

# Least Squares Estimation of 3D Shape and Motion of Rigid Objects from Their Orthographic Projections

Yiannis Xirouhakis, *Student Member, IEEE*, and  
Anastasios Delopoulos, *Member, IEEE*

**Abstract**—The extraction of motion and shape information of three-dimensional objects from their two-dimensional projections is a task that emerges in various applications such as computer vision, biomedical engineering, and video coding and mining especially after the recent guidelines of the Motion Pictures Expert Group regarding MPEG-4 and MPEG-7 standards. Present work establishes a novel approach for extracting the motion and shape parameters of a rigid three-dimensional object on the basis of its orthographic projections and the associated motion field. Experimental results have been included to verify the theoretical analysis.

**Index Terms**—3D motion, 3D structure, structure from motion, orthography.

## 1 INTRODUCTION

MUCH work has been done recently for determining three-dimensional motion and structure of moving rigid objects viewed at different time points and/or by multiple cameras [8], [17]. In particular, the extraction of motion and shape parameters of a moving rigid 3D object from a 2D image sequence (often named as the Structure From Motion problem) has received considerable attention lately in relation to the new object-based coding standards MPEG-4 and MPEG-7. Various approaches have been proposed to this problem, which differ in the projection model assumed, the feature correspondences and the input measurements employed, and the adopted data-processing technique [16]. As far as the 2D features are concerned, line, curve, and point correspondences have been utilized [6], with the latter being the most popular. Two well-known projection models mainly considered in the literature are the perspective and orthographic, with the latter assumed when the object is far away from the camera [8], [17]. Exact and approximate mathematical solutions have been reported for both cases, including for example, [10], [19], [21], [13] for the perspective and [7], [1], [18] for the orthographic case. A few deviations based on approximations of these models, such as: the weak perspective/paraperspective and orthoperspective projections have also been treated (see, e.g., [5] and references therein).

Regarding the case of orthographic projections, Ullman, in his classical work [20], proved that four point correspondences over three frames are sufficient to yield a unique solution to motion and structure up to a reflection. Huang and Lee in [7], proposed a linear algorithm to obtain the 3D motion and structure parameters. Aizawa et al. in [1], use small Eulerian angles and a two-step iteration for 3D motion and depth estimation, where they assume initial depth estimates based on an a priori known 3D model. In [2], Bozdagi et al. define an error criterion and propose a gradient search add-on. Later approaches on the Structure From Motion

problem under orthography include: the factorization method of Tomasi and Kanade [18] and its derivatives [14], [11], as well as the epipolar methods of Shapiro et al. [15], Ostuni and Dunn [12], and Xu and Sugimoto [23]. The epipolar methods generally refer to weak perspective images, however, it can be seen that this model differs from the orthographic one, only in the sense that it permits a scale change between different views. Tomasi and Kanade's solution in [18] is based on a camera-centered problem representation, which may incorporate an arbitrary number of point correspondences and frame transitions to achieve robustness in the presence of noise. Shapiro et al. rely on the affine epipolar line properties and solve the affine epipolar line equation. The next step determines all unknown camera motion parameters (see [15] and references therein). In the same manner, Xu and Sugimoto [23] solve the epipolar equation and determine the three rotation angles in a second step. Ostuni and Dunn [12] utilize the epipolar equation as well along with a different parametrization for the rotation matrix.

In this work, we follow the orthographic projections approach. First, we prove that the elements of the rotation matrix can be computed via the eigenvalues and eigenvectors of appropriately defined  $2 \times 2$  matrices; the latter are simple expressions of four motion vectors in two successive transitions. Some of the key-points of the analysis were first introduced in [4]. In the sequel, we point out that the derived expressions can be used for unbiased, least squares estimation of the involved rotation matrices in the case that more than four motion vectors per transition are available. Finally, appropriate simulation results have been included to both verify the theoretical propositions and test the algorithm's performance against popular algorithms for the estimation of 3D motion and structure under orthography [18], [23], [15].

## 2 PRELIMINARY DEFINITIONS

Any movement of a rigid object in the 3D space is a superposition of a 3D rotation and a 3D translation. Consequently, when a point  $\mathbf{p} = [x, y, z]^T$  on the object moves to  $\mathbf{p}' = [x', y', z']^T$ , it holds:

$$\mathbf{p}' = \mathbf{R}\mathbf{p} + \mathbf{T}, \quad (1)$$

where  $\mathbf{R}$ ,  $\mathbf{T}$  are the rotation and translation matrices, respectively. In the sequel, whenever we refer to their elements, we assume that  $\mathbf{R} = [r_{mn}]$  and  $\mathbf{T} = [t_1 t_2 t_3]^T$ . Rotation matrix elements  $r_{mn}$  obey to orthogonality equations imposed by the fact that  $\mathbf{R}$  is unitary.

We define matrices  $\tilde{\mathbf{K}}$ ,  $\mathbf{K}$  as,

$$\Delta \mathbf{r}' = \mathbf{K} \Delta \mathbf{r}, \quad \Delta \mathbf{v} = \tilde{\mathbf{K}} \Delta \mathbf{r}, \quad (2)$$

where

$$\Delta \mathbf{r} = [\mathbf{r}_3 - \mathbf{r}_1 \quad \mathbf{r}_2 - \mathbf{r}_1], \quad \Delta \mathbf{r}' = [\mathbf{r}'_3 - \mathbf{r}'_1 \quad \mathbf{r}'_2 - \mathbf{r}'_1],$$

and

$$\Delta \mathbf{v} = [\mathbf{v}_3 - \mathbf{v}_1 \quad \mathbf{v}_2 - \mathbf{v}_1],$$

vectors  $\mathbf{v}_i$ ,  $\mathbf{r}_i$  being the motion and reference vectors, respectively, for three points on the first frame and vectors  $\mathbf{r}'_i$  being the reference vectors of the corresponding points on the subsequent frame. It can be seen that matrices  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$  are the same for any three points on the same planar 3D surface (*face*). Moreover, it can be verified that

• The authors are with the Image, Video, and Multimedia Systems Laboratory, Department of Computer and Electrical Engineering, National Technical University of Athens, 9, Iroon Polytechniou str., GR-15773 Athens, Greece. E-mail: {jxiro, adelo}@image.ece.ntua.gr.

