Euclidian Shape and Motion from Multiple Perspective Views by Affine Iterations

Stéphane Christy, Radu Horaud

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Euclidian Shape and Motion from Multiple Perspective Views by Affine Iterations

Stéphane Christy and Radu Horaud

N° 2421

Décembre 94

PROGRAMME 4

Robotique,

image

et vision

Rapport de recherche

1994

ISSN 0249-6399

Euclidian Shape and Motion from Multiple Perspective Views by Affine Iterations *

Stéphane Christy and Radu Horaud

Programme 4 — Robotique, image et vision
Projet MOVI (LIFIA–CNRS)

Rapport de recherche n° 2421 — Décembre 94 — 30 pages

Abstract: In this paper we describe a method for solving the Euclidean reconstruction problem with a perspective camera model by incrementally performing an Euclidean reconstruction with either a weak or a para perspective camera model. With respect to other methods that compute shape and motion from a sequence of images with a calibrated perspective camera, this method converges in a few iterations, is computationnaly efficient, and does not suffer from the non linear nature of the problem. With respect to factorization and/or affine-invariant methods, this method solves for the sign (reversal) ambiguity in a very simple way and provides much more accurate reconstructions results. We give a detailed account of the method, analyse its convergence based on numerical and experimental considerations, and test its efficiency both with synthetic and real data.

Key-words: perspective, weak perspective, and para perspective camera models, 3-D Euclidean and affine reconstruction, shape and motion from multiple views.

(Résumé : tsvp)

*This work has been supported by "Société Aérospatiale".

Unité de recherche INRIA Rhône-Alpes
46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1 (France)
Téléphone : (33) 76 57 47 77 – Télécopie : (33) 76 57 47 54

Détection du mouvement et de la structure d'une scène à partir d'une séquence d'images perspectives par itérations affines

Résumé : Dans cet article nous décrivons une méthode de résolution du problème de reconstruction euclidienne avec un modèle perspectif de caméra en effectuant une reconstruction euclidienne incrémentale avec un modèle de caméra perspective faible ou para perspective. Par rapport à d'autres méthodes qui calculent la structure et le mouvement à partir d'une séquence d'images, cette méthode converge en quelques itérations, est efficace d'un point de vue calculatoire, et ne souffre pas de la nature non linéaire du problème traité. Par rapport à des methodes telles que la factorisation ou les invariants affines, notre méthode résout le problème de l'ambiguité de signe d'une façon très simple et fournit des résultats plus précis. Nous décrivons la nouvelle méthode en détail, analysons sa convergence sur la base de considérations numériques et expérimentales et nous testons son efficacité avec des données synthétiques et réelles.

Mots-clé : perspective, perspective faible, para perspective, reconstruction 3D affine et euclidienne, mouvement et structure, séquence d'images.

1 Introduction and background

The problem of computing 3-D shape and motion from a long sequence of images has received a lot of attention for the last few years. Previous approaches attempting to solve this problem fall into several categories, whether the camera is calibrated or not, and/or whether a projective or an affine model is being used. With a calibrated camera one may compute Euclidian shape up to a scale factor using either a perspective model [4], [14], [3], [13], or a linear model [5], [15], [16], [19], [10], [12], [11]. With an uncalibrated camera the recovered shape is defined up to a projective transformation [7], [2], [8], [18], or up to an affine transformation [7], [8]. One can therefore address the problem of either Euclidian, affine, or projective shape reconstruction. In this paper we are interested in Euclidian shape reconstruction with a calibrated camera. In that case, one may use either a perspective camera model or an affine approximation – orthographic projection, weak perspective, or para perspective.

