3-D Motion Estimation and Wireframe Adaptation Including Photometric Effects for Model-Based Coding of Facial Image Sequences

Gözde Bozdağı, *Member, IEEE*, A. Murat Tekalp, *Senior Member, IEEE*, and Levent Onural

Abstract—We propose a novel formulation where 3-D global and local motion estimation and the adaptation of a generic wireframe model to a particular speaker are considered simultaneously within an optical flow based framework including the photometric effects of the motion. We use a flexible wireframe model whose local structure is characterized by the normal vectors of the patches which are related to the coordinates of the nodes. Geometrical constraints that describe the propagation of the movement of the nodes are introduced, which are then efficiently utilized to reduce the number of independent structure parameters. A stochastic relaxation algorithm has been used to determine optimum global motion estimates and the parameters describing the structure of the wireframe model. Results with both simulated and real facial image sequences are provided.

I. INTRODUCTION

DUE to growing interest in very low bit rate digital video (about 10 kbits/s), a significant amount of research has focused on model-based video compression [1]–[12]. Scientists became interested in model-based coding because the quality of digital video obtained by hybrid coding techniques, such as CCITT Rec. H.261 [13], is deemed unsatisfactory at these very low bit rates. Studies in model-based coding employ object models ranging from general purpose 2-D or 3-D models [2], [10], [11] to application specific wireframe models [3]–[9]. One of the main applications of model-based coding has been the videophone, where scenes are generally restricted to head and shoulder type images. In videophone applications, it is common to represent the head and shoulders of the speaker by a specific wireframe model which is present at both the receiver and the transmitter. Then, 3-D motion and structure estimation techniques are employed at the transmitter to track the motion of the wireframe model and the changes in its structure from frame to frame. The estimated motion and structure (depth) parameters along with changing texture information are sent and used to synthesize the next frame in the receiver side.

Traditionally, the scaling of a generic wireframe model to the actual speaker and motion estimation have been handled

separately. Many of the existing methods consider scaling a generic wireframe to the actual speaker using only the initial frame of the sequence [3], [14]. Thus, the scaling in the z-direction (depth) is necessarily approximate. The scaling procedure first proposed by Aizawa *et al.* [3] involves manual steps, while Reinders *et al.* [14] later proposed an automated procedure. For subsequent frames, first the 3-D global motion of the head is estimated under rigid body assumption, using either point correspondences [3], [6], [12] or optical flow based formulations [7], [8]. A general overview of 3-D rigid body motion and structure estimation methods can be found in [15]. Then, local motion (due to facial expressions) is estimated making use of action units (AU) described by Facial Action Coding System (FACS) [8]. Recently, Li *et al.* [9] proposed a method, to recover both the local and global motion parameters from the spatio-temporal derivatives of the image, which also requires *a priori* knowledge of the AU's. Finally, provisions for texture update for regions where the model-based approach fails are considered.

In this paper, we propose a novel formulation where 3-D global and local motion estimation and the adaptation of the wireframe model are considered simultaneously within an optical flow based framework including the photometric effects of motion. Although, the utility of photometric cues in 3-D motion and structure estimation are well-known [16]–[19], photometric information was not used in the context of motion estimation for videophone applications before. The main contributions of this paper are: (i) a flexible 3-D wireframe model has been used where the X, Y and Z coordinates of the nodes of the wireframe model are allowed to vary from frame to frame so as to minimize the error in the optical flow equation, and (ii) photometric effects are included in the optical flow equation. The proposed adaptation of the wireframe model serves for two purposes that cannot be separated: to reduce the misfit between the wireframe model and the facial region in frame $k-1$, and to account for the local motion deformations from frame $k-1$ to frame k without using any *a priori* information about the AU's. The simultaneous estimation formulation is motivated by the fact that estimation of the global motion, local motion and adaptation of the wireframe model including the depth values are mutually related; thus a combined optimization approach is necessary to obtain the best results. Because an optical flow based criterion function is utilized, computation of the synthesis error is not necessary from iteration to iteration;

thus, resulting in an efficient implementation. The synthesis error at the conclusion of the iterations is used to validate the estimated parameters, and to decide whether a texture update is necessary.

In Section II, we review the optic-flow based approach to 3-D motion and structure estimation and briefly discuss a photometric image formation model. Estimation of the illuminant direction is summarized in Section II.C. The formulation of simultaneous motion estimation and wireframe adaptation problem including the photometric effects of motion is presented in Section III. The algorithm for the proposed simultaneous estimation including an efficient method to update the nodes of the wireframe model is given in Section IV. Experimental results on simulated and real video sequences are presented in Section V to demonstrate the effectiveness of the proposed methods.

II. BACKGROUND

A. 3-D Motion and Structure from Optical-Flow

In this section, we will review the equation describing the relation between the 3-D motion and structure and the corresponding 2-D velocity field (optical flow), which is related to the projection of the 3-D motion onto the image plane under certain assumptions [16]. By using additional constraints regarding the 3-D structure of the scene, it is possible to recover the parameters of the 3-D motion from the associated optical flow field [15].

Let $[X_s(t) \quad Y_s(t) \quad Z_s(t)]^T$ be the vector of the 3-D coordinates of a particular point s of a moving object at time t and S refers to the object which is the set of all such points. If we assume that the object is rigid and subject to small rotation, we can express the position of s at time $t + \Delta t$ given its position at time t as,

$$\begin{bmatrix} X_s(t + \Delta t) \\ Y_s(t + \Delta t) \\ Z_s(t + \Delta t) \end{bmatrix} = \begin{bmatrix} 1 & \omega_Z & -\omega_Y \\ -\omega_Z & 1 & \omega_X \\ \omega_Y & -\omega_X & 1 \end{bmatrix}$$
$$\times \begin{bmatrix} X_s(t) \\ Y_s(t) \\ Z_s(t) \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}, \quad \forall s \in S \quad (1)$$

where ω_X, ω_Y, and ω_Z are the rotational displacements about the X, Y and Z axes, respectively, and T_X, T_Y, and T_Z are the translational displacements along the X, Y and Z axes, respectively. Then, the 3-D velocities are approximated by

$$\begin{bmatrix} \dot{X}_s(t) \\ \dot{Y}_s(t) \\ \dot{Z}_s(t) \end{bmatrix} = \begin{bmatrix} 0 & \omega_Z & -\omega_Y \\ -\omega_Z & 0 & \omega_X \\ \omega_Y & -\omega_X & 0 \end{bmatrix} \begin{bmatrix} X_s(t) \\ Y_s(t) \\ Z_s(t) \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}, \quad (2)$$

where $\dot{X}_s(t) = X_s(t+\Delta t) - X_s(t)$, $\dot{Y}_s(t) = Y_s(t+\Delta t) - Y_s(t)$, $\dot{Z}_s(t) = Z_s(t+\Delta t) - Z_s(t)$, and $\Delta t = 1$. Under orthographic projection along the z-direction, we can represent the 2-D velocity field associated with the projection of the point s as

$$\dot{x}_s(t) = \omega_Z y_s(t) - \omega_Y Z_s(t) + T_X$$
$$\dot{y}_s(t) = -\omega_Z x_s(t) + \omega_X Z_s(t) + T_Y \quad (3)$$

where $x_s(t) \triangleq X_s(t)$, $y_s(t) \triangleq Y_s(t)$.

