International Journal of
Applied Mathematics,
Electronics and Computers
ISSN: 2147-8228 http://ijamec.atscience.org

Advanced Technology and Science

Original Research Paper

Tracking feature points of the talking face

Noureddine Cherabit*[1], Amar Djeradi[1], Fatma Zohra Chelali[1]

Accepted 05th May 2016

Abstract: This paper presents a method of tracking points on a speaking face and the reconstruction of a 2D face model from speaking descriptive vectors followed. After capturing a video of a talking face using a CCD camera, his facial expression changes in a video sequences. The essential of this work is to track the feature points on the face using the Lucas-Kanade method, then use descriptive vectors up to be exploited for the reconstruction of a model of speaking face. The tracking method used is based on the theorem of Lucas-Kanade and facial reconstruction based on the method of delaunay triangulation.

Keywords: *talking head, automatic tracking points, coding of visual speech, delaunay triangulation, Lucas-Kanade.*

1. Introduction

Tracking points on the face image from a video sequence is the object to provide possibility for computers to generate descriptive vectors on the position of these points over time.

The input of such system is a set of points on the face of the first frame of a video sequence (chosen on the face) and the output is to present the vectors describing the new position points followed by succession the video images over time. The major challenge of tracking is to identify the most similar region to the original points surrounded regions of face.

We're interested in our work to track feature point on the face using Lucas-Kanade methode. A lot of methods have been developed in this research area including 2D and 3D reconstruction. All these methods used the principal of motion estimation (ME).

Several studies have been implemented to tracking video and 3D facial reconstruction [21], [22], [23].

Motion estimation (M.E.) has been studied by many authors like Wang, Weiss, Adelson and Bergen [5, 6, 7], Heeger [8], Horn and Schunk [9] Weber and Malik [10], Wu and Kanade [11], and more recently Leduc [12], Bernard [13], Lee [14].

It should be noted first that the motion estimation is a quantification of the simple motion, directed by translation vectors pixel block or object, or more complex, implementing methods of calculating trajectory in curly or not systems. Applications of M.E. are especially reducing temporal redundancy for compression and scene analysis. Several methods return primarily in M.E. like:

- The mapping of blocks.
- The spatio-temporal filtering (with or without compensation).
- Measuring dense field or optical flow.

Vaillant [2] proposed a method of faces surface reconstruction by active stereovision. Following the principle of active stereovision, his work is based on the projection of a pattern on the observed surface, and then uses well points chosen on this basis to achieve the 3D reconstruction of the surface. Indeed, a fairly comprehensive review of the work on this subject can be found in [3].

The major classes of proposed methods are [2]:

- Direct use of the face image: after some processing, the image is supplied to a discrimination algorithm which has been previously trained with a database. The major drawback of this type of method is sensitive to lighting conditions: the image of a face will be significantly different.
- Use of a profile image: the image acquired is an image from the face profile. The contour of this profile is extracted. It is then possible to extract some features which can be provided to a discrimination algorithm (the objective in the article). Although priori, the profile image of a face is less discriminating than the image of the face itself. The advantage of this strategy is that the data used are independent of the lighting because seeks separate regions on the succession of images, and can easily be normalized to be made independent of the face position and orientation in the image.

Our article is presented as follows:

Section II presents the video acquisition, section III describes the lucas-canade algorithm for optical flow estimation, and section VI describes the application of monitoring points on a face. Section V shows the simulation results obtained.

2. Video Acquisition

In our work for video acquisition, we used a digital camera canon MV530i, equipped with internal analog to digital converter for digitizing the video signal. An acquisition card firewire IEEE 1394 high speed (400 Mbps), computer Pentium3 2.1 Ghz clock frequencies for video acquisition and processing. We used a projector and a transparent grid to project on speaker's face.

In general, current systems used for video acquisition is presented in Figure 2.

[1] *Speech communication and signal processing laboratory Faculty of Electronic engineering and computer science University of Science and Technology Houari Boumedienne (USTHB). Box n°:32 El Alia, 16111, Algiers, Algeria.*

* *Corresponding Author: Email: nourchera@yahoo.fr*

Note: This paper has been presented at the International Conference on Advanced Technology&Sciences (ICAT'14) held in Antalya (Turkey),

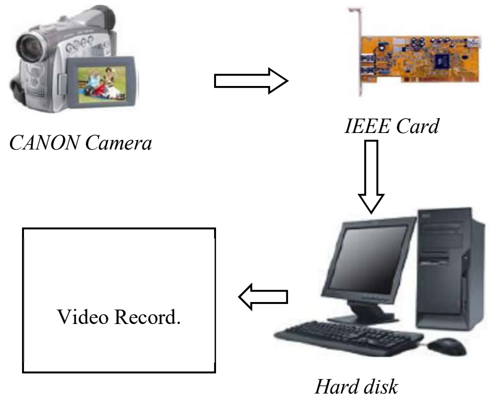


Figure 1. Block diagram of the experiment used for video capture.

