A Face Detection Solution Integrating Automatic and User Assisted Tools

João Ascenso[1], Paulo Correia[2], Fernando Pereira[3]

Instituto Superior Técnico – Instituto de Telecomunicações

Av. Rovisco Pais, 1049-001 Lisboa, Portugal

email: [1] Joao.Ascenso@lx.it.pt, [2] Paulo.Correia@lx.it.pt, [3] Fernando.Pereira@lx.it.pt

Abstract: Knowledge about the number, position and size of human faces present in a given image or video is valuable information for their content-based description. This paper proposes a two-stages solution for the face detection problem using both colour and shape information. The first stage consists in a pixel-based colour segmentation algorithm of the HSV colour space. For this purpose, a study about the skin-colour distribution of faces in natural images is first presented. The second stage recognises that an ellipse can approximate the outline of a human face shape in several positions. In this stage, an algorithm based on the Hough Transform is used to perform ellipse detection with some shape constraints. Since the automatic detection tool does not perform satisfactorily all the time, this tool is complemented with a user-assisted tool allowing the user to improve the automatic detection results. The system implemented has been tested for a wide variety of natural sequences and the experimental results show that the algorithm is robust to changes of position, size, orientation and number of faces in an image.

Keywords: Face Detection, Colour Segmentation, Hough Transform.

1 Context and Objectives

The detection of faces is a task performed daily by human beings. However, the automatic detection of human faces in natural images is a fairly complex task, and it has been intensively studied for many years. In recent years, the research activities in this area have been intensified as a result of the easy acquisition, communication and storage of visual information. The main task of these systems is to find the location of a person's face in an image or in a video sequence. The terms *face detection*, *face localization*, *face extraction* or *face segmentation* are often found in the literature to refer to this problem; *face detection* is the term to be used along this paper.

The significant advances of the last decade in the area of audiovisual technology, both in terms of software and hardware products, made easier the acquisition, processing and distribution of audiovisual content. This resulted in a huge amount of information to be handled by image and video databases, thus requiring efficient indexing and search engines. Nowadays a major objective of multimedia research and development is to allow the user to quickly and efficiently identify (search, filter, etc.) various types of audiovisual content according to his needs. Human faces assume a major role in many image and video databases, as they can be used for describing important content in photographs, news clips, interviews, home-made videos, films, etc. Taking into account the specific characteristics of facial information due to its importance for humans, there are three major areas of application for a face detection tool:

1) **Face recognition** – Robust face detection must typically be performed to appropriately locate the faces to be used by a face recognition module.

2) **Content-based coding** – By segmenting faces in images, a face detection tool is useful for their content-based coding such as supported by the MPEG-4 standard.

3) **Content-based description** – Face detection information, e.g. where, when, what, can be used to index and search images in multimedia databases, such as those foreseen by the MPEG-7 standard.

This paper proposes a face detection algorithm, which in its automatic part consists in two main stages associated to the colour and shape data. First, the distribution of the skin-colour in natural images is investigated, and normalised histograms for the HSV colour space are presented together with some detection results. Then a technique to perform colour segmentation and some post-processing operations to improve the shape of the colour candidate regions and to represent them in a suitable manner to be used in the shape processing module are introduced. Also the problems and limitations of this technique when used alone are discussed. The second stage consists in an ellipse localisation algorithm, exploiting the knowledge that the shape of human faces is approximately elliptic. The ellipse detection algorithm developed is based on the Hough Transform method for curve detection [6]. This module processes the outcome of the colour processing stage, selecting the regions with approximately elliptic shape. Experimental results for several types of test sequences are presented. Complementing the fully automatic face detection algorithm, an interactive tool for correcting automatic results is presented. This tool allows the user to correct any mistakes or oversights that the automatic tool may have produced, by supporting *add*, *remove* and *correct* face actions. Finally, the conclusions are presented and items for future research proposed.

2 Automatic Face Detection

The face detection system presented in this paper aims to detect faces in unconstrained natural image sequences, and its architecture is represented in the block diagram of Figure 1.

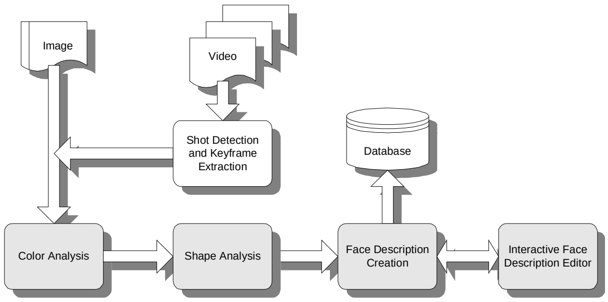


Figure 1 – Architecture of the face detection system

The grey boxes correspond to the modules described in this paper. The shot detection and keyframe extraction modules are used to select the representative images of a video sequence where face detection should be performed. Then colour analysis is performed, identifying an initial set of regions with skin-like colour, from which the face-like shaped regions are selected. The output of the algorithm is a face description, to be stored in a database, containing the number of faces present in each keyframe, together with their position, size and orientation. The automatic description can be further refined by using the interactive face description editor. The automatic face detection modules are described in this section.

