

Sławomir Jeżewski\*, Piotr Duch\*

## **Algorytmy segmentacji obrazów barwnych w rozpoznawaniu obiektów na obrazach satelitarnych i lotniczych**

### **1. Wprowadzenie**

Percepcja koloru odgrywa istotną rolę w badaniach nad analizą i przetwarzaniem obrazu. Między innymi dlatego, że kolor stanowi bogate źródło informacji dla procedur segmentacji oraz dostarcza użytecznych danych do algorytmów ogniskowania uwagi i śledzenia obiektów. W literaturze podkreśla się fakt, że informacja kolorystyczna jest stosunkowo stała przy zmianach natężenia oświetlenia i kąta patrzenia, a to ułatwia zadanie wykrywania obiektów [1, 2]. Źródłem tej stałości upatruje się głównie w separacji informacji kolorystycznej na zmienną iluminację i względnie stałą chrominancję. W [3] przedstawiono wyczerpujący przegląd procedur segmentacji obrazów barwnych, a w ramach niego ogólny podział procedur segmentacji na metody pikselowe, otoczeniowe i fizykalne. Podobny przegląd przedstawił Cheng i współautorzy [4]. Przedstawiając mnogość algorytmów segmentacji autorzy nie wskazali jednoznacznie na najlepszy z nich, głównie dlatego, że użyteczność regionów posegmentacyjnych silnie zależy od sposobu ich wykorzystania.

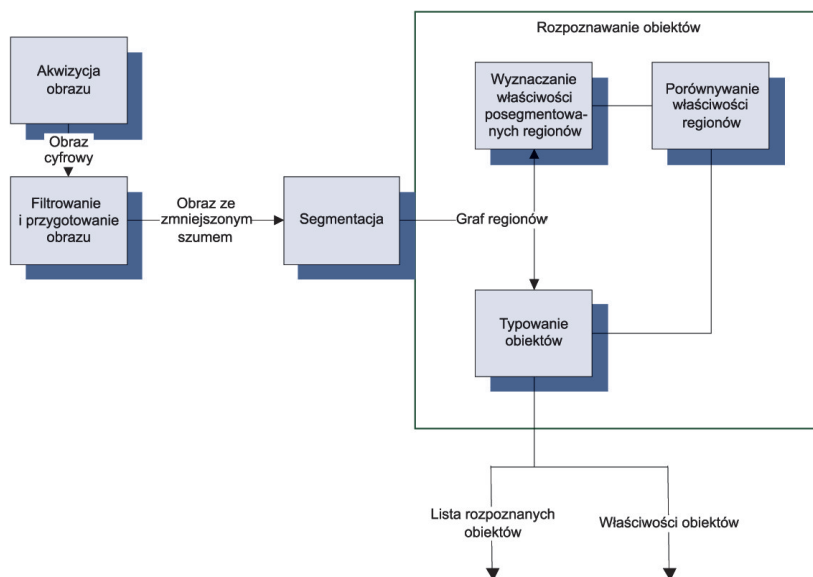
### **2. Dwie koncepcje segmentacji obrazów kolorowych**

Według uznanej definicji segmentacja jest to proces podziału obrazu na rozłączne obszary nazywane regionami. Segmentacja prowadzona jest przy wykorzystaniu arbitralnie wybranego kryterium funkcyjnego działającego jako kryterium „podobieństwa” pikseli lub przeciwnie jako kryterium różnicowe. Kryteria te w ten czy inny sposób powstają w wyniku statystycznej, analizy całego obrazu, jego fragmentu lub sekwencji obrazów. Propozycje algorytmów segmentacji można znaleźć w wielu podręcznikach i artykułach naukowych, by wymienić tylko [5, 6]. Wyczerpujący przegląd procedur segmentacji dla obrazów monochromatycznych przedstawia między innymi [7]. Większość autorów projektując i opisując procedury segmentacji, przyjmuje, że segmentacja obrazu i rozpoznawanie obiektów to odrębne operacje oraz że zbiór posegmentowanych regionów jest podstawowym typem

