Tridimensional pose estimation of a person head.

Elisa Perez Berenguer¹, Carlos Soria², Oscar Nasisi², Vicente Mut²
¹ Gabinete de Tecnología Médica.
² Instituto de Automática.
Facultad de Ingeniería- Universidad Nacional de San Juan, Argentina.
E-mail: eperez@gateme.unsj.edu.ar

Abstract. In this work, we present a method for estimating 3-D motion parameters; this method provides an alternative way for 3D head pose estimation from image sequence in the current computer vision literature. This method is robust over extended sequences and large head motions and accurately extracts the orientation angles of head from a single view. Experimental results show that this tracking system works well for development a human-computer interface for people that possess severe motor incapacity.

1. Introduction
The tracking of the face movement in an image sequences is a complex task in computer vision and has great application in the following areas: human-computer interface, face recognition and identification, virtual communication, virtual reality, graphical computation. The tracking of the face is defined as the process to estimate the parameters that modeling the position and orientation 3D of the face, in an image sequences [1].
The systems of face tracking must consider the rigid and no rigid movement of the face in images sequences. They can be classified in:
- Methods based on features, which extract image features and make a tracking of their movements from an image to another one. The image features are description such as regions, contours and landmarks. The tracking of regions very complicated that have small changes between images can cause very different segmentations between the images. On the other hand the tracking of curved contours is very difficult. The face trackers more used are those using landmarks; such landmarks can be: eyes, nose and mouth, among others [2],[3],[4].
- Method based on Optical Flow, which uses space and temporary partial derivates that consider the displacements of the gray pixels levels or blocks of pixels from an image to another. The optical flow is used to calculate the movement of the object projected onto the image and exist several methods for this estimate [5],[6],[7].
- Methods based on correlation are very used in objects tracking. They use the sum of the differences between a pattern and an area of pixels intensities in the image. The disadvantage of these methods is that they varying with the illumination conditions [8].
In this work a 3D tracking model of a face is implemented, which considers 6 parameters of translation and orientation in 3D in real time. The objective is to apply it in the development of
human-computer interface for people that possesses severe motor incapacity. A form of integrating these individuals to the society is to look for mechanisms of high technology that allow him to develop suitably and to feel inside the productive sector of the society. That is to say, to be able to provide to a person with this affection the appropriate ways so that it can communicate and to displace in their environment, replacing the function of some traditional devices as for example the keyboard, mouse, joystick, trackball, etc. Also its possible extension to the substitution of the device command to help the displacements of people with motor incapacity, for example a wheelchair. This paper is organized as follows: Section II present Methods. Section III presents some experimental results. Finally conclusions are given in Section IV.

1. Methodology.
The projection of the 3D movement of a rigid object in the time forms a movement field on the image.

1.1. Movement field of rigid objects.
The motion field is the 2D vector field of speeds of the image points, induced by the relative motion between the viewing camera and the observed scene [9]. The movement field can think like the projection of the 3D speed field on the image plane. In order to analyze the relation between the 3D world and the image plane, the coordinates system shown in Figure 1 is used.

![Figure 1. Coordinate system.](image)

The system shown in the Fig. 1 has the origin of coordinates at the focal point of the sensor, and the Z-axis is that it crosses through the origin and it is perpendicular to the image plane.

Let \( P = [X~Y~Z]^T \), a point in the 3D space, its projection on the image plane is given by:

\[
\begin{align*}
  i_x &= f \frac{X}{Z}, \\
  i_y &= f \frac{Y}{Z}
\end{align*}
\]

(1)

where \( f \) is the focal length. The relative movement between the point \( P \) in the world space and the camera can be described by:

\[
V = \frac{dP}{dt} = -T - \omega \times P
\]

(2)

where \( T = [T_x~T_y~T_z]^T \) is the translational component, and \( \omega = [\omega_x~\omega_y~\omega_z]^T \) is the rotational component. Because the movement is rigid, \( T \) and \( \omega \) are the same for any point \( P \).

In (2), \( V \) is the movement of the point \( P \) in the 3D space respect to this system world coordinates. This equation can be expressed as the function of these components:
To obtain the relationship between the velocity $P$ in the world space and the corresponding velocity $p$ in the image plane, the derivatives in both sides of (1) are taken, take in account that the projection of $p = [X \ Y \ Z^T]$ in the image plane is denoted by $p = [x \ y]^T$, obtaining an important equation sets. The movement field in the image plane is:

$$v = \frac{Z \cdot V - V_z \cdot P}{Z^2} \tag{4}$$

Reemplacing (3) in (4) we obtain the instant velocity components in the image plane like:

$$u = \frac{T_{z,y} - T_{x,y}}{Z} - \omega_y \frac{f}{x} + \omega_z \frac{f^2}{f}$$

$$v = \frac{T_{z,x} - T_{x,y}}{Z} + \omega_x \frac{f}{y} - \omega_z \frac{f^2}{f} + \omega_y \frac{f}{f} \tag{5}$$

