

A Colour Tracking Procedure for Low-Cost Face Desktop Applications

F.J. Perales, R. Mas, M. Mascaró, P. Palmer, A. Igelmo, A. Ramírez

Computer Graphics and Vision Group
Department of Computer Science
Universitat de les Illes Balears (UIB)
{paco.perales,ramon.mas,mascport,pere.palmer,angel.igelmo}@uib.es

Abstract.

In this paper we present an environment for the tracking of a human face obtained from a real video sequence. We will describe the system and discuss the advantages and disadvantages of our approximation. We mainly focus on the situation of the main attributes of the human face (eyes, eyebrows, nose and mouth). The tracking algorithm and the ulterior animation of the synthetic model must guarantee the real time response without the need of any additional markup of the actor. Due to the complexity of the process, we make an initial selection of the facial attributes involved without any efficiency or robustness loss. We define a probabilistic model of skin face area and we would like to track this region in the sequence of images. In parallel we propose additional criteria to search inside this tracked area main features in human face (as lips, eyes, eyebrows, nose, etc.). The tracking algorithm is based in a efficient implementation of continuously adaptive mean shift procedure (CAMSHIFT) and this process is improved also with the second step with feature detections. In this paper only we present the whole process, the tracking background criteria and lips detection procedure. The synthesis phase is out scope of this paper and we generate the facial animations parameters (FAP) as input to a compliant MPEG-4 facial animation engine (FAE). This system is designed as a computer interface for controlling commercial computer applications which include avatar or clones in real time.

1. Introduction

A lot of work has been devoted in the literature to the tracking of human face and which somehow contribute in the approach we present in this paper [1,2]. Having in mind the existing previous work, we want to perform a realistic animation of a synthetic actor from a real sequence of images captured using a low-cost low-performance video camera. We have mainly focused in the main attributes of the human face although we have also included additional information like blobs, templates and a relational graph of visual entities.

The tracking and animation techniques for the facial model must guarantee the real time requirement with the constraint of not using any kind of additional markup for

the human actor being tracked. However, we impose some constraints also to the kind of allowed occlusions of the regions of interest (ROI) as can be the hair on the eyebrows, the glasses or the beard. Due to the initial complexity of the problem we have used a previous manual selection (only in the first frame) of the face attributes involved without any noticeable loss of neither efficiency nor robustness of the system. We plan in future to do automatically this initialization step.

In the next sections we describe the architecture of our system and their different modules. Finally we conclude our work with some examples of face tracking and axis visualization.

2. System architecture

Figure 1 shows a global view of the system, where we can see each of the different modules and tasks interacting. The union nexus is a simple but easy-to-use and fully functional user interface. The interface mainly allows the visualizing and the capture of the real actor and the animation of the virtual avatar.

Each module is intended to have the greatest possible degree of autonomy to minimize the influence of local algorithm modifications. The main modules are:

- the capture and visualization of real images module
- the geometrical characteristics recognition and tracking module
- the synthetic visualization module

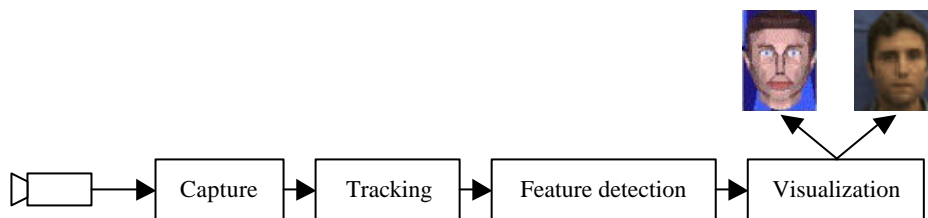


Fig. 1. The system architecture

The first module takes the input of both video and audio from a low-cost video camera (USB interfaced, IEEE 1394, etc.) We have also tried a medium-cost SONY EVI-D31 Pan/Tilt Zoom camera to test the validity of the approximation. The capture and visualization module provides the input to the second module at the same time than visualizes the input sequence. Both processes must be fully synchronized. The second module is the kernel of detection, tracking and recognition process. We have implemented a tracking method based on basic visual entities. At the same time, we have designed a tracking algorithm for the shape of the human actor based on color information using the HSL (hue, saturation and luminance) model although images are captured in RGB. The theoretical model is statistic; the symmetry axis and inertial moments are computed over the color distribution (hue) in order to be able to identify

in every moment, the precise position and orientation of the human face. There are some constraints affecting the position and orientation of the face respect to the camera but when the system detects a loss of tracking it asks for a recovery position. We assume that the person is in a frontal plane respect to the camera.

