

A Segmentation System for Soccer Robot Based on Neural Networks

Carmelo Amoroso¹, Antonio Chella^{1,2}, Vito Morreale¹, and Pietro Storniolo³

¹ Dip. di Ingegneria Automatica e Informatica, Univ. of Palermo, Italy

² CERE-CNR, Palermo, Italy

³ Dip. di Ingegneria Elettrica, Univ. of Palermo, Italy

{amoroso,chella,vito,piero}@csai.unipa.it

Abstract. An innovative technique for segmentation of color images is proposed. The technique implements an approach based on thresholding of the hue histogram and a feed-forward neural network that learns to recognize the hue ranges of meaningful objects. A new function for detecting valleys of the histogram has been devised and tested. A novel blurring algorithm for noise reduction that works effectively when used over hue image has been employed. The reported experimental results show that the technique is reliable and robust even in presence of changing environmental conditions. Extended experimentation has been carried on the framework of the Robot Soccer World Cup Initiative (RoboCup).

1 Introduction

The task of identification of objects placed at different locations in gray level images is usually difficult unless uniform lighting conditions have been achieved. Uneven illumination leads to variations in light intensity reflected from objects. This is true in the RoboCup framework [10], where light conditions vary during the day and due to different game fields. Considering this problem, in [4] a method for calibrating lighting conditions in the scene is proposed, in which a brightness map of the area is employed to select areas for which different color calibrations are needed. But in general, it is necessary to resort to additional criteria, such as size or shape in order to identify objects of interest. With this respect, color becomes a powerful descriptor that greatly simplifies the object identification and extraction tasks.

Several authors analyzed the advantages of color image analysis with respect to gray level image analysis [15]. The most widely adopted approaches are the statistical methods. They are based on the assumption that an analytic expression for the distribution functions of the pattern classes is known, such as the Gaussian probability density function. This assumption may be unrealistic, especially in the case of images acquired by a vision system placed on a mobile robot, due to the presence of illumination changes, shadowing and obstacles in the robot working environment.

The starting point of our segmentation technique is the HSI color space and particularly the hue component. In the last RoboCup edition many teams employed intensity (or value), saturation and hue components of pixels in order to

label them. In [1] a real-time color detection system based on HSV information of pixels using designed LSI is described. Also the Italian robot Rullit [5] adopting an omnidirectional vision sensor, performs the color classification by using the HSV color space. Alternative color models have been used by other teams during RoboCup games. For example, in [6], authors adopt a particular decision tree using the rgG color space, where r and g are red and green normalized components and G is roughly related to brightness.

To accomplish the task of a robust segmentation of color images oriented to the vision system of a mobile robot, as in the RoboCup framework, we employ a novel approach based on a feed-forward neural network that learns to recognize the hue range of the objects of interest. Unlike statistical methods, a neural network does not require the exact knowledge of the statistical distribution of input items, as it generates an internal representation of input data distribution by a suitable learning process. Furthermore, a neural network deals effectively with noisy and partial data, as it is able to generalize to new and unexpected situations. Connectionist approach has often been used to perform the segmentation step in an image analysis process providing high performances [3].

Ranges to be recognized have been extracted by using an effective technique for thresholding based on analysis of the hue histogram. Thresholding has been used by the Agilo RoboCuppers Team also in its vision system in which the YUV color space is used to distinguish different color classes: particularly color clusters are found in the UV -plane and a color label is assigned to them according to their position in that plane.

Images acquired by optical or electronic means are likely to be degraded by imperfection of sensing mechanism, that occurs in the form of a noise. Several filters have been reported in the literature to remove noise from images. As our analysis is based on the analysis of the hue component of HSI color model, we propose a novel technique to remove noise from hue images that performs well when works on angular values.

The described approach has been tested in the framework of the Robot Soccer World Championship Initiative. In this framework, a mobile robot acquires the color images by a CCD camera fixed on its top (see Fig. 1).

The images are then processed by a PC on board of the robot. Although the color of the objects in the scene are clear and well defined, several disturbing conditions, such as shadowing, nonuniform lighting and light reflections arise during the RoboCup competition, that impose tight constraints to the segmentation step. Therefore different game fields may be characterized by similar but not equal color, e.g., the green color of the training field may not be exactly equal to the green color of the game field.

