

**Akademia Górniczo-Hutnicza**  
**im. Stanisława Staszica w Krakowie**

---

Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki



**PRACA MAGISTERSKA**

**Maciej Dudek**

**AUTOMATYCZNA PARAMETRYZACJA I KLASYFIKACJA  
EKSPRESJI EMOCJI NA PODSTAWIE DYNAMICZNYCH  
ZMIAN MIMICZNYCH W OBRAZIE WIDEO**

Promotor:  
dr Adrian Horzyk



Katedra Automatyki, Kraków 2011

## **OŚWIADCZENIE AUTORA PRACY**

OŚWIADCZAM, ŚWIADOMY ODPOWIEDZIALNOŚCI KARNEJ ZA POŚWIADCZENIE NIEPRAWDY, ŻE NINIEJSZĄ PRACĘ DYPLOMOWĄ WYKONAŁEM OSOBIŚCIE I SAMODZIELNIE, I NIE KORZYSTAŁEM ZE ŹRÓDEŁ INNYCH NIŻ WYMIENIONE W PRACY.

.....

PODPIS

**AGH**

**University of Science and Technology in Krakow**

---

Faculty of Electrical Engineering, Automatics, Computer Science and  
Electronics



**MASTER OF SCIENCE THESIS**

**Maciej Dudek**

**AUTOMATIC PARAMETRIZATION AND CLASSIFICATION  
OF EMOTION EXPRESSIONS AFTER DYNAMIC MIMETIC  
CHANGES IN VIDEO STREAM**

Supervisor:  
Adrian Horzyk Ph.D



Department of Automatics, Krakow 2011

## Podziękowanie

PRAGNĘ PODZIĘKOWAĆ PROMOTOROWI

*dr Adrianowi Horzykowi*

ZA CIERPLIWOŚĆ, WYROZUMIAŁOŚĆ, POŚWIĘCONY CZAS,  
INTERESUJĄCE POMYSŁY ORAZ DOBRE RADY, KTÓRYMI SŁUŻYŁ  
W TRAKCIE PISANIA TEJ PRACY.

# Spis treści

<b>1</b>	<b>Wstęp</b>	<b>7</b>
1.1	Emocje	7
1.2	Relacje człowiek - komputer	7
1.3	Zastosowania praktyczne	8
1.4	Cel pracy	8
1.5	Części pracy	8
<b>2</b>	<b>Wprowadzenie</b>	<b>9</b>
2.1	Analiza zagadnienia	9
2.2	Ludzka twarz	9
2.3	Podział emocji	10
2.3.1	Radość	12
2.3.2	Smutek	12
2.3.3	Złość	14
2.3.4	Niesmak	15
2.3.5	Strach	15
2.3.6	Zaskoczenie	16
2.4	Parametry emocji	18
2.5	Funkcje emocji	19
2.6	Istniejące rozwiązania	19
2.6.1	Facial Acting Coding System	19
2.6.2	Facial Expression.Awareness.Compassion.Emotions.	19
2.6.3	Face Meaning Sensivity Test	20
2.6.4	emoRate	20
2.6.5	Candide	20
2.6.6	FaceLab	22
<b>3</b>	<b>Opis rozwiązania</b>	<b>23</b>
3.1	Koncepcja systemu	23
3.2	Używane algorytmy	25
3.2.1	Connected Component Labeling	25
3.2.2	RGB vs HSV	26
3.2.3	Operacje morfologiczne	27
3.2.4	Filtracja obrazów	28
3.2.5	Algorytm Lucasa Kanade	30
3.2.6	Histogram	32
3.2.7	Color slicing	32
3.3	Detekcja twarzy	34
3.4	Wykrywanie brwi	36
3.5	Wykrywanie oczu	37
3.6	Wykrywanie ust	38
3.6.1	Wykrycie odsłonięcia zębów	43

3.7	Śledzenie zmian . . . . .	43
3.8	Graf twarzy . . . . .	43
3.9	Detekcja gestów twarzy . . . . .	44
3.9.1	Brwi uniesione/obniżone . . . . .	44
3.9.2	Ściśnięcie brwi . . . . .	45
3.9.3	Otwarcie/zamknięcie oczu . . . . .	46
3.9.4	Otwarcie ust . . . . .	46
3.9.5	Podniesienie/opuszczenie kątek ust . . . . .	46
3.10	Identyfikacja emocji . . . . .	47
<b>4</b>	<b>Implementacja</b>	<b>48</b>
4.1	Budowa programu . . . . .	49
4.1.1	Wykorzystywane biblioteki . . . . .	49
4.1.2	Stworzone klasy . . . . .	50
4.2	Funkcje programu . . . . .	52
4.2.1	Menu programu . . . . .	52
4.2.2	Obraz wideo . . . . .	54
4.2.3	Gesty . . . . .	54
4.2.4	Emocje . . . . .	55
4.2.5	Wykres zmian emocji w czasie . . . . .	55
<b>5</b>	<b>Wyniki</b>	<b>56</b>
5.1	Detekcja gestów . . . . .	56
5.1.1	Detekcja brwi . . . . .	56
5.1.2	Detekcja oczu . . . . .	57
5.1.3	Detekcja ust . . . . .	58
5.1.4	Błędy wynikające z oświetlenia . . . . .	60
5.2	Klasyfikacja emocji . . . . .	60
5.2.1	Porównanie wyników . . . . .	63
<b>6</b>	<b>Podsumowanie</b>	<b>64</b>

# 1 Wstęp

*“Najbardziej interesującą powierzchnią w świecie jest twarz ludzka.”*  
Georg Christoph Lichtenberg

## 1.1 Emocje

“Emocja wygląda tak samo, niezależnie czy jesteś gospodynią domową czy zamachowcem-samobójcą. **Prawda jest wypisana na naszych twarzach**” twierdził Cal Lightman, specjalista od wykrywania kłamstw i mowy ciała, w popularnym serialu ”Lie to Me”. Czym są emocje i dlaczego są tak ważne w naszym życiu?

Emocje są pierwotnymi uczuciami. Towarzyszą nam od czasów prehistorycznych. Dlatego tak trudno je zdefiniować. Jako pierwszy naukowo zajął się tym Karol Darwin, który, badając ile wspólnego ma rasa ludzka ze zwierzętami, odkrył, że niektóre reakcje naszego organizmu są wspólne wszystkim naczelnym. W 1967 V.J. Wukmir zaproponował następującą definicję: ”Emocja to natychmiastowa odpowiedź organizmu, która informuje o stopniu przyjemności (*favorability*) zaistniałej sytuacji”. [29]

## 1.2 Relacje człowiek - komputer

Wydawać by się mogło, że emocje są sferą zarezerwowaną tylko dla istot żywych i nie mają nic wspólnego z komputerami, które na początku były tylko bardzo dużymi kalkulatorami wykonującymi wiele obliczeń matematycznych w krótkim czasie. Przekaz był jednostronny. Człowiek wydawał komendę i otrzymywał odpowiedź. Jednak wraz z postępem technicznym naukowcy chcą uczynić z komputerów maszyny jak najbardziej podobne do człowieka, tak, aby w pewnym momencie monolog zamienił się w inteligentny dialog między człowiekiem i komputerem. Ale aby tego dokonać komputer musi w pełni zrozumieć wiadomości jakie przekazuje człowiek. A jeśli aż 55% z tego przekazu to informacje niewerbalne to komputery muszą się nauczyć odczytywać i rozumieć ten przekaz.

Niniejsza praca jest wstępem do badań nad zrozumieniem emocji ludzkich przez komputery. Kto wie, może w przyszłości będą one umiały rozumieć ludzi i reagować adekwatnie do naszych stanów emocjonalnych?

### 1.3 Zastosowania praktyczne

Nauka, która jest tylko sztuką dla sztuki, szybko ginie zapomniana. Wsparcie ze strony biznesu, który zainteresowany jest praktycznymi zastosowaniami wiedzy sprawia że, nauka rozkwita bardzo szybko. Automatyczne wykrywanie emocji, mimo że jest dziedziną bardzo młodą może mieć wiele zastosowań w życiu codziennym. Część już funkcjonuje:

- marketing dostosowany do nastroju klienta
- pomoc osobom niepełnosprawnym, mających problemy z komunikacją w sposób tradycyjny
- bezpieczeństwo - wyszukiwanie niebezpiecznych czy zachowujących się podejrzanie osób na lotniskach czy stadionach

### 1.4 Cel pracy

Celem pracy jest stworzenie programu przetwarzającego obraz wideo ze zwykłej kamery internetowej postawionej przed użytkownikiem, poszukującego oznak ekspresji emocji na twarzy użytkownika, oraz analizującego te emocje pod kątem ich intensywności, przebiegu, czasu trwania, ich autentyczności czy też występowania tak zwanych mikroekspresji.

**Mikroekspresja** to nieświadome występowanie pełnej ekspresji mimicznej w bardzo krótkim okresie czasu, w odpowiedzi na występowanie określonej emocji. Odkrył ją dr Paul Ekman obserwując w zwolnionym tempie zapis wideo chorej psychicznie pacjentki, która kłamała, aby wydostać się ze szpitala i popełnić samobójstwo. Odpowiadając na pytanie o plany na przyszłość, wykonała ona krótkotrwały wyraz mimiczny rozpacz, tak krótki, że specjaliści nie dostrzegli go przy obserwacji wideo w normalnej prędkości [10].

### 1.5 Części pracy

Pierwszy rozdział poświęcono opisowi istniejących rozwiązań w zakresie rozpoznawania ludzkich emocji. Już teraz istnieją udane próby analizy komputerowej ludzkich zachowań, z czego niektóre z nich zostały nawet wdrożone do rozwiązań biznesowych. Dowodzi to, iż wspomniana tematyka nie jest zagadnieniem czysto teoretycznym, ale ma duże zapotrzebowanie praktyczne. Część druga została poświęcona opisowi zaproponowanego przeze mnie sposobu rozwiązania podjętego problemu. Postaram się przedstawić w tej części algorytm detekcji twarzy, wykrywania położenia brwi, ust, oczu, jak również zastosowaną metodę wykrywania i analizy emocji na podstawie wykrytych ekspresji. W rozdziale 4, poświęconym implementacji, opisano sposób zastosowania wiedzy teoretycznej zastosowano w stworzonej przez autora aplikacji komputerowej. Następny rozdział opisuje przeprowadzone testy stworzonego oprogramowania oraz ich wyniki. Ostatni (6) rozdział podsumowuje całość prac przeprowadzonych w ramach niniejszej pracy dyplomowej.



## 2 Wprowadzenie

### 2.1 Analiza zagadnienia

Wydawać by się mogło, iż nie ma nic bardziej odległego od techniki niż ludzkie emocje. Jednak analiza zagadnień z dziedziny anatomii i psychologii pozwala na szczegółowy opis powstawania oznak emocji na ludzkiej twarzy. Dzięki takiemu opisowi możliwe jest stworzenie algorytmów do automatycznej analizy emocji. Trzeba przy tym pamiętać o niezwyklej złożoności procesów, jakie zachodzą w trakcie pojawiania się emocji u człowieka, co powoduje, iż takie programy obciążone są dużym prawdopodobieństwem pomyłki. Dobrym przykładem jest tutaj wariograf (potocznie wykrywacz kłamstw), który analizuje stan napięcia badanego, jego potliwość czy tętno w celu wykrycia możliwości zaistnienia kłamstwa. Wariograf jednak bardzo często wskazuje na niewinne osoby, które zestresowały się samym podejrzeniem o możliwość popełnienia przestępstwa[10].

Jak już wspomniano badania nad emocjami zainicjował Karol Darwin, który szukał pierwotnych cech wspólnych dla ludzi i zwierząt. Jednak pierwszą osobą, która zajęła się tematem ludzkich emocji w sposób systematyczny i naukowy, był psycholog amerykański Paul Ekman. Stworzył on w 1976 roku system Face Acting Coding System (FACS)[11], katalogujący poszczególne ruchy twarzy (tzw Action Unity) i przypisujący je do konkretnym emocjom.

### 2.2 Ludzka twarz

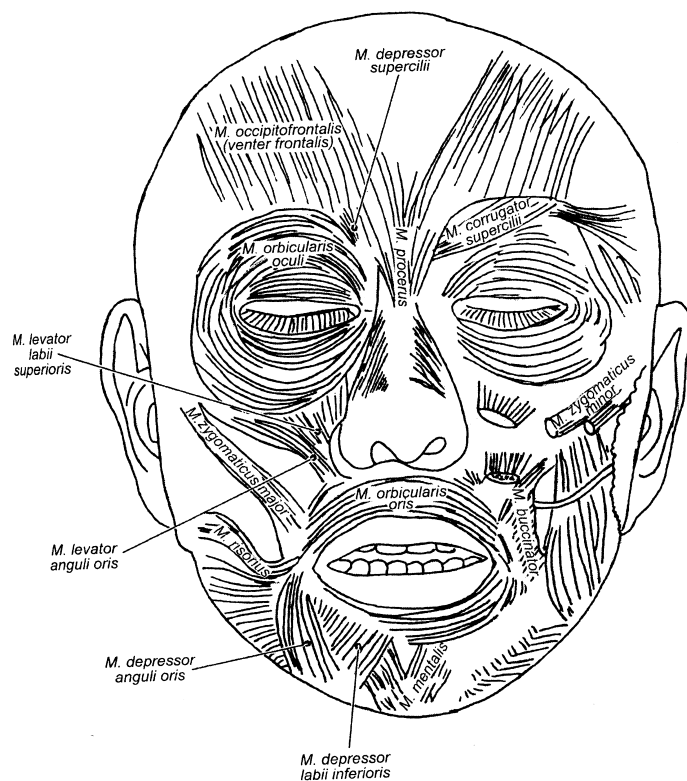
Twarz może dostarczyć całej gamy komunikatów niewerbalnych, ponieważ znajduje się na niej 80 mięśni, dzięki którym możemy utworzyć więcej niż siedem tysięcy wyrazów twarzy, czyli min. [4]

Mięśnie głowy można podzielić na pięć głównych grup: mięśnie szczękowe, mięśnie wargowe, mięśnie gałki ocznej, mięśnie czoła i mięśnie szyi[5].

Twarz jest odpowiedzialna za ponad połowę przekazywanych informacji. Za jej pośrednictwem wyrażane są uczucia i emocje. Sygnały twarzy są dość podobne w społeczeństwach znacznie różniących się kulturowo.[4]

Twarz ludzka pełni wiele ważnych funkcji jeśli chodzi o komunikowanie emocji:

1. Przez przyjemny bądź nieprzyjemny wyraz twarzy komunikujemy, czy bieżący obiekt naszego zainteresowania oceniamy jako dobry czy zły.
2. Twarz komunikuje zainteresowanie lub brak zainteresowania innymi ludźmi lub środowiskiem.
3. Twarz komunikuje intensywność i stopień zaangażowania w daną sytuację.
4. Twarz komunikuje, w jakim stopniu ludzie kontrolują własną ekspresję.



Rysunek 1: Mięśnie twarzy [2]

5. Twarz prawdopodobnie komunikuje również intelektualny aspekt, wskazujący na rozumienie sytuacji lub jego brak. [24]

## 2.3 Podział emocji

Włodzimierz Szewczuk w [39] zaproponował podział emocji na 4 rodzaje:

**Emocje zagrożenia:** niepokój, lęk, strach, przerażenie,

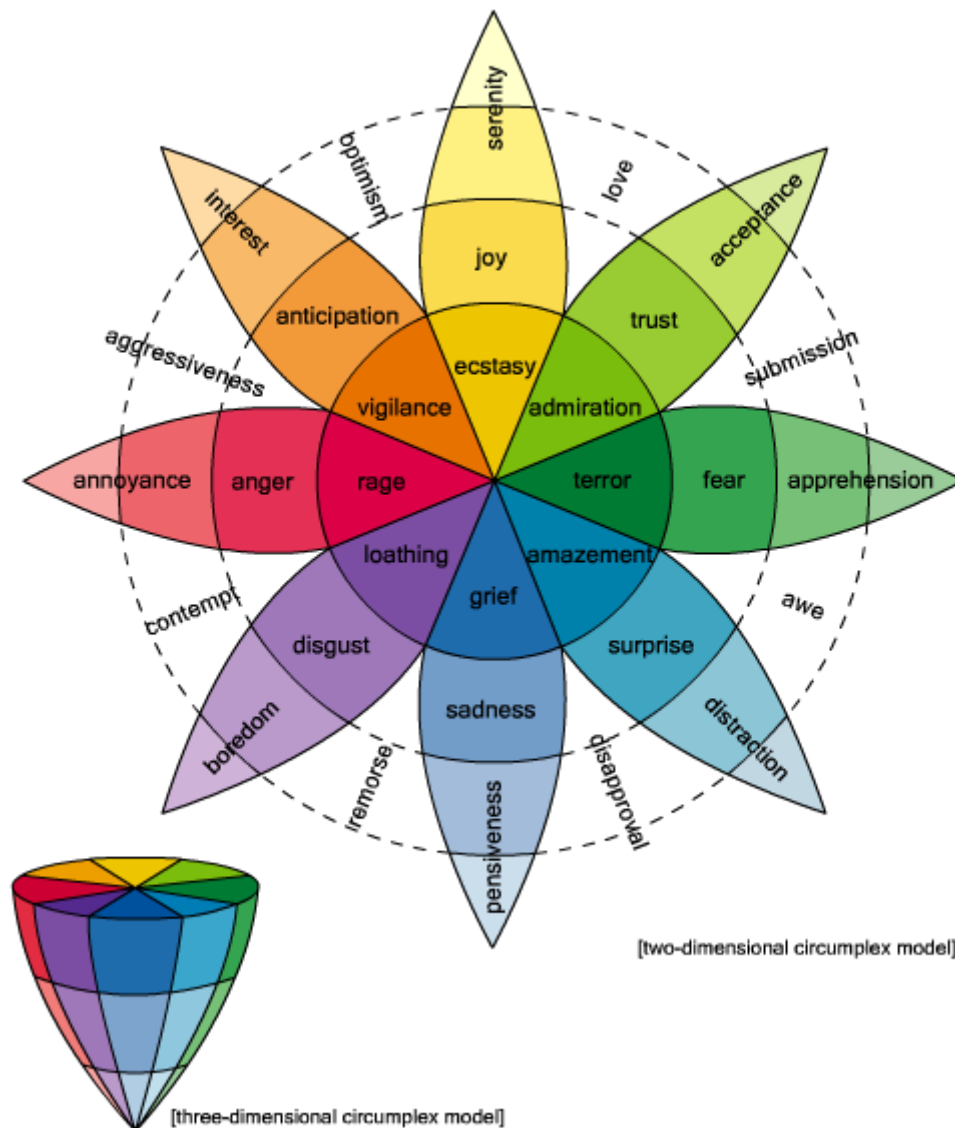
**Emocje ograniczenia:** rozdrażnienie, złość, gniew, wściekłość,

**Emocje pozbawienia:** smutek, żałość, cierpienie, rozpacz,

**Emocje zaspokojenia:** radość, uniesienie, szczęście, ekstaza.

Robert Plutchik zaproponował w [35] podział emocji na podstawowe i złożone (będące połączeniem wielu emocji podstawowych). Wśród wyróżnionych przez badacza 8 emocji podstawowych znalazły się radość, złość, strach, wstęś, smutek, zaskoczenie, zaciekawienie, akceptacja. Uczony zaproponował również ciekawą formę prezentacji wizualnej tego podziału w postaci kolorowego koła emocji (Rysunek 2).

## Plutchik's Wheel of Emotions



Rysunek 2: Koło emocji wg Plutchika. Autor Ivan Akira

Paul Ekman w [8] zaproponował 9 cech, które charakteryzują emocje podstawowe:

1. uniwersalne, odróżniające sygnały,
2. ekspresje porównywalne u innych zwierząt,
3. fizjologia specyficzna dla konkretnej emocji,
4. uniwersalne pierwotne korzenie,
5. spójność w odpowiedzi organizmu,
6. błyskawiczne pojawienie się oznak emocji,

7. krótkotrwałość,
8. automatyczna, mechaniczna odpowiedź organizmu
9. mimowolne, niekontrolowane występowanie

W niniejszej pracy skupię się na 6 podstawowych emocjach wyróżnionych przez Ekmana. Wymienione poniżej *Action Unit* (AU) przedstawione są w rozdziale 2.6.1

### 2.3.1 Radość

”Zdobycie, zachowanie i odzyskanie szczęścia jest tak naprawdę dla większości ludzi w każdej epoce sekretnym motywem każdego działania” - zauważył William James. Nietrudno to zrozumieć, ponieważ stan szczęścia lub nieszczęścia nadaje barwę wokół nas. Ludziom szczęśliwym świat wydaje się bezpieczniejszy, szczęśliwi łatwiej podejmują decyzje, korzystniej oceniają kandydatów do pracy i uważają się za bardziej zadowolonych ze swojego życia[30].

Na podstawie [13], [5] oraz [11], w niniejszej pracy przyjęto, iż radość na twarzy objawia się występowaniem następujących cech:

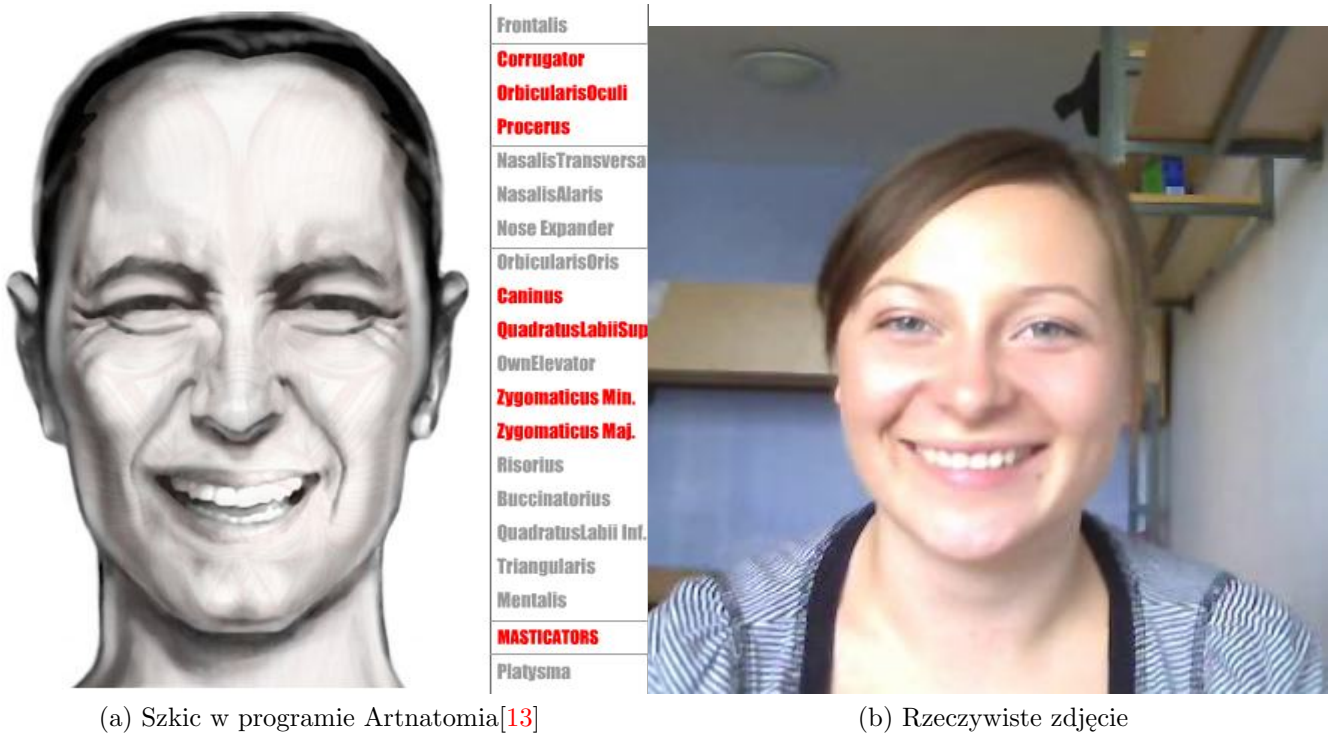
- przymrużenie powiek AU 7
- uniesieniem brwi AU 2
- uniesienie kątek ust (i ewentualnie całej górnej wargi) AU 12
- ukazaniem się zębów (opcjonalne, wzmacnia siłę emocji)

### 2.3.2 Smutek

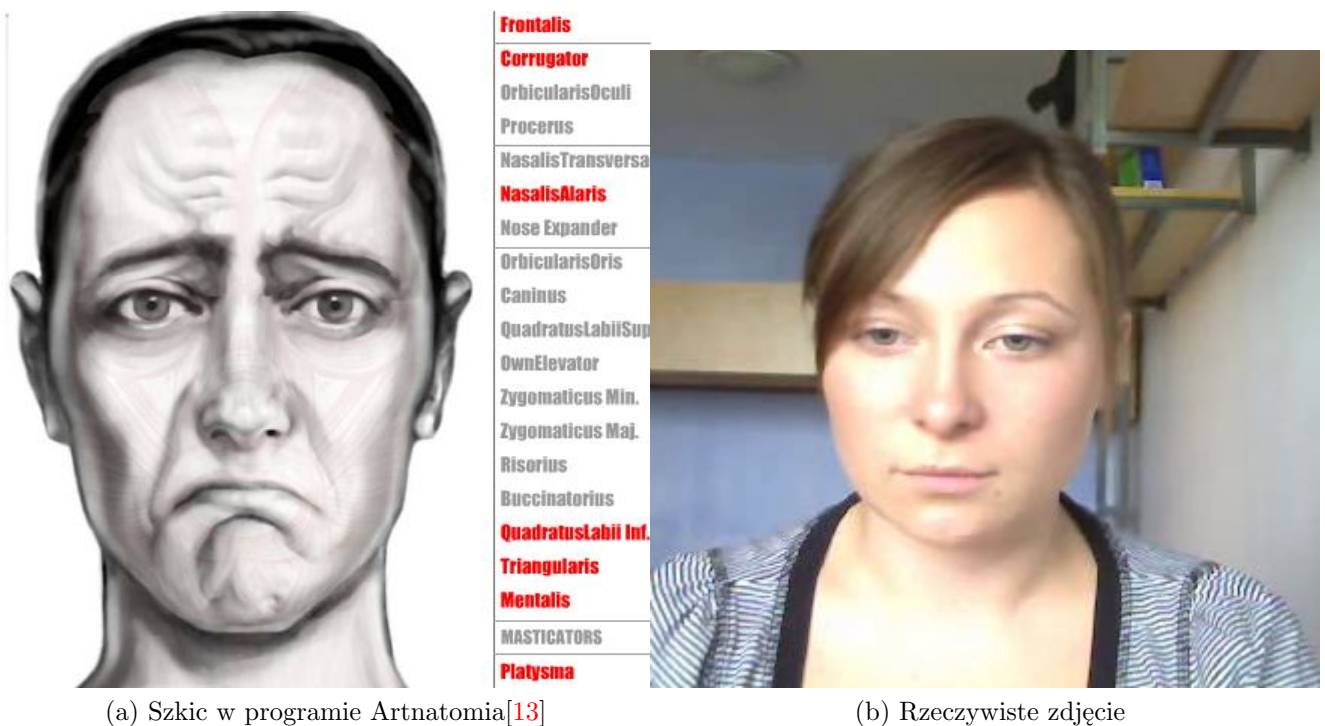
Smutek jest emocją doświadczaną w obliczu wydarzenia opisywanego jako przykre. Smutek traktuje się jako reakcję na utratę celu lub nieosiągnięcie celu. Smutek różni się od strachu tym, iż jest reakcją na wydarzenie, które już zaszło, podczas gdy strach antycypuje wydarzenie. Smutek powoduje skupianie się człowieka na sobie, a nie na czynnikach zewnętrznych. Jednostka skupia się na konsekwencjach związanych z nieosiągnięciem wyznaczonych celów[27].

