Multispectral Image Segmentation by Chromaticity Classification

JANDER MOREIRA[1]
LUCIANO DA FONTOURA COSTA[2]

[1]Universidade Federal de São Carlos
Departamento de Computação
GAPIS - Architecture, and Image & Signal Processing Research Group
jander@power.ufscar.br

[2]Universidade de São Paulo
IFSC - Instituto de Física de São Carlos
Cybernetic Vision Research Group
Luciano@ifqsc.sc.usp.br

Abstract. This paper describes a color segmentation technique, based on the k-nearest-neighbor classification scheme, which operates on a normalized version of the color image known as the chromaticity image. An investigation was carried out in order to evaluate how the classification behaves for different number of neighbors (k), for distinct window sizes (in which an average of a sample feature is taken), and for various numbers of samples per class. The results, which are experimentally assessed by comparing the obtained classifications with a standard reference (segmented by human), shows that the method provides good overall accuracy and robustness. The class space for the test image is also presented in graphical form.

INTRODUCTION

The human visual system is capable of performing a series of very complex tasks, such as discerning occluded objects even in the presence of geometric distortion and great amount of noise. It is known that these difficulties are not easily solved by artificial visual systems, in such a way that such systems must incorporate a high quantity of sophisticated techniques and procedures.

Biologically-inspired artificial vision systems, the so called cybernetic vision systems, are an ambitious goal to be achieved through computer based systems. This endeavor depends on a plenty of image analysis techniques, which have being developed during the past decades[1]. Nevertheless, the majority of related reported works have concentrated on gray level images, where the principal component is the scene luminance. Considerably fewer developments have addressed the segmentation of color images, which are known to encode substantially more information than their corresponding monochromatic versions. Colors incorporate a broad variety of additional information, including radiance and object spectral composition. Most of the expert performance of the human visual system is accounted for by the extensive use of color as basic attribute for segmentation, thus leading to better recognition and more accurate classification.

The human eyes are the first stage of the most complex and sophisticated pattern recognition system. Although the eyes do not make any severely complicated processing, early processing takes place on retina which includes data reduction (from millions of receptor to hundreds of optic fibers) with no substantial loss of information, edge detection on luminance data, and even color processing through the so called RGB photoreceptors. Later, into the *lateral geniculate nucleus* (LGN) and further to the visual cortex, the visual information undergoes a more elaborated processing. Contrast, orientation, motion, color opponency, and simultaneous contrast are analyzed in a combined fashion in such a way that information from some of these attributes can be used by other subsystem as aid to their own tasks. This sophisticated cooperative assemblage taking place at the visual cortex has been known to present a hierarchical modular integrated architecture.

Inspired on such a natural approach, an attempt is being made toward the development of a cybernetic vision system which brings together many of the characteristics previously discussed[2]. Named *Cyvis-1* (that stands for *Cybernetic Vision System*, first version), this system consists of collection of processing specializations, each performing dedicated tasks and sharing information with other modules. Each module is therefore able to interchange pieces of information with other modules and to consult not only its own database, but databases associated to other modules as well. The main tasks being considered include depth examination (through stereo analysis), edge and region detection, texture, and color labeling. These cooperative modules generate an intermediate description of the scene to be processed at a higher level, and also produce control information to correct and adjust camera operation (*Cyvis-1* incorporates active vision).

The present work deals with the color subsystem in *Cyvis-1*, which though currently not yet inter-

Anais do VIII SIBGRAPI (1995) 119-125

changing information with other specializations, has the difficult task of segmenting a scene using its individual channel data (e.g. the RGB channels). Therefore the attention here is directed to the treatment of images in a three-dimensional vector field. While gray scale images consist of a two-dimensional function with single valued data (the gray level itself), color images include triple-valued data (for instance the RGB or HSI components) — also called 3-D vector value.

During the past years some attempts have been made with the aim of using such additional color information to segment multispectral images. Different approaches have been tried for color segmentation, which can be classified in: (a) edge and line oriented segmentation; (b) region growing methods; (c) clustering; and (d) region splitting methods. The first type of procedures can work on the separate parts of the color image, and then join the partial result, which can be considered a difficult task, or address the color space as a whole, considering a two-dimensional vector space. In this way, a gradient on the vector space is defined in such a manner that edges are determined by high slope gradients[3,4]. Once the edges are extracted, they are postprocessed to create segments and objects, representing the objects in the scene. Edges can be also identified by a Hueckel's edged-based variant technique[5].

