A SIMPLE AND EFFICIENT FACE DETECTION ALGORITHM FOR VIDEO DATABASE APPLICATIONS

Alberto Albiol†, *Luis Torres*‡[*]

Politechnic University of Valencia, Spain †
e-mail: alalbiol@dcom.upv.es
Politechnic University of Catalonia, Spain ‡
e-mail: luis@gps.tsc.upc.es

Charles A. Bouman, Edward J. Delp

Purdue University, USA
email: {bouman,ace}@ecn.purdue.edu

ABSTRACT

The objective of this work is to provide a simple and yet efficient tool to detect human faces in video sequences. This information can be very useful for many applications such as video indexing and video browsing. In particular the paper will focus on the significant improvements made to our face detection algorithm presented in [1]. Specifically, a novel approach to retrieve skin-like homogeneous regions will be presented, which will be later used to retrieve face images. Good results have been obtained for a large variety of video sequences.

1. INTRODUCTION

An increasing amount of audio-visual material is becoming available in digital form in more and more places around the world. With the increasing availability of potentially interesting material, the problem of identifying and indexing multimedia information is becoming more difficult. The new standard MPEG-7 [2] will provide a standardized description of multimedia content that can be used in image and video databases. However, it is very important to note that the tools needed to access the video information will not be part of the standard. This means that there will be a continuous need to provide new video analysis tools once the MPEG-7 standard is accepted. These tools will help the user to identify and locate video content, as a first step towards their description

The objective of this work is to provide a simple and yet efficient tool to detect human faces in the context of video sequences. This information can be very useful for many applications such as video indexing and video browsing. Face recognition applications have been also proposed combined with face detection [3]. In particular we will focus on the improvements made to our face detection algorithm presented in [1], specifically we present a novel approach to retrieve skin-like homogeneous regions, which will be later used to retrieve face areas.

This new approach speeds up significally the process of face detection while keeping the performance of the system, and making possible to use the algorithm in every single frame of the sequence in a reasonable amount of time. This opens the door to exploit the temporal redundancy present in video sequences in order to reduce the error rate.

We are well aware that much more information is present in video sequences such as closed captions, audio and motion, which should be also studied and exploited for video indexing applications. However, our objective in this paper is to concentrate on the image information and, at a later stage, to combine it with other information sources.

The next section describes the proposed face detection system and section 3 will present some results and conclusions.

2. FACE DETECTION SYSTEM

Our approach for face detection is designed to be robust to variations that can occur in face illumination, shape, color, pose, and orientation. To achieve this goal, our method integrates information regarding face color, shape, position and texture to identify the regions which are most likely to contain a face.

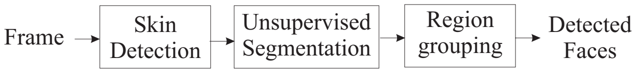


Fig. 1. Processing steps used to detect faces.

[*] This work was partially supported by the grants TIC 98-0442 and TIC 98-0335 of the Spanish Goverment

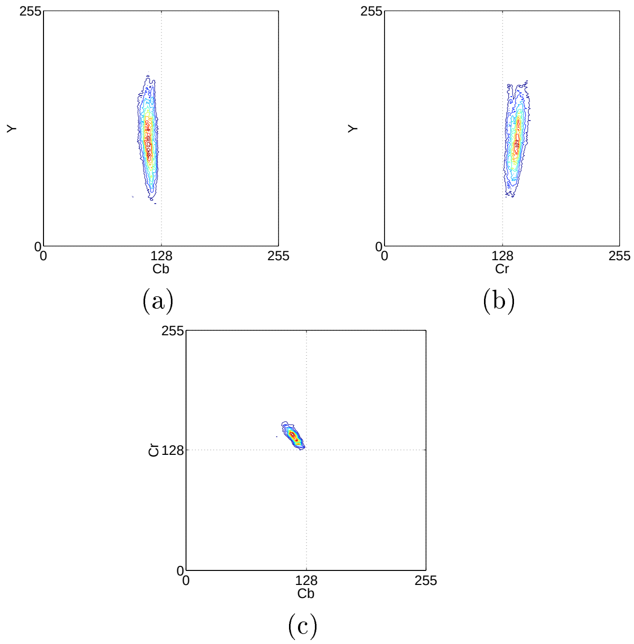


Fig. 2. Projections of the histogram of a representative set of skin pixels in the YC_bC_r color space.