---

\* Katedra Informatyki Stosowanej, Politechnika Łódzka

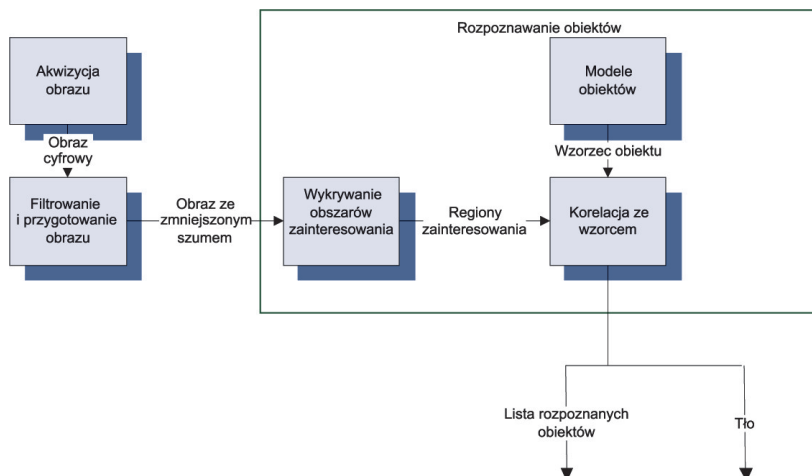
danych do rozpoznania obiektów (rys. 1). W fazie rozpoznawania obiektów wyznaczają właściwości posegmentowanych regionów i dokonują klasyfikacji obiektów w oparciu o otrzymany wektor właściwości. Przy takim podejściu istotnym jest by segmentacja wiernie odtworzyła krawędzie pomiędzy fizycznymi obiektami w scenie. Jeśli tak nie jest, algorytm segmentacji jest oceniany jako nieadekwatny do postawionego zadania.



**Rys. 1.** Segmentacja w ujęciu tradycyjnym

Podejściu tradycyjnemu można przeciwstawić podejście, w którym proces segmentacji i proces rozpoznawania obiektów stanowi nierozdzielalną całość (rys. 2). Wynikiem tak zdefiniowanej segmentacji jest podział na obiekty rozpoznane oraz obiekty nierozpoznane (tło). Rozpoznanie obiektu opiera się na modelu obiektu powstającym w wyniku uczenia sieci neuronowej lub systemu klasyfikatorów. Ten typ rozpoznawania obiektów odbywa się bez jawnej fazy wyznaczania właściwości regionów, stąd opozycja do tradycyjnego modelu segmentacji. Podstawą tego typu segmentacji najczęściej są korelacyjne metody porównawcze, inspirowane badaniami neurobiologicznymi. Wynik działania wielowymiarowej korelacji można traktować jako funkcję gęstości prawdopodobieństwa i łączyć ją z wynikami innych obliczeń, dając podstawę do wnioskowania wielokryterialnego. Najtrudniejszym problemem związanym z podejściem korelacyjnym jest spory nakład obliczeniowy proporcjonalny do wielkości obrazu i ilości rozpoznawanych wzorców. Współczesne maszyny obliczeniowe nie radzą sobie z wielowymiarową korelacją w czasie rzeczywistym, szczególnie te implementowane w pojazdach. Aby ominąć ten problem, część autorów postuluje wytworzenie mapy regionów zainteresowania (saliency map) w obrazie, aby wykluczyć z przeszukiwania korelacyjnego te części obrazu, które na pewno nie zawierają interesujących obiektów.

W sensie formalnym mapa zainteresowania jest dwuwartościową funkcją skrótu z wielowymiarowej przestrzeni poszukiwań a więc segmentacją w tradycyjnym ujęciu tego terminu. W dalszej części artykułu ta forma segmentacji będzie zwana presegmentacją.



Rys. 2. Segmentacja korelacyjna

Presegmentacja i segmentacja tradycyjna mają podobną koncepcję, lecz całkowicie różne cele, co przekłada się na inny dobór rozwiązań. W przypadku podejścia tradycyjnego najważniejsze jest dokładne wyznaczenie położenia regionów, ich krawędzi. Algorytmy ocenia się pod kątem korespondencji pomiędzy krawędziami regionów wyznaczonymi przez człowieka a tymi wyznaczonymi przez algorytmy. Oczekuje się, że krawędzie pozorne *false edges* powstające w wyniku światłocieni na zakrzywionych powierzchniach bądź teksturowania, zostaną pominięte.