The presented technique is part of the vision system adopted by the soccer player robots of the Italian robot team ART (Azzurra Robot Team) [11], a robot team in "F2000" category (middle size league). The technique is also employed as a first step of a more complex vision system for an autonomous robot aimed at the description of static and dynamic scenes [7, 8].

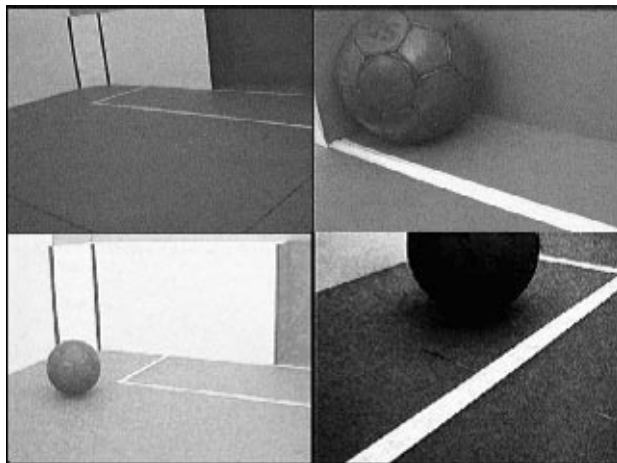


Fig. 1. Images of the adopted experimental framework. The soccer field is green, the walls and the field lines are near white, the goal area is blue and the ball is red.

2 Color imaging

Color images are widely used due to the superior characteristics over gray-level images. They contain object spectral information and provide valuable details for quick objects identification. Color information processing provides more reliable and accurate results for machine perception and scene analysis.

Andreadis and Tsalides [2] showed that the *HSI* color space compensates for illumination variations, hence contributing to simplified segmentation. According to the Phong shading model, hue has the property of multiplicative/scale invariance and additive/shift invariance [13]. According to these features, the hue parameter is our starting point for developing image processing algorithms based on color information. This choice allows us to work on a 1D space, rather than in a 2D (such as *IQ*, $a * b*$, $u * v*$) or a 3D (such as *RGB*, *CieXYZ*) color space. This simplifies the search for significant clusters that identify meaningful regions of pixels.

3 Noise removing by circular blurring

The purpose of pre-processing is to improve the image in ways that increase the chances for success of the subsequent processes. The alteration of the image may be generated in a number of ways, such as interference, noise, light reflection, etc. To maximize the noise attenuation is only one aspect of image pre-processing. On the other hand, it might happen that noise is attenuated so hardly that the useful image details are lost.

Many methods for scalar image processing have been developed for gray-level image processing. A direct application of such methods to the hue image can be

performed, but this generalization exhibits some drawbacks, the most important of which lies in not considering exact relationship between hue values. Hence, the development of new methods becomes necessary to realize efficient low-pass filtering of hue image. In this section, a new algorithm for hue image blurring and noise reduction is presented. This technique is based on mask processing.

Throughout this section let us denote the hue levels of pixels under the mask at any location by h_1, h_2, \dots, h_N (let N be the number of pixels within the mask) and suppose that hue values lie in the range $[0, L-1]$, due to sampling performed over this range. The only requirement on L is that it should be integer, positive, and finite.

Defining n as the largest integer value lower than k/L , that is

$$n = \left\lfloor \frac{k}{L} \right\rfloor, \quad (1)$$

we introduce the following *circular indexing function*

$$I_L(k) = k - nL \quad \forall k \in \mathbb{Z} \quad (2)$$

This operator allows us to perform any operation over the domain of the hue values considering automatically its circular nature, without changing the form of equations.

Given $I_L(k)$, the counterclockwise distance between h_i and h_j is defined as:

$$D(h_i, h_j) = I_L(h_j - h_i) \quad (3)$$

The most important properties of this function are:

$$D(h_i, h_i) = 0 ; \quad (4)$$

$$D(h_j, h_i) = L - D(h_i, h_j) . \quad (5)$$

Defining

$$d_i = \min_{\substack{1 \leq j \leq N \\ i \neq j}} \{D(h_i, h_j)\} \quad i = 1, 2, \dots, N, \quad (6)$$

the reference value of the set $\{h_1, h_2, \dots, h_N\}$ is the value B such that

$$B = h_b \quad (7)$$

where

$$b = \arg \max_{1 \leq i \leq N} \{d_i\} . \quad (8)$$

We can define a new sequence of relative hue values $\mathbf{H} = \{H_1, H_2, \dots, H_N\}$, where

$$H_i = D(B, h_i) \quad i = 1, 2, \dots, N . \quad (9)$$

Now, both median and linear filtering may be used to obtain the new relative hue level of the pixel located at the center of the considered mask. Median filter are best suited to double-exponential and impulsive noise but may not be effective for Gaussian noise, while linear filters are optimal in removing additive white noise (Gaussian or non-Gaussian) but will produce rather poor results in an impulsive noise environment.