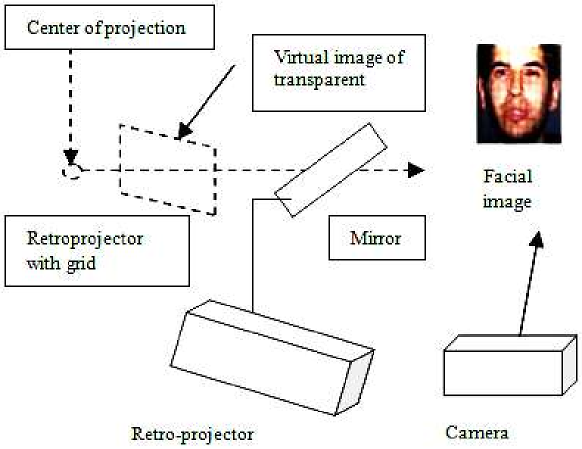


Figure 2. Projection of a transparent grid on face.

We selected feature points on the face in order to follow the reconstruction of a 2D model of the face.

1) The selected points are sufficient to obtain a good representation of the facial surface; we have to detect about 48 points (standardized face key points) on the face speaker.

2) The points are distributed evenly over the face.

3) Points are easily detectable in the image: the idea is the use of a grid formed by a set of projected lines (figure 2).

4) The points can be identified: we will need to match the pattern points and those of the pattern extracted from the image. One way to identify these points is to make a label (number) on each point.

3. Lucas-Kanade Method

In computer vision, the Lucas-Kanade method is a differential method widely used for optical flow estimation developed by Bruce Lucas [1]. They assume that the **flow** is essentially constant in a local neighborhood of the pixel under consideration and solve the basic equations of optical flow for all pixels in the area by the least squares criterion. By combining information from several neighboring pixels, the Lucas-Kanade method can often solve the inherent ambiguity in the optical flow equation.

3.1. Concept

The Lucas-Kanade method assumes that the movement of the image content between two frames is small and almost constant in a neighborhood of the point P to study. Thus, the optical flow equation can be assumed true for all pixels in a window centered at p. namely, the local flow of the image is a vector (V_x, V_y) satisfy the following equation:

$$\begin{aligned} I_x(q_1)V_x + I_y(q_1)V_y &= -I_t(q_1) \\ I_x(q_2)V_x + I_y(q_2)Vy &= -I_t(q_2) \\ \ldots\ldots\ldots\ &\ldots\ldots\ldots \\ \ldots\ldots\ldots\ &\ldots\ldots\ldots \\ I_x(q_n)V_x + I_y(q_n)V_y &= -I_t(q_n) \end{aligned}$$

Where: q_1, q_2, ... q_n are the pixels in the window, and Ix (qi), Iy (qi), It (qi) are the partial derivatives of the image with respect to the x, y position and t is time, measured at the point q_i and the current time.

These equations can be written in matrix define by: Av=b, where:

$$A = \begin{bmatrix} I_x(q_1) & Iy(q_1) \\ I_x(q_2) & Iy(q_2) \\ \ldots\ldots\ldots\ldots & \\ \ldots\ldots\ldots\ldots & \\ I_x(q_n) & Iy(q_n) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ -I_t(q_n) \end{bmatrix}$$

3.2. Principle of the lucas-kanade ALGORITHM

The Lucas-Kanade algorithm is one of the most popular algorithms iteratively followed; try to minimize the difference between the image and a deformed model. The technique can be used for tracking and motion estimation.

The basic idea of the Lucas-Kanade algorithm is based on three assumptions [1]:

- Constant brightness: the brightness is preserved between two successive images.
- Temporal persistence or "small movements»: the movement from one frame to the other must be "small".
- Spatial coherence: the neighbours of a point must remain the same.

One of the problems of the algorithm is due to the use of small windows. If the movements are great, there is a risk to move the points outside the local window and by the inability of the algorithm to find them. So we chose to use windows larger than 15 pixels.

