REGION GROWING COLOUR IMAGE SEGMENTATION APPLIED TO FACE DETECTION

I. Grinias, Y. Mavrikakis, and G. Tziritas

Department of Computer Science, University of Crete
P.O. Box 2208, 71409 Heraklion, Greece
E-mails: {grinias,mavrik,tziritas}@csd.uoc.gr

ABSTRACT

A colour segmentation method is described which is based on a seeded region growing (SRG) segmentation algorithm. First the 3-D histogram in the *Lab* colour coordinate system is extracted. The histogram bins are then grouped using a hierarchical clustering algorithm. The number of classes can be automatically determined, while a *k-means* algorithm is used for determining the final list of dominant colours. The classification according to the dominant colours determines the initial seeds for a region growing segmentation algorithm. Segmentation is followed by a merging procedure which is based on colour and boundary information of regions. The proposed algorithm has been tested on facial images, for the needs of face detection.

1. INTRODUCTION

Image segmentation is an important and, at the same time, difficult task in image analysis. In this paper the analysis of in-door and out-door images for face detection is taken as the application domain for the segmentation algorithm. Colour image segmentation may be used in image/video post-production, and it could provide the required reduction of the spatio-temporal redundancy which characterizes the visual information. It also provides essential information for image description in visual indexing and retrieval by content.

Image segmentation techniques can be roughly grouped into five categories: (a) local filtering approaches such as the Canny edge detector [1]; (b) snakes [2]; (c) level sets [3] [4]; (d) region growing and merging techniques [5]; (e) global optimization approaches based on energy functions or Bayesian [6] and Minimum Description Length (MDL) [7] criteria. An interesting approach to automatic image segmentation using a combination of region-based and contour-based techniques is proposed by Zhu and Yuille [8].

Considering colour segmentation, the problem which arises from the light reflectance on polished surfaces (such as metal objects or the human skin) is that—in most of the perceptual colour systems—large lightness variations on object surfaces appear, which may not necessarily be accompanied by similar variations of object chrominance. On the other hand, in many cases "highlighted" regions appear with no colour information at all. In the current approaches to object colour modeling assumptions are made about the object material and the lightness conditions [9], parameters that are difficult to test when the goal is to segment outdoor images.

In what follows, a new colour segmentation method is presented, based on the Seeded Region Growing [5] segmentation algorithm. The original algorithm was initialized by user-selected region seeds, whereas in our approach the seeds are determined by an automatic clustering and classification algorithm described in the next section. In Section 3 we present the results obtained by the algorithm's application to facial images. The proposed method is tested in a face detection system [10], whose first stage is colour classification and/or segmentation. Other approaches using a skin colour classification and segmentation technique can be found in [11] and [12].

2. COLOUR SEGMENTATION USING SRG

2.1. Description

In this section, an automatic colour segmentation system is described. In Figure 1 the five stages of the algorithm are shown. These stages are summarized here and detailed in the following subsections. First the 3-D histogram is extracted, providing the initial bins for a hierarchical clustering algorithm. When the algorithm is applied to facial images and is used for face detection, a skin colour classification is initially applied, thus limiting the data space. The histogram bins are sequentially grouped in order to determine the number of distinct dominant colours present in the image. Finally the dominant colours are determined using the *k-means* algorithm with the Euclidean distance in the *Lab* colour space as dissimilarity measure. The most confident decisions are taken according to the distance from the dominant colours in order to label the initial region sets. These labels are propagated using a known seeded region growing algorithm. As an oversegmenation may result from the previous stages, the last step is the automatic merging of adjacent similar regions. The colour dissimilarity measures for both region growing and merging are defined in the *Lab* colour system, because of its uniformity which permits the effective use of the Euclidean distance between colours.

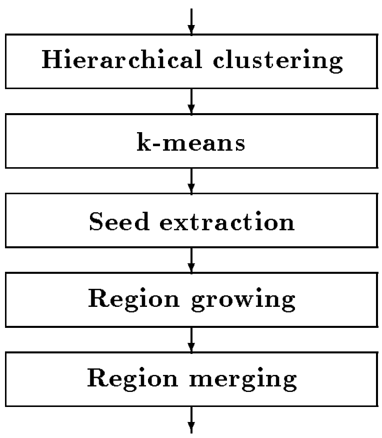

Fig. 1. The stages of the proposed segmentation algorithm.

2.2. Skin Colour Classification

In [10] an empirical technique for skin colour classification was presented. In [13] a mixture of Gaussians is proposed for modeling the skin colour space. In the current paper we present a skin colour detector based on a support vector machine [14].

