

Mirosław Jabłoński^{*}, Jaromir Przybyło^{*}, Paweł Wołoszyn^{*}

Automatyczna segmentacja twarzy dla potrzeb interfejsu człowiek-komputer**

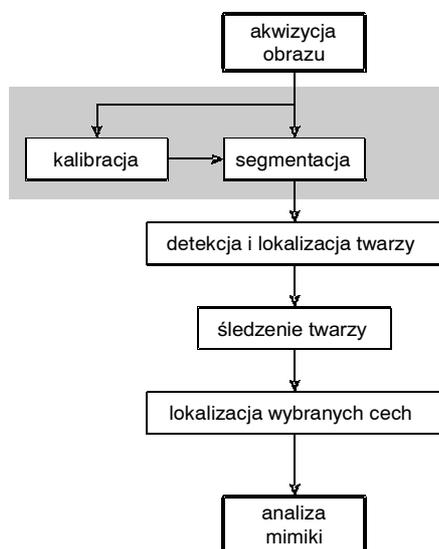
1. Wprowadzenie

W realizacji wizyjnych interfejsów człowiek-komputer opartych na analizie mimiki twarzy występuje konieczność odszukania na obrazie zarówno twarzy człowieka, jak i jej charakterystycznych cech, które będą następnie podstawą do dalszej analizy. Algorytmy poszukiwania twarzy charakteryzują się dużym zapotrzebowaniem na moc obliczeniową, ze względu na konieczność analizy całego obrazu [15, 12]. W praktyce uniemożliwia to działanie systemu w czasie rzeczywistym (czyli z założoną dla interfejsu szybkością). Poszukiwanie twarzy w każdej klatce sekwencji obrazów możemy jednak zastąpić poprzez etap początkowego jej zlokalizowania podczas kalibracji systemu, a następnie zastosowanie efektywnych algorytmów śledzenia do pracy ciągłej [7, 2] (rys. 1). Algorytmy śledzenia charakteryzują się niewielką złożonością obliczeniową, lecz wymagają precyzyjnej lokalizacji obszaru śledzonego obiektu, w tym wypadku twarzy, w pierwszej fazie działania, jak również później, okresowo dla adaptacji do zmiennych warunków otoczenia. Najczęściej używaną do zlokalizowania twarzy techniką są algorytmy segmentacji barwnej obrazu [1, 11], oparte na wykorzystaniu informacji o określonym zakresie kolorów, które mogą przyjmować piksele twarzy.

Segmentacja twarzy, oparta na klasyfikacji pikseli obrazu według modelu koloru skóry człowieka, jest wystarczająca w przypadku, gdy obiekty należące do tła posiadają inny rozkład kolorów w stosunku do poszukiwanego obiektu. W praktyce założenie takie często nie jest spełnione. Skutkuje to zaklasyfikowaniem części tła jako twarzy (artefakty) lub zaklasyfikowaniem części poszukiwanego obiektu jako tła. W przypadku gdy zadaniem segmentacji jest wyodrębnienie twarzy, utrudnieniem są również pojawiające się w polu widzenia kamery inne części ciała (np. dłonie) o takim samym modelu kolorów. Pod uwagę należy również wziąć zmiany rozkładu kolorów skóry człowieka, wynikające z właściwości kamery oraz zmienności oświetlenia. Aby zwiększyć skuteczność segmentacji barwnej, stosowane są różne metody adaptacji modelu koloru skóry [5].

^{*} Katedra Automatyki, Akademia Górniczo-Hutnicza, Kraków

^{**} Praca wykonana w ramach badań własnych – umowa AGH nr. 10.10.120.39



Rys. 1. Początkowe etapy procesu analizy mimiki twarzy prowadzonego przez interfejs człowiek-komputer. Szary obszar obejmuje etap segmentacji i automatycznej kalibracji proponowany w niniejszej pracy

W niniejszej pracy podjęto próbę opracowania algorytmu lokalizacji twarzy dla wizyjnego interfejsu człowiek-komputer. Algorytm realizuje segmentację obszaru twarzy w obrazie barwnym na podstawie automatycznie wyznaczonego modelu koloru skóry. Kalibracja współczynników modelu odbywa się w oparciu o analizę zmian jasności obrazu podczas intencjonalnego oświetlenia twarzy przez monitor komputerowy. Detekcja twarzy dla potrzeb interfejsu opiera się na założeniu, że użytkownik znajduje się w niewielkiej odległości od monitora komputera.

2. Akwizycja obrazu

Stanowisko badawcze, a zarazem docelowe, wykorzystywane do realizacji postawionego zadania składa się z powszechnie dostępnych komponentów:

- komputera PC,
- monitora ekranowego,
- kamery kolorowej,
- karty tunera telewizyjnego.

Jako podstawowe założenie związane ze stanowiskiem akwizycji, przyjęto zastosowanie ogólnie dostępnych i niedrogich kamer, takich jak wykorzystana mikrokamera płytkowa M10733 lub popularne kamery USB. Wiąże się to jednak z koniecznością pracy z obrazem o niskiej jakości. Kamera z szerokokątnym obiektywem zamocowana jest na monitorze.