In another direction, an investigation on the significant features that could be extracted from color images was carried out, which determined, experimentally, that the features $(R+G+B)/3$, $R-B$, and $(2G-R-B)/2$ are, in this order, the most effective[6]. Such a color space is known as the Ohta's coordinate system. The approach to segmentation used in that work was based on a region splitting method with color feature calculations by the Karhunen-Loève transform, and thresholding. Another method by Ohlander, Price, and Reddy has a similar approach, using any color scheme representation (e.g. RGB, HSI, or YIQ), that makes use of significant thresholding to split regions[7]. The histogram for a region to be split is computed for all components. The one which has the "best" peaks is adopted, if actually any "best" peaks is obtained. The process is carried out for connected regions after a "region smoothing" is used to eliminate thin connections and small holes.

Most of clustering-type methods deals with fuzzy clustering for segmentation. The fuzzy *c*-means is largely employed in order to identify different classes, given that the number of existing clusters are known in most cases[8,9]. Trivedi and Bezdek[10] use a region splitting method with fuzzy *c*-means based clustering, in distinct resolution levels. A slightly different strategy was adopted by another method by Lim and Lee[11], where a coarse-to-fine technique is employed to reduce computational cost, and the histograms are analyzed by a tool based on the scale-space filter before thresholding. Many color spaces are addressed by this work, including the Ohta's coordinate system.

There are also approaches with use non-fuzzy, classical, clustering techniques, utilizing, for instance circular-cylindrical decision elements in the 1976 CIE (L*, a*, b*)-uniform color coordinate system[12]. Yet another interesting approach was proposed by Caelly and Reye[13], in which color, texture and shape are handled jointly. An alternative attempt was presented by Schlünzen *et al*[14], where a multi-layer neural network is used in the classification of multispectral remote sensing data. The work investigates the learning phase of the Perceptron Network and compare results to well-known classification methods.

The approach described here can be classified into the region-segmentation by classification group, since it uses a well-defined classifier in order to obtain the regions. The next section presents the developed classification technique and experimental results respectively to a real image, as well the discussion of these, are presented in the subsequent sections.

CLASSIFICATION METHOD

The first step in the segmentation process is to preprocess the image to reduce noise, which can be done simply by averaging each of the three bands according to a 3×3 window. This results in a perceptual color smoothing. Then, the image is normalized with respect to color, i.e. each component (say R, G, and B) is converted to the normalized chromaticity components, *r*, *g*, and *b*. Such a transformation is attained by

$$\begin{cases} r = \dfrac{R}{R+G+B} \\ g = \dfrac{G}{R+G+B} \\ b = \dfrac{B}{R+G+B} \end{cases} \tag{1}$$

This manipulation results in a floating point data image, whose values range from zero to 1.0 for each of the three bands, thus ensuring that $r+g+b=1.0$. The choromaticity is obtained in order to reduce the influence of the luminance contribution to the scene. The result, except for problems during image acquisition, is free from the brightness variations, thus yielding a representation in terms of "pure color" components (hue and saturation). It should be however observed that this normalization will not produce good results for the situation in which the components are too small. This shortcoming can be avoided by simple preprocessing, but such cases are rarely encountered.

After the pre-processing described above takes place, the region segmentation itself is carried out by classifying each image pixel accordingly to predefined classes, which are manually chosen, based on the visual

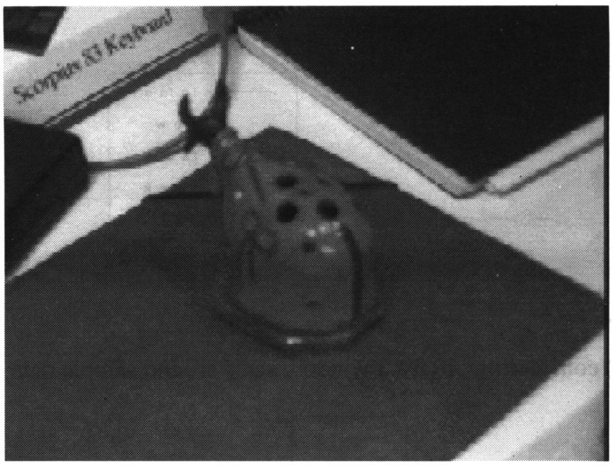

Figure 1. The 24-bit (true RGB) image used in the experiments. (a cores, na página 325)