The module for the visualization of the synthetic model represents the motion of the tracked points and regions. An easy and practical extension in order to make an open system would be to code the resulting motion using the FAT (facial animation tables) of the MPEG-4 standard. The synthetic actor or clone uses the information extracted by the geometrical characteristics computation and tracking module from the real images to automatically reproduce the motion of the real actor. The clone can be easily used as a videoconferencing low-band interlocutor. Only few points are needed to update the clone modification between frames.

To minimize the errors in the segmentation and in the tracking of the face, the registration process takes place in a controlled illumination background. The camera is supposed to be in near front view and the variations in the orientation of the human actor with respect to the image plane must be moderate (less than 5 degrees). In case a high variation is produced, the tracking algorithm can gather position and orientation errors, which can drive the system to an erroneous state. This constraint is constantly controlled using a predefined threshold.

3. The tracking module

One important step of the system is the tracking of the ROI in the face of the person. To get real time tracking we have to solve multiple problems as the solutions can not be as computationally expensive as those usually used in the face recognition systems. One of the main troubles deals with the great variety in the physical appearance of the studied subjects: skin color, eyes color, beard, glasses, hair and many other variables. Moreover, the facial attributes can be totally or partially occluded or in shadows which make the edges of the features indistinguishable. All this set of complications lead to an increase in the difficulty to recognize the face and the ulterior tracking of the ROI considered.

To minimize the effect of those problems in our system, the key points considered and their associated information are manually identified in an initialization or learning process. The initialization process is trivial and adjustable depending on the level of precision required by the application. We can select areas in the tracked attributes (the face, the nose, the mouth, the eyes and the eyebrows). We have stored a logical structure between face attributes which allows the validation of the initialization phase to help the tracking and recognition. Once the areas have been identified using boxes, the system captures the color and edges characteristics of the regions.

The tracking subsystem is divided in several phases according to the primitives or attributes considered. The main consideration of the tracking system is that the face has an homogeneous color distribution (skin distribution). To take profit of this fact we can define robust parameters and probabilistic distributions which have the necessary properties to allow efficient and noise tolerant in the adjustment process.

The algorithm implemented is based in a robust non-parametric technique (Mean shift) which augments the gradients density to find the nearest dominant mode [1].

We have detected that the algorithm works well when we have a near constant illumination and the color of the face occupies most of the images with no similar colors in its neighborhood. Although not useful for the tracking of the attributes of the face themselves, the algorithm has provided to be good for a global positioning and orientation of the face, as can be seen in the results shown in figure 2. Notice in green colour the dynamic window used, and the centroid and axis on the face.

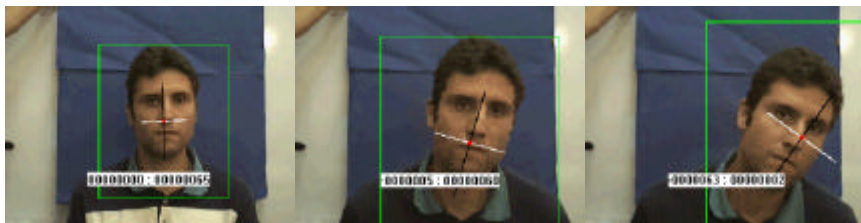


Figure 2. Three steps of the tracking process: moving the face laterally

To track the mouth, the eyes and the eyebrows we have had to adapt the color distribution to match the particularities of the minor regions. The synthetic essays have shown to be encouraging. In the real images we experience relatively frequent losses of the goal, if a hand or similar skin region move over face region. But the system recover the main region after this overlapping state. Moreover we would like to recognize the facial features on face so we need additional control criteria. Let see how we deal with the mouth. We have some procedures for eyes and eyes brows but are not yet fully tested. So we concentrate in this paper in lips study.

