

New Framework of HSL System Based Color Clustering Algorithm

Vasile Pătrașcu

Tarom Information Technology, Otopeni, Romania

patrascu.v@gmail.com

Abstract

This paper proposes a construction of a new distance in the *HSL* color space by a combination of the distances of *HSL* scalar components and by using the fuzzy c-means framework to derive a fuzzy color clustering method. The main contribution consists in introducing two particular parameters to control the balance between the hue distance and the luminosity distance in the color Euclidean distance. The principal idea is to decrease the luminosity-distance influence and to increase the hue-distance influence when saturation values increases, and vice-versa, so that for instance, the colored pixels with similar hue, similar saturation and different luminosity values finally belong to the same cluster. These parameters are computed from the saturation values and for that reason, the proposed clustering method supplies good results also for gray-level pixels. Results are presented on some color images to illustrate the new approach and are compared to those obtained using the standard Euclidean *HSL* distance.

Keywords: Color distance, fuzzy color clustering and HSL perceptual color system.

1. Introduction

In many situations of image analysis and image processing, we are faced with two color comparison subject. We want to know how similar or how different two colors are. In order to do this comparison, we need to have a good coordinate system for color representation and also, we need to define an efficient inter-color distance measure in the considered system. In this paper we will consider the particular case of color representation by the perceptual system *HSL* (hue, saturation and luminosity) [2], [4]. The obtained degree of similarity or dissimilarity is dependent on the used inter-color distance [1], [3], [5]. In addition, the formulae used for saturation and luminosity calculation are very important. In the *HSL* color comparison procedure, the three color components have not the same importance. Thus, the most important is the hue, the saturation is the next and finally, the luminosity is less important. Because of this reason, many authors have proposed some color clustering methods that use the bi-dimensional space *HS*. On this way, they have neglected the third component, the luminosity. This method is very good for those images that are quite saturated. In the same time, this method has the disadvantage of supplying erroneous clusters for the less

saturated colors because it does not take into account the achromatic component, the luminosity. On the other side, when the luminosity is taken into account in the distance formula, there are arising situations when two colors are placed in different clusters because of the difference between their luminosities despite of their strong similarity in the chromatic space *HS*. In order to minimize the two disadvantages mentioned above, in this paper one proposes a new distance formula for the color space *HSL*. This distance is constructed as a quasi-linear combination of the three standard distances of *H, S, L* scalar components.

In the following sections, the paper is thus organized: section 2 presents the particular form of the system *HSL* used in this paper; section 3 presents the new color distance; section 4 presents the color clustering algorithm based on the proposed distance and related to the fuzzy c-means algorithm; section 5 presents some experimental results while section 6 outlines the conclusions.

2. The perceptual color system HSL

The most part of the color images are represented by the *RGB* color system. In this paper we suppose that $R, G, B \in [0, 1]$. Starting from the *RGB* system, it was defined other system for color representation. One of them is the *HSL* system where *H* is the hue, *S* is the saturation and *L* is the luminosity. The *HSL* system belongs to the perceptual system category because it is very close to the human color perception. There exist many formulae for *H, S, L* components calculation.

For the hue *H*, the most part of the definitions are closed to the following formula:

$$H = \text{atan2}\left(\frac{B - G}{\sqrt{2}}, \frac{2R - B - G}{\sqrt{6}}\right) \quad (1)$$

and $H \in (-\pi, \pi]$.

For the luminosity calculation, most frequently, one uses one of these two following formulae:

$$L = \frac{R + G + B}{3} \quad (2)$$

$$L = \frac{M + m}{2} \quad (3)$$

where $M = \max(R, G, B)$ and $m = \min(R, G, B)$.

In this paper we will use for luminosity and saturation calculation the formulae presented in [11]. In order to obtain the luminosity formula, one constructs a bifuzzy set [6] related to the color brightness. In the same time, its complement will be related to the color darkness. One consider two colors $Q_1, Q_2 \in [0,1]^3$, where $Q_1 = (R_1, G_1, B_1)$ and $Q_2 = (R_2, G_2, B_2)$. One uses for the space RGB the well known distance:

$$d(Q_1, Q_2) = \sqrt[n]{\frac{(R_1 - R_2)^n + (G_1 - G_2)^n + (B_1 - B_2)^n}{3}} \quad (4)$$

where $n \in (1, \infty)$.

