jest preferowana, KNN sprawdza się lepiej.

tabela 1 - Zestawienie ilościowe danych używanych w testowaniu.

Nazwa zestawu	Ilość zdjęć (twarzy) w zestawie testowym	Ilość zdjęć (twarzy) w zestawie treningowym	Ilość emocji
Twarz znana	70 (10)	143 (10)	7
Twarz nieznaną	22 (1)	191 (9)	7

tabela 2 – Uzyskane wyniki.

Na zwa zesta wu	Zaznacza nie punktów	Baye s	KNN ¹⁷	Sieć RBF	Śred nia wazo na ¹⁸	Głoso wanie
Twarz znana	automaty czne	52.8 5%	71.4 2%	65.7 1%	74.2 8%	75.71%
	ręczne	64.2 8%	85.7 1%	75.7 1%	90.0 0%	87.14%
Twarz nie znana ¹⁹	Automaty czne	54.5 4%	63.6 3%	86.3 6%	81.8 1%	72.72%
	ręczne	63.6 3%	86.3 6%	77.2 7%	86.3 6%	81.81%

6. Podsumowanie

Jest jeszcze miejsce na ulepszenia, szczególnie jeśli chodzi o część zajmującą się klasyfikacją. Można oczekiwać lepszych rezultatów jeśli zastosuje się:

- kompensację rotacji w osi Y (twarz nie ułożona en face),
- ilość neuronów jest zredukowana na podstawie podobieństwa twarzy,
- użyte zostaną odległości Mahalanobisa zamiast Euklidesowych,

¹⁷ Dla twarzy znanych przyjęto sąsiedztwo wynoszące 1, a dla twarzy nieznanymi sąsiedztwo wynoszące 11. Sąsiedztwa dobrane zostały eksperymentalnie.

¹⁸ Tylko klasyfikator KNN i sieć RBF.

¹⁹ Testy przeprowadzone zostały na twarzy YM, zestaw trenujący składał się ze wszystkich pozostałych twarzy oprócz twarzy YM.

- sposób klasyfikacji zostanie zmodyfikowany w celu użycia innych narzędzi dla twarzy różnego typu (znanej, nieznannej).

200 lat temu Karol Darwin wskazał jak dużą wagę ma wyrażanie emocji w komunikacji międzyludzkiej. Nowoczesne techniki pozwalają na interakcję pomiędzy człowiekiem a komputerem. Rezultaty tej pracy wykazały, że użycie cech antropometrycznych ułatwia zadanie rozpoznawania emocji. Punkty, zawierające wystarczającą informację do rozpoznawania siedmiu podstawowych emocji mogą być precyzyjnie lokalizowane.

Potencjalne zastosowania aplikacji rozpoznającej emocje to m.in. programy e-learningowe, dopasowujące styl prezentacji do obserwowanych emocji studenta, wspomaganie serwisów zdrowotnych w określaniu stanu emocjonalnego klienta, urządzenia towarzyszące jak robotyczne zwierzęta, które dzięki umiejętności określenia emocji mogą dostarczyć większego stopnia realizmu interakcji i autonomii. Może być to również monitoring emocji np. w samochodzie, pomagający w zastosowaniu dodatkowych środków bezpieczeństwa w zależności od wykrytych emocji pasażerów. Innym z zastosowań jest HCI - interakcje człowieka z komputerem. Przykładowo może być to program dobierający muzykę do humoru słuchacza, czy też program ostrzegający przed wysłaniem wściekłego e-maila. Istnieje również pewien potencjał zastosowania aplikacji rozpoznającej emocje w rozwoju technologii komunikacyjnych dla ludzi z autyzmem, którzy nie są w stanie rozpoznawać emocji z taką samą łatwością jak ludzie zdrowi, co utrudnia im interakcje ze społeczeństwem.

Oczywiście system opracowany na potrzeby niniejszej pracy jest zaledwie małym wkładem w ogólnoświatowe badania nad tego typu zagadnieniami, które w przyszłości mogą całkowicie odmienić sposób, w jaki ludzie porozumiewają się z komputerami. Przy kontynuacji badań, możliwy będzie dalszy rozwój tej technologii.

Bibliografia

- [1] Sohail, A.S.M., Bhattacharya, P.: "Detection of Facial Feature Points Using Anthropometric Face Model", in: Damiani, E. et al. (Ed.), *Signal Processing for Image Enhancement and Multimedia Processing*, ISBN 978-0-387-72499-7, Springer (2007)
- [2] Viola P., Jones M.J., "Robust Real-Time Object Detection", *Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing and Sampling*, Vancouver, Canada (2001)
- [3] Sohail, A.S.M., Bhattacharya P., "Support Vector Machines Applied to Automated Categorization of Facial Expressions", in: Prasad B. (Ed.): *Proceedings of the 3rd Indian International Conference on Artificial Intelligence IICAI 2007*, ISBN 978-0-9727412-2-4, Pune, India, December 17-19 (2007)
- [4] The Japanese Female Facial Expression (JAFFE) Database <http://www.kasrl.org/jaffe.html>
- [5] OpenCV Wiki-pages: <http://opencv.willowgarage.com/wiki/>
- [6] Jobson, D.J., Rahman, Z., Woodell, G.A. "A multiscale retinex for bridging the gap between color images and the human observation of scenes", *Image Processing, IEEE Transactions on*, Vol. 6, Issue 7, 965 – 976 (1997)
- [7] Lienhart R., Maydt J., "An Extended Set of Haar-like Features for Rapid Object Detection", *Proceedings of the International Conference on Image Processing, IEEE* (2002)
- [8] Castrillón-Santana M., Déniz-Suárez O, Antón-Canalís L., J. LorenzoNavarro "Face And Facial Feature Detection Evaluation", <http://gias720.dis.ulpgc.es/Gias/Publications/visapp-2008-1.pdf> (2008)
- [9] Freund Y., Schapire R.E., "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences* vol. 55(1) (1997)
- [10] Oukhellou L., Aknin P., "Optimization of Radial Basis Function Network for Classification Tasks", *Neurap 1998 IV International Conference on Neural Networks and their Applications*, Marseille (1998)
- [11] Ritter, G.X., Wilson, J.N.: "Handbook of Computer Vision Algorithms in Image Algebra", CRC Press, Boca Raton, USA (1996)
- [12] P. Kuo, J. M. Hannah, "An Improved Eye Feature Extraction Algorithm Based on Deformable Templates", *IEEE International Conference on Image Processing*, Genova, Italy, pp. 1206-1209. (2005)
- [13] Otsu, N.: "A Threshold Selection Method from Gray Level Histograms", *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 9, No. 1 (1979) 62–66
- [14] Bouguet, J.-Y., "Pyramidal Implementation of the Lucas Kanade Feature Tracker", *Intel Corporation, Microprocessor Research Labs* (2002)
- [15] Yoav Freund, Robert E. Schapire "A short Introduction to Boosting" In: *Journal of Japanese Society for Artificial Intelligence*, September 1999. At&T Labs – Research Shannon Laboratory 180 Park Avenue.
- [16] Jacek Rumuński, „Metody reprezentacji, przetwarzania i analizy obrazów w medycynie” Katedra Elektroniki Medycznej i Ekologicznej – Politechnika Gdańska