Manuscript received 2 Nov. 1998; revised 21 Sept. 1999; accepted 26 Oct. 1999.

Recommended for acceptance by Y.-F. Wang.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108161.

$$\mathbf{K} = \mathbf{R}_{2 \times 3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ p & q \end{bmatrix}, \quad (3)$$

where, for  $\xi = [\xi_x \ \xi_y \ \xi_z]^T$  representing the unit vector perpendicular to the face, the scalars  $p = -\xi_x/\xi_z$ ,  $q = -\xi_y/\xi_z$  contain the orientation information of the face, and  $\mathbf{R}_{2 \times 3}$  contains the first two rows of the  $3 \times 3$  rotation matrix (see also [4]). In the sequel, adjoint matrix  $\mathbf{L} = \text{adj}(\mathbf{K}) \equiv \text{adj}(\tilde{\mathbf{K}} + \mathbf{I})$  will be used to obtain simpler equations.

### 3 COMPUTATION OF ROTATION AND SHAPE PARAMETERS

Four  $\mathbf{L}$ -matrices contain sufficient information in order to determine the motion and structure parameters of a rigid object. In particular, if  $\mathbf{L}_{aj}$ ,  $a = R, S$ ,  $j = 1, 2$  are available (indices  $R, S$  denoting the rotation matrices corresponding to movement from frame 1 to 2 and 1 to 3, respectively and  $j = 1, 2$  denoting two faces) exact recovery of  $\mathbf{R} = [r_{mn}]$  and  $\mathbf{S} = [s_{mn}]$  is possible. In addition, relative 3D coordinates of the projected points of the rigid object can be computed. In this section, we establish a set of propositions that allow for the estimation of rotation matrices  $\mathbf{R}$  and  $\mathbf{S}$  w.r.t. the four aforementioned matrices  $\mathbf{L}_{aj}$ .

Based on (3), Propositions 1, 2, and 3 allow the computation of rotation matrices  $\mathbf{R}, \mathbf{S}$  on the basis of  $\mathbf{L}_{Rj}$  and  $\mathbf{L}_{Sj}$  of the faces  $j = 1, 2$ . As a first step, Proposition 1 allows for the computation of vectors  $\mathbf{r}_1 \triangleq [r_{13} \ r_{23}]^T$  and  $\mathbf{r}_2 \triangleq [r_{31} \ r_{32}]^T$  within a scalar ambiguity factor  $\rho$ , based on information contained solely in the motion vectors (equivalently matrices  $\mathbf{L}_{Rj}$ ) corresponding to the rotation  $\mathbf{R}$  (two frames). Similarly, vectors  $\mathbf{s}_1 \triangleq [s_{13} \ s_{23}]^T$  and  $\mathbf{s}_2 \triangleq [s_{31} \ s_{32}]^T$  are separately computed within a common scalar ambiguity factor  $\sigma$  based on the rotation  $\mathbf{S}$ . As a second step, Proposition 2 provides a simple expression of the ratio  $w = \frac{\sigma}{\rho}$  of the unknown factors  $\rho$  and  $\sigma$ . This is a necessary, intermediate result. The ratio  $w$  is used next in Proposition 3 in order to compute  $r_{33}$  and  $s_{33}$  by solving a  $2 \times 2$  system of linear equations. Then,  $\rho^2 = 1 - (r_{33})^2$  and  $\sigma^2 = 1 - (s_{33})^2$  and all rotation elements are computed.

**Proposition 1.** Let  $\mathbf{r}_1 \triangleq [r_{13} \ r_{23}]^T$ ,  $\mathbf{r}_2 \triangleq [r_{31} \ r_{32}]^T$ ,

$$\mathbf{s}_1 \triangleq [s_{13} \ s_{23}]^T, \mathbf{s}_2 \triangleq [s_{31} \ s_{32}]^T,$$

and

$$\Delta \mathbf{L}_R \triangleq \mathbf{L}_{R1} - \mathbf{L}_{R2}, \quad \Delta \mathbf{L}_S \triangleq \mathbf{L}_{S1} - \mathbf{L}_{S2}.$$

Then,

1. It holds,  $\|\mathbf{r}_1\|_2^2 = \|\mathbf{r}_2\|_2^2 = \rho^2 = 1 - (r_{33})^2$  and

$$\|\mathbf{s}_1\|_2^2 = \|\mathbf{s}_2\|_2^2 = \sigma^2 = 1 - (s_{33})^2,$$

and also

$$\mathbf{L}_{Rj}\mathbf{r}_1 = -\mathbf{r}_2, \quad \mathbf{L}_{Sj}\mathbf{s}_1 = -\mathbf{s}_2, \quad j = 1, 2.$$

2. If  $\mathbf{c}_1, \mathbf{d}_1$  are the unit-norm eigenvectors of the rank-1 positive semidefinite matrices  $(\Delta \mathbf{L}_R)^T (\Delta \mathbf{L}_R)$ ,  $(\Delta \mathbf{L}_S)^T (\Delta \mathbf{L}_S)$ , respectively, that correspond to their zero eigenvalues, then  $\mathbf{r}_1, \mathbf{s}_1$  can be computed within a scalar ambiguity from  $\mathbf{r}_1 = \rho \mathbf{c}_1$ ,  $\mathbf{s}_1 = \sigma \mathbf{d}_1$ .

3. Also  $\mathbf{r}_2, \mathbf{s}_2$  can be computed within the same scalar factors as  $\mathbf{r}_2 = \rho \mathbf{c}_2$ ,  $\mathbf{s}_2 = \sigma \mathbf{d}_2$ , where  $\mathbf{c}_2 \triangleq -\mathbf{L}_{Rj}\mathbf{c}_1$ ,  $\mathbf{d}_2 \triangleq -\mathbf{L}_{Sj}\mathbf{d}_1$  for both  $j = 1, 2$ .