W kadrze mieści się cała twarz użytkownika systemu. Obiektów został dobrany i ustawiony tak, aby obszar twarzy pokrywał ok. 20% rozmiaru ramki. Naturalne ruchy i przemieszczenia głowy użytkownika systemu są dopuszczalne w zakresie wyznaczonym przez kadr kamery.

Przyjęte parametry akwizycji obrazów to:

- głębia kolorów 24 bity,
- rozdzielczość 320×240 pikseli.

Sygnal obrazu transmitowany jest w formacie PAL.

Rezygnacja z wyspecjalizowanego wyposażenia optycznego i użycie popularnych urządzeń podyktowane jest zamiarem dostosowania wymagań systemu do uwarunkowań ekonomicznych przeciętnego użytkownika. Interfejs człowiek-komputer, będący nadrzędnym celem prowadzonych prac, z założenia powinien być dostępny dla każdej osoby posiadającej komputer osobisty wyposażony w typowe i niedrogi urządzenia peryferyjne. System nie może wymagać instalacji specjalistycznych kamer, obiektywów lub oświetlenia, dlatego też już na obecnym etapie jego tworzenia jest on dostosowywany do jakości obrazu oferowanej przez ogólnodostępne kamery.

Należy zauważyć, że popularny sprzęt wizyjny instalowany w zestawach komputerowych nie został zaprojektowany dla potrzeb zaawansowanego przetwarzania i analizy obrazu. Wiąże się z tym szereg problemów wymagających uwzględnienia przy opracowywaniu wizyjnego interfejsu człowiek-komputer o niewygórowanych wymaganiach sprzętowych.

Problemy te skupiają się wokół następujących aspektów akwizycji obrazu:

- szumu,
- ostrości i rozdzielczości obrazu,
- balansu bieli,
- oświetlenia i parametrów ekspozycji.

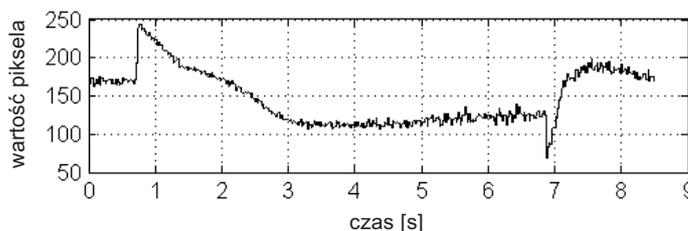
Obecność szumu, generowanego już przez sam sensor kamery, wprowadza stochastyczną zmienność koloru pikseli w następujących po sobie ramkach obrazu, przy czym losowe wariacje koloru pojedynczego piksela mogą być większe niż zmiany wynikające z przemieszczania się obiektów w kadrze. W przypadku tanich kamer internetowych stosunek sygnału do szumu jest mały i pogarsza się szybko w miarę zmniejszania oświetlenia, gdy układ automatycznej regulacji wzmocnienia dostosowuje kamerę do pracy w ciemniejszych warunkach. Wymusza to wprowadzenie do opracowywanych algorytmów segmentacji pewnych metod oddzielających szum od istotnych zmian koloru pikseli, na przykład poprzez lokalne uśrednianie wartości sąsiednich pikseli obrazu.

Ostrość obrazu pochodzącego z niedrogiej kamery jest warunkowana nie tylko jakością jej układu optycznego, niejednokrotnie bardzo uproszczonego i pozbawionego możliwości regulacji, ale także interpolacją obrazu na różnych etapach jego akwizycji, zależną od fizycznej rozdzielczości przestrzennej przetwornika, formatu ramki transmitowanej do urządzenia przechwytyjącego obraz oraz rozdzielczości docelowej. Z punktu widzenia segmentacji barwnej ważniejsza jest jednak rozdzielczość barwna obrazu wynikowego (głębina kolorów), która może być ograniczona do 5 bitów na kanał koloru, czyli 32 pozio-

mów jasności. Przetwarzanie takiego obrazu, realizowane w standardowej 8-bitowej rozdzielczości kanałów, prowadzi do uzyskania między innymi nieciągłych histogramów wartości pikseli.

Typowe kamery internetowe, wykorzystywane powszechnie przez użytkowników komputerów, są wyposażone w układ regulacji balansu bieli, pomagający unikać przebarwień obrazu w przypadku zmian temperatury barwowej oświetlenia. Działanie tego układu nie zawsze prowadzi jednak do pożądaných rezultatów i nie zapewnia powtarzalności kolorów rejestrowanych w kadrze obiektów szczególnie przy oświetleniu mieszanym (np. światło żarowe i światło monitora komputerowego). Z tego powodu nie jest możliwe z góry dokonanie założenia wartości koloru skóry twarzy i wykorzystanie go w segmentacji barwnej, gdyż w zależności od kontekstu i oświetlenia sceny odcień skóry przesuwają się w szerokim zakresie widma barw.