Using significant size windows is not a solution to the problem. And the use of large windows is contrary to one of the three assumptions of the algorithm precisely the "spatial coherence". The solution is given by the hierarchical or pyramidal approach (Figure 2) [18]. This method meets to all the assumptions of the Lucas-Kanade algorithm. The first window in the upper part of the pyramid is smaller, than the following, but provides less detail than the following. Then, as we move the base of the pyramid over the level of detail provided increases.

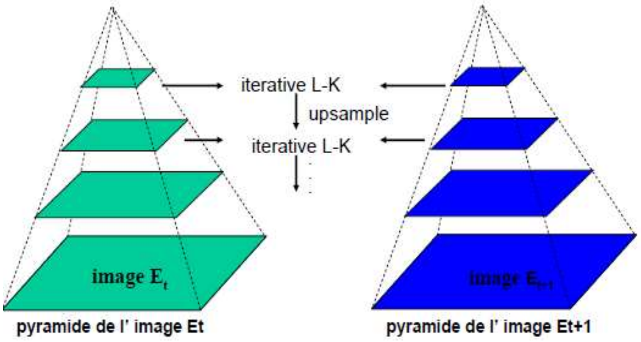


Figure 3. Hierarchcal aproach [18].

4. Application of Monitoring Points On A Face

Using the Canon camera, the recording was made with 25 frames/seconde, of size 576×720 pixels. The data are then transformed into computer using IEE1394 card. Our corpus consists of a repetition of video sequence produced by five speakers.
We choose then the best video in terms of significant change in facial expression that are important for our application.

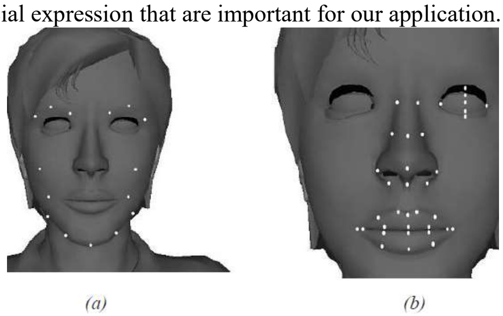


Figure 4. Key points of a standardized face [4].

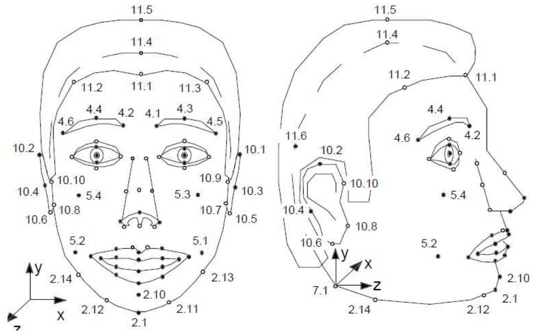


Figure 5. Key points of a standardized face (detailed) [4].

5. Feature Extract

We took several videos. Each speaker repeat the sentence (on est le 11 juin 2011, il est 15 heure: it's June 11, 2011, it is 15 o'clock) more times.

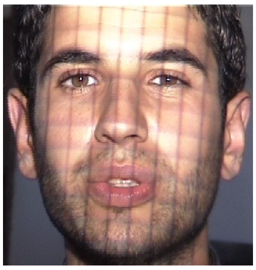


Figure 6. First vidéo image.

As described in figure 7, we chose 48 points to generate the entire surface of the face, especially around the mouth and the area around the eyes describing the important changes for a talking face.
We create a region around a point situated at the center of this region. To obtain good results, we choose the region size of 21×21 pixels.
The following figure shows the 48 selected points of the first frame.