Smutek na ludzkiej twarzy charakteryzuje się występowaniem następujących cech:

- zwężeniem brwi AU 1
- uniesieniem brwi AU 2
- przymknięciem oczu AU 7
- obniżeniem kątek ust AU 15
- ściśnięciem ust AU 24



Rysunek 3: Wizualizacja radości



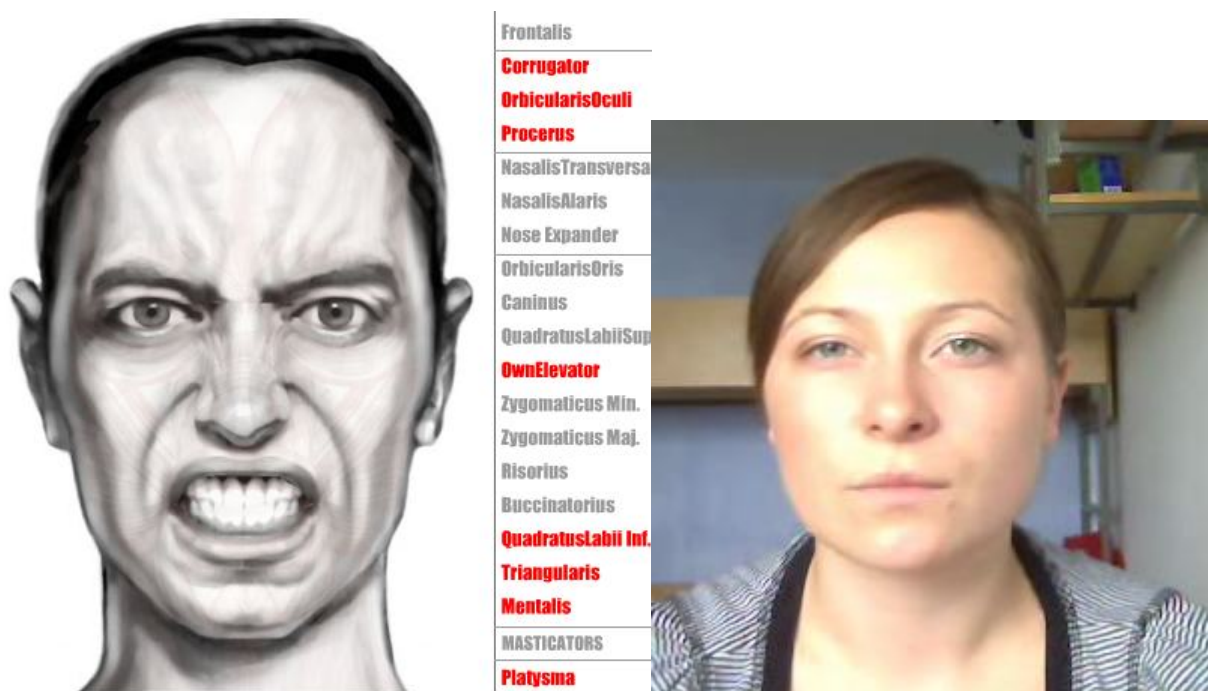
Rysunek 4: Wizualizacja smutku

### 2.3.3 Złość

Złość organizuje i reguluje wewnętrzne procesy fizjologiczne i psychologiczne związane z samoobroną i dominacją, a także z kontrolowaniem zachowań społecznych i interpersonalnych. Złość sygnalizuje coś ważnego o relacjach jednostki ze środowiskiem i wpływa na reakcję podmiotu na sytuację[27].

Cechy charakteryzujące występowanie złości na twarzy to:

- obniżenie brwi AU 4
- zwężenie brwi AU 1
- przymrużenie oczu AU 7
- obniżenie kątek ust AU 15
- obniżenie dolnej wargi (opcjonalne) AU 26
- ukazanie zębów (opcjonalne, wzmacnia siłę emocji)



(a) Szkic w programie Artnatomia[13]

(b) Rzeczywiste zdjęcie

Rysunek 5: Wizualizacja złości

### 2.3.4 Niesmak

Darwin w swoim klasycznym dziele *"O wyrazie uczuć i człowieka i zwierząt"* zdefiniował wstręt jako dotyczący "czegoś odrażającego, szczególnie, w odniesieniu do zmysłu smaku, co jest rzeczywiście postrzegane lub plastycznie wyobrażane; a po drugie dotyczący czegokolwiek, co budzi podobne uczucie poprzez zmysł zapachu, dotyku lub nawet wzroku". Darwin łączył wstręt nie tylko z odczuwaniem odrazy, ale i z charakterystyczną ekspresją mimiczną. Wstręt jest stosowany, by człowiek mógł bronić się przed psychicznym zetknięciem się z odrażającym obiektem lub jakimkolwiek zbliżeniem się do niego. Wstręt przejawia się jako dystansowanie się od jakiegoś obiektu, wydarzenia lub jakiejś sytuacji i daje się scharakteryzować jako odrzucenie. Tylko wstręt łączy się ze specyficznym stanem fizjologicznym - mdłościami[27].

Paul Rozin, Jonathan Haidt i Clark McCauley w książce "Disgust", proponują 5 stopniowe etapy wstrętu[27]:

**Awersja** - ochrona ciała przed trucizną,

**Podstawowy wstręt** - ochrona ciała przed chorobą/zakażeniem,

**Wstręt związany ze zwierzęcą naturą** - ochrona ciała i duszy, zaprzeczenie śmiertelności,

**Wstręt interpersonalny** - ochrona ciała, duszy i porządku społecznego,

**Wstręt moralny** - ochrona porządku społecznego.

Niesmak na twarzy charakteryzuje się występowaniem następujących cech:

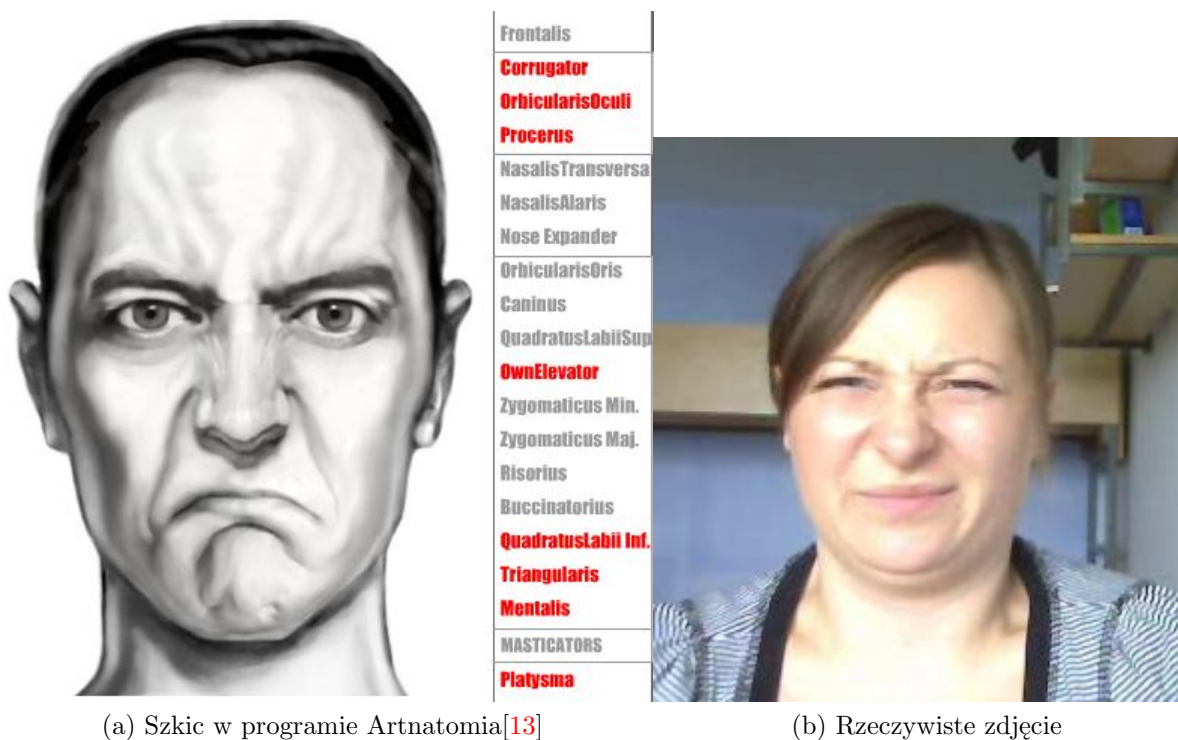
- obniżeniem brwi AU 4
- zwężeniem brwi AU 1
- obniżenie kącików ust AU 15

### 2.3.5 Strach

Z perspektywy klinicznej lęk to niesprecyzowane i nieprzyjemne przeczucie zagrożenia, związane ze zmianami fizjologicznymi, przejawiającymi się zarówno w sferze somatycznej, jak i autonomicznej. Aspekt behawioralny lęku obejmuje ucieczkę i unikanie. Strach różni się od lęku przede wszystkim tym, że jest wywoływany przez rozpoznawalny bodziec[27].

Lęk jest reakcją przystosowawczą, ponieważ przygotowuje organizm do ucieczki przed niebezpieczeństwem. Lęk przed prawdziwymi lub wyimaginowanymi wrogami skupia ludzi w rodziny, plemiona, narody. Lęk przed karą lub zemstą powstrzymuje przed wyrządzeniem krzywdy innym. Lęk wywołuje troskę, dzięki której umysł skupia się na problemie i analizuje strategie rozwiązania.[30]

Można wnioskować występowanie emocji strachu na ludzkiej twarzy po zaobserwowaniu następujących czynników:



Rysunek 6: Wizualizacja niesmaku

- zwężenie brwi AU 1
- uniesienie górnych powiek (szeroko otwarte oczy) AU 5
- obniżenie kątek ust AU 15
- obniżenie dolnej wargi AU 26

### 2.3.6 Zaskoczenie

Zaskoczenie jest najkrótszą ze wszystkich emocji, trwającą do maksymalnie kilku sekund. W chwili, kiedy zaskoczenie mija, kiedy zorientujemy się co się dzieje, przeistacza się ono w strach, ulgę, radość, złość, zniesmaczenie itd., w zależności co nas zaskoczyło, lub zamienione na brak emocji, jeśli zaskakujące wydarzenie było bez konsekwencji [9].

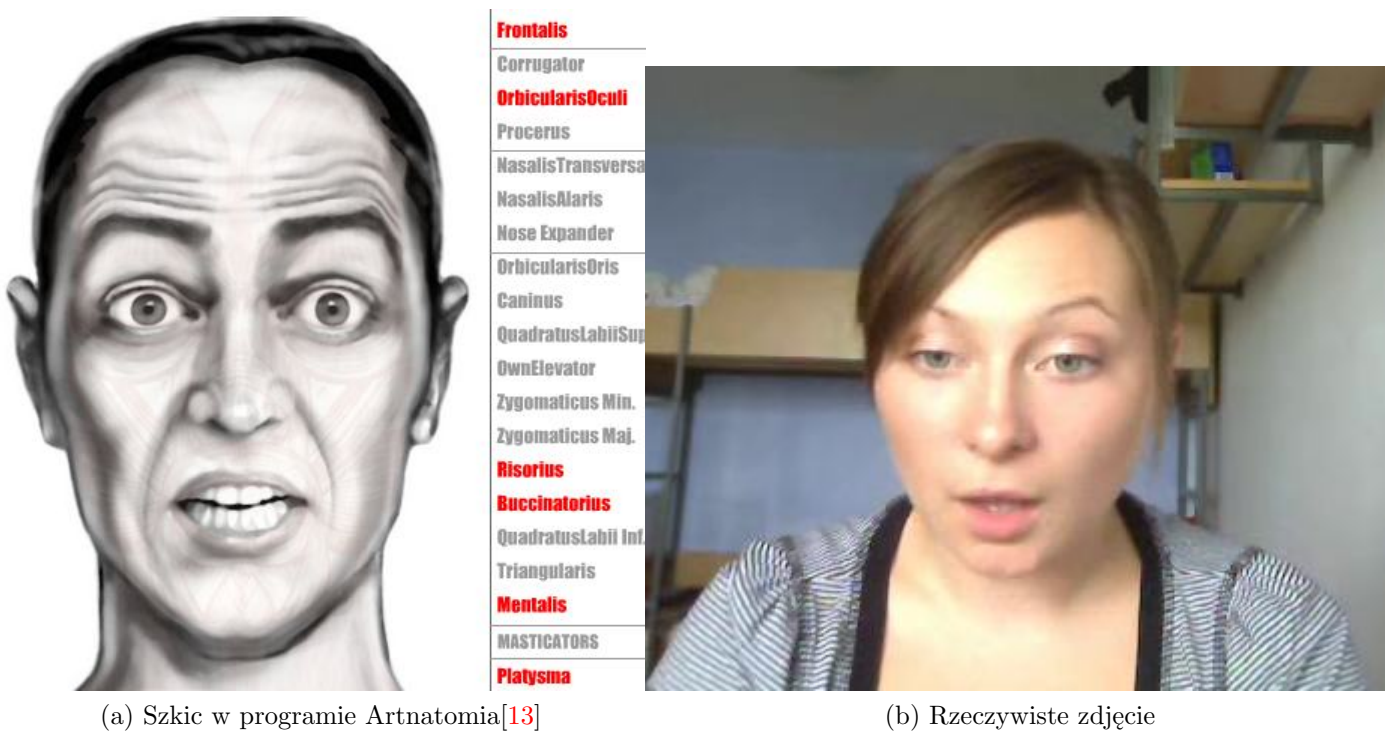
Zaskoczenie na twarzy charakteryzuje się występowaniem następujących cech:

- uniesieniem brwi AU 2
- otwarciem oczu AU 5
- obniżeniem dolnej wargi AU 26





Rysunek 7: Wizualizacja strachu



Rysunek 8: Wizualizacja zaskoczenia

## 2.4 Parametry emocji

Włodzimierz Szewczuk w [39] podaje 5 cech charakteryzujących emocje:

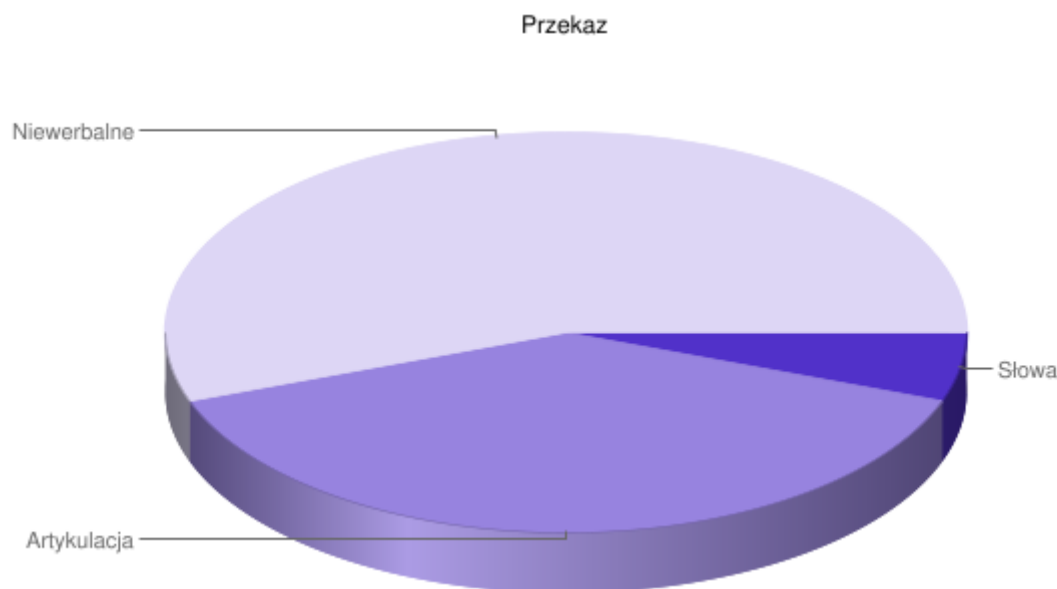
**Wartość** - jest zależna od znaczenia przedmiotu (zjawiska, sytuacji) dla osobnika. Jeżeli przedmiot, zakłócający równowagę, ma pozytywne znaczenie dla osobnika, zaspokaja jego potrzebę, nasila jego moc osobniczą - emocja ma wartość dodatnią. W odwrotnym przypadku, gdy przedmiot zagraża osobnikowi, zmniejsza jego moc osobniczą - emocja ma wartość ujemną. Nie ma emocji "zerowych". Gdy przedmiot nie ma żadnego znaczenia dla osobnika, nie może powstać emocja.

**Kierunkowość emocji** - ma, podobnie jak wartość biegunowy charakter: ku przedmiotowi lub od przedmiotu. Jeżeli przedmiot ma dodatnią wartość dla osobnika, zmierza on w kierunku tego przedmiotu. W przeciwnym wypadku jest odwrotnie.

**Ruchliwość emocji** - oznacza tę jej właściwość, która polega na mobilizowaniu osobnika, na zwiększaniu jego gotowości do działania, na zasilaniu jego prężności działaniowej, jego aktywności. Z jednej strony, mamy emocje (jak radość, gniew, miłość), które aktywizują osobnika, z drugiej, emocje dezaktywizujące, jak smutek, lęk zgrzyzota.

**Intensywność emocji** - jest związana z wartością przedmiotu dla osobnika i wyraża się zaangażowaniem osobnika w działanie, w charakterze zmian wewnątrzorganicznych, nasileniu doznania i stopnia odporności na czynniki hamujące.

**Czas trwania** - waha się w bardzo dużych granicach, od milisekund, po całe dni.



Rysunek 9: Procentowy rozkład przekazu ludzkiego

## 2.5 Funkcje emocji

Pojawia się naturalne pytanie o cel i funkcje, jakie pełnią emocje w życiu ludzi. Psycholog Michael Lewis określił to w ten sposób: "Nasze emocje i emocje innych są źródłem więzi społecznej i źródłem ludzkiego interesu." [27] On też wskazuje na to, że emocje mogą być przejawem cech atawistycznych, wynikających z naszych pierwotnych korzeni. "Kolejna ważna hipoteza dotyczy tego, że rozwój intelektualny i kulturalny człowieka wyprzedził ewolucję. Emocje pełniły funkcję przystosowawczą przy radzeniu sobie z zagrożeniami i szansami, jakie stwarzała sawanna, i przy posługiwaniu się pięściami oraz narzędziami kamiennymi. Mogą nie pełnić dłużej takiej funkcji przy radzeniu sobie z aktualną infrastrukturą techniczną i niemal nieograniczoną dostępnością zasobów. Współczesna złość i współczesna chciwość przybrały wynaturzoną postać, ponieważ system emocjonalny nie rozwinął się wraz z tymi okolicznościami kulturowymi." [27]

Emocje przekazują pewien komunikat innym ludziom w sposób niewerbalny i zrozumiały dla każdej osoby na Ziemi. Jedną z podstawowych potrzeb człowieka jest komunikowanie się. Według badań Alberta Mehrabiana tylko 7 procent naszego przekazu stanowi część werbalna (słowa), 38 procent to artykulacja (ton głosu, barwa, rytm, głośność) i aż 55 procent z komunikatów niewerbalnych (gestów, spojrzeń, wyrazów twarzy) [32].

## 2.6 Istniejące rozwiązania

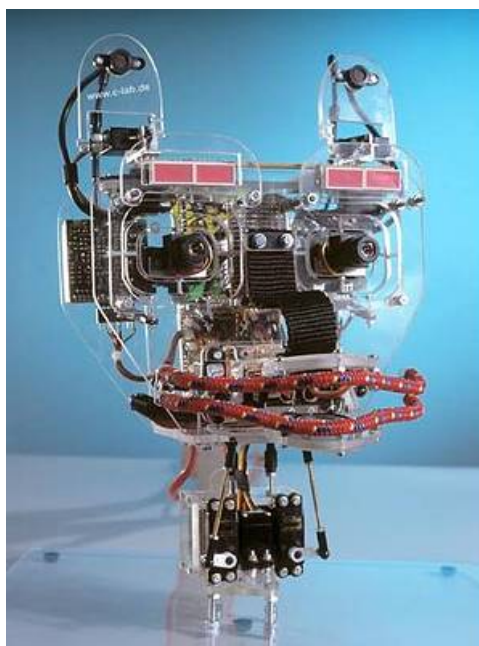
Naukowcy od wczesnych lat 70 starali się skodyfikować i zapisać stany emocjonalne w sposób możliwy do przetworzenia przez komputery, tak aby w przyszłości maszyny były w stanie rozumieć emocje, jak również je imitować (przykład na Rysunku 10). Jako pioniera w tej dziedzinie można uznać Paula Ekmana, amerykańskiego psychologa, który opisał każdą emocję w sposób umożliwiający implementację odpowiednich algorytmów komputerowych.

### 2.6.1 Facial Acting Coding System

Jest to system stworzony w 1978 roku przez Paula Ekmana and Wallace'a V. Friesena, do opisu wszystkich możliwych odróżnialnych ekspresji mimicznych na ludzkiej twarzy. Wychodząc z założenia, że każdy ruch na twarzy jest wynikiem pracy mięśni twarzy, badali oni, jak każdy mięsień wpływa na zmianę wizerunku. Każdy odróżnialny ruch twarzy oznaczyli jednostką *Action Unit (AU)*, która grupuje wszystkie mięśnie które biorą udział w danym ruchu [1]. Zebrane *Action Units* przedstawiono w Tabeli 1.

### 2.6.2 Facial Expression.Awareness.Compassion.Emotions.

F.A.C.E to stworzone przez Paula Ekmana rozwinięcie poprzedniego systemu, zawierające odkrycie Ekmana - mikroekspresje, czyli występowanie jednych emocji w czasie trwania innych.



Rysunek 10: MEXI (Machine with Emotionally eXtended Intelligence). Źródło: [31]

### 2.6.3 Face Meaning Sensivity Test

Test wrażliwości na znaczenie wyrazu twarzy (FMST - Face Meaning Sensivity Test) jest szczegółowym i precyzyjnym narzędziem mierzącym umiejętność dekodowania ekspresji twarzy. Część I testu FMST zawiera 10 fotografii reprezentujących 10 podstawowych klas znaczenia twarzy. Część II testu składa się z 30 zdjęć wyrazów twarzy, po 3 na każdą emocję. W III etapie te same zdjęcia należy skategoryzować przy użyciu bardziej specyficznych rodzajów ekspresji (np. zakłopotanie, strapienie czy upór). [24]

### 2.6.4 emoRate

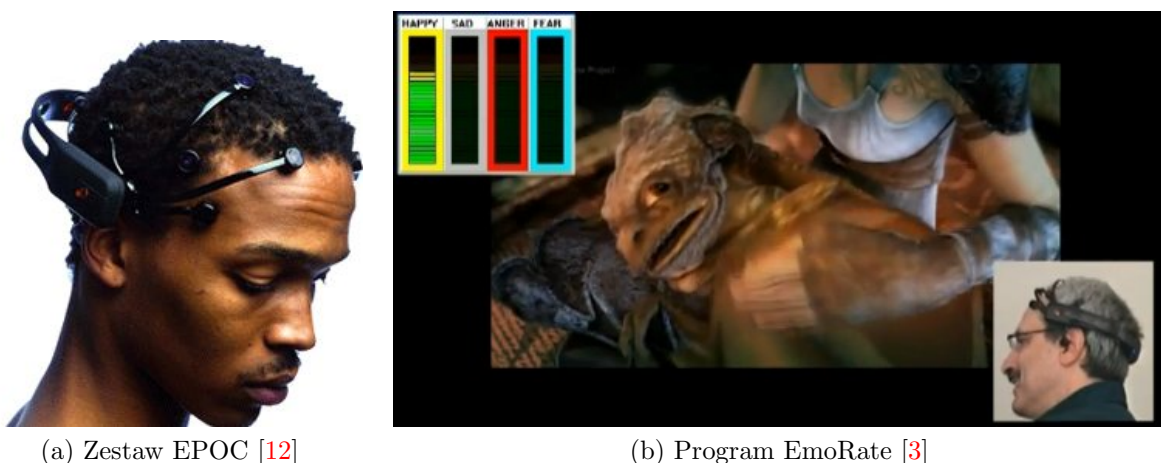
EPOC firmy emotiv to urządzenie zawierające 14 sensorów będących elektroencefalografem (EEG), badające bioelektryczne czynności mózgu. Urządzenie to wspólnie z oprogramowaniem EmoRate, prezentuje możliwości tak zwanego Affective Computing, technologii która pozwala na detekcję i reakcję na ludzkie emocje. EmoRate wykrywa 4 emocje, które zdaniem autora są najsilniej reprezentowane przez ludzki organizm: radość, smutek, strach i złość.

