COLOR-BASED RETRIEVAL OF FACIAL IMAGES

Yannis Avrithis, Nicolas Tsapatsoulis and Stefanos Kollias

Department of Electrical and Computer Engineering
National Technical University of Athens
Heroon Polytechniou 9, 157 73 Zographou, GREECE
Tel: +301 7723038; Fax: +301 7722492
e-mail: {iavr,ntsap}@image.ntua.gr

ABSTRACT

Content-based retrieval from image databases attracts increasing interest the last few years. On the other hand several recent works on face detection based on the chrominance components of the color space have been presented in the literature showing promising results. In this work we combine color segmentation techniques and color based face detection in an efficient way for the purpose of facial image retrieving. In particular, images stored in a multimedia database are analyzed using the M-RSST segmentation algorithm and segment features including average color components, size, location, shape and texture are extracted for several image resolutions. An adaptive two-dimensional Gaussian density function is then employed for modeling skin-tone chrominance color component distribution and detecting image segments that probably correspond to human faces. This information is combined with object shape characteristics so that robust face detection is achieved. Based on the above, a query by example framework is proposed, supporting a highly interactive, configurable and flexible content-based retrieval system for human faces. Experimental results have shown that the proposed implementation combines efficiency, robustness and speed, and could be extended to generic visual information retrieval or video databases.

1 INTRODUCTION

The interest in multimedia applications has increased rapidly during the last years, leading to a growing demand for efficient storage, management and browsing in multimedia databases. New tools for summarization, content-based query, indexing and retrieval capabilities have received considerable attention especially for browsing digital video databases, due to the huge amount of information involved. Such tools are of major importance in the context of the emerging MPEG-4 and MPEG-7 multimedia standards [8].

Several frameworks have been proposed in the recent literature for content-based retrieval from image or video databases [1][10], and a lot of prototype systems have emerged, providing content-based image query and retrieval capabilities. Some of these systems, including for example VIRAGE, QBIC, Photobook and VisualSEEk, have already been implemented and are now in the stage of evaluation or commercial exploitation. In most cases, content information is handled by video object modeling and segmentation, and subsequent extraction of object attributes including color, motion, texture, shape as well as spatial and temporal relation between objects [2][3].

However, it has been often pointed out that existing systems lack the ability to extract and retrieve semantic information, which is naturally due to the fact that such capabilities always require a priori knowledge and can only be achieved in the context of specific applications. Examples of such applications with increasing interest include retrieval of images containing human faces as well as subsequent face detection, segmentation or recognition. Several approaches have been proposed for this purpose, including face registration using detection of the principal facial features (eyes, nose and mouth), template matching as well as color thresholding using the chrominance components of color space and a suitable distribution of the skin color [4][11]. The latter has received considerable attention as it combines very fast implementation with promising results.

In the context of this paper, we combine color segmentation techniques and color based face detection in an efficient way for the purpose of facial image retrieving. Images and video stored in a multimedia database are analyzed using the M-RSST segmentation algorithm and segment features including average color components, size, location, shape and texture are extracted for several image resolutions. An adaptive two-dimensional Gaussian density function is then employed for modeling skin-tone chrominance color component distribution and detecting image segments that probably correspond to human faces. This information is combined with shape characteristics extracted from object contours so that robust face detection is achieved, independent of image resolution, illumination, scaling and rotation. Finally, a query by example framework is proposed, supporting a highly interactive, configurable and flexible content-based retrieval system for human faces.

2 A WORKING SCENARIO

In the proposed framework, a user presents to the system a color facial image as an example for retrieving facial images. The system performs color segmentation, using the M-RSST algorithm described in Section 3, and presents the results to the user who is able to select the face segment that he requires to retrieve. The mean chrominance components of the selected face segment are computed and used to adapt a predefined probabilistic model that describes the chrominance distribution of the face skin, as explained in Section 4.

In the image database, apart from the raw image data indexing information is also stored, including the number of color segments, the average chrominance components of each segment, the percentage of the whole image area that they cover, as well as an indexed field that describes their contour shape and particularly their similarity to an ellipsis, since face shape is supposed to be elliptical. During the searching procedure image segments whose chrominance components match the adapted probabilistic model, giving a high probability of being face segments, are assessed against their shape to verify that they are not unrelated objects with color similar to that of the skin. Images that contain at least one face segment with a high probability are presented to the user. Ranking of the retrieved images can be performed in several ways and is discussed in Section 5.