Let  $\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ . After some manipulations, using the rotation matrix properties and the definitions of Proposition 1, Propositions 2 and 3 can be obtained.

**Proposition 2.** The ratio  $w^2 = (\frac{\sigma}{\rho})^2$  equals  $\frac{\lambda_S}{\lambda_R}$ , where  $\lambda_R, \lambda_S$  are the nonzero eigenvalues of matrices  $(\Delta \mathbf{L}_R)^T (\Delta \mathbf{L}_R)$ ,  $(\Delta \mathbf{L}_S)^T (\Delta \mathbf{L}_S)$ , respectively. Thus,  $w = \sigma/\rho$  can be determined within a sign ambiguity. It can be shown that one of the solutions is rejected.

**Proposition 3.** The unknown elements  $r_{33}, s_{33}$  of the rotation matrices can be uniquely determined by:

$$[w\mathbf{J}\mathbf{c}_2 \quad -\mathbf{J}\mathbf{d}_2] \begin{bmatrix} r_{33} \\ s_{33} \end{bmatrix} = [\mathbf{L}_{Sj} \quad \mathbf{L}_{Rj}] \begin{bmatrix} \mathbf{J}\mathbf{d}_1 \\ -w\mathbf{J}\mathbf{c}_1 \end{bmatrix}. \quad (4)$$

At this point, rotation matrices  $\mathbf{R}, \mathbf{S}$  are computed. It can be pointed out here that  $\mathbf{r}_1, \mathbf{r}_2$  accept two real valued solutions corresponding to reflection of the rigid object w.r.t. the image plane. Once  $\mathbf{R}, \mathbf{S}$  have been computed, the estimation of the translation vectors and the structure parameters is straightforward. In order to suppress the projection model's ambiguities in absolute point depth ( $z$ -coordinate) and in translation along the  $z$ -axis, the object geometric center is assumed to coincide with the world origin  $(0, 0, 0)$  without loss of generality. Then, translation  $\mathbf{T}$  is computed for each scene and relative depth information for all available points is given from (1).

The proposed solution to the SFM problem under orthography, as introduced by Propositions 1, 2, and 3 verifies the assertion of [7] that four accurate point correspondences over three frames are sufficient for the computation of  $\mathbf{R}$  and  $\mathbf{S}$ . In practice, motion vector estimation techniques may introduce errors in matrices  $\mathbf{L}_{Rj}, \mathbf{L}_{Sj}$ . These errors are either due to limitations of the estimation procedure or due to the resolution of the image plane grid. For this purpose, it is essential that more than four point correspondences per transition are utilized for the estimation of 3D motion parameters. In this sense, more than two  $\mathbf{L}$ -matrices and one  $\Delta \mathbf{L}$ -matrix should be employed. In the following, index  $j = 1 \dots N$  will refer to  $\mathbf{L}$ -matrices, whereas index  $k = 1 \dots M$  will refer to  $\Delta \mathbf{L}$ -matrices. Scalars  $M, N$  will represent the number of  $\Delta \mathbf{L}_k$  and  $\mathbf{L}_j$  matrices involved in the computations, while  $P$  will denote the number of points (and associated motion vectors) used. The relation between  $M$  and  $N$  varies along with the formation strategy of  $\Delta \mathbf{L}_k$  from  $\mathbf{L}_j$  matrices, while the relation between  $N$  and  $P$  depends on the formation strategy of  $\mathbf{L}_j$  matrices from the available points. In general,  $M = M(N) \cong N/2$  and  $N = N(P) \cong P/3$ , since we choose that

- A1. Each  $\mathbf{L}_j$  is expressed on distinct reference points and motion vectors,
- A2. Each  $\Delta \mathbf{L}_k$  is expressed on the basis of distinct matrices  $\mathbf{L}_j$  (see also [22]).

Under the Assumptions A1 and A2 in the formulation of  $\mathbf{L}_j, \Delta \mathbf{L}_k$  matrices, it can be shown that additive error terms in matrices  $\Delta \mathbf{L}_k^T \Delta \mathbf{L}_k$  are in the form of  $\sigma_k^2 \mathbf{I}$ , where

$$\sigma_k^2 \triangleq \pi_j^2 + \pi_i^2, \quad (5)$$

$$\pi_j^2 \triangleq \frac{\sigma_e^2 (\|\mathbf{r}_3^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_2^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_3^j - \mathbf{r}_2^j\|^2)}{|\det[\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]|^2}, \quad (6)$$

and  $\mathbf{r}_1^j, \mathbf{r}_2^j, \mathbf{r}_3^j$  is a collection of reference point triplets ( $j = 1 \dots N$ ). Scalar  $\sigma_e^2$  denotes the error variance on each motion vector component, assuming that the error variance is identical to both directions.

As it can be seen, the power of error terms  $\sigma_k^2, \pi_j^2$  is greatly affected by the strategy adopted in the formulation of  $\mathbf{L}_j$  and  $\Delta\mathbf{L}_k$  matrices. Moreover, as it will be shown in the sequel, such terms are conveniently encapsulated into distinguishable quantities in the proposed method.

In fact, Propositions 1, 2, and 3 can be slightly modified to include more than four point correspondences per transition. To this end, we define the following  $2 \times 2$  matrices,

$$\mathbf{Z}^{(N)} \triangleq \frac{1}{N} \sum_{j=1}^N \mathbf{L}_j, \quad (7)$$

$$\mathbf{Y}^{(M)} \triangleq \frac{1}{M} \sum_{k=1}^M \Delta\mathbf{L}_k^T \Delta\mathbf{L}_k. \quad (8)$$

Matrices  $\mathbf{Y}^{(M)}, \mathbf{Z}^{(N)}$  are defined for each transition; indices  $R, S$  in the rotation matrices will be omitted for simplicity when possible. Based on these definitions, the following Propositions 4, 5, and 6 establish a procedure that allows for the computation of rotation matrices  $\mathbf{R}$  and  $\mathbf{S}$  on the basis of finitely many motion vector estimates. In this sense, Propositions 4, 5, and 6 extend (in fact, replace) Propositions 1, 2, and 3.