The perspective model has associated with it, in general, non linear reconstruction techniques. This naturally leads to non-linear minimization methods which require some form of initialization [4], [14], [3], [13], [2], [8], [18]. If the initial "guess" is too faraway from the true solution then the minimization process is either very slow or it converges to a wrong solution. Affine models lead, in general, to linear resolution methods [5], [16], [19], [10], [11], but the solution is defined only up to a sign (reversal) ambiguity. Moreover, both these two solutions are just an approximations of the true solution. For example, Figure 1 shows four Euclidian shapes. The first one (top-left) is the theoretical shape. The second one (top-right) is the reconstructed shape using a perspective camera model and the method outlined in this paper. The third and fourth ones (bottom-left and bottom-right) are the reconstructed shapes using a weak perspective camera model.

One way to combine the perspective and affine models could be to use the linear (affine) solution in order to initialize the non-linear minimization process associated with perspective. However, there are several drawbacks with such an approach. First, such a resolution technique does not take into account the simple link that exists between the perspective model and its linear approximations. Second, there is no mathematical evidence that a non-linear least-squares minimization method is "well" initialized by a solution that is ob-

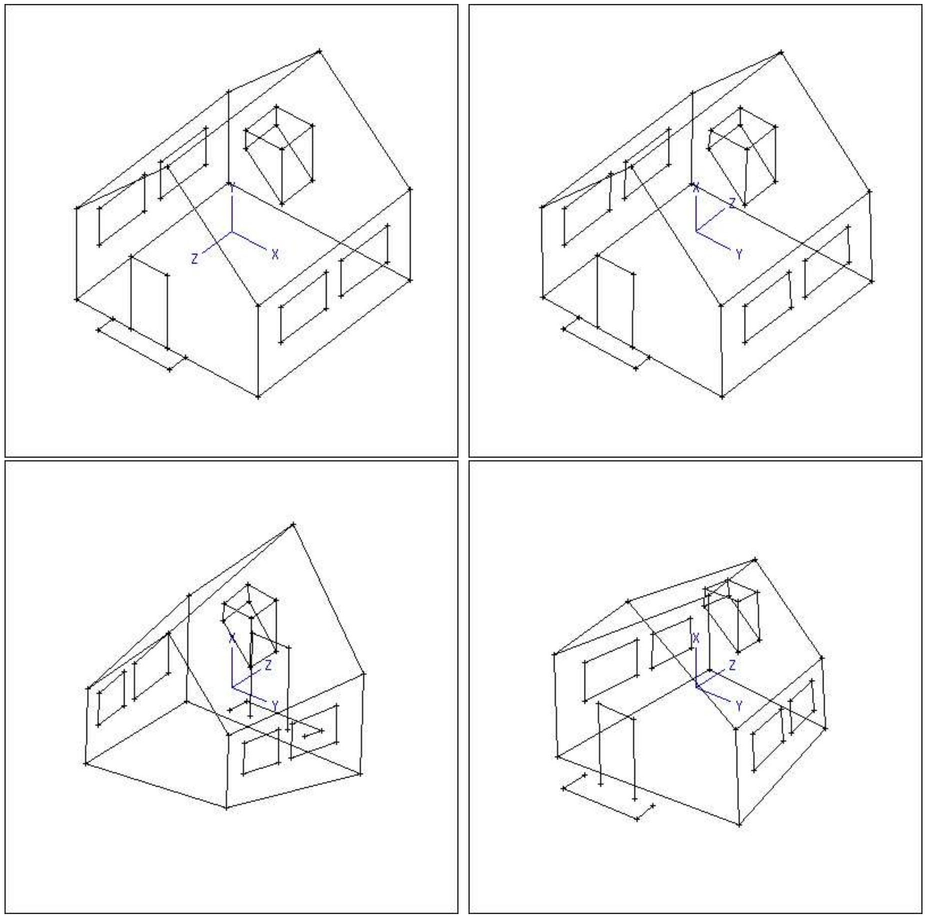

Figure 1: This figure shows a theoretical 3-D shape (top-left) and three reconstructions of this shape from 10 views. The first reconstruction (top-right) was obtained using a perspective camera model and the method described in this paper. The second reconstruction (bottom-left) and its reversal (bottom-right) were obtained using a weak perspective camera model and the factorization method of Tomasi & Kanade.

tained linearly. Third, there are two solutions associated with the affine model and it is not clear which one to choose.