Wpływ oświetlenia na jakość obrazu uzyskiwanego z kamery obejmuje nie tylko stosunek sygnału do szumu, ale także nasycenie kolorów. Zakładając typowe warunki użycia projektowanego systemu, należy oczekiwać przypadkowego i słabego oświetlenia, przy którym nasycenie rejestrowanych kolorów jest niskie, a co za tym idzie, małe zmiany wartości poszczególnych składowych koloru (w tym również spowodowane szumem) powodują duże zmiany odcienia barwy pikseli. W trakcie prac badawczych zauważono również zakłócenia wprowadzane przez wspomniany wyżej układ automatycznego wzmocnienia stosowanej kamery. Lokalne zmiany jasności obrazu powodują globalne zmiany parametrów układu przetwornika wizyjnego kamery. Skutkuje to przesunięciem poziomu średnich wartości wszystkich pikseli w czasie. Zostało to zilustrowane na rysunku 2.



Rys. 2. Zmiana intensywności wybranego piksela wynikająca z działania układu automatycznego wzmocnienia kamery

3. Analiza zmienności oświetlenia

Jednym z problemów, jakie należy rozwiązać podczas automatycznej kalibracji etapu segmentacji twarzy, jest ustalenie modelu koloru skóry charakteryzującego twarz aktualnie obserwowanego użytkownika. Zmienność oświetlenia oraz aktywność układów automatycznej regulacji parametrów kamery powoduje, że modelowej wartości koloru skóry nie można ustalić arbitralnie. Aby wyznaczyć charakterystykę kolorystyczną skóry w sposób

automatyczny, konieczne jest wyodrębnienie z obrazu pikseli należących do skóry twarzy, a więc dokonanie segmentacji tego obiektu. Pod tym względem etap kalibracji zależny jest od etapu segmentacji, sam zaś dostarcza parametrów do jego przeprowadzenia. Z oczywistych zatem powodów podczas kalibracji nie można posłużyć się zwykłą segmentacją barwną, należy ją oprzeć na dodatkowym kryterium, niewykorzystywanym później podczas przetwarzania kolejnych ramek obrazu.

Zaproponowanym przez autorów dodatkowym kryterium segmentacji, wykorzystywanym tylko podczas kalibracji systemu, jest zmienność oświetlenia obiektów wywołana intencjonalnymi zmianami emisji światła przez monitor komputera, przed którym znajduje się użytkownik. Impulsowe oświetlenie twarzy przez ekran monitora stwarza możliwość odróżnienia jej od pozostałych obiektów znajdujących się w kadrze. Twarz użytkownika jest obiektem znajdującym się bezpośrednio przed ekranem w polu widzenia kamery zamontowanej na monitorze. Programowo sterowane rozjaśnienie i wygaszenie ekranu monitora powoduje zmiany jasności znajdujących się przed nim obiektów, tym większe, im bliżej monitora znajduje się dany obiekt. Efekt ten jest wyraźny, gdyż strumień światła emitowanego przez ekran maleje z kwadratem odległości od niego.

Wykorzystanie tego rodzaju aktywnego oświetlenia dla odróżnienia obiektu pierwszoplanowego od tła pozostaje w zgodzie z ogólnymi założeniami ekonomiczno-funkcjonalnymi systemu: nie wymaga użycia dodatkowego oświetlacza, niewchodzącego w skład wyposażenia typowego komputera osobistego. Z drugiej strony ze względów bezpieczeństwa użytkownika migotanie ekranu monitora musi zostać ograniczone do jednorazowego etapu wstępnego (oraz ewentualnie rzadko powtarzanej rekalkibracji) i nie może być powtarzane wielokrotnie. Te zastrzeżenia dotyczą przede wszystkim użytkowników z zaburzeniami neurologicznymi, u których migotanie ekranu może spowodować napad padaczkowy. Nie stanowi to jednak przeszkody w wykorzystaniu tej techniki do jednorazowej kalibracji systemu, pomiędzy rozjaśnieniem i wygaszeniem monitora może upłynąć czas rzędu sekundy, zaś użytkownik może w tym czasie mieć zamknięte oczy.

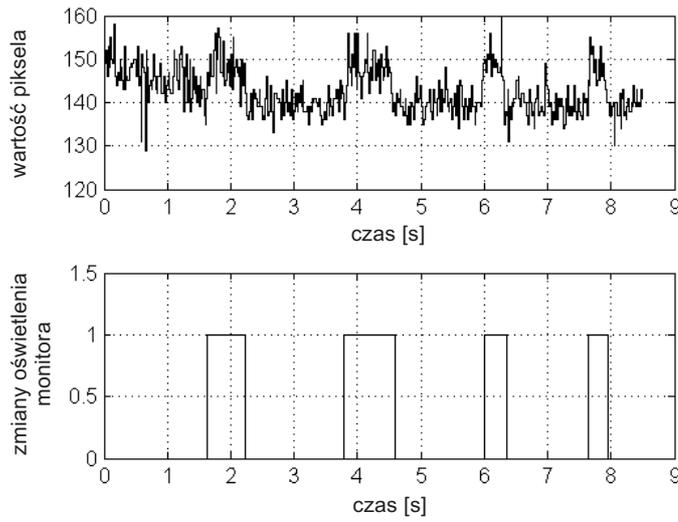
W celu weryfikacji takich założeń algorytmu kalibracji przeprowadzona została analiza zmian intensywności pikseli z wybranych obszarów obrazu zachodzących pod wpływem impulsowych zmian jaskrawości monitora. Pod uwagę brano składową czerwoną koloru.