Now, let us represent the intensity at point $(x_s(t), y_s(t))$ by $I(x_s(t), y_s(t), t)$. Assuming that $I(x_s(t), y_s(t), t)$ does not change in time, we can write,

$$I(x_s(t), y_s(t), t) = \text{constant}; \quad \forall t, \text{ and } \forall s \in S. \quad (4)$$

or in other words,

$$\frac{dI(x_s(t), y_s(t), t)}{dt} = 0; \quad \forall s \in S. \quad (5)$$

Since the above equations are valid for all $s \in S$, from now on we will drop the subscript s to simplify the equations. Using the chain rule of differantiation, we can write (5) as

$$I_x \dot{x} + I_y \dot{y} + I_t = 0, \quad (6)$$

where I_x, I_y, and I_t are the partial derivatives of image intensity, I, in x, y and t, respectively. Substituting (3) into (6), we get

$$I_x(\omega_Z y - \omega_Y Z + T_X) + I_y(-\omega_Z x + \omega_X Z + T_Y) + I_t = 0. \quad (7)$$

Eq. 7 is a constraint that relates the spatio-temporal image gradients to the 3-D motion $\omega_x, \omega_y, \omega_z, T_x, T_y$ and structure (Z) parameters under the assumption that the variation in image intensity pattern is solely due to the 3-D motion of the underlying scene.

Eq. (7) alone is not sufficient to determine the 3-D motion and structure parameters if they are allowed to change freely and independently at every point. Several approaches exist in the literature to compute the motion and structure parameters from (7) under piecewise rigid scene assumption (constraining the variation of the 3-D motion parameters) and with certain surface structure models (constraining the variation of Z with (X, Y)) [15]–[21].

B. Photometric Model of Image Formation

The above formulation neglects the effects of changing shading due to the 3-D motion of the scene. For example, in the case of rotational motion because the surface normals change, the shading of the objects varies even if the external illumination remains constant. Recently Pentland [18] showed that the changes in image intensity because of photometric effects can dominate intensity changes due to geometric effects of motion (changes in projected surface geometry due to 3-D motion). Similar discussions can be found in [16], [17], [19]. Here we briefly discuss a photometric model of image formation with the aim of incorporating photometric effects into the above optical flow based formulation.

The image intensity associated with the point s (described in the previous section) can be represented through a reflectance map

$$I(x_s(t), y_s(t), t) = \mathcal{R}(p_s(t), q_s(t)) \quad (8)$$

where $\mathcal{R}$ denotes the reflectance map function, and $p = \frac{\partial Z}{\partial x}$ and $q = \frac{\partial Z}{\partial y}$ are the partial derivatives of depth $Z_s(t)$ with respect to the image coordinates x and y respectively. Under the assumptions of orthographic projection onto the image

plane and a Lambertian surface with constant albedo, ρ, we can express the reflectance map as [22]

$$\mathcal{R}(p_s(t), q_s(t)) = \rho \vec{N_s}(t) \cdot \vec{L} \qquad (9)$$

where $\vec{L} = (L_x, L_y, L_z)$ is the unit vector in the mean illuminant direction and $\vec{N}$ is the unit surface normal of the object at point s given by

$$\vec{N_s}(t) = (-p_s(t), -q_s(t), 1)/(p_s^2(t) + q_s^2(t) + 1)^{1/2}. \quad (10)$$

Note that the illuminant direction can also be expressed in terms of tilt and slant angles as

$$\vec{L} = (L_x, L_y, L_z) = (\cos\tau\sin\sigma, \sin\tau\sin\sigma, \cos\sigma) \quad (11)$$

where τ, the tilt angle of the illuminant, is the angle between $\vec{L}$ and the $X - Z$ plane, and σ, the slant angle, is the angle between $\vec{L}$ and the positive Z axis.

The model (9) is widely used in the computer vision literature for estimating the object shape and the illuminant direction from shading in case of diffuse reflection [18]. Moreover, its simple functional form makes it popular than specular models which are mainly used when the surface is shiny.

C. Estimation of Illumination Direction

In order to incorporate the photometric effects of motion into 3-D motion estimation and wireframe adaptation, the illuminant direction $\vec{L}$ must be known or estimated from the available frames. In this study, we use the method proposed by Zheng et al. [25] to estimate the illuminant direction.

The method to estimate the tilt angle is based on approximating the surface by spherical patches. The reason for using locally spherical patches is to begin from an initial state close to the true surface shape in order not to be stuck in a local minimum. An estimate of the tilt angle is given by

$$\tau = \arctan\left(\frac{E\{\hat{x}_L/\sqrt{\hat{x}_L^2 + \hat{y}_L^2}\}}{E\{\hat{y}_L/\sqrt{\hat{x}_L^2 + \hat{y}_L^2}\}}\right) \qquad (12)$$

where $E\{.\}$ denotes ensemble averages over the spatial variables (which is computed by averaging the results obtained over 4×4 windows in our implementation), and $\hat{x}_L$ and $\hat{y}_L$ are the x and y components of the local estimate of the tilt of the illuminant, respectively, computed as

$$\begin{bmatrix} \hat{x}_L \\ \hat{y}_L \end{bmatrix} = (B^t B)^{-1} B^t \begin{bmatrix} \delta I_1 \\ \delta I_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta I_N \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} \delta x_1 & \delta y_1 \\ \delta x_2 & \delta y_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \delta x_N & \delta y_N \end{bmatrix}.$$

Here, δI_i is the difference in image intensity along a particular direction $(\delta x_i, \delta y_i)$, and N is the number of directions (in our implementation we have set $N = 8$ for each 4×4 window).

The slant angle σ can be uniquely estimated from

$$\frac{E\{I\}}{E\{I^2\}} = f_3(\sigma) \qquad (13)$$

since $f_3(\sigma)$ (defined in [25]) is a monotonically decreasing function of σ, where $E\{I\}$ and $E\{I^2\}$ are the expected values of the image intensities and the square of the image intensities, respectively, estimated from the image area where the wireframe model is fitted. Finally, the surface albedo can be estimated from

$$\rho = \frac{E\{I\} \cdot f_1(\sigma) + \sqrt{E\{I^2\} \cdot f_2(\sigma)}}{f_1^2(\sigma) + f_2(\sigma)}, \qquad (14)$$

where $f_1(\sigma)$ and $f_2(\sigma)$ are seventh order polynomials in $\cos\sigma$ as defined in [25].

III. PROBLEM FORMULATION

In this section we present the proposed formulation for simultaneous estimation of 3-D global motion parameters and the adaptation of the wireframe model which also takes the presence of local motion into account. The inclusion of the photometric effects into the optical flow equation is discussed in Section III.A. In Section III.B, we introduce geometrical constraints about the structure of the wireframe model and formulate the problem.

A. Incorporation of the Photometric Effects

Since we represent the 3-D structure of the head and shoulders scene by a wireframe model and the surface of the wireframe model is composed of planar patches, the variation in the intensity of a pixel due to photometric effects of motion will be related to a change due to motion in the normal vector of the patch that this pixel belongs to using the image formation model (9).