2.1 Colour Analysis

Human skin-colour is a powerful cue in detecting and localizing faces in images and has been used efficiently in many domains from face segmentation to hand tracking. One of the main requirements for a face detection system is that it can handle the variations that occur in natural image sequences. These variations can be in the image conditions (size, location, lighting, orientation, viewpoint), in the face appearance (glasses, beard, moustache, makeup, expression), or in the image contents (complexity of the background, occlusion by other objects, number of faces). The human skin-colour is one of the most robust face features regarding the variations mentioned above, and is used as the basis for the colour segmentation model developed in this paper.

2.1.1 Colour Space

The efficacy of the colour segmentation stage depends essentially on the colour space chosen. The human skin-colour distribution should be as confined as possible, even when variations due to the pigmentation or to the illumination of the face in a particular image occur. The colour space that produces the clearest separation between skin-colour and other colours must be chosen in order to obtain the best results. The HSV colour space has been selected for this purpose, as it produces a good separation between skin-colour and other colours with varying illumination [4].

The HSV is a cylindrical coordinate system and the subset of this space defined by the colours is a hexcone or a six-sided pyramid. The Hue value (H) is given by the angle around the vertical axis with red at 0°, yellow at 60°, green at 120°, cyan at 180°, blue at 240°, and magenta at 300°. Note that complementary colours are 180° apart. The Saturation value (S) varies between 0 and the maximum value of 1 in the triangular sides of the hexcone and refers to the ratio of purity of the corresponding hue to its maximum purity when S equals 1. The Value (V) also ranges between 0 and 1 and is a measure of relative brightness. At the origin, when the V component equals 0, the H and S components become undefined and the resulting colour is black. The set of equations that transform a point in the *RGB* colour space to a point in the *HSV* colour space is presented in [1].

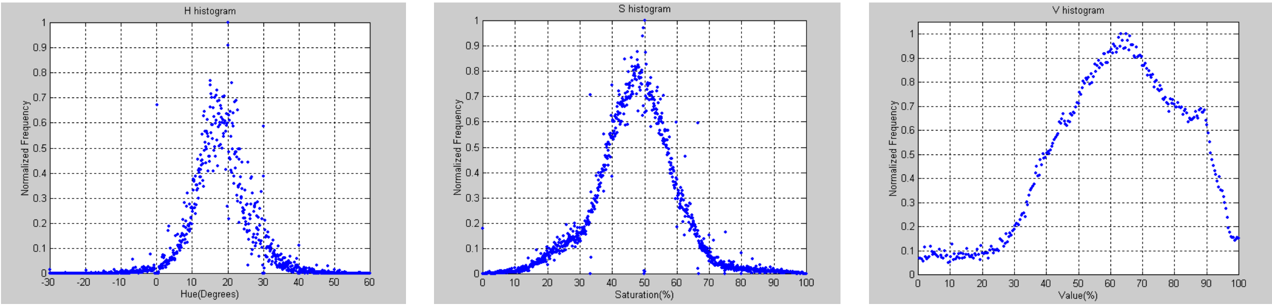
2.1.2 Colour Distribution of Skin Regions

Before performing any colour segmentation, the distribution of skin-colour within several representative face images has to be studied. To build colour histograms, it is necessary to identify the colours associated with each face, which are affected by several factors such as the pigmentation (varies among persons), concentration of blood in the face and the number and characteristics of the light sources.

To characterise the distribution of the skin-colour in the HSV colour space, a set of face images was extracted from the MPEG-4 test sequences, from the MPEG-7 test sequences and also from the IST face database. The selected images cover a considerable set of skin-colours and lighting conditions (both outdoor and indoor). Over 80 images were used as the training set for a Caucasian race skin-colour distribution model. As it can be expected, this model will not work well e.g. for black people and a new model would have to be specified. Some examples are presented in Figure 2.

Figure 2 - Examples of faces used to study the distribution of face colour

To generate the histograms for the HSV components of the faces used in this study, all faces where first segmented with a user-controlled tool. The H, S and V values were computed for each face pixel and the histograms were built. The total number of skin-colour pixels was used to normalise all histograms and the Hue histogram was clipped since most of the values are within the range [-30; 60]. Analysing the histogram for the Hue (see Figure 3 a), it is clear that the skin-colour pixel values of this component are confined to a limited range, mainly between 0 (Red) and 50 degrees (Yellow).



a) b) c)

Figure 3 – Normalized histograms for a) Hue; b) Saturation; c) Value

The Saturation histogram shows that the samples taken are mainly between the range of 20% to 70%; this suggests that the colours are neither deeply saturated nor deeply unsaturated. Finally, the Value or brightness component histogram ranges from approximately 25% to the maximum value of 100%.