### 2.6.5 Candide

Candide to program komputerowy generujący sparametryzowaną maskę ludzkiej twarzy. Każdy z około 100 wieloboków tworzących siatkę twarzy, pozwala na odzworowanie poszczególnych jednostek (*Action Unit*) stworzonych przez Ekmana. Mimo bardzo prostej konstrukcji, pro-

AU	Czynność	Mięśnie twarzy
1	Podniesienie wewnętrznej części brwi	<i>Frontalis, pars medialis</i>
2	Podniesienie zewnętrznej części brwi	<i>Frontalis, pars lateralis</i>
4	Obniżenie brwi	<i>Corrugator supercilii, Depressor supercilii</i>
5	Podniesienie górnej powieki	<i>Levator palpebrae superioris</i>
6	Podniesienie policzków	<i>Orbicularis oculi, pars orbitalis</i>
7	Zwężenie powiek	<i>Orbicularis oculi, pars palpebralis</i>
9	Zmarszczenie nosa	<i>Levator labii superioris alaquae nasi</i>
10	Podniesienie górnej wargi	<i>Levator labii superioris</i>
11	Rozszerzenie płatków nosowych	<i>Zygomaticus minor</i>
12	Podniesienie kącików ust	<i>Zygomaticus major</i>
13	Wydęcie policzków	<i>Levator anguli oris (a.k.a. Caninus)</i>
14	Dołeczki na policzkach	<i>Buccinator</i>
15	Obniżenie kącików ust	<i>Depressor anguli oris (a.k.a. Triangularis)</i>
16	Obniżenie dolnej wargi	<i>Depressor labii inferioris</i>
17	Podwyższenie podbródka	<i>Mentalis</i>
18	Wydęcie ust	<i>Incisivii labii superioris, Incisivii labii inferioris</i>
20	Rozciągnięcie ust	<i>Risorius w/ platysma</i>
22	Otwarcie ust	<i>Orbicularis oris</i>
23	Zwężenie ust	<i>Orbicularis oris</i>
24	Ścisnięcie ust	<i>Orbicularis oris</i>
26	Opad szczęki	<i>Masseter, relaxed Temporalis, internal Pterygoid</i>

Tabela 1: Action Unit w systemie FACS[36]

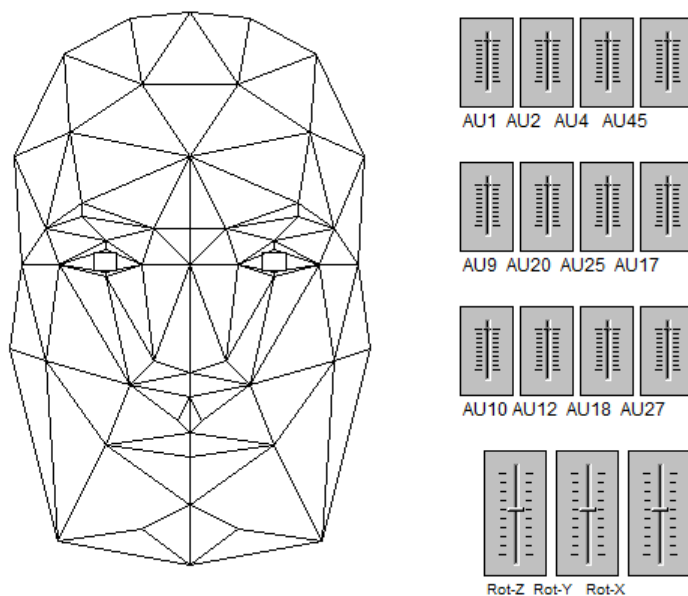


(a) Zestaw EPOC [12]

(b) Program EmoRate [3]

Rysunek 11: Zestaw EPOC z programem emoRate

gram ten jest często wykorzystywany przez środowiska naukowe do wizualizacji poszczególnych stanów emocjonalnych.



Rysunek 12: Program Candide. Źródło: [28]

### 2.6.6 FaceLab

Pakiet faceLAB firmy Seeing Machines jest zestawem składającym się z 2 kamer ustawionych równoległe w niedalekiej odległości od siebie (tak jak jest to wykorzystywane w nagrywaniu filmów 3D) oraz oprogramowania komputerowego. Po skalibrowaniu zestaw pozwala na uzyskanie szczegółowych danych na temat:

- pozycji głowy i jej skrzywienie/pochylenie względem pionu,
- ruchu oczu,
- rozmiaru źrenic,
- ruchu powiek,
- ruchu ust,
- ruchu brwi.



Rysunek 13: Zestaw faceLAB. Źródło: [26]

## 3 Opis rozwiązania

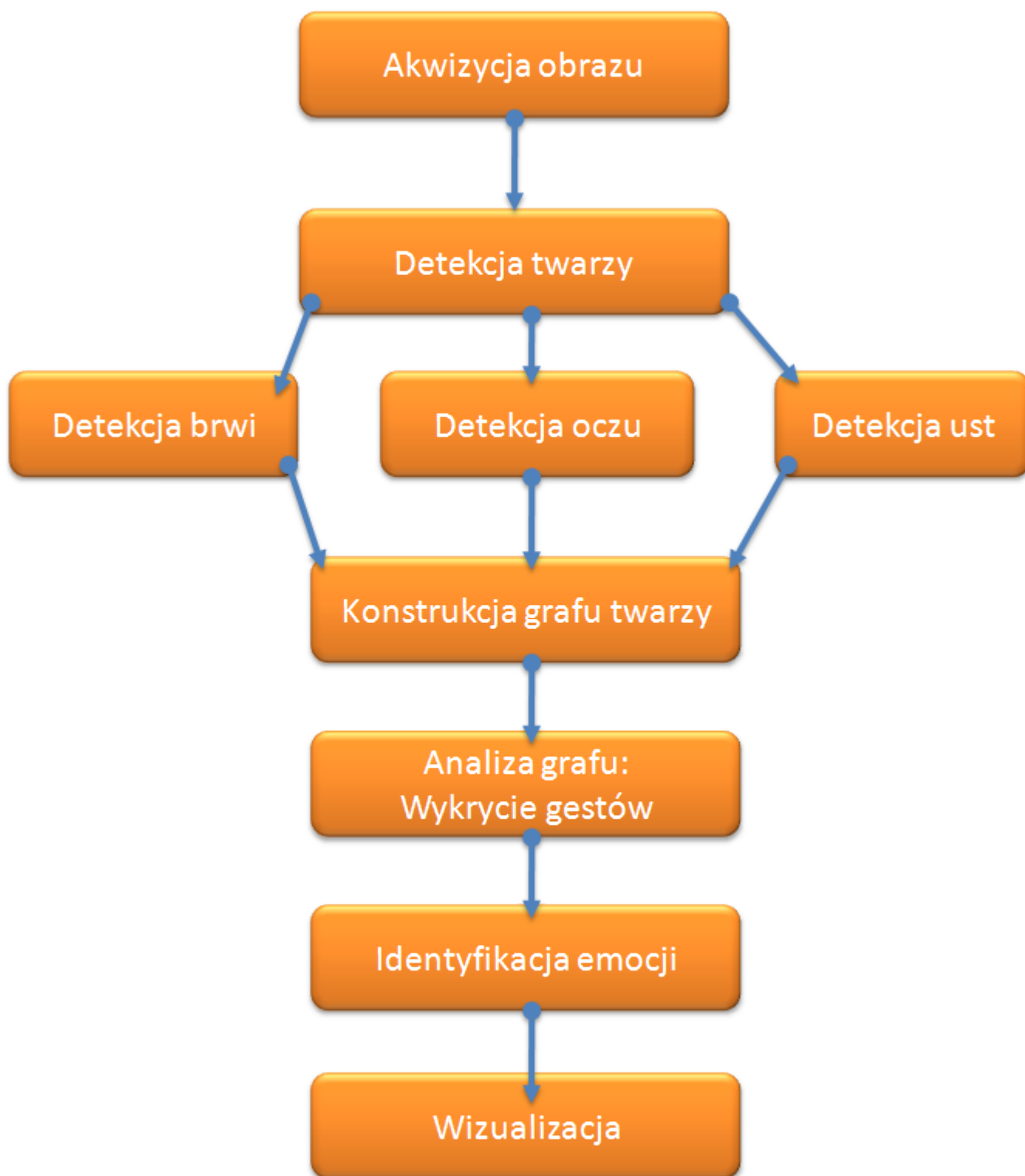
### 3.1 Koncepcja systemu

Stworzony system automatycznej parametryzacji i klasyfikacji emocji bazuje na gestach twarzy wykrytych na obrazie wideo. Schemat działania systemu przedstawiony jest na Rysunku 14.

Dane do analizy dostarcza się przy użyciu strumienia wideo, na którym nagrany jest użytkownik siedzący przed kamerą internetową. Film nagrany jest z prędkością 15 klatek na sekundę, co pozwala wychwycić najdrobniejsze zmiany emocjonalne na twarzy. Taka ilość ramek wpływa znacząco na prędkość ich analizy, dlatego program pozwala ustawienie przetwarzania na co którąś klatkę, co przyspiesza działanie algorytmu i pozwala na działanie w czasie rzeczywistym, odbywa się to jednak kosztem dokładności.

Innym aspektem technicznym wpływającym na dokładność i szybkość przetwarzania jest rozdzielczość przetwarzanego obrazu wideo. W dotychczasowych pracach (jak na przykład w [22]) stosowano rozdzielczość 320x240 pikseli. Pozwala to na szybką analizę obrazu, jednak gubione są istotne szczegóły twarzy. Oczy reprezentowane są jako czarne punkty, gorzej również są rejestrowane szczegóły dotyczące ust, podobne regiony łączone są jedność co utrudnia ekstrakcje danych. Z tych powodów w niniejszej pracy zdecydowano się na użycie obrazu o rozdzielczości 640x480 gdzie wyżej wymienione szczegóły są dobrze reprezentowane. Przy mocy obliczeniowej obecnych komputerów oraz odpowiedniej optymalizacji zastosowanych algorytmów konfiguracja taka zapewnia dużo lepszą dokładność przy akceptowalnym czasie przetwarzania.

System prezentuje na wyjściu wyniki w postaci wykrytych elementów twarzy, zarejestrowanych gestów oraz zidentyfikowanych emocji.



Rysunek 14: Schemat działania algorytmu



## 3.2 Używane algorytmy

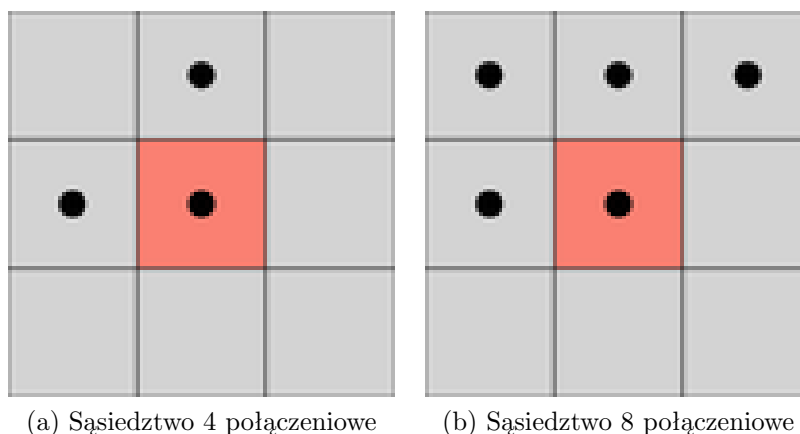
### 3.2.1 Connected Component Labeling

Connected Component Labeling jest algorytmem służącym do nadania unikalnych etykiet dla rejonów, które są ze sobą połączone. Zazwyczaj stosuje się go do obrazów binarnych, gdzie piksel o wartości 0 należy do tła, natomiast ten o wartości 1 jest obiektem zainteresowania. Istnieją jednak również algorytmy do etykietowania dla obrazów w skali szarości czy w pełnej przestrzeni barw. Algorytmy CCL można podzielić na 4 typy[37]:

1. metody, które przechodzą do przodu a następnie w tył przez dane,
2. metody, które przechodzą dwukrotnie przez dane - wymagają one osobnych tablic do przechowywania odpowiedników etykiet,
3. metody, które używają struktur drzewiastych do przechowywania danych,
4. metody, które używają algorytmów równoległych.

Niech  $I$  będzie obrazem binarnym a  $F, B$  podzbiórmi  $I$  odpowiadającymi odpowiednio wartościami oraz pikselami tła. Połączone komponenty (*connected components*)  $I$  oznaczone jako  $C$  jest podzbiorem  $F$  o maksymalnym rozmiarze takim że wszystkie piksele  $C$  są ze sobą połączone[25].

Dwa piksele,  $P$  i  $Q$ , są ze sobą połączone jeśli istnieje ścieżka pikseli  $(p_0, p_1, \dots, p_n)$  takich że  $p_0 = P$ ,  $p_n = Q$  i  $\forall 1 \leq i \leq n$   $p_{i-1}$  i  $p_i$  są sąsiadami[25]. Sąsiedztwo pikseli może być 4 połączeniowe lub 8 połączeniowe (Rysunek 15)

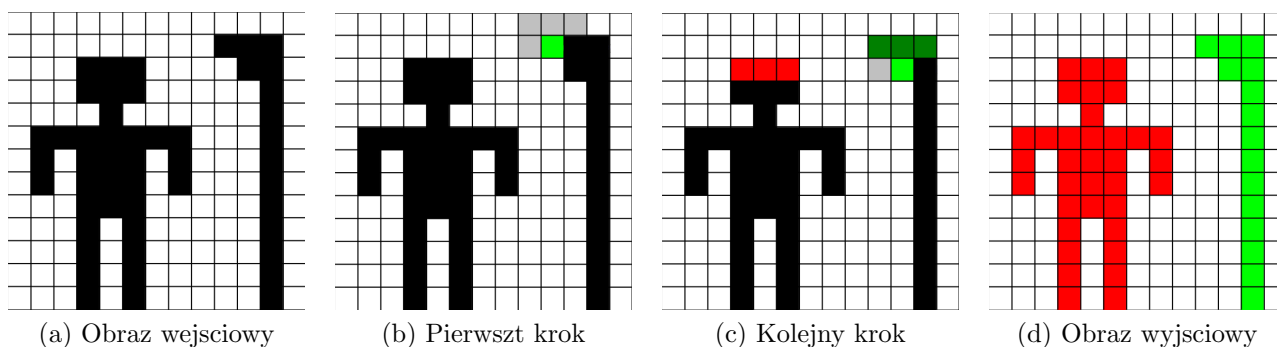


Rysunek 15: Sąsiedztwa pikseli w algorytmie CCL. Źródło: [40]

Klasyczny algorytm sekwencyjny opiera się na 2 następujących po sobie przejściach przez  $I$ . Za pierwszym razem przypisywane są tymczasowe etykiety dla wszystkich pikseli w zbiorze  $F$ , bazując na wartościach pikseli z sąsiedztwa już przetworzonego. Oznaczając jako  $N_F$  zbiór wszystkich już odwiedzonych sąsiedztw należących do  $F$  pierwszy etap może zostać opisany następująco. Jeśli  $N_F$  jest pusty, piksel  $x$ , który ma być zaetykietowany, dostaje przypisaną

nową etykietę ( $x$  jest pierwszym elementem nowego komponentu), jeśli wszystkie piksele w  $N_F$  mają etykietę  $l$ , wtedy przypisujemy  $x$  etykietę  $l$  ( $x$  staje się częścią komponentu opisanego etykietą  $l$ ), jeśli z kolei dwa piksele mają różne etykiety, wtedy  $x$  przypisana jest jedna z tych etykiet a etykiety związane z pikselami w  $N_F$  są zapisywane jako równe sobie ( $x$  staje się łącznikiem dwóch poprzednio rozdzielonych rejonów)[25].

Jako rezultat pierwszego etapu różne etykiety mogą być przypisane do tego samego rejonu. Dlatego też następuje wtedy drugie przejście, w którym przypisuje się unikalne oznaczenia dla tymczasowych etykiet.



Rysunek 16: Przykład działania algorytmu CCL. Źródło: [40]

### 3.2.2 RGB vs HSV

Oko ludzkie widzi kolory jako złożenie:

**nasylenia (Hue)** który opisuje 'czysty' kolor,

**saturacji (Saturation)** opisuje jego rozmycie przez białe światło,

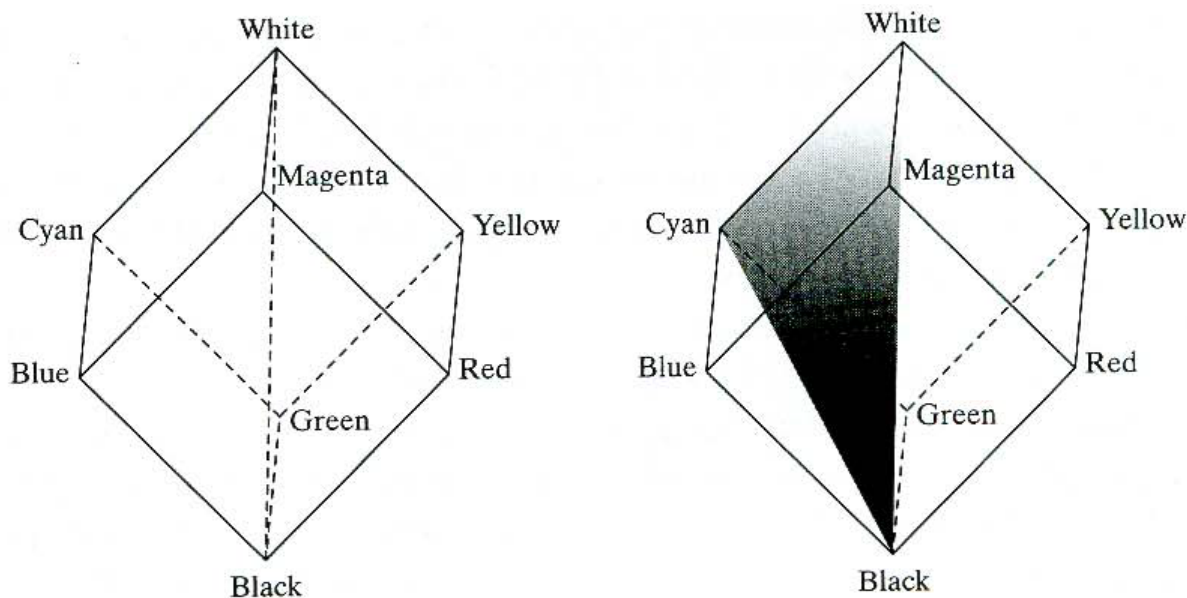
**wartość (Value)** która oznacza jasność.

Opis ten może służyć znacznie lepiej do niektórych zastosowań, niż tradycyjnie używany w programach komputerowych schemat RGB (Red - Green - Blue). Rafael Gonzalez w [15] podaje następujący algorytm pozwalający zamienić model RGB na HSV. Aby otrzymać składową nasylenia (H - hue) ze składowych HSV używamy następującego wzoru:

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B > G \end{cases}$$

gdzie

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}} \right\}$$



Rysunek 17: Relacja między przestrzeniami barw RGB i HSV. Źródło:[15]

składową saturacji uzyskujemy wyliczając:

$$S = 1 - \frac{3}{R + G + B} [\min(R, G, B)]$$

i wreszcie składową intensywności (V - value, lub I - intensity):

$$I = \frac{1}{3}(R + G + B)$$

### 3.2.3 Operacje morfologiczne

Operacje morfologiczne to operacje matematyczne służące do pozyskania części obrazu, które są użyteczne w celu reprezentacji i opisu kształtu regionu (krawędzie, szkielety, otoczenie) [15].

Operacja **dylatacji** dla obrazów  $A$  i  $B$  ze zbioru  $Z^2$  definiowana jest następująco:

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (1)$$

Dylatacja jest więc takim przekształceniem, gdzie  $\hat{B}$  i  $A$  nachodzą na siebie w przynajmniej jednym punkcie. Zbiór  $B$  jest zazwyczaj nazywany elementem strukturalnym. Dylatacja jest najczęściej stosowana w celu połączenia rozłączonych elementów lub uwypuklenia danego elementu.

Operacja **erozji** dla obrazów  $A$  i  $B$  ze zbioru  $Z^2$  definiowana jest jako:

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (2)$$

Innymi słowy erozja zbioru  $A$  zbiorem  $B$  jest zbiorem wszystkich punktów  $z$  takich że  $B$  przetworzone przez  $z$  zawiera się w  $A$ . Jednym z głównych zastosowań erozji jest eliminacja detali i szumów (elementów o jednostkowej, bądź małej, powierzchni).

Dwie kolejne istotne operacje będące połączeniem poprzednich operacji to otwarcie i zamknięcie. **Otwarcie** zdefiniowane jako

$$A \circ B = (A \ominus B) \oplus B \quad (3)$$

jest operacją dylatacji wyniku erozji zbioru  $A$  elementem strukturalnym  $B$ . Otwarcie wygładza kontury obiektu, łamie ostre krawędzie oraz eliminuje cienkie występy.

Operacja **zamknięcia**, czyli erozja poprzedzona dylatacją:

$$A \bullet B = (A \oplus B) \ominus B \quad (4)$$

wyostrza ostre krawędzie, eliminuje małe dziury oraz wypełnia przestrzenie w konturze.



(a) Obraz oryginalny

(b) Dylatacja

(c) Erozja

(d) Otwarcie

(e) Zamknięcie

Rysunek 18: Operacje morfologiczne

### 3.2.4 Filtracja obrazów

Filtracja obrazów dostarcza szereg możliwości wydobycia z obrazu oryginalnego szeregu informacji lub ułatwia jego obróbkę. Filtrację obrazów cyfrowych w dziedzinie przestrzennej uzyskuje się wykorzystując operację splotu (mnożenie dwóch transformat w dziedzinie częstotliwości tj. transformaty obrazu i filtru, jest równoważne splotowi obrazu z filtrem w dziedzinie przestrzennej). Operacja splotu oblicza nową wartość piksela obrazu na podstawie wartości pikseli sąsiadujących. Każda wartość piksela sąsiadującego jest odpowiednio wagowana i wpływa na końcową wartość piksela obrazu po filtracji zgodnie ze wzorem [38] :

$$P_i = \left[ \frac{\sum_{k=1}^K \sum_{l=1}^K P_{kl} * F_{kl}}{N} \right] \quad (5)$$

gdzie:

$P_i$  - wartość kolejnego piksela po filtracji,

$K$  - rząd macierzy filtru, (dla  $K=3$ , filtr  $3 \times 3$ ),

$P_{kl}$  - kolejna wartość piksela obrazu oryginalnego,

$F_{kl}$  - kolejna wartość wagi filtru,

$N$  - suma wartości wag filtru, lub 1 gdy suma wynosi 0.

Filtry możemy podzielić na [33]:

- przekształcenia kontekstowe: filtry konwolucyjne, logiczne i statystyczne
- przekształcenia widmowe: wykorzystujące transformatę Fouriera

Filtr Sobela jest filtrem konturowym zaprojektowanym w celu wykrywania krawędzi pionowych i poziomych, z filtrem konwolucji o rozmiarze 3

1	2	1	-1	0	1
0	0	0	-2	0	2
-1	-2	-1	-1	0	1

Tabela 2: Filtr Sobela

Filtr Robertsza został stworzony do szybkich obliczeń krawędziowych, których wynikiem są pochodne kierunkowe dla kierunków  $135^\circ$  i  $45^\circ$ .

1	0	0	1
0	-1	-1	0

Tabela 3: Filtr Robertsza

Filtr Prewitta jest operatorem podobnym do filtra Sobela, służącego do wykrywania krawędzi wertykalnych i horyzontalnych.

1	2	1	-1	0	1
0	0	0	-2	0	2
-1	-2	-1	-1	0	1

Tabela 4: Filtr Prewitta

Filtr Laplace'a wykorzystuje operator laplasjana, będącego sumą pochodnych cząstkowych drugiego rzędu danej funkcji względem każdego wymiaru:

$$L[f(x, y)] = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (6)$$

Filtr Laplace'a wykrywa regiony dużej zmiany jasności obrazu, może więc być wykorzystywany do wykrywania krawędzi na obrazie

0	1	0	1	1	1	-1	2	-1
1	-4	1	1	-8	1	2	-4	2
0	1	1	1	1	1	-1	2	-1

Tabela 5: Filtry Laplace'a



Rysunek 19: Przykłady rezultatów działania filtrów obrazów. Źródło: [34]

### 3.2.5 Algorytm Lucasa Kanade

Mając 2 obrazy  $I$  oraz  $J$ , oraz punkt na obrazie  $I$   $u = [u_x \ u_y]^T$  celem algorytmu do śledzenia cech jest znalezienie położenia punktu  $v = u + d = [u_x + d_x \ u_y + d_y]^T$  na drugim obrazie  $J$ . Wektor  $d = [d_x \ d_y]^T$  jest prędkością zmian sekwencji obrazów w punkcie  $x$ , znanym również jako przepływ optyczny w punkcie  $x$ .