Jako wybrane obszary obrazu przyjęto:

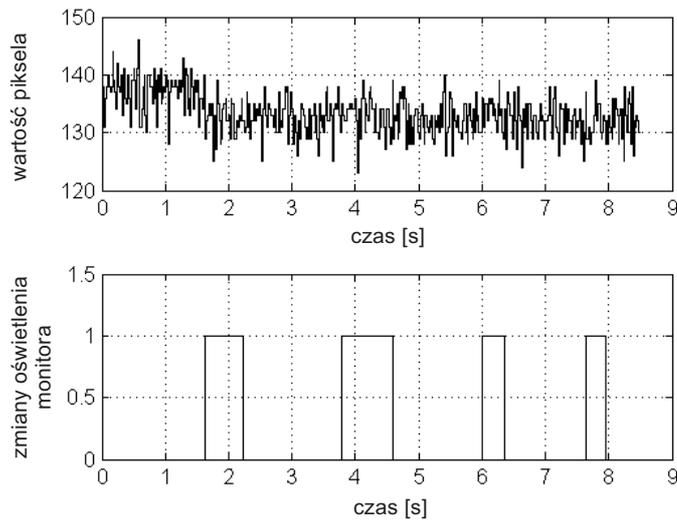
- pojedynczy piksel zlokalizowany w obszarze twarzy (rys. 3),
- pojedynczy piksel zlokalizowany w obszarze tła (rys. 4),
- zbiór pikseli obszaru należących do twarzy (rys. 5),
- zbiór pikseli obszaru należących do tła (rys. 6).

Dla pojedynczych pikseli należących do obiektu zmiana intensywności wywołana aktywnym oświetleniem jest niewielka i porównywalna z szumami wprowadzanymi przez czujnik CCD kamery. Natomiast średnia wartość pikseli dla całego obszaru obiektu (piksele zaklasyfikowane do obiektu na podstawie wstępnej segmentacji barwnej) pozwala już na

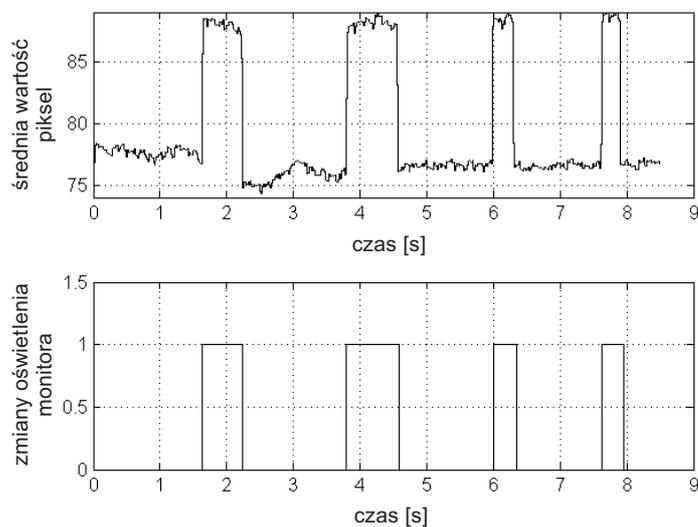
przyjęcie określonego kryterium klasyfikacji i odróżnienie go od obszaru tła. Zwiększenie stosunku sygnału do szumu jest już wyraźnie zauważalne nawet dla obszarów o rozmiarze 2×2 i 3×3 . Przeprowadzona analiza pozwala stwierdzić, iż intencjonalne zmiany oświetlenia są wykrywalne, jeśli zostanie uwzględniony kontekst otoczenia pikseli.



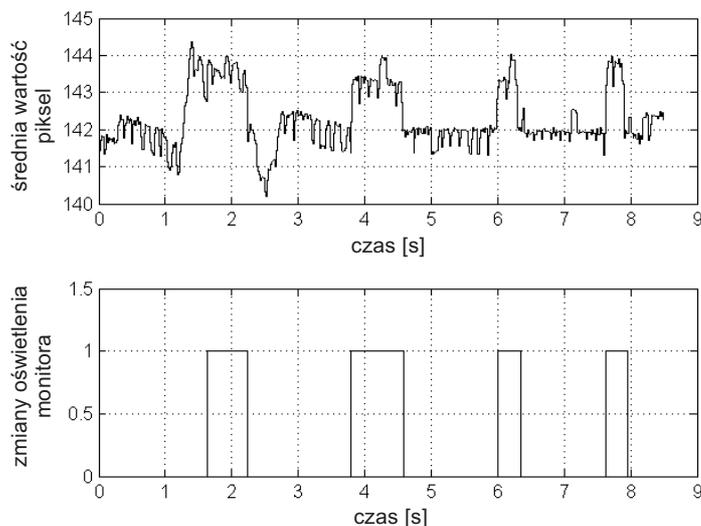
Rys. 3. Zmiany intensywności pojedynczego piksela należącego do twarzy podczas impulsowych zmian jaskrawości monitora (0 – wygaszony, 1 – rozjaśniony)