2.1.3 Colour Segmentation

The goal of colour segmentation in this context is to identify the skin-colour regions present in the image. If the various image regions have distinctive and well-separated colours, this task is straightforward, but if the image has many regions with colours very similar to the skin-colours the problem may become quite challenging.

The method proposed here for the colour segmentation belongs to a class of techniques called pixel-based segmentation. These techniques rely on *a priori* knowledge of the distribution of the skin-colour. The histograms of Figure 3 suggest that the Hue component has the most discriminative power in extracting the face regions since it is confined to the smallest range. However this component is unreliable in two situations. First, when the brightness in a scene is low, i.e. for images with low lighting or with shadows, and second when the face regions under consideration have low saturation components, i.e. for images with specular reflections due to non-uniform illumination. To take into account these effects, the Value and Saturation ranges mentioned above have been slightly changed. : for the Value component, a lower threshold of 35% was selected and for the Saturation component lower and higher thresholds of 20% and 85%, respectively, were selected. Similar results have been determined in [2]. In conclusion, the skin-colour model proposed in this paper is defined by a polyhedron in the HSV space, which is obtained by applying the following five thresholds:

$$Thr_{hue}^{1} \le H \le Thr_{hue}^{2} \quad \text{with} \quad Thr_{hue}^{1} = 0, \quad Thr_{hue}^{1} = 50$$

$$Thr_{sat}^{1} \le S \le Thr_{sat}^{2} \quad \text{with} \quad Thr_{sat}^{2} = 85\%, \quad Thr_{sat}^{1} = 20\%$$

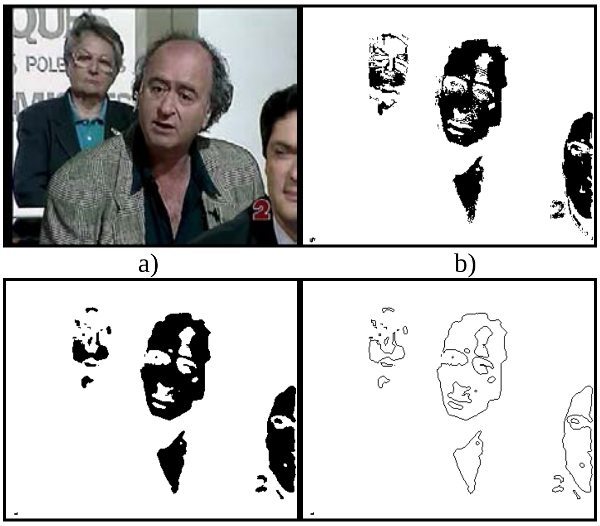
$$V \ge Thr_{value} \quad \text{with} \quad Thr_{value} = 35\%$$

Pixels within the polyhedron are classified as skin-colour pixels while the others are classified as belonging to the background. This segmentation method can be implemented using five comparisons per pixel, so it is computationally inexpensive to perform. An example of colour segmentation is included in Figure 4 b). The output format of this stage is a binary image that identifies the colour candidate regions to be further processed. The skin-colour model presented above was built for Caucasian people, but it can be easily extended to deal with a wider range of skin-colours if a database of representative faces is available.

2.1.4 Post-Processing: Median Filtering and Contour Extraction

The post-processing block refines the shape of the skin-colour regions extracted and represents them in a format suitable for the shape-processing module. This block performs median filtering and contour extraction.

Median filtering is a very common filter used to remove binary noise in images and video sequences [3], due to his good performance. In this case, the median filter is applied to the output of the colour segmentation in order to remove unwanted isolated pixels, while preserving the general spatial appearance. The output of this stage is presented in Figure 4 c). The contour extraction step identifies the boundaries of the candidate face regions, by removing the candidate face pixels for which all neighbours are also candidate face pixels. A representative output is presented in Figure 4 d).



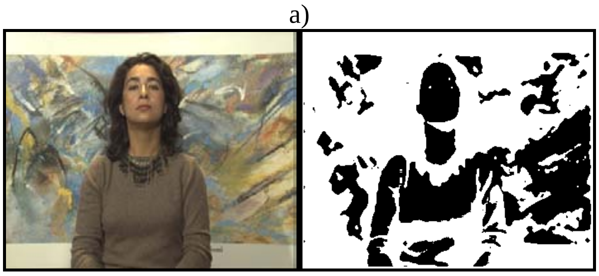
a) b) c) d)

Figure 4 – Colour segmentation and post-processing steps: a) Original image; b) Colour segmentation; c) Median filtering; d) Contour extraction