Using linear filtering, the relative hue value of the pixel located at the center of the current mask is replaced by

$$m_{lin} = \mathbf{w} \cdot \mathbf{H} = \frac{1}{N} \sum_{i=0}^N w_i H_i \quad (10)$$

while the median of set \mathbf{H} is defined as the value m_{med} such that

$$m_{med} = \arg \min_{H_j \in \mathbf{H}} \left\{ \sum_{i=1}^N |H_i - H_j| \right\} . \quad (11)$$

The absolute hue level of the center pixel in the defined window is replaced by

$$M = I_L(m + B) , \quad (12)$$

where m is either m_{lin} or m_{med} according to the choice of the used filter.

4 Thresholding

We employed a thresholding technique to perform low-level image processing. According to the definition of the hue in the *HSI* model, the hue histogram is a periodic function in the same range of the hue.

Fig. 2 shows three “typical” hue histograms of color images of the chosen framework. It should be noted that the color modes (the peaks in the histograms) extend themselves within similar hue ranges. Each color mode corresponds to an object in the scene: the first one is related to the ball, the second one indicates the wall and the lines, the third one is related to the game field and the last one indicates the game goal. In the figure, the hue ranges corresponding to the color modes are highlighted.

In the literature several techniques for searching valleys and peaks of gray level histogram are presented [14, 9, 12, 16]. These methods fail when directly used for the analysis of the hue histogram, because the hue is periodic. For this reason we developed a special technique for histogram analysis, in order to find meaningful modes of the hue histogram.

Hue histogram is defined as follows:

$$q(k) = \frac{r(k)}{R \cdot S} \quad k = 0, 1, \dots, L - 1 \quad (13)$$

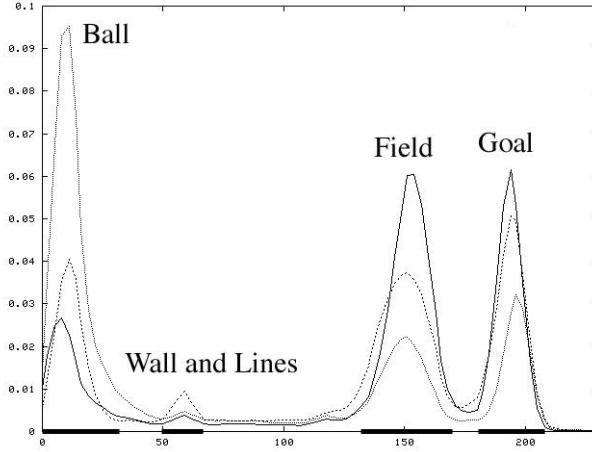


Fig. 2. Typical hue histograms related to images of the game field.

where $r(k)$ is the number of pixels with hue value equal to h_k , i.e., the k -th sampled hue value, and L is the number of hue samples. R and S are, respectively, width and height of the image.

To be able to deal correctly with hue histogram, for searching valleys and peaks, it is necessary to redefine the histogram function, according to the circular domain. Hence, starting from the histogram function $q(\cdot)$, it is possible to define a new function $p(\cdot)$, called *extended histogram function*, by means of functional composition with the function $I_L(\cdot)$ defined in (2)

$$p(k) = q(k) \circ I_L(k) = q[I_L(k)] \quad \forall k \in \mathbb{Z} . \quad (14)$$

To improve the performances of the following histogram analysis, the hue histogram is first smoothed to remove noise by means of:

$$\hat{p}(k) = p(k) * w_\sigma(k) = \sum_{i=-\sigma}^{\sigma} p(k-i)w_\sigma(i) \quad \forall k \in \mathbb{Z} \quad (15)$$

where $w_\sigma(\cdot)$ is a Gaussian smoother, defined as

$$w_\sigma(k) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{k^2}{2\sigma^2}\right) & |k| \leq \sigma \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in \mathbb{Z} \quad (16)$$

and σ is the standard deviation related to the width of the smoother.