Rys. 4. Zmiany intensywności pojedynczego piksela należącego do tła podczas impulsowych zmian jaskrawości monitora



Rys. 5. Zmiany średniej intensywności pikseli twarzy podczas impulsowych zmian jaskrawości monitora



Rys. 6. Zmiany średniej intensywności pikseli tła podczas impulsowych zmian jaskrawości monitora

4. Procedura kalibracji i segmentacji

Podstawowe algorytmy segmentacji jako dane wejściowe pobierają obraz monochromatyczny. oraz barwny pobrany z kamery niesie jednak znacznie więcej informacji niż reprezentowany w skali szarości.

Dodatkowe dane mogą być użyteczne w wyznaczaniu obszaru, gdzie znajduje się analizowany obiekt, czyli twarz użytkownika systemu. Metody oparte na segmentacji obrazu monochromatycznego zostały odrzucone jako nieefektywne, ze względu na zmienne warunki akwizycji obrazu. Twarz może pojawić się na tle innych obiektów o podobnym poziomie jasności. Co więcej, jasność tła i obiektu może się w dużym zakresie zmieniać ze względu na czynniki zewnętrzne, wpływ dodatkowego oświetlenia lub odbłask monitora umieszczonego przed użytkownikiem systemu.

Format RGB jest jednym z podstawowych formatów danych obrazu, szeroko wykorzystywany w torach obróbki sygnału wizyjnego: w akwizycji, przetwarzaniu, analizie, ale również w wizualizacji. Przestrzeń kolorów RGB stanowi również punkt wyjścia do transformacji pikseli obrazu do innych przestrzeni barw, takich jak HSV, YUV, LUV, normalizowany rgb. Zagadnienia związane z wyznaczaniem składowych w poszczególnych przestrzeniach kolorów oraz ich własności są opisane szczegółowo w pracy [13].

Zarówno w grafice komputerowej, jak i analizie obrazów, obok przestrzeni RGB często stosuje się również przestrzeń HSV ze względu na jej naturalną interpretację. W odróżnieniu od przestrzeni RGB, poszczególne składowe nie są skojarzone bezpośrednio z przedziałami widma obrazu, lecz korespondują z subiektywną percepcją promieniowania widzialnego przez ludzkie oko. Składowa H reprezentuje skalę odcieni, S odpowiada nasyceniu barwy, a V reprezentuje jasność piksela. Ograniczeniem analizy w przestrzeni HSV jest duża złożoność obliczeniowa transformacji RGB do HSV. Ma to szczególne znaczenie ze względu na wymóg szybkości przetwarzania ramek obrazu oraz udostępnienia określonej mocy obliczeniowej procesora dla znacznie bardziej wymagających algorytmów rozpoznawania i analizy twarzy.

Mając na uwadze powyższy wymóg, w ramach niniejszej pracy zrealizowano segmentację obszaru twarzy na podstawie obrazu w przestrzeni RGB, połączoną z procedurą automatycznej kalibracji parametrów, która wykorzystuje przestrzeń HSV. Transformacja przestrzeni RGB do HSV dokonywana jest dzięki temu jedynie podczas kalibracji i nie musi być powtarzana dla każdej ramki obrazu. Z kolei sama procedura kalibracji wykorzystuje obraz przeniesiony do przestrzeni HSV, gdyż w ten sposób łatwiej jest określić charakterystykę barwną skóry twarzy.

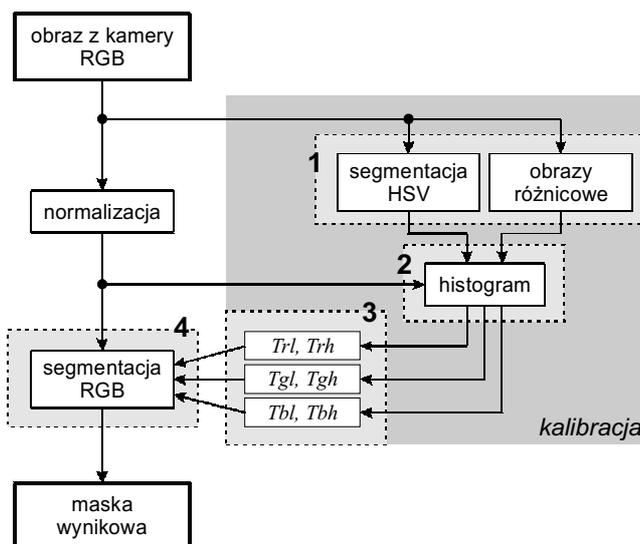
W przetwarzaniu obrazów mianem segmentacji określa się identyfikację jednolitych pod wybranym względem obszarów. Spośród istniejących metod segmentacji (np. przez progowanie, klasteryzację, wykrywanie krawędzi, rozrost, podział obszarów) wybrano najprostszą i jednocześnie najszybszą metodę: progowanie. Pozostałe metody wykorzystują rekurencję lub składają się z kilku etapów, przez co są zbyt czasochłonne w rozważanym zastosowaniu.