2.1.5 Limitations of the Skin-colour Segmentation Algorithm

Figure 5 shows results of the colour segmentation process for two images: one is extracted from a sequence of the MPEG-7 test data set, *Le Balladin du Monde Ocidental*, and the other from the MPEG-4 test data set, *Silent Voice*. For the first image, the colour segmentation algorithm performs rather well in extracting skin-coloured pixels, including the hands. However, for the second one, the face is extracted along with several other regions that belong to the painting behind the woman.

a)



b)

Figure 5 – Examples of skin-colour segmentation: a) *Le Balladin du Monde Ocidental;* b) *Silent Voice*

After extensive analysis of the results of the proposed colour segmentation algorithm, it is possible to conclude that its performance may degrade under several conditions:

1) **Non-face skin-colour regions** – If there are regions of the background or belonging to other parts of the body (e.g. hands, neck, and arms) that exhibit similar colour characteristics as face regions they will be wrongly detected as face region(s).

2) **Low Contrast** – If the contrast between the background and the skin-colour is not high, parts of the background may be classified as face.

3) **Varying lighting conditions** – The number and characteristics (intensity, distance) of light sources may pose a problem in detecting face regions. For example, when a “bright spot” or a dark shadow falls on the subject’s face, the colour segmentation process may fail since some parts of the face have their colour characteristics very much changed.

In view of these problems, additional processing allowing to improve the discrimination capabilities between the face and the other skin-colour regions is needed. This method is presented in the following and consists in the shape analysis of the detected skin-colour regions.

2.2 Shape Analysis

This section describes the shape analysis of the colour segmentation module output to improve the detection of faces. The proposed algorithm is based on the fact that the outline of the human face can be generally described as being roughly elliptic in nature. This feature remains unchanged whatever the size, position, expression and orientation of the face. The proposed shape analysis module is based on the idea of fitting ellipses to the results of the colour segmentation, selecting some candidate ellipses, and deciding after if they correspond or not to faces by applying certain constraints. The ellipse detection algorithm and the face selection algorithm are iteratively applied a pre-defined number of times.

2.2.1 Ellipse Detection Algorithm

A robust algorithm to detect elliptical shapes in images based on the Randomised Hough Transform (RHT) [6] is used in this paper. The algorithm allows the detection of regions whose exact shape is unknown, but for which there is enough *a priori* knowledge to build an approximate model of the ellipse. This method has many desirable features, as it is able to recognise partial and slightly deformed ellipses, and it may find several occurrences of ellipses during the same processing step. Thus, the task of ellipses detection consists in three major steps:

1) *Find the ellipses present in the image, limited to 5.*

2) *For each ellipse found, determine the parameters that describe it, namely the coordinates of the centre of the ellipse, the major and minor axis and the orientation.*

3) *Classify each pixel in the image as belonging to one of the ellipses found or as a pixel that does not belong to any ellipse (background pixel).*

In order to make the presentation of this algorithm as precise as possible and to facilitate the understanding of the framework build to detect faces, it is necessary to define what is an ellipse in this context. An ellipse c_i is expressed by a parametric function $f\,(\mathrm{a},\mathrm{d})$ with $\mathrm{a} = [\alpha_1,\ldots,\alpha_5]$ containing five ellipse parameters and $\mathrm{d} = \mathrm{d}(x,y)$ being the coordinates of a contour point of the ellipse. A δ-band of c_i is defined as a subset of the space, around the ellipse c_i with width δ. So, the elliptical shape mentioned before is defined as an elliptical band around the ellipse subject to $f\,(\mathrm{a},\mathrm{d}) = 0$.

Using the RHT, the ellipse detection algorithm works in the parameter domain. For each triplet of pixels in the image, a set of 5 ellipse parameters is computed (centre of the ellipse, the major and minor axis and the orientation) that vote in the parameter space; these votes are recorded in a structure called the accumulator. The idea is to accumulate evidence for different ellipse equation parameters in the "parameter space" for each triplet of pixels in the "image space”. The most likely set of parameters will define the candidate ellipses.

This type of algorithm requires the existence of an **accumulator**, since the accumulated votes will determine where ellipses are found. The choice of voting triplets is done by **random sampling** controlled by a **convergence mapping**, and the votes are stored in the **accumulator** after a **quantization**. Once ellipses have been detected, a finer description of their parameters is obtained by applying **clustering procedures** to all the votes that contributed to the corresponding quantised values. Finally, a check whether the detected shapes really correspond to an ellipse is performed (**Face Detection Constraints**) as described in the next section. More details about the various steps of the ellipse detection algorithm are given below:

Random sampling – The implementation of RHT consists in a series of random trials. For each trial, a triplet of pixels is randomly sampled from the image and then mapped using convergence mapping into a point $a \in \Re^n$ in the parameter space by solving the equation of the ellipse. An important restriction to the triplets of points randomly chosen is that they remain in the same skin-colour contour connected component. This restriction is not present in the original formulation of the RHT but is very useful to make the algorithm more robust to noise in the accumulator space, caused by choosing pixels that belong to different regions in the image.