The perspective projection can be modelled by a projective transformation from the 3-D projective space to the 2-D projective plane. Weak perspective and para perspective are the most common affine approximations of perspective. Weak perspective may well be viewed as a zero-order approximation: $1/(1+\epsilon) \approx 1$. Para perspective [1] is a first order approximation of full perspective: $1/(1+\epsilon) \approx 1-\epsilon$. Recently, in [6] a method has been proposed for determining the pose of a 3-D shape with respect to a single view by iteratively improving the pose computed with a weak perspective camera model to converge, at the limit, to a pose estimation using a perspective camera model. At our knowledge, the method cited above, i.e., [6] is among one of the first computational paradigms that link linear techniques (associated with affine camera models) with a perspective model. In [9] an extension of this paradigm to para perspective is proposed. The authors show that the *iterative para perspective* pose algorithm has better convergence properties than the *iterative weak perspective* one.

In this paper we describe a new Euclidian reconstruction method that makes use of affine reconstruction in an iterative manner such that this iterative process converges, at the limit, to a set of 3-D Euclidian shape and motion parameters that are consistent with a perspective model. The novelty of the method that we propose is twofold: (i) it extends the iterative pose determination algorithms described in [6] and in [9] to deal with the problem of shape and motion from multiple views and (ii) it is a generalization to perspective of the factorization methods [16], [11] and of the affine-invariant methods [19]. More precisely, the *affine-iterative reconstruction* method that we propose here has a number of interesting features:

- It solves the sign (or reversal) ambiguity that is inherent with affine reconstruction;
- It is fast because it converges in a few iterations (3 to 5 iterations), each iteration involving simple linear algebra computations;
- We show that the quality of the Euclidian reconstruction obtained with our method is only weakly influenced by camera calibration errors. The only intrinsic camera parameter that has a crucial effect on the quality of the reconstruction is the ratio between the horizontal pixel size and vertical pixel size – ratio which is known to be very stable [17];
- It allows the use of either weak or para perspective camera model approximations which are used iteratively, and
- It can be combined with almost any affine shape and motion algorithm. In particular we show how our method can be combined either with factorization methods [16], [11] or with affine-invariant methods [19], [10].

1.1 Paper organization

The remainder of this paper is organized as follows. Section 2 describes the relationship between full, para, and weak perspective and establishes the relationships between the perspective projection of a 3-D point and its weak and para perspective projections. Section 3 describes how to perform reconstruction with a perspective camera model by iterating either a weak perspective reconstruction or a para perspective reconstruction algorithm. Section 4 outlines two affine reconstruction algorithms and the conversion of affine reconstruction into Euclidian reconstruction using either weak or para perspective. Section 5 describes how to solve for the reversal ambiguity associated, in general, with an affine camera. Section 6 explains why the method described in this paper is not sensitive to camera calibration errors and Section 7 provides an analysis of the convergence of the perspective reconstruction algorithm. Finally, Section 8 gives some results obtained both with simulated and real data, and Section 9 provides a short discussion and a proposal for future work along the lines developped in this paper.

2 Camera models

Let us consider a pin hole camera model. We denote by P_i a 3-D point with Euclidian coordinates X_i, Y_i, and Z_i in a frame that is attached to the object – the object frame. The origin of this frame may well be the object point P_0. An object point P_i projects onto the image in p_i with image coordinates u_i and v_i and we have ($\mathbf{P}_i$ is the vector $\overrightarrow{P_0 P_i}$ from point P_0 to point P_i):

$$\begin{pmatrix} su_i \\ sv_i \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_c \\ 0 & \alpha_v & v_c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{i}^T & t_x \\ \mathbf{j}^T & t_y \\ \mathbf{k}^T & t_z \\ 0\,0\,0 & 1 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

The first matrix describes the affine transformation between the camera coordinates and the images coordinates. The second matrix describes the projective transformation between the 3-D projective space and the 2-D projective image plane. The third matrix describes the rigid transformation (rotation and translation) between the object frame and the camera frame.