Idea proponowanej metody segmentacji ściśle wiąże się ze wykorzystywanym stanowiskiem akwizycji obrazów.

Na wyodrębnienie obszaru twarzy w obrazie uzyskanym z kamery składają się następujące etapy:

1. wykrycie obszaru twarzy na podstawie impulsowego oświetlenia monitorem,
2. wyznaczenie charakterystyki barwnej skóry,
3. obliczenie progów binaryzacji,
4. segmentacja.

Etapy 1, 2 i 3 wykonywane są podczas kalibracji systemu, etap 4 stanowi właściwą procedurę roboczą systemu (rys. 7).



Rys. 7. Etapy procedury segmentacji twarzy. Etapy 1–3 składają się na kalibrację systemu, etap 4 stanowi właściwą procedurę roboczą systemu

4.1. Kalibracja systemu – segmentacja HSV

W omawianym algorytmie kalibracji zastosowano metodę segmentacji obrazu barwnego po uprzedniej transformacji formatu RGB do HSV. Jak wspomniano wcześniej, w dziedzinie HSV kolor piksela jest jednoznacznie reprezentowany przez składową Hue. Binarystacja dwuprogowa składowej H pozwala jednoznacznie zaklasyfikować piksele mieszczące się w określonym zakresie barw zgodnie z zależnością

$$\mathbf{M}(x, y) = \begin{cases} 1: H(x, y) \in (TH_L, TH_H) \\ 0: else \end{cases} \quad (1)$$

gdzie TH_L i TH_H są odpowiednio dolnym i górnym progiem składowej H.

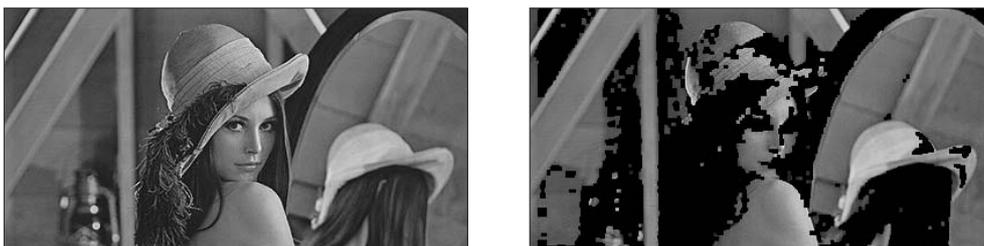
Eksperymenty wskazują, że takie progowanie pozwala wyodrębnić z obrazu zwarte obszary skóry twarzy. Okazuje się również, że w niektórych przypadkach włosy mieszczą się w tym samym zakresie barw. Znacząca różnica w stosunku do skóry uwidacznia się jedynie w jasności (składowa V). Wynika to z obecności zarówno w skórze, jak i we włosach tego samego pigmentu, melaniny, nadającej obu obszarom zbliżony odcień.

Wprowadzenie dodatkowego kryterium minimalnego nasycenia barwy (składowa S) pozwala wyeliminować z obrazu wynikowego punkty o zbyt niskim nasyceniu, gdzie odcień nie jest dobrze określony. Odrzucony przedział odpowiada pikselom szarym. Ostatecznie kryterium selekcji barw określa równanie

$$\mathbf{Mhs}(x, y) = \begin{cases} 1: H(x, y) \in \langle TH_L, TH_H \rangle \wedge \\ S(x, y) > TS_L \\ 0: else \end{cases} \quad (2)$$

gdzie TS_L jest dolnym progiem składowej S.

Wynikiem segmentacji w dziedzinie HSV jest obraz maski \mathbf{Mhs} , której elementy korespondują z pikselami obrazu o kolorze twarzy. Wadą takiego rozwiązania jest fakt, iż w rzeczywistych warunkach w widoku kamery mogą się pojawić również inne obiekty, których piksele spełniają równanie (2), a nie są punktami twarzy, np. meble. Rysunek 8 obrazuje wyniki tak przeprowadzonej segmentacji: poprzez dobór opisanych wyżej progów nie jest możliwe oddzielenia twarzy od tła.



Rys. 8. Rezultat segmentacji twarzy opartej na progowaniu HSV. Po lewej obraz oryginalny, po prawej wynik segmentacji programi $TH_L = 88$, $TH_H = 114$, $TS_L = 63$

Ponieważ segmentacja w dziedzinie HSV nie pozwala na jednoznaczne wyodrębnienie twarzy od tła, konieczne jest dodatkowe kryterium, które pozwoli precyzyjniej ustalić ten obszar. Autorzy wykorzystali tu omówione już wcześniej zmiany oświetlenia twarzy wywołane przez zmiany jasności ekranu monitora i uwidaczniające się w obrazach różnicowych kolejnych ramek. Do wyznaczania obrazu różnicowego wykorzystywany jest sygnał monochromatyczny uzyskany ze składowych RGB. Pozwala to zwiększyć stosunek sygnału do szumu. Dodatkowe polepszenie sygnału uzyskuje się poprzez odpowiednią analizę kolejnych obrazów różnicowych, biorąc pod uwagę kontekst pikseli. Wyznaczanie obiektu pierwszoplanowego polega na binaryzacji wynikowego obrazu różnicowego progiem Td oraz wskazaniu obiektu, który spełnia kryterium największego pola powierzchni Ad . Rezul-

tatem tego etapu jest maska **Md**, która odpowiada obiektowi umieszczonemu w bliskiej odległości od źródła światła. Złożenie