Once the parameters of the ellipse are estimated, they are quantized and the accumulator is incremented. In Figure 6, a three-dimensional graphic of an accumulator is presented containing the accumulated values for the centre of the ellipse.

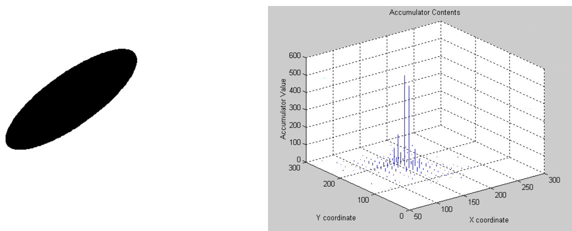


Figure 6 – Image of an ellipse and the corresponding accumulator for the centre of the ellipse

Convergence mapping – This property guarantees that the parameterisation of the ellipse can be solved analytically. For some combinations of pixels, infinite solutions may exist (e.g. when the three pixels randomly chosen lead to linearly dependent equations) or no solution exists at all (e.g. when three collinear pixels are chosen). It may also happen that the three randomly chosen pixels can only be adapted to a hyperbola or parabola. In these cases, the 3 pixels are simply discarded and another triplet is randomly sampled. If the parameterisation is successful, there are a couple of obvious constraints to the computed values (centre, major and minor axis and orientation) as follows: i) *the centre of the ellipse has to remain within the image, and ii) the major axis has to be smaller than the vertical dimension of the image.*

Clustering – In order to achieve salient peaks in the accumulator, a coarse quantization of the ellipse parameters is applied. However, this leads to errors in the definition of the ellipse's position, size and orientation. In order to have a more accurate location of these ellipses, it is necessary to store the parameters before quantization for all voting triplets. Then, in a final stage, more precise values for the ellipses are calculated in the 5D parameter space, within the selected quantization interval.

Among all elements that have been accumulated, the parameter array is searched for peaks; this means the most likely set of parameters and thus the most likely ellipse locations. The top 5 maximum peaks are taken as candidate ellipses for further processing by the face selection algorithm, according to some chosen face constraints.

2.2.2 Face Selection Algorithm

Since some of the candidate ellipses do not always correspond to faces because face-specific constraints were not applied, it is necessary to filter the candidate ellipses according to some adequate criteria, finally selecting the faces detected. In this context, a 'face' is a candidate ellipse that fulfils all the face selection criteria previously set.

In order to perform the face selection, a procedure is applied to each candidate ellipse, which consists in the checking of the following criteria:

Angle sector criterion *–This criterion evaluates the amount of contour skin-colour processed pixels under the δ-band of each candidate ellipse in order to check if it corresponds to a face according to the angle sector criteria. The candidate ellipse is divided into a limited number of angular sectors and for each sector it is checked if it contains contour skin-colour processed pixels or not. The ratio between the number of sectors containing at least one pixel and the total number of sectors gives an indication if the candidate ellipse is adequately covered by contour pixels in its extension. This ratio is compared to an ellipse acceptability threshold.*

Orientation criterion – *This criteria evaluates if a real ellipse is under consideration by calculating for each contour that the ellipse δ-band contains, the angle between the normal to the ellipse at that pixel and the line that links it with the ellipse centre. If the angle is less than a specified threshold this pixel is valid. A ratio between the number of valid pixels and the total number of pixels has to be inferior to another specified threshold.*

Aspect ratio criterion – *Given the geometry of the human face, it is assumed that the aspect ratio between the major and minor axis of the ellipse must fall within a specified range. Experimentation with the same images used to obtain the distribution of the skin-colour in the HSV space, led to the conclusion that the aspect ratio (height/width) has an average value of 1.5. Analysing the variation of the face aspect ratio of all faces in the IST face database, and taking into account that the colour segmentation is not perfect, a range of 1.25 to 2 for the aspect ratio is accepted.*

Every candidate ellipse fulfilling the face selection criteria above described is finally detected as a face.

2.2.3 Global Face Detection Architecture

Due to the random nature of the ellipse detection algorithm, the ellipse detection-face selection modules have to be cycled a number of times. After a face is detected in each ellipse detection-face selection iteration, the corresponding contour skin-colour processed pixels are removed. When the 5 candidate ellipses for one iteration have been processed by the face selection algorithm, the next iteration of the ellipse detection algorithm may start. In our case, the results are taken after 4 iterations of the ellipse detection-face selection combination without detecting faces (stop condition). Figure 7 presents the ellipse detection-face selection cycle architecture in the form of a block diagram; this cycle is run until the stop condition is fulfilled. The output of the 2-stages automatic face detection algorithm is an XML (eXtensible Markup Language) description that contains the information (descriptors in MPEG-7 language) needed to describe faces in an audiovisual sequence, namely the location of keyframes, number of faces per keyframe and the ellipse parameters for each face detected.