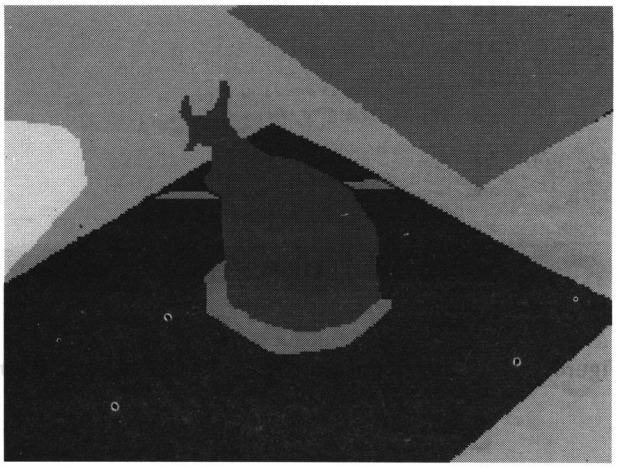

Figure 2. Segmented image obtained from human-assisted segmentation of image in Figure 1.

analysis to the image. The adopted classification is based on the *k*-nearest-neighbor technique[15]. Firstly, a supervised learning takes place in which sample data are selected from the image and used to define a database containing the classes to be used during the classification procedure. Each pixel to be classified is then compared to each of the elements in the database by means of a specific metric (the Euclidean distance in the three-dimensional chromaticity space has been adopted). The *k* nearest neighbors of each point to be classified are thus identified and the dominant class is taken as the outcome of the classification. Should two dominant classes be obtained (i.e. containing the same number of elements), the nearest neighbor is taken instead.

RESULTS AND DISCUSSION

The procedure described in the previous section was applied to the real color image shown in Figure 1, which was acquired by a standard camera. This is an indoor scene in a fluorescent light illuminated laboratory. Some color distortions can be observed, which can be attributed to the equipment and illumination conditions. These are more noticeable on the book (upper-left side) where an artificial texture was introduced, on the cow itself where the colors were separated in "stripes", and on the folder below the cow, at its right most extreme, where an artificial brown pigmentation can be noticed.

In order to evaluate the proposed segmentation method, a parameter was established to express the classification error. This error was obtained by comparing the results produced by the algorithm with a standard result obtained through human-assisted segmentation. The error metric, representing the percentage of mismatches occurred between the reference image and the algorithm-generated one, considering the spatial distribution, is named mismatch ratio. It should be observed that such a standard segmentation was made based not only on the color attribute itself, but also on high-level knowledge about the objects in the image (the yellow cow, the folder, a high color-textured book — in the right-upper corner, and part of a diskette box — in the left). The standard image is shown in Figure 2.

The sample data set was chosen arbitrarily over the image in order to define the associated classes. The number of points of each class is arbitrary as well. Considering that the color of the image can vary locally, an average of the color components was taken around a rectangular window. The size of this windows was also allowed to vary. Therefore, the tests carried out depend on three main parameters: the number of sample data (i.e. the points selected by the operator during the training stage), the size of the window around these points (from which the average local color is defined), and the number *k* of neighbors selected.

The tests were performed by comparing the misclassifications occurred between the segmented image and the reference segmented image. The obtained experimental data is expressed in terms of misclassification percentage, calculated by the ratio between the actual number of pixels belonging to a class and the number of pixels the algorithm classified into the respective classes.