From (10) and (11), we can write the change in the normal vector associated with the point s due to the 3-D motion as

$$\begin{aligned} d\vec{N_s}(t) &= \vec{N}(X_s(t+\Delta t), Y_s(t+\Delta t), Z_s(t+\Delta t)) \\ &\quad - \vec{N_s}(X_s(t), Y_s(t), Z_s(t)) \\ &= \frac{(-p_s(t+\Delta t), q_s(t+\Delta t), 1)^T}{(p_s(t+\Delta t)^2 + q_s(t+\Delta t)^2 + 1)^{1/2}} \\ &\quad - \frac{(-p_s(t), -q_s(t), 1)^T}{(p_s(t)^2 + q_s(t)^2 + 1)^{1/2}}, \end{aligned} \qquad (15)$$

where using orthographic projection $p_s(t) = \frac{\partial Z_s(t)}{\partial x_s(t)}$, $q_s(t) = \frac{\partial Z_s(t)}{\partial y_s(t)}$ and

$$\begin{aligned} p_s(t+\Delta t) &= \frac{\partial Z_s(t+\Delta t)}{\partial x_s(t+\Delta t)} = \frac{-\omega_Y + p_s(t)}{1 + \omega_Y p_s(t)} \\ q_s(t+\Delta t) &= \frac{\partial Z_s(t+\Delta t)}{\partial y_s(t+\Delta t)} = \frac{\omega_X + q_s(t)}{1 - \omega_X q_s(t)}. \end{aligned} \qquad (16)$$

Assuming that the mean illuminant direction $\vec{L} = (L_x, L_y, L_z)$ remains constant, we can represent the change in intensity due to photometric effects of motion using (9) as

$$\frac{dI(x_s(t), y_s(t), t)}{dt} = \rho\vec{L} \cdot d\vec{N_s}(t). \qquad (17)$$

Then, combining (6) and (17), dropping the subscript s for simplicity and denoting $p(t + \Delta t)$ as p', $q(t + \Delta t)$ as q', $p(t)$

$$p_i = -\frac{(Y_2^{(i)} - Y_1^{(i)})(Z_3^{(i)} - Z_1^{(i)}) - (Z_2^{(i)} - Z_1^{(i)})(Y_3^{(i)} - Y_1^{(i)})}{(X_2^{(i)} - X_1^{(i)})(Y_3^{(i)} - Y_1^{(i)}) - (Y_2^{(i)} - Y_1^{(i)})(X_3^{(i)} - X_1^{(i)})},$$

$$q_i = -\frac{(Z_2^{(i)} - Z_1^{(i)})(X_3^{(i)} - X_1^{(i)}) - (X_2^{(i)} - X_1^{(i)})(Z_3^{(i)} - Z_1^{(i)})}{(X_2^{(i)} - X_1^{(i)})(Y_3^{(i)} - Y_1^{(i)}) - (Y_2^{(i)} - Y_1^{(i)})(X_3^{(i)} - X_1^{(i)})},$$

and

$$c_i = Z_1^{(i)} + X_1^{(i)} \frac{(Y_2^{(i)} - Y_1^{(i)})(Z_3^{(i)} - Z_1^{(i)}) - (Z_2^{(i)} - Z_1^{(i)})(Y_3^{(i)} - Y_1^{(i)})}{(X_2^{(i)} - X_1^{(i)})(Y_3^{(i)} - Y_1^{(i)}) - (Y_2^{(i)} - Y_1^{(i)})(X_3^{(i)} - X_1^{(i)})}$$
$$+ Y_1^{(i)} \frac{(Z_2^{(i)} - Z_1^{(i)})(X_3^{(i)} - X_1^{(i)}) - (X_2^{(i)} - X_1^{(i)})(Z_3^{(i)} - Z_1^{(i)})}{(X_2^{(i)} - X_1^{(i)})(Y_3^{(i)} - Y_1^{(i)}) - (Y_2^{(i)} - Y_1^{(i)})(X_3^{(i)} - X_1^{(i)})}.$$

$$\begin{aligned}
e_i(x,y) = &-I_x(\omega_z y - \omega_y(p_i x + q_i y + c_i) + T_x) \\
&+ I_y(-\omega_z x + \omega_x(p_i x + q_i y + c_i) + T_y) \\
&+ I_t - \rho(L_x, L_y, L_z) \\
&\cdot \left(\frac{(-\frac{-\omega_y + p_i}{1+\omega_y p_i}, -\frac{\omega_x + q_i}{1-\omega_x q_i}, 1)}{((\frac{-\omega_y + p_i}{1+\omega_y p_i})^2 + (\frac{\omega_x + q_i}{1-\omega_x q_i})^2 + 1)^{1/2}} - \frac{(-p_i, -q_i, 1)}{(p_i^2 + q_i^2 + 1)^{1/2}} \right),
\end{aligned}$$

$$(21)$$

---

as p, $q(t)$ as q, we include the photometric effects into the optical flow based formulation as

$$I_x(\omega_z y - \omega_y Z + T_x) + I_y(-\omega_z x + \omega_x Z + T_y) + I_t$$
$$= \rho \vec{L} \cdot \left[\frac{(-p', -q', 1)^T}{\sqrt{p'^2 + q'^2 + 1}} - \frac{(-p, -q, 1)^T}{\sqrt{p^2 + q^2 + 1}} \right]. \quad (18)$$

The term on the right hand side of (18) may be significant especially if the change in the surface normal has components either toward or away from the illuminant direction [18].

B. Structure of the Wireframe Model and Problem Statement

The wireframe model is composed of triangular patches which are characterized by the (X, Y, Z) coordinates of their respective vertices. Given the (X, Y, Z) coordinates of the vertices of a patch, we can write the equation of the plane containing this patch. Let $P_1^{(i)} = (X_1^{(i)}, Y_1^{(i)}, Z_1^{(i)})$, $P_2^{(i)} = (X_2^{(i)}, Y_2^{(i)}, Z_2^{(i)})$ and $P_3^{(i)} = (X_3^{(i)}, Y_3^{(i)}, Z_3^{(i)})$ denote the vertices of the ith patch, and $P^{(i)} = (X_i, Y_i, Z_i)$ be any point on this patch. Then,

$$\vec{P^{(i)}P_1^{(i)}} \cdot (\vec{P_1^{(i)}P_2^{(i)}} \times \vec{P_1^{(i)}P_3^{(i)}}) = 0$$

gives the equation of the plane containing $P_1^{(i)}$, $P_2^{(i)}$ and $P_3^{(i)}$, where $\vec{P^{(i)}P_1^{(i)}}$, $\vec{P_1^{(i)}P_2^{(i)}}$, and $\vec{P_1^{(i)}P_3^{(i)}}$ are the vectors from the former point to the latter, respectively. We can express this equation in the form

$$Z_i = p_i X_i + q_i Y_i + c_i, \quad (19)$$

where (see top of page).

Using (19), the Z coordinate of any point on the ith patch can be expressed in terms of the parameters p_i, q_i and c_i and the X and Y coordinates of the point. Then, we can eliminate Z from (18) by substituting (19) into (18) with $X_i = x_i$, $Y_i = y_i$, where the patch index i is determined for each (x, y) according to the orthographic projection.

