AN UNSUPERVISED COLOR IMAGE SEGMENTATION ALGORITHM FOR FACE DETECTION APPLICATIONS

Alberto Albiol

Politechnic University of Valencia, Spain
e-mail: alalbiol@dcom.upv.es

Luis Torres†, Edward J. Delp‡

Politechnic University of Catalonia, Spain†
e-mail: luis@gps.tsc.upc.es
Purdue University, USA‡
e-mail: ace@ecn.purdue.edu

ABSTRACT

This paper presents an unsupervised color segmentation technique to divide skin detected pixels into a set of homogeneous regions which can be used in face detection applications or any other application which may require color segmentation. The algorithm is carried out in a two stage processing, where the chrominance and luminance informations are used consecutively. For each stage a novel algorithm which combines pixel and region based color segmentation techniques is used. The algorithm has proven to be effective under a large number of test images.

1. INTRODUCTION

Skin-color based face detection approaches have several advantages compared to other methods since under constant lighting conditions color is almost invariant against changes in size, orientation and partial occlusion of the face. Moreover the processing of color information has proven to be much faster than processing of other facial features which is an important point when dealing with video sequences [2, 1]. Many of the skin color based face detection schemes assume that the face can be modeled as a connected set of skin-like pixels adding a few shape and texture constraints. However, many objects commonly found in the background may also contain skin-like tones that could mistakenly be considered to be part of the facial region, as it is shown in Figure 1. In order to cope with such situations we proposed a general face detection scheme in [2] which is depicted in Figure 2. First, skin-like pixels are detected in order to discard many regions based on the color information. Once the pixels of interest are located, unsupervised segmentation is used to separate these pixels into smaller regions which are homogeneous in color. This is a very important step because the skin detection will produce non-homogeneous regions often containing more than one object. Finally, a region merging step is introduced since the unsupervised segmentation can split the face regions into smaller homogeneous regions which has been greatly improved with respect to the work presented in [2].

This work was partially supported by the grants TIC 98-0442, TIC 98-0335 of the Spanish Government and the Programa Hispano-Norteamericano de Cooperación Científica y Tecnológica

(a) (b)

Fig. 1. (a) Original image. (b) Gray areas represent pixels detected as skin

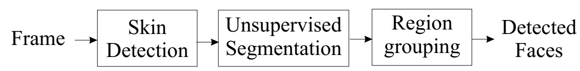

Fig. 2. (a) Original image. (b) Pixels detected as skin

This paper focuses only on the second step, the unsupervised color segmentation. This is a crucial step because if the number of regions that it provides is too small then background objects can be merged to the face area, and if it is too high it will be almost impossible to reconstruct the face from its subregions. The rest of the paper is organized as follows: First an overview of the proposed algorithm is introduced in section 2. The details of the different steps are described in sections 3 and 4. Finally some results and conclusions are shown in section 5.

2. ALGORITHM OVERVIEW

The proposed unsupervised segmentation is performed efficiently in a a two stage processing, where the chrominance and luminance information are used consecutively. For each stage a novel algorithm which combines pixel and region based color segmentation techniques is used. Figure 3 shows an outline of the proposed algorithm. The in-

Fig. 3. Algorithm outline

put of the system is an image where all the skin-like pixels have been previously labeled. Several algorithms and color spaces have been proposed for skin detection [2, 3, 4]. In this paper a simple colormap based skin detection using the YCbCr color space which provides good results will be used [2]. First, chrominance information is used alone to split the skin-like regions as detailed in section 3. Secondly, this partition is resegmented using the luminance information as detailed in section 4, this step is especially useful to segment regions with similar colors and different illumination conditions such as face and the neck areas.

3. UNSUPERVISED SEGMENTATION USING CHROMINANCE INFORMATION

The algorithm proposed for this unsupervised segmentation is a combination of pixel and region based techniques. First, the color space is clustered using the chrominance histogram of the skin detected pixels as described in section 3.1. This yields a first pixel-based segmentation which is then improved using the watershed algorithm [5] as it will be shown in section 3.2.

3.1. Histogram clustering

As mentioned above our algorithm first starts by clustering the color space using the histogram of the skin detected pixels. Several clustering techniques could be used to that end (see [6]). In our system a watershed based technique will be used to cluster the color space. To that end the histogram of the skin detected pixel is treated as a gray scale image which will be segmented. Two of the advantages of this watershed approach is that the algorithm automatically provides the number of clusters in the color space and it is very efficient computationally since the skin pixels are located in a small portion in the color space and therefore the flooding process can be performed very fast. The watershed algorithm was also used in [7] to cluster the color space. However, in that system the markers used for the watershed were set to be all the local maxima of the color histogram. Taking into account that the number of markers equals to the number of different classes, this usually yields an excessive number of classes. This paper introduces a more restrictive criteria to locate significant local maxima which will be later used as markers for the watershed algorithm as will be shown in section 3.1.2.