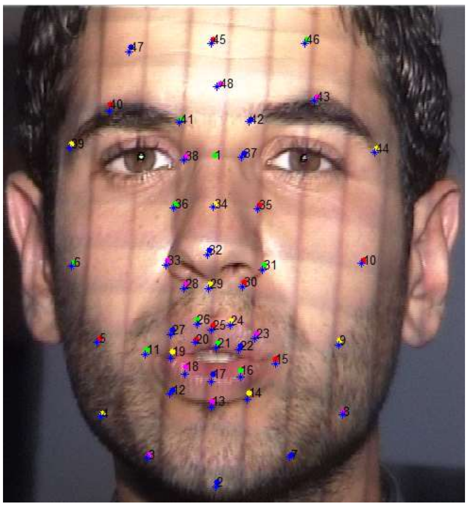
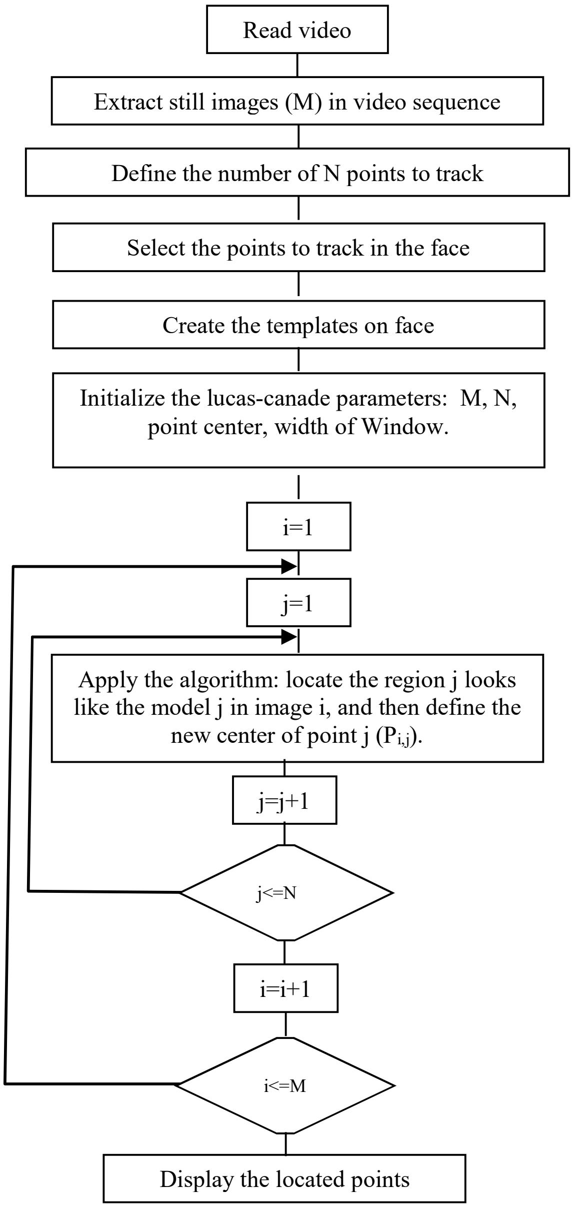


Figure 7. The 48 points selected on the first video image of face.

Figure 8 shows the general diagram describing the Lucas-canade applied for our algorithm.



Read video

Extract still images (M) in video sequence

Define the number of N points to track

Select the points to track in the face

Create the templates on face

Initialize the lucas-canade parameters: M, N, point center, width of Window.

i=1

j=1

Apply the algorithm: locate the region j looks like the model j in image i, and then define the new center of point j ($P_{i,j}$).

j=j+1

j<=N

i=i+1

i<=M

Display the located points

Figure 8. Diagram of the lucas-canade algorithm application.

6. Results Obtained for Tracking Points On Face

The system scans the video frame by frame. In parallel, it displays tracking points corresponding to each image. In this operation we compute the Euclidian distance between the points situated between the eyes (point number 1) and the points tracked on the face. In progress of this operation, this takes few minutes to generate all tracking points of the video. Figure 9 presents the tracking movement of 48 points on the 6th image.

Figure 10 presents tracking movement of all points in the video sequence. For example, taking the point number 14 located on the lower lip, the figure 11 present Tracking movement of this point on 20 images. From this figure, this point undergoes a linear change between the first image and the 10th. After this last image, the point undergoes an important displacement.

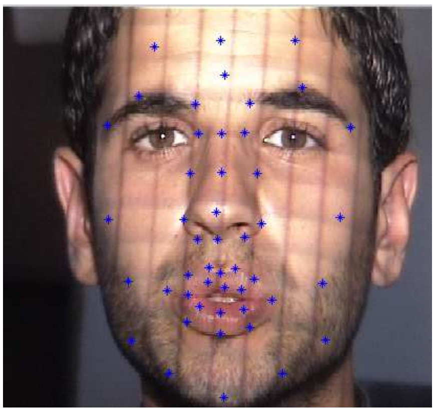


Figure 9. Tracking of 48 points on the 6th image.

Figure 10 presents tracking movement of all points in the video sequence. The tracking error is calculated by method used in [24]. Tracking error result is presented on figure 11.