We denote the white color $W = (1,1,1)$ and the black color $K = (0,0,0)$. One defines the membership and the non-membership functions by:

$$\mu(Q) = d(Q, K) \quad (5)$$

$$\nu(Q) = d(Q, W) \quad (6)$$

Thus, we constructed a bifuzzy set where the negation ν of μ does not verify the well-known equality $\nu = 1 - \mu$, that is happened in the particular case of fuzzy set. The color luminosity is described by the following bi-dimensional vector:

$$\lambda = (\mu, \nu) \quad (7)$$

From the triangle inequality it results:

$$d(Q, K) + d(Q, W) \geq d(K, W) \quad (8)$$

and

$$\mu + \nu \geq 1 \quad (9)$$

Thus, the obtained bifuzzy set is a paraconsistent one [6]. In literature, there exist many methods that transform a bifuzzy set into a fuzzy one [8], [9]. In this paper, we consider the following two formulae [10]:

$$L = \frac{\mu + 1 - \nu}{2} \quad (10)$$

$$L = \frac{\mu}{\mu + \nu} \quad (11)$$

For $n = 1$, the formula (9) becomes equality and the constructed bifuzzy set becomes a fuzzy set. From formulae (10) and (11) one obtains formula (2) as particular case. For $n = \infty$, from (10) it results formula (3) and from (11) it results the following formula for luminosity calculation:

$$L = \frac{M}{1 + M - m} \quad (12)$$

Regarding to the saturation calculation, many variants are defined by the distance between the values $M = \max(R, G, B)$ and $m = \min(R, G, B)$. Frequently are used the following two well-known formulae:

$$S = M - m \quad (13)$$

$$S = \frac{M - m}{1 + |M + m - 1|} \quad (14)$$

In the paper [7] one defined the following distance for the interval $[0,1]$.

$$D(x, y) = \frac{2|x - y|}{1 + |x - y| + |x + y - 1|} \quad (15)$$

In order to increase the distance value, one replaces the sum $|x - y| + |x + y - 1|$ with $\max(|x - y|, |x + y - 1|)$.

One obtains:

$$D(x, y) = \frac{2|x - y|}{1 + \max(|x - y|, |x + y - 1|)} \quad (16)$$

But $\max(|x - y|, |x + y - 1|) = |x - 0.5| + |y - 0.5|$ and finally one obtains the saturation calculation formula:

$$S = \frac{2 \cdot (M - m)}{1 + |M - 0.5| + |m - 0.5|} \quad (17)$$

Formula (17) supplies for saturation, values that are greater than the values supplied by formulae (13) and (14). Using a calculation formula that leads to saturation values increasing, one amplifies the role played by saturation among the HSL scalar components. Finally, from (12) and (17) it results that $L, S \in [0,1]$.

3. The new HSL color distance

In Cartesian coordinate system $(x, y, z) \in R^3$, for two vectors $v_1 = (x_1, y_1, z_1)$, $v_2 = (x_2, y_2, z_2)$ one defines the Euclidean distance by:

$$D^2(v_1, v_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \quad (19)$$

In order to obtain the variant of Euclidean distance for the cylindrical coordinate system one uses the following substitution:

$$x = \rho \cos(\varphi), \quad y = \rho \sin(\varphi), \quad z = l.$$

It results:

$$D^2(v_1, v_2) = 4\rho_1\rho_2 \cdot d^2(\varphi_1, \varphi_2) + d^2(\rho_1, \rho_2) + d^2(l_1, l_2) \quad (20)$$

where:

$$d^2(\varphi_1, \varphi_2) = \sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right)$$

$$d^2(\rho_1, \rho_2) = (\rho_1 - \rho_2)^2$$

$$d^2(l_1, l_2) = (l_1 - l_2)^2$$

The coordinate system HSL is a cylindrical one and for two colors $Q_1 = (H_1, S_1, L_1)$ and $Q_2 = (H_2, S_2, L_2)$, the Euclidean distance becomes [2]:

$$D_E^2(Q_1, Q_2) = 4S_1S_2 \cdot d^2(H_1, H_2) + d^2(S_1, S_2) + d^2(L_1, L_2) \quad (21)$$

Where:

$$d^2(H_1, H_2) = \sin^2\left(\frac{H_1 - H_2}{2}\right)$$

$$d^2(S_1, S_2) = (S_1 - S_2)^2$$

$$d^2(L_1, L_2) = (L_1 - L_2)^2$$

The color Euclidean distance (21) has three terms: the distance between hues, the distance between saturations and the distance between luminosities. In addition, the distance between hues is multiplied by a factor that depends on the color saturations. Thus, when the saturation values increase, the hues distance influence increases in framework of inter color distance (21). When the saturation values decrease, the hue distance influence decreases.