To visualizing better the equation above, we write it in matrix form:

$$\begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \frac{1}{Z(x,y)} \begin{bmatrix} -f & 0 \\ 0 & -f \end{bmatrix} \begin{bmatrix} T_{x,y} \\ T_{y,x} \end{bmatrix} + \frac{1}{Z(x,y)} \begin{bmatrix} \omega_x & \omega_z & 0 \\ \omega_x & \omega_z & 0 \\ \omega_y & \omega_z & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{6}$$

In (6) two terms can be distinguished: the first one describes the component of the movement field of the image due to the translation of the object, and as it is seen, it depends on the depth inverse in each point, and the second is the component of the movement field of the image due to the rotation of the object, which is independent of the depth. Once obtained the relation between the three-dimensional space and its projection on the image plane, the concept of optical flow is presented.

2.2. Determining the Optical Flow.

Now the problem is to consider the movement field from a sequence of images, this is from the world space and temporary variations of the brightness intensity of the image. This estimation not always is possible to obtain it satisfactorily, even in noise absence, therefore the concept of optical flow or image speed arises: that it is the bidimensional field of apparent speeds in the image plane associated with the variation of patterns of brightness intensity in an image. That is to say, the field can be due to the movement of the observer, the movement of the objects in the scene or the apparent movements.

If we denominated $l = (x, y, t)$ to the brightness of the image in the point $(x, y)$ in the image plane at the instant $t$, and in addition one assumes that this remains constant under the movement, we can say that:

$$\frac{dl}{dt} = 0 \tag{7}$$

Applying the partial derivate we obtain:
Partial space derivatives from the brightness intensity of the image are simply the space gradient of the image, \( \nabla I \), and the time derivates, \( \frac{dx}{dt} \) \( \frac{dy}{dt} \), are the components of the movement field \( v \). Rewriting the previous equation we obtain the fundamental equation of the optical flow or movement restriction equation which is:

\[
I_x u + I_y v + I_t = 0
\]

(9)

Analyzing the (9) one sees that we have two unknown parameters that are \((u, v)\) and only have one equation by each point of the image, therefore only the speed component in the direction of the brightness gradient can be determined.

2.3. Point Features.
In order to obtain the values of \((u, v)\) firstly the point features obtained. The chosen point features do not belong to any geometric organization observed in the image of the face, the selected features are the corners, since in them the gradient vector of the image will have an elevated magnitude, and we have this way the certainty that the flow vector has better estimate.

Of the image of the face the space gradient calculates \( \begin{bmatrix} E_x & E_y \end{bmatrix} \), and considering a point \( p \) and a generic neighbor region \( Q \) of \( p \) of size of 5x5, the matrix \( C \) is defined like:

\[
C = \begin{bmatrix}
\Sigma E_x^2 & \Sigma E_x E_y \\
\Sigma E_x E_y & \Sigma E_y^2
\end{bmatrix}
\]

(10)

where the sum is made in the region \( Q \). As \( C \) is symmetric can be diagonalize in the following way:

\[
C = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}
\]

(11)

Interpreting of geometric way the values of the eigenvectors of \( C \) we can determine if we have a pixel that represents a corner. We know that a corner is formed by the intersection of two strong contours, therefore the eigenvectors are the quite great, and all those pixels that have a value of greater intensity to the minor eigenvalue their coordinates belong to a corner. By this way the features points were obtained.

2.4. Estimation of the Head Movement.
The procedure to calculate the 3D position of the head consists of the following steps:
- Calculate of the point features in the image \( I_t \), using (11):
With n calculated point features, as showed in fig. 2, it is come to consider the values of optical flow, implementing the method of [10].

- Estimation of optical flow: implementing the method of [10], this uses the Kalman filter to accomplish a data fusion step, to compute a robust optical flow. The optical flow calculated, is applied to the features points selected (it is assumes that only there is movement of the head in the image, all the background remains static), see fig. 3. We suppose that the obtained images of face have background that is not in movement, the only movement that takes place is the head). The estimation is made between two successive frames, \( I_t, I_{t+1} \).

- Estimation of the movement 3D: Recovery of the parameters of successive rotation and translation between two frames using movement field in 2D. The three steps have been explained in the previous sections, now will be explained how the calculated optical flow is related to the 6 parameters to calculate. As it were explained in section 2.1 the projection of the movement in 3D in the movement field 2D is given by (3). Supposing that the movement of the rigid body between two frames is infinitesimal, we can directly replace the angular velocities by the value of the angles; to do this the variation of the value of the angles does not have to be up to 5° [11]. This is totally possible since webcam in where the capture time between two frames is about to 100 msec, a very small time to make great movements of the head. Then we can rewrite (3) by the following way:

\[
\begin{align*}
V_x &= -T_x - \beta Z + \gamma Y \\
V_y &= -T_y - \gamma X + \alpha Z \\
V_z &= -T_z - \alpha Y + \beta X
\end{align*}
\]

(12)

Where \( \alpha \) is the rotation angle respect to the x axis, \( \beta \) is the rotation angle respect to the y and \( \gamma \) is the rotation angle respect to the z axis.