Wektor  $d$  jest zdefiniowany jako wektor, który minimalizuje funkcję:

$$\epsilon(d) = \epsilon(d_x, d_y) = \sum_{x=u_x-\omega_x}^{u_x+\omega_x} \sum_{y=u_y-\omega_y}^{u_y+\omega_y} (I(x, y) - J(x + d_x, y + d_y))^2 \quad (7)$$

Dwoma najważniejszymi czynnikami dla algorytmów jest jego dokładność oraz odporność na zmiany. Aby zapewnić dokładność należy wybrać jak najmniejszą wartość okna integracji  $(\omega_x, \omega_y)$  tak aby nie 'wygładzić' detali zawartych na obrazach. Z kolei odporność na zmiany odpowiada za czułość algorytmu na zmiany oświetlenia czy skali ruchu na obrazie. Aby móc to realizować należałoby wybrać jak największe okno integracji. Jak widać, wybór odpowiednich wartości  $(\omega_x, \omega_y)$  należy do ważnego czynnika poprawności działania algorytmu i odpowiedniej relacji między jego dokładnością a odpornością na zmiany. Zaproponowana przez Jeana-Yves Bouguet'a, z Microprocessor Research Labs z firmy Intel, w [6] piramidalna reprezentacja obrazów zapewnia dobry kompromis między tymi czynnikami. Obraz oryginalny będzie poziomem zerowym obrazu  $I^0 = I$ . Kolejne poziomy budowane są w sposób rekursywny. Obraz na poziomie  $L$  jest zdefiniowany następująco:

$$\begin{aligned}
I^L(x, y) &= \frac{1}{4}I^{L-1}(2x, 2y) \\
&+ \frac{1}{8}(I^{L-1}(2x-1, 2y) + I^{L-1}(2x+1, 2y) + I^{L-1}(2x, 2y-1) + I^{L-1}(2x, 2y+1)) \\
&+ \frac{1}{16}(I^{L-1}(2x-1, 2y-1) + I^{L-1}(2x+1, 2y-1) \\
&\quad + I^{L-1}(2x-1, 2y+1) + I^{L-1}(2x+1, 2y+1))
\end{aligned} \tag{8}$$

Przyjmując

$$A(x, y) \doteq I^L(x, y) \tag{9}$$

$$B((x, y) \doteq J^L(x + g_x^L, y + g_y^L) \tag{10}$$

gdzie  $g^L = [g_x^L \quad g_y^L]^T$  to przepływ optyczny na poziomie  $L$  dążymy do tego aby równanie

$$\epsilon(\bar{\nu}) = \epsilon(\nu_x \nu_y) = \sum_{x=p_x-\omega_x}^{p_x+\omega_x} \sum_{y=p_y-\omega_y}^{p_y+\omega_y} (A(x, y) - B(x + \nu_x, y + \nu_y))^2 \tag{11}$$

dla wektora przemieszczenia  $\bar{\nu} = [\nu_x \quad \nu_y]^T$  w pierwszej pochodnej  $\epsilon$  wynosiło zero.

$$\frac{\partial \epsilon(\bar{\nu})}{\partial \bar{\nu}} \Big|_{\bar{\nu}=\bar{\nu}_{opt}} = [0 \quad 0] \tag{12}$$

Obraz różnicowy  $\delta I$  definiujemy jako:

$$\delta I(x, y) \doteq A(x, y) - B(x, y) \tag{13}$$

Pochodne kierunkowe oznaczamy jako:

$$I_x(x, y) = \frac{\partial A(x, y)}{\partial x} = \frac{A(x+1, y) - A(x-1, y)}{2} \tag{14}$$

$$I_y(x, y) = \frac{\partial A(x, y)}{\partial y} = \frac{A(x, y+1) - A(x, y-1)}{2} \tag{15}$$

Oznaczając dodatkowo:

$$G \doteq \sum_{x=p_x-\omega_x}^{p_x+\omega_x} \sum_{y=p_y-\omega_y}^{p_y+\omega_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{16}$$

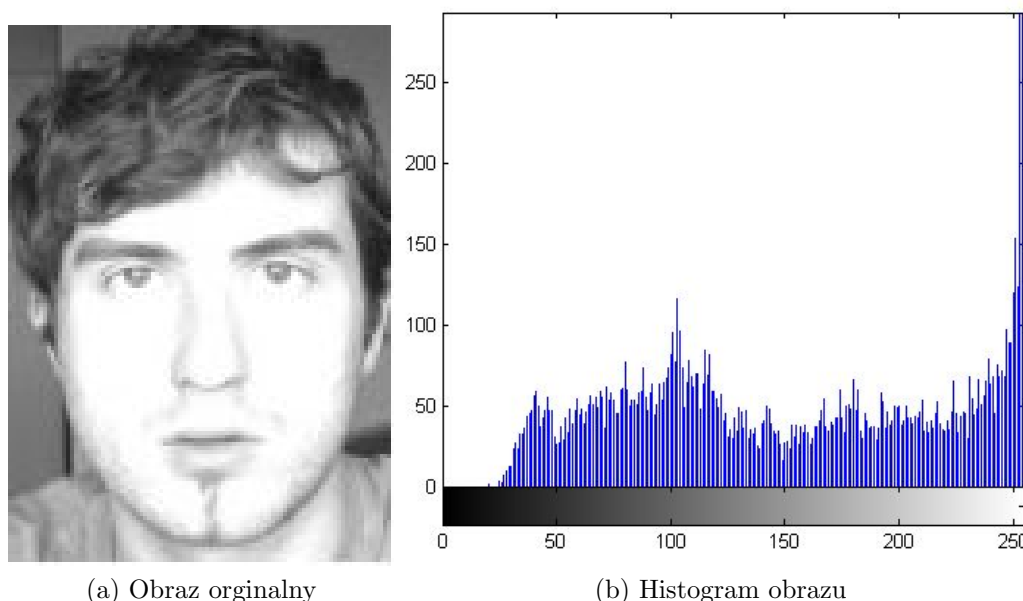
$$\bar{b} \doteq \sum_{x=p_x-\omega_x}^{p_x+\omega_x} \sum_{y=p_y-\omega_y}^{p_y+\omega_y} \begin{bmatrix} \delta I I_x \\ \delta I I_y \end{bmatrix} \tag{17}$$

Po serii przekształceń równania 12 otrzymujemy warunek optymalnego przepływu optycznego:

$$\bar{\nu}_{opt} = G^{-1} \bar{b} \tag{18}$$

### 3.2.6 Histogram

Histogram obrazu w skali szarości w przedziale  $[0, L - 1]$  jest dyskretną funkcją  $h(r_k) = n_k$  gdzie  $r_k$  jest k-tym poziomem szarości a  $n_k$  jest liczbą pikseli na obrazie o poziomie szarości  $r_k$  [15].



Rysunek 20: Histogram obrazu

Ze specjalnym przypadkiem histogramu, mamy do czynienia, gdy procesujemy obraz binarny. Możemy wtedy zliczać występowanie pikseli oznaczonych jako 1 w każdej kolumnie. Mówiąc ściślej, projekcją (histogramem) horyzontalnym jest funkcja

$$h_h(x) = \sum_{y=y_{min}}^{y_{max}} I(x, y) \quad (19)$$

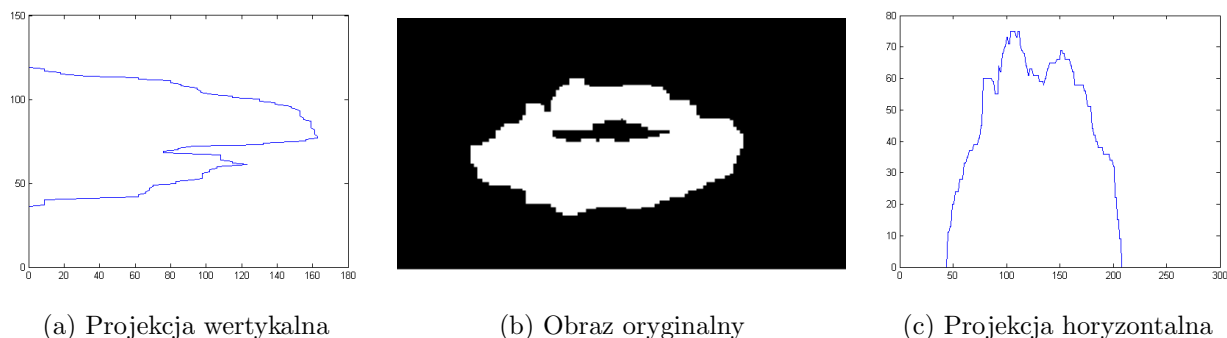
Analogicznie projekcją (histogramem) wertykalnym jest suma wystąpień pikseli w poszczególnych wierszach:

$$h_v(y) = \sum_{x=x_{min}}^{x_{max}} I(x, y) \quad (20)$$

### 3.2.7 Color slicing

Zaproponowana przez Rafaela Gonzaleza w [15] metoda wyróżniania określonych kolorów na obrazie może służyć do odseparowania obiektów od ich tła. Metoda ta polega na oznaczeniu





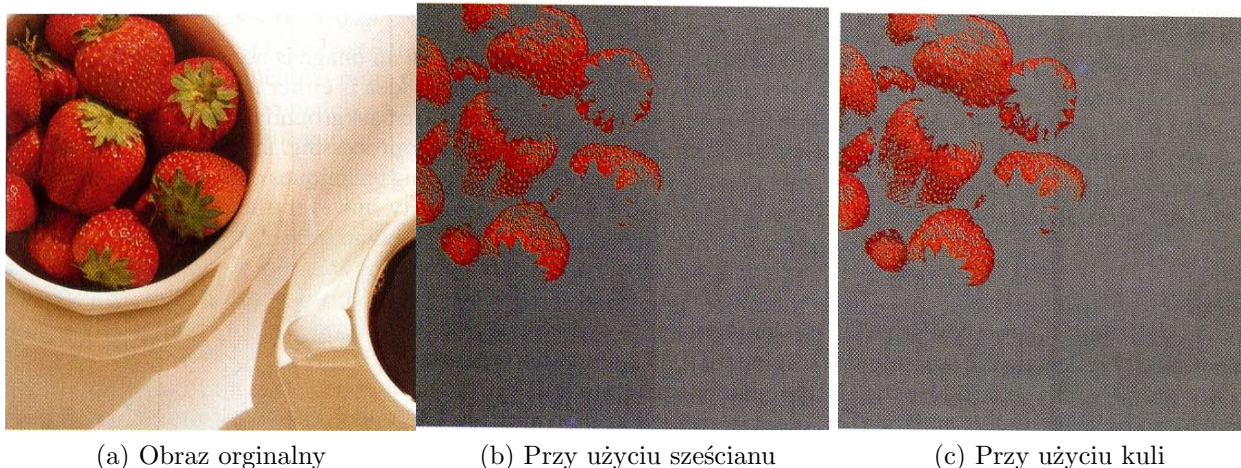
Rysunek 21: Histogramy wertykalny oraz horyzontalny

kolorów spoza określonego zakresu, kolorem neutralnym (np. szarym albo czarnym). Jeśli kolory zaprezentujemy jako sześcian o szerokości  $W$  o środku o kolorze o składowych  $(a_1, a_2, a_3)$  to przekształcenie wyróżniające określony kolor wygląda następująco:

$$s_i = \begin{cases} 0.5 & \text{jeśli } [|r_j - a_j| > \frac{W}{2}]_{1 \leq j \leq n} \\ r_i & \text{w przeciwnym wypadku} \end{cases} \quad (21)$$

Możemy również osiągnąć ten efekt zastępując sześcian kulą o promieniu  $R$ . Wtedy zależność ta przyjmuje postać:

$$s_i = \begin{cases} 0.5 & \text{jeśli } \sum_{j=1}^n (r_j - a_j)^2 > R^2 \\ r_i & \text{w przeciwnym wypadku} \end{cases} \quad (22)$$



Rysunek 22: Rezultat działania algorytmu Color Slicing. Źródło: [15]

### 3.3 Detekcja twarzy

Dla poprawnego działania całej aplikacji niezbędne jest wykrycie twarzy na obrazie wideo. Sposobem na uzyskanie pożądanego efektu jest analiza różnicowa poszczególnych klatek. Metoda ta bazuje na założeniu, iż głowa jest najwyżej położonym punktem ciała (wyklucza to udział rąk w okazywaniu emocji) oraz, że jest w nieustannym (nawet minimalnym) ruchu. Przy normalnej sytuacji, użytkownika siedzącego przed komputerem założenia te są spełnione.

W najprostszym przypadku można by odejmować kolejne klatki od siebie. Metoda taka jednak obciążona jest sporym błędem z powodu delikatnych zmian oświetlenia, które zachodzą nieustannie. Dużo lepszym rozwiązaniem będzie wyznaczenie osobno tła, a następnie odejmowanie poszczególnych klatek od tego tła. W idealnej sytuacji najpierw nagrywano by tło ujęcia (najlepiej o jednolitym tle), a następnie do tak przygotowanego “studia” wchodziłby ochotnik i badano by jego stan emocjonalny. W przypadku tej pracy nie dysponujemy obrazem tła, gdyż użytkownik od razu siedzi przed kamerą, oraz sekwencje wideo nagrywane są w różnych pomieszczeniach. Tło należy więc wyznaczyć na podstawie poprzednich klatek, uznając, iż wszystkie piksele, które na przestrzeni czasu nie zmieniły swojej wartości są punktami tła, natomiast wszystkie pozostałe są regionami poruszającego się ciała (lub punktami gdzie przed chwilą to ciało było).

Należy tutaj wziąć poprawkę na możliwe zmiany wartości punktów będących tłem, wynikających z zachodzenia samoistnych zmian oświetlenia na obrazie wideo mających dwojakie pochodzenie. Po pierwsze może to wynikać z różnic w świetle naturalnym (np. przesłonięciem Słońca przez chmury), po drugie może to wynikać z zachowania matrycy CCD oraz algorytmów automatycznej korekcji balansu bieli na obrazie wideo, występujących w większości obecnie dostępnych kamer internetowych. Problemy z oświetleniem przy realizacji tej pracy są szerzej opisane w Rozdziale 5.1.4.

Tło należy więc wyliczyć jako zbiór punktów, które w czasie nie ulegały dużym zmianom (bądź ulegały im sporadycznie). Można by to osiągnąć wyliczając średnią z ostatnich kilkunastu klatek. Wymagałoby to jednak dużego nakładu pamięciowego jak i wykonywałoby spory narzut obliczeniowy związany w wyliczaniem średniej klatek w każdym roku. Aby więc zaoszczędzić na pamięci ramek oraz przyspieszyć działanie algorytmu zastosowano następujące rozwiązanie. Obraz różnicowy jest wynikiem odejmowania aktualnej klatki od aktualnego obrazu tła:

$$I_{diff} = I_i - B_i \quad (23)$$

Wartość tła natomiast aktualizowana jest w każdym kroku jako część starego tła uzupełniana resztą informacji z bieżącej klatki:

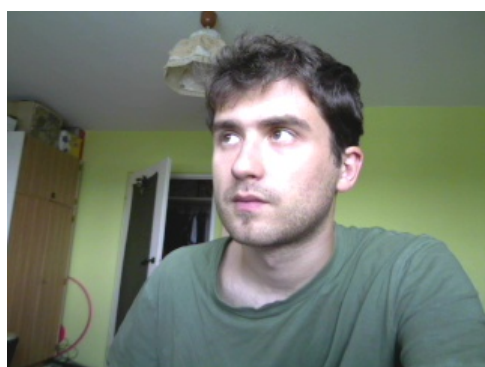
$$B_{i+1} = \alpha * I_i + (1 - \alpha) * B_i; \quad \alpha \in (0, 1) \quad (24)$$

Na podstawie testów przyjęto wartość  $\alpha = 0.85$ .

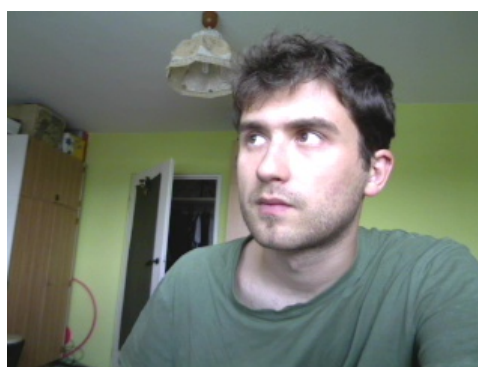
Tak otrzymany obraz różnicowy jest następnie poddawany algorytmowi Connected Component Labeling w celu wykrycia największego poruszającego się elementu na obrazie (w założeniu ciała użytkownika). Następnie wybierany jest najwyżej położony punkt z tego regionu, który zostaje

uznany za czubek głowy. Jako górny punkt twarzy oznaczany jest punkt będący na wysokości 10 % wysokości wykrytego regionu. Pomijamy w ten sposób włosy, które nie wnoszą żadnych interesujących informacji i ograniczamy rejon analizy stanów emocjonalnych od czoła, gdzie stany te mogą się pojawić. Oznaczony w ten sposób rejon może być zbyt szeroki ze względu na możliwość zarejestrowania ruchu barków użytkownika. Aby wyznaczyć faktyczną szerokość twarzy w hipotetycznym miejscu występowania oczu, na  $1/3$  wysokości regionu, badamy skrajne występowanie pikseli z etykietą i oznaczamy te punkty jako szerokość twarzy. W niniejszej pracy przyjęto, iż ludzka twarz proporcje szerokości do długości ma w stosunku  $3/4$ . Z proporcji tej oraz otrzymanej przed chwilą szerokości możemy wyliczyć długość twarzy. Przycinając znaleziony na początku region do tych obostrzeń, otrzymujemy miejsce występowania twarzy na obrazie wideo.

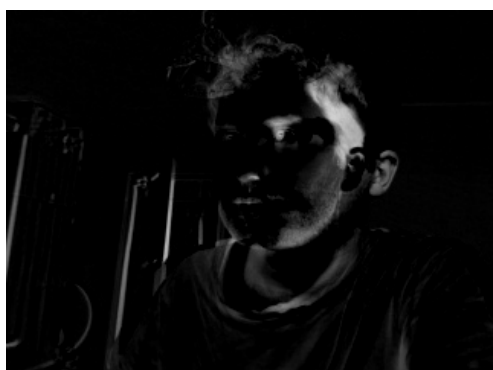
W przypadku niepowodzenia wykrycia twarzy (region byłby zbyt mały, lub występował w znacznej różnicy od poprzednio wykrytego regionu) za region twarzy przyjmowany jest region z poprzedniej klatki.



(a) Pierwsza klatka



(b) Druga klatka



(c) Obraz różnicowy



(d) Wynik działania algorytmu

Rysunek 23: Działanie algorytmu detekcji twarzy

### 3.4 Wykrywanie brwi

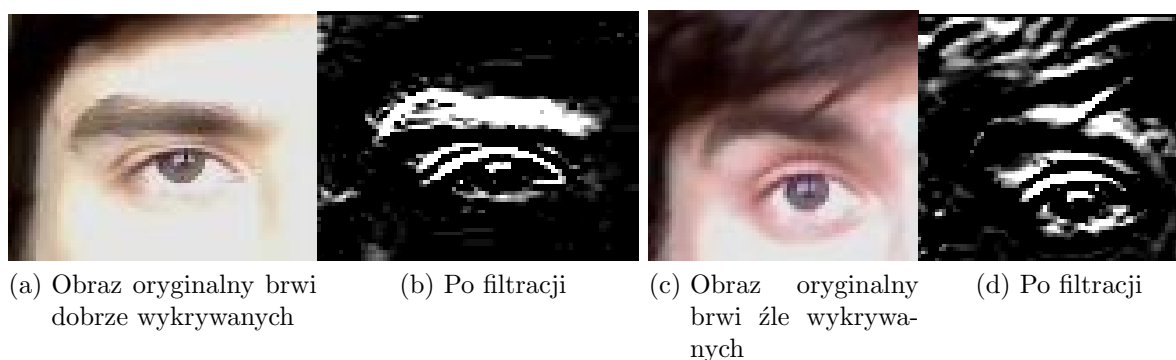
Łuk brwiowy wyznacza środkową linię twarzy. Powszechnym sposobem określania rozmiarów łuku brwiowego jest wykorzystanie rozmiarów oczu. Szerokość twarzy wzdłuż łuku brwiowego powinna być równa pięciokrotnej szerokości oka [5].

Wykrywanie brwi początkowo dokonywane jest przy założeniu, iż są to ciemne horyzontalne linie. Wykrywane są one przy użyciu filtru Sharra.

$$\text{Element strukturalny filtru Sharra} = \begin{array}{|c|c|c|} \hline 3 & 10 & 3 \\ \hline 0 & 0 & 0 \\ \hline -3 & -10 & -10 \\ \hline \end{array}$$

Tak uzyskany obraz podawany jest operacji zamknięcia (erozji a następnie dylatacji) w celu wyostrenia różnic i likwidacji szumów.

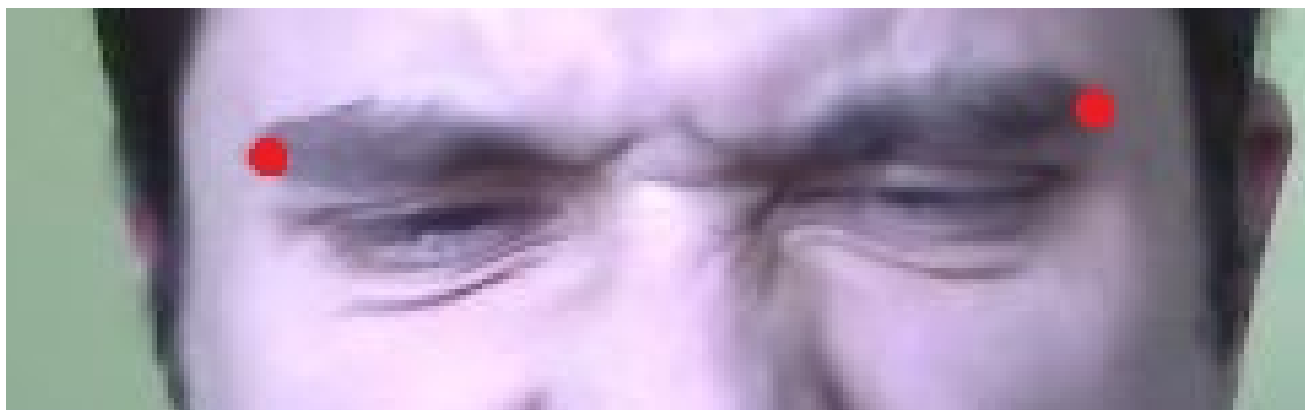
Algorytm ten nie wykrywa jednak dokładnie całości brwi, gdyż ze względu na ich naturalną budowę bądź też aktualny stan emocjonalny, zawierają one również linie skośne.



Rysunek 24: Działanie algorytmu detekcji brwi przy użyciu filtru Sharra

Aby poprawić wykrywanie brwi, w następnych klatkach, kiedy znane jest już położenie brwi, używany jest algorytm bazujący na segmentacji koloru, opisany w Rozdziale 3.2.7. Kolor bazowy może być ustawiony osobno dla brwi lewej i prawej, ze względu na możliwość wystąpienia bocznego oświetlenia, które może spowodować znaczne różnice w kolorze między lewą stroną twarzy a prawą (prześwietlenia od strony doświetlonej oraz cienie od strony bez źródła światła).

Obraz po binaryzacji i operacji otwarcia, poddawany jest operacji etykietowania, opisanej w Rozdziale 3.2.1, w celu odnalezienia 2 najszerszych regionów, będących, w zależności od wzajemnego położenia, brwią lewą i prawą. Każda brew opisana poprzez 3 punkty: skrajnie lewy punkt  $L$ , najwyżej położony piksel  $T$ , oraz skrajnie prawy  $R$  (Rysunek 26).



(a) Kolor zaznaczonego punktu (140,129,163)

(b) Kolor zaznaczonego punktu (74,67,81)

Rysunek 25: Różnice w kolorach dla twarzy z oświetleniem bocznym



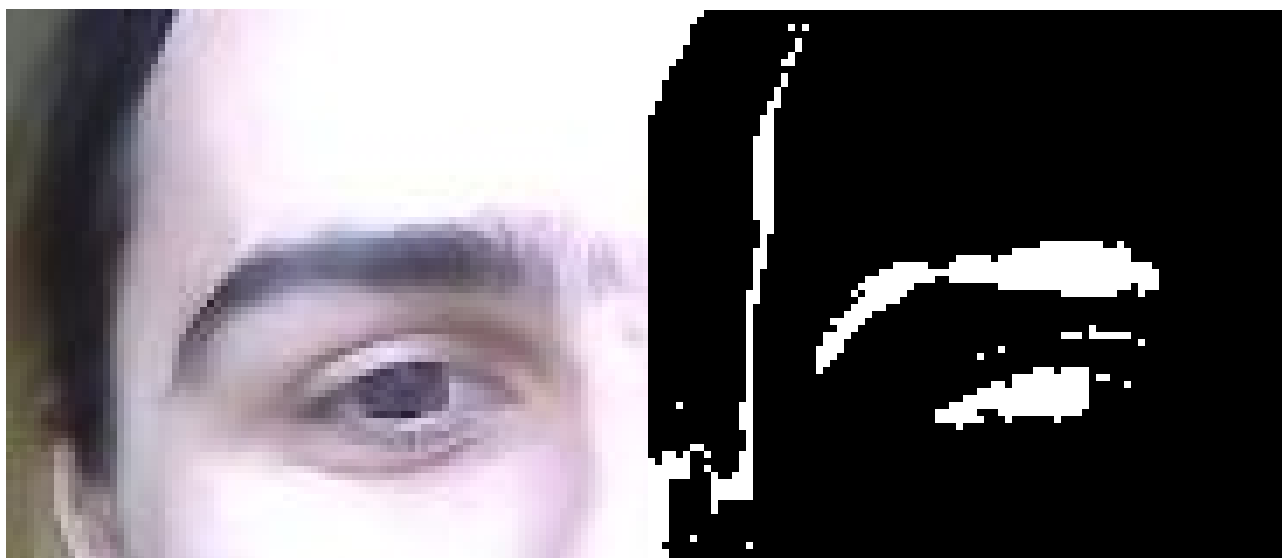
Rysunek 26: Punkty opisujące każdą brew

### 3.5 Wykrywanie oczu

Oczodół rozciąga się wzdłuż połowy długości nosa, dochodząc do szczytu kości policzkowej (jarzmowej). Gałka oczna ma średnicę około 3cm i jest prawie idealnie okrągła. Odległość dzieląca obie gałki oczne równa jest szerokości jednego oka[5].