W przypadku presegmentacji dokładność odwzorowania krawędzi nie jest najistotniejsza. Wystarczy znaleźć przybliżone położenie regionów oraz określić, czy mają one zostać poddane późniejszym badaniom. Przy presegmentacji ważna jest szybkość działania algorytmów, jako że będą wykonywane w systemie wielokrotnie, z różnymi ustawieniami i tyczyć będą całego obrazu. Krawędzie pozorne pojawiające się w obrazie nie stanowią przeszkody, jednak pomnażają punkty zainteresowania.

Wykazana wyżej rozbieżność w oczekiwaniach wobec algorytmów segmentacyjnych i presegmentacyjnych skłoniła autorów do ponownej oceny istniejących algorytmów segmentacji.

### 3. Kodowanie kolorów i presegmentacja

Poprawne działanie procedur segmentacji i presegmentacji obrazu zależy od wielu czynników towarzyszących procesowi akwizycji obrazu, również od sposobu wstępnej cyfrowej obróbki obrazu. Przykładowo „rozdzielczość” procedur segmentacji, czyli zdolność

do rozróżniania drobnych detali zdeterminowana jest przede wszystkim szumem przetworników analogowo cyfrowych i jego charakterystykami spektralnymi, doбором procedur segmentacji. Podobny wpływ na działanie procedur segmentacji posiada sposób kodowania koloru i ewentualne całkowitoliczbowe przekształcenia kolorów. W trakcie procesu przetwarzania obrazu kodowanie koloru można zmienić, lecz zwykle wiąże się to z utratą dokładności wyników za sprawą całkowitoliczbowego zapisu wyników i szumu kwantowania. Poniżej przedstawiono charakterystykę podstawowych systemów kodowania koloru oraz ich ocenę pod kątem zastosowań presegmentacyjnych.

Przestrzeń kolorów **RGB** naśladuje sposób postrzegania kolorów w oku człowieka i jest najczęściej stosowanym w technice sposobem kodowania kolorów. Jak większość metod kodowania koloru wywodzi się z podstawowego założenia kolorymetrii o addytywnym mieszanii kolorów. Podstawowe kolory R, G, B stosowane w technice mają zakresy spektralne określone normą i są traktowane jako kolory referencyjne. Poniższe zależności przedstawiają odpowiedź sensora o funkcji czułości spektralnej  $C(R(\lambda), G(\lambda), B(\lambda))$  na światło padające o radiacji  $E(\lambda)$

$$R = \int_{\lambda} E(\lambda) R(\lambda) d\lambda \quad G = \int_{\lambda} E(\lambda) G(\lambda) d\lambda \quad B = \int_{\lambda} E(\lambda) B(\lambda) d\lambda \quad (1)$$

W wyniku digitalizacji wszystkie wynikowe wektory (R, G, B) posiadają wartości całkowitoliczbowe w zakresie  $0-C_{\max}$ . Ponieważ funkcje czułości  $C(\lambda)$  poszczególnych kamer i skanerów różnią się, obraz tego samego obiektu obserwowany dwoma różnymi urządzeniami może być reprezentowany inną trójką RGB. Jeżeli przestrzeń RGB będzie traktowana jako przestrzeni euklidesowa, wówczas obliczenie odległości dwóch kolorów od siebie sprowadza się do obliczenia odległości euklidesowej między dwoma punktami w przestrzeni trójwymiarowej. Im ta odległość jest mniejsza, tym kolory są bardziej do siebie podobne. Niestety tak obliczone podobieństwo kolorów słabo odwzorowuje doznania kolorystyczne człowieka, dlatego stosowana jest zmodyfikowana wersja tej odległości z dodanymi wagami [2]. Wartość wagi zależy od tego, jak bardzo różnią się badane kolory w kanale czerwonym. Odległość ( $\Delta C$ ) pomiędzy kolorami  $C1 = (R1, G1, B1)$  i  $C2 = (R2, G2, B2)$  (gdzie każdy z nich posiada kanał czerwony, zielony i niebieski z wartościami z zakresu 0–255) wyraża się zależnościami:

$$\Delta C = \sqrt{\left(2 + \frac{r}{256}\right) \Delta R^2 + 4 \Delta G^2 + \left(2 + \frac{255-r}{256}\right) \Delta B^2}$$

$$r = \frac{R_1 + R_2}{2}$$

Przestrzeń RGB jest często transformowana do innego systemu kolorów zarówno w przemysłowych, jak i naukowych zastosowaniach obrazów. Część z tych transformacji ma charakter liniowy.