From now on we will be assuming that the intrinsic camera parameters are known and therefore we consider the relationship between camera coordinates and image coordinates:

$$u_i = \alpha_u x_i + u_c \tag{1}$$

$$v_i = \alpha_v y_i + v_c \tag{2}$$

In these equations α_u and α_v are the vertical and horizontal scale factors and u_c and v_c are the image coordinates of the intersection of the optical axis with the image plane. It will be shown that the reconstruction method described here depends only on the ratio α_u/α_v and that the reconstruction obtained with our method is not sensitive to errors in u_c and v_c. The relationship between object points and camera points can be written as:

$$x_i = \frac{\mathbf{i} \cdot \mathbf{P}_i + t_x}{\mathbf{k} \cdot \mathbf{P}_i + t_z} \tag{3}$$

$$y_i = \frac{\mathbf{j} \cdot \mathbf{P}_i + t_y}{\mathbf{k} \cdot \mathbf{P}_i + t_z} \tag{4}$$

We divide both the numerator and the denominator of eqs. (3) and (4) by t_z. We introduce the following notations:

- $\mathbf{I} = \mathbf{i}/t_z$ is the first row of the rotation matrix scaled by the z-component of the translation vector;
- $\mathbf{J} = \mathbf{j}/t_z$ is the second row of the rotation matrix scaled by the z-component of the translation vector;
- $x_0 = t_x/t_z$ and $y_0 = t_y/t_z$ are the camera coordinates of p_0 which is the projection of P_0 – the origin of the object frame, and
- We denote by ϵ_i the following ratio:

$$\epsilon_i = \frac{\mathbf{k} \cdot \mathbf{P}_i}{t_z} \tag{5}$$

We may now rewrite the perspective equations as:

$$x_i = \frac{\mathbf{I} \cdot \mathbf{P}_i + x_0}{1 + \epsilon_i} \tag{6}$$

$$y_i = \frac{\mathbf{J} \cdot \mathbf{P}_i + y_0}{1 + \epsilon_i} \tag{7}$$

Whenever the object is at some distance from the camera, the ϵ_i are small compared to 1. We may therefore introduce two approximations of the perspective equations: weak and para perspective.

2.1 Weak perspective

Weak perspective assumes that the object points lie in a plane parallel to the image plane passing through the origin of the object frame, i.e., P_0. This is equivalent to a zero-order approximation:

$$\frac{1}{1 + \epsilon_i} \approx 1 \;\; \forall i, \; i \in \{1 ... n\}$$

With this approximation, eqs. (6) and (7) become:

$$x_i^w - x_0 = \mathbf{I} \cdot \mathbf{P}_i \tag{8}$$

$$y_i^w - y_0 = \mathbf{J} \cdot \mathbf{P}_i \tag{9}$$

In these two equations x_i^w and y_i^w are the camera coordinates of the weak perspective projection of the point P_i. By identification with eqs. (6) and (7) we obtain the relationship between the weak perspective and the perspective projections of P_i:

$$x_i^w = x_i(1 + \epsilon_i) \tag{10}$$

$$y_i^w = y_i(1 + \epsilon_i) \tag{11}$$

These equations allow us to determine the quality of the weak perspective approximation with respect to the perspective projection. Indeed the error between the weak perspective projection and the "true" projection is:

$$\Delta x^w = |x_i^w - x_i| = |x_i \epsilon_i| \tag{12}$$
$$\Delta y^w = |y_i^w - y_i| = |y_i \epsilon_i| \tag{13}$$

Hence the quality of the approximation depends both on the value of ϵ_i AND on the position of the point in the image. Let's consider for example a 512×512 image. Approximate values for the intrinsic parameters are: $\alpha_u = \alpha_v = 1000$ and $u_c = v_c = 256$. Therefore using eq. (1) and (2) we have:

$$0 \leq x_i \, , \; y_i \leq 0.25$$

We conclude that for objects that are quite closed to the camera, the weak perspective approximation is still valid provided that the object lies in the neighbourhood of the optical axis.