**Proposition 4.** Let  $\lambda_{Rk}, \lambda_{Sk}$  be the maximum eigenvalues of  $(\Delta\mathbf{L}_{Rk})^T (\Delta\mathbf{L}_{Rk}), (\Delta\mathbf{L}_{Sk})^T (\Delta\mathbf{L}_{Sk})$  for  $k = 1 \dots M$ . Let also  $\sigma_{Rk}^2, \sigma_{Sk}^2$  be defined as in (5). Then,

1. The eigenvalues of  $\mathbf{Y}_R^{(M)}$  are  $\lambda_{Rmax} \cong \lambda_R + \sigma_R^2$  and  $\lambda_{Rmin} \cong \sigma_R^2$  corresponding to the unit-norm eigenvectors  $\mathbf{Jc}_1$  and  $\mathbf{c}_1$ , respectively, where

$$\lambda_R = \frac{1}{M} \sum_{k=1}^M \lambda_{Rk} \text{ and } \sigma_R^2 = \frac{\sum_{k=1}^M \sigma_{Rk}^2}{M}.$$

2.  $\mathbf{c}_2 = -\mathbf{Z}_R^{(N)} \mathbf{c}_1$  if we replace  $\mathbf{L}_{Rj}$  by  $\mathbf{Z}_R^{(N)}$ .
3. Similar relations hold for matrix  $\mathbf{S}$  with appropriately defined  $\mathbf{Y}_S^{(M)}, \mathbf{Z}_S^{(N)}$ .

It must be pointed out here that the estimation of  $\mathbf{c}_1$  along the lines of the above method is the least-squares solution of the over-determined system,

$$[(\Delta\mathbf{L}_{R1}^T \Delta\mathbf{L}_{R1})^T \quad \dots \quad (\Delta\mathbf{L}_{RM}^T \Delta\mathbf{L}_{RM})^T]^T \mathbf{c}_1 = \mathbf{0},$$

subject to the constraint  $\|\mathbf{c}_1\|^2 = \mathbf{c}_1^T \mathbf{c}_1 = 1$ .

**Proposition 5.** If  $\lambda_{Rmax}$  ( $\lambda_{Smax}$ ) and  $\lambda_{Rmin}$  ( $\lambda_{Smin}$ ) are defined as in Proposition 4, scalar  $w$  is now given by

$$w^2 = \frac{\lambda_{Smax} - \lambda_{Smin}}{\lambda_{Rmax} - \lambda_{Rmin}}. \quad (9)$$

It can be seen that  $w$  obeys (9) in the error-free case as well, where  $\lambda_{Rmin} = \lambda_{Smin} = 0$ .

The following Proposition 6 substitutes Proposition 3 for the estimation of the unknown elements  $r_{33}, s_{33}$  of the rotation matrices. Beforehand, some definitions are needed. Equation (4) can be rewritten in the form:

$$- [w \mathbf{JL}_{Rj} \mathbf{c}_1 \quad -\mathbf{JL}_{Sj} \mathbf{d}_1] \begin{bmatrix} r_{33} \\ s_{33} \end{bmatrix} = [\mathbf{L}_{Sj} \quad \mathbf{L}_{Rj}] \begin{bmatrix} \mathbf{Jd}_1 \\ -w \mathbf{Jc}_1 \end{bmatrix}.$$

If  $\theta = [r_{33} \quad s_{33}]^T$  is the vector containing the unknowns, then, we can rewrite the previous equation as  $\mathbf{A}_j \theta = \mathbf{b}_j$ . Utilizing  $N$  triplets of points,

$$\mathbf{A} \theta = \mathbf{b} \quad (10)$$

which is to be solved in the least-squares sense.

**Proposition 6.** The least-squares solution of (10) can be asymptotically approximated by the estimator

$$\theta_{LS}^{(N)} = \frac{1}{N} \left( \frac{1}{N} \mathbf{A}^T \mathbf{A} - \begin{bmatrix} w^2 \pi_R^2 & 0 \\ 0 & \pi_S^2 \end{bmatrix} \right)^{-1} \mathbf{A}^T \mathbf{b}, \quad (10)$$

where

$$\pi_R^2 = \frac{\sum_{j=1}^N \pi_{Rj}^2}{N}, \quad \pi_S^2 = \frac{\sum_{j=1}^N \pi_{Sj}^2}{N},$$

are correction factors which account for the presence of error in the estimates. Under the assumption that  $\Delta\mathbf{L}_k$ -matrices are computed on distinct  $\mathbf{L}$ -matrices,

$$\pi_R^2 = \frac{M}{N} \sigma_R^2 \equiv \frac{M}{N} \lambda_{Rmin}$$

and

$$\pi_S^2 = \frac{M}{N} \sigma_S^2 \equiv \frac{M}{N} \lambda_{Smin}.$$

Table 1 summarizes the steps of the algorithm.

## 4 SIMULATIONS

The experimental results presented in this section exhibit the ability of the proposed approach to recover rotation and shape information even on the basis of noisy motion vector fields.

In Fig. 1a, a computer generated object consisting of two polygons is depicted. Considering that the object's reference position is that of Fig. 1a, the object was rotated by  $\phi_R = 4^\circ$  and  $\phi_S = 7^\circ$  w.r.t. axes  $\mathbf{u}_R = [.9129 \quad .3651 \quad .1826]^T$  and  $\mathbf{u}_S = [.6172 \quad .7715 \quad .1543]^T$ . The projected motion fields were next fed to the algorithm. Fig. 1b depicts an indicative portion of the motion field corresponding to movement from frame one to two. In accordance to the theoretic results of Section 3, in this experiment the algorithm yields exact estimates. The motion parameters are estimated to be identical to the ones used to produce the three scenes and the reconstructed object is identical to that depicted in Fig. 1a.

