Detection and Segmentation of Faces using Binary Partition Trees[1]

Verónica Vilaplana[(1)] and Ferran Marqués[(2)]

[(1)] Universidad de Buenos Aires.
Departamento de Computación, FCEyN
Universidad de Buenos Aires
Ciudad Universitaria. Pabellón I. Buenos Aires 1428, Argentina
Email: veronica@dc.uba.ar

[(2)] Universitat Politècnica de Catalunya
Campus Nord - Mòdul D5 - C/ Gran Capità, Barcelona 08034, Spain
Tel: (343) 401 64 50, Fax: (343) 401 64 47
Email: ferran@gps.tsc.upc.es

Abstract

In this paper we improve the face detection and segmentation technique proposed in [1]. In order to obtain the shape of the face, we use a region based approach and find the face as a set of regions from a generic segmentation.

The original image is segmented and a partition tree is created by merging regions from this partition. Facial descriptors and a similarity measure to faces are computed for each node. The analysis is performed using information from the regions represented by the node and also information from neighboring regions. The new method overcomes the rigidity of the tree structure and allows the extraction of new facial regions that are not represented as nodes in the tree.

A search algorithm selects the nodes associated to faces. The use of information from neighboring regions significantly improves the performance of the algorithm and avoids the post-processing step used in our previous work to completely extract the facial regions.

Keywords: face detection, face segmentation, binary partition trees, principal component analysis

[1] This work was partially supported by the FOMEC grant 877-P376-E459802, the project Interface IST-1999-10036 and the Spanish Government through the project CICYT TIC98-0422.

1. Introduction

Automatic face detection and segmentation are key issues for many applications. The more traditional ones are surveillance or user identification, but there are many new applications such as key frame selection, video indexing, content-based sequence edition, selective coding and MPEG4 applications. In this context the analysis algorithms should not only detect the faces, that is, to find the location, size and orientation of all the faces in the scene, but also segment them, obtaining their actual shapes. Different approaches to face detection and segmentation are discussed in [2,…,7].

A face is an object, an entity with semantic meaning. It can be associated to a set of homogeneous regions, with contours well defined in the original image. Consequently, it should be possible to define a face by selecting a set of regions from a correctly segmented image [9].

In order to obtain the shape of the face, we use a region based approach and try to find the face as a set of regions from a generic segmentation.

In this paper we improve the technique proposed in [1]. The original image is segmented and a partition tree is created by merging regions from this partition.

Facial descriptors and a similarity measure to faces are computed for each node. The analysis is performed using information from the regions represented by the node and also information from neighboring regions.

A search algorithm selects the nodes associated to faces. The use of information from neighboring regions significantly improves the performance of the algorithm and avoids the post-processing step used in our previous work to completely extract the facial regions.

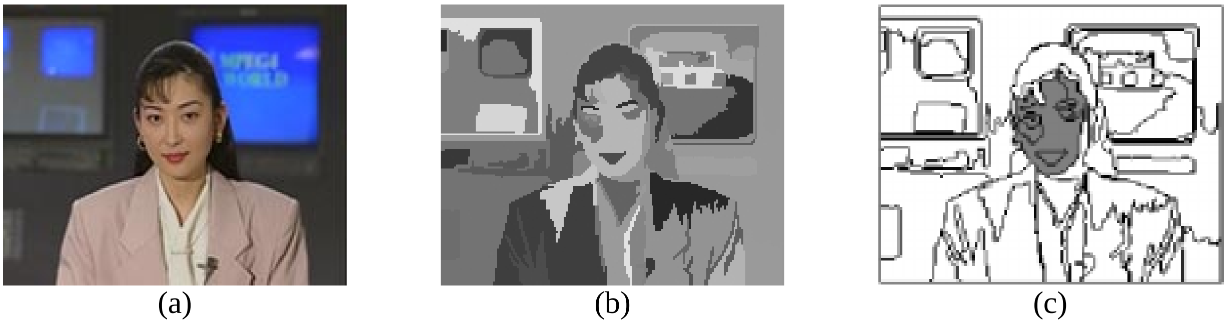
The organization of the paper is the following. Section 1 has stated the problem and briefly presented our algorithm. Section 2 describes the creation of an initial segmentation. Section 3 is devoted to the creation of the Binary Partition Tree (BPT), a region based representation that is appropriate for the segmentation task. Section 4 details the analysis of the BPT, and Section 5 explains the face extraction step. In Section 6, several results are presented to assess the quality of the proposed technique. Finally, Section 7 comments the current work.