3 COLOR SEGMENTATION

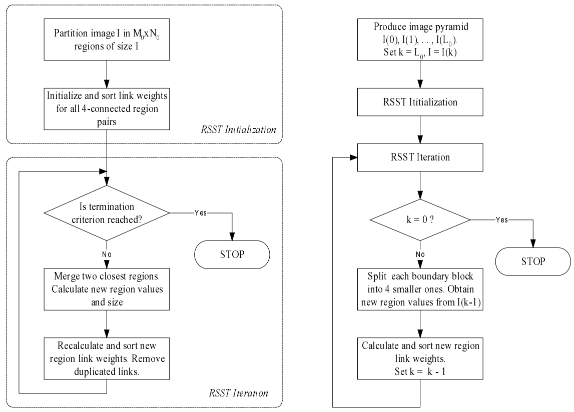
The Multiresolution Recursive Shortest Spanning Tree (M-RSST) algorithm, first introduced in [3], is our basis for color segmentation. The M-RSST is an efficient, multiresolution implementation of the conventional RSST [6] algorithm, which is considered as one of the most powerful tools for image segmentation, compared to other techniques such as color clustering, pyramidal region growing and morphological watershed [7].

The flowchart of the RSST algorithm is depicted in Figure 1(a). Initially an image I of size M_0xN_0 pixels, is partitioned into M_0xN_0 regions (segments) of size 1 pixel and links are generated for all 4-connected region pairs. Each link is assigned a weight equal to the distance between the two respective regions, which is in general defined as the Euclidean distance between the average color components of the two regions, using a bias for merging small regions. Using, for example, the YCrCb color space, a distance measure between two adjacent regions X and Y is defined as

$$d(X,Y) = \| \mathbf{C}_X - \mathbf{C}_Y \| \frac{A_X A_Y}{A_X + A_Y} \quad (1)$$

where $\mathbf{C}_X$ = [Y_X, Cr_X, Cb_X]^T contains the average color components of region X and A_X is the number of pixels within the region. All link weights are then sorted in ascending order, so that the least weighed link corresponds to the two closest regions. The iteration phase of the RSST is then initiated, where neighboring regions are recursively merged by applying the following actions in each iteration: *(i)* the two closest regions are merged and the new region color components and size are calculated, *(ii)* the new region link weights from all neighboring regions are recalculated and sorted, and *(iii)* any duplicated links are removed. The iteration terminates when either the total number of regions or the minimum link weight (distance) reaches a target value (threshold). A distance threshold is in general preferable since it provides a result that is independent of the image content.

The execution time of the RSST is heavily dependent upon the choice of the sorting algorithm, which is certainly a bottleneck of the algorithm. For this reason, M-RSST approach is employed, which recursively applies the RSST algorithm on images of increasing resolution, as depicted in the flowchart of Figure 1(b). Initially a multiresolution decomposition of image I is performed with a lowest resolution level of L_0 so that a hierarchy of frames $I(0)=I$, $I(1),\ldots,I(L_0)$ is constructed, forming a truncated image pyramid, with each layer having a quarter of the pixels of the layer below. The RSST initialization takes place for the lowest resolution image $I(L_0)$ and then an iteration begins, involving the following steps: *(i)* regions are recursively merged using the RSST iteration phase, *(ii)* each boundary pixel of all resulting regions is split into four new regions, whose color components are obtained from the image of the next higher resolution level, *(iii)* the new link weights are calculated and sorted. This "split-merge" procedure is repeated until the highest resolution image $I(0)$ is reached.



(a) (b)

Figure 1: Flowchart of the M-RSST segmentation algorithm.

At each resolution level of the M-RSST, the total number of segments for the next iteration is considerably reduced compared to the initial number of segments of the conventional RSST algorithm at the same level, yielding much faster execution. The execution times of both algorithms are compared in [3], where it is shown that the M-RSST is approximately 400 times faster than the RSST for a typical image size of 720x576 pixels.

Also, it is observed that very small segments cannot be found by the algorithm at the initial (lowest) resolution, and, since no segments are created or destroyed at each iteration, these segments are also eliminated from all resolution levels. For example, even if the target number of segments for an image containing a human face is high enough, some facial details cannot produce separate segments and a single segment is retained for the whole face. Such filtering according to object size is desirable in the context of face detection, since it achieves a high level of video content representation.