In the sequel, the motion fields were artificially disturbed by 10db uniformly distributed random noise. The SNR level is in our setup defined as  $SNR_v = 10 \log_{10}(\frac{\sigma_v^2}{\sigma_e^2})$ , where  $\sigma_v^2 \triangleq \frac{1}{P} \sum_{i=1}^P \mathbf{v}_i^T \mathbf{v}_i$  and  $\sigma_e^2 \triangleq E\{\mathbf{e}_i^T \mathbf{e}_i\}$ ,  $\mathbf{e}_i$  being the error term added to  $\mathbf{v}_i$ . In fact, the performance of the proposed algorithm is influenced directly from the SNR level at the differential motion field rather than the original motion field itself. This indicates that  $SNR_{dv} = 10 \log_{10}(\frac{\sigma_{dv}^2}{\sigma_{de}^2})$ , where

$$\sigma_{dv}^2 \triangleq \frac{1}{P} \sum_{i \neq j} (\mathbf{v}_i - \mathbf{v}_j)^T (\mathbf{v}_i - \mathbf{v}_j)$$

and

$$\sigma_{de}^2 \triangleq E\{(\mathbf{e}_i - \mathbf{e}_j)^T (\mathbf{e}_i - \mathbf{e}_j)\} = 2\sigma_e^2,$$

TABLE 1  
The Algorithm in Steps

<p><b>For each transition:</b></p> <p>(a) Divide available point correspondences in (at least two) triplets and form <math>\mathbf{L}_j</math> (equiv. <math>\mathbf{K}_j</math>) matrices using eqs. (2).</p> <p>(b) Group <math>\mathbf{L}_j</math>-matrices in (at least one) pairs and form <math>\Delta\mathbf{L}_k</math>-matrices using Proposition 1.</p> <p>(c) Form matrix <math>\mathbf{Y}</math> from eq. (8) and compute eigenvalues <math>\lambda_{max}</math>, <math>\lambda_{min}</math> and eigenvector <math>\mathbf{c}_1</math> w.r.t. Proposition 4.</p> <p><b>For a pair of transitions:</b></p> <p>(a) Compute <math>w^2</math> from eq. (9) of Proposition 5.</p> <p>(b) Compute <math>r_{33}</math> for both rotation matrices from eq. (11) of Proposition 6.</p> <p><b>For each transition:</b></p> <p>(a) Compute all rotation matrix elements,</p> <p>(b) Compute translation vector under the assumptions for geometric center,</p> <p>(c) Estimate relative depth from eq. (1).</p>
---

corresponding to a reference differential motion field obtained for step size  $x = 1$  (for example,  $\Delta\mathbf{v}(n, m) = \mathbf{v}(n, m) - \mathbf{v}(n + 1, m)$ ). In fact, as it will be pointed out in the sequel, improved SNR is attained for the same motion field estimates by increasing step size  $x$ . In Fig. 1c, the noise-contaminated counterpart of Fig. 1b is depicted.

Following the methodology of Section 3, relatively accurate estimates of the rotation matrices are obtained, even in the presence of noise. In particular, in this experiment, the estimated rotation parameters were  $\phi_R = 4.2^\circ$ ,  $\phi_S = 7.4^\circ$  w.r.t. and  $\mathbf{u}_R = [.9157 \ .3645 \ .1691]^T$ ,  $\mathbf{u}_S = [.6162 \ .7765 \ .1319]^T$ . The reconstructed object in the first scene is depicted in Fig. 1d. Although the estimated surface is of almost correct orientation, it suffers from significant irregularity in comparison to the original one (Fig. 1a) due to the noise involved in motion estimates that affects the solution of (1) w.r.t. the depth. Various techniques can be considered for smoothing the reconstructed surface. Constraints and/or a prior knowledge regarding the local variation of the 3D surface may be useful to this direction.

In the same context, the reconstructed object for the 3D smooth surface of Fig. 1e is given in Fig. 1f, where again the true motion fields were disturbed by 10db i.i.d. noise. The true motion parameters were identical to the one utilized for the object consisting of two planar surfaces, while the estimated ones were  $\phi_R = 3.7^\circ$ ,  $\phi_S = 6.96^\circ$  w.r.t. and  $\mathbf{u}_R = [.9179 \ .3351 \ .2125]^T$ ,  $\mathbf{u}_S = [.6382 \ .7557 \ .1471]^T$ .

The aforementioned results were obtained after appropriate choice of point triplets,  $\mathbf{L}_j$  and  $\Delta\mathbf{L}_k$  matrices. As proposed in Section 3, it is essential that  $\mathbf{L}_j$  matrices are formed on the basis of distinct triplets of points. In addition, it can be seen that point triplets must be chosen so that the respective  $\pi_j^2$  parameter is minimized. Minimization of  $\pi_j^2$  (and, of  $\sigma_k^2$ , consequently) for a

given  $\sigma_e^2$  is achieved by maximizing the differential reference vectors and, thus, increasing the area of the triangles defined by the point triplets employed. Supposing that point triplets are chosen on orthogonal triangles and that all differential reference vectors are equal to a step  $x$ . Fig. 2 confirms the effectiveness of this assertion. In Fig. 2a and Fig. 2b, the mean (solid line) and standard deviation (dash-dotted line) estimates of the rotation angles  $\phi_R$ ,  $\phi_S$  for 50 Monte Carlo runs are depicted w.r.t. increasing differential reference vectors. In this experiment, the object consisting of two planar surfaces was rotated by angles  $\phi_R = 20^\circ$ ,  $\phi_S = 40^\circ$ .