Figure 1 shows the block diagram of the proposed algorithm. The basic blocks do not differ much from our initial proposal presented in [1]. In that proposal, a Gaussian mixture distribution was used to model skin pixels and then a multiscale segmentation algorithm (SMAP) was used to detect the skin pixels. Once the pixels of interest were identified, unsupervised segmentation was used to separate these pixels into smaller regions which were homogeneous in color. This is important because the skin detection will produce non-homogeneous regions often containing more than a single object. The EM algorithm was used in [1] to cluster the skin detected pixels in the color space using a Multivariate Gaussian Mixture Distribution. The unsupervised segmentation usually further partitioned the skin detected areas into smaller homogeneus regions making necessary the use of a region merging stage to extract the faces. The next subsections will describe in detail the changes made in these basic blocks which improve the overall performance of the whole system.

2.1. Skin detection

The first step, skin detection, is used to segment regions of the image which potentially correspond to face regions based on pixel color. Under normal illumination conditions skin colors fall into a small region of the color space and it is possible to use this information to classify each pixel of the image as *skin-like* or *non skin-like* [4, 5, 6, 7]. The projections of the 3-D histogram of a representative set of manually extracted skin pixels are plotted in the Figure 2. Figures 2.a and 2.b show that the luminance is uncorrelated respect to the C_bC_r components and Fig. 2.c shows that C_bC_r are highly correlated and define a small cluster on the C_bC_r plane.

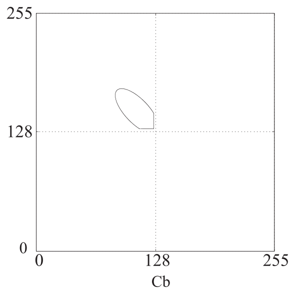


Fig. 3. Region bounding the skin-like colors on the C_bC_r color space.

Then, a pixel will be labeled as skin-like if its crominance vector falls into the region plotted in the Figure 3 and its luminance value is within the interval $45 < Y < 235$. These values were chosen empirically to reduce the miss detection rate since it is impossible to reduce the false alarm rate produced by skin-like colored background objects. The negative effect of this false alarm should be solved with additional processing [7]. The most important advantage of this simple algorithm is that the skin detection can be implemented with a LUT what makes the algorithm extremely fast.

Figure 6.a shows a challenging example, where the background is formed by many skin-like colored objects. The result of the skin detection is presented in Fig. 6.b, where the non skin-like pixels are drawn in black. We can see that skin pixels are detected however many skin-like objects are also selected. Further processing is necessary to alleviate this problem.

2.2. Unsupervised segmentation

Once the pixels of interest are identified, unsupervised segmentation is used to separate these pixels into smaller regions which are homogeneous in color. We present a novel approach for the unsupervised segmentation stage using the watershed algorithm [8] to cluster the skin detected pixels in the color space. To that end, once the skin-like pixels are detected, a 2D histogram in the C_b-C_r color space is constructed. Then, this histogram is treated as a gray-scale image and the watershed segmentation algorithm is applied on the histogram. The markers used for the watershed algorithm

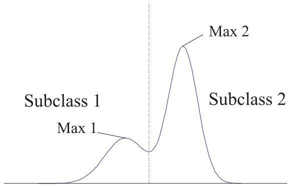


Fig. 4. Watershed algorithm can be used to find the support regions of two mixed classes.