The (7) can be rewritten in the following form:
\[
\begin{bmatrix}
  u_i \\
  v_i
\end{bmatrix} =
\begin{bmatrix}
  f & 0 & -x_i & \frac{x_i y_i}{f^2} & \frac{f + x_i^2}{f} & -y_i \\
  0 & f & -y_i & \frac{f + y_i^2}{f} & \frac{x_i y_i}{f} & x_i
\end{bmatrix}
\begin{bmatrix}
  t_x \\
  t_y \\
  t_z \\
  \alpha \\
  \beta \\
  \gamma
\end{bmatrix}
\]  

(13)

where \( t_x = \frac{T_x}{Z}, \) \( t_y = \frac{T_y}{Z} \) and \( t_z = \frac{T_z}{Z} \), by this way we have a system of equations of the way:

\[
Ax = b
\]  

(14)

where \( A = \begin{bmatrix}
  f & 0 & -x_i & \frac{x_i y_i}{f^2} & \frac{f + x_i^2}{f} & -y_i \\
  0 & f & -y_i & \frac{f + y_i^2}{f} & \frac{x_i y_i}{f} & x_i
\end{bmatrix} \), \( b = \begin{bmatrix}
  u_i \\
  v_i
\end{bmatrix} \) and \( x = \begin{bmatrix}
  t_x \\
  t_y \\
  t_z \\
  \alpha \\
  \beta \\
  \gamma
\end{bmatrix}^T \).

The system is resolved using recursive least mean square, minimizing the error:

\[
\|Ax - b\| = E
\]  

(15)

By this way, one has an approach of robust estimate of the parameters of the head 3D.

2. RESULTS
The experiments were realized with real image sequences captured at 10 frames per second with a webcam. The size of the image is 320x240 pixels. In Fig 4 is shown the optical flow estimate from the image sequence (a). In Fig. 5, the evolution of \( \alpha, \beta \) and \( \gamma \), corresponding to the image sequence of the face (a) is shown. In Fig 6 is shown the optical flow estimate from the image sequence (b). The Fig. 7 depicts the evolution of \( \alpha, \beta \) and \( \gamma \), corresponding to the image sequence of the face (b).

![Figure 4. Image sequence (a)](image-url)
Figure 5. 3D rotation parameter estimation results for the image sequence (a).

Figure 6. Image sequence (b)

Figure 7. 3D rotation parameter estimation results for the image sequence (b).

3. CONCLUSIONS
In this work, we have presented a method for estimating 3D motion parameters; this method provides an alternative way for 3D head pose estimation from image sequence in the current computer vision literature. We have shown that this method is robust over extended sequences and large head motions, the image sequence captured by the webcam it does not have any preprocessing and accurately
extracts the orientation angles of head from a single view. Experimental results show that this tracking system works well for development a human-computer interface for people that possess severe motor incapacity.

**Acknowledgments**
This work was funded in part by Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET).

**References**

[1] S. Z Li. and A. K. Jain. Handbook of Face Recognition, Springer Science+Business Media (2005)

[2] F. Dornaika and J. Ablberg. Face and facial feature tracking using deformable models. International Journal of Image and Graphics, vol4, pp 499-532, 2004.

[3] J.B.Chen, Tiddeman, E. “Robust Facial Feature Tracking System”, Advanced Video and Signal Based Surveillance. AVSS 2005, IEEE Conference, 445-449, 2005.

[4] A. Yilmaz and M.A. Shah, “Automatic feature detection and pose recovery for faces”, The 5th Asian Conference on Computer Vision, Melbourne, Australia, 2002.

[5] B. K.P. Horn, and Shunk, B.G. “Determining optical flow,” Artificial Intelligence, vol 17, pp185-203, 1981.

[6] K.I. Diamantaras, T., Papadimitriou, M.G. Strintzis and M. Roumeliotis., “Total least squares 3-D motion estimation,” Image Processing, vol 1, 923-927, 1998.

[7] T. Chun- Jen, , N.P. Galatsanos and A.K. Katsaggelos, “Optical flow estimation from noisy data using differential Techniques,” International Conference on Acoustics, Speech and Signal Processing, Phoenix, AZ,USA, pp 3393-3396, 1999.

[8] G.D Hager and P.N. Buelhumer, “ Real Time Tracking of Image Regions with Changes in Geometry and Ilumination,” IEEE Conference on Computer Vision and Pattern Recognition, pp 403-410, 1996.

[9] E. Trucco and A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998.

[10] C. M. Soria, R. Carelli and M. Sarceinelli-Filho, Optical Flow Estimation Using Data Fusion, VI Simposio Brasileiro de Automacao Inteligente. Bauru, pp. 259-264, 2003.

[11] Y. Zhu, and Fujimora,K. “ 3D Head Pose Estimation with Optical Flow and Depth Constraints”, Proceedings of the Fourth Intenational Conference on 3D Digital Imaging an Modeling, Computer Society IEEE 2003.