At the same time, as it is intuitively expected, improved estimates of the rotation parameters are obtained for increasing number of points  $P$  (equiv.  $\mathbf{L}_j$ s). In Fig. 2c and Fig. 2d, the mean (solid line) and standard deviation (dash-dotted line) estimates of the rotation angles  $\phi_R$ ,  $\phi_S$  for 50 Monte Carlo runs are depicted w.r.t. increasing number of employed matrices. For this experiment, point triplets were chosen on orthogonal triangles and differential reference vectors were set equal to step  $x = 30$ .

The proposed algorithm was tested against existing approaches for the estimation of 3D motion and structure under orthography. For this purpose, a factorization method [18] and a method utilizing the epipolar constraint [23], [15] were implemented. The proposed approach seemed to be superior in nearly all simulated experiments held. In Fig. 3, estimates of the rotation angles  $\phi_R$ ,  $\phi_S$  for varying SNR level are depicted for the smooth surface. The latter was rotated to produce two sets of motion fields, which were in turn, artificially contaminated with i.i.d. noise. In Figs. 3a-f, the mean (solid line) and standard deviation (dash-dotted line) estimates using the factorization, the epipolar, and the proposed method, respectively, are depicted. The true rotation angles for

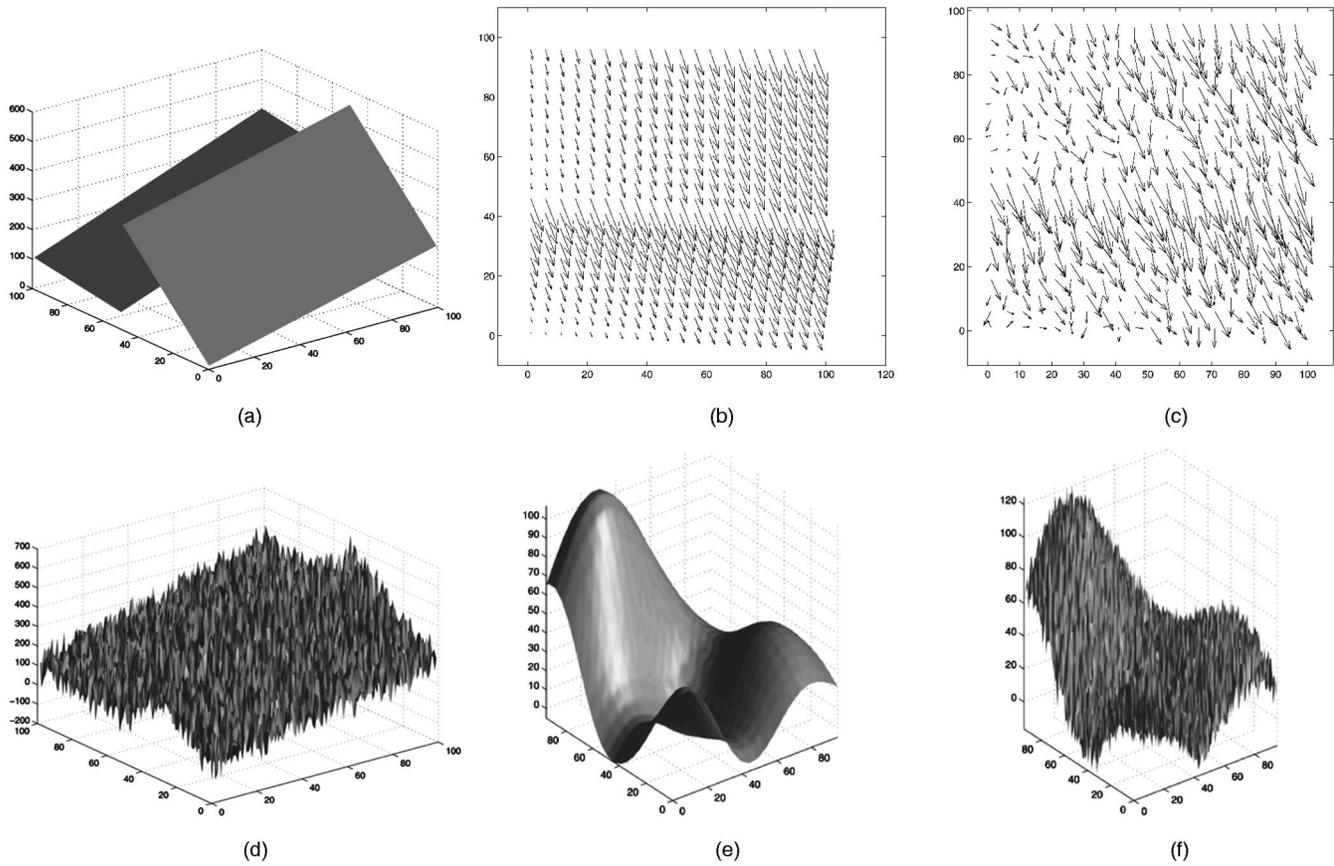


Fig. 1. Reconstruction in the presence of noise: (a) object consisting of two planar surfaces, (b) true motion field, (c) 10db perturbed motion field, (d) reconstructed scene, (e) smooth 3D surface, and (f) reconstructed scene.