The neural network was trained by a data set collected from a large base of images provided by the Institut National Audiovisuel (INA, France) and by Greek Radio-Television (ERT) for the needs of the DiVAN project [15]. We have trained and tested the support vector machine for five colour systems: *Luv, Lab, YCbCr, HSV, RGB*. The best results were obtained for the *Luv* and *Lab* colour systems, for which a Euclidean distance may be used for measuring colour dissimilarities. In Fig. 2 we present the classification result on the (a, b), (u, v), (C_b, C_r), and (G, B) planes. The skin colour domain is depicted in red.

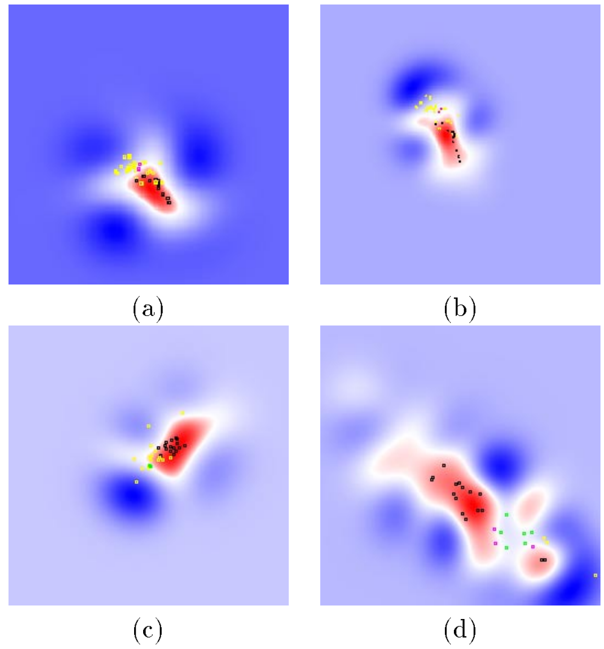

Fig. 2. The skin colour domain in the: (a) (a, b) plane for L about 70, (b) (u, v) plane for L about 70, (c) (C_b, C_r) plane for Y about 170, and (d) (G, B) plane for R about 200.

2.3. Dominant Colours

The histogram is computed in the *Lab* colour space. The cells of the histogram are defined using a constant step in the L component and a constant number of bins per L value. In the (a, b) plane polar coordinates are used for defining the quantization cells. Therefore for a given value of L, a constant step is used for the magnitude $\sqrt{a^2 + b^2}$ and for the hue component $\arctan \frac{b}{a}$. The histogram cells are represented by their centroid. The histogram bins are sequentially merged according to the minimum inertia increase criterion [16]. The inter-class distance is therefore determined by

$$\Delta_{i,j} = \frac{n_i n_j}{n_i + n_j} \left((L_i - L_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2 \right) \tag{1}$$

for two colour classes (C_i, C_j) represented by their respective mean colour vector and containing (n_i, n_j) image points. A threshold on the inter-class distance automatically determines the number of classes.

When the hierarchical classification algorithm terminates, the *k-means* algorithm is used for determining the final set of dominant colours. In this stage the Euclidean distance is used for measuring the dissimilarity of a colour vector from the associated dominant colours. Therefore the dominant colours are the centroids of the clusters they represent.

2.4. Automatic Seed Extraction

The dominant colours guides the initial pixel labeling. Decisions with high confidence are taken for points whose colour vector is similar to one of the dominant colours. For more confident and spatially homogenous classifications the neighborhood around a point may be tested. The goal of the initialization is the extraction of initial regions which:

- do not contain object boundaries,
- are large enough in order to efficiently compute the colour parameters of each one of them, and
- are homogeneous.

Once dominant colours have been extracted in the Lab colour space, the following steps are performed to obtain the initial sets:

- First, a binary decision array s is created. Pixels whose the colour vector is similar to one of the dominant colours are labeled and $s(p)$ is set to 1. Therefore the decision is based on the distance of the pixel's colour vector $(L(p), a(p), b(p))$ from the nearest dominant colour C to which p has been assigned:

$$\delta(p, C) = \min_i \frac{\sum_{q \in W(p)} d(q, C_i)}{|W(p)|} \tag{2}$$

  where $|W(p)|$ is the cardinality of a window $W(p)$ centered at p and $(L(q), a(q), b(q))$ is the colour vector of q with Euclidian distance from dominant colour C_i equal to $d(q, C_i)$. If the above distance is below a threshold t_a, we set $s(p) = 1$, otherwise $s(p) = 0$. In this manner, we obtain uniform colour regions which are not spatially connected.

- Then, the binary array s is scanned in order to extract the M_0 distinct regions resulting from the sets of connected points p with $s(p) = 1$.
- Finally regions of small size are rejected. The remaining M sets-regions $(R_1, R_2, \ldots, R_M)$ are the initial sets which are given as input to SRG.

When the above steps are applied to facial images in which a skin colour classifier has been already used, the first step may be bypassed. Otherwise the rectangular window W may be a square as large as 11×11 pixels and a typical value for the threshold is $t_a = 10$.

