RULE-BASED FACE DETECTION IN FRONTAL VIEWS

Constantine Kotropoulos *Ioannis Pitas**

Department of Informatics
Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece
{ costas, pitas }@zeus.csd.auth.gr

ABSTRACT

Face detection is a key problem in building automated systems that perform face recognition. A very attractive approach for face detection is based on multiresolution images (also known as *mosaic images*). Motivated by the simplicity of this approach, a rule-based face detection algorithm in frontal views is developed that extends the work of G. Yang and T.S. Huang. The proposed algorithm has been applied to frontal views extracted from the European ACTS M2VTS database that contains the videosequences of 37 different persons. It has been found that the algorithm provides a correct facial candidate in all cases. However, the success rate of the detected facial features (e.g. eyebrows/eyes, nostrils/nose, and mouth) that validate the choice of a facial candidate is found to be 86.5 % under the most strict evaluation conditions.

1. INTRODUCTION

Face recognition has been an active research topic in computer vision for more than two decades. A critical survey of the literature on human and machine face recognition is given in [1]. One of the key problems in building automated systems that perform face recognition tasks is face detection. Recently, the research on model-assisted coding schemes [2] in addition to the need for multimodal verification techniques in tele-services and teleshopping applications [3] has reinforced the interest on face detection algorithms. Many algorithms have been proposed for face detection in still images that are based either on texture, depth, shape and color information or a combination of them. A very attractive approach for face detection is based on multiresolution images (also known as *mosaic images*) attempting to detect a facial region at a coarse resolution and subsequently to validate the outcome by detecting facial features at the next resolution level [4]. Towards this goal, the method employs a hierarchical knowledge-based pattern recognition system. Fuzzy theory has been used in [5] in order to cope with the inexact knowledge about the face and to improve the performance of the method. Mosaic images have been employed in detecting an unknown human face in input imagery and recognizing the facial expression as well [6]. Another closely related method that is based on a coarse-fine detection scheme uses nonlinear sampling lattices and simple models for magnocellular ganglion cells to extract salient regions and to find the bounding rectangle of a face [7].

Motivated by the simplicity of the face detection approach in [4] and the need for a mechanism that controls the placement of a sparse grid over a face in order to store a model for each person in face recognition algorithms based on Dynamic Link Architecture [8], we propose a variant of the method in [4] that has the following features:

1. It allows for rectangular cells in contrast to the square cells used in [4].
2. It is equipped with a preprocessing step that determines an estimate of the cell dimensions and the offsets so that the mosaic model fits the face image of each person.
3. It has very low computational demands compared to the original algorithm, because the iterative nature of the latter is avoided due to the preprocessing step that has been used.
4. It employs more general rules that are close to our intuition for a human face.

Accordingly, the work presented in this paper extends previously reported work [4]. The proposed algorithm has been applied to frontal views extracted from the European ACTS M2VTS database [9]. One set of 37 frontal views, one for each person has been used to evaluate the performance of the proposed method. The experimental results indicate that the algorithm can provide a correct facial candidate in all cases. However, the true success rate of detected facial features (e.g. eyebrows/eyes, nostrils/nose, and mouth) that validate the choice of a facial candidate is found to be 86.5 % under the most strict evaluation conditions.

The outline of the paper is as follows. The proposed face detection algorithm is described in Section 2. Experimental results are reported in Section 3. In this section, the performance of the proposed algorithm is also evaluated. Finally, conclusions are drawn in Section 4.

2. FACE DETECTION BASED ON MOSAIC IMAGES

To begin with let us describe the framework proposed in [4]. The original algorithm is based on mosaic images of reduced resolution that attempt to capture the macroscopic features of the human face. It is assumed that there is a

*This work is supported by the European ACTS M2VTS project.

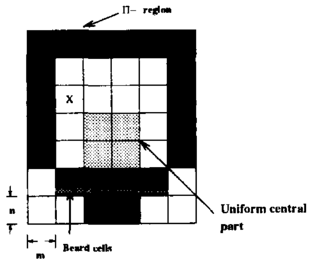
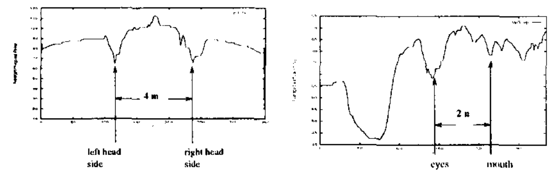


Figure 1. Abstract face model at the resolution level of the quartet image.