Ich rolą jest wzmacnianie, a w wielu przypadkach także modyfikowanie wyrazu twarzy utworzonego przez usta[5].

W celu wykrycia oczu zastosowano metodę przedstawioną w pracy [17]. Metoda ta bazuje na założeniu, iż oko ludzkie składa się zarówno z jasnych, jak i ciemnych obszarów. Używając operacji dylatacji oraz erozji z kulistym elementem strukturalnym otrzymujemy region określający



(a) Obraz oryginalny

(b) Wynik działania algorytmu

Rysunek 27: Działanie algorytmu detekcji brwi przy użyciu algorytmu Color Slicing dla sześciangu o środku w (95,90,112) i szerokości  $W=35$

położenie oczu, przy pomocy wzoru:

$$EyeMap = \frac{Y(x, y) \oplus g(x, y)}{Y(x, y) \otimes g(x, y)}$$

gdzie  $\oplus$  to operacja dylatacji w skali szarości,  $\otimes$  to operacja erozji w skali szarości, a  $g(x, y)$  to okrągły element strukturalny o wymiarach  $5 \times 5$ .

$$g(x, y) = \begin{array}{|c|c|c|c|c|} \hline 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline 0 & 1 & 1 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 \\ \hline \end{array}$$

Otrzymany obraz poddawany jest binaryzacji, a następnie algorytmowi CCL, w celu odnalezienia 2 największych regionów. Mając wykryte brwi można ograniczyć rejon poszukiwań do regionu poniżej poziomu każdej brwi oraz w ramach jej szerokości. Odnaleziony region oka indeksowany jest poprzez 2 punkty: lewy górny róg  $LT$  oraz prawy dolny  $RB$  (Rysunek 29).

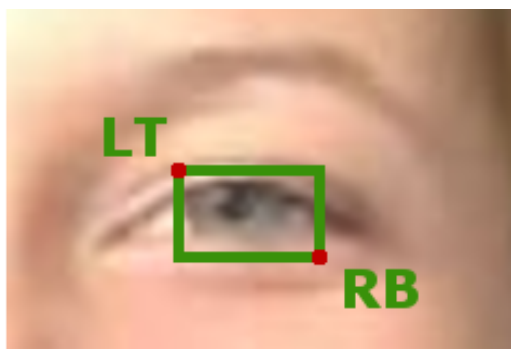
### 3.6 Wykrywanie ust

Usta zajmują w sumie dwie trzecie powierzchni rozciągającej się pomiędzy czubkiem nosa i podbródkiem. Kąciki ust znajdują się zwykle w jednej linii ze środkiem odpowiadającej im gałki ocznej[5].



(a) Obraz oryginalny      (b) Wynik działania algorytmu      (c) Wynik po binaryzacji

Rysunek 28: Działanie algorytmu detekcji oczu



Rysunek 29: Punkty opisujące każde oko

Usta są elementem definiującym wyraz twarzy, czyli jego podstawą. Tworzone przez nie kształty mają zawsze to samo znaczenie, podczas gdy oczy nie są wcale konsekwentne w wyrażaniu emocji[5].

W celu wykrycia ust posłużono się metodą zaproponowaną w [22], polegającą na wstępnym wykryciu położenia ust, bazującym na następującym założeniu:

1. Oznaczmy jako  $(x_L, y_L)$  współrzędne położenia środka lewego oka, i  $(x_R, y_R)$  dla prawego oka
2. Wyznaczamy  $d = \sqrt{(x_R - x_L)^2 + (y_R - y_L)^2}$  odległość między środkami oczu
3. Prowadzimy prostą prostopadłą, o długości  $d$ , ze środka odcinka między oczami
4. Koniec tego odcinka, punkt  $M$  o współrzędnych  $(x_M, y_M)$  wyznacza teoretyczny środek obszaru ust



Rysunek 30: Wstępne wykrywanie obszaru ust

5. Wstępny obszar ust wyznacza prostokąt o wierzchołkach o współrzędnych:

- $x_{MU} = x_M - d * 0.1$
- $y_{MU} = y_M - d * 0.55$
- $x_{MD} = x_M + d * 0.7$
- $y_{MD} = y_M + d * 0.55$

Tak wykryty obszar poddawany jest następnie dalszej analizie w celu wykrycia dokładnego kształtu ust. Wykorzystywane jest tutaj spostrzeżenie, iż usta dość istotnie różnią się kolorystycznie od pozostałych części twarzy.

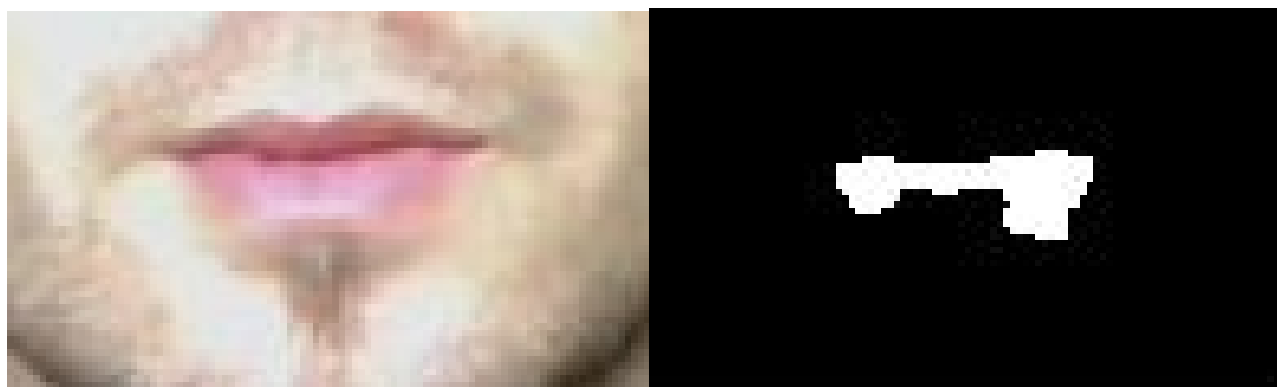
Można zastosować tutaj algorytm Color Slicing z wykrywaniem rejonów o kolorze zbliżonym do koloru czerwonego. Podejście to ma jednak sporo wad, ze względu na różnice w kolorach pomiędzy wargą górną a dolną, występowanie ciemniejszych rejonów w przestrzeni między wargami, jak i w kącikach ust. Poważnym problemem w wykrywaniu ust jest również występowanie na ustach miejsc prześwietlonych (jasnych obszarów), wynikających ze złego zogniskowania kamery w przypadku poruszenia się obrazu (przykłady na Rysunku 32)

Lepszym rozwiązaniem w celu wykrycia obszaru ust jest zastosowanie przestrzeni barw HSV (Hue - Saturation - Value). Składowa nasycenia dobrze pokazuje różnice dla 'czystych' kolorów. W przypadku ust kolor czerwony ukazywany jest w sposób bardzo zdecydowany (Rysunek 34).

Na wstępnie wykrytym obszarze ust dokonuje się binaryzacji tablicy wartości składowej nasycenia, z przyjętym progiem binaryzacji. Tak wykryty obszar poddawany jest operacji etykietowania, a następnie usuwane (oznaczane jako tło) są wszystkie rejony za wyjątkiem jednego z największą liczbą etykiet (największą powierzchnią).

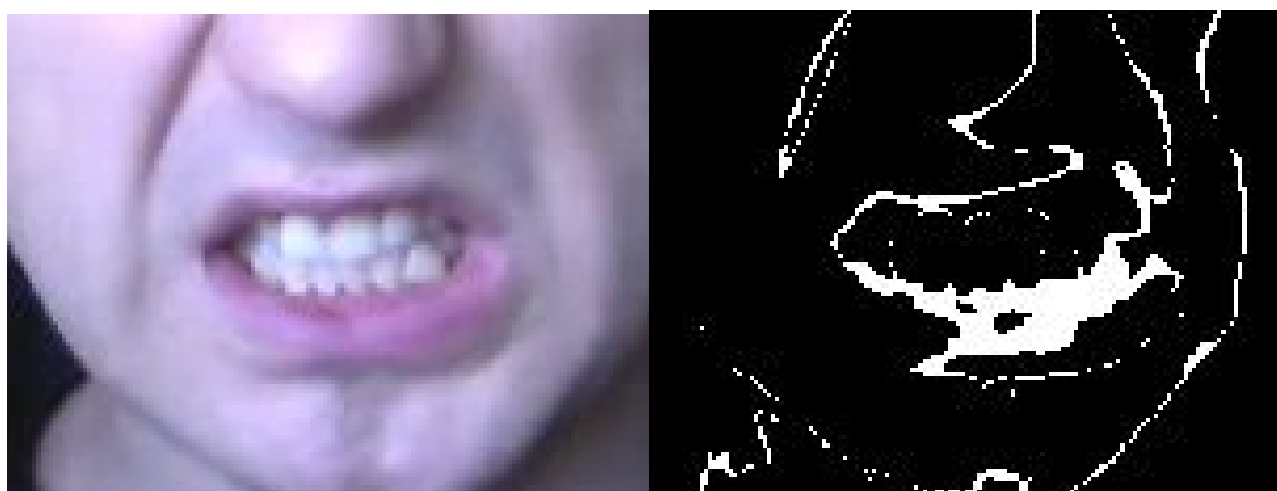
W celu wyznaczenia położenia kącików ust wyliczana jest projekcja horyzontalna wyznaczonego obszaru. Pierwsze oraz ostatnie niezerowe wartości tej projekcji wyznaczają położenie pionowe odpowiednio kącików lewego i prawego (odpowiednio punkty  $L$  i  $R$ ). W celu wyznaczenia współ-





(a) Obraz oryginalny

(b) Wynik działania algorytmu Color Slicing



(c) Obraz oryginalny

(d) Wynik działania algorytmu Color Slicing

Rysunek 31: Przykłady działania algorytmu Color Slicing na obszarze ust

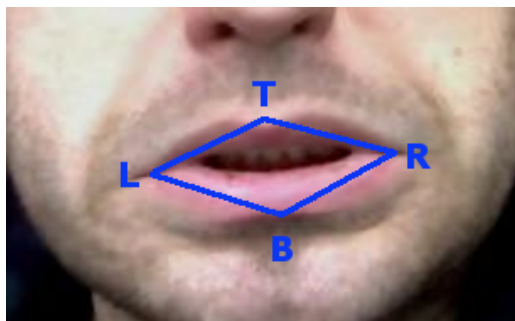


Rysunek 32: Przykłady prześwietlenia na obszarze ust

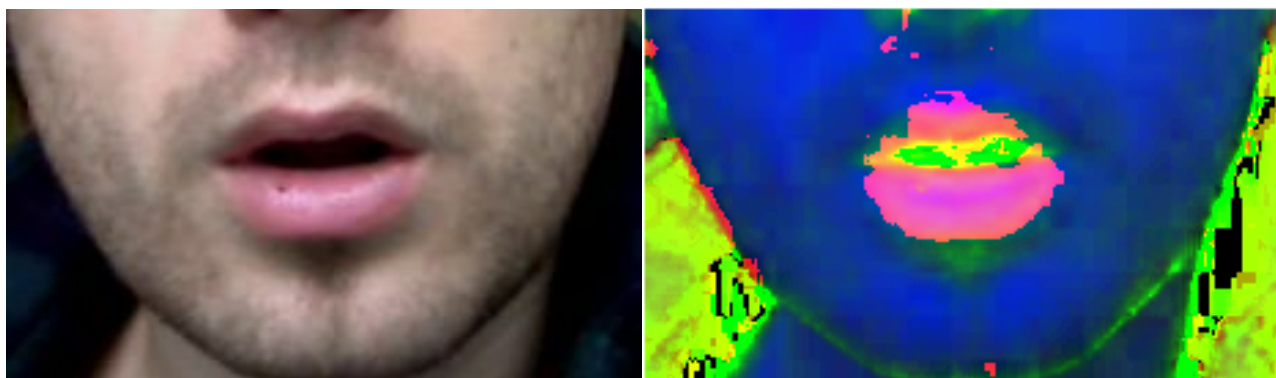
rzędnej poziomej wyszukiwane (od góry do dołu) jest pierwsze wystąpienie pikseli oznaczonych jako nie-tło.

Ze względu na budowę ust założono, iż współrzędna pionowa wargi dolnej (punkt  $B$ ) i górnej (punkt  $T$ ) znajduje się w połowie odległości między kącikami ust. Rozwiązanie takie pozwala

na uniknięcie wykrywania jako najwyższego punktu górnej wargi na przemian lewego i prawego wznórka na wardze. Położenie w poziomie wyznaczone jest jako pierwsze (ostatnie) wystąpienie zaetykietowanego piksela na wysokości ustalonej w poprzednim kroku.

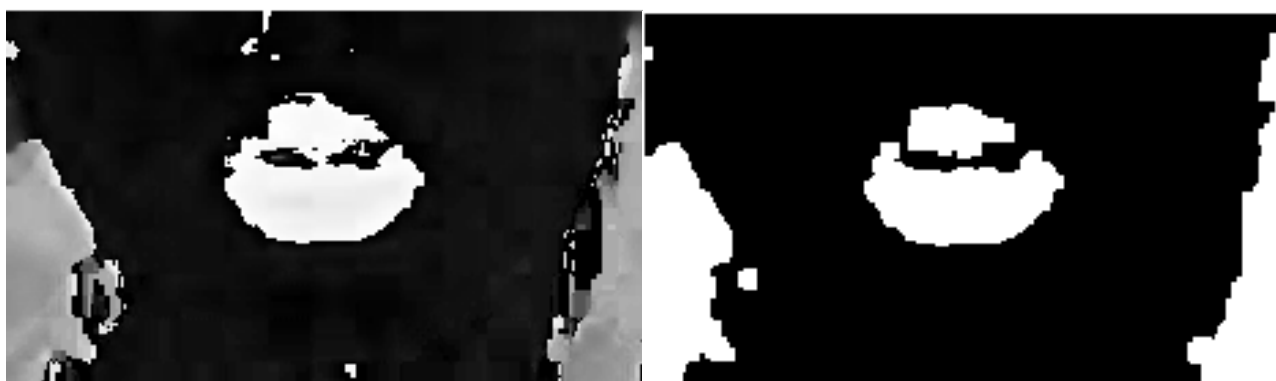


Rysunek 33: Punkty opisujące usta



(a) Obraz oryginalny w modelu RGB

(b) Obraz w przestrzeni barw HSV



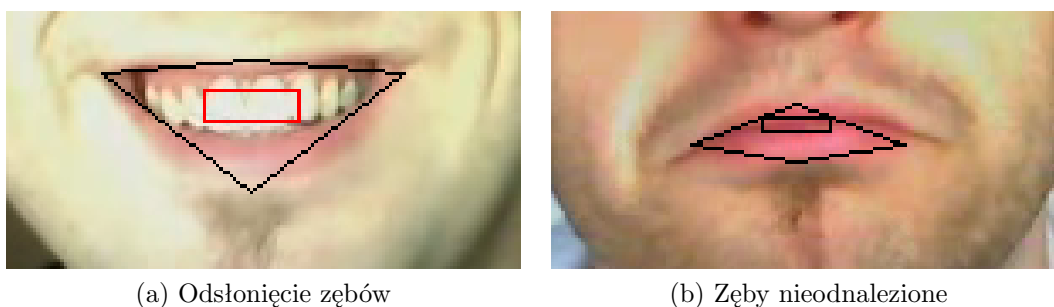
(c) Wartość nasycenia (składowa Hue)

(d) Obraz po binaryzacji i operacji otwarcia

Rysunek 34: Działanie algorytmu detekcji ust

### 3.6.1 Wykrycie odsłonięcia zębów

Pojawienie się zębów oznacza zazwyczaj istnienie silnej emocji ("szczerzymy zęby" w agresji, bądź też pokazujemy je w szerokim uśmiechu), dlatego też wykrycie takiego zjawiska może być bardzo użyteczne. Aby tego dokonać skorzystano z obserwacji, iż zęby znacząco różnią się kolorystycznie od reszty regionu ust i mają bardzo jasny kolor. W prostokącie o środku w środku regionu ust o szerokości równej  $1/3$  szerokości ust, i wysokości równej  $1/4$  wysokości regionu ust zliczany jest średni poziom jasności pikseli (wartość piksela po przejściu na skalę szarości). Jeśli średnia ta przekracza poziom tolerancji (przyjęty na 210), to przyjmuje się, iż zęby pojawiły się na obrazie.



Rysunek 35: Detekcja odsłonięcia zębów (kolor czerwony oznacza pojawienie się zębów)

## 3.7 Śledzenie zmian

Aby z każdą kolejną klatką nie przeprowadzać tych samych operacji na sporym obszarze, które są dość czasochłonne, stosowane są algorytmy śledzenia wykrytych już cech. Dostępne są 2 podejścia. Pierwsze, prostsze, wylicza obszar potencjalnego występowania konkretnego elementu twarzy, jako region tego elementu z poprzedniej klatki powiększony o 10%. Na tak przygotowanym obszarze uruchamiane są algorytmy zaprezentowane wcześniej. Daje to spory zysk w odciążeniu pracochłonności procesora, gdyż badane obszary są wielkości kilkuset pikseli a nie jak wcześniej paru tysięcy pikseli. Potencjalne zagrożenie dla takie podejścia są gwałtowne ruchy, które mogą spowodować "wyskoczenie" danego elementu poza przewidywany rejon.

Drugim podejściem jest zastosowanie zaprezentowanego w Rozdziale 3.2.5 algorytmu śledzenia przepływu optycznego Lucasa - Kanade. Algorytm ten daje dobre rezultaty, jednak wadami jego jest szybkość działania, która jest porównywalna z oryginalnymi algorytmami detekcji, oraz wrażliwość na wyjściowy punkt poszukiwań, który musi spełniać określone warunki, aby śledzenie było skuteczne, co nie jest spełnione w przypadku niektórych cech.

## 3.8 Graf twarzy

Na potrzeby tej pracy, użyto koncepcji zaproponowanej przez promotora pracy - dr Adriana Horzyka, grafu reprezentującego mimicznie charakterystyczne elementy/punkty twarzy wraz z

jego dynamiką, składającego się z 14 charakterystycznych węzłów:

- 3 węzły na każdą brew,
- 2 węzły na każde oko,
- 4 węzły na usta.

Na początku działania programu tworzone są 3 grafy:  $G_E$ ,  $G_H$ ,  $G_W$ , gdzie wartość krawędzi  $E(V_1, V_2)$  pomiędzy wierzchołkami  $(V_1, V_2)$  to lista zawierająca

$G_E$ : odległość euklidesowa pomiędzy  $V_1$  i  $V_2$  :  $E_E(V_1, V_2) = \sqrt{(V_{1x} - V_{2x})^2 + (V_{1y} - V_{2y})^2}$ ,

$G_H$ : wysokość pomiędzy  $V_1$  i  $V_2$  :  $E_H(V_1, V_2) = |V_{1y} - V_{2y}|$ ,

$G_W$ : szerokość pomiędzy  $V_1$  i  $V_2$  :  $E_W(V_1, V_2) = |V_{1x} - V_{2x}|$ .

Po zakończonej detekcji brwi, oczu i ust na twarzy, wykryte współrzędne są normalizowane względem odległości między środkami gałek ocznych  $EyeDist$  według wzoru:

$$V_N = \frac{V}{EyeDist} \quad (25)$$

co pozwala uniezależnić wyniki od cech biometrycznych użytkownika, a następnie dodawane są do listy krawędzi. Pozwala to na wyciągnięcie danych historycznych czy statystycznych (takich jak średnia wartość danej krawędzi, wartość maksymalna czy minimalna), które przydadzą się w późniejszej analizie stanu grafu.

### 3.9 Detekcja gestów twarzy

Mając sporządzony aktualny graf twarzy, można przystąpić do detekcji gestów, jakie pojawiają się na twarzy. Za gesty uznaje się stany, w jakich mogą znajdować się brwi, oczy i usta. Na podstawie badań literatury oraz własnych przemyśleń zaproponowano zestaw gestów wraz z warunkami ich wystąpienia. Wszystkie podane poniżej wartości progowe, czy też skale referencyjne można regulować z poziomu aplikacji.

Wszystkie wartości dla gestów dwubiegunowych (np brwi uniesione - brwi opuszczone), są skalowane do przedziału  $[-100, 100]$  według funkcji:

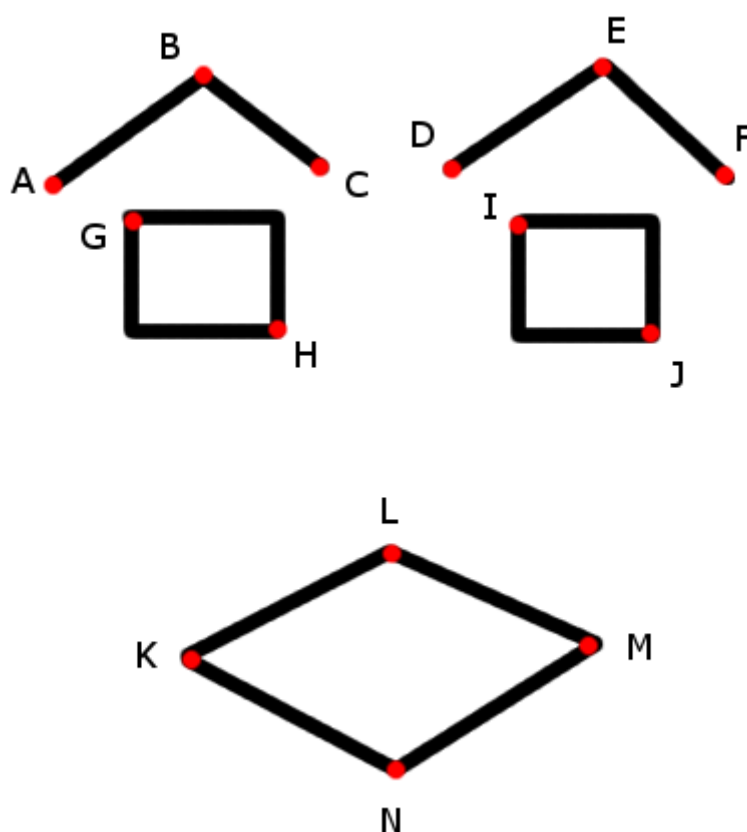
$$S(Ref_{Min}, Ref_{Max}, x) = \frac{200 * (x - Ref_{Min})}{Ref_{Max} - Ref_{Min}} - 100 \quad (26)$$

#### 3.9.1 Brwi uniesione/obniżone

Badana jest wysokość pomiędzy górnym punktem brwi  $TopEyebrow$  a podstawą oka  $BottomEye$ , dla każdej brwi osobno.

$$EyebrowElevated = E_H(TopEyebrow, BottomEye) \quad (27)$$

Wynik jest porównywany do przedziału progowego (0.2; 0.4) według wzoru 26.



Rysunek 36: Schemat grafu twarzy użyty w niniejszej pracy

### 3.9.2 Ściśnięcie brwi

W celu zbadania czy brwi zostały ściśnięte, sprawdzane jest, czy przestrzeń między brwiami pomarszczyła się. Badany jest prostokąt o końcach:

- w prawym wierzchołku lewej brwi  $RightEyebrow_{Left}$ ,
- lewym wierzchołku brwi prawej  $LeftEyebrow_{Right}$ ,
- wysokości równej  $\min(TopEyebrow_{Left}, TopEyebrow_{Right})$ .

Tak otrzymany obszar konwertowany jest do skali szarości a następnie, równoległe wykonywane są 3 filtry Sobela: wertykalny i skośny w obu kierunkach, które wyszukują pionowych bądź też ukośnych zmarszczek które występują wtedy pomiędzy brwiami. Otrzymane wyniki dodaje się do siebie binarnie, a następnie zlicza się obszary z wartością logiczną "1". Jeśli przekroczyła ona określony próg (ustalony na 9) wtedy brwi uznawane są za ściśnięte.



(a) Obraz oryginalny

(b) Obraz po binaryzacji

Rysunek 37: Wykrycie ściśnięcie brwi

### 3.9.3 Otwarcie/zamknięcie oczu

Aby stwierdzić poziom otwarcia czy też przymrużenia oczu, badana jest wysokość pomiędzy górną powieką  $TopEye$  a podstawą oka  $BottomEye$ :

$$EyeOpen = E_H(TopEye, BottomEye) \quad (28)$$

Otrzymany wynik skalowany jest w przedziale (0.05; 0.2).