2. Initial Partition

The main goal of this step is the simplification of the analysis. The analysis starts at region level, so the image is segmented into homogeneous regions using the merging strategy proposed in [8]. In this case, we merge neighboring regions from the partition of flat zones of the original image, modeling each region with its mean value in each color component.

The order of the mergings is given by the Euclidean distance in the YCbCr color space between regions. Regions are merged until a certain PSNR is achieved, typically 30 dB. Ideally, the obtained partition should be fine enough to contain, among others, all the regions that form the face.

In Figure 1 an example of an initial partition is presented. Figure (a) shows the original frame #0 of the *Akiyo* sequence, Figure (b) presents the initial partition, with 70 regions. The partition is well oriented to the face segmentation task since contains a set of regions that preserve the contours of the object of interest, as can be seen in (c).



(a) (b) (c)

Figure 2 - (a) Original frame #0 of the sequence *Akiyo*, (b) Color based partition, (c) Face regions

3. Binary Partition Tree Creation

The algorithm will find every face as a set of regions from the initial partition. But it is impossible to analyze all regions and all possible unions of regions; some simplification is needed. A proposal of candidates has to be made.

We use a tree structure, called Binary Partition Tree [8] as a region-based representation that is appropriate for the face segmentation. The Binary Partition Tree (BPT) is created by merging neighboring regions from the initial partition until only one region remains.

For each par of neighboring regions, a homogeneity measure is assessed. The merging algorithm then starts merging the pair of neighbors whose distance is lower. Then distances are recomputed and the process is iterated until one region is obtained. The tree represents the sequence of mergings. Leaves are related to the regions in the initial partition, and the remaining nodes are associated to the regions that are created during the merging process. Once the tree is built, every node can be separately analyzed.

For the face segmentation, the tree has to be created in such a way that the most meaningful regions are represented in its nodes. Faces contain many regions that are homogeneous in chrominance. This observation leads to the use of a chrominance criterion to build the tree. The merging order is defined by the Euclidean distance in the CrCb space between the mean of each region.

All possible mergings of regions from the initial partition are not represented in the tree. However, this representation allows the detection of all the faces present in the scene. If the initial partition is fine enough, the core regions of the faces will be nodes in the tree.

Figure 2 shows the binary partition tree associated to the partition presented in Figure 1.

4. Binary Partition Tree Analysis

The BPT is analyzed in order to select the nodes that represent the core components of the faces in the image. The first part of the analysis is performed by rejecting or deactivating the nodes that cannot represent faces using simple criteria such as size, color and aspect ratio. The use of these criteria greatly simplifies the search since, on average, 80% of the nodes are deactivated.

The rejection of a node means that the associated region is not the core component of a face. However, it may well be one of the regions that form the face (for example, a region associated to an eye).

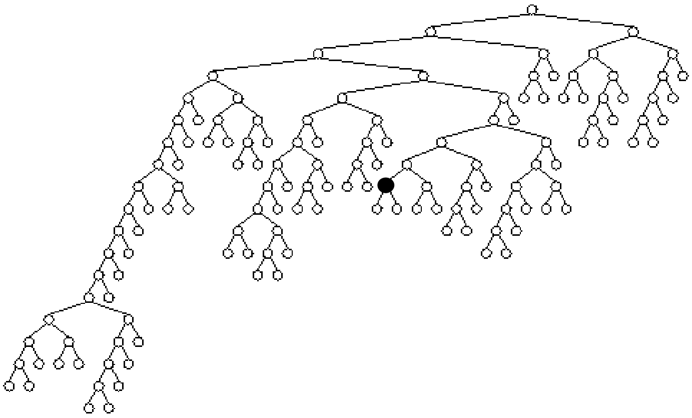


Figure 2: Binary partition tree associated to the partition in Figure 1-b

For the remaining nodes, a measure of distance to a face class is estimated. In [1], the information used to evaluate these nodes was restricted to their associated regions. However, this approach had many problems:

(i) other facial regions were not analyzed, so candidate nodes partially represented the face regions
(ii) information about the type of object being sought, shape for instance, was not used
(iii) the distance computation used incomplete data
(iv) selected nodes partially represented the faces, so a refinement step was needed

To overcome these problems and provide with more flexibility to the BPT structure, we propose to analyze every active node in the tree using information from its associated regions as well as from neighboring ones.