2.5. Region Growing

The initial regions $R_1, R_2, \ldots, R_M$ provided by the automatic seed extraction process are expanded using the SRG algorithm. The pixels belonging to each set are used for the estimation of the segmentation features. Then, yet unlabeled image pixels, which are neighbours of each set, are dynamically inserted into a queue sorted according to a dissimilarity measurement criterion $\delta(p; R_i)$ between each of these pixels p and the unique set R_i onto which p adjoins. If p adjoins onto more that one set, it is inserted into the set of minimal dissimilarity and also into the set B of boundary pixels. Pixels are labeled according to this queue and the queue is updated after each new label assignment. In each step of the algorithm exactly one pixel is labeled, the one currently at the head of the queue. The queue was implemented using AVL binary trees [17] in order to reduce the computational cost incurred when new adjacencies arise and pixels must be promoted based on their new dissimilarity.

The final segmentation result is the set map $i(p)$ of each pixel p to the set it belongs to. The dissimilarity $\delta(p; R_i)$ may be defined, as in the previous stages, by the Euclidean distance in the Lab colour space. For facial images, our experiments showed that a more suitable dissimilarity measure is defined by

$$\delta(p; R_i) = \sqrt{\alpha(a(p) - a_i)^2 + \beta(L(p) + b(p) - L_i - b_i)^2} \tag{3}$$

better taking into account the lighting conditions and shadow effects. The data analysis of colour vectors extracted from face regions led us to set typically $\alpha = 0.8$ and $\beta = 0.2$. If the dissimilarity measure exceeds an experimentally predefined threshold, which in our implementations is about 20–25, the concerned pixel is excluded from the waiting list for labeling.

2.6. Region Merging

As the number of regions at the end of the growing process is equal to the initial number of sets, an over-segmentation is very likely to result. The last step of the overall implementation is therefore the merging of similar adjacent regions. This is a difficult task, since it is known that regions of the same object may have different colour parameters, because of possible lightness variations over an object's surface due to its geometry, light reflectance or shadows. Indeed facial surfaces exhibit such variations. Therefore for the merging of face regions, we often use *a priori* colour information or other characteristics such as the shape or texture of objects (which are known for faces) [10].

We propose here a merging method based only on the colour of regions and on their boundaries. First, the adjacency map of regions R_i ($1 \leq i \leq M$) of the segmentation map is computed. Then, a distance measure $\Delta(R_i, R_j)$

$$\Delta(R_i, R_j) = \sqrt{\alpha(a_i - a_j)^2 + \beta(L_i - L_j + b_i - b_j)^2} \tag{4}$$

between the mean colour vectors of each pair of regions R_i, R_j is determined. Each step of the merging procedure involves the merging of the regions R_i, R_j with the minimum distance $\min_{(i,j)} \Delta$:

- if $\min_{(i,j)} \Delta(R_i, R_j) < t_m$, where t_m is a given small threshold (about 5),
  - the regions R_i, R_j are merged, if in addition $B_{ij}/(B_i + B_j) > t_b$, where B_i, B_j are the boundary lengths of regions R_i, R_j respectively, B_{ij} is the length of their common boundary and t_b is a given threshold, typically about 0.05. The above condition prevents the merging of regions whose common boundary is small compared to their total boundary. Then, the mean vector of the new set is computed and, considering it to replace one of the sets R_i, R_j in the adjacency map, the colour distance from its adjacent sets is estimated.
- otherwise the procedure terminates.

At the end of this procedure we test whether there exist regions which are almost completely encompassed by another region. A region is considered as included into another one, if $B_{ij}/\min(B_i, B_j) > t_c$, where the threshold t_c represents a large fraction (about 0.9). If the condition is satisfied the regions are merged.

3. EXPERIMENTAL RESULTS

The segmentation algorithm has been applied to facial images which have been acquired under various lighting conditions and over various complex and uncontrolled backgrounds. All images were extracted from video data provided by INA and ERT for the needs the DiVAN project [15].

In what follows, the results obtained using the automatic segmentation system on the face images are presented (Figure 3). The goal was to detect and locate faces using the proposed segmentation algorithm for uniform colour region extraction. After the skin colour classification stage, the dominant colours are extracted only from the possible face colours, and for this implementation their number is fixed to eight. These are shown in the top-left image. The pixels in black are classified as no-skin, and therefore the dominant colours are extracted only from the possible facial colours. In the top-right image the result of the region growing algorithm is shown, while the region merging result is given in the bottom-left image. We also give the face detection result obtained by applying the method described in [10], suitably modified to take into consideration the improvements due to the segmentation method proposed in this paper. The main stages following the colour segmentation consist of testing the shape and the image texture of the candidate face. The texture is measured by the wavelet packet coefficients as described in [10]. There is no constraint on the number of faces in the image, but the size of face frame is lower bounded by a rectangle of 80 pixels height and 48 pixels width.