4 AN ADAPTIVE PROBABILISTIC MODEL FOR SKIN –TONE COLOR DISTRIBUTION

It is stated in some classic studies [5][9] that skin-tone colors are spread over a small area of the *Cr-Cb* chrominance plane of the [*Y, Cr, Cb*] color model. This fact has been successfully used in some recent studies for face detection in color images and video sequences [4][11]. In our work, we approximated skin-tone color distribution using a two-dimensional Gaussian density function. Assuming that the mean vector $\boldsymbol{\mu}_0$ and the covariance matrix $\mathbf{C}$ are robustly estimated, the likelihood of an input pattern $\mathbf{x}$ is given by:

$$P(\mathbf{x} \mid \boldsymbol{\mu}_0, \mathbf{C}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_0)^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}_0)\}}{(2\pi)^{\frac{k}{2}} \cdot |\mathbf{C}|^{\frac{1}{2}}} \quad (2)$$

where k = 2 is the number of chrominance components. For the estimation of $\boldsymbol{\mu}_0$ and $\mathbf{C}$ we used as training data facial pixels of different races, obtained from regular TV clips, color images and personal video cameras.

Color segments represented by their mean chrominance components can be assessed using the above model to give a probability of being skin segments. Quantities like the elliptical form of the segment can be exploited to separate the face and body segments.

Although the proposed Gaussian model is rather efficient for the classification of segments as skin and non-skin ones, it is much more valuable if it is used in an adapted form. The face segment selected by the user can be exploited for the re-estimation of the model's parameters and especially the mean vector $\boldsymbol{\mu}_0$. In particular, we re-estimate the $\boldsymbol{\mu}_0$ according to the following equation:

$$\hat{\boldsymbol{\mu}}_0 = m \cdot \boldsymbol{\mu}_0 + (1-m) \cdot \boldsymbol{\mu} \quad (3)$$

where $\boldsymbol{\mu}$ is the vector of the mean chrominance components of the selected face segment and m is a memory tuning constant. In a similar way we can also re-estimate the covariance matrix $\mathbf{C}$. However, re-estimation of $\mathbf{C}$ has been left out since it increases the computational complexity without a corresponding improvement of the system's performance.

The memory constant m is a critical parameter for the model. As it is stated in Section 5, appropriate choice of m can be used to interpret the queries and produce meaningful and satisfactory for the user results. Small value of m can be used when the user looks for faces similar to the one he/she presents to the system while a higher value can be utilized when the user queries the database for facial images in general.

Since objects unrelated to human faces but whose color chrominance components are still similar to those of the skin might be present in an image, we also provide a description of the object contour shape. In particular, the similarity of object shape to an ellipsis is calculated, since face shape is supposed to be elliptical. Thus, image segments whose chrominance components match the adapted probabilistic model are assessed against their shape to verify that they correspond to human faces.

5 RETRIEVAL RESULT RANKING

The proposed system gives the user the ability to select the criterion that is to be used for ranking of the retrieved images. There are three basic alternatives:

Similarity with the presented face segment. The memory constant m of the probabilistic model is kept small and therefore the model is strongly adapted to the presented face segment. In addition, the retrieved images are ranked with respect to their segment probabilities. Since in some retrieval results there are more than one probable face segments, retrieval results are represented by the segment with the highest probability.

Facial scale. In this case the user is interested in retrieving facial images that contain a face with scale similar to that of the selected face segment. The system keeps a high value for the memory constant m of the probabilistic model and image segments whose probability exceeds a given threshold can be selected. However, these segments are ranked according to the percentage of the image area they cover.

Number of face Segments. The user is interested in retrieving faces with a specific number of face segments. The system keeps a high value for the memory constant m and images that contain the required number of face segments are retrieved. Ranking of the retrieved images is performed according to their representative segment probability.

6 EXPERIMENTAL RESULTS

Some preliminary results are presented in this Section. The database that has been used in our experiments was created in the framework of project PHYSTA of the Training Mobility and Research Program of the European Community. 200 images have been selected from various video-clips recorded from BBC's broadcasted program, of which 156 contain at least one face. For experimental purposes we stored at least two instances of the same face in the database.

Figure 2: An image presented to the system.

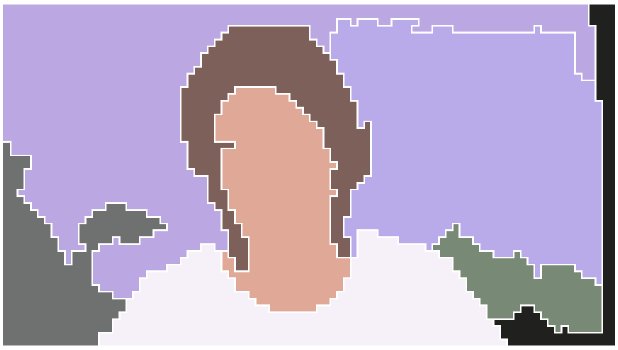


Figure 3: Color segmentation of the presented image.