Figure 3 shows the classification error respectively to distinct window sizes. While the error was as small as 3% for the folder (red), more than 70% was obtained for the box at the left hand side (almost-black object). The small area of the image covered by this latter object is probably responsible for this effect. The class table was the one for which the best performance was achieved, with the respective error remaining below 1% during most of the classification procedure. For the other classes, the errors range between 10% and 25%. The cow (yellow-orange object) classification can be considered good, even though the entire cow, which is

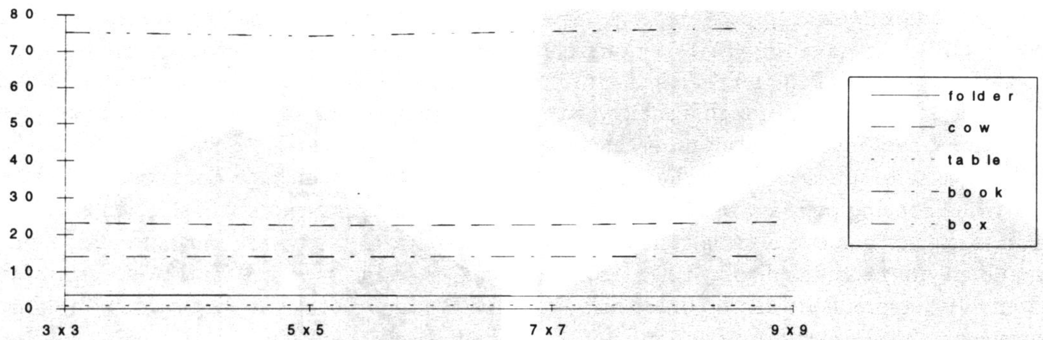

Figure 3. Error due to misclassifications (the mismatch error) considering different window sizes around sample data.

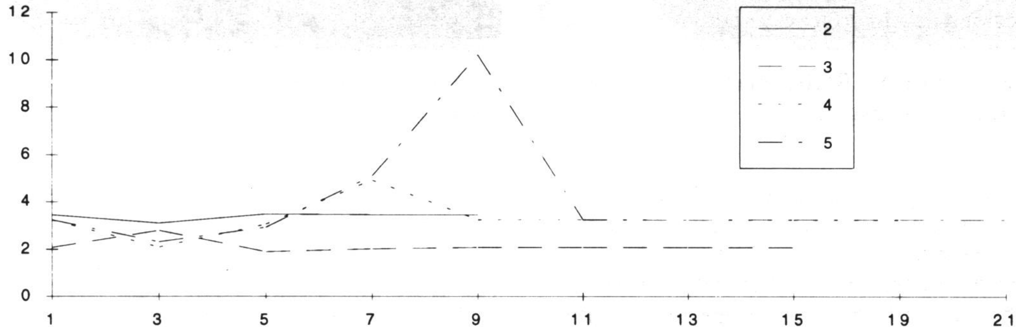

Figure 4. Mismatch ratio considering class folder, for different number of sample per class (2, 3, 4, and 5 samples) for distinct number of neighbors. The window used is 3×3.

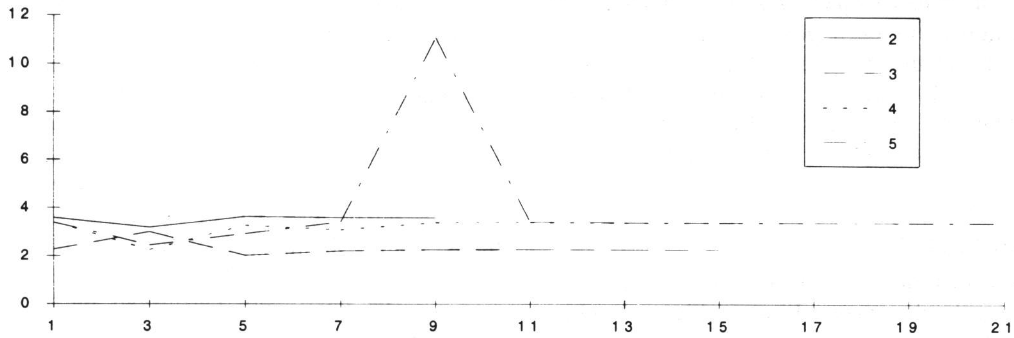

Figure 5. Same as Figure 4, using window 5×5.

known to include some small color details and reflections, had been assigned to a single class during the training stage. Conversely, the error respective to the classification of pixels belonging to the book can be attributed to the texture introduced at the image capture stage.

Another feature that is clear in Figure 3 regards the fact that the error varied little for the different sizes of windows that were tried. The tendency to increase the error above 9×9 windows can be nevertheless observed, which is expected given that, for large windows, the samples tend to cover areas not belonging to the selected object. Such an effect has been verified to be less marked for the situations where the objects are larger.