### 3.9.4 Otwarcie ust

Badając czy usta są otwarte, czyli de facto poziom opadnięcia szczęki, mierzymy wysokość pomiędzy górną wargą  $TopLip$  a dolną wargą  $BottomLip$

$$LipsOpen = E_H(TopLip, BottomLip) \quad (29)$$

Skalą porównawczą w tym przypadku jest (0.2; 0.5)

### 3.9.5 Podniesienie/opuszczenie kąćków ust

Stan kąćków ust bada się w dwóch etapach. W pierwszym korzysta się z obserwacji iż, kąćki ust kiedy wykrzywają się w górę albo w dół rozszerzają również usta na boki. Badana jest więc szerokość pomiędzy lewym kąćkiem ust  $LeftCornerLips$  a prawym  $RightCornerLip$

$$LipsStreched = E_W(LeftCornerLips, RightCornerLip) \quad (30)$$

Jeśli wartość ta przekracza poziom progowy ustawiony na 0.63, usta oznaczane są jako *rozszerzone*, co jest warunkiem koniecznym dla algorytmu do przystąpienia do drugiego etapu rozpoznania stanu kąćków ust. W etapie tym badana jest wysokość pomiędzy kąćkami ust a górną i dolną wargą:

$$TopLipCorners = avg(E_H(LeftCornerLips, TopLip), E_H(RightCornerLip, TopLip)) \quad (31)$$

$$BottomLipCorners = avg(E_H(LeftCornerLips, BottomLip), E_H(RightCornerLip, BottomLip)) \quad (32)$$

Następnie sprawdza się czy *TopLipCorners* jest mniejsze od wartości referencyjnej (ustalonej na 0.1). Jeśli tak - oznacza to, że kąciki ust są uniesione. W przeciwnym wypadku wartość *BottomLipCorners* porównywana jest z wartością referencyjną (0.2) i zapisywana jako wartość obniżenia kącików ust.

### 3.10 Identyfikacja emocji

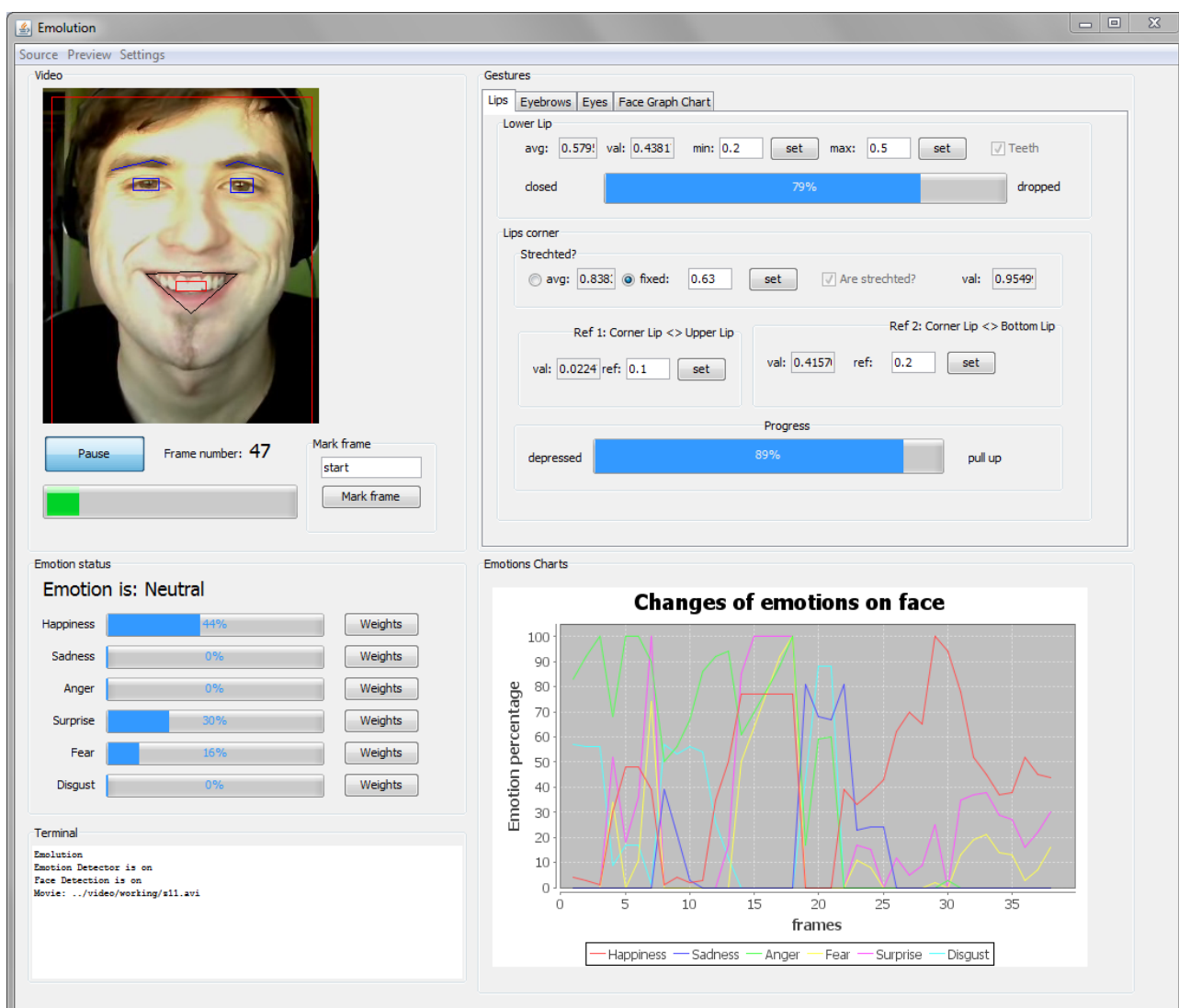
Na podstawie wiadomości zamieszczonych w Rozdziale 2.3 zaimplementowano system regułowy, który stara się zidentyfikować emocje na podstawie wykrytych gestów twarzy. System ten wylicza procentową siłę wystąpienia danej emocji na podstawie średniej ważonej z poszczególnych *Action Unitów*. Wagi te są konfigurowalne z poziomu aplikacji.

Emocja	Gest	Waga
Szczęście	Uniesienie brwi	2
	Przymrużenie oczu	3
	Uniesienie kącików ust	5
	Pojawienie się zębów	1
Zaskoczenie	Uniesienie brwi	1
	Otwarcie oczu	1
	Otwarcie ust	1
Odraza	Obniżenie brwi	4
	Ścisnięcie brwi	4
	Obniżenie kącików ust	5
Złość	Obniżenie brwi	5
	Przymrużenie oczu	1
	Obniżenie kącików ust	3
	Otwarcie ust	3
	Ścisnięcie brwi	5
	Pojawienie się zębów	2
Smutek	Uniesienie brwi	1
	Przymrużenie oczu	4
	Obniżenie kącików ust	4
	Zamknięcie ust	3
	Ścisnięcie brwi	1
Strach	Uniesienie brwi	5
	Otwarcie oczu	6
	Obniżenie kącików ust	2
	Otwarcie ust	5
	Ścisnięcie brwi	1
Emocja	Gest	Waga

Tabela 6: Wagi gestów dla poszczególnych emocji

## 4 Implementacja

Analiza teoretyczna zagadnienia rozpoznawania emocji, przedstawiona we wcześniejszych rozdziałach posłużyła do skonstruowania programu komputerowego "Emolution". Jest to samodzielna (typu *stand-alone*) aplikacja napisana w całości w języku Java. Język ten dzięki podejściu obiektowemu, ścisłej typizacji, dużemu wsparciu ze strony społeczności komputerowej, mnogości dostępnych, gotowych do wykorzystania bibliotek, rozbudowanych narzędzi wspierających tworzenie oprogramowania, pozwala na konstrukcje rozbudowanych systemów.

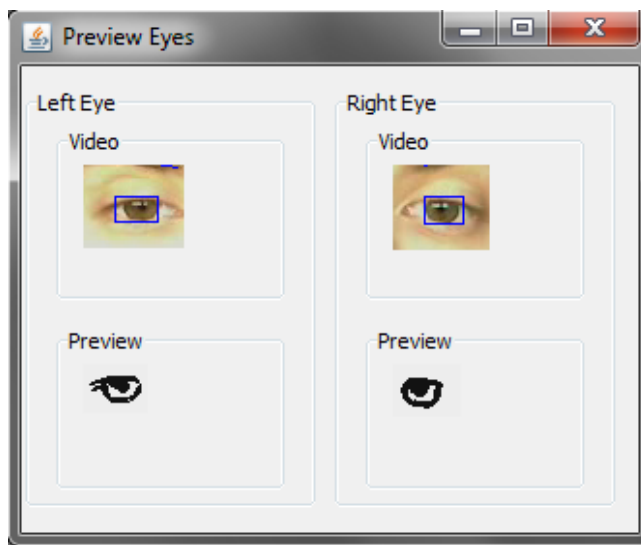


Rysunek 38: Zrzut ekranu aplikacji "Emolution"



## 4.1 Budowa programu

Program “Emolution” to około 14 000 linii kodu podzielonego na 50 klas napisanych w języku Java. Większość algorytmów została zaimplementowana samodzielnie na podstawie wcześniejszych badań przy użyciu programu Matlab, który umożliwia łatwe operacje na obrazach przy użyciu gotowych procedur. W niektórych przypadkach, opisanych poniżej posłużono się zewnętrznymi bibliotekami, które miały już zaimplementowane poszczególne algorytmy, bądź funkcjonalności.



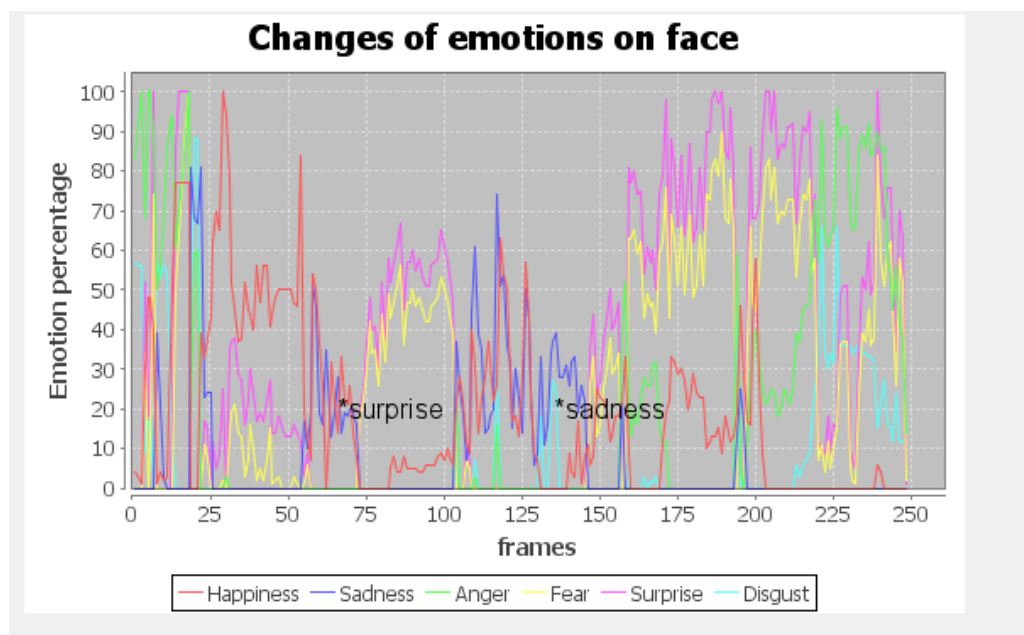
Rysunek 39: Podgląd działania algorytmu detekcji oczu

### 4.1.1 Wykorzystywane biblioteki

W czasie pracy nad aplikacją wykorzystano szereg bibliotek ułatwiających obliczenia czy wizualizację działania algorytmów. Poniżej przedstawiono najważniejsze z nich:

*Java Media Framework* [19] (JMF) to podstawowa biblioteka Javy do obsługi w niej multimediów, stworzona przez firmę Sun. Niestety biblioteka ta nie jest rozwijana od 2001 roku, co powoduje szereg problemów z utrzymaniem oraz działaniem biblioteki na najnowszych kodekach. Pozwala ona bowiem na obsługę jedynie formatu .vmw. Skorzystano więc z rozwiązania o nazwie: *Fobs4JMF* [14] będącego wrapperem oryginalnego frameworka, dodającego szereg nowych kodeków dzięki połączeniu z biblioteką *ffmpeg*. Wiąże się to niestety ze skomplikowaną konfiguracją oraz powolnym działaniem. Według zapewnień firmy Oracle, która wykupiła Suna wraz z Javą, w wersji 7.0 języka (która nie ukazała się przed zakończeniem prac nad niniejszą pracą) obsługa multimediów ma zostać przepisana od nowa, co pozwoliłoby na znaczne przyspieszenie działania algorytmów i programów do rozpoznawania emocji.

W celu wizualizacji wyników zastosowano standardową bibliotekę do tworzenia graficznych interfejsów użytkownika - *Swing*. Dzięki wbudowanym w środowiska takie jak NetBeans tworzenie



Rysunek 40: Przebieg zmian emocji w czasie

aplikacji z GUI jest stosunkowo proste i pozwala skupić się na poprawności działania zastosowanych algorytmów.

Aby zaprezentować zmiany procesowanych danych w czasie zastosowano bibliotekę *JFreeChart* [20], która w przystępny sposób pozwala na tworzenie zaawansowanych wykresów różnego typu.

Zastosowane algorytmy wymagały wielu operacji na macierzach pikseli. Do operacji, między innymi transpozycji, skalowania czy mnożenia, zastosowano bibliotekę *Java Matrix Package* [18].

W celu implementacji algorytmu Lucasa Kanade, autor przepisał kod źródłowy z matlaba autorstwa Sohaiba Khana [21], oraz posłużył się kodem Paula Breeuwsma'y [7] w celu uzyskania interpolacji bikubicznej (*bicubic interpolation*), niezbędnej do powrotu z reprezentacji piramidalnej obrazu do oryginalnych rozmiarów obrazu.

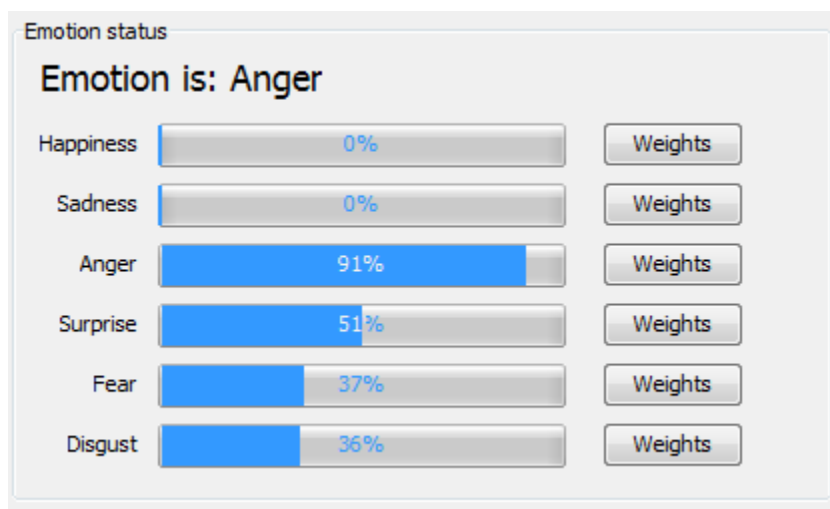
#### 4.1.2 Stworzone klasy

Aplikacja ma budowę modułową, która pozwala na jego łatwą rozbudowę, jak również na wykorzystanie napisanych w ramach tego projektu klas, w innych programach. Najważniejsze napisane klasy to:

**Emolution** - klasa nadrzędna uruchamiająca cały proces.

**FaceDetector** - klasa odpowiedzialna za detekcje twarzy na klatce filmu.

**EyeDetector** - klasa poszukująca na przestrzeni twarzy oczu.



Rysunek 41: Wizualizacja aktualnego stanu emocjonalnego

**EyeBrowDetector** - klasa, która wyszukuje brwi.

**LipsDetector** - klasa odpowiedzialna za detekcję ust.

**VideoPlayer** klasa, która procesuje strumień wideo przy użyciu biblioteki JMF, a następnie przekazuje poszczególne klatki do analizy do klasy EmotionDetector.

**EmotionDetector** - klasa, która na podstawie wykrytych cech twarzy zapisanych w grafie twarzy dokonuje identyfikacji emocji.

**emoGUI** - klasa odpowiedzialna za wizualizację danych przy użyciu biblioteki Swing.

**ImageFilter** - klasa pomocnicza, realizująca podstawowe operacje na obrazach, takie jak dylatacja, erozja, filtracja, normalizacja tła.

**LabelRegion** - klasa implementująca algorytm Connected Component Labeling.

**LucasKanade** - klasa implementująca algorytm śledzenia przepływu optycznego Lucasa - Kanade.

**GraphFace** - klasa implementująca reprezentację grafu twarzy.

Schemat działania programu prezentuje się następująco. Wybrany plik wideo przekazywany jest przez klasę Emolusion do klasy VideoPlayer, która stara się otworzyć i przeprocować dany plik wideo. Jeśli się to powiedzie, tworzona jest nowa instancja klasy EmotionDetector, do której przekazywane są poszczególne klatki pliku wideo. Dana klatka na początku jest analizowana przez klasę FaceDetector w poszukiwaniu twarzy. Fragment klatki obrazu na której wykryto twarz, przekazywany jest do dalszej analizy w celu poszukiwania brwi (EyeBrowsDetector), oczu (EyeDetector) oraz ust (LipsDetector). Odnalezione 14 charakterystycznych punktów na twarzy służą do aktualizacji stanu grafu twarzy (GraphFace). Następnie graf ten analizowany jest przez klasę EmotionDetector, która identyfikuje i parametryzuje emocje. Wyniki działania algorytmów wyświetlane są na interfejsie użytkownika, będącym instancją klasy emoGUI.

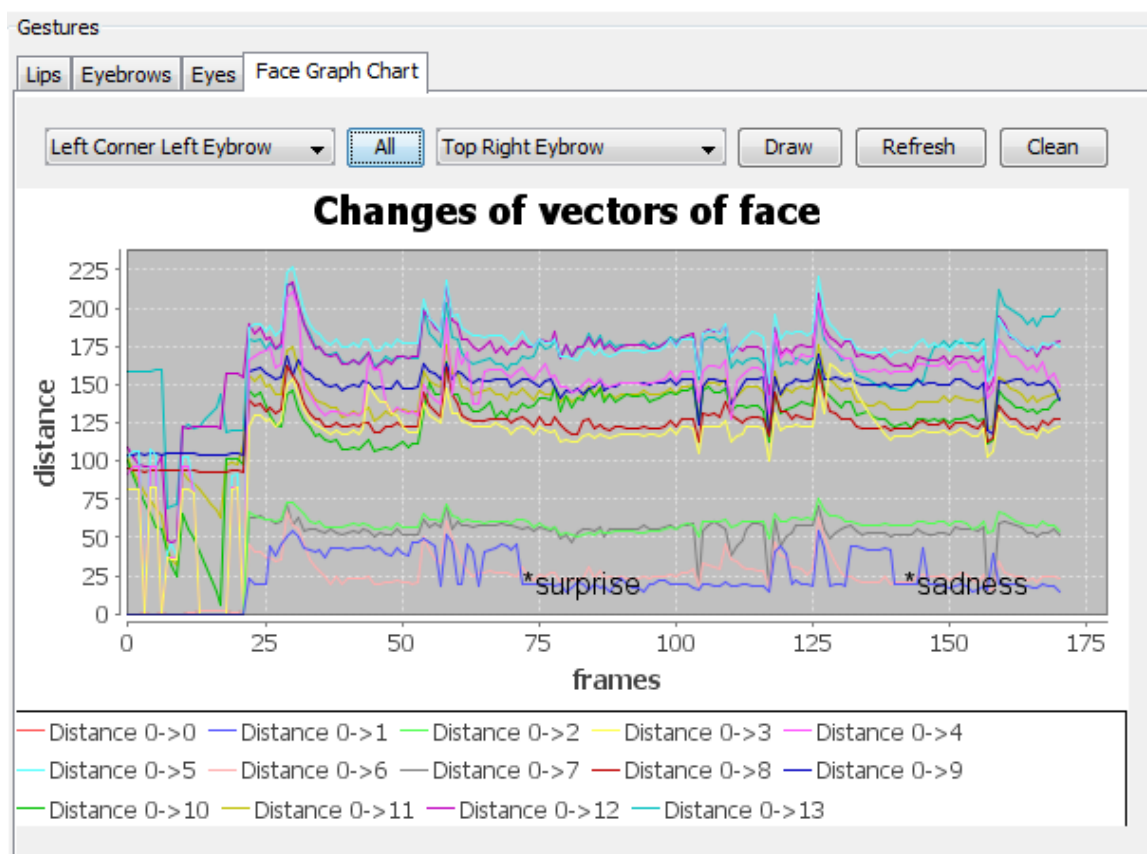
## 4.2 Funkcje programu

Po uruchomieniu aplikacji pojawia się okno pokazane na Rysunku 38. Aplikacja została podzielona na 5 głównych bloków funkcjonalnych:

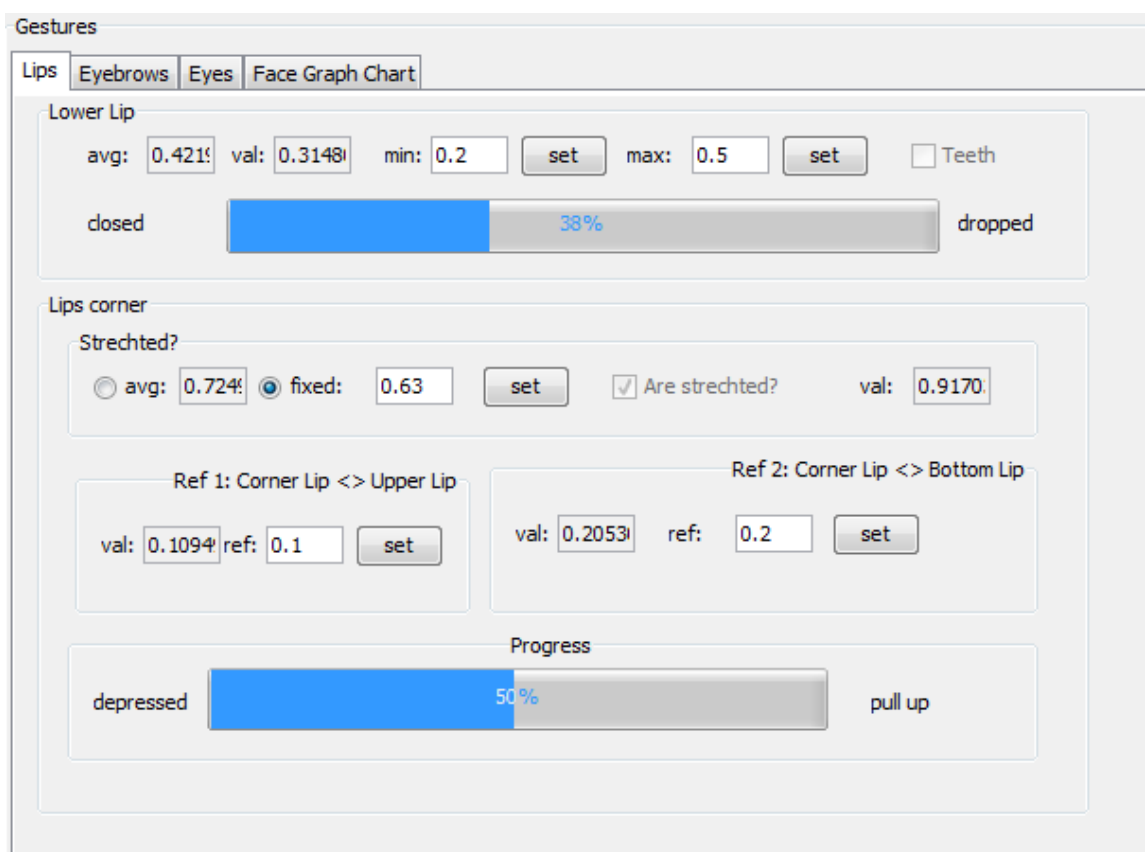
### 4.2.1 Menu programu

W zakładce **Source** mamy możliwość wybrania źródła strumienia wideo w postaci pliku wideo w formacie `.avi` lub `.wmv`. Ze względu na ograniczenia biblioteki *JMF*, nie wszystkie pliki w formacie `.avi` są poprawnie odgrywane, ponieważ zostały skompresowane różnymi kodekami. Odwołując się do doświadczenia autora można stwierdzić, że poprawnie (i jednocześnie w miarę szybko) odgrywane są filmy skompresowane kodekiem *XVID*.

Jest również możliwość wyboru, czy w części, gdzie wyświetlany jest podgląd wideo ma być prezentowana cała klatka filmu, czy też jedynie odnaleziony region twarzy, co jest zachowaniem domyślnym (w celu przyspieszenia pracy). Ostatnim elementem tej zakładki jest możliwość zakończenia pracy programu.