### Transformacje liniowe

W tabeli 1 przedstawiono częściej wykorzystywane transformaty liniowe RGB.

**Tabela 1**  
Liniowe przekształcenia kolorów

YIQ	$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.253 & -0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$
YUV	$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.437 \\ 0.615 & 0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$
$I_1 I_2 I_3$	$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$
NTSC XYZ	$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$

Wiążą one ze sobą wyniki działania różnych urządzeń akwizycji (kamery PAL, NTSC, drukarki i inne) i pozwalają na kolorystyczne „dopasowanie” do siebie tych urządzeń.

W segmentacji i rozpoznawaniu obiektów liniowe transformacje koloru wydają się mało użyteczne. Zastosowanie nieosobliwego przekształcenia liniowego transformującego trójwymiarową przestrzeń w inną trójwymiarową przestrzeń nie wpływa na rozdzielność klas w tej przestrzeni. Jeśli dwie klasy są separowane za pomocą powierzchni  $F(R, G, B) = 0$ , to po liniowej transformacji pozostają separowane z użyciem innej powierzchni  $F_1(R, G, B) = 0$ , pozostałe warunki rozdzielczości pozostają bez zmian. Liniowa transformacja koloru nie mając sensu matematycznego, ma jednak sens z obliczeniowego punktu widzenia. Jeśli wynikowa powierzchnia separująca posiada prostsze obliczeniowo równanie lub zanika w niej jedna z trzech współrzędnych, obliczenia w fazie segmentacji będą przebiegały znacznie szybciej, co jest bardzo istotne w procesie presegmentacji. Zysk czasowy z zastosowania liniowej transformacji koloru będzie szczególnie duży, jeśli system akwizycji będzie w stanie sprzętowo transformować kolor z zadowalającą dokładnością.

### Transformacje nieliniowe

W przypadku ogólnym powierzchnie separujące klasy  $F(R, G, B) = 0$  nie są płaszczyznami, lecz powierzchniami wyższego rzędu. Stosując nieliniową transformację, można uprościć powierzchnię separującą do postaci płaszczyzny lub lepiej jej wycinka. Powstałe w ten sposób przekształcenia nieliniowe często posiadają użyteczne właściwości. Powszechnie znane są przekształcenia rgb, HSV, CIE Lab.

Trójwymiarową przestrzeń RGB można zredukować o jeden wymiar poprzez przekształcenie normalizujące:

$$r = \frac{R}{R+G+B} \quad b = \frac{B}{R+G+B} \quad g = \frac{G}{R+G+B} \quad (2)$$

Suma  $Y = R + G + B$  ma wymiar natężenia światła, para  $r, g$  lub  $r, b$  stanowi wektor chromatyczny. Analiza w przestrzeni  $rg$  stanowi podstawę separacji kolorów np. przy wykrywaniu cieni. Pomimo istotnych zalet przestrzeni  $rg$  posiada istotną wadę związaną z osobliwością przestrzeni w punkcie  $R = G = B = 0$ . W okolicy tego punktu błędy wynikające z szumu akwizycji czy błędy kwantowania istotnie zmieniają wartości, pozbawiając wiarygodności wyniki segmentacji.

Druga z przestrzeni, przestrzeń HSV jest skutkiem badań nad percepcją kolorów przez człowieka. Ten typ przedstawiania kolorów lepiej niż RGB dopasowuje matematyczne podobieństwo kolorów do intuicyjnego odczucia człowieka. Nazwa modelu pochodzi od pierwszych liter angielskich nazw dla składowych opisów barwy: *hue* (barwa światła), *saturation* (nasycenie światła) oraz *value* (wartość). Przestrzeń HSV opisuje kolor jako punkt w stożku, w którym centralna oś ma zakres od barwy czarnej na szczycie do białej przy podstawie z kolorami w odcieniach szarości między nimi, kąt wokół osi odpowiada barwie, odległość od osi odpowiada nasyceniu, a odległość wzdłuż osi odpowiada jasności i jest o tej samej intensywności co światło kolorowe. Transformacja HSV wprowadza metrykę podobieństwa, która jest nie mniej skomplikowana niż  $rg$  czy RGB (3), mimo to wymaga wyliczania funkcji trygonometrycznych i pierwiastków [8].