(a) (b)

Figure 2. (a) Horizontal profile. (b) Vertical profile.

resolution level where the main part of the face occupies an area of about 4×4 cells. Accordingly, a mosaic image can be created for this resolution level. It is the so called *quartet* image. The grey level of each cell equals the average value of the grey levels of all pixels included in the cell. An abstract model for the face at the resolution level of the quartet image is depicted in Figure 1. The main part of the face corresponds to the region of 4×4 cells having as origin the cell marked by "X". By subdividing each quartet image cell to 2×2 cells of half dimensions the *octet* image results, where the main facial features such as the eyebrows/eyes, the nostrils/nose and the mouth are detected. Therefore, an hierarchical knowledge-based system can be designed that aims at detecting facial candidates by establishing rules applied to the quartet image and subsequently at validating the choice of a facial candidate by establishing rules applied to the octet image for detecting the key facial features mentioned above.

As can be seen, the underlying idea in [4] is very simple and very attractive, because it is close to our intuition for the human face. However, the implementation proposed in [4] is computationally intensive. The algorithm is applied iteratively for the entire range of possible cell dimensions in order to determine the best cell dimensions for creating the quartet image for each person. Another limitation is that only square cells are employed.

In order to avoid the iterative nature of the original method, we propose to estimate the cell dimensions in the quartet image by processing the horizontal and the vertical profile of the image. Let us denote by n and m the vertical and the horizontal quartet cell dimensions, respectively.

The horizontal profile of the image is obtained by averaging all pixel intensities in each image column. By detecting abrupt transitions upwards, two significant local minima are determined in the horizontal profile. These local minima correspond to the left and right side of the head (i.e., roughly speaking to the cheeks). The distance between them is set to $4m$, as can be seen in Figure 2a. Accordingly, the quartet cell dimension in the horizontal direction can easily be estimated.

Similarly, the vertical profile of the image is obtained by averaging all pixel intensities in each image row. The significant local minima in the vertical profile correspond to the hair, eyebrows, eyes, mouth and chin. It is fairly easy to locate the row where the eyebrows/eyes appear in the image by detecting the local minimum after the first abrupt transition in the vertical profile. Searching for the row where the upper lip appears in the image, first the nose tip should be detected. It corresponds to a significant maximum that occurs below the eyes. Then, the steepest minimum below the nose tip is associated to the upper lip. By setting the distance between the rows where the eyes and the upper lip have been found to $2n$, the quartet cell dimension in the vertical direction can be estimated, as is depicted in Figure 2b. It is evident that the proposed preprocessing step overcomes also the drawback of square cells, because the cell dimensions are adapted to each person separately.

Having estimated the quartet cell dimensions, we proceed to the description of facial candidate detection rules. Since the system remains hierarchical, as in [4], it is more preferable to decide that a face exists in a scene although there is no actually a face than not to detect a face that exists indeed. The decision whether or not a region of 4×4 cells is a facial candidate is based on:

- the detection of a homogeneous region of 2×2 cells in the middle of the model that are shown in light grey color in Figure 1, and,
- the detection of homogeneous connected components having significant length in the Π-shaped region shown in black color in Figure 1, or,
- the detection of a beard region shown in dark grey color in Figure 1.

Moreover, a significant difference in the average cell intensity between the central 2×2 region and the Π-shaped region must be detected. For the sake of completeness, we note that if there are not adequate cells in the vertical direction, the Π-shaped region may have a total length of 12 cells instead of 14 cells. This is done by reducing the height of the Π to 4 cells. We have found that the above-described rules are more successful in detecting a facial candidate than those proposed in [4]. It is also worth noting that beard detection rules are not included in [4].

Subsequently, eyebrows/eyes, nostrils/nose and mouth detection rules are developed to validate the facial candidates determined by the procedure outlined above.

Eyebrows/eyes detection rules. First, the local extrema in the vertical profile of the octet image are determined. If there is one local minimum inside an admissible search interval in the upper half portion of the octet image, then its coordinate in the vertical profile indicates the row where octet cells corresponding to eyebrows/eyes should be searched for. If there are more than one local minima, the local maxima are arranged with respect to the local minima and the local minimum with the steepest slope between itself and the preceding local maximum is found. The row where such a local minimum is detected, is searched for octet cells that correspond to eyebrows/eyes. Alternatively, the brightest octet cell in the upper part of the octet image is determined and we check if it lies in the middle. Additionally, we examine if there is a pair of octet cells that possess similar grey level values and if a significant difference in grey level values exists between that pair of cells and the brightest octet cell. Then, the most symmetric octet image row is determined. That is, we determine the row where the sum of absolute grey level differences before and after the brightest octet is minimal. In this row, octet cells corresponding to eyebrows/eyes are to be searched for.