The area of support of a node is determined by the shape of the object to detect, frontal faces in our case. Therefore for each node in the tree, its area of support is formed by the regions of the initial partition contained in an elliptical mask placed on the node regions. An ellipse is used as a first approximation to a face shape.

Figure 3(a) is a mask showing the area of support of the black node in Figure 2. Black regions are the regions represented by the node, light gray regions are new regions inside the elliptical mask (in dark gray).

Figure 3(b) shows the information from the original image associated to the node, 3(c) shows the information within the area of support of the node, and 3(d) shows the regions associated to its parent node in the tree. This example illustrates the convenience of the method: we can find a set of facial regions that is not represented as a node in the BPT.

Once each active node is extended by the computation of its area of support, more pruning criteria such as compactness, aspect ratio and circularity can be used to deactivate some of them. This allows the rejection, on average, of 95% of the original nodes.

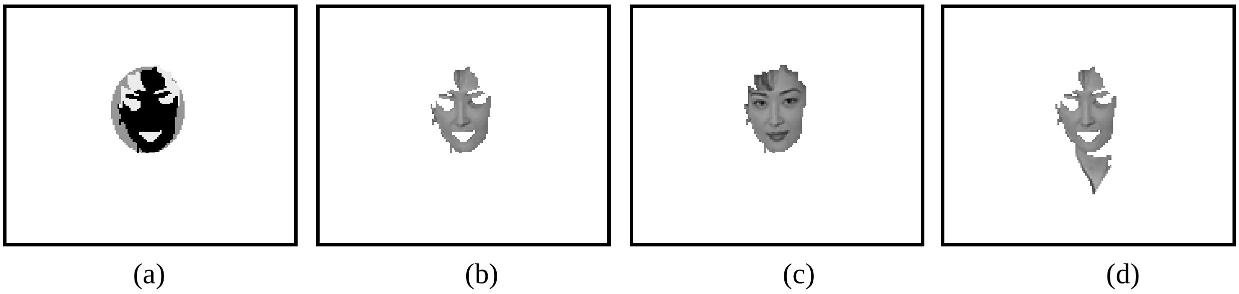


Figure 3 - (a) Mask of supporting area, (b) Node regions, (c) Area of support, (d) Parent node regions

Then for each active node, we compute a similarity measure [2] between its area of support and the face class. The face class Ω is characterized using a database of face images I$_i$, $i=1,...,K$, K being the size of the database. Each $n \times m$ image I$_i$ is read by rows in a N-dimensional vector x$_i$ in R^N, where $N = nm$. In the N dimensional space, the mean face image $\bar{x}$ and the covariance matrix Σ are estimated, and the covariance matrix is diagonalized $\Lambda = \Phi^T \Sigma \Phi$ where Φ is the eigenvector matrix of Σ and Λ is the diagonal matrix of eigenvalues.

Principal Component Analysis (PCA) can then be used to obtain the projection matrix Φ_M of the N dimensional space into the M dimensional subspace spanned by the M eigenvectors with largest eigenvalues ($M << N$). This gives a decomposition of the space in two complementary subspaces, the principal or face subspace and its orthogonal complement. The projection of a mean normalized image vector $\tilde{x} = x - \bar{x}$ into this subspace is $[y_1,..., y_M] = \Phi_M^T (x - \bar{x})$.

The component of any normalized image $\tilde{x}$ in the orthogonal complement, called Distance From Feature Space (DFFS), can be computed using the Euclidean norm of x and the principal coefficients $y_1,..., y_M$.

$$DFFS(x,\Omega) = \|\tilde{x}\|^2 - \sum_{i=1}^{M} y_i^2$$

The DFFS is an Euclidean distance that can be effectively used for the face detection task.

To compute the distance, an auxiliary image containing a scaled version of the information from the original image is created. First, the subimage given by the tightest rectangle bounding the area of support of the node is found. This subimage is then warped to a new image whose size is the same as the database size (see Figure 4). This warping preserves the aspect ratio of the initial subimage. The areas outside the area of support are set to a constant value. Finally the auxiliary image is projected into the face subspace and the distance to the subspace is calculated.