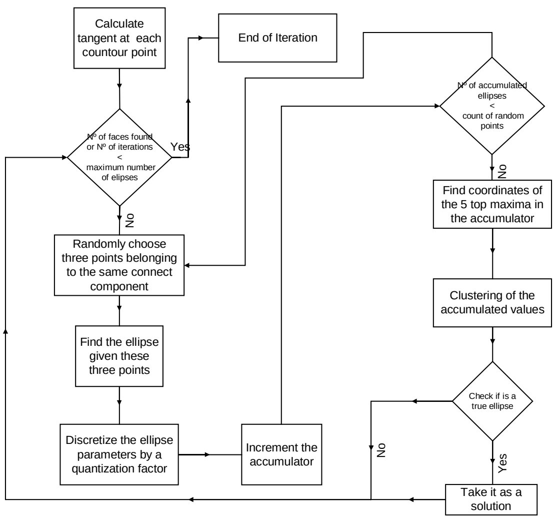


Figure 7 – Block diagram of the shape analysis algorithm

3 Automatic Face Detection Results

In order to evaluate the performance of the face detection system proposed, several images were chosen with no restrictions on the pose, expression, illumination, background and number of faces. These sequences were chosen from the MPEG-4 and MPEG-7 test data, in the CIF format: *Carphone* (a video-telephony type of sequence), *Silent Voice* (a sequence containing gestual language), *Polemic* (an excerpt of a talk show), *Le Balladin du Monde Ocidental (*an excerpt of a theatre play), *Camilo e Filhos* (an excerpt of a sitcom)**,** and *Riscos* (an excerpt of a soap opera).

The algorithm parameters used were:

1) Number of triplets randomly chosen to detect an ellipse: 4 times the number of contour skin-colour processed pixels in the image.

2) Quantization values for the ellipse parameters: 15% of the maximum dimension of the image for the centre, 10% for the major and minor axis, and 36 degrees for the orientation.

3) Threshold value for accepting a candidate ellipse as a face: 65% for the angle sector criteria and 50% for the orientation criteria.

Figure 8 shows the results for head and shoulders type of images, namely for a video-telephony sequence and for an anchorperson shot. In both images, the system was able to correctly detect the position of the face.

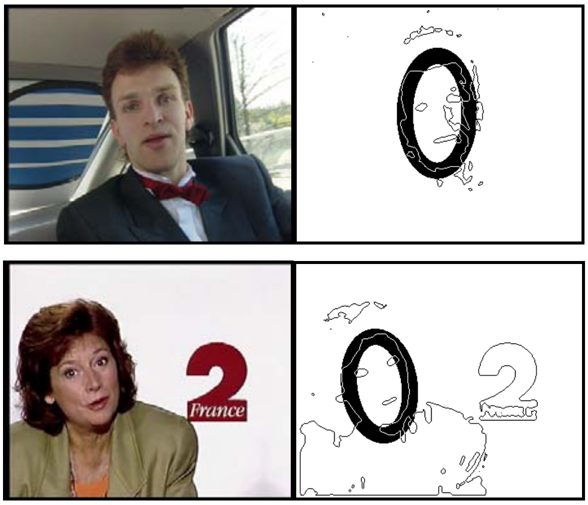


Figure 8 – Head and shoulders images: original, contour colour segmentation and shape-based facial detection

In Figure 9, results for body and face type of images are presented. In these images, the faces are smaller than in the head and shoulders type images and very often the hands are present in the image. In both cases, the system correctly detected the faces.

Figure 9 – Body and face type images: original, contour colour segmentation and shape-based facial detection

In Figure 10, results for face images in a talk show are presented. These include several faces within the same image. In the case, the two faces facing the camera are correctly identified. The faces that are behind are eliminated due to their small size and bad illumination. Notice that the faces are detected, even if the colour segmentation output consists in several non-connected regions corresponding to the same face.