If the row where eyebrows/eyes appear in the octet image is determined, then the octet cells that correspond to eyebrows/eyes can be found by detecting the local extrema in that row. Several possibilities can be examined. For example, a significant maximum occurs in the middle with the octet cells before and after it possessing a symmetry. Another case is when more than one local extrema occur in the row under examination. In such a case, two local minima at a suitable distance are detected that do not differ so much in grey level values. In conclusion, the rules developed enhance the ones proposed in [4] by highlighting the key role of symmetry.

Nostrils/Nose detection rules. Searching for nostrils/nose candidates is done in all columns between the eyes and below them. Local minima in an admissible interval are searched for in those columns. Then, we check if the local minima are close to the central part of the octet image. If this is the case, the row that corresponds to the local minimum in the vertical direction is searched for local minima in the horizontal direction as well. Our objective is to detect local minima that coincide in both directions. However, a small offset is tolerated in matching local minima in vertical and horizontal directions. If local minima in the vertical direction can not be found, then a local maximum is determined below the eyes. The rows in an admissible search interval below the row where the local maximum is detected, are additionally searched for local minima. The octet cells corresponding to such minima are considered as eligible nostrils/nose candidates as well.

Mouth detection rules. Mouth candidates are detected as homogeneous regions of 2 or 3 octet cells at most three rows below the nostrils/nose candidates in the octet image.

Nostril/nose candidates that fall inside mouth regions and yield other mouth candidates at a distance greater than 6 rows from the eyes in the octet image are discarded. Mouth candidates that fall inside nostrils/nose regions and are close to the eyes are filtered out as well. Finally, symmetry is imposed on the facial candidate by correcting the initial offset estimates provided by the preprocessing step employed.

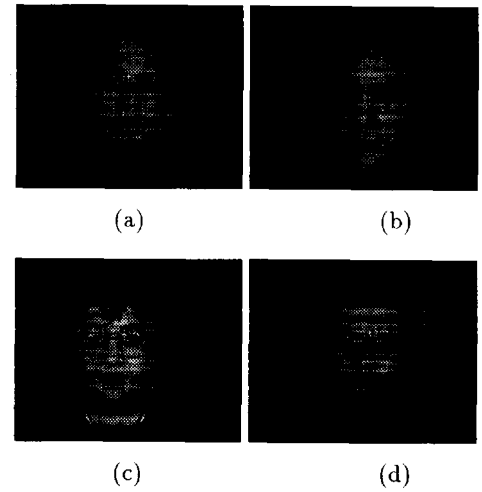


Figure 3. Problems in facial feature detection. (a) Wrong nostrils/nose detection. (b) Wrong eyebrows/eyes detection. (c) Wrong nostrils/nose detection. (d) Wrong mouth detection.

3. EXPERIMENTAL RESULTS

The proposed algorithm has been applied to the European ACTS project M2VTS database. The database includes the videosequences of 37 different persons. The algorithm provides a correct facial candidate in *all* cases. However, the detected facial features that validate the choice of the facial candidate are not always correct, as can be seen in Figure 3.

Table 1 summarizes the rate of success/failure for each detected facial feature. The true rate of success of the proposed method is 86.5 % under the most strict evaluation conditions. Results of face detections are shown in Figures 3 and 4. In all figures, the octets for the facial features are shown overlaid. The octets for eyebrows/eyes and nostrils/nose are shown as white overlaid rectangles. Mouth candidates are shown as black overlaid rectangles. The white cross indicates the characteristic bright octet between the eyes. Examples of false facial feature constellations are shown in Figure 3. The facial candidates in Figure 4 are successfully detected and the facial features are located correctly.

Table 1. Evaluation of the face and facial feature detection rules in M2VTS database (37 persons).

Feature	Success	
	Number of frames	[%]
Left eyebrow/eye	35	94.6
Brightest octet	34	94.4
Right eyebrow/eye	37	100
Nostrils/nose	36	97.3
Mouth	35	94.6
Eyebrows/eyes, Nostrils/nose, Mouth	32	86.5

4. CONCLUSIONS

A rule-based detection algorithm in frontal views has been proposed in this paper. The proposed algorithm extends the work of G. Yang and T. Huang [4]. It has been found by experiments that the algorithm has a true success rate in face detection of 86.5 % under the most strict evaluation conditions. The output of the algorithm has been successfully used to control the placement of the grid in dynamic link matching [10].