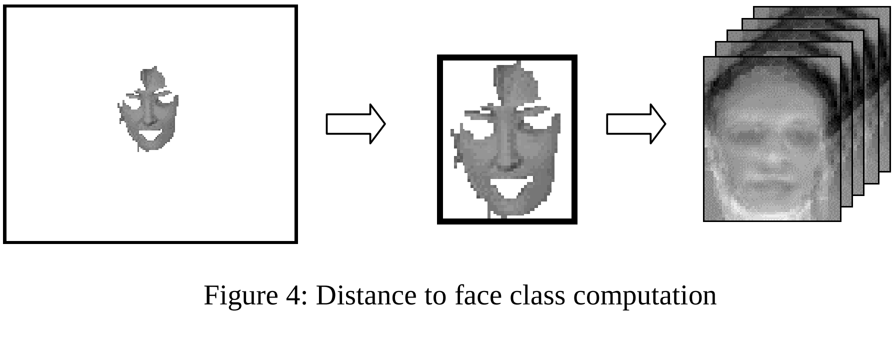


Figure 4: Distance to face class computation

5. Face Extraction

Once the distances are computed for all active nodes, a search algorithm selects the nodes associated to faces. The algorithm analyzes the tree recursively. Since different nodes may lead to the same area of support, the search algorithm has to take care of that. It also has to select the best candidate when two active nodes (with different areas of support) are associated to the same object.

Using general facial descriptors (obtained from the face image database) such as color and shape, and the computed distances, the algorithm finds a threshold that allows the detection of all the faces present in the scene. This is very useful, since the great variability in test images makes impossible the estimate of a unique, predefined threshold. Note that the computational load of the distance computation and search algorithm steps is low, because most of the nodes have been previously rejected using simple criteria.

6. Results

The algorithm has been tested with a large number of images representing very different conditions. Some of the results are presented in Figure 5. The *Olivetti Research Lab.* database of face images has been used to characterize the face class with M=5 eigenvectors.

The first three examples are test images from MPEG-7 database, with two faces per image. In all cases both faces are correctly detected and segmented, despite they are not purely frontal and have very different sizes. There are not false detections although the images contain many other skin colored regions. The second example shows that our technique is able to find the face of the girl with the instrument in spite it partially occludes her face. The first plane face in the third example contains also the neck; this is a problem of the initial segmentation that does not represent exactly the facial contour.

Next three examples are also MPEG-7 test images, with one face per image. They have different cluttered backgrounds and lighting conditions and in all cases the faces are correctly segmented.

The last ones are images from the MX2VTS[2] database. The algorithm performs well in all cases, being robust to differences in skin color, facial expressions and to the presence of additional elements (glasses).

7. Conclusions

The proposed algorithm successfully performs in a large set of images. Our current work focuses in different aspects such as the reduction of the computational load, and the use of the XM2VTS for the validation of the technique. Also, as we mentioned, the technique relies on the quality of the initial partition: initial regions must correctly represent the face boundaries; we are planning to use skin probability maps to solve this problem.

[2] The authors acknowledge the use of the Extended Multimodal Face Database and associated documentation. Further details can be found in: K.Messer, J.Matas, J.Kittler, J.Luettin and G.Maitre; “XM2VTSbd: The Extended M2VTS Database, Proc. 2nd Conf. on Audio and Video based Biometric Personal Verification”, Springer Verlag, New York, 1999. http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb.

