Depth Independent Facial Movement Estimation

Haibo Li and Robert Forchheimer
Department of Electrical Engineering,
Linköping University, S-581 83 Linköping, Sweden

Abstract

A theoretical framework on 3D facial movement estimation for model-based image coding is presented in this paper. 3D facial movement estimation is a difficult problem to solve because both facial motion parameters and depth parameters are required to estimate simultaneously. When the conventional approaches are used to handle this problem, initial estimates of the depth parameters must be provided. The performances of these approaches are seriously affected by the accuracy of the initial estimates of the depth parameters. In this paper we present a depth independent facial movement estimation method using the subspace technique. In this method a constraint equation related to only the rotation motion component is built. The rotation motion can be recovered from this equation without any knowledge on facial depth parameters. Once the rotation motion has been recovered the translation motion and depth parameters can be easily obtained by an LS method. Due to no knowledge on facial depth parameters being needed, a better performance of facial motion estimation can be expected.

1 Introduction

3D facial movement estimation plays a very important role in model-based image coding[2][4]. The main difficult in 3D facial motion estimation lies in the fact that both facial motion parameters and depth parameters are required to estimate simultaneously. This makes the motion estimation here become a high-dimensional (more than six) nonlinear optimization problem. Obviously, directly solving the nonlinear optimization problem is very impractical. Several approximate approaches have been proposed to handle this problem[5][6]. A typic work had been done by K.Aizawa, et al [6]. In their work, the depth is initially given a rough estimate based on the adjusted general facial wireframe model. The motion parameters are then estimated using the rough depth estimate. Once the motion parameters are obtained, the rough depth estimate is further refined. These two steps have to be iterated in order to obtain a good solution. The main problem with this approach is that it is difficult to give an exact initial estimation of the facial depth. The errors in the estimate of the facial depth maybe lead to the divergence of the iteration process. Therefore, the performance of facial motion estimation is seriously affected by the initial depth estimate.

Motivated by the work of D.Heeger and A.Jepson[12], a depth independent facial movement estimation method using the subspace technique is proposed. In this method a constraint equation related to only the rotation motion component is built. The rotation motion can be recovered from this equation without any knowledge on facial depth parameters. Once the rotation motion has been recovered the translation motion and depth parameters can be easily obtained by an LS method. Due to no knowledge on facial depth parameters being needed, a better performance of facial motion estimation can be expected.

2 Facial Movement Model

Facial movement is a highly nonrigid motion, which contains not only the head motion but also the facial expression change. Based on the CANDIDE model[10] — a parameterized face model, facial movement can be modelled as follows[4]

$$\mathbf{s}' = \mathbf{R}\mathbf{s} + \mathbf{T} + \mathbf{R}\mathbf{E}\mathbf{\Phi} \tag{1}$$

where $\mathbf{s}= (x, y, z)^T$ is a position vector of arbitrary point in the face, after facial movement it moves to a new position $\mathbf{s}'= (x', y', z')^T$. $\mathbf{R}$ is called the rotation matrix. In the real visual communication case, considering that the frame rate is relatively high with respect to the motion of the face, the rotation matrix can be approximately represented as follows

$$\mathbf{R} = \mathbf{I} + \begin{pmatrix} 0 & \Omega_z & -\Omega_y \\ -\Omega_z & 0 & \Omega_x \\ \Omega_y & -\Omega_x & 0 \end{pmatrix} \tag{2}$$

Where $\mathbf{I}$ is the identity matrix; The Ω_x, Ω_y, and Ω_z are angular velocities about the x, y, and z axis, respectively.

$\mathbf{T}$ is the translation matrix.

$$\mathbf{T} \;=\; [T_x, T_y, T_z]^T \tag{3}$$

Where T_x, T_y, and T_z are the three velocity components of the translation motion.

m facial expression movement parameters are collected in the vector $\mathbf{\Phi} = (\phi_1, \phi_2, ...\phi_m)^T$. These real valued movement parameters are computed from the larger set of AU's (Action Unites) described in [10]. The $3 \times m$ matrix $\mathbf{E}$ determines how a certain point $\mathbf{s}$ is affected by $\mathbf{\Phi}$.