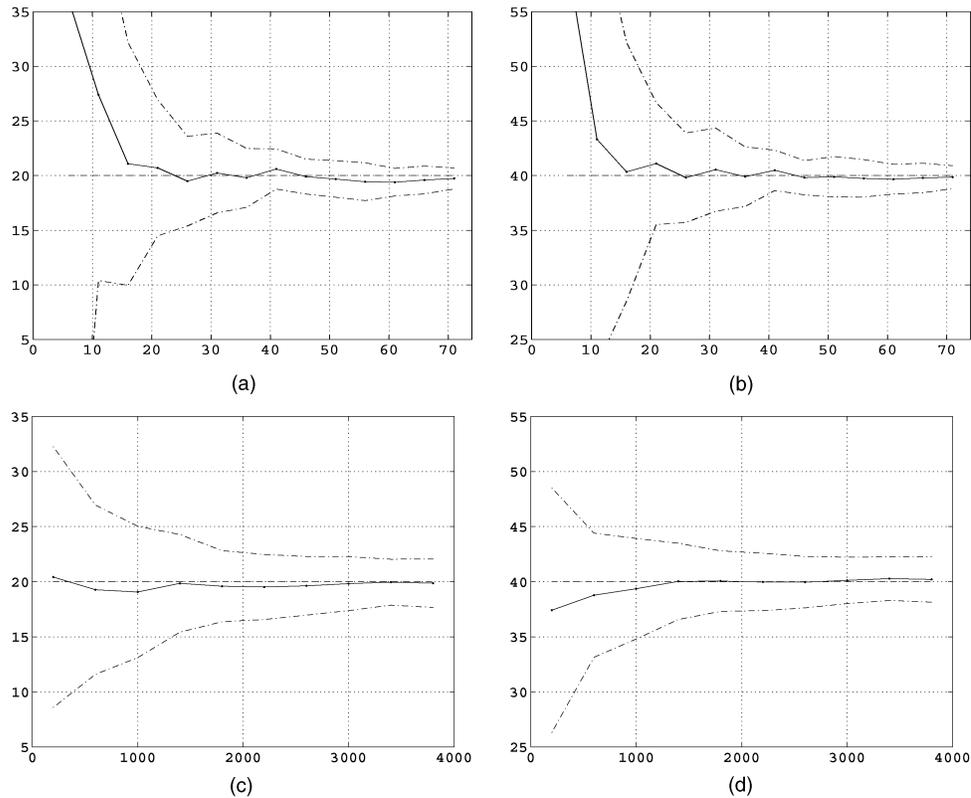


Fig. 2. Motion parameters estimation for object consisting of two planar surfaces for variant step and number of samples, (a) and (b) estimates of rotation angles  $\phi_R, \phi_S$  w.r.t. step size  $x$ , (c) and (d) estimates of rotation angles  $\phi_R, \phi_S$  w.r.t. number of samples ( $L_j$  matrices).

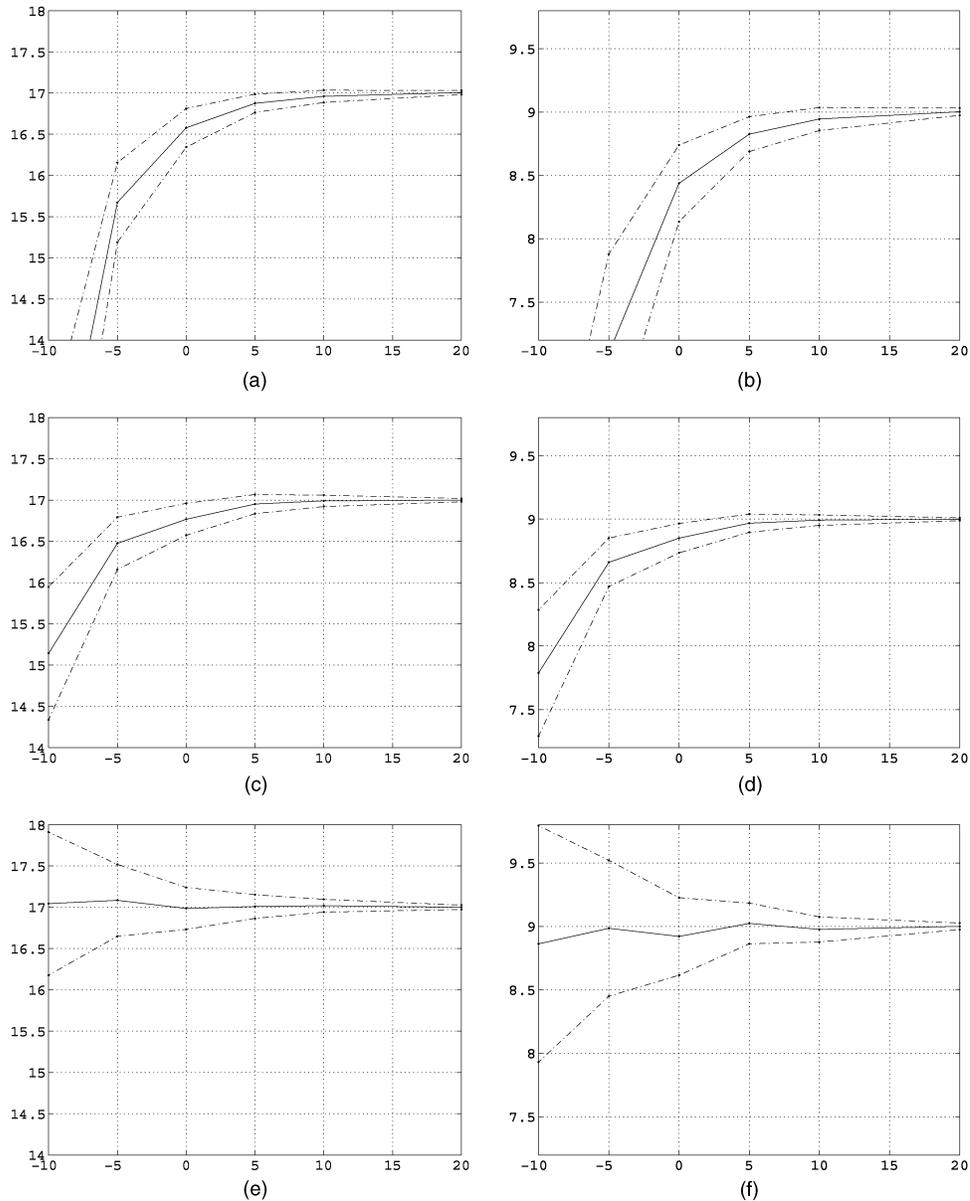


Fig. 3. Estimates of rotation angles for varying SNR level in differential motion fields, for smooth surface: (a), (c), and (e)  $\phi_R$  using the factorization, the epipolar and the proposed method, respectively, (b), (d), and (f)  $\phi_S$  using the factorization, the epipolar and the proposed method, respectively.