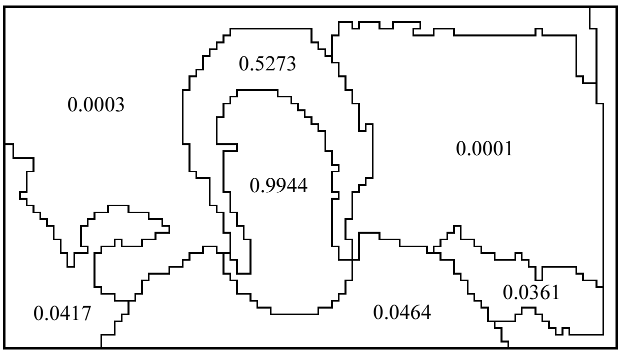


Figure 4: Face probabilities of the color segments.

Figure 2 illustrates an image that a user presents to the system in a query by example framework. Figure 3 illustrates the color segmentation of the presented image, while Figure 4 presents the probabilities of the various segments of being face segments. Figure 5 depicts similarity-based retrieval. A pretty small memory constant (*m=0.3*) has been chosen and the 7 best matches are presented. Below of each image, the corresponding probability of containing a face segment is illustrated. Figure 6 and 7 present scale based retrieval and retrieval based on the number of face segments respectively.

Figure 5: Similarity based retrieval.

Figure 6: Facial scale based retrieval.

Figure 7: Retrieval based on required face segments.

7 CONCLUSION

Color segmentation has proved a powerful tool for object extraction from images, especially in the case of human faces that are usually characterized by uniform color. The M-RSST algorithm is able eliminate facial details and provide a single object for each face. Moreover, chrominance components provided with a probabilistic model can be used in an efficient way for retrieving facial images from image databases. The interactive form of the proposed system adapts the model to the needs of the user and consequently leads to much more meaningful retrieval results.

8 ACKNOWLEDGMENTS

This work is funded by the project PHYSTA of the Training Mobility and Research Program of the European Community. The authors are within the team of the project, where speech and psychological cues are also used for emotion classification. We would also like to thank the BBC for allowing us to use video sequences recorded from its broadcasted program.

9 REFERENCES

[1] D. Androutsos, K. N. Plataniotis and A. N. Venetsanopoulos, "Extraction of Detailed Image Regions for Content-Based Image Retrieval," *Proceedings of IEEE ICASSP*, Seattle WA, USA, May 1998.

[2] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel and T. Sikora, "Image Sequence Analysis for Emerging Interactive Multimedia Services - The European Cost 211 Framework," IEEE Trans. Circuits and Systems for Video Technology, Vol. 8, No. 7, pp. 802- 813, Nov. 1998.

[3] Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," *Computer Vision and Image Understanding*, Vol 75, No. 1/2, pp. 3-24, July/August 1999.

[4] C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," *IEEE Trans. on Multimedia*, vol. 1, no. 3, pp.264-277, September 1999.

[5] L. A. Harwood, "A Chrominance Demodulator IC with Dynamic Flesh Correction," *IEEE Trans. Consumer Electron.*, Vol. CE-22, pp. 111-117, Feb 1976

[6] O. J. Morris, M. J. Lee and A. G. Constantinides, "Graph Theory for Image Analysis: an Approach based on the Shortest Spanning Tree," *IEE Proceedings*, Vol. 133, pp.146-152, April 1986.

[7] P. J. Mulroy, "Video Content Extraction: Review of Current Automatic Segmentation Algorithms," *Proceedings of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS)*, Louvain-la-Neuve, Belgium, June 1997.

[8] Y. Rui, T. S. Huang and S.-F. Chang, "Digital Image/Video Library and MPEG-7: Standardization and Research Issues," *Proc. of IEEE ICASSP*, pp. 3785-3788, Seattle, WA, May 1998.

[9] T. Rzeszewski, "A Novel Automatic Hue Control System," *IEEE Trans. Consumer Electron.*, Vol. CE-21, pp. 155-162, May 1975

[10] Special Issue on Segmentation, Description and Retrieval of Video Content, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, 1998.

[11] H. Wang and S. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video Sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, special issue on Multimedia Systems and Technologies, August 1997.