The problem of simultaneously estimating the 3-D global motion parameters $\omega_x, \omega_y, \omega_z, T_x, T_y$, describing the global motion of the head, and the parameters p_i, q_i, c_i, that describe the structure of the wireframe model, can then be formulated as to minimize the sum square error in the optical flow equation (18) over all pixels in a frame given by

$$E = \sum_i \sum_{(x,y) \in i^{\text{th}} \text{patch}} e_i^2(x, y) \quad (20)$$

where (see (21) above) with respect to the variables $\omega_x, \omega_y, \omega_z, T_x, T_y, p_i, q_i, c_i$ and $i = 1, \ldots$, number of patches. It is important to note that the surface normals p_i, q_i and the translation (in the Z direction) parameters c_i of each planar patch of the wireframe are not completely independent of each other. Each triangular patch is either surrounded by two (if it is on the boundary of the wireframe) or three other triangles. The requirement that the neighboring patches must intersect at a straight line imposes a constraint on the structure parameters of these patches in the form

$$p_i x_{ij} + q_i y_{ij} + c_i = p_j x_{ij} + q_j y_{ij} + c_j \quad (22)$$

where p_j, q_j and c_j denote the structure parameters of the jth patch, and (x_{ij}, y_{ij}) denote the coordinates of a point that

lie on the intersection of the ith and j^{th} patches. Eq. 22 not only enables us to preserve the structure of the wireframe during the adaptation, but also facilitates the reduction of the number of independent unknowns in the optimization process. An efficient algorithm to reduce the number of independent unknowns using (22) is given in Section IV.

The perturbation of the structure parameters p_i, q_i and c_i for each patch i results in a change in the coordinates of the nodes of the updated wireframe. The new coordinates (X_n, Y_n, Z_n) of the node n can be computed given the updated structure parameters of three patches that intersect at node n. Let the patches i, j and k intersect at node n. Then, the relations

$$p_i X_n + q_i Y_n + c_i = p_j X_n + q_j Y_n + c_j$$
$$p_i X_n + q_i Y_n + c_i = p_k X_n + q_k Y_n + c_k \qquad (23)$$

specify X_n and Y_n. The coordinate Z_n can be computed using (19) for any of the patches i, j or k. It is this updating of the coordinates of the nodes that allows the adaptation of the wireframe model to lower the misfit error and accomodate the presence of local motion, such as the motion of the eyes and the mouth.

IV. METHOD

In this Section a stochastic relaxation algorithm to estimate the global motion parameters $\omega_x, \omega_y, \omega_z, T_x, T_y$, and the structure parameters p_i, q_i, c_i, by minimizing the error in the optical flow equation E given by (20), is proposed using the values of ρ and (L_x, L_y, L_z) that are estimated according to Section II.C. Stochastic relaxation is an iterative technique to find the global optimum for complex optimization problems. Each iteration consists of perturbing the state of the system in some random fashion [26]. In our experiments, the independent unknowns are perturbed at each iteration, where the perturbations are generated as samples of Gaussian distributed numbers with the variance of the distribution adjusted according to the value of the cost function E (given by (20)). Further, the magnitude of perturbations is reduced with the number of iterations, so that convergence should result.

As stated in Section III.B, not all of the parameters p_i, q_i, c_i are independent due to the geometrical constraints defining the structure of the wireframe model given by (22). The dependent structure parameters are determined as follows: At each iteration cycle, we visit all the patches of the wireframe model in sequential order. If, at the present iteration cycle, none of the neighboring patches of patch i has yet been visited for updating their structure parameters (e.g., the initial patch), then p_i, q_i, c_i are all independent and are perturbed. If only one of the neighboring patches, say patch j, has been visited (p_j, q_j, c_j have already been updated), then two of the parameters, say p_i and q_i are independent and perturbed. The dependent variable c_i is computed from (22) as

$$c_i = p_j x_{ij} + q_j y_{ij} + c_j - p_i x_{ij} - q_i y_{ij}, \qquad (24)$$

where x_{ij} is one of the nodes common to both patches i and j, that is either in the boundary or has been already updated in the present iteration cycle. If two of the neighboring patches, say

patches j and k, have already been visited, i.e., the variables p_j, q_j, c_j and p_k, q_k, c_k have been updated, than only one variable, say p_i, is independent and perturbed. Then, c_i can be found from (24), and q_i can be evaluated as

$$q_i = \frac{p_k x_{ik} + q_k y_{ik} + c_k - p_i x_{ik} - c_i}{y_{ik}}, \qquad (25)$$

where x_{ik} is one of the nodes common to both patches i and k, that is either in the boundary or has been already updated in the present iteration cycle.

The change in the structure parameters p_i, q_i, c_i affect the location of the nodes. At each iteration cycle, the coordinates of a node are updated as soon as the structure parameters of three patches that intersect at that node are updated using (23). Thus, the new X and Y coordinates of the nodes are given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} p_i - p_j & p_j - p_k \\ q_i - q_j & q_j - q_k \end{bmatrix}^{-1} \begin{bmatrix} c_j - c_i \\ -c_j + c_k \end{bmatrix} \qquad (26)$$

The new Z coordinate can be computed from (19) given the X and Y coordinates and the p_i, q_i, c_i for any patch passing through that node.

In the following we summarize the proposed algorithm as:
1) Estimate the illumination direction using (12) and (13), and the surface albedo using (14).
2) Initialize the coordinates of the nodes (X_n, Y_n, Z_n), for all n, using an approximately scaled initial wireframe model. Set the iteration counter $m = 0$.
3) Determine the initial motion parameters using the stochastic relaxation method described in [12] (or any other point correspondence method to compute the motion parameters using a set of selected nodes given their depth values).
4) Compute the value of the cost function E given by (20).
5) If $E < \epsilon$, stop.
   Else, set $m = m + 1$ and
   perturb the motion parameters $\omega = [\omega_x, \omega_y, \omega_z, T_x, T_y]^T$ as

$$\omega_{(m)} \longleftarrow \omega_{(m-1)} + \alpha^m \Delta, \qquad (27)$$

where α is a constant, $\Delta = N(0, \sigma^{2(m)})$, i.e., zero mean Gaussian with variance $\sigma^{2(m)}$, where $\sigma^{2(m)} = E$, and the structure parameters p_i, q_i and c_i as

```
Define count_i as the number of neigh-
boring patches to patch i whose struc-
ture parameters have been perturbed.
Set count_i=0, for all patches i.
Perturb p_1, q_1, c_1 as
```

$$p_{1(m)} \longleftarrow p_{1(m-1)} + \alpha^m \Delta_1,$$
$$q_{1(m)} \longleftarrow q_{1(m-1)} + \alpha^m \Delta_1,$$
$$c_{1(m)} \longleftarrow c_{1(m-1)} + \alpha^m \Delta_1, \qquad (28)$$

where $\Delta_i = N_i(0, \sigma_i^{2(m)})$, i.e., zero mean Gaussian with variance $\sigma_i^{2(m)}$, where $\sigma_i^{2(m)} = \sum_{(x,y) \in \text{patch} i} e_i^2(x,y)$.

```
for (i=2) to number of patches
 {
 if(count_i = 1){
        perturb p_i and q_i
        increment count_m, for all m denoting neighbors of patch i
        Compute c_i using Eq. 27 where the x and y coordinates are that
        of a fixed or a precomputed node on the intersection line
        between patch i and j
        }

 if(count_i == 2)){
        perturb p_i
        increment count_m, for all m denoting neighbors of patch i
        Compute c_i using Eq. 27 and q_i using Eq. 28 where x_ij,y_ij
        and x_ik,y_ik are coordinates of a fixed or a precomputed node
        on the intersection line between patches i,j and i,k
        respectively.
        }
 If p_i,q_i, and c_i for at least three patches intersecting at a node are
 updated, then update the coordinates of the node by using Eq. 28.
 }
```

```
        increment count_j, for all j denoting
        neighbors of patch 1 (con't top of this page).
```
6) Go to (4).