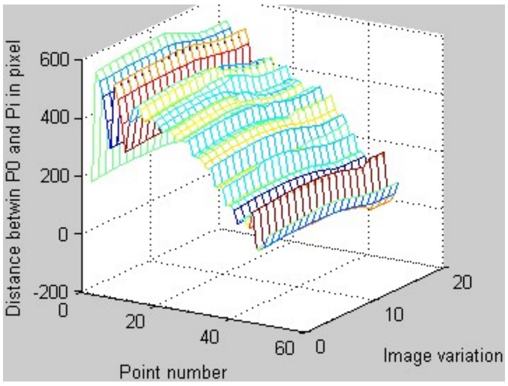


Figure 10. Movement tracking of all point in video sequence.

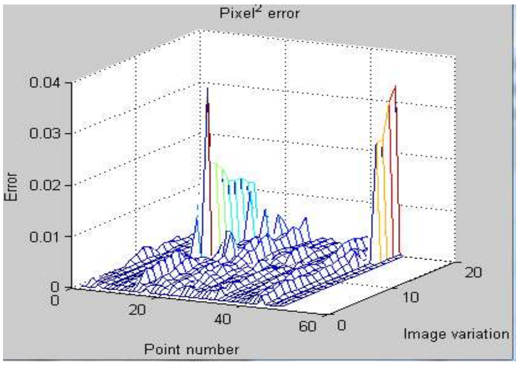


Figure 11. Error movement tracking.

Figure 12 presents tracking movement of point 14 located on the lower lip in video. The tracking movement error of this point is presented on figure 13.

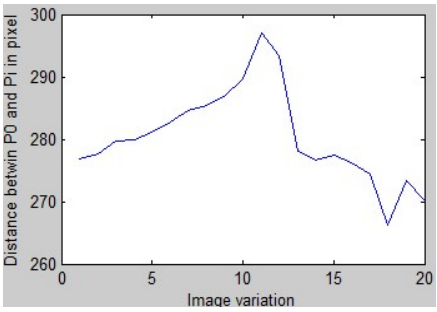


Figure 12. Movement tracking of point number 14 on 20 images.

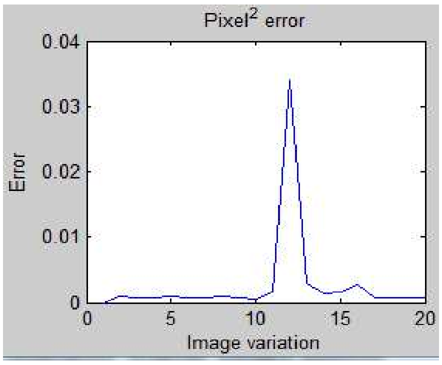


Figure 13. Error of movement tracking for point 14.

7. Principle of face reconstruction using delauny triangulation.

The principle of the face model reconstruction is based on the delauny triangulation (DT).

Called triangulation of a set of points in the plane P, the data of a set of triangles whose peaks points P such that two distinct triangles have their interiors empty intersection, and maximizes the number of edges [19], [20].

Let P be a set of points in the plane, the Delaunay triangulation is the triangulation that maximizes the lexicographical order on the angles [19].

The following figure shows an example of reconstruction of a face model from vectors motion tracking points on the face using the method of delauny triangulation.

So, we can reconstruct the model face in 2D using only the point described earlier.

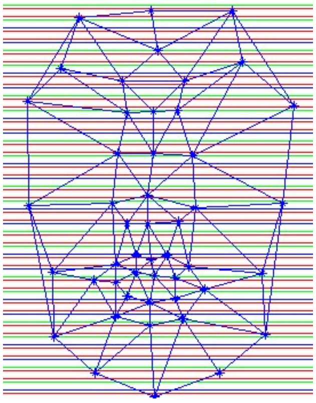


Figure 14. Application of the Delaunay triangulation on points tracked in an image.

8. Conclusion

We propose in this paper a method of tracking feature points on face. Using a video acquisition and a grid, it makes easier to locate key points on the studied face.

A lot of points have been used in order to have good performance to represent correctly the face shape using MPEG-4 key points.

After selection of key points on the first image of the video, then we apply the lucas-kanade algorithm, an optical flow estimation, which assumes that the flow is essentially constant in a local neighbourhood of the pixel under consideration. To verify the position i on the picture we have make a label on all selected points, where good results were obtained by using a window of 21 pixels. The points inside the lip contour haven't been correctly tracked because of the loss information during opening and closing lip.