It is therefore possible search the valley points in the histogram, according to its circular property using a *valley detection signal* $r(\cdot)$ generated from the extended histogram function by convoluting it with a *valley detection kernel* $v_P(\cdot)$:

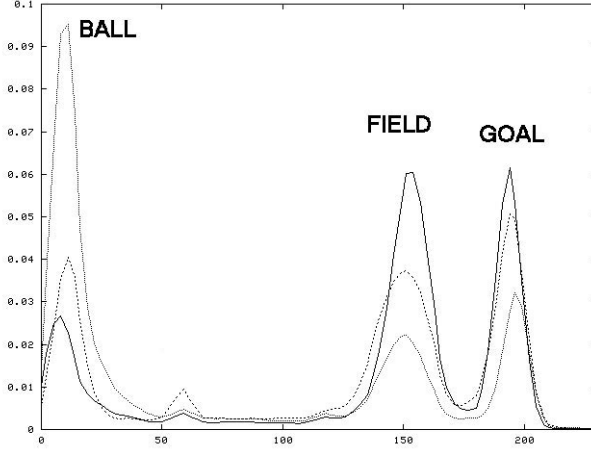


Fig. 3. An example of hue histogram and corresponding valley detection signal.

$$r(k) = \hat{p}(k) * v_P(k) = \sum_{i=-\lfloor P/2 \rfloor}^{\lfloor P/2 \rfloor} \hat{p}(k-i)v_P(i) \quad k = 0, 1, \dots, L-1 \quad (17)$$

where the valley detection kernel $v_P(\cdot)$ is defined as

$$v_P(k) = \begin{cases} \frac{k}{|k|} \frac{1}{\sqrt{\left(1 - \frac{|k|}{k_0}\right)^2 + 4\frac{k^2}{k_0^2}\zeta^2}} & |k| \leq \lfloor \frac{P}{2} \rfloor, k \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

P is the dimension of the kernel mask, ζ and k_0 are adjustable parameters (typically, $\zeta \in [0.1 - 2]$ and $k_0 \in [1 - 3]$). Valleys are placed on zero crossing to negative points of the $r(\cdot)$ signal.

In Fig. 3 an example of a hue histogram and the corresponding valley detection kernel is shown: the above mentioned zero crossing points are drawn as painted black dots. Only meaningful valleys are found, due to the robustness of the kernel with respect to small and irrelevant jitters.

5 The neural networks approach

As previously explained, the hue range and the shape of a meaningful color mode of the histogram are not clearly defined because they greatly depend on illumination conditions, shadows effect and so on.

We compared several techniques for classifying the modes extracted from the hue histogram. The statistical minimum distance classifier and the optimum

Bayes classifier have not been provide satisfactory results, due to the variability of the histogram. Because of this, we have been chosen a neural network for the classification step.

To code the hue ranges of the histogram to feed them as input to the neural network, we compared three different approaches: the *bit range*, the *Gaussian* and the *value* coding. Let L be the number of the considered hue sampled levels. Each level corresponds to one of the input units of the neural network. In the *bit range* coding, we set to 1 all the input units corresponding to the hue levels belonging to the considered range, and we set the other units to 0. In the *Gaussian* coding, we consider a Gaussian distribution of probability over the considered hue range. Then, we set to 1 the input units corresponding to the hue levels belonging to the range $[m - \sigma, m + \sigma]$, where m is the mean value and σ is the standard deviation of the hue of the considered color mode. The units outside the range are set to 0. In the *value* coding, we set the input units to the effective values of the color mode of the considered hue range; the units that do not belong to the considered hue range are set to 0.

The choice of the number L of sampled hue levels corresponding to the input nodes of the network is an important issue. We have chosen $L = 32$ according to the width and the resolution of the hue segments, corresponding to the classes of objects of interest. Experiments showed that for $L > 32$ the network does not improve its performances, while for $L < 32$ the hue segments overlap and the classification ratio degrades.

In the considered application, we recognize the following four classes of object: the *field*, the *wall and lines*, the *goal* and the *ball*. Hence, we chosen the number Out of the units in the output layer so that each node represents one of the above mentioned classes; i.e., $Out = 4$.