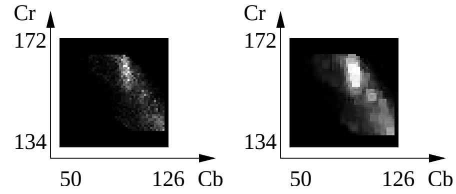

Fig. 4. (a) Chrominance histogram of the skin detected pixels of figure 6.a. (b) Preprocessed histogram

3.1.1. Histogram preprocessing

Prior to any clustering process over the chrominance histogram, it is necessary to smooth it to remove the histogram estimation error. To that end a morphological dilation using a 4×4 structuring element is performed. The size of this structuring element was chosen after determining the standard deviation of the Cb and Cr color components of 100 manually segmented faces. The values obtained were 4.3 and 3.9 respectively. Figure 4 shows the chrominance histogram of the pixels labeled as skin in Figure 6.a before and after the preprocessing. It can be noticed how the preprocessing has simplified the histogram while preserving the significant peaks.

3.1.2. Markers extraction

In order to select the significant histogram peaks, the normalized contrast is obtained for each local maxima. The normalized contrast is defined as follows:

$$\text{normalized contrast} = \frac{Contrast}{Height} \tag{1}$$

where Contrast is defined as the minimum height difference necessary to reach a higher maximum and Height is the maximum amplitude. Figure 5 presents an example where the Contrast and Height of each maximum is shown. Then all the local maxima where the normalized contrast is higher than a suitable threshold (typically 10%) are set to be markers for the watershed algorithm. It is important to emphasize that normalizing the contrast is very important to make the criteria independent of the number of pixels.

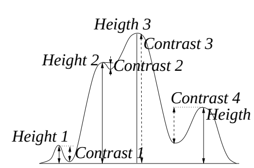

Fig. 5. Height and contrast for a histogram peak

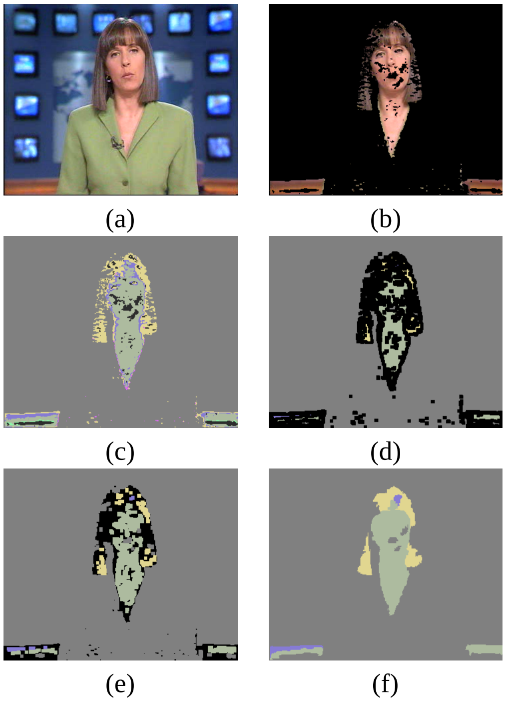

Fig. 6. (a) Original image. (b) Pixels detected as skin. (c) Segmentation after histogram clustering. (d) Label erosion of size 2 of c. (e) Label dilation of size 2 of d. (f) Watershed segmentation using the markers of e.

3.1.3. Clustering and first segmentation

Once the histogram markers are extracted, they are used to cluster the histogram using the watershed algorithm. Finally, these clusters are used to segment the image. Figure 6.b shows the pixels detected as skin from Fig. 6.a. The result of the segmentation using the chrominance information is shown in Fig. 6.c. It can be seen that the algorithm is able to split the hair and face areas, however it can not separate the face and neck areas since the chrominance in both areas is very similar since they are the same skin.

3.2. Watershed refinement

Unfortunately the previous segmentation has an associated classification error which is especially noticeable in the region boundaries as shown in Fig. 6.c. To reduce this effect a morphological filtering is applied over the resulting segmented image. The next subsections will describe this process in detail.

3.2.1. Morphological filtering

The following morphological operators are defined in order to remove spurious regions:

Label erosion, if a pixel has a label j and any of its neighbors has a label $k \neq j$, then the pixel is set to unassigned.

Label dilation, if all the assigned pixels in the neighborhood of an unassigned pixel have the same label j then the label j is assigned to the unassigned pixel.

Label erosion/dilation of size n, is done by repeating a label erosion/dilation n times.

Fig. 6.d shows the result of a label erosion of size two of Fig. 6.c. It can be noticed that spurious small regions or thin areas are set to unassigned. Fig. 6.e shows the result of a label dilation of size two of Fig. 6.d.