This work tracks the movement of facial talking face, so we create a 3D matrix containing the position (x, y) of each point of standard MPEG-4 at different moments. The second task of this work is to generate a mesh grid to reconstruct models of faces using the method of triangulation delauny. The results are satisfactory. From a neutral face, we can generate new positions of key points.

The future work is to analyse several visual phonemes then create a database containing the position of each point in order to generate a talking face. Our method can be used in analysis/synthesis of talking face.

Acknowledgements

I thank Prof DJERADI Amar for his help and orientations. The work is part of a project of MMC in LCPTS laboratory at USTHB University.

References

[1] Lucas, Bruce david. Generalized image matching by the method of differences. Phd thesis, university microfilms international. 1985.

[2] R.vaillant et I. surin. Face Reconstruction Through Active Stereovision. Traitement du Signal 1995.

[3] Ashok Samal and Prasana Iyengar. Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognition, 25(1) 65-77, 1992.

[4] N. Sarris and M. G. Strintzis. 3D Modeling and Animation: Synthesis and Analysis Techniques for the Human Body. Published in the United States of America by IRM Press. 2005.

[5] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of vision. Journal of Optical Society of America, A2: 284{299, February 1985.

[6] J.Y.A. Wang and E.H. Adelson. Spatio-temporal segmentation of video data. In Proceedings of SPIE on Image and Video Processing II, 2182, pages 120{131, San Jose, February 1994.

[7] Y. Weiss and E.H. Adelson. Perceptually organized em: A framework for motion segmentation that combines information about form and motion. Technical Report 315, MIT Media Lab Perceptual Computing Section TR, 1994.

[8] D.J. Heeger. Optical flow using spatiotemporal fillters. International Journal of Computer Vision, vol 1, pp 279-302, 1988.

[9] B.K.P Horn and B.G. Schunck. Determining optical flow. Artificial Intelligence, 17 :185-204, 1981.

[10] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. Computer Vision, 14 :67-81, 1995.

[11] Y.T. Wu, T. Kanade, J. Cohn, and C.C. Li. Optical flow estimation using wavelet motion model. In Sixth International Conference on Computer Vision, pages 992-998 Narosa Publishing House, 1998.

[12] J.P. Leduc, F. Mujica, R. Murenzi, and M.J.T. Smith. Spatiotemporal wavelets: A group-theoretic construction for motion estimation and tracking. Siam Journal of Applied Mathematics, 61(2) :596-632, 2000.

[13] C.P. Bernard. Ondelettes et problèmes mal posés : la mesure du flot optique et l'interpolation irrégulière. PhD thesis, Ecole Polytechnique, CMAP, Centre de Mathématiques Appliquées., Palaiseau, France, Novembre 1999.

[14] C.M. Lee. Joint source-Channel Coding Tools for Robust Transmission of Video Sequences; Application to H.263+ and H.264. PhD thesis,

[15] Mathew A. Turk and Alex P. Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3(1): 72-86, 1991.

[16] A. Pentland, B. Moghaddam, O. Starner, T. Oliyide and M. Turk. View-Based and Modular Eigenspaces for Face.

[17] D. Beymer. Face Recognition under Varying Pose. In Computer and Pattern Recognition, June 1994.

[18] S. Mechkour, R. Boesch et R. Oprea. Interfaces Multimodales. Université de Fribourg – Département d'Informatique. 2009.

[19] Ma. Pillet. Triangulation de Delaunay. Ecole Normale Supérieure de Cachan, antenne de Bretagne, avril 2010.

[20] S. W. SLOAN. A fast algorithm for constructing Delaunay triangulations in the plane. Department of C~vil Engineering and Surveying, The University of Newcastle, NSW 2308, Australia, Adv. Eng. Software, 1987, Vol. 9, No. 1.

[21] R. Cutler and Larry S. Davis. Fellow.Robust Real-Time Periodic Motion Detection, Analysis, and Applications, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8, AUGUST 2000.

[22] M. Julien Richefeu, Détection et analyse du mouvement sur système de vision à base de rétine numérique. DOCTEUR de l'UNIVERSITE PARIS 6. Décembre 2006.

[23] M. gotla and Z. huang. A minimalist approach to facial reconstruction. 1999.

[24] S. baker AND I. Matthews. Lucas-Kanade 20 Years On: A Unifying Framework. International Journal of Computer Vision 56(3), 221–255, 2004.