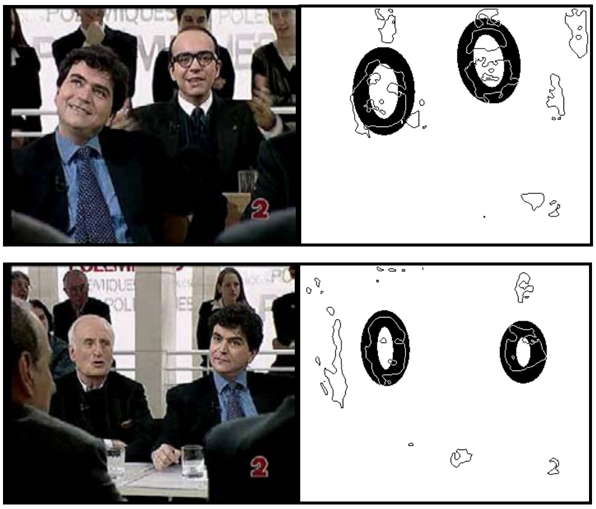


Figure 10 – Talk show face images: original, contour colour segmentation and shape-based facial detection

Figure 11 shows results for two images containing three faces each. In both cases, the system has failed to detect all the faces present. In the first image, and due to illumination problems, one face is not detected since the output of the colour segmentation module could not find a sufficiently large region of support. The other face was missed because it is half clipped, thus not fitting the ellipse shape. For the second image, since the size of the two missing faces is too small and the illumination poor, the colour regions extracted do not give enough information for the shape processing module to succeed.

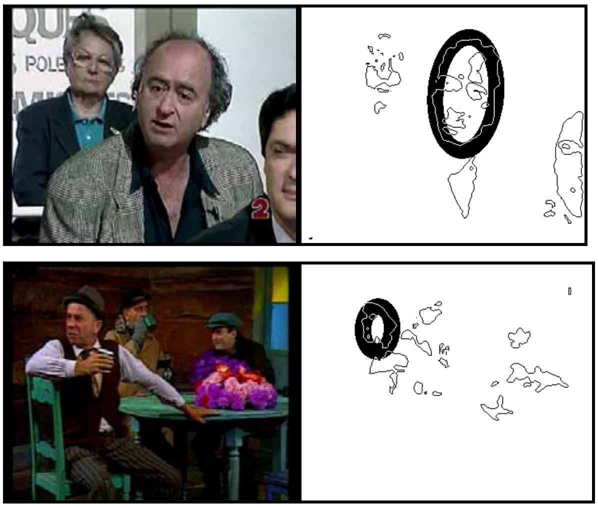


Figure 11 – Images with 3 faces: original, contour colour segmentation and shape-based facial detection

Figure 12 shows some failure results of the face detection system. In both these images, the colour analysis module outputs rather bad results in terms of the number, size and deformation of the candidate facial regions extracted. For the first image, the strong shadow present in the face, leads to a heavily deformed output of the colour-processing module, making impossible the adaptation of an ellipse. For the second image, several background regions are also extracted and thus the shape-processing module adapts ellipses to non-face regions that have ellipsoid shapes.

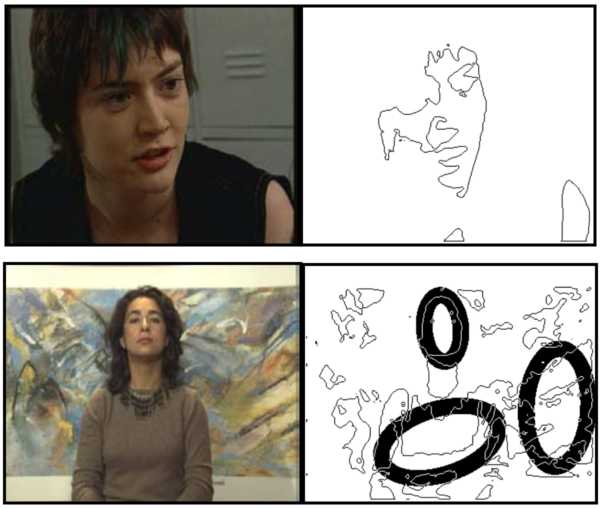


Figure 12 – False negative and false positives: original, contour colour segmentation and shape-based facial detection

In conclusion, two major types of errors may occur:

1) **False negatives** – If the face region(s) extracted are heavily deformed due to insufficient chromatic information or occlusions, the shape-processing module eliminates them since is not able to find any acceptable ellipsoid form.

2) **False positives** – If the number and area of the regions extracted that belong to non-facial areas are significant, the noise present in the accumulator of the ellipse detection module is high and consequently the ellipses may be adapted to other skin-colour regions than faces.

As already mentioned, the problems shown above highlight the importance of complementing the automatic face detection tools with some user interactive tools to allow the refinement of the detection results creating the description. Of course, it is desirable that the automatic detection tools give the best possible results and the user interactive editor is used as less as possible.

4 Interactive Face Description Editor

This section presents the interactive framework where the user has the possibility to interact with the automatic face detection results in order to correct them as desired. This combination of automatic and interactive tools leads to a system that is able to: i) browse the keyframes per shot or individually and visualize the output of the automatic face detection algorithm and ii) in case of failure of the automatic face detection algorithm, it is possible to edit the description and correct the results. The integration of these two types of tools allows the detection of any faces, in any type of images, without any constraints. Depending on how difficult the images are, the final detection results will more or less require the use of the face detection editor. In the system proposed, the editor parses the XML description previously generated by the automatic tools and provides means for the user to interact with it, notably by:

1) Open and Save face descriptions;

2) Visualize the face ellipses in the description, superimposed on each keyframe individually or per shot;

3) Delete and Draw faces in a keyframe;

4) Modify the size, orientation and position of an ellipse in a keyframe.