3.2.2. Watershed segmentation

Once the morphological filtering is finished a color watershed algorithm is used to fill the unassigned areas (black areas in Fig. 6.e). The necessary markers are set to be the remaining regions obtained after the morphological filtering. This process will propagate the labels until no unassigned pixel remains as shown in Fig. 6.f. The idea of this step is to give spatial coherence to the first pixel based classification.

4. UNSUPERVISED SEGMENTATION USING LUMINANCE INFORMATION

Once the unsupervised segmentation using the chrominance information is done, the resulting partition is resegmented using the luminance information. This is useful in situations where the chrominance information is not enough, for instance to split the neck and face areas. The segmentation procedure is very similar to the one described in the previous section. First the luminance histogram is obtained for each class using the previous chrominance segmentation. Each histogram is smoothed using a morphological dilation again. In this case the structuring element size is set to be 12 since luminance standard deviation for the face regions was found to be around 15. Figure 7, shows two of the luminance histograms and their smoothed versions. Then, the luminance histograms are clustered using the watershed algorithm. The procedure used to find the markers is the same that the one described in section 3.1. Once the new classes are obtained the segmented image using the chrominance information is resegmented using these new classes as shown in 8.g. Finally, a watershed refinement is done using the same procedure described in section 3.2. Figures 8.h and 8.i show the result of the morphological filtering and Figure 8.j shows the final segmentation.

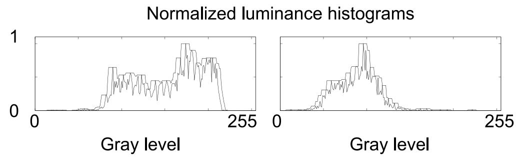

Fig. 7. Luminance histograms of two classes of Fig. 6.g and their smoothed versions

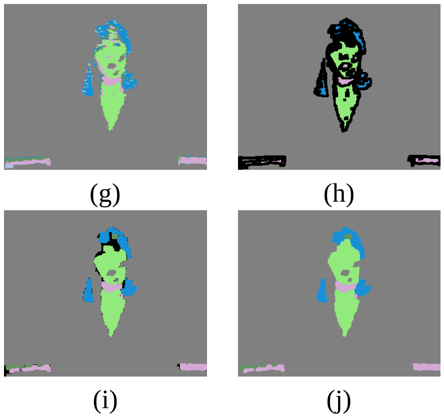

Fig. 8. (g) Segmentation after clustering the luminance histograms. (h) Label erosion of size 2 of g. (i) Label dilation of size 2 of j. (k) Final segmentation

5. RESULTS AND CONCLUSIONS

The algorithm proposed in this paper has been tested over a large number of images extracted from the VIBE video database [8]. Figure 9 shows the segmentation results for a small sample set. It can be noticed that the algorithm is able to separate the face areas from the skin colored backgrounds although in some examples the face regions are further split into two or more regions. This is not a big problem for our face detection application since usually the number of subregions is small or the regions are small compared to the face size. Furthermore, our face detection scheme [2] is designed to cope with this situation and it incorporates a region merging prior the detection. Although the algorithm proposed in this paper has been designed for this specific application, its philosophy is quite general and can be extended easily to other applications.

Fig. 9. Segmentation results

6. REFERENCES

[1] 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. 615–628, August 1997.

[2] A. Albiol, L. Torres, C.A. Bouman, and E. J. Delp, "A simple and efficient face detection algorithm for video database applications," in *Proceedings of the IEEE International Conference on Image Processing*, Vacouver,Canada, September 2000, vol. 2, pp. 239–242.

[3] D. Chai and K. N. Ngan, "Foreground background video coding using h.261," in *SPIE Visual Communications and Image Processing*, San Jose, CA, January 1998, pp. 434–445.

[4] M. J Jones and J. M. Rehg, "Statistical color models with application to skin detection," Technical report, Cambridge Res. Lab., Compaq Comput. Corp., 1998.

[5] Serge Beucher and Ferdinand Meyer, *Matehmatical Morphology in Image Processing*, chapter 12. The morphological Approach to Segmentation: The Watershed Transformation, pp. 433–481, Marcel Dekker Inc., 1993.

[6] A.N. Venetsanopoulos and K. PlataniotisJ, *Colour Image Processing and Applications*, Springer Verlag, 2000.

[7] M Petrou L Shafarenko and J Kittler, "Histogram-based segmentation in a perceptually uniform color space," *IEEE Trans Image Processing*, vol. 7, pp. 1354–1358, 9 1998.

[8] J. Chen, C. M. Taskiran, A. Albiol, C. A. Bouman, and E. J. Delp, "Vibe: A video indexing and browsing environment," in *IEEE- SPIE Conference on Multimedia Storage and Archiving Systems IV*, Boston (USA), September 1999.