We chose the number Hid of hidden nodes by considering the variation of performances of the network by varying Hid . The number of hidden nodes was assumed to be the lowest one that achieves the minimum error on the test set. Experiments showed that the best performances were reached with $Hid = 16$.

In conclusions, we experimented the best performances with a feed-forward neural network with 32 input nodes, 16 hidden nodes and 4 output nodes. We chose the sigmoid activation function for every node. The learning rate was 0.5.

After the classification of the hue ranges performed by the neural network, the obtained classified ranges are used to label each pixel in the image.

6 Experimental results

In this section, the performances of our segmentation technique based on neural network are analyzed. To produce the training set that has trained the network, we used a sequence of images drawn by some films realized on a soccer field during testing. In these images, there were all of the objects of interest.

Image pre-processing has been made by using the technique described in the Sect. 3: experimental results showed that a linear mean filter performs better than a median filter.

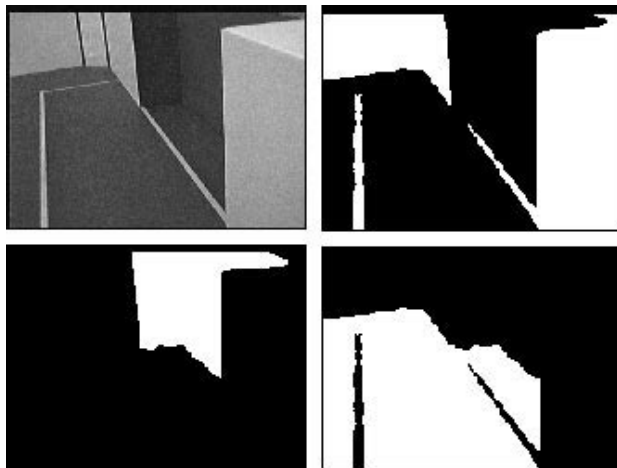


Fig. 4. An example of the operation of the thresholding (see text).

Thresholding of training and testing images has been made by using the efficient technique based on the convolution of the hue histogram with the suitable defined valley detection kernel, as described in Sect. 4.

A visual example of the obtained results are reported in Fig. 4 that shows the acquired image and *wall*, *field*, and *goal* images in which white patches represent selected regions.

To analyze in details the performances of the network, we accomplished three different tests: *Test A*, *Test B* and *Test C*, described below. In all of the tests, the training and the test sets are non-overlapping sets of patterns. We repeated the tests for all the three input coding described in the previous section.

In *Test A*, we analyzed the performances of the network by using training and testing images drawn by the same soccer field. To generate the training and the testing sets, we used 300 images acquired by the robot moving in the soccer field built in our laboratory. Different lighting conditions have been considered during acquisition in order to provide the network with a high degree of robustness.

We divided the training set into 5 mutually exclusive groups of 60 patterns. In each group, there were 15 patterns of each of the four classes. The network has been trained using 4 of these groups and tested using the fifth group. This train-and-test process has been repeated 5 times, using all groups of 60 test patterns. At the end, a mean misclassification error has been evaluated: the results are shown in the first row of Tab. 1. Results show the high reliability of the network when it works with pattern generated from images acquired in the same environment of the training phase.

To verify the capabilities of the network to work in different environmental conditions, during *Test B* we used the same training set of *Test A* but the test set was made up by images acquired by the robot during the previous RoboCup competition held in Paris in 1998. Therefore, environmental conditions, tonality

Table 1. Misclassifications of the neural network (see text).

	<i>Bit range</i>	<i>Gaussian</i>	<i>Value</i>
Test A (%)	0.33	0.67	0.67
Test B (%)	10.0	19.0	17.0
Test C (%)	6.17	7.83	9.33

of colors and illuminations in test set were very different with respect to the training set. The corresponding misclassification errors are shown in second row of Tab. 1. It should be noted that the error is maintained within acceptable limits. These results confirm that our method is satisfactory with respect to severe environmental changes.

To evaluate the performances of the network working on noisy data, during *Test C* we used the same training set of the previous tests, and a test set made up by noisy images. We produced the test set as follows: each pattern of the training set has been modified by operating a right shift and a left shift of the input values. Hence, for any training pattern, we produced two new test patterns. Altogether, 600 new patterns have been generated, on which the network has produced the results shown in third row of Tab. 1. Once again, experimental results highlight that the network is robust with respect to the noise.