Rysunek 42: Zmiany wartości na poszczególnych krawędziach grafu twarzy



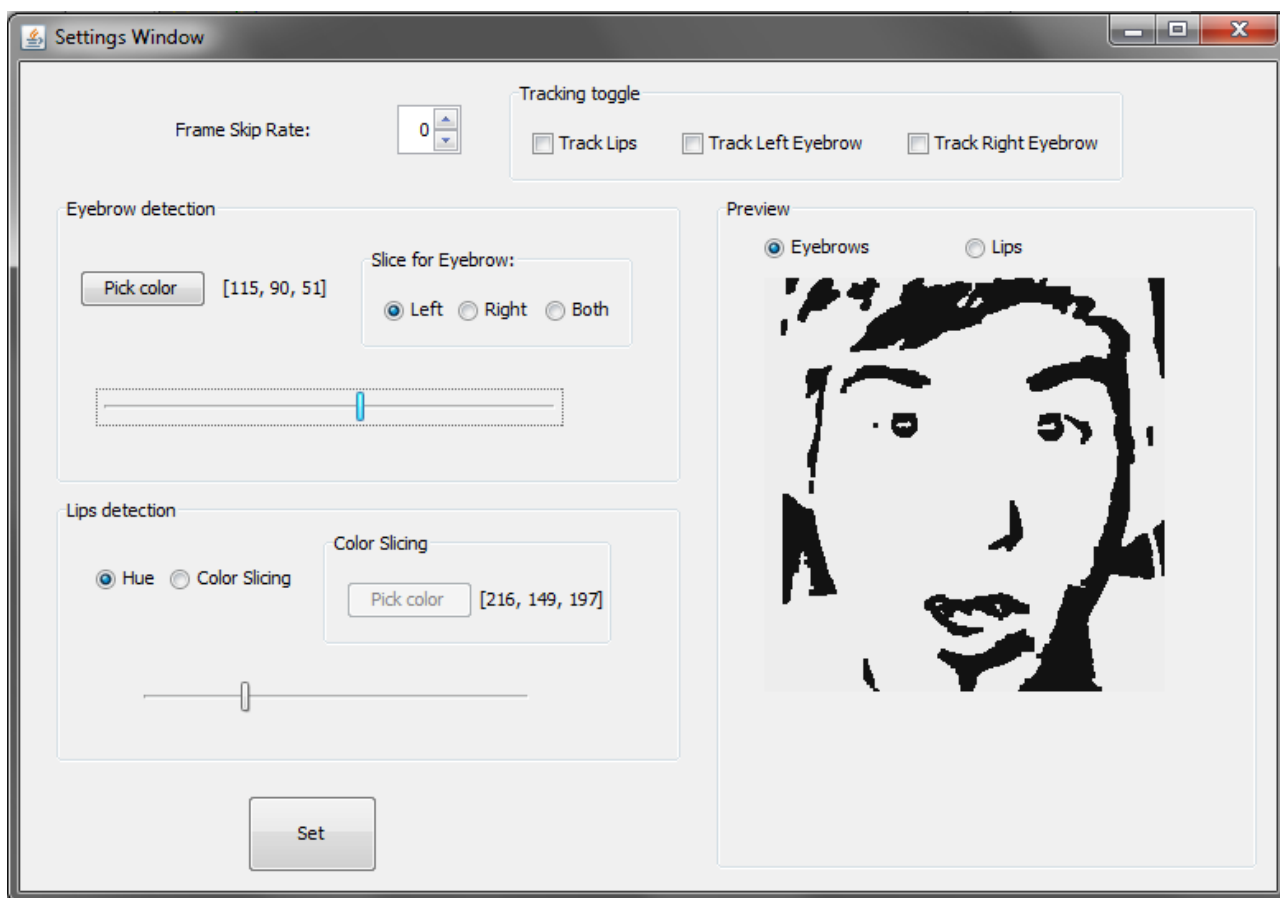
Rysunek 43: Wartości oraz parametry poszczególnych gestów

W kolejnej zakładce o nazwie **Preview**, dostępne są 4 przyciski wywołujące podgląd działania algorytmów detekcji odpowiednio:

- twarzy,
- oczu,
- brwi,
- ust.

Przykładowe okienko z podglądem działania algorytmu detekcji oczu znajduje się na Rysunku 39.

W ostatniej zakładce **Settings** dostępne są ustawienia służące do sterowania algorytmami detekcji poszczególnych części twarzy. Mamy tutaj możliwość zmiany liczby ramek, które nie są przetwarzane (domyślnie **Frame Skip Rate** = 0 co oznacza, że analizowane są wszystkie klatki), ustawienia koloru bazowego oraz szerokości tolerancji dla algorytmu **Color Slicing** dla wykrywania każdej z brwi, oraz wybór między algorytmem detekcji ust na podstawie binaryzacji składowej nasycenia lub algorytmu **Color Slicing**. Jest również dostępne okno, w którym można ręcznie wprowadzić wszystkie wartości cech, oraz okno z informacjami na temat autora i tematu pracy.



Rysunek 44: Ustawienia aplikacji, z możliwością kalibracji algorytmów detekcji

#### 4.2.2 Obraz wideo

W tej części dostępny jest podgląd aktualnego obrazu wideo, przycisk *Pause*, umożliwiający wstrzymanie działania algorytmów w celu analizy aktualnego stanu, licznik pokazujący numer aktualnej klatki filmu, pasek postępu pokazujący ile klatek zostało już przetworzone oraz możliwość zaetykietowania konkretnej ramki, w której wydarzyło się coś szczególnego (np zaczęła się jakaś emocja, bądź algorytm źle wykrył jakąś cechę). Etykieta ta jest zapisana w systemie i następnie prezentowana na wykresach obrazujących zmiany na grafie twarzy jak i zmiany emocji w czasie.

#### 4.2.3 Gesty

Dostępne są tutaj 3 zakładki odpowiedzialne za wizualizacje parametrów i wyników algorytmów detekcji gestów: brwi, oczu i ust. Dla każdego gestu mamy podgląd aktualnej wartości parametru, jego średnią wartość, możliwość ustawienia jego wartości progowej oraz *progressbar* informujący o aktualnej wartości gestu.

Ostatnia 4 zakładka (Rysunek 42) daje możliwość rysowania wykresów poszczególnych krawędzi grafu twarzy. Z pierwszego menu rozwijanego wybiera się pierwszy wierzchołek, z kolejnego menu drugi wierzchołek, po wciśnięciu przycisku **Draw**, rysowany jest przebieg zmian w czasie wartości krawędzi pomiędzy wybranymi wierzchołkami. Za pomocą przycisku **All** możliwe jest wyrysowanie historii wszystkich krawędzi połączonych z wybranym wierzchołkiem. Przycisk **Clean** czyści pole wykresu. W przypadku wprowadzenia etykiet klatek, w części związanej z obrazem, etykiety te wyświetlają się na wykresie w miejscu odpowiadającym na osi czasu danej klatce.

#### 4.2.4 Emocje

Zilustrowane jest tutaj natężenie wszystkich stanów emocjonalnych w postaci procentowych pasków postępu. Dla każdej emocji możliwa jest zmiana wag ważności poszczególnych gestów po naciśnięciu przycisku **Weights**. U góry znajduje się napis informujący o aktualnie wiodącej emocji. Jest ona wyliczana jako najsilniejsza emocja spośród wszystkich, jeśli wartość natężenia danej emocji przekroczyła 50%. W przeciwnym razie ustawiany jest *neutralny* stan emocjonalny.

Na samym dole znajduje się małe pole tekstowe, pełniące rolę terminala z aktualnymi informacjami technicznymi generowanymi przez algorytm, służącymi do weryfikacji poprawności działania aplikacji.

#### 4.2.5 Wykres zmian emocji w czasie

Ostatnim elementem aplikacji jest wykres prezentujący zmiany wszystkich stanów emocjonalnych w czasie. Jeśli zapamiętane były jakieś etykiety to wpisywane są na tym wykresie w miejscu odpowiadającym wprowadzeniu danej etykiety.

## 5 Wyniki

Przeprowadzono testy przy użyciu zestawu 16 filmów, gdzie każdy zestaw zawierał sekwencje poszczególnych emocji. Filmy nagrane były w różnych pomieszczeniach, w różnych warunkach oświetleniowych (światło naturalne i sztuczne), na próbie 8 osób. Nagrano 2 kobiety oraz 6 mężczyzn. Jedna osoba była pochodzenia azjatyckiego (pozostali uczestnicy byli rasy kaukaskiej) oraz jedna osoba była starsza od pozostałych. Osoby, które noszą okulary, nagrywały próbki zarówno w okularach jak i bez nich.

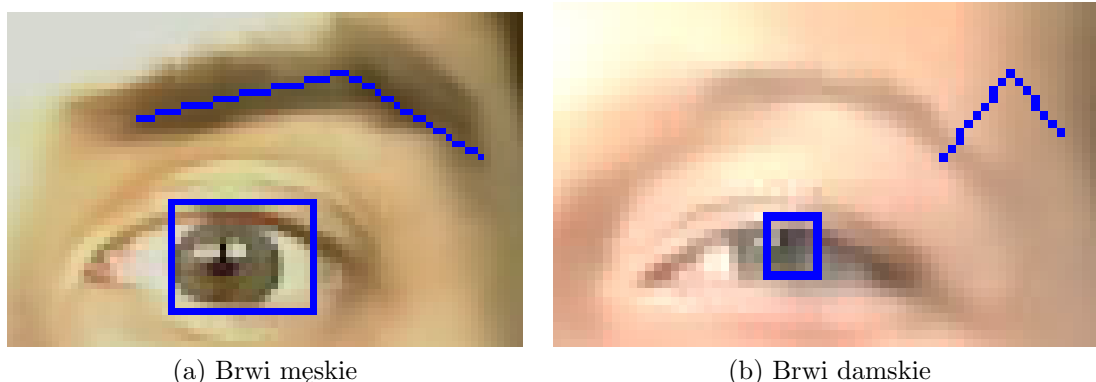


Rysunek 45: Próbkę osób

### 5.1 Detekcja gestów

#### 5.1.1 Detekcja brwi

Spore problemy w detekcji brwi, aplikacja wykazała w przypadku użytkowników płci żeńskiej. Kobiety mają zazwyczaj brwi dużo cieńsze niż mężczyźni, jedynie ledwo zarysowane, co w przypadku przyjętej strategii detekcji na podstawie filtra wykrywającego krawędzie i duże zmiany w kolorach, daje słabe rezultaty.



(a) Brwi męskie

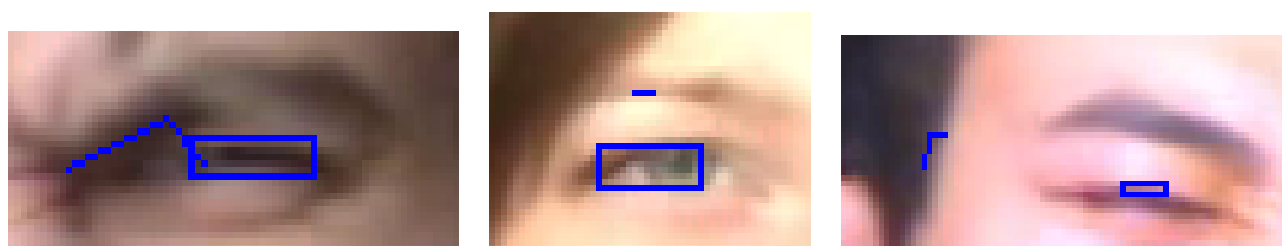
(b) Brwi damskie

Rysunek 46: Różnice w grubości brwi między brwiami męskimi a żeńskimi





Rysunek 47: Przykłady poprawnie rozpoznanych brwi



(a) Ciemne oświetlenie

(b) Cienkie brwi

(c) Skręt głowy

Rysunek 48: Przykłady niedokładnie rozpoznanych brwi

Gest	Odraza	Złość	Średnio
<b>Obniżenie brwi</b>	86,8%	87,4%	<b>87,1%</b>

Tabela 7: Wyniki dla gestu obniżenia brwi

Gest	Zaskoczenie	Strach	Szczęści	Średnio
<b>Uniesienie brwi</b>	98,7%	91,6%	74,5%	<b>88,7%</b>

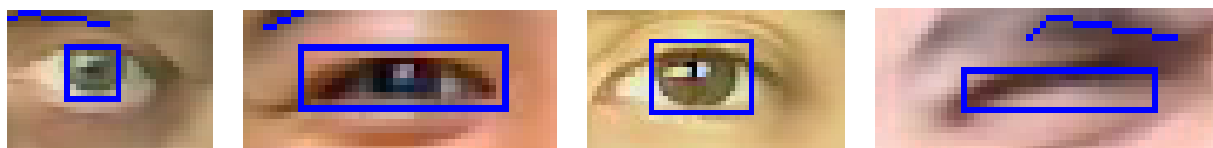
Tabela 8: Wyniki dla gestu uniesienia brwi

Gest	Odraza	Złość	Średnio
<b>Ściśnięcie brwi</b>	53,8%	77,7%	<b>63,6%</b>

Tabela 9: Wyniki dla gestu ściśnięcia brwi

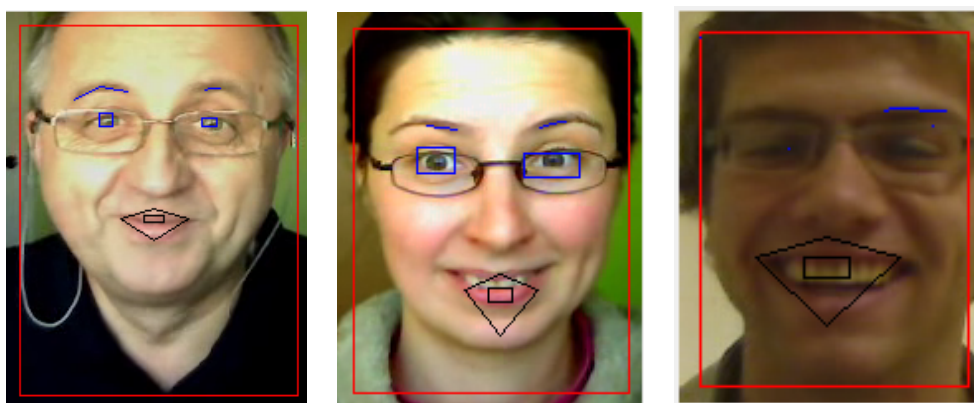
### 5.1.2 Detekcja oczu

Zaproponowany algorytm detekcji oczu, okazał się być bardzo skutecznym i jednocześnie mało wymagającym obliczeniowo. Pewną niedogodnością jest fakt, iż dla różnych osób różne obszary oka uznawane są za wykryty region, co wynika z samego algorytmu, który wyszukuje obszary o dużej zmienności jasności pikseli. Dla osób z ciemną tęczówką oznaczany jest obszar całej tęczówki, natomiast dla osób o jasnej tęczówce, wykrywany jest region jedynie źrenicy. Dla oczu przymkniętych, bądź też dla osób pochodzenia azjatyckiego (z tak zwaną fałdą mongolską) oznaczany jest obszar całej gałki ocznej. W przypadku niedostatecznego oświetlenia, z powodu występujących cieni, za region oczu może zostać uznany cały oczodół.



Rysunek 49: Przykłady rezultatów detekcji oczu

Testy wykazały dużą odporność algorytmu na występowanie artefaktów na twarzy w postaci okularów, co jest sporym problemem w większości algorytmów wykrywania oczu, gdyż okulary powodują odbicia światła, zaburzające naturalne kształty oka [16].



Rysunek 50: Przykłady rezultatów detekcji oczu dla osób noszących okulary

Gest	Zaskoczenie	Strach	Średnio
<b>Otwarcie oczu</b>	85,2%	66,5%	<b>77,2%</b>

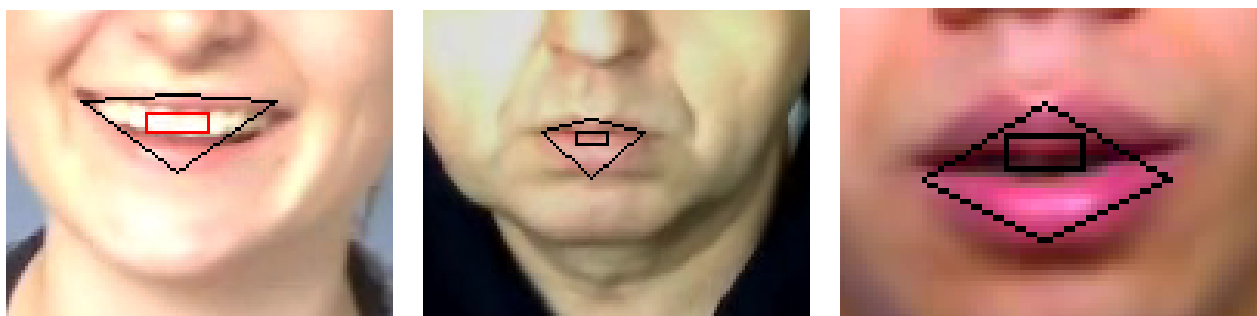
Tabela 10: Wyniki dla gestu otwarcia oczu

Gest	Szczęście	Złość	Średnio
<b>Przymrużenie oczu</b>	57%	53,1%	<b>54,8%</b>

Tabela 11: Wyniki dla gestu przymrużenia oczu

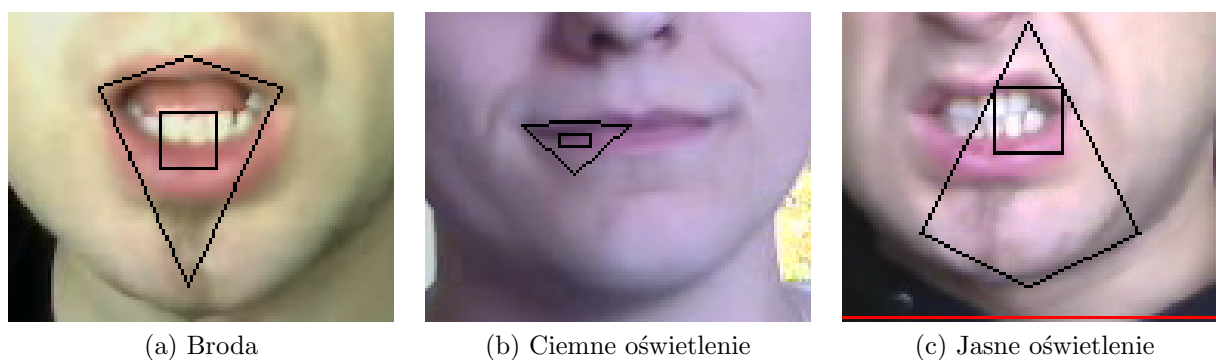
### 5.1.3 Detekcja ust

Problemy z detekcją ust wynikały głównie z 2 czynników: oświetlenia oraz zarostu. Jak było już wspomniane wcześniej, rodzaj i jakość światła padającego na twarz, może zmienić w znaczący sposób rzeczywisty kolor. Dla obrazów z ciemnym oświetleniem usta mogły mieć fioletowy kolor na obrazie wideo, w jasnym były natomiast bardzo blade, prawie nie odróżnialne od koloru skóry. Drugim problemem jest występowanie na twarzy przesłoneń na przykład w postaci brody,



Rysunek 51: Przykłady poprawnie rozpoznanych ust

gdzie, tak jak w przypadku autora, często była ona uznawana za część ust. Problematiczne mogą okazać się również zbyt mocne szminki jakie stosują kobiety.



(a) Broda

(b) Ciemne oświetlenie

(c) Jasne oświetlenie

Rysunek 52: Przykłady błędów w detekcji ust

Gest	Zaskoczenie	Złość	Strach	Średnio
<b>Otwarcie ust</b>	91,6%	62,5%	56,5%	<b>71%</b>

Tabela 12: Wyniki dla gestu otwarcia ust

Gest	Szczęście	Średnio
<b>Uniesienie kącików ust</b>	78,5%	<b>78,5%</b>

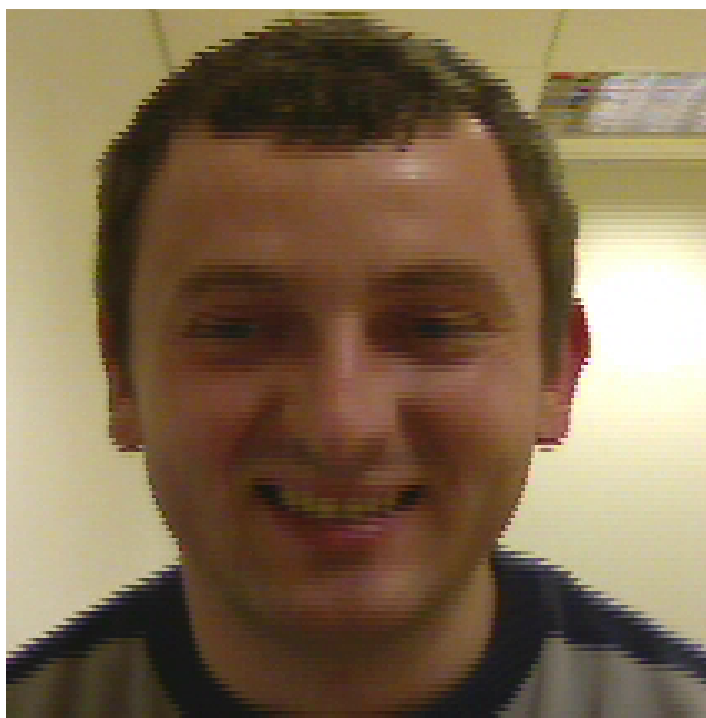
Tabela 13: Wyniki dla gestu uniesienia kącików ust

Gest	Odraza	Złość	Strach	Smutek	Średnio
<b>Obniżenie kącików ust</b>	53,5%	49,3%	50%	51,1%	<b>51,3%</b>

Tabela 14: Wyniki dla gestu obniżenia kącików ust

#### 5.1.4 Błędy wynikające z oświetlenia

Największą przeszkodą napotkana w czasie prac, było zdecydowanie oświetlenie. Dużo lepsze wyniki osiągnięto przy świetle naturalnym, nie mogło być ono zbyt mocne, co groziłoby prześwietleniem obrazu i kompletnym zachwianiem balansu kolorystycznego na twarzy. Oświetlenie sztuczne zazwyczaj było punktowe, co powodowało nadmierne naświetlenie części twarzy i cienie na reszcie ciała (przykład różnicy w kolorach dla oświetlenia bocznego na Rysunku 25). Bardzo złe wyniki dają nagrania przy oświetleniu jarzeniowym (popularnych świetlówkach), gdyż światło to rejestrowane na kamerze w postaci jaśniejszych i ciemniejszych fal oraz *postrzępionego* obrazu po kompresji (przykład na Rysunku 53).

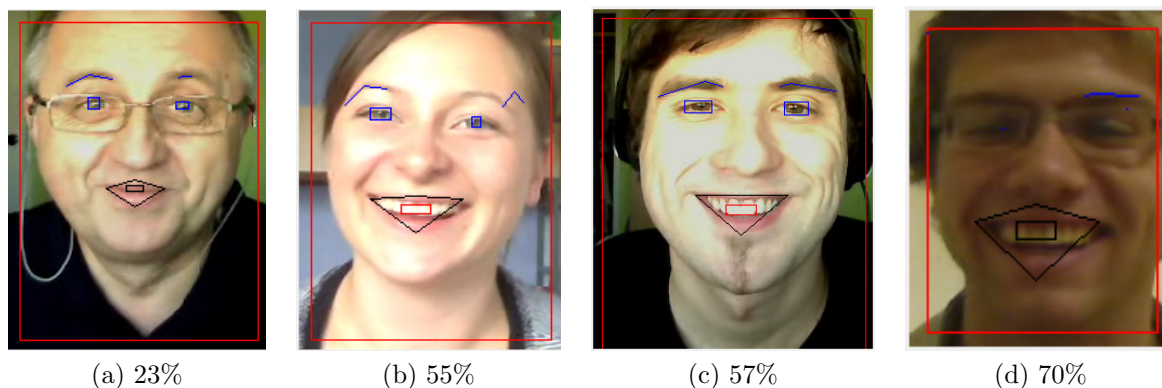
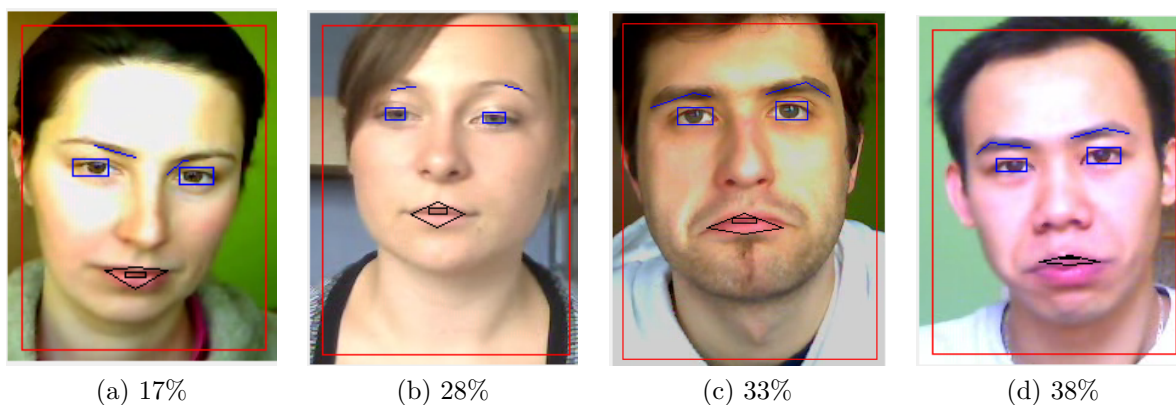
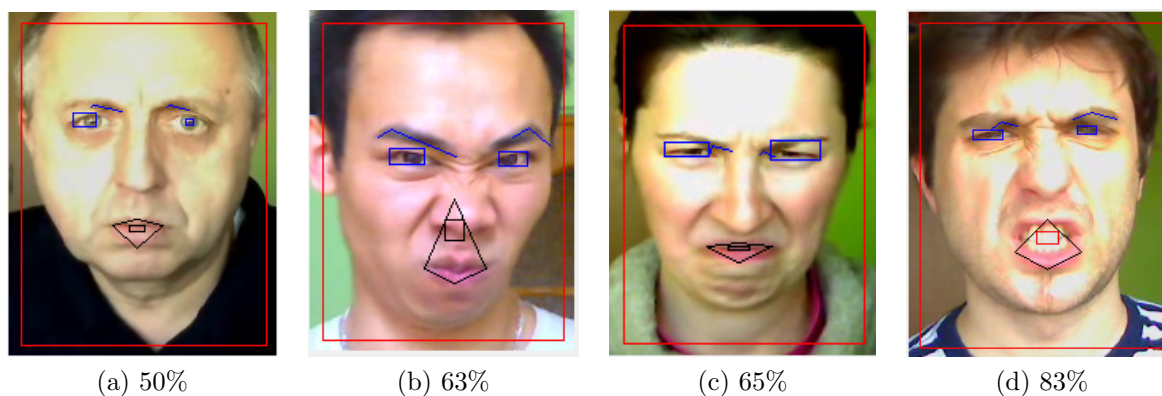


Rysunek 53: Nagranie z kamery przy oświetleniu jarzeniowym

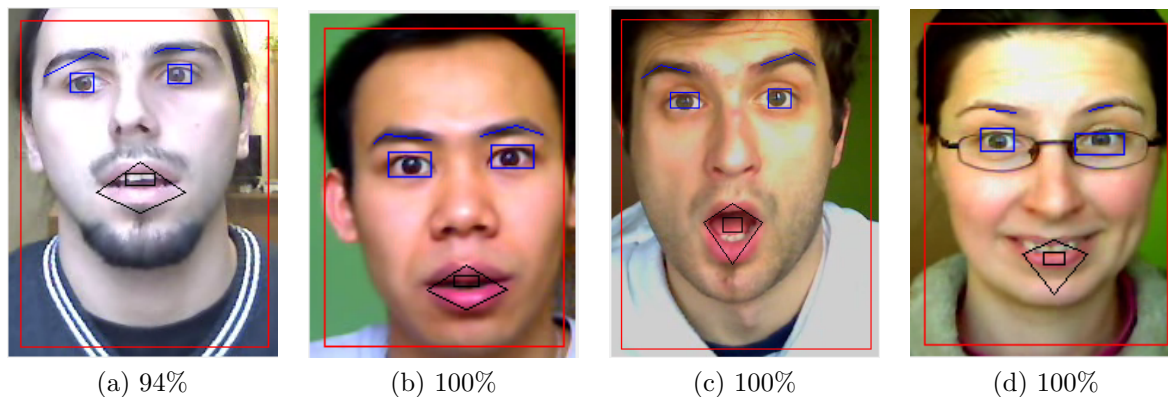
## 5.2 Klasyfikacja emocji

Na podstawie przeprowadzonych testów sporządzano raport poprawności identyfikacji emocji zebrany w Tabeli 15.