Experimental results will be presented in the next section to demonstrate the performance of the proposed method.

V. RESULTS

We have demonstrated the proposed method with both real and synthetic image sequences. The real image sequences are "Miss America" and "Claire" where each frame consists of 256×256 and 352×288 pixels, respectively. The synthetic sequence is generated by moving (the vertices corresponding to the head area) and shading the wireframe (Fig. 1) which is an extension of the CANDIDE wireframe [27] and composed of 217 triangles and 144 nodes [5]. In the simulations, we set $\alpha = 0.95$ to obtain the reported results. In each case, the convergence is achieved in about 100 iterations. The experimental results are summarized in the following subsections.

A. Simulation Results

According to our model, the synthesis error in model-based image synthesis originates from: (i) misfit of the wireframe, (ii) error in global motion estimation, (iii) error in local motion estimation, and (iv) the photometric effects of the motion. The following set of simulation experiments is intended to test each of these causes in a controlled manner. The experiments with the synthetic image sequences show the performance of the algorithm in case of global motion only, global and local motion, with and without incorporating illumination effects. On the other hand, the real image sequences used in the experiments have both global motion, local motion and illumination effects. These set of experiments enables us to
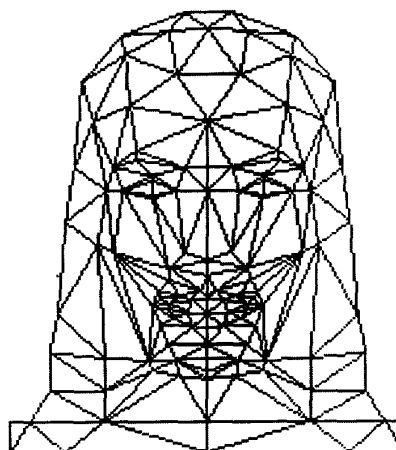

Fig. 1. The wireframe model.

see whether the consideration of illumination is useful and effective for the real image sequences for motion and structure estimation.

Simulations With a Synthetic Image Sequence: The first simulation tests the accuracy of the proposed method for global motion estimation by eliminating all other sources of error. To this effect, we generate an image frame by taking the orthographic projection of the wireframe in Fig. 1, and painting its patches by black and white alternatingly. Since the image is obtained directly from the wireframe itself, there is no fitting problem, and therefore, no misfit. The next frame is obtained by rotating and translating the wireframe using a given set of motion parameters and computing the projection of the wireframe in this new position. No local motion or shading effects are simulated. Table 1 shows the
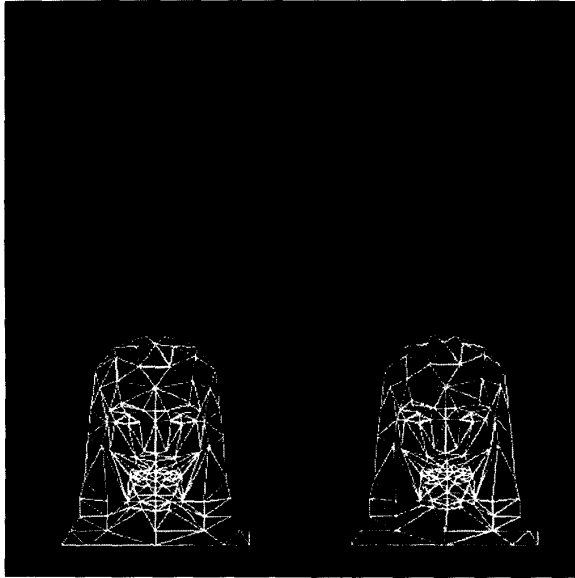
Fig. 2. clockwise (a) The first and (b) the second frames of the synthetic image sequence with global motion; (c) the initial wireframe and (d) the rotated wireframe through the estimated motion parameters pasted on the first and the second frames respectively.

Fig. 3. clockwise (a) The first and (b) the second frames of the synthetic image sequence with globaland local motion; (c) The initial wireframe and (d) the rotated wireframe through the estimated motion parameters pasted on the first and the second frames respectively.

TABLE I
GLOBAL MOTION ESTIMATION WITH THE SYNTHETIC SEQUENCE

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1046	4.61
ω_y	0.35	0.3368	0.3526	0.76
ω_z	-0.03	-0.0113	-0.030641	2.13
T_x	6	4.962	5.9860	0.23
T_y	-3	-2.8999	-2.9791	0.69

TABLE II
GLOBAL AND LOCAL MOTION ESTIMATION WITH THE SYNTHETIC SEQUENCE

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1072	7.15
ω_y	0.35	0.3368	0.3461	1.104
ω_z	-0.03	-0.0113	-0.02724	9.19
T_x	6	4.962	5.8697	2.17
T_y	-3	-2.8999	-2.9737	0.87

true and estimated motion parameters. "Initial point" indicates the motion parameters at the beginning of the iterations, and "Result" shows the estimated parameters at convergence. The "Percentage error" gives the percentage deviation of our results from the true values. The first and the second frames are shown in Fig. 2(a) and 2(b). The initial wireframe, and, the rotated wireframe with the estimated motion parameters, are shown on their corresponding frames in Fig. 2(c) and 2(d). The mean square synthesis error between the actual second frame $I_a(x, y)$ and the synthesized second frame $I_s(x, y)$ is computed according to

$$\text{MSE} = \left(\frac{1}{N \times M} \sum_{x,y} (I_a(x, y) - I_s(x, y))^2 \right)^{\frac{1}{2}}, \quad (29)$$

where N and M show the x and y extends of the image, respectively. The MSE in this case is 5.062.

The next simulation tests the ability of our method to track the local motion deformations. For this purpose, in addition to the above global motion, the action units (AU) 17 and 46, corresponding to chin raiser and winking, are also synthesized. Table II shows the true and estimated global motion parameters. The first and the second frames together with the initial wireframe and the rotated and modified wireframe through the
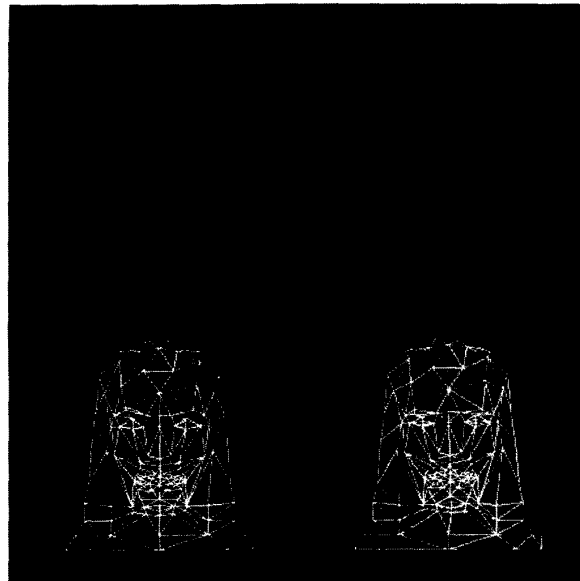
estimated motion parameters pasted on the first and the second frames respectively are also shown in Fig. 3. The MSE is computed to be 5.579.