Other important aspect to be examined is how the number of neighbors, i.e. k, influence the classification. Figure 4 shows, with respect to the class folder, how this influence occurs. Each line in that figure represents a different number of samples per class, from 2 to 5, while always using a 3×3 window. The small influence of the number of neighbors onto the obtained results is clearly observed from Figure 4. A strange peak is also presented, which has been considered to be an

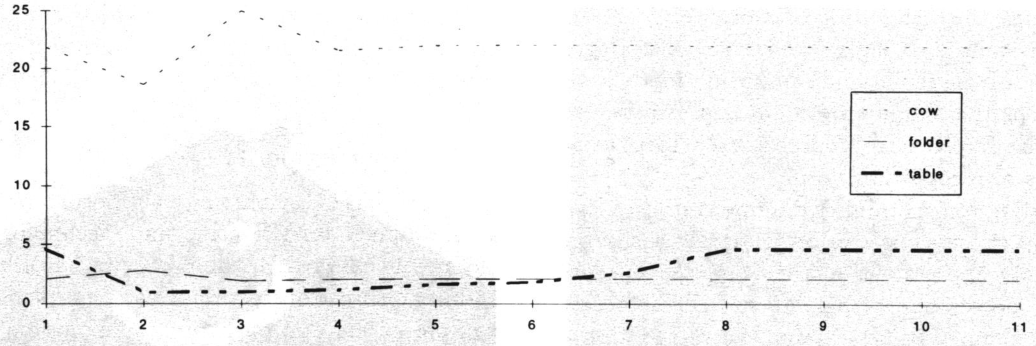

Figure 7. Comparing three different classes for distinct numbers of neighbors. The mismatch error is presented.

spurious effect due to the choice of the sample points. This same feature was also produced in the situation shown in Figure 5, which is respective to a 5×5 window.

Figure 4 and Figure 5 display other interesting characteristics which arise by influence of the number of samples. It is noticed that 3 samples give the best results, except when the classification uses 3 neighbors, in which case the best results are produced by using 4 and 5 samples per class.

Figure 7 presents the misclassification error in terms of k with respect to three classes (cow, table and folder), using 3 samples per class and a 3×3 window. For the class folder no significant disturbances can be observed and the error is almost constant, thus indicating little influence of k. This can be attributed to the fact that the folder in the image is quite uniform and contain a small degree of irregularities. Substantially higher errors were obtained for the class table, which indicates that this class presents some overlap with other class(es). The class cow presents an interesting behavior, varying for small values of k and becoming more stable for larger numbers of neighbors. Considering that the cow itself is not an homogeneous object, since its head has a distinct color and there are many different colors in small details, higher classification error values could be expected, which are more discernible at the lower values of k. Using more neighbors during the classification stage, the error has been minimized.

Another aspect that must be observed concerns how the colors are distributed in the "chromaticity" space, or the class space. Figure 6 presents projections for the "raw" color image onto the orthogonal planes RG, RB, and GB. While the colors were enhanced for visual purposes, efforts were made in order to preserve their original hue perception. It can be observed that the colors are distributed over the color space, tending to form irregular clusters, although similar colors seem to be close to each other, as expected. The short line segments that were obtained have been attributed to a digitization side-effect, since the original image shown if Figure 1 shows a noticeable "stripe" effect, probably due to illumination and sensor characteristics.

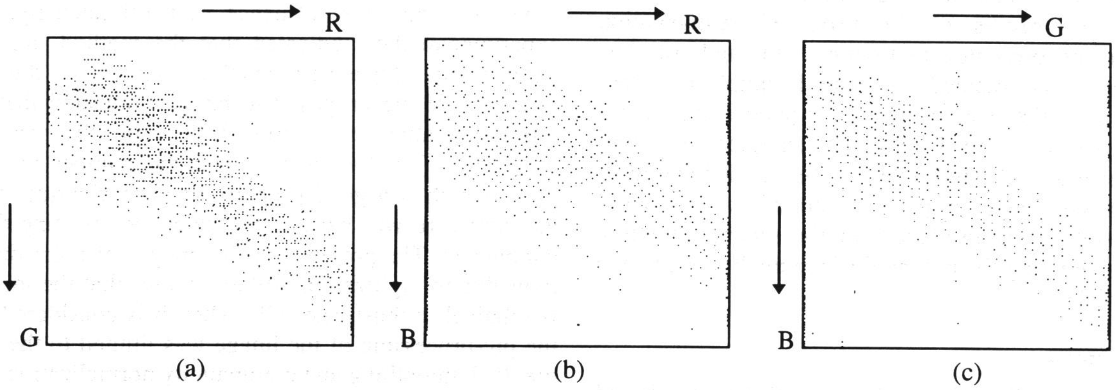

Figure 6. Projection of the color space on the orthogonal-planes: (a) RG-plane; (b) RB-plane; (c) GB-plane. The origin is at the upper-left corner for the three images.