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ e_{31} & e_{32} & \cdots & e_{3m} \end{bmatrix} \tag{4}$$

Equation (1) is a general facial movement model. For some specific application situations, the complex movement model can be simplified. We now list the alterations of facial movement model in several special cases:

Case I: Φ_i=0, no facial expression change occurs in the two successive frames. The facial movement is caused by only the head movement. In this case the facial movement model can be simplified as

$$\begin{aligned} x' &= x + \Omega_z y - \Omega_y z + T_x \\ y' &= y - \Omega_z x + \Omega_x z + T_y \\ z' &= z + \Omega_y x - \Omega_x y + T_z \end{aligned} \tag{5}$$

Case II: $\Phi_i \Omega_j \approx 0$. When the change in facial expressions between two successive frames is not too larger, and the rotation motion is also small at the same time the facial motion can be viewed as an affine motion

$$\begin{aligned} x' &= x + \sum_{i=1}^{m} e_{1i}\phi_i + \Omega_z y - \Omega_y z + T_x \\ y' &= y + \sum_{i=1}^{m} e_{2i}\phi_i - \Omega_z x + \Omega_x z + T_y \\ z' &= z + \sum_{i=1}^{m} e_{3i}\phi_i + \Omega_y x - \Omega_x y + T_z \end{aligned} \tag{6}$$

This affine motion model has been successfully used in the facial motion tracking system[4].

Case III: the general case, which is suitable to a larger change in facial expressions:

$$\begin{aligned} x' &= x+\sum_{i=1}^{m} e_{1i}\phi_i+\sum_{i=1}^{m} e_{2i}\phi_i\Omega_z-\sum_{i=1}^{m} e_{3i}\phi_i\Omega_y+\Omega_z y-\Omega_y z+T_x \\ y' &= y+\sum_{i=1}^{m} e_{2i}\phi_i-\sum_{i=1}^{m} e_{1i}\phi_i\Omega_z+\sum_{i=1}^{m} e_{3i}\phi_i\Omega_x-\Omega_z x+\Omega_x z+T_y \\ z' &= z+\sum_{i=1}^{m} e_{3i}\phi_i+\sum_{i=1}^{m} e_{1i}\phi_i\Omega_y-\sum_{i=1}^{m} e_{2i}\phi_i\Omega_x+\Omega_y x-\Omega_x y+T_z \end{aligned} \tag{7}$$

Which model is chosen depends on the problem at hand. No matter which model is employed, the common objective is to estimate facial motion parameters (Ω_x, Ω_y, Ω_z, T_x, T_y, T_z, $\phi_1, \phi_2, ...\phi_m$), as well as depth parameters z_i.

3 Motion Parameter Estimation

In this section we discuss how to estimate facial motion parameters and depth parameters from the optical flow field or directly from the spatiotemporal derivatives of image intensity.

3.1 Optical Flow Field Induced by 3D Facial Motion

It is assumed that the geometrical projection from the three dimensional space onto the two dimensional image plane can be approximated by the orthogonal projection (the facial motion estimation under perspective projection is considered in [4]). We further assume the focus length $f = 1$. Then the projection relationship becomes

$$\begin{aligned} X &= x \\ Y &= y \end{aligned} \tag{8}$$

where (X, Y) are coordinates of the 2D image plane. For the sake of convenience, the 2D plane coordinates are still represented by (x, y).

The optical flow field (u, v) induced by 3D facial movements in several special cases are as follows:

Case I

$$\begin{aligned} u &= \Omega_z y - \Omega_y z + T_x \\ v &= -\Omega_z x + \Omega_x z + T_y \end{aligned} \tag{9}$$

where (u, v) is the optical flow field.

Case II

$$\begin{aligned} u &= \sum_{i=1}^{m} e_{1i}\phi_i + \Omega_z y - \Omega_y z + T_x \\ v &= \sum_{i=1}^{m} e_{2i}\phi_i - \Omega_z x + \Omega_x z + T_y \end{aligned} \tag{10}$$

Case III

$$\begin{aligned} u &= \sum_{i=1}^{m} e_{1i}\phi_i + \sum_{i=1}^{m} e_{2i}\phi_i\Omega_z - \sum_{i=1}^{m} e_{3i}\phi_i\Omega_y + \Omega_z y - \Omega_y z + T_x \\ v &= \sum_{i=1}^{m} e_{2i}\phi_i - \sum_{i=1}^{m} e_{1i}\phi_i\Omega_z + \sum_{i=1}^{m} e_{3i}\phi_i\Omega_x - \Omega_z x + \Omega_x z + T_y \end{aligned} \tag{11}$$

3.2 Motion Parameter Estimation

Now we discuss how to recover motion parameters from optical flow field. We first examine the Case I.