2.2 Para perspective

Para perspective may be viewed as a first-order approximation of perspective. Indeed, with the approximation:

$$\frac{1}{1+\epsilon_i} \approx 1 - \epsilon_i \;\; \forall i, \; i \in \{1...n\}$$

we obtain the para perspective projection of P_i:

$$\begin{aligned} x_i^p &= (\mathbf{I} \cdot \mathbf{P}_i + x_0)(1 - \epsilon_i) \\ &\approx \mathbf{I} \cdot \mathbf{P}_i + x_0 - x_0 \epsilon_i \\ &= \frac{\mathbf{i} \cdot \mathbf{P}_i}{t_z} + x_0 - x_0 \frac{\mathbf{k} \cdot \mathbf{P}_i}{t_z} \end{aligned}$$

where the term in $1/t_z^2$ was neglected. There is a similar expression for y_i^p.

Finally, the para perspective equations are:

$$x_i^p - x_0 = \frac{\mathbf{i} - x_0 \, \mathbf{k}}{t_z} \cdot \mathbf{P}_i \tag{14}$$
$$y_i^p - y_0 = \frac{\mathbf{j} - y_0 \, \mathbf{k}}{t_z} \cdot \mathbf{P}_i \tag{15}$$

In order to obtain the relationship between the para perspective and the perspective projections of P_i we can write these equations as follows:

$$\begin{aligned} x_i^p &= x_0 + \mathbf{I} \cdot \mathbf{P}_i - x_0 \epsilon_i \\ y_i^p &= y_0 + \mathbf{J} \cdot \mathbf{P}_i - y_0 \epsilon_i \end{aligned}$$

Again, by identification with eqs. (6) and (7) we obtain the relationship between the para perspective and the perspective projections of P_i:

$$x_i^p = x_i(1+\epsilon_i) - x_0\epsilon_i \tag{16}$$
$$y_i^p = y_i(1+\epsilon_i) - y_0\epsilon_i \tag{17}$$

As in the previous section, we can easily estimate the error between the para perspective and perspective projections:

$$\Delta x^p = |x_i^p - x_i| = |(x_i - x_0)\epsilon_i| \tag{18}$$
$$\Delta y^p = |y_i^p - y_i| = |(y_i - y_0)\epsilon_i| \tag{19}$$

Whenever an object point P_i is far from the optical axis the weak perspective model is a poor approximation. However, a proper choice of the origin, i.e, P_0, and the use of the para perspective model can compensate and provide a good approximation even if ϵ_i is not small.

3 Reconstruction with a perspective camera

Let us consider again the perspective equations (6) and (7). These equations may be also written as:

$$x_i(1+\epsilon_i) - x_0 = \mathbf{I}\cdot\mathbf{P}_i \tag{20}$$
$$y_i(1+\epsilon_i) - y_0 = \mathbf{J}\cdot\mathbf{P}_i \tag{21}$$

Let us subtract the *para perspective term* from both the left and right sides of equations (20) and (21). We obtain:

$$x_i(1+\epsilon_i) - x_0 - x_0\underbrace{\frac{1}{t_z}\mathbf{k}\cdot\mathbf{P}_i}_{\epsilon_i} = \frac{1}{t_z}\mathbf{i}\cdot\mathbf{P}_i - x_0\frac{1}{t_z}\mathbf{k}\cdot\mathbf{P}_i$$

$$y_i(1+\epsilon_i) - y_0 - y_0\underbrace{\frac{1}{t_z}\mathbf{k}\cdot\mathbf{P}_i}_{\epsilon_i} = \frac{1}{t_z}\mathbf{i}\cdot\mathbf{P}_i - y_0\frac{1}{t_z}\mathbf{k}\cdot\mathbf{P}_i$$