By analyzing Tab. 1 with respect to the input coding, it should be noted that the *bit range* always performed better in terms of misclassification error with respect to the other coding. This is due to the fact that the *bit range* coding do not make any assumption on the form of the hue histogram. It therefore provides a low sensibility with respect to the form of the color modes. The result is an improvements in terms of robustness and adaptation to the environmental situations.

7 Conclusions

The problem of automatic segmentation of color images acquired by a camera placed on a mobile robot platform has been discussed. A new algorithm for noise removing in hue image has been presented. A special technique for dealing with multiple threshold selection from hue histograms has been developed. Here, a new valley detection signal is proposed. A feed-forward neural network has been employed to classify hue ranges corresponding to hue modes extracted from histograms.

Experimental results showed that the proposed approach is effective and robust with respect to changes of environmental conditions, such as shadows and illumination variations. Although this method has been applied in the RoboCup experimental framework, it is an effective intermediate level step in generic image interpretation tasks where the color is a meaningful descriptor.

We suppose that, training the network with a new training set according to the new working environment, this method is suitable to be used in different

environments. This fact provides our method a high degree of adaptability re-defining the correspondences between the chromatic components and objects by the neural network. Furthermore, we think that the approach based on to feed the representation of the most meaningful modes of the histogram as input to the neural network may be easily extended to N -dimensional analysis.

Acknowledgments

Authors would like to thank E. Ardizzone and S. Gaglio for stimulating discussions, and D. Nardi and all the persons involved in the ART (Azzurra Robot Team) for their encouragement. This work has been partially supported by CNR and MURST “Progetto Cofinanziato CERTAMEN” and ex 60%.

References

1. J. Akita. Real-time color detection system using custom lsi for high-speed machine vision. In *Proc. of the Third International Workshop on RoboCup*, pages 1–4. IJCAI Press, 1999.
2. I. Andreadis and P. Tsalides. Coloured object recognition using invariant spectral features. *Journal of Intelligent and Robotic Systems*, (13):93–106, 1995.
3. E. Ardizzone, A. Chella, and R. Pirrone. A neural based approach to image segmentation. In *Proc. of the International Conference on Artificial Neural Networks*, pages 1152–1156. Springer-Verlag, London, 1994.
4. J. Baltès. Practical camera and color calibration for large rooms. In *Proc. of the Third International Workshop on RoboCup*, pages 11–16. IJCAI Press, 1999.
5. A. Bonarini. The body, the mind or the eye, first? In *Proc. of the Third International Workshop on RoboCup*, pages 40–45. IJCAI Press, 1999.
6. J. Brusey and L. Padgham. Techniques for obtaining robust, real-time, colour-based vision for robotics. In *Proc. of the Third International Workshop on RoboCup*, pages 63–67. IJCAI Press, 1999.
7. A. Chella, M. Frixione, and S. Gaglio. A cognitive architecture for artificial vision. *Artificial Intelligence*, (89):73–111, 1997.
8. A. Chella, M. Frixione, and S. Gaglio. An architecture for autonomous agents exploiting conceptual representations. *Robotics and Autonomous Systems*, 25(3-4):231–240, 1998.
9. L. Hayat, M. Fleury, and A. F. Clark. Candidate functions for a parallel multi-level thresholding technique. *Graphical Models and Image Processing*, 58(4):360–381, July 1996.
10. H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. Robocup: A challenge problem for AI. *AI Magazine*, 18(1):73–85, 1997.
11. D. Nardi, G. Clemente, and E. Pagello. Art: Azzurra robot team. Technical report, DIS Universita’ di Roma *La Sapienza*, 1998.
12. N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
13. F. Perez and C. Koch. Toward color image segmentation in analog VLSI: Algorithm and hardware. *International Journal of Computer Vision*, 12(1):17–42, 1994.

14. E. Saber, A. M. Tekalp, R. Eschbach, and K. Knox. Automatic image annotation using adaptive color classification. *Computer Models and Image Processing*, 58(2):115–126, 1996.
15. G. Sharma and H. J. Trussel. Digital color imaging. *IEEE Transactions on Image Processing*, 6(7):901–932, July 1997.
16. J. Yen, F. Chang, and S. Chang. A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, 4(3):370–378, March 1995.