Simulations With the "Miss America" Sequence

With no photometric effects

After getting successful results from the previous experiment we tested our algorithm, again for global and local motion tracking performance, using a textured 3-D model instead of the previous bare wireframe model. The textured model is obtained by mapping a single frame from the "Miss America" sequence to the initial wireframe model. The mapping is accomplished after scaling the wireframe model approximately to the location and the size of the face by positioning four extreme points, interactively. A second frame is obtained from the first one by rotating and translating it. Then we applied our algorithm to check its performance in finding these already known motion parameters. The results of global motion estimation with no local motion and no photometric effects is presented in Table III. In addition, we have synthesized a new frame from the first frame using the estimated motion parameters and computed the difference between this synthesized frame and the second frame. The MSE is found to be 6.989.
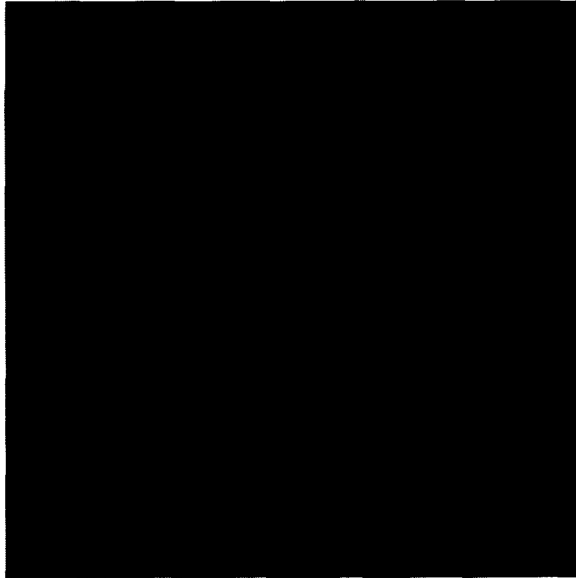
Fig. 4. clockwise (a) The first frame of "Miss America", (b) simulated second frame with global and localmotion (without photometric effects); (c) synthesized second frame using the estimated motion and structure parameters.

TABLE III
GLOBAL MOTION ESTIMATION WITH THE SIMULATED MISS
AMERICA SEQUENCE WITHOUT THE PHOTOMETRIC EFFECTS

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1054	5.4
ω_y	0.35	0.3368	0.3369	3.74
ω_z	-0.03	-0.0113	-0.02717	9.43
T_x	6	4.962	5.7126	4.79
T_y	-3	-2.8999	-3.0796	2.65

We repeat the above experiment by including local motion specified by the AU2, AU17 and AU46 which correspond to outer brow raiser, chin raiser and winking. By this way, we are able to test how local deformations affect the 3-D global motion and structure parameter estimation and whether we can track these deformations. Table IV shows the true and estimated global motion parameters. The first, second, and the synthesized frame in this case are shown in Fig. 4. The MSE between the second and the synthesized frames is now equal to 7.554.

With photometric effects: The purpose of this set of simulations is to see whether there is an improvement or not when the illumination effects are also considered. To simulate the photometric effects, we estimate the direction of the illuminant from the first frame. The estimated tilt and slant angles are 141.131 and 72.5064 which correspond to $L = (-0.74257, 0.59851, 0.33059)$. The 3-D object of the first frame is again rotated and translated (with or without local motion) as done in the previous experiment, but this time the object is shaded using (8) according to the change in the surface normal vectors due to rotation (assuming the direction of the illuminant remains constant). Tables V and VI show the results of global motion estimation both without and with
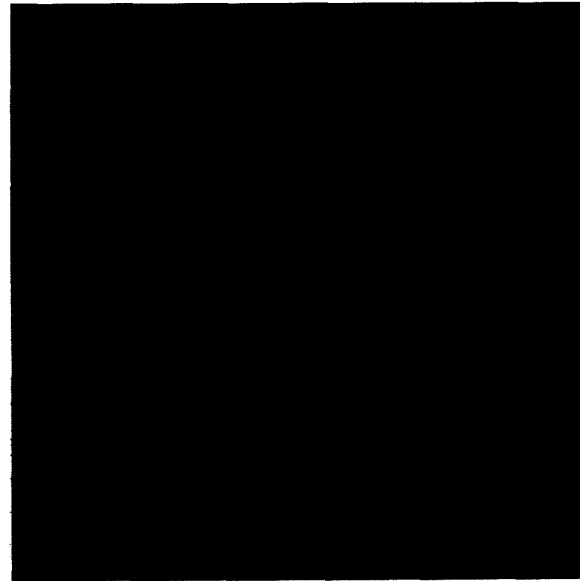

Fig. 5. clockwise (a) The first frame of "Miss America", (b) simulated second frame with global and local motion, and the photometric effects; (c) synthesized second frame using the estimated motion and structure parameters.

TABLE IV
GLOBAL AND LOCAL MOTION ESTIMATION WITH THE SIMULATED MISS
AMERICA SEQUENCE WITHOUT THE PHOTOMETRIC EFFECTS

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1083	8.3
ω_y	0.35	0.3368	0.33446	4.44
ω_z	-0.03	-0.0113	-0.02683	10.56
T_x	6	4.962	5.4719	8.80
T_y	-3	-2.8999	-2.7853	7.15

TABLE V
GLOBAL MOTION ESTIMATION WITH THE SIMULATED MISS
AMERICA SEQUENCE INCLUDING THE PHOTOMETRIC EFFECTS

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1052	5.2
ω_y	0.35	0.3368	0.3482	0.53
ω_z	-0.03	-0.0113	-0.02801	6.64
T_x	6	4.962	6.2038	3.39
T_y	-3	-2.8999	-3.0702	2.3

local motion. In the case of local motion, we again use the AU 2, 17 and 46 which correspond to outer brow raiser, chin raiser and winking, respectively. The second and synthesized frames using estimated parameters are shown in Figs. 5 in the case with local motion. The MSE are found to be 5.839 and 7.101, for the experiments without and with local motion, respectively.

B. Results With Real Image Sequences

Experiments With Two Frames of the "Miss America" Sequence: Now, we test our algorithm using the first and the tenth frames from the"Miss America" sequence as the first
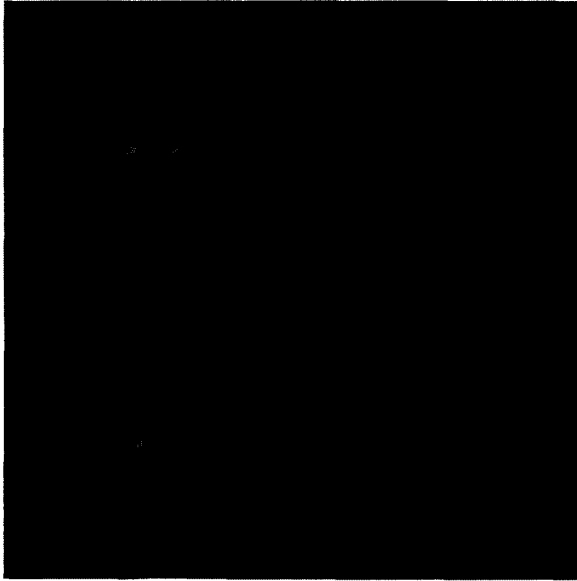
Fig. 6. clockwise (a) The first, (b) tenth, andthe (c) synthesized tenth frame of the real "Miss America" sequenceincluding the photometric effects.
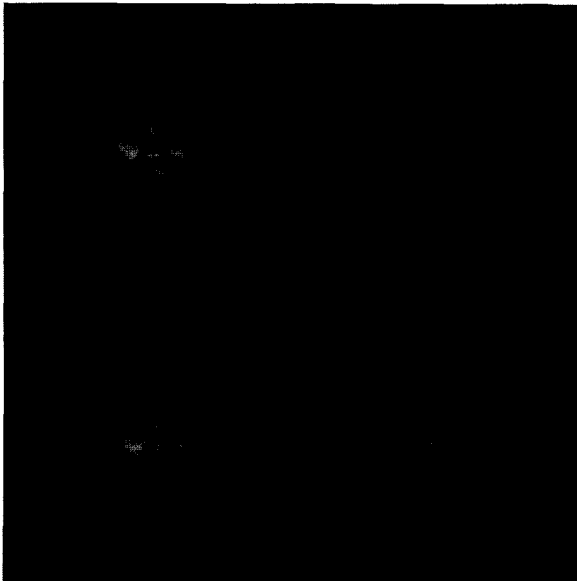