From here, the idea to multiply the luminosity distance with a similar factor comes up. This factor will have the following behavior: when the saturation values increase, the luminosity distance influence decreases and when the saturation values decreases, the luminosity distance influence increases. We can generalize (21) using two real and positive parameters α_H , α_L , by:

$$D_P^2(Q_1, Q_2) = \alpha_H \cdot d^2(H_1, H_2) + d^2(S_1, S_2) + \alpha_L \cdot d^2(L_1, L_2) \quad (22)$$

Thus, for two chromatic colors, the main components are given by the *HS* space, for two achromatic colors the main components are given by the *LS* space. We can see that, when a color is chromatic and the other is achromatic, the main component is given by the *S* space.

In [11] it was defined the hue and saturation relevancies by:

$$c(S) = \sin\left(\frac{\pi}{2} S\right) \quad (23)$$

$$a(S) = \cos\left(\frac{\pi}{2} S\right) \quad (24)$$

and a color similarity having the following structure:

$$\sigma(Q_1, Q_2) = \sigma(H_1, H_2) + \sigma(L_1, L_2) \quad (25)$$

where

$$\sigma(H_1, H_2) = c(S_1) \cdot c(S_2) \cdot \cos^2\left(\frac{H_1 - H_2}{2}\pi\right) \quad (26)$$

represents the hue similarity and

$$\sigma(L_1, L_2) = a(S_1) \cdot a(S_2) \cdot \cos^2\left(\frac{L_1 - L_2}{2}\pi\right) \quad (27)$$

represents the luminosity similarity.

In this paper, we will use for the hue and saturation relevancies the following functions:

$$c(S) = \sqrt{S} \quad (28)$$

$$a(S) = \sqrt{1 - S} \quad (29)$$

Finally, for the two multipliers it results the following particular functions:

$$\alpha_H = \sqrt{S_1} \sqrt{S_2} \quad (30)$$

$$\alpha_L = \sqrt{1 - S_1} \sqrt{1 - S_2} \quad (31)$$

We must underline that the distance defined by (22) has three terms, while the color similarity defined by (25) have only two terms. The main difference is the fact that in (22), in addition, it was taken into account the saturation distance.

4. The fuzzy color clustering algorithm

We consider n colors Q_1, Q_2, \dots, Q_n that must be separated into k clusters. Each cluster j is characterized by the membership coefficients $w_{j1}, w_{j2}, \dots, w_{jn}$ for the considered n colors and the cluster center defined by the color $q_j = (h_j, s_j, l_j)$. For color clustering we will construct an algorithm that is similar to the fuzzy c-means algorithm [1].

The membership function w_{ij} is calculated using formula (32), where the distance D_P is calculated with formula (22).

$$\begin{cases} \forall i \in [1, n] \\ \forall j \in [1, k] \end{cases},$$

$$w_{ij} = \frac{1}{1 + \sum_{\substack{m=1 \\ m \neq j}}^k \left(\frac{D_P(Q_i, q_j)}{D_P(Q_i, q_m)} \right)^{\frac{2}{\alpha-1}}} \quad (32)$$

The functions w_{ij} verify the condition of the partition of unity, namely:

$$\forall i \in [1, n], \quad w_{i1} + w_{i2} + \dots + w_{ik} = 1$$

The cluster center components (h_j, s_j, l_j) are calculated with formulae (33), (34) and (35).

$$\forall j \in [1, k],$$

$$h_j = \text{atan2} \left(\frac{\sum_{i=1}^n w_{ij}^\alpha C_i \cdot \sin(H_i)}{\sum_{i=1}^n w_{ij}^\alpha C_i}, \frac{\sum_{i=1}^n w_{ij}^\alpha C_i \cdot \cos(H_i)}{\sum_{i=1}^n w_{ij}^\alpha C_i} \right) \quad (33)$$

where $C_i = \sqrt{S_i}$.

$$\forall j \in [1, k], \quad l_j = \frac{\sum_{i=1}^n w_{ij}^\alpha A_i \cdot L_i}{\sum_{i=1}^n w_{ij}^\alpha A_i} \quad (34)$$

where $A_i = \sqrt{1 - S_i}$.

$$\forall j \in [1, k], \quad s_j = \frac{\sum_{i=1}^n w_{ij}^\alpha \cdot S_i}{\sum_{i=1}^n w_{ij}^\alpha} \quad (35)$$

where α is a fuzzification-defuzzification parameter and also, $\alpha \in (1, 1.5)$.

If α approaches 1, then the fuzzy algorithm approaches a crisp one.