The optical flow field (9) can be rewritten as the following matrix form

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} y & 1 & 0 & -\Omega_y \\ -x & 0 & 1 & \Omega_x \end{bmatrix} \begin{bmatrix} \Omega_z \\ T_x \\ T_y \\ z \end{bmatrix} = \mathbf{A}'\mathbf{b}' \quad (12)$$

It is worth noticing that motion parameters and depth parameters are arranged into two different matrixes $\mathbf{A}'$ and $\mathbf{b}'$. An important observation is that the matrix $\mathbf{A}'$ contains only the rotation components Ω_x and Ω_y. The aim of such an arrange is to remove the vector $\mathbf{b}'$ in which depth parameters are contained. When total optical flow field is available the above matrix equation can be extended as

$$\begin{bmatrix} u_1 \\ v_1 \\ \cdot \\ \cdot \\ u_n \\ v_n \end{bmatrix} = \begin{bmatrix} y_1 & 1 & 0 & -\Omega_y & 0 & .. & 0 \\ -x_1 & 0 & 1 & \Omega_x & 0 & .. & 0 \\ y_2 & 1 & 0 & 0 & -\Omega_y & .. & 0 \\ -x_2 & 0 & 1 & 0 & \Omega_x & .. & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ y_n & 1 & 0 & 0 & .. & .. & -\Omega_y \\ -x_n & 0 & 1 & 0 & .. & .. & \Omega_x \end{bmatrix} \begin{bmatrix} \Omega_z \\ T_x \\ T_y \\ z_1 \\ z_2 \\ .. \\ z_n \end{bmatrix} \quad (13)$$

Equation (13) can be further rewritten as a more compact form

$$\mathbf{u} = \mathbf{A}\mathbf{b} \quad (14)$$

If the $\mathbf{A}$ is available, the optimum estimate of $\mathbf{b}$ in the mean-square sense is

$$\mathbf{b} = \mathbf{A}^{+}\mathbf{u} \quad (15)$$

where $\mathbf{A}^{+} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^{\mathbf{T}}$ is the generalized inversion of $\mathbf{A}$.

Substituting (15) into (14) yields the following constraint equation

$$\mathbf{u} = \mathbf{A}\mathbf{A}^{+}\mathbf{u} \quad (16)$$

In this constraint equation only rotation components Ω_x, Ω_y are concerned, which implies that the effect of the depth parameters on the rotation parameter estimation is removed. According to this fact a three-step approach to estimate facial motion parameters and depth parameters is designed

(1)**Estimation of Ω_x, Ω_y**

We seek the choice of Ω_x, Ω_y that minimizes the following expression over all candidate Ω_x, Ω_y.

$$E(\Omega_x, \Omega_y) = ||\mathbf{A}^{\perp}\mathbf{u}||^2 \quad (17)$$

Where $\mathbf{A}^{\perp} = \mathbf{I} - \mathbf{A}\mathbf{A}^{+}$.

(2)**Estimation of other motion parameters and depth parameters**

Once the parameters Ω_x, Ω_y are obtained, we can directly obtain other parameters

$$(\Omega_z, T_x, T_y, z_1, ..., z_n)^T = \mathbf{A}^{+}\mathbf{u} \quad (18)$$

(3)**Refinement of the depth value**

The obtained depth $\mathbf{z}$ using (18) is noisy because (18) is sensitive to the noise in optical flow field. In order to obtain a reliable depth estimate, the depth of the triangular vertex of the wireframe is refined by an LS method. This can be referred to [14].

The above method is only suitable to the case that the optical flow field is available. We now give an approach to estimate these parameters directly from the spatiotemporal derivatives of image intensity.