Fig. 7. clockwise (a) The first, (b) tenth, and the (c) synthesized tenth frame of the real "Miss America" sequencewithout the photometric effects.

and second frames in our algorithm. Here, there exists both global and local motion. Further, since the second frame is not artifically synthesized, there is also additional wireframe misfit error. To see the importance of incorporating the photometric effects into the optical flow equation, we repeat the experiment once ignoring the photometric effects and once taking them into account. Tables VII and VIII show the estimated global

TABLE VI
GLOBAL AND LOCAL MOTION ESTIMATION WITH THE SIMULATED MISS AMERICA SEQUENCE INCLUDING THE PHOTOMETRIC EFFECTS

	True motion	Initial point	Our method	Percentage Error
ω_x	-0.1	-0.08894	-0.1079	7.90
ω_y	0.35	0.3368	0.34001	2.85
ω_z	-0.03	-0.0113	-0.0272	9.28
T_x	6	4.962	6.4660	7.76
T_y	-3	-2.8999	-2.8852	3.82

TABLE VII
THE ESTIMATED GLOBAL MOTION PARAMETERS WITH THE REAL MISS AMERICA SEQUENCE INCLUDING THE PHOTOMETRIC EFFECTS

	Initial point	Our method
w_x	-0.053	-0.095
w_y	0.114	0.235
w_z	0.035	0.024
T_x	0.908	0.857
T_y	1.4827	2.1505

TABLE VIII
THE ESTIMATED GLOBAL MOTION PARAMETERS WITH THE REAL MISS AMERICA SEQUENCE WITHOUT THE PHOTOMETRIC EFFECTS

	Initial point	Our method
w_x	-0.053	-0.1098
w_y	0.114	0.204
w_z	0.035	0.0226
T_x	0.908	0.7648
T_y	1.4827	2.6744

motion parameters with and without considering the photometric effects, respectively. The synthesized images using the estimated motion and structure parameters are depicted in Fig. 6 and Fig. 7 and the MSE are 5.81658 and 6.23229, respectively. The MSE's are calculated only using the region of the image pasted on the wireframe.

Motion Tracking Experiments With the "Claire" Sequence: Finally, we test our method on a longer sequence using the first 16 frames of the "Claire" sequence. Every second frame of this sequence is shown in Fig. 8. In the beginning Claire looks straight into the camera. Later, she turns her head to the left. There are also some facial movements such as blinking and openning of the mouth. A frame consists of 352 × 288 pixels (which is the CIF picture format). Only the luminance component is used in the experiments. Our task is to track the motion of the head and extract the facial expressions. Fig. 9 shows the synthesized image sequence using the estimated parameters. The MSE for the frames used in the experiment are 6.98, 7.17, 8.06, 8.27, 8.21, 8.77, 9.48. Upon viewing the two sequences, the difference is hardly noticable.

VI. DISCUSSION AND CONCLUSION

Our goal in this paper has been to estimate the 3-D structure and motion of an head and shoulder type scene with the highest possible accuracy based on two frames in order to predict the second frame from the first. We start with a wireframe model developed for a generic head and shoulders type scene. We address the problem of fitting this generic wireframe model to the actual speaker in the scene, by flexibly adjusting its configuration in both X, Y, and Z directions,

Fig. 8. Every second frame of the first 16 frames of the original "Claire" image sequence.
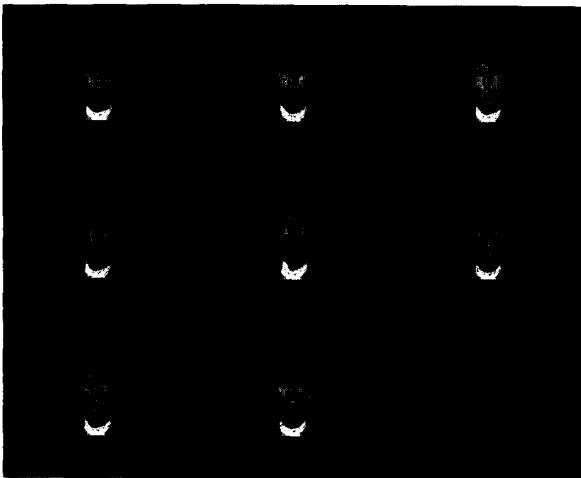

Fig. 9. The synthesized "Claire" sequence using the estimated motion and structure parameters.

simultaneously with estimating its 3-D motion based on two frames. The proposed algorithm not only minimizes the misfit of the wireframe model to the speaker in the first frame but also adjusts the configuration of the wireframe model to adapt for the local motion deformations from the first to the second frame. The simultaneous estimation framework developed in this paper promises the highest possible estimation accuracy due to the inherent connection between the global motion estimation, local motion estimation and wireframe model adaptation in 3-D model-based coding.

This problem could also have been formulated to minimize the synthesis (prediction) error between the actual second frame and its 3-D model-based synthesis. However, this would require a synthesis step after each perturbation which may be costly in terms of the computational load of the algorithm. The proposed optical flow based formulation avoids this synthesis step while allowing us to incorporate the photometric effects

of the motion into the formulation. The representation of the structure of the wireframe model in terms of the surface normal and the displacement in the z-direction of each patch facilitates incorporation of the photometric effects while allowing us to flexibly adjust the structure of the wireframe model. Note that the structure of the wireframe (the location of the nodes) enters the cost function (21) in terms of the Z coordinate of each point on the wireframe.

The validity of the estimated 3-D motion and structure parameters are tested in terms of the sythesis error at convergence. The synthesis error is in general due to (i) misfit of the wireframe model to the actual speaker, (ii) 3-D global motion estimation error, (iii) 3-D local motion estimation error, (iv) error in the estimation of photometric effects of the motion, and (v) change in the texture as a result of the motion. Note that the formulation proposed in this paper takes the first four causes into account while the texture changes would require texture updating which is not addressed here. Experimental results indicate that the resulting synthesis errors are within acceptable limits for both simulated and real image sequences.

Finally, we note that the goal of this work has been to compress head and shoulder type image sequence at about 10 kbit/s. However, we have not addressed the actual coding of the model parameters in this paper. Work in this area is in progress.

REFERENCES

[1] R. Forchheimer and T. Kronander, "Image coding-from waveforms to animation," *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. 37, no. 12, pp. 2008–2023, Dec. 1989.

[2] H. G. Musmann, M. Hotter and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images," *Signal Processing: Image Communication*, vol. 1, pp. 117–138, 1989.