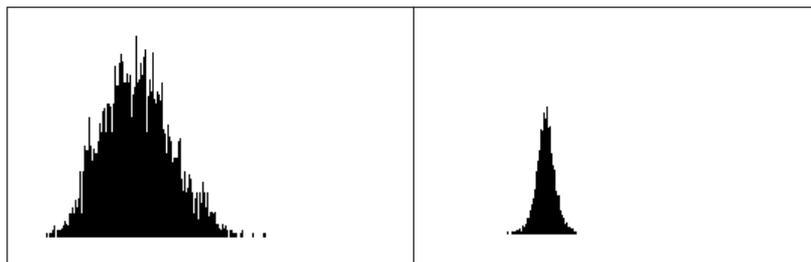
$$\mathbf{Mt} = \mathbf{Mhs} \wedge \mathbf{Md} \quad (3)$$

masek **Mhs** i **Md** daje w efekcie maskę **Mt** która może być zastosowana jako wiarygodne szacowanie obszaru twarzy.

4.2. Wyznaczanie progów binaryzacji

Zastosowane w pierwszym etapie progi TH_L , TH_H i TS_L pozwalają dokonać wstępnej klasyfikacji pikseli na należące do tła oraz do twarzy użytkownika. Parametry te przyjmują stałe wartości, odpowiadające granicznym zakresom barwnym analizowanych twarzy. Charakterystyka barwna skóry konkretnego użytkownika może jednak stanowić węższy podzbiór przedziału wyznaczonego przez te parametry. Określenie tej charakterystyki polega na pomiarze statystycznych parametrów obrazu źródłowego mieszczącego się w masce **Mt**.

Analiza histogramów składowych R, G i B wyznaczonego obszaru twarzy wskazuje, że charakteryzują się one wyraźnie unimodalnym charakterem (rys. 9). Cecha ta zostaje zachowana również dla składowych znormalizowanych r , g i b .



Rys. 9. Histogramy składowej R wyznaczonego obszaru twarzy (maska **Mt**). Po lewej histogram w przestrzeni RGB, po prawej w przestrzeni rgb (znormalizowana RGB). Skale pionowe histogramów są różne

Przedstawione rozkłady, z dużą dokładnością mogą być opisane funkcją Gaussa. Jako parametry reprezentatywne dla wykrytego obszaru wytypowano więc wartość średnią obrazu Avg oraz standardowe odchylenie Sdv . Zakładając zaobserwowaną unimodalność, można uznać, że zdecydowana większość wartości pikseli mieści się w granicach określonych przez progi:

$$\begin{aligned} T_L &= Avg - k \cdot Sdv \\ T_H &= Avg + k \cdot Sdv \end{aligned} \quad (4)$$

Wartości progów bezpośrednio zależą od statystycznych własności pikseli twarzy. Współczynnik k ustalono doświadczalnie na wartość 1,5.

4.3. Segmentacja w dziedzinie rgb

Jest pożądaną, aby właściwa segmentacja obrazu kolorowego realizowana była możliwie szybko. Konwersja obrazu z przestrzeni RGB do HSV jest złożona obliczeniowo i czasochłonna, z tej racji rzadko stosowana. Z kolei zastosowanie binaryzacji dwuprogowej składowych RGB w oparciu o wartości T_L i T_H nie gwarantuje poprawnego działania algorytmu przy zmianie oświetlenia. Podobnie, treść wyświetlana na monitorze wpływa na parametry statystyczne danych obrazowych Avg i Sdv . Do klasyfikacji pikseli wybrano zatem binaryzację dwuprogową obrazu w przestrzeni znormalizowanej rgb. Poszczególne składowe wyznaczane są według zależności:

$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \\ b &= \frac{B}{R+G+B} \end{aligned} \quad (5)$$

Zasadniczą zaletą tej przestrzeni jest niezmienniczość ze względu na proporcjonalne zmiany wszystkich składowych RGB [13]. Do ostatecznego wyznaczenia obszaru twarzy wykorzystywane są więc progi dla poszczególnych składowych rgb: Tr_L , Tr_H , Tg_L , Tg_H , Tb_L , Tb_H , określone na podstawie zależności (4). Wartości te wyznaczane są we wcześniej opisanej procedurze kalibracji na podstawie analizy obszaru twarzy wykrytego za pomocą impulsowego oświetlenia monitorem.

Rysunki 10 i 11 ilustrują przedstawione wyżej etapy przetwarzania obrazów. Na rysunku 10 przedstawiono obraz oryginalny RGB oraz maski, których iloczyn logiczny daje obszar uwzględniany w kalibracji progów binaryzacji rgb. Rysunek 11 obrazuje klasyfikację pikseli na obszary tła i twarzy. Kolor znajdujących się w otoczeniu mebli nie odbiega zauważalnie od barwy skóry.