$$dC = \sqrt{\Delta H^2 + \Delta C^2} \quad (3)$$

$$\Delta C = \sqrt{S_1^2 + S_2^2 - 2S_1S_2 \cos(\theta)}$$

$$\theta = \begin{cases} |\Delta H| & \text{jeżeli } |\Delta H| > \pi \\ 2\pi - |\Delta H| & \text{jeżeli } |\Delta H| < \pi \end{cases}$$

Trzeci system kodowania kolorów – przestrzeń barw CIE  $L^*a^*b^*$  (CIELab) została opracowana i zatwierdzona przez International Commission on Illumination. Opisuje wszystkie kolory widziane przez ludzkie oko. Przestrzeń ta jest opisana jako trójwymiarowy system koordynat reprezentowany przez jasność koloru ( $L^* = 0$  jest kolorem czarnym,  $L^* = 100$  wskazuje na kolor biały), jego położeniem między kolorem czerwonym i zielonym ( $a^*$ , ujemne wartości wskazują na kolor zielony, a dodatnie na czerwony) oraz położenie między kolorem żółtym a niebieskim ( $b^*$ , ujemne wartości wskazuje na kolor niebieski, a dodatnie na żółty). Względna różnicę dwóch kolorów w przestrzeni barw CIELab można aproksymować, traktując każdy kolor jako punkt w trójwymiarowej przestrzeni i obliczając odległość euklidesową pomiędzy tymi punktami [8–9].

$$dC = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2} \quad (4)$$




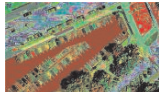
















Badania literaturowe wskazują na wysoką użyteczność nieliniowych transformacji koloru w procesie rozpoznawania obiektów. Niestety ich wadą jest duży nakład obliczeniowy

i punkty osobliwe przekształceń. Duży koszt obliczeniowy ogranicza zastosowanie znanych transformacji nieliniowych w procesie presegmentacji, nie mniejszym problemem jest istnienie punktów osobliwych Przekształcenia nieliniowe mimo że kosztowne obliczeniowo, są użytecznym elementem presegmentacji.


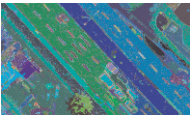
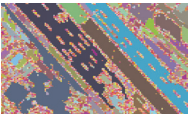
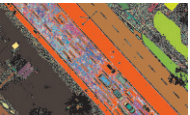


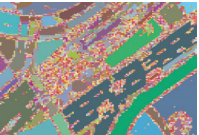


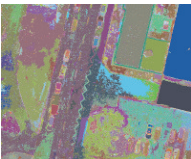

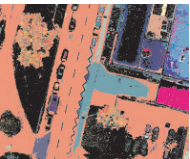




#### 4. Wyniki badań

Badania skuteczności algorytmów segmentacji obrazów barwnych zostały przeprowadzone na serii zdjęć lotniczych pobranych z GoogleMaps oraz na zdjęciach wysokiej rozdzielczości. Każde zdjęcie zostało przetworzone przez trzy algorytmy segmentacyjne. Ustawienia wejściowe parametrów algorytmów zostały ustalone przed rozpoczęciem eksperymentu. Do eksperymentu zostały zastosowane autorskie wersje popularnych algorytmów quadtree oraz watershed. Algorytmy te wykorzystują przestrzeń barw HSV. Tak jak już było wspomniane we wcześniejszym rozdziale, głównym zadaniem tych algorytmów jest wyznaczenie mapy obszarów zainteresowań. Wyniki zostały zamieszczone w tabelach 2–4.