5. Experimental results

The proposed algorithm was applied to images: “flower1” (fig. 1a), “flower2” (fig. 2a), “house” (fig. 3a), “parrots” (fig. 4a), “girl” (fig. 5a), “dog” (fig. 6a), “bird” (fig. 7a) and “horses” (fig. 8a). The clustered images obtained using the new distance can be seen in figures 1b, 2b, 3b, 4b, 5b, 6b, 7b, 8b and those obtained using the Euclidean distance can be seen in figures 1c, 2c, 3c, 4c, 5c, 6c, 7c, 8c. In the case of images “flower1”, “flower2” and “girl”, using the Euclidean distance it was obtained two clusters for background. These two clusters have the same hue but the luminosities are different (figs. 1c, 2c, 5c). Using the proposed new distance one obtained only one cluster for the entire background (figs. 1b, 2b, and 5b). For images “house” and “dog”, using the Euclidean distance, the white and bright blue colors were not separated (fig. 3c, 6c). In the case of image “parrots”, the yellow and red colors were not separated (fig. 4c) and the clustering is different from that obtained using the proposed distance (fig. 4b). For the image “bird” using the Euclidean distance, the orange and grey colors were not separated (fig. 7c) while for the image “horses” the white color was not separated (fig. 8c). The obtained results show the advantage of using this new inter-color distance measure in color clustering algorithms.

6. Conclusions

In this paper one presents a fuzzy c-means algorithm enhancement for the particular case of color clustering. It was used the perceptual color system *HSL* for color representation.

The main step is represented by definition of a new distance in the *HSL* color space. In this construction, there were used two multipliers that make the balance between the hue weight and luminosity weight in the framework of this three-term color distance. The obtained experimental results were compared with those

obtained by using the Euclidean distance. This comparison shows the effectiveness of the proposed clustering algorithm using.

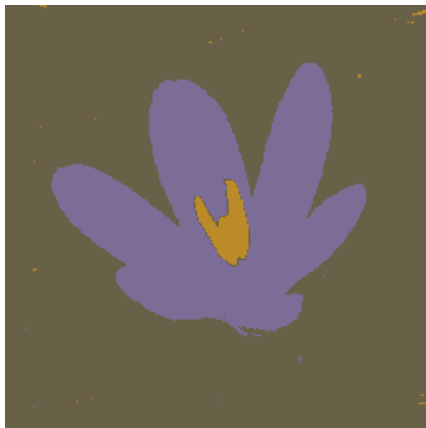
We can conclude that the new inter-color distance D_p (22), the two multipliers α_H (30) and α_L (31), the particular form of luminosity L (12) and saturation S (17) construct a new and useful framework for color clustering procedures.

References

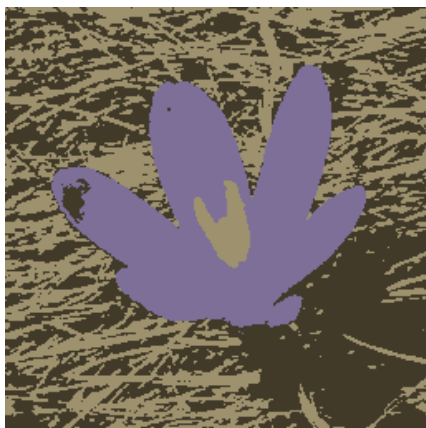
- [1] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Functions*, New York: Plenum Press, 1981.
- [2] R.C. Gonzalez, and R.E.Woods, *Digital Image Processing*. Prentice Hall. pp. 407–413, 2007.
- [3] Y.W. Lim, and S.U.Lee, On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recognition*, vol. 23, no.9, pp.935-952, 1990.
- [4] A.R. Smith, Color Gamut Transform Pairs, *Computer Graphics* 12 (3): 12, 1978.
- [5] M. Trivedi, M., and J.C. Bezdek, Low-level segmentation of aerial images with fuzzy clustering, *IEEE Trans. On Systems, Man, and Cybernetics*, vol. 16, no. 4, pp.589-598, 1986.
- [6] V. Pătraşcu, A New Penta-valued Logic Based Knowledge Representation, *Proceedings of IPMU 2008 Conference*, pp. 17-22, Malaga, Spain, June 22-27, 2008.
- [7] V. Pătraşcu, Cardinality and Entropy for Bifuzzy Sets, *Proceedings of IPMU 2010 Conference, Part I*, pp. 656-665, Springer-Verlag, Dortmund, Germany, June 28-July 02, 2010.
- [8] K. T. Atanassov, Remark on a Property of the Intuitionistic Fuzzy Interpretation Triangle, *Notes on Intuitionistic Fuzzy Sets* 8, pp. 34-36, 2002.
- [9] A.Q. Ansari, A. Shadab Siddiqui Javed and A. Alvi, “Mathematical Techniques To Convert Intuitionistic, Fuzzy Sets Into Fuzzy Sets”, Vol. 1, *Notes on Intuitionistic Fuzzy Sets*, pp. 13-17, 2004.
- [10] V. Pătraşcu, Fuzzy Membership Function Construction Based on Multi-Valued Evaluation, In vol. *Uncertainty, Modeling in Knowledge Engineering and Decision Making. Proceedings of the 10th International FLINS Conference*, World Scientific Press, pp. 756-761, Istanbul, Turkey, 2012.
- [11] V. Pătraşcu, New fuzzy color clustering algorithm based on *hsl* similarity, In *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress (IFSA 2009)*, Lisbon, Portugal, July 20-24, 2009.