[3] K. Aizawa, H. Harashima and T. Saito, "Model-based analysis-synthesis image coding (MBASIC) system for a person's face," *Signal Processing: Image Communication*, vol. 1, pp. 139–152, 1989.

[4] T. S. Huang, S. C. Reddy and K. Aizawa, "Human facial motion modeling, analysis, and synthesis for video compression," *SPIE Visual Comm. and Image Proc'91*, vol. 1605, pp. 234–241, Nov. 1991.

[5] W. J. Welsh, "Model-based coding of videophone images," *Electronics and Communication Engineering Journal*, pp. 29–36, Feb. 1991.

[6] M. Kaneko, A. Koike and Y. Hatori, "Coding of facial image sequence based on a 3D model of the head and motion detection," *J. Visual Comm. and Image Rep.*, vol. 2, no. 1, pp. 39–54, Mar. 1991.

[7] H. Morikawa and H. Harashima, "3D structure extraction coding of image sequences," *J. Visual Comm. and Image Rep.*, vol. 2, no. 4, Dec. 1991, pp. 332–344.

[8] K. Aizawa, C. S. Choi, H. Harashima and T. S. Huang, "Human facial motion analysis and synthesis with application to model-based coding," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, ed., Kluwer Academic Publishers, 1993.

[9] H. Li, P. Roivainen and Forcheimer, "3-D Motion Estimation in Model-Based Facial Image Coding," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 15, pp. 545–555, Jun. 1993.

[10] N. Diehl, "Model-Based Image Sequence Coding," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, ed., Kluwer Academic Publishers, 1993.

[11] J. Ostermann, "Analysis-synthesis Coder based on Moving Flexible 3-D Objects," in *Picture Coding Symp.* (PCS-93) p. 2.8, Mar. 1993.

[12] G. Bozdağı, A. M. Tekalp and L. Onural, "An Improvement to MBASIC Algorithm for 3-D Motion and Depth Estimation," accepted for publication in the *IEEE Trans. Image Proc.*, Special Issue on Image Sequence Processing, Jun. 1994.

[13] CCITT Recommendation H.261, Video Codec for Audiovisual Services at $p \times 64$Kbit/s, COM XV-R 37-E, 1990.

[14] M. J. T. Reinders, B. Sankur and J. C. A. van der Lubbe, "Transformation of a general 3D facial model to an actual scene face," *11th Int. Conf. Pattern Recog.*, pp. 75–79, 1992.

[15] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images—A review," *Proc. IEEE*, vol. 76, no. 8, pp. 917–935, Aug. 1988.

[16] A. Verri and T. Poggio, "Motion Field and Optical Flow: Qualitative Properties," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-11, no. 5, pp. 490–498, May 1989.

[17] D. Pearson, "Texture Mapping in Model-Based Image Coding," *Signal Processing: Image Comm.* Vol. 2, pp. 377–395, 1990.

[18] A. Pentland, "Photometric Motion," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-13, no. 9, pp. 879–890, Sept. 1991.

[19] J. N. Driessen, "Motion Estimation for Digital Video," *Ph.D. Thesis*, Delft University of Technology, Department of Electrical Engineering, Information Theory Group, Delft, The Netherlands, Sept. 1992.

[20] B. K. P. Horn and B. G. Schunk, "Determining optical flow," *Artif. Intell.*, 17, pp. 185–203, 1981.

[21] K. Kanatani, "Structure and motion from optical flow under orthographic projection," *Comp. Vision Graph. Image Proc.*, 35, pp. 181–199, 1986.

[22] B. Klaus and P. Horn, Robot Vision, MIT Press, 1986.

[23] A. Pentland, "Finding the illuminant direction," *J. Opt. Soc. Am.*, vol. 72, no. 4, pp. 448–455, Apr. 1982.

[24] H. Lee and A. Rosenfeld, "Improved methods of estimating shape from shading using light source coordinate system," in *Shape from Shading*, (B. K. P. Horn and M. J. Brooks, eds.), Cambridge, MA: MIT Press, pp. 323–569, 1989.

[25] Q. Zheng and R. Chellappa, "Estimation of illuminant direction, albedo, and shape from shading," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-13, no. 7, pp. 680–702, Jul. 1991.

[26] K. Zeger and A. Gersho, "Stochastic relaxation algorithm for improved vector quantizer design," *Electronic Letters*, vol. 25, no. 14, pp. 96–98, Jul. 1989.

[27] M. Rydfalk, "CANDIDE: A parametrised face," Dept. Elec. Eng. Rep. LiTH-ISY-I-0866, Linköping Univ., Oct. 1987.

A. Murat Tekalp (S'84–M'85–SM'91) was born in 1958. He received B.S. degree in Electrical Engineering, and B.S. degree in Mathematics from Boğaziçi University, Istanbul, Turkey in 1980, with the highest honors, and M.S. and Ph.D. degrees in Electrical, Computer and Systems Engineering from Rensselaer Polytechnic Institute (RPI), Troy, New York, in 1982 and 1984, respectively.

From December 1984 to August 1987, he was a research scientist, and then a senior research scientist at Eastman Kodak Company, Rochester, New York. He joined the Electrical Engineering Department at the University of Rochester, Rochester, New York, as an assistant professor in September 1987, where he is currently an associate professor. He has been a visiting assistant professor at Rensselaer Polytechnic Institute from January 1987 to June 1987, and a visiting Associate Professor at Bilkent University, Ankara, Turkey from September 1992 to June 1993. His current research interests are in the area of digital image and video processing, including image restoration, motion and structure estimation, model-based coding, image segmentation, and magnetic resonance imaging.

Dr. Tekalp is a senior member of IEEE, and a member of SPIE and Sigma Xi. He has been a scholar of the Scientific and Technical Research Council of Turkey between 1978-80. He received the NSF Research Initiation Award in 1988, and IEEE Rochester Section Awards in 1989 and 1992. He has served as an Associate Editor for the IEEE Transactions on Signal Processing (1990-1992), and as the Chair of the Technical Program Committee for the 1991 MDSP Workshop sponsored by the IEEE Signal Processing Society. He has been the organizer and first Chairman of the Rochester Chapter of the IEEE Signal Processing Society. At present he is the Vice-Chair of the IEEE Signal Processing Society Technical Committee on Multidimensional Signal Processing, and an Associate Editor for IEEE Trans. on Image Processing, and the Multidimensional Systems and Signal Processing Journal. He is also the Vice-Chair of the Rochester Section of IEEE.

Gözde Bozdağı was born in Ankara, Turkey in 1967. She received B.S. degree from the Middle East Technical University, Ankara, Turkey, M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1988, 1990 and 1994, respectively, all in electrical and electronics engineering. She is currently a visiting research associate at the Electrical Engineering Department, University of Rochester. Her current research interests include image coding, 3D motion estimation and holography. She is a member of IEEE, ACM and SPIE.

Levent Onural was born in İzmir, Turkey in 1957. He received the B.S. and M.S. degrees in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1979 and 1981, respectively, and the Ph.D. degree in electrical and computer engineering from the State University of New York at Buffalo in 1985. He was a Fulbright Scholar between 1981 and 1985. After a Research Assistant Professor position at the Electrical and Computer Engineering Department of the State University of New York at Buffalo, he joined the Electrical and Electronics Engineering Department of Bilkent University, Ankara, Turkey, where he is an Associate Professor at present. His present research interests are in signal processing, image processing, holography, and fractal models.