3.1 Detection of the mouth

The mouth is one of the most difficult facial attributes to analyze and track. Its shape is very versatile and multiple facial muscles take part in its motion. As an added problem, the beard, moustache, tongue and teeth can be disturbing appearing and disappearing properties. To simplify this casuistic we will make the assumption that no beard nor moustache are present and we may include structural information about the anatomy of the person. Lets consider that:

1. The upper teeth are fixed in the skull and than its relative position is constant.
2. The lower teeth move downwards from its initial position following the rotation of the lower mandible.
3. The shape of the moth depends on the lower maxilar motion.
4. Facial muscles can contribute to the shape of the moth.

The initial selection of the main areas of the tracked features makes the initial control points or lines more easier. With this consideration in mind, the detection of

the moth must take into account the detection of each of its parts and their partial occlusions. So we base our algorithm in the following main points:

1. ROI selection in the initialization process.
2. Filtering and space color translation (RGB to HSL).
3. Lips edges detection.
4. Lips center line detection. Minimum luminance.
5. Detection of the corners of the mouth. Finding of the mean point and of the Cupid point.
6. Determination of the equivalent lower point.
7. Interpolation using a spline to get the full curve of the lips.
8. Finding the best elements combination to adjust the found edges. Initially, mouth closed.
9. Definition of the similitude function from the candidates and the obtained image. Maximize this function.

The main idea is to compute the edges of the vertical transitions of a line from the nose to the chin. We sample several lines and we verify the computed edges. Simultaneously we consider the HUE as in some cases the edge of the lower lip is difficult to detect. The final comparison is based in the three generated curves, which provides added robustness to the matching process, and considering the variation of the length of the moth in each basic position (closed or open). This can be seen in the figure 2.

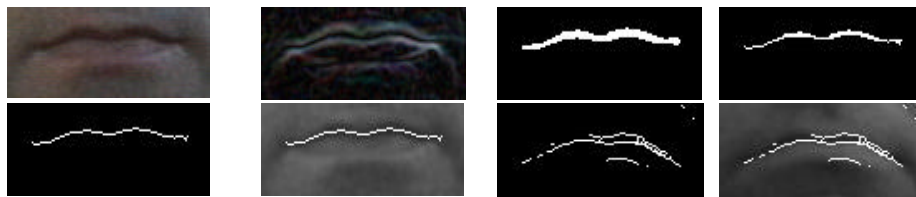


Figure 3: Original image, edges and mean line recognized of the mouth

The computed data are temporarily stored in the face structure. We are now working in a mouth database to create a model to allow better initial position detection. The model of the open mouth can be seen in the figure 4. When the mouth is closed, points 17 and 30, 20 and 27, 23 and 24 all superpose. From these characteristics the vertical curve will have a different shape depending on the status of the mouth:

- 1) Open mouth: the central contour is strong and the edges corresponding to the upper and lower lips are less pronounced respect to the skin of the face. Teeth are not visible.
- 2) Closed moth: the teeth are visible, contours are more visible in the lip-teeth and teeth-lip unions. The extern transitions lip-face are softer. In some cases, some of the intern transitions are not detectable when the lip occludes teeth.

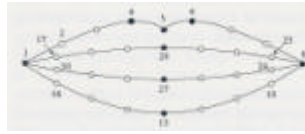


Figure 4. The model of the mouth

We use fourteen points for both, the model of the open mouth and of the closed mouth despite some of them may superpose when closing the mouth. The relative distance between the coincident points determine the degree of opening of the mouth.

4. The interaction module

Even though the main part of this project is the tracking of the human face, it is necessary to design a graphical user interface in order to provide an efficient interaction with the system. The most important parts are (see figure 4):

- A window where the user face got from the camera will be shown. On this window the key points will be displayed, giving an easy to handle feedback.
- A system with which the user can manually select and adjust the key points considered.
- Some button where the user can activate the manual calibration, start and stop the tracking and finish the application.
- A message area where the system can show information and give indications to the user. Also with a light signal mechanism with which to mark the general situation.