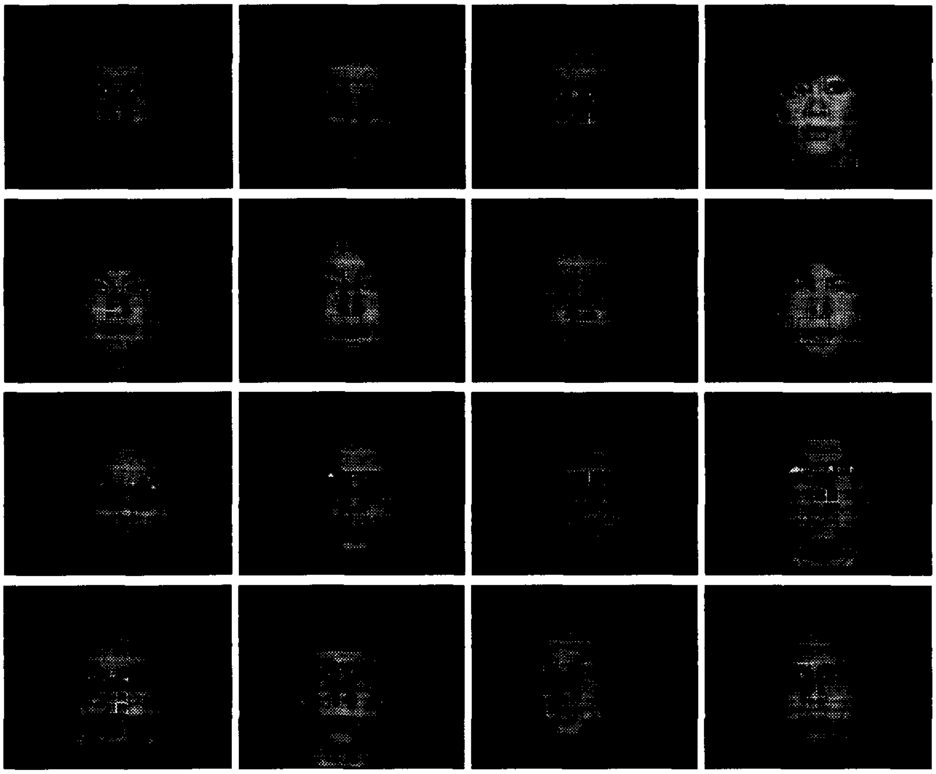


Figure 4. Correct results of face detection.

REFERENCES

[1] R. Chellapa, C.L. Wilson and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

[2] E. Elefteriadis and A. Jacquin, "Automatic face location, detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates," *Signal Processing: Image Communication*, vol. 7, no. 3, pp. 231–248, July 1995.

[3] M. Acheroy, C. Beumier, J. Bigün, G. Chollet, B. Duc, S. Fischer, D. Genoud, P. Lockwood, G. Maitre, S. Pigeon, I. Pitas, K. Sobottka, and L. Vandendorpe, "Multi-modal person verification tools using speech and images," in *Proc. of the European Conference on Multimedia Applications, Services and Techniques*, pp. 747–761, Louvain-La-Neuve, Belgium 1996.

[4] G. Yang and T.S. Huang, "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.

[5] H. Wu, Q. Chen and M. Yachida, "An application of fuzzy theory:face detection," in *Proc. of the Int. Workshop on Automatic Face- and Gesture- Recognition* (M. Bichsel, Editor), pp. 314–319, Zurich, 1995.

[6] K. Matsuno, C.-W. Lee, S. Kimura and S. Tsuji, "Automatic recognition of human facial expressions," in *Proc. of the IEEE Fifth Int. Conf. on Computer Vision*, pp. 352–359, Massachusetts, U.S.A., 1995.

[7] B. Takács, and H. Wechsler, "Face location using a dynamic model of retinal feature extraction," in *Proc. of the Int. Workshop on Automatic Face- and Gesture-Recognition* (M. Bichsel, Editor), pp. 243–247, Zurich, 1995.

[8] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P Würtz, and W. Konen, "Distortion invarinat object recognition in the Dynamic Link Architecture," *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, March 1993.

[9] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," in *Proc. of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 97)* to appear, Crans-Montana, Switzerland, 12-14 March 1997.

[10] C. Kotropoulos, I. Pitas, S. Fischer, and B. Duc, "Face Authentication using Morphological Dynamic Link Architecture," in *Proc. of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 97)* to appear, Crans-Montana, Switzerland, 12-14 March 1997.