After the changes, always visible in the GUI (Graphical User Interface), a new XML description is generated. The GUI has been developed in a PC-Windows environment, using Java to ensure maximum flexibility and portability for other environments.

5 Conclusions and Future Work

The results of the automatic face detection algorithm proposed in this paper are promising for unconstrained images, detecting a wide range of face poses and sizes without degrading the performance of the system. Also, other body parts and background regions not connected to the face are eliminated most of the times, notably if they do not form a shape similar to an ellipse. When other skin-colour regions are connected to the face, such as the neck or a coat, the shape processing system adapts the ellipse to the part of the shape that is more ellipsoidal, leading sometimes to localization errors. The algorithm developed is also capable of successfully detecting ellipsoidal shapes that are composed of several skin-colour regions (see Figure 10). A typical situation where the facial region is divided into separate regions is in the presence of partial occlusions, due to either dark glasses or facial hair. These last two cases are an advantage of this system in comparison with algorithms that evaluate each region individually [5]. The automatic face detection tool is however sensitive to strong shadows and bad illumination, to the characteristics of the background and to some image acquisition conditions. These limitations are mainly associated to the colour segmentation module. After several experiments with different images, the scenarios where the automatic face detection tool has best performance are news bulletins, talk shows and interviews.

When errors occur, the interactive face detection tool provides a solution for correcting the results by editing the output description of the automatic face detection tool. Possible directions of future research are to use better colour models and better colour segmentation techniques in order to improve the results of the colour segmentation stage, retrieving better segmented regions, especially when changes occur in illumination and acquisition conditions of the video to analyse.

Acknowledgements

The authors acknowledge the support of the European Commission under the ACTS project DICEMAN.

References

[1] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips, "Introduction to Computer Graphics", *Addison Wesley,* 1994, pp. 413-416.

[2] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, "Automatic Location and Tracking of the Facial Region in Color Video Sequences", *Signal Processing: Image Communication*, Vol. 14, pp.359-388, March 1999.

[3] A. Jain, "Fundamentals of Digital Image Processing", Prentice-Hall International Editions, 1989, pp. 246-249.

[4] C. H. Lee, J. S. Kim, and K. H. Park, "Automatic Human Face Location in a Complex Background Using Motion and Color information, *Pattern Recognition*, Vol. 29, Nº 11, pp. 1877-1889, 1996.

[5] J. Terrillon, M. David, and S. Akamatsu, "Automatic Detection of Human Faces in Natural Scenes Images by Use of a Skin Color Model and of Invariant Moments", *3rd International Conference on Face & Gesture Recognition*, April 1998, Japan.

[6] L. Xu and E. Oja, "Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms, and Computational Complexities", *CVGIP: Image Understanding*, Vol. 57, Nº 2, March 1993, pp 131-152.

[7] L. Xu, E. Oja, and P. Kultanen, "A New Curve Detection Method: Randomized Hough Tranform", *Pattern Recognition Letters*, Vol. 11, Nº 5, May 1990, pp 331-33

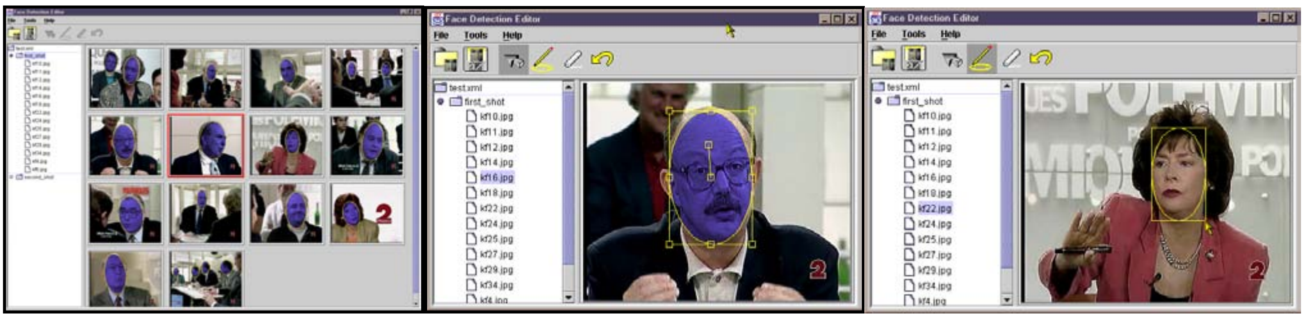


Figure 13 – The graphical user interface: a) browsing mode; b) editing mode; c) drawing mode