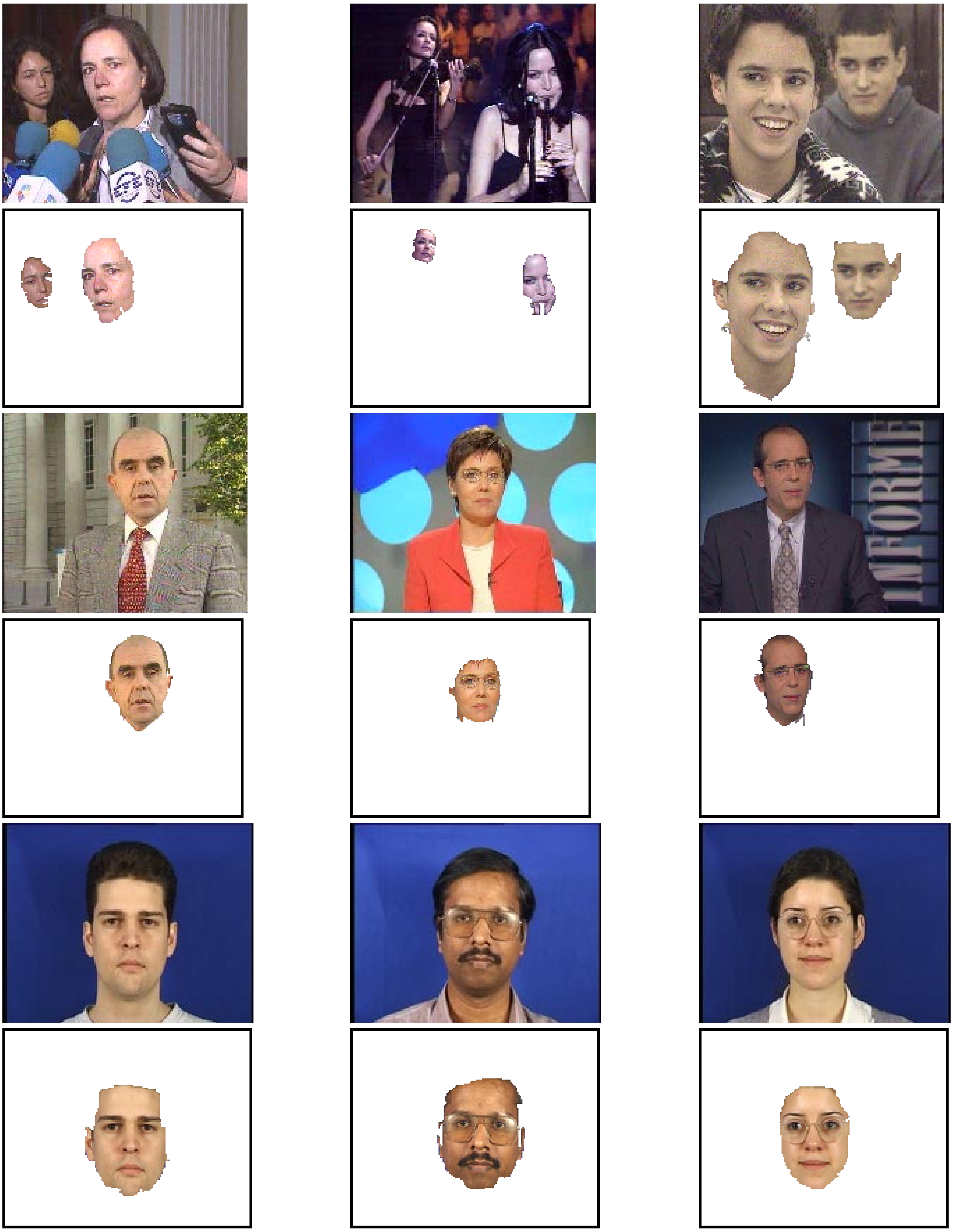


Figure 4- Examples of face segmentation in MPEG-7 and XM2VTS image

8. References

[1] F. Marqués and V. Vilaplana. Face segmentation and tracking based on connected operators and partition projection. *Pattern Recognition*, 2001.

[2] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7): 696-710, Jul 1997.

[3] H. Rowley, S. Baluja and T. Kanade. Neural network based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1): 23-38, Jan 1998.

[4] K.K. Sung and T. Poggio. Example based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1): 39-51, Jan 1998.

[5] A. Colmenarez and T.S. Huang. Face detection with information-based maximum discrimination. *Int. Conference on Computer Vision and Pattern Recognition*, 782-787, 1997.

[6] K. Sobotka, I. Pitas. A novel meted for automatic face segmentation, facial feature extraction and tracking. *Signal Processing: Image Communication*, (12): 263-281, 1999.

[7] D. Chai and K. Ngan. Face segmentation using skin color map in videophone applications. IEEE Transactions on Circuits and Systems for Video Technology, 9(4): 551-564, Jun 1999.

[8] P. Salembier and L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 9 (4): 561-576, April 2000.

[9] P. Salembier and F. Marqués. Region-based representations of image and video: Segmentation tools form multimedia services. *IEEE Transactions on Circuits and Systems for Video Technology*, 9 (8): 1147-1167, Dec. 1999.