Rys. 10. Etapy kalibracji. Po lewej obraz źródłowy z kamery, pośrodku maska wyznaczona na podstawie progów HSV, po prawej maska obiektu na pierwszym planie



Rys. 11. Wynik segmentacji. Maska zawiera piksele zaklasyfikowane jako należące do twarzy

5. Wnioski

Zagadnieniem poruszonym w niniejszej pracy była segmentacja obszaru twarzy na podstawie obrazu barwnego, połączona z procedurą automatycznej kalibracji modelu koloru skóry. Przeprowadzona analiza zmian intensywności pikseli obrazu pozwoliła na wyspecyfikowanie założeń algorytmu automatycznej kalibracji. Opracowany algorytm był podstawą automatycznego wyznaczania parametrów dla segmentacji obrazu w przestrzeni normalizowanej RGB. Krótka procedura kalibracji umożliwia łatwe ustalenie parametrów w oparciu o kolorowy obraz.

W przypadku gdy obiekty należące do tła posiadają rozkład kolorów różny od poszukiwanego obiektu, segmentacja barwna daje zadowalające rezultaty. W rzeczywistych warunkach założenie powyższe jednakże często jest niespełnione. Aby zwiększyć skuteczność detekcji twarzy, wykorzystuje się dodatkowe założenia *a priori*, na przykład dotyczące cech morfologicznych głowy człowieka. Przykładem może być rozwiązanie zaproponowane w [4], w którym analizowany jest kształt wyodrębnionych z obrazu obszarów (dopasowanie elipsy o zadanych parametrach). Często dodatkowa analiza opiera się na detekcji oczu i ust w hipotetycznym obszarze twarzy. Na przykład w pracach [8, 10] autorzy wykorzystali efekt jasnej źrenicy, powstający po odbiciu światła od siatkówki oka oświetlonego aktywnym źródłem podczerwieni. Niektóre rozwiązania opierają się na informacji o ruchu obiektu, tak jak w pracy [9], gdzie przedstawiono sposób inicjalizacji algorytmu śledzenia oparty na detekcji ruchu brwi.

Dalszym etapem prac będzie poszukiwanie metod analizy obrazu pozwalających na wiarygodną klasyfikację pikseli obrazu oraz połączenie ich z algorytmami śledzenia wybranych cech twarzy.

Literatura

- [1] Chai D., Ngan K.N.: *Face segmentation using skin color map in videophone applications*. IEEE Transactions on Circuits and Systems for Video Technology, t. 9, nr 4, 1999, 551–564
- [2] Comaniciu D., Ramesh V., Meer P.: *Real-time tracking of nonrigid objects using mean shift*. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, t. II, 2000, 142–149

- [3] Cutler R., Davis L.: *View-based detection and analysis of periodic motion*. International Conference on Pattern Recognition, Brisbane, Australia, 1998, 495–500
- [4] Eleftheriadis A., Jacquin A.: *Model-assisted coding of video teleconferencing sequences at low bit rates*. Proc. ISCAS '94, 1994
- [5] Fritsch J., Lang S., Kleinhagenbrock M., Fink G.A., Sagerer G.: *Improving adaptive skin color segmentation by incorporating results from face detection*. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), Berlin, 2002, 337–343
- [6] Heikkila J., Silven O.: *A real-time system for monitoring of cyclists and pedestrians*. Second IEEE Workshop on Visual Surveillance, Fort Collins, Colorado, 1999, 74–81
- [7] Isard M., Blake A.: *CONDENSATION – conditional density propagation for visual tracking*. Int. J. Computer Vision, t. 29, nr 1, 1998, 5–28
- [8] Kapoor A., Picard R.W.: *Real-Time, Fully Automatic Upper Facial Feature Tracking*. Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washinton, 2002
- [9] Lombardi J., Betke M.: *A Self-initializing Eyebrow Tracker for Binary Switch Emulation*. Boston University, Computer Science Department, Technical Report BUCS-TR-2002-023, 2002
- [10] Przybyło J., Jabłoński M., Wołoszyn P.: *Wizyjny interfejs człowiek-komputer dla użytkowników niepełnosprawnych*. Półrocznik AGH Automatyka, t. 7, z. 3, 2003, 385–398
- [11] Schwerdt K., Crowley J.L.: *Robust Face Tracking Using Color*. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000
- [12] Shokoufandeh A., Dickinson S., Jonsson C., Bretzner L., Lindeberg T.: *On the Representation and Matching of Qualitative Shape at Multiple Scales*. European Conference on Computer Vision, Kopenhaga, 2002
- [13] Skarbek W.: *Segmentation of Colour Images*. Warszawa, PAN 1994
- [14] Stauffer Ch., Grimson W.E.L.: *Adaptive background mixture models for real-time tracking*. CVPR99, Fort Colins, 1999
- [15] Turk M., Pentland A.: *Eigenfaces for recognition*. Journal of Cognitive Neuroscience, t. 3, nr 1, 1991, 71–86