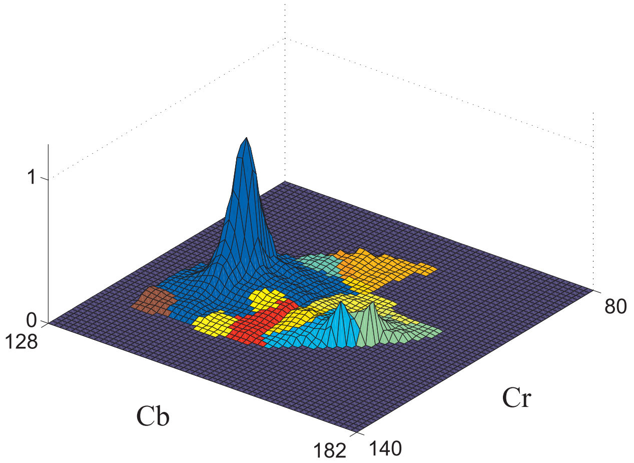


Fig. 5. Histogram of the skin detected pixels on fig. 6.a and the clusters found using the watershed algorithm.

are set to be all the local maxima in the histogram. The histogram is previously smoothed with a 3 × 3 linear filter to avoid over-segmentation.

Figure 4 illustrates the process for the one dimensional case. In this example two different Gaussian classes have been mixed. Once the local maxima are located, the watershed algorithm is started using these local maxima as markers and then two different subclasses are found. We can see that in this example the number of local maxima corresponds exactly with the number of subclasses. It can also be noticed that the threshold used in this simple example to classify the pixels into each subclass corresponds to the threshold of a MAP algorithm. Figure 5, shows the histogram for the skin-like detected pixels of 6.b and the subclasses found by the watershed algorithm. Figure 6.c shows the results of the unsupervised segmentation using these subclasses. It can be seen how the algorithm is able to successfully separate the face region from the background.

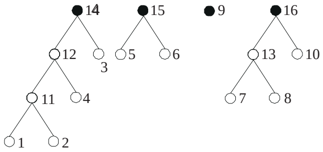


Fig. 7. Binary tree structure resulting from the homogeneous skin-like region merging.

2.3. Region merging and face extraction

The unsupervised segmentation described in the previous section can split the face regions into smaller homogeneous regions. Therefore, we must incorporate a way to merge regions into the system. To that end, we have modeled the behavior of a manually segmented set of face regions. The model takes into account parameters regarding shape, size, position and texture of the face regions and it is described in detail in [1].

Once we get the regions of the unsupervised segmentation, we search each pair-wise merging of regions to find the merged region which best fits the face model. Then, if the new region fits the model better than the original regions, they are merged. The process is repeated until we can not find any merging which fits the model better than the original regions. At this point, the merging of any two regions will only reduce the quality of the match to the face model. Figure 7 illustrates how this recursive merging process progresses. Each node represents a region of the image with the internal nodes representing regions which result from merging. The merging process terminates in a set of nodes, in this example nodes 9, 14, 15 and 16. Any of these nodes which contains less than 600 pixels are discarded, and the region that best fits the face hypothesis is used to compute the face label, which we will use to index the video sequence as described in [1].

3. RESULTS AND CONCLUSIONS

The proposed algorithm has been checked using sequences belonging to the ViBE video database [1] and to the MPEG-7 data set. Some results are shown in Figure 8. It can be seen how the algorithm is able to detect a variety of different faces in spite of the difficulty associated to different illumination conditions and different face poses. Examples of a false alarm and a miss detection are also shown in 8.g and in 8.h respectively.