Figs. 3a, c, and e and Figs. 3b, d, and f were  $\phi_R = 17^\circ$  and  $\phi_S = 9^\circ$ , respectively.

As it can be seen, the proposed approach was superior to the both the factorization and the epipolar method. In particular, the other approaches failed for very low SNR levels in all experiments carried out; especially for the object consisting of planar surfaces. Although it can be pointed out that both the factorization and the epipolar method illustrated smaller standard deviation estimates, compared to the proposed method for higher SNR levels (more than 5dB in the case of the smooth surface).

## 5 CONCLUSIONS AND FURTHER RESEARCH

We presented a novel method for computing the rotation matrices that characterize the motion of a 3D rigid object on the basis of its orthographic 2D projections. The obtained

expressions are very simple to compute via eigenvalue decomposition of symmetric  $2 \times 2$  matrices that involve the corresponding motion vectors. When the available motion vectors are noise-free, computations yield the exact values of the rotation and 3D shape of the object, even when only four motion vectors per transition are available. A slight modification of the expressions of the aforementioned propositions may allow for the inclusion of an arbitrarily large number of motion vectors in the computations. On top of that, both the theoretic analysis and the simulations indicate that the quality of the estimates can be improved if a certain strategy is being followed in the selection of the motion vectors that are used in the involved matrices.

The results presented in this work provide the groundwork for further improvements in the following manner:

1. Improvement in the estimates of  $\mathbf{R}$  and  $\mathbf{S}$  by using appropriate weighting of the involved least-squares schemes. Preliminary results on this subject can be found in [22].
2. Improvement in the reconstructed 3D shape (wireframe) which tends to be rather messy when the disturbing noise is of low SNR. A post-smoothing procedure, possibly based on partial a priori knowledge of the object, may be useful in this direction.
3. Reestimation of the motion vectors by using the algorithm presented in this paper as a constraint imposed by the rigid nature of the considered 3D object.
4. Improvement in the shape (depth) estimates can also be achieved by means of time (i.e., frame)—recursions, as it has been reported in [9], [3].

## REFERENCES

- [1] K. Aizawa, H. Harashima, and T. Saito, "Model-Based Analysis-Synthesis Image Coding (MBASIC) System for a Person's Face," *Signal Processing: Image Comm.*, no. 1, pp. 139-152, Oct. 1989.
- [2] G. Bozdogi, A.M. Tekalp, and L. Onural, "An Improvement to MBASIC Algorithm for 3D Motion and Depth Estimation," *IEEE Trans. Image Processing*, vol. 3, pp. 711-716, June 1994.
- [3] A. Briassouli, Y. Xirouhakis, and A. Delopoulos, "Recursive 3D Reconstruction Under Orthography Using Kalman Filtering," *Proc. Int'l Workshop Synthetic-Natural Hybrid Coding and 3D Imaging*, Sept. 1999.
- [4] A. Delopoulos and Y. Xirouhakis, "Robust Estimation of Motion and Shape Based on Orthographic Projections of Rigid Objects," *Proc. IEEE Image and Multidimensional Digital Signal Processing Workshop*, pp. 151-154, July 1998.
- [5] D. DeMenthon and L.S. Davis, "Exact and Approximate Solutions of the Perspective-Three-Point Problem," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp. 1,100-1,104, Nov. 1992.
- [6] O. Faugeras, *Three-Dimensional Computer Vision*. MIT Press, 1993.
- [7] T.S. Huang and C.H. Lee, "Motion and Structure from Orthographic Projections," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 536-540, May 1989.
- [8] T.S. Huang and A.N. Netravali, "Motion and Structure from Feature Correspondences: A Review," *Proc. IEEE*, vol. 82, pp. 252-269, Feb. 1994.
- [9] Y.S. Hung and H.T. Ho, "A Kalman Filter Approach to Direct Depth Estimation Incorporating Surface Structure," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 570-575, June 1999.
- [10] H.C. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene from Two Projections," *Nature*, vol. 293, pp. 133-135, Sept. 1981.
- [11] T. Morita and T. Kanade, "A Sequential Factorization Method for Recovering Shape and Motion from Image Streams," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 8, pp. 858-867, Aug. 1997.
- [12] J. Ostuni and S. Dunn, "Motion from Three Weak Perspective Images Using Image Rotation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 64-69, Jan. 1996.
- [13] J. Philip, "Estimation of Three-Dimensional Motion of Rigid Objects from Noisy Observations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 61-66, Jan. 1991.
- [14] C.J. Poelman and T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 206-218, Mar. 1997.
- [15] L.S. Shapiro, A. Zisserman, and M. Brady, "3D Motion Recovery via Affine Epipolar Geometry," *Int'l J. Computer Vision*, vol. 16, pp. 147-182, 1995.
- [16] S. Soatto and P. Perona, "Reducing Structure From Motion: A General Framework for Dynamic Vision, Part 2: Implementation and Experimental Assessment," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 943-960, Sept. 1998.
- [17] M. Tekalp, *Digital Video Processing*. Prentice Hall, 1995.
- [18] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: a Factorization Method," *Int'l J. Computer Vision*, vol. 9, no. 2, pp. 137-154, 1992.
- [19] R.Y. Tsai and T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 13-27, Jan. 1984.
- [20] S. Ullman, *The Interpretation of Visual Motion*. Cambridge, Mass.: MIT Press, 1979.
- [21] J. Weng, T.S. Huang, and N. Ahuja, "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451-476, May 1989.
- [22] Y. Xirouhakis, G. Tschepnakis, and A. Delopoulos, "User Choices for Efficient 3D Motion and Shape Extraction from Orthographic Projections," *Proc. Sixth IEEE Int'l Conf. Electronics, Circuits, and Systems*, Sept. 1999.
- [23] G. Xu and N. Sugimoto, "A Linear Algorithm for Motion from Three Weak Perspective Images Using Euler Angles," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 54-57, Jan. 1999.