Detection quality	Proposed method	Method in [10]	CMU method
Well framed	88	90	86
faces	92.63%	91.84%	96.63%
Moderate framed	7	8	3
faces	7.37%	8.16%	3.37%
Detected faces	95	98	89
	91.35%	94.23%	85.58%
False	9	6	15
dismissals	8.65%	5.77%	14.42%
False alarms	3	20	9

Table 1. Results of face detection on the test data set.

The test data set contains 100 images, most of them being extracted from advertissements, news and movies. The data set contains 104 faces with size above the minimal one. Ten images do not contain any face. About the half of the views are frontal face views, while the other are either semi-frontal, or viewed from one side, or tilted. Bearded people or with moustache is included, as well as people with glasses. There are some black people too, whose skin is correctly detected. In Table 1 we summarize the obtained results on face detection. In this Table the obtained results are compared to previous results of our team [10] and to results of the CMU face detector [18].

Compared to the results in [10] we can observe that there exist an improvement on the rate of false alarms, while the detection rate remains high. Almost all detected frames are accurately framed. Compared to the CMU face detector, the algorithm presented in this paper performs better in both false alarm and detection

Fig. 3. Results on three facial images. For each of them: top-left, dominant colours; top-right, region growing algorithm segmentation result; bottom-left, region merging final segmentation; and bottom-right, face detection result.

Fig. 4. Results with detection errors. For each of them: top, original; medium, region growing algorithm segmentation result; and bottom, face detection result. Illustartion of: (a) bad location of face frame, (b) false alarm, (c) and (d) missed faces.

rate, while the CMU face detector is sligtly better on the accuracy of face localization. The main weakness of our approach, the rate of missed faces (9 over 104), is due to the lighting conditions, when shadows and highlights are strongly present. Two such cases are shown in Fig. 4 (c) and (d). In addition in Fig. 4 (a) is illustrated the case of a badly located face frame, due to the beard and to the lighting conditions. A case of false detection is also shown in Fig. 4 (b).

4. SUMMARY

We proposed a skin colour classifier based on a support vector machine. We presented an extension of the SRG algorithm in automatic colour image segmentation. In all stages of the segmentation method the *Lab* colour system was used. SRG's seed extraction is based on the dominant colours of the image, which are extracted by a sequential hierarchical clustering algorithm followed by the *k-means* algorithm. We have also proposed a region merging technique for confronting some drawbacks, such as oversegmentation. The algorithm has been applied to face detection and localization with very good results.

Acknowledgment: This work has been funded in part by the Greek PENED "MPEG-4 Authoring Tools" project.

5. REFERENCES

[1] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Intern. J. of Computer Vision*, vol. 4, pp. 7–38, 1990.

[3] J.A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Applied Mathematics*, vol. 93, pp. 1591–1595, Feb. 1996.

[4] E. Sifakis, C. Garcia, and G. Tziritas, "Bayesian level sets for image segmentation," *J. of Visual Communication and Image Representation*, 2001 (to appear).

[5] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641–647, June 1994.

[6] D. Geman, S. Geman, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 609–628, 1990.

[7] Y.G. Leclerc, "Constructing simple stable descriptors for image partitioning," *Intern. J. of Computer Vision*, vol. 3, pp. 73–102, 1989.

[8] S.C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 884–900, Sept. 1996.

[9] G.J. Klinker, S.A. Shafer, and T. Kanade, "A physical approach to color image understanding," *Intern. J. of Computer Vision*, vol. 4, pp. 7–38, 1990.

[10] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. on Multimedia*, vol. 1, pp. 264–277, Sept. 1999.

[11] Y. Wang and B. Yuan, "A novel approach for human face detection from color images under complex background," *Pattern Recognition*, vol. 34, pp. 1983–1992, 2001.

[12] H. Wu, Q. Chen, and M. Yashida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 557–563, June 1999.

[13] H. Greenspan, J. Goldberger, and I. Eshet, "Mixture model for face-color modeling and segmentation," *Pattern Recognition Letters*, vol. 22, pp. 1525–1536, 2001.

[14] V. Vapnik, "An overview of statistical learning theory," *IEEE Trans. on Neural Networks*, vol. 10, pp. 988–999, Sept. 1999.

[15] P. Bouthemy, C. Garcia, R. Ronfard, G. Tziritas, E. Veneau, and D. Zugah, "Scene segmentation and image feature extraction in the divan video indexing and retrieval architecture," *Proc. of Visual Conf.*, 1999.

[16] R.C. Dubes and A.K. Jain, *Algorithms for clustering data*, Prentice-Hall, 1988.

[17] R. Kruse, *Data Structures and program design*, Prentice-Hall, 1987.

[18] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 23–38, Jan. 1998.