**Tabela 2**  
Wyniki działania algorytmów na zdjęciach pobranych z GoogleMaps

Oryginalny obraz	Watersheed (HSV)	Quadtree (HSV)	Tominaga (HSV, RGB, Lab)
802 x 504 (JPEG) 			
827 x 506 (JPEG) 			
803 x 486 (JPEG) 			
816 x 486 (JPEG) 			
824 x 443 (JPEG) 			




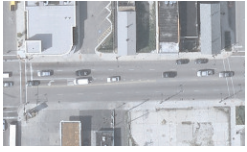

**Tabela 3**  
Wyniki działania algorytmów na zdjęciach wysokiej rozdzielczości

Oryginalny obraz	Watersheed (HSV)	Quadtree (HSV)	Tominaga (HSV, RGB, Lab)
634 x 379 (TIFF) 			
670 x 457 (TIFF) 			
638 x 529 (TIFF) 			
622 x 481 (TIFF) 			

Wyniki przedstawione w tabelach 2 oraz 3 pokazują, że rozdzielczość liniowa zdjęć ma niewielki wpływ na jakość segmentacji. Znacznie większy wpływ ma sposób kodowania i kompresji kolorów. W przypadku kompresji JPEG wyniki działania algorytmów są znacznie gorsze niż dla nieskompresowanych zdjęć w formacie TIFF. Duży wpływ na jakość segmentacji ma także przestrzeń barw. Najlepsze wyniki uzyskuje się przy zastosowaniu kilku różnych przestrzeni barw, tak jak to zostało zrobione w algorytmie Tominaga. Jednakże czas wykonywania obliczeń w takim wypadku jest nieakceptowalnie długi. Algorytm watersheed najlepiej nadaje się do wykorzystania, jeżeli celem jest uzyskanie saliency map. Szybkość jego działania oraz otrzymywane wyniki spełniają zadania stawiane tego typu segmentacji.



**Tabela 4**  
Czasy przetwarzania obrazów przez poszczególne algorytmy

Oryginalny obraz	Watersheed	Quadtree	Tominaga
	75.9125 s	0.975447 s	830.025 s
	26.0794 s	0.969249 s	240.594 s
	15.3993 s	0.935809 s	899.737 s
	26.3286 s	0.951589 s	402.197 s
	30.9053 s	0.892231 s	326.524 s

## 5. Podsumowanie

W niniejszej publikacji porównano trzy algorytmy segmentacji obrazów kolorowych. Dwa z nich są autorskimi wersjami znanych uprzednio algorytmów, natomiast trzeci jest znany z literatury. Porównanie algorytmów dotyczyło czasu oraz jakości segmentacji.

Porównanie wykazało, że algorytmy różnią się między sobą istotnie. Najlepsze wizualnie wyniki daje algorytm Tominaga, jednakże czas analizy zdjęć jest nieakceptowanie duży. Pozostałe algorytmy dają „gorsze” wizualnie rezultaty, ale są od 10 do 100 razy szybsze. Wyniki działania algorytmów silnie zależy od źródła danych. Zdjęcia o pozornie takiej samej jakości i rozdzielczości segmentowane są w diametralnie inny sposób. Postulowane jest wypracowanie szeregu algorytmów segmentacji. Wśród tych wskaźników na pewno powinien być zawarty sposób kodowania kolorów.

## Literatura

- [1] Ozyildiz E., Krahnstover N., Sharma R., *Adaptive texture and color segmentation for tracking moving objects*. Pattern Recognition, 35, 2002, 2013–2029.
- [2] Greenspan H., Goldberger J., Eshet I., *Mixture model for face-color modeling and segmentation*. Pattern Recognition Letters, 22, 2001, 1525–1536.
- [3] Skarbek W., Koschan A., *Colour image segmentation a survey*. Technical Report, Tech. Univ. of Berlin, October 1994.
- [4] Cheng H.D., Jiang X.H., Sun Y., Wang J., *Color image segmentation: advances and prospects*. Pattern Recognition, 34, 2001, 2259–2281.
- [5] Pitas I., *Digital Image Processing Algorithms*. Prentice Hall, Hertfordshire, England, 1993.
- [6] Pratt W.K., *Digital Image Processing*. 2nd ed., Wiley, New York, 1991, 548–553.
- [7] Pal N.R., Pal S., *A review on image segmentation techniques*. Pattern Recognition, 26, 1993, 1277–1294.
- [8] Koshan A., Abidi M., *Digital color image processing*. 2008.
- [9] Sharma G., *Digital color imaging handbook*. 2003.