(a)



(b)



(c)

Fig. 1: The image „flower1” (a) and its 3-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)

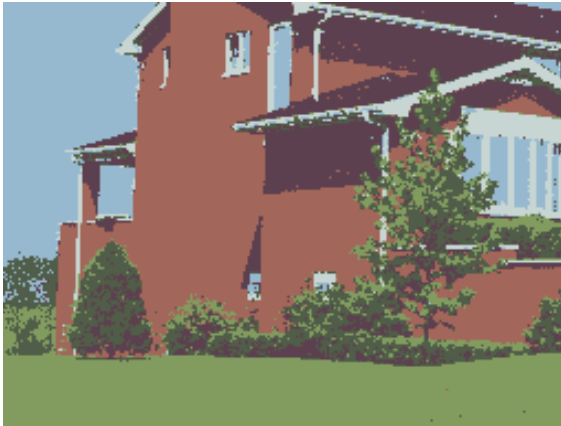


(c)

Fig. 2: The image „flower2” (a) and its 3-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)



(c)

Fig. 3: The image „house” (a) and its 6-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)



(c)

Fig. 4: The image „parrots” (a) and its 6-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)

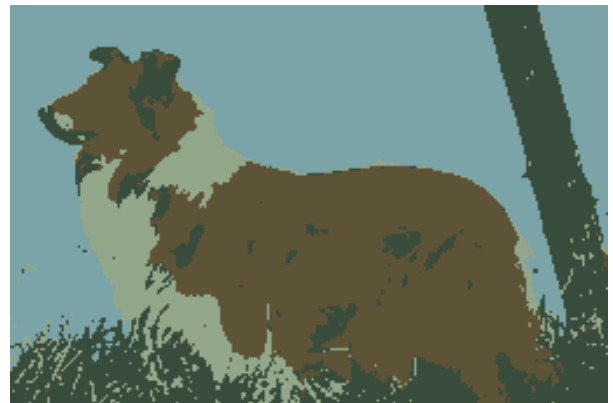


(c)

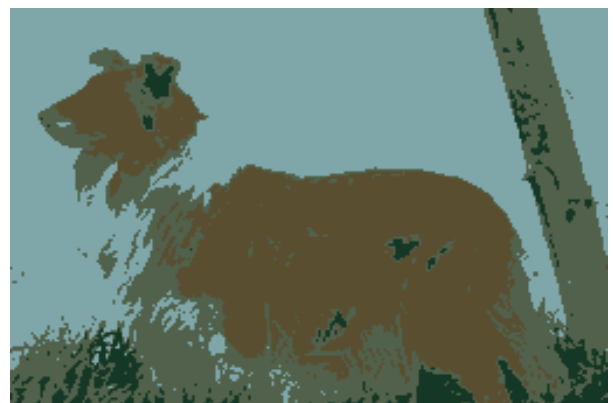
Fig. 5: The image „girl” (a) and its 8-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)



(c)

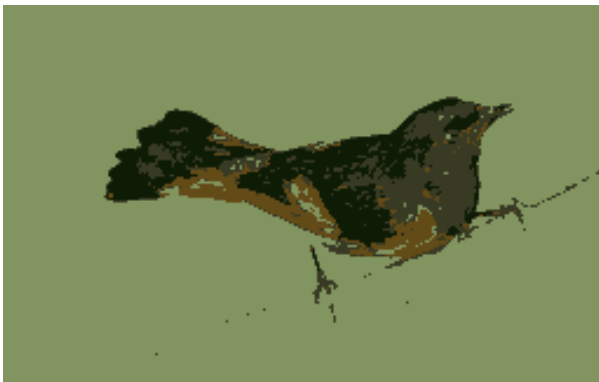
Fig. 6: The image „dog” (a) and its 4-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)



(c)

Fig. 7: The image „bird” (a) and its 4-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).



(a)



(b)



(c)

Fig. 8: The image „horses” (a) and its 5-color representation based on proposed HSL distance (b) and based on HSL Euclidean distance (c).