We further rewrite the equation(12) as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} y & 1 & 0 & -\Omega_y \\ -x & 0 & 1 & \Omega_x \end{bmatrix} \begin{bmatrix} \Omega_z \\ T_x \\ T_y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \end{bmatrix} \mathbf{b}' \quad (19)$$

As we know the famous optical flow constraint equation[13][11] is

$$I_x u + I_y v + I_t = 0 \quad (20)$$

Substituting (19) into (20) yields

$$(\mathbf{a_1} I_x + \mathbf{a_2} I_y)\mathbf{b}' = -I_t \quad (21)$$

From this constraint equation we are able to obtain the similar formula to compute motion parameters using the same strategy as the above optical flow based approach.

As for other two cases, we can make the same algebraic manipulations except that the optical flow fields are rewritten as follows,

Case II

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} y & 1 & 0 & e_{11} & .. & e_{1m} & -\Omega_y \\ -x & 0 & 1 & e_{21} & .. & e_{2m} & \Omega_x \end{bmatrix} \begin{bmatrix} \Omega_z \\ T_x \\ T_y \\ \phi_1 \\ . \\ \phi_m \\ z \end{bmatrix} \tag{22}$$

Case III

$$\begin{bmatrix} u - y\Omega_z \\ v + x\Omega_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & e_{11} + e_{21}\Omega_z + e_{31}\Omega_y & .. & -\Omega_y \\ 0 & 1 & e_{21} - e_{11}\Omega_z + e_{31}\Omega_x & .. & \Omega_x \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ \phi_1 \\ . \\ \phi_m \\ z \end{bmatrix} \tag{23}$$

respectively.

4 Conclusion

We have presented a depth independent facial movement estimation approach using the subspace technique. In this approach the 3D facial motion estimation is achieved through the following three steps: The first step is to estimate partial rotation parameters; the second step is to recover other motion parameters; and the final step is to refine the depth estimation. The most main advantage of this approach is that no knowledge on the depth is required.

5 Acknowledgement

This work is financially supported by NETUK.

References

[1] R.Forchheimer and O.Fahlander,"Low bit-rate coding through animation," in Proc. Picture Coding Symp. (PCS-83), Davis, Mar. 1983, pp.113-114.

[2] R.Forchheimer,"The motion estimation problem in semantic image coding,"in Proc. Picture Coding Symp. (PCS-87), Stockholm,June 1987, pp.171-172.

[3] R.Forchheimer and T.Kronander,"Image coding–from waveforms to animation", IEEE Trans. on ASSP, Vol.37, No.12, December 1989

[4] Haibo Li, P.Roivainen and R.Forchheimer, "3D motion estimation in model-based facial image coding", Dep. Elec. Eng. Rep. LiTH-ISY-I-1278, Linköping Univ., Oct 1991.

[5] P.Roivainen,"Motion estimation in model-based coding of human faces", Licentiate Thesis LIU-TEK-LIC-1990:25, ISY, Linköping Univ.,Sweden, 1990.

[6] K.Aizawa et al., "Model-based synthesis image coding system–modeling a person,s face and synthesis of facial expressions", in Proc. GLOBECOM-87 Nov. 1987, pp.45-49, Paper 2.3.

[7] A.N.Netravali and J.Salz,"Algorithms for estimation of three-dimensional motion", AT&T Technical Journal, vol.64, No.2, Feb 1985.

[8] Bill Welsh,"Model-based coding of images", Ph.D. dissertation, British Telecom Research Lab., Jan. 1991.

[9] H.G.Musmann, M.Hotter and J.Ostermann,"Object-oriented analysis-synthesis coding of moving images," Image Communication, 1,No.2, 117-138, Oct 1989.

[10] M.Rydfalk,"CANDIDE: A Parameterized face," Dep.Elec.Eng. Rep.LiTH-ISY-I-0866, Linköping Univ., Oct.1987.

[11] Haibo Li, Computation of optical flow considering changes in image intensity, Proc. Sym. on Image Analysis, Stockholm, March 1991

[12] D.Heeger and A.Jepson,"Subspace methods for recovering rigid motion I: algorithm and implementation," International Journal of Computer Vision, 7:2, 95-117, 1992.

[13] B.K.Horn, Robot Vision, MIT Press 1986

[14] Haibo Li and R.Forchheimer, "Depth refinement in model-based facial image coding", to be published.