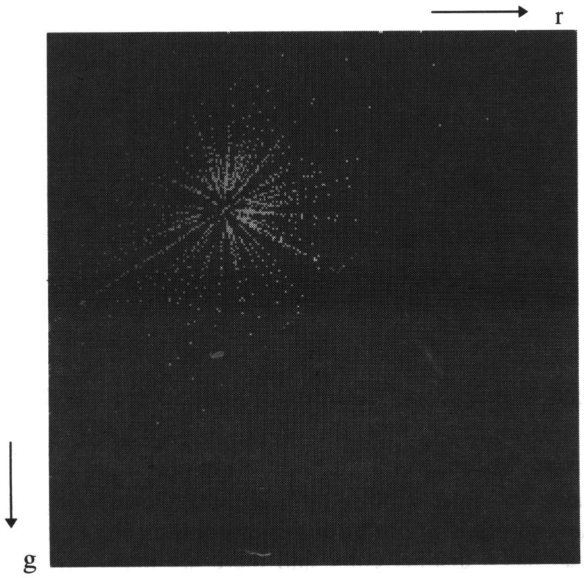

Figure 8. Projection of the chromaticity space on the orthogonal plane rg. (a cores, na página 325)

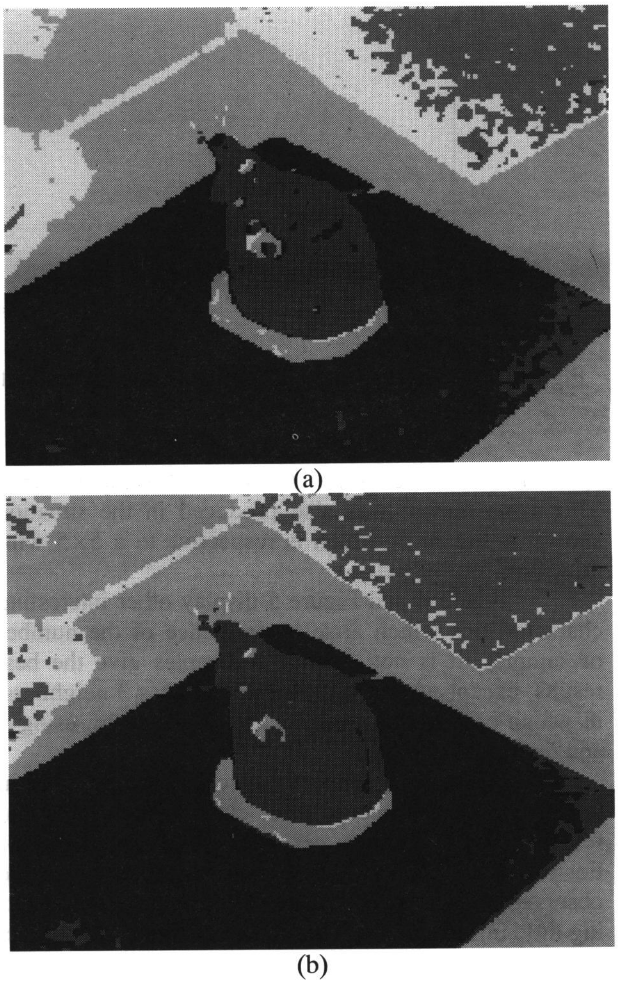

Figure 9. (a) Resulting segmentation using 1 neighbor, and 1 samples per class; (b) same as (a), using 15 neighbors, and 3 samples per class. Both samples were taken over a 3×3 window.