The user interface must be as simple as possible, only the valid actions are available. The message area is the only one with text. The other parts follow the icon-based paradigm to avoid the complexity. The aspect is closer to the multimedia systems actually in use. The whole interface is configurable using skinning techniques. Before to use the system it is necessary select an image source from the list of all capture devices detected by the operating system; the user can select one of them from a list in a modal window. After this first step the main window appears. The message area shows a welcome text and the user image obtained from the selected camera is displayed in the feedback window. The only action the user can do at this time is calibrating the system. The calibration requires another auxiliary window where the user can select one by one the different areas to fit in the feedback window. Once the user has picked over one area, he can select over his own face the corresponding region. The user image keeps static during this procedure in order to make an easy selection. The areas of interest are: the whole face, eyes, mouth and nose.

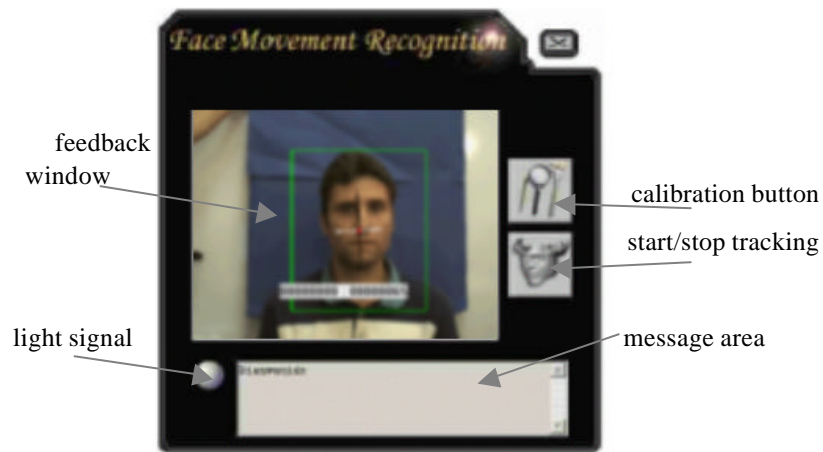


Figure 5: the main interaction window

After this step, when all areas have been selected and correctly marked over the user image, the light signal changes to green colour if the areas are valid or to red colour if there is an invalid area. When the calibration window is closed the user image becomes dynamic once again and before the start the tracking procedure the user should fit his face over the areas defined in the calibration procedure.

5. Conclusions and future work

In this paper we have presented a new approach of human face tracking and facial feature detection. We would like to reach all the process in near real time. The tracking process is based on an improved version of CAMSHIFT procedure and feature detection is considered on eyes, eyebrows and lips. Due to space, we only present the lips procedure and some initial results. Also the tracking procedure is presented and fully tested. Also, our system is completed with an interactive procedure interface for non experts end users. Our future goal is to extend the system, to cope with the analysis of all of the facial features. Also, at the moment we export the points detected to a FAE MPEG-4 compliant animation engine. We are working in a stereo version to improve the errors in Z measurement.

6. References

1. Computer Vision Face Tracking for Use in a Perceptual User Interface. Gary R. Bradsky. Microcomputer Research Lab, Intel Corporation. 2001.2.
2. Robust object location detection – Automatic head contour detection. Multimedia communications research laboratory. Bell Labs. 2000.3.

F.J. Perales, R. Mas, M. Mascaró, P. Palmer, A. Igelmo, A. Ramírez

3. Rapid Design of MPEG-4 Compliant Animated Faces and Bodies.
Erich Haratsch. Technical University of Munich. 1997.
4. Intel © Image Processing Library & Open Source Computer Vision Library. Reference
Manuals. Intel Corporation. 2001.5.ISO/IEC JTC1/WG11 N1902.Text for CD 14496-2
Visual Fribourg meeting, November 1997.