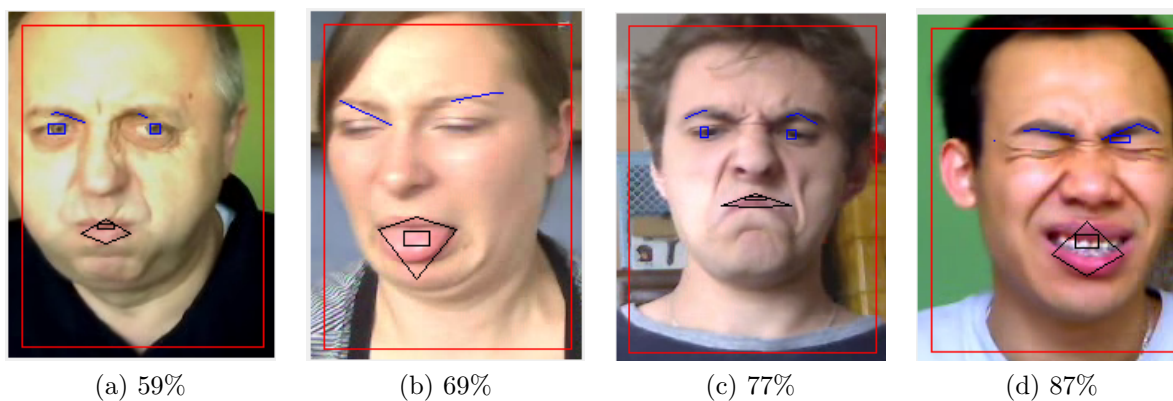
Najlepiej ze wszystkich emocji odwzorowywane jest zaskoczenie, które dla filmów prezentujących tę emocję miały średnio 98% natężenia. Wynika to zapewne z intensywności tej emocji, która objawia się na większości twarzy w podobny sposób. Na drugim biegunie znajduje się smutek. Uzyskał on jedynie średnio około 22% siły emocji. Może mieć to wytłumaczenie w tym, iż smutek zazwyczaj przeżywamy bezobjawowo na twarzy, bardziej wewnętrznie, *duchowo*, co z wiadomych przyczyn jest niemierzalne przez komputery.

Rysunek 54: Przykłady wyników dla emocji **Szczęścia**Rysunek 55: Przykłady wyników dla emocji **Smutku**Rysunek 56: Przykłady wyników dla emocji **Złości**

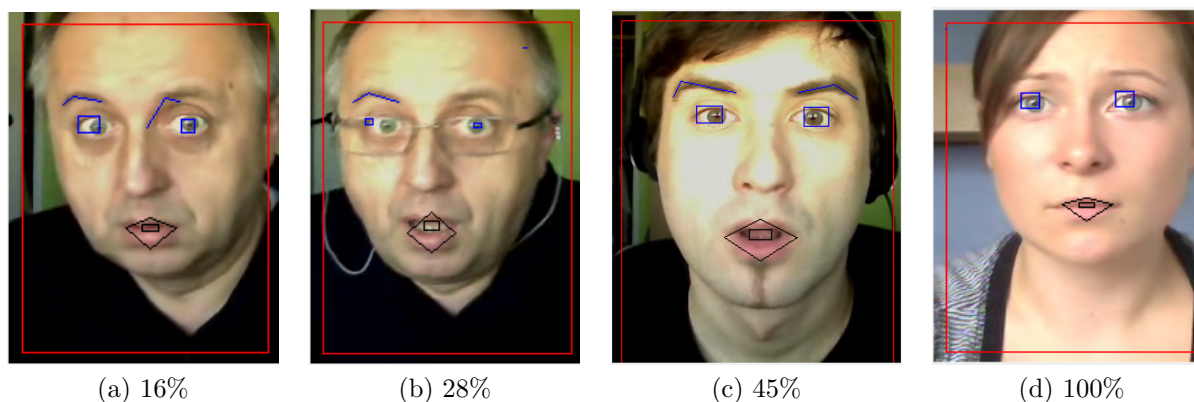
Testy wykazały duże skorelowanie ze sobą niektórych emocji. Na filmach prezentujących złość emocja ta miała średnią 61%, ale tuż za nią identyfikowana była odraza (54%). Jeszcze bardziej zbieżność ta była widoczna na sekwencjach prezentujących odrazę - tutaj obie emocje były



Rysunek 57: Przykłady wyników dla emocji **Zaskoczenia**



Rysunek 58: Przykłady wyników dla emocji **Odraza**



Rysunek 59: Przykłady wyników dla emocji **Strachu**

identyfikowane prawie równomiernie (odraza 48,3%, złość 48%). Mogłoby to wskazywać na podobne pochodzenie obu emocji.

Emocja rzeczywista	Emocje wykryte					
	Szczęście	Smutek	Złość	Zaskoczenie	Strach	Odraza
<b>Szczęście</b>	<b>39,7%</b>	3%	10,85%	<b>34,14%</b>	24%	0,71%
<b>Smutek</b>	3,3%	<b>21,7%</b>	6,7%	17,7%	16,5%	10,2%
<b>Złość</b>	2,44%	5,33%	<b>61,33%</b>	23,77%	23,77%	<b>54,77%</b>
<b>Zaskoczenie</b>	8,62%	0%	2,87%	<b>98%</b>	<b>90,12%</b>	0,75%
<b>Strach</b>	24,83%	14,16%	0%	<b>62,5%</b>	<b>54,5%</b>	0%
<b>Odraza</b>	2,23%	12%	<b>48,07%</b>	5,23%	5%	<b>48,38%</b>

Tabela 15: Podsumowanie wyników klasyfikacji emocji

Podobne skorelowanie widać na przykładzie pary strach - zaskoczenie. Dla filmów prezentujących strach, mocniej wykrywane było zaskoczenie (62%) niż sam strach (54%). Wynik taki może mieć dwójakie uzasadnienie. Po pierwsze, trudno jest wywołać, na potrzeby badań naukowych, prawdziwy strach, jaki wywołują traumatyczne przeżycia (takie jak zagrożenie zdrowia, życia czy spokoju psychicznego). Po drugie, obie te emocje w początkowej fazie są ze sobą powiązane. Z punktu widzenia psychologicznego, boimy się tego co nieznane. Można więc przyjąć, że najpierw okazywane jest zaskoczenie, a dopiero potem, kiedy nasz umysł zrozumie, iż jest w niebezpieczeństwie emocja przybiera formę strachu.

Powodem ograniczonych wyników rozpoznawania może być również fakt, iż emocje były wywoływane sztucznie a nie rzeczywiście przeżywane, gdyż trudno byłoby otrzymać materiał z faktycznie przeżywanym na przykład strachem. Otrzymane wyniki mogą się więc częściowo różnić od ich autentycznych odpowiedników.

### 5.2.1 Porównanie wyników

Otrzymane wyniki porównano z pracą dyplomową Tomasza Kotlińskiego [22], pisaną również pod opieką dr Adriana Horzyka. Zastosowano w niej sieci neuronowe do rozpoznawania emocji. Jakość rozpoznawania poszczególnych emocji w pracy [22] oraz niniejszej zebrana jest w Tabeli 16.

Praca	Szczęście	Smutek	Zaskoczenie	Złość	Strach	Odraza
<b>Tomasz Kotliński</b>	91,7%	25%	61,7%	72%	66,7%	91,7%
<b>Maciej Dudek</b>	39,7%	21,7%	98%	61,3%	54%	43%

Tabela 16: Porównanie jakości rozpoznawania emocji

Wyniki pokazują, w większości przypadków, stosunkowo lepsze wyniki pracy Tomasza Kotlińskiego. Może to sugerować poprawność zastosowania sieci neuronowych do badań nad rozpoznawaniem emocji na ludzkiej twarzy. Należy jednak pamiętać iż wyniki w pracy [22] otrzymane były na mniejszej i mniej reprezentatywnej próbie osób. Niezależnie od przyjętej metodologii, obie prace pokazują wysoką skuteczność w detekcji "silnych" emocji "pierwotnych" takich jak zaskoczenie, strach czy niesmak oraz problemy w rozpoznaniu smutku.

## 6 Podsumowanie

“Smutek był formą, a szczęście treścią.  
Szczęście wypełniało przestrzeń smutku.”

Milan Kundera

*Niežnośna lekkość bytu*

Niniejsza praca poświęcona została badaniu kwestii automatycznego rozpoznawania i opisywania emocji na podstawie zmian mimiki na ludzkiej twarzy.

Zagadnienie poruszane w tej pracy łączy ze sobą psychologię i technikę. Zanim więc przystąpiono do prac związanych z badaniem konkretnych algorytmów, sporo czasu i uwagi poświęcono na zbadanie zagadnień związanych z psychologią i naturą ludzkich emocji oraz mowy ciała. Bardzo przydatne okazały się wieloletnie badania prowadzone w związku z tym tematem przez amerykańskiego psychologa Paula Ekmana, który zajmował się dokładnym sprawdzaniem i opisaniem zmian anatomicznych i psychologicznych w czasie trwania emocji.

Jako że temat łączenia emocji z komputerami, jak i całej dziedziny *Human-Computer Interaction* jest stosunkowo nowym działem, który wśród wielu osób budzi wątpliwości na temat jego powodzenia, na początku tej pracy przedstawiłem dotychczasowe osiągnięcia nauki w tych zagadnieniach. Istniejące już rozwiązania pozwalają na rozpoznawanie emocji przez maszyny i odpowiednie reakcje na stany emocjonalne człowieka. Wraz z rozwojem robotyki temat rozpoznawania emocji ludzkich będzie intensywnie badany i będą powstawać coraz dokładniejsze systemy pozwalające na wykrywanie uczuć człowieka.

W dalszej części pracy opisano metody rozpoznawania ludzkich emocji na podstawie zmian mimicznych przy użyciu zestawu gestów opisanych przez Paula Ekmana jako *Action Units*. Poszczególne gesty wykrywane były na podstawie analizy zmian skonstruowanego na potrzeby niniejszej pracy Grafu Twarzy, zawierającego jako wierzchołki 14 kluczowych punktów na twarzy, a wartościami krawędzi były odległości pomiędzy poszczególnymi wierzchołkami grafu.

W celu konstrukcji takiego grafu niezbędne było odnalezienie dokładnych i jednocześnie odpowiednio szybkich algorytmów detekcji położenia twarzy, brwi, oczu i ust. Algorytmy oparte na wykrywaniu regionów o podobnych kolorach, zastosowanie detekcji brwi (Color Slicing) i ust (analiza składowej nasycenia przestrzeni barw HSV), okazały się być dość mocno podatne na zmiany oświetlenia, które powodowały że te same regiony mogły mieć w różnych chwilach kolory o bardzo odmiennych wartościach. Wprowadzono dlatego możliwość ręcznej kalibracji tych algorytmów, tak aby dostosować je do aktualnych warunków oświetleniowych. Z kolei algorytm wykrywania oczu, po przeprowadzonych testach okazał się być bardzo skutecznym, również w



przypadku osób noszących okulary. Nie wymagał on przy tym pracochłonnych obliczeń, jak większość rozwiązań proponowana w literaturze.

Zaproponowany system rozpoznawania emocji dał zadowalające rezultaty. Średnio rozpoznawane emocje miały natężenie 53%, przy czym, znaczące różnice w wynikach pojawiły się pomiędzy poszczególnymi emocjami. Emocja zaskoczenia była prawie zawsze (98%) oznaczana z maksymalną intensywnością, natomiast emocja smutku osiągała zaledwie 21% intensywności. Różnice te wynikają ze sposobu przeżywania różnych emocji, jak również ze skuteczności wykrywania gestów.

Mimo zapewnień doktora Ekmana, sposób wyrażania emocji w poszczególnych kręgach kulturowych może znacząco różnić się od siebie. Jak pisał znany polski psychiatra, doktor Antoni Kępiński w [23]: *Twarz nie kłamie*; łatwiej jest pokierować reakcjami ruchowymi innego typu, np. ruchami rąk, całego ciała itp. reakcjami werbalnymi niż mimiką twarzy. Mimo bardzo rozległej reprezentacji korowej w porównaniu z innymi grupami mięśniowymi mięśnie mimiczne twarzy wymykają się łatwo spod kontroli wolicjonalnej. Stopień ujarzmienia twarzy przez wolę zależy od osobniczej historii jednostki i od tradycji kulturowej. W niektórych kręgach jest on większy niż w innych. Inaczej wyraża radość czy smutek Europejczyk, a inaczej Chińczyk, inaczej Anglik niż Włoch, inaczej w “wyższych sferach”, a inaczej u chłopów czy robotników.



(a) Tsunami w Japonii. Źródło: The Telegraph



(b) Wybuch bomby w Bagdadzie. Źródło: Guardian

Rysunek 60: Reakcje rozpaczki w różnych kręgach kulturowych w obliczu tragedii

Możliwe kierunki rozwoju systemu mogą koncentrować się na poprawie działania algorytmów przy zmiennych warunkach oświetleniowych, usprawnienie bądź stworzenie bardziej dokładnych algorytmów detekcji poszczególnych cech (w szczególności ust, które są głównym źródłem informacji o stanie emocjonalnym człowieka). Można również użyć bardziej skomplikowanych algorytmów z teorii grafów w celu lepszej analizy stanu grafu i identyfikacji emocji na jej podstawie.

Po rozbudowie i testach przedstawione w pracy rozwiązanie mogłoby zostać zastosowane jako element w systemach bezpieczeństwa np na lotniskach do identyfikacji osobników zachowujących się podejrzanie, lub też, w bardziej komercyjnych celach np w tak zwanych awatarach, czyli wirtualnych biurach obsługi, które analizując stan emocji petenta mogłyby lepiej reagować na jego potrzeby.

## Spis rysunków

1	Mięśnie twarzy[2]	10
2	Koło emocji wg Plutchika. Autor Ivan Akira	11
3	Wizualizacja radości	13
4	Wizualizacja smutku	13
5	Wizualizacja złości	14
6	Wizualizacja niesmaku	16
7	Wizualizacja strachu	17
8	Wizualizacja zaskoczenia	17
9	Procentowy rozkład przekazu ludzkiego	18
10	MEXI (Machine with Emotionally eXtended Intelligence). Źródło: [31]	20
11	Zestaw EPOC z programem emoRate	21
12	Program Candide. Źródło: [28]	22
13	Zestaw faceLAB. Źródło: [26]	23
14	Schemat działania algorytmu	24
15	Sąsiedztwa pikseli w algorytmie CCL. Źródło: [40]	25
16	Przykład działania algorytmu CCL. Źródło: [40]	26
17	Relacja między przestrzeniami barw RGB i HSV. Źródło:[15]	27
18	Operacje morfologiczne	28
19	Przykłady rezultatów działania filtrów obrazów. Źródło: [34]	30
20	Histogram obrazu	32
21	Histogramy wertykalny oraz horyzontalny	33
22	Rezultat działania algorytmu Color Slicing. Źródło: [15]	33
23	Działanie algorytmu detekcji twarzy	35
24	Działanie algorytmu detekcji brwi przy użyciu filtru Sharra	36
25	Różnice w kolorach dla twarzy z oświetleniem bocznym	37
26	Punkty opisujące każdą brew	37
27	Działanie algorytmu detekcji brwi przy użyciu algorytmu Color Slicing dla sześciangu o środku w (95,90,112) i szerokości $W=35$	38
28	Działanie algorytmu detekcji oczu	39
29	Punkty opisujące każde oko	39
30	Wstępne wykrywanie obszaru ust	40
31	Przykłady działania algorytmu Color Slicing na obszarze ust	41
32	Przykłady prześwietlenia na obszarze ust	41
33	Punkty opisujące usta	42
34	Działanie algorytmu detekcji ust	42
35	Detekcja odsłonięcia zębów (kolor czerwony oznacza pojawienie się zębów)	43
36	Schemat grafu twarzy użyty w niniejszej pracy	45
37	Wykrycie ściśnięcia brwi	46
38	Zrzut ekranu aplikacji "Emolution"	48
39	Podgląd działania algorytmu detekcji oczu	49
40	Przebieg zmian emocji w czasie	50
41	Wizualizacja aktualnego stanu emocjonalnego	51

42	Zmiany wartości na poszczególnych krawędziach grafu twarzy . . . . .	52
43	Wartości oraz parametry poszczególnych gestów . . . . .	53
44	Ustawienia aplikacji, z możliwością kalibracji algorytmów detekcji . . . . .	54
45	Próbka osób . . . . .	56
46	Różnice w grubości brwi między brwiami męskimi a żeńskimi . . . . .	56
47	Przykłady poprawnie rozpoznanych brwi . . . . .	57
48	Przykłady niedokładnie rozpoznanych brwi . . . . .	57
49	Przykłady rezultatów detekcji oczu . . . . .	58
50	Przykłady rezultatów detekcji oczu dla osób noszących okulary . . . . .	58
51	Przykłady poprawnie rozpoznanych ust . . . . .	59
52	Przykłady błędów w detekcji ust . . . . .	59
53	Nagranie z kamery przy oświetleniu jarzeniowym . . . . .	60
54	Przykłady wyników dla emocji <b>Szczyścia</b> . . . . .	61
55	Przykłady wyników dla emocji <b>Smutku</b> . . . . .	61
56	Przykłady wyników dla emocji <b>Złości</b> . . . . .	61
57	Przykłady wyników dla emocji <b>Zaskoczenia</b> . . . . .	62
58	Przykłady wyników dla emocji <b>Odrzy</b> . . . . .	62
59	Przykłady wyników dla emocji <b>Strachu</b> . . . . .	62
60	Reakcje rozpaczy w różnych kręgach kulturowych w obliczu tragedii . . . . .	65

## Spis tabel

1	Action Unit w systemie FACS[36] . . . . .	21
2	Filtr Sobela . . . . .	29
3	Filtr Robertsa . . . . .	29
4	Filtr Prewitta . . . . .	29
5	Filtry Laplace'a . . . . .	30
6	Wagi gestów dla poszczególnych emocji . . . . .	47
7	Wyniki dla gestu obniżenia brwi . . . . .	57
8	Wyniki dla gestu uniesienia brwi . . . . .	57
9	Wyniki dla gestu ściśnięcia brwi . . . . .	57
10	Wyniki dla gestu otwarcia oczu . . . . .	58
11	Wyniki dla gestu przymrużenia oczu . . . . .	58
12	Wyniki dla gestu otwarcia ust . . . . .	59
13	Wyniki dla gestu uniesienia kąćików ust . . . . .	59
14	Wyniki dla gestu obniżenia kąćików ust . . . . .	59
15	Podsumowanie wyników klasyfikacji emocji . . . . .	63
16	Porównanie jakości rozpoznawania emocji . . . . .	63

## Literatura

- [1] Description of facial action coding system (facs). <http://face-and-emotion.com/dataface/facs/description.jsp>.
- [2] J. Gorczyca J. Skrzat A. Skawina, J. Walocha. *Szyja i głowa*. Wydawnictwo Uniwersytetu Jagiellońskiego, 2003.
- [3] Inc Android Technologies. Emorate. <http://www.androidreview.com/html/emorate.php>.
- [4] Antoni Benedikt. *Mowa ciała*. Astrum, 2002.
- [5] Darris Dobbs Bill Fleming. *Animacja cyfrowych twarzy*. Helion, 2002.
- [6] Jean-Yves Bouguet. Pyramidal implementation of the lucas kanade feature tracker. description of the algorithm.
- [7] Paul Breeuwsma. Bicubic interpolation. <http://www.paulinternet.nl/?page=bicubic>.
- [8] Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 1992.
- [9] Paul Ekman. *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Times Books, 2003.
- [10] Paul Ekman. *Kłamstwo i jego wykrywanie w biznesie, polityce i małżeństwie*. Wydawnictwo Naukowe PWN, 2008.
- [11] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. *Facial Action Coding System. A Human Face*, 2002.
- [12] Emotive. emotive epoc headset. <http://www.emotiv.com>.
- [13] Victoria Contreras Flores. Artnatomy. <http://www.artnatomia.net>.
- [14] Fobs4jmf. <http://fobs.sourceforge.net/>.
- [15] Rafael C. Gonzalez. *Digital Image Processing*. Prentice Hall, 2002.
- [16] Yepeng Guan. Robust eye detection from facial image based on multicue facial information. In *IEEE International Conference on Control and Automation*, 2007.
- [17] Hamid R. Pourreza Jalal A. Nasiri, Sara Khanchi. Eye detection algorithm on facial color images. *Second Asia International Conference on Modelling & Simulation*, 2008.
- [18] Java matrix package. <http://math.nist.gov/javanumerics/jama/>.
- [19] Java media framework. <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-140239.html>.
- [20] Jfreechart. <http://www.jfree.org/jfreechart/>.
- [21] Sohaib Khan. Hierarchical lucas kanade method. [http://server.cs.ucf.edu/~vision/Code/Optical\\_Flow/](http://server.cs.ucf.edu/~vision/Code/Optical_Flow/).

- 
- [22] Tomasz Kotliński. Automatyczne śledzenie twarzy i rozpoznawanie gestów na twarzy użytkownika komputera siedzącego przed kamerą. Master's thesis, AGH, 2009.
- [23] Antoni Kepiński. *Poznanie chorego*. Wydawnictwo Literackie, 2002.
- [24] Dale G. Leathers. *Komunikacja niewerbalna. Zasady i zastosowania*. Wydawnictwo Naukowe PWN, 2007.
- [25] Andrea Bulgarelli Luigi Di Stefano. A simple and efficient connected components labeling algorithm. *Image Analysis and Processing*, 1999.
- [26] Seeing Machines. facelab. <http://www.seeingmachines.com/product/facelab/>.
- [27] Jeannette M. Haviland-Jones Michael Lewis. *Psychologia emocji*. Gdańskie Wydawnictwo Psychologiczne, 2005.
- [28] Mårten Strömberg Mikael Rydfalk. Candide - a parameterized face. <http://www.bk.isy.liu.se/candide/>.
- [29] C. Pons P. Marteles M.P. González, E. Barrull. What is emotion? *Biopsychology.org*, 1998.
- [30] David. G. Myers. *Psychologia*. Wydawnictwo Zysk i S-ka, 2003.
- [31] University of Paderborn. Machine with emotionally extended intelligence. <http://www.c-lab.de/en/operation-fields/computing/organic-computing/mexi-a-robot-with-emotions/index.html>.
- [32] Allan Pease. *Mowa ciała. Jak odczytywać myśli innych ludzi z ich gestów*. Jedność, 2005.
- [33] prof. dr hab. inż. Małgorzata Kujawińska. *Cyfrowe metody przetwarzania obrazu*. Politechnika Warszawska.
- [34] Srikanth Rangarajan. Algorithms for edge detection.
- [35] Plutchik Robert. What is an emotion? *The Journal of Psychology*, 61, 1965.
- [36] Carnegie Mellon University Robotics Institute. Facs - facial action coding system. <http://www-2.cs.cmu.edu/afs/cs/project/face/www/facs.htm>.
- [37] Koichi Harada Roshan Dharshana Yapa. Connected component labeling algorithms for gray-scale images and evaluation of performance using digital mammograms. *International Journal of Computer Science and Network Security*, 8(6):33–41, June 2008.
- [38] Jacek Rumiński. *Metody reprezentacji, przetwarzania i analizy obrazów w medycynie*. Uniwersytet Łódzki.
- [39] Włodzimierz Szewczuk. *Encyklopedia psychologii*. Fundacja Innowacja, 1998.
- [40] Wikipedia. Connected component labeling. [http://en.wikipedia.org/wiki/Connected\\_Component\\_Labeling](http://en.wikipedia.org/wiki/Connected_Component_Labeling).