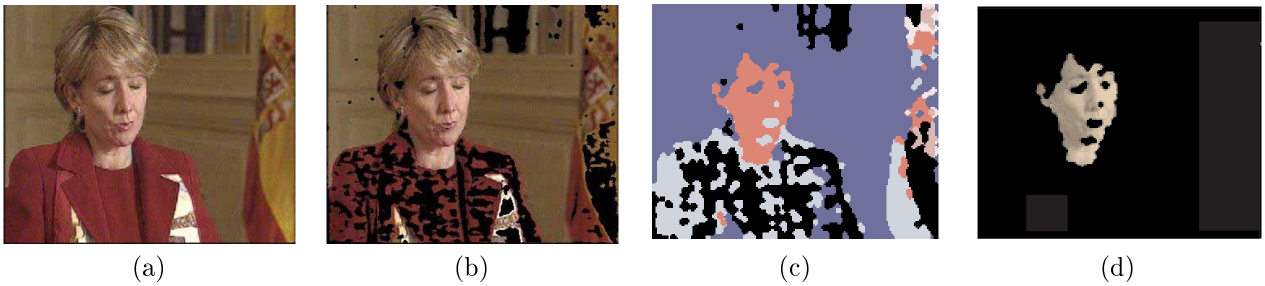
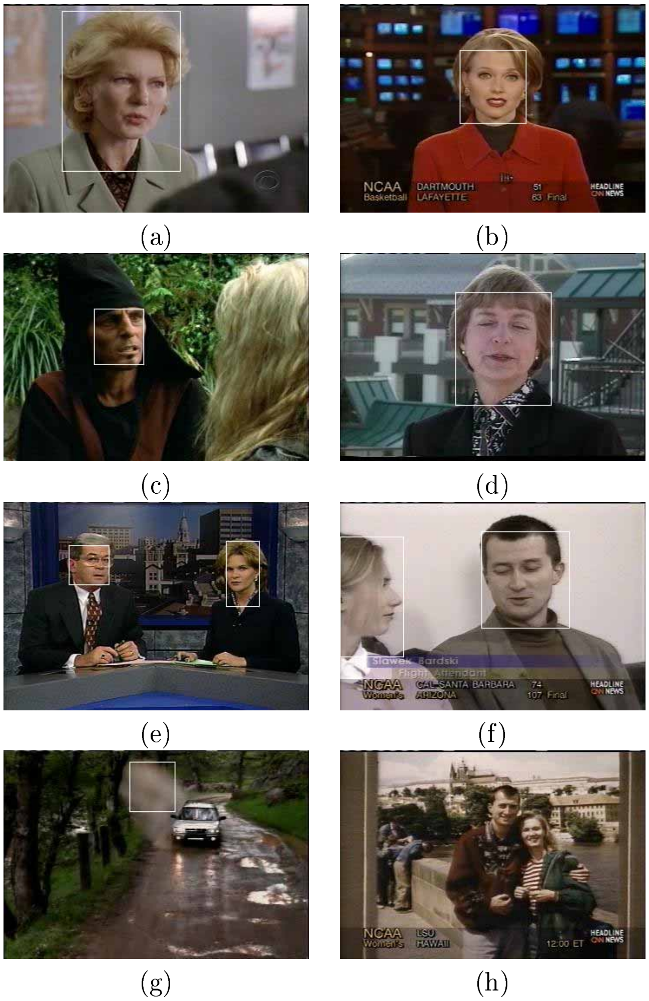

(a) (b) (c) (d)

Fig. 6. (a) Original image. (b) Skin-like detected pixels. (c) Homogeneous skin-like regions. (d)Detected face.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 8. Examples of detected faces.

4. REFERENCES

[1] A. Albiol, C.A. Bouman, and E.J. Delp, “Face detection for pseudo-semantic labeling in video databases,” in *IEEE Int. Conference on Image Processing*, Kobe, Japan, October 1999.

[2] MPEG Requirements Group, “Applications for MPEG-7,” in *Doc. ISO/MPEG N2462*, MPEG Atlantic City Meeting, October 1998.

[3] L. Torres, F. Marques, L. Lorente, and V. Vilaplana, “Face location and recognition for video indexing in the hypermedia project,” in *European Conference on Multimedia Applications, Services and Techniques*, Spain, May 1999.

[4] H. Wang and S-F. Chang, “A highly efficient system for automatic face region detection in mpeg video,” *IEEE Transactions on circuits and system for video technology*, vol. 7, no. 4, pp. 13, August 1997.

[5] M-H Yang and N. Ahuja, “Detecting human faces in color images,” in *IEEE International Conference on Image Processing*, Chicago, IL, October 4-7 1998, pp. 127–130.

[6] V. Vilaplana, F. Marques, P. Salembier, and L. Garrido, “Region-based segmentation and tracking of human faces,” in *European Signal Processing*, Rhodes, September 1998, pp. 593–602.

[7] C. Garcia and G. Tziritas, “Face detection using quantized skin color regions, merging and wavelet packet analysis,” *IEEE. Transactions on multimedia*, vol. 1, no. 3, pp. 264–277, September 1999.

[8] S. Beucher and F. Meyer, *Mathematical Morphology in Image Processing*, chapter 12. The morphological Approach to Segmentation: The Watershed Transformation, pp. 433–481, Marcel Dekker Inc., 1993.