Figure 8 shows the projection of the normalized color space onto the planes rg, representing the "chromaticity space". The projection shows that similar colors tend to be close one another, forming the expected clusters. However, it should be noticed that such clusters are irregular and overlapped. Some misclassifications should clearly be attributed to these properties. Another characteristic observed is the overall line-shaped arrangement of some points, which has been attributed to discretization errors due to the finite representation of the values R, G and B (i.e. some points in the rgb space cannot be produced by the transformation operating over discrete values of R, G and B). It is also interesting to observed that the spherical shape of the normalized feature space has its center at the coordinates $(1/3, 1/3, 1/3)$, which is the achromatic point, where all components contribute equally.

Two results of the proposed segmentation technique are presented in Figure 9, (a) and (b). The first image was obtained by using 1 sample per class and $k = 1$. In the second image, the parameters were 3 samples per class and $k = 15$. The samples were considered over a 3×3 window. It can be verified that most of the classes were properly separated, with slightly better results being produced for the second situation, thus substantiating the potential of the method presented in this article.

CONCLUSIONS

This work describes the application of a region-based classification method as a means of segmenting images based on the color attribute. The investigation was performed assuming normalized colors (chromaticity) in order to reduce the influence of pixel intensity. The experimental data indicated that the method provides good results. An experimental assessment of the performance of the proposed technique indicated that the misclassification rate varies little with the involved parameters (window size, k, and the number of samples), which is a good property of the technique, since no extra care will have to be taken for selecting those parameters. The presence in the images of color distortions due to digitization artifacts shows that the method is relatively robust, specially when it is considered that the pre-processing of the image was limited to the simple 3×3 smoothing and chromaticity normalization.

However the presented methodology provides a robust and effective subsidy for semi-automated image analysis in terms of the color attribute, it would be interesting to extend the proposed technique to include self-organization of the classes, thus leading to automatic image segmentation. Investigation of alternative feature spaces other than the chromaticity one, such as HSI (hue, saturation, and intensity) or the Ohta's space are also being considered.

REFERENCES

[1]M. D. Levine. *Vision in man and machine*. McGraw Hill Publ. Co., New York, 1985.

[2]L. da F. Costa, V. O. Roda, and R. Köberle. A biologically-inspired system for visual pattern recognition. *IEEE International Symposium on Industrial Electronics*, Santiago-Chile, 1994.

[3]A. Cumani. Edge detection in multispectral images. *CVGIP Graph. Models and Im. Process.*, *53*:40–51, 1991.

[4]H.-C. Lee and D. R. Cok. Detecting boundaries in a vector field. IEEE *Trans. Sign. Process.*, *39*:1181–1194, 1991.

[5]R. Nevatia. A color edge detector and its use in scene segmentation. *IEEE Trans. Syst. Man Cybern.*, *7*:820–826, 1977.

[6]Y.-I. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, *13*:222–241, 1980.

[7]R. Ohlander, K. Price, and D. R. Reddy. Picture segmentation using a recursive region splitting method. *Comp. Graph. Image Process.*, *8*:313–333, 1978.

[8]T. L. Huntsberger, C. L. Jacobs, and R. L Cannon. Iterative fuzzy image segmentation. *Patt. Recogn.*, *18*:131–138, 1985.

[9]T. L. Huntsberger, C. Rangarajan, and S. N Jayaramamurthy. Representation of uncertanty in computer vision using fuzzy sets. *IEEE Trans. Computers*, *35*:145–156, 1986.

[10]M. M. Trivedi and J. C. Bezdek. Low-level segmentation of aerial images with fuzzy clustering. *IEEE Trans. Syst. Man Cybern*, *16*:589–598, 1986.

[11]Y. W. Lim, and S. U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy *c*-means techniques. *Patt. Recog.*, *23*:935–952, 1990.

[12]M. Celenk. A color clustering technique for image segmentation. *Comp. Vis. Graph. Image Process.*, *52*:145–170, 1990.

[13]T. Caelli and D. Reye. On the classificaton of image regions by colour, texture and shape. *Patt. Recog.*, *26*:461–470, 1993.

[14]E. T. M. Schlünzen *et al.* Classificação de dados multiespectrais utilizando redes neurais: a influência da amostragem no processo de treinamento. *Anais do Workshop sobre Visão Cibernética.* São Carlos, Agosto 1994.

[15]R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. John Wiley and Sons, Inc., New York, 1973.