These equations can be written more compactly as:

$$(x_i - x_0)(1+\epsilon_i) = \mathbf{I}_p\cdot\mathbf{P}_i \tag{22}$$
$$(y_i - y_0)(1+\epsilon_i) = \mathbf{J}_p\cdot\mathbf{P}_i \tag{23}$$

with:

$$\mathbf{I}_p = \frac{\mathbf{i} - x_0\,\mathbf{k}}{t_z} \tag{24}$$
$$\mathbf{J}_p = \frac{\mathbf{j} - y_0\,\mathbf{k}}{t_z} \tag{25}$$

To summarize, we have two different sets of equations that describe the *same* perspective camera model. The first set, i.e., equations (20) and (21) establish the link between perspective and weak perspective while the second set, i.e., equations (22) and (23) establish the link between perspective and para perspective. If the values of ϵ_i are set to zero than we obtain equations that approximate the camera either with weak perspective or with para perspective. The crucial point is that if the ϵ_i's are set to some fixed non zero values, then the equations remain linear. In particular, there are values for the ϵ_i's for which these equations are consistent – up to some measurement noise – with the full perspective model. The key idea of our reconstruction method is to iteratively estimate values for the ϵ_i's such that the perspective reconstruction problem becomes an "affine iterative" reconstruction problem.

Let us consider now k views of the same scene points. We assume that image-to-image correspondences have already been established. Both equations (20) and (21) or equations (22) and (23) can be written as:

$$\underbrace{\mathbf{s}_{ij}}_{2\times 1} = \underbrace{A_j}_{2\times 3}\underbrace{\mathbf{P}_i}_{3\times 1} \tag{26}$$

In this formula the subscript i stands for the i^{th} point and the subscript j for the j^{th} image. The 2-vector $\mathbf{s}_{ij}$ is equal to (weak perspective):

$$\mathbf{s}_{ij} = \begin{pmatrix} x_{ij}(1+\epsilon_{ij}) - x_{0j} \\ y_{ij}(1+\epsilon_{ij}) - y_{0j} \end{pmatrix} \tag{27}$$

or to (para perspective):

$$\mathbf{s}_{ij} = \begin{pmatrix} (x_{ij} - x_{0j})(1+\epsilon_{ij}) \\ (y_{ij} - y_{0j})(1+\epsilon_{ij}) \end{pmatrix} \tag{28}$$

In these equations ϵ_{ij}, i.e., eq. (5) is defined for each point and for each image:

$$\epsilon_{ij} = \frac{\mathbf{k}_j \cdot \mathbf{P}_i}{t_{z_j}} \tag{29}$$

The reconstruction problem is now the problem of simultaneously solving $2 \times n \times k$ equations of the form of eq. (26). We introduce now a method that solves these equations by linear iterations. More precisely, this method can be summarized by the following algorithm:

1. $\forall i$, $i \in \{1...n\}$ and $\forall j$, $j \in \{1...k\}$ set: $\epsilon_{ij} = 0$;
2. Update the values of $\mathbf{s}_{ij}$ using ϵ_{ij};
3. Perform an Euclidian reconstruction with an affine camera;
4. $\forall i$, $i \in \{1...n\}$ and $\forall j$, $j \in \{1...k\}$ estimate new values for ϵ_{ij};
5. Check the values of ϵ_{ij}:

if $\forall(i,j)$ the values of ϵ_{ij} just estimated at this iteration are identical with the values estimated at the previous iteration,
then stop;

else go to step 2.

The most important step of this algorithm is step 4: estimate new values for ϵ_{ij}. This computation can be made explicit if one considers into some more detail step 3 of the algorithm which can be farther decomposed into: (i) Affine reconstruction and (ii) Euclidian reconstruction.

The problem of affine reconstruction is the problem of determining both A_j and $\mathbf{P}_i$, for all j and for all i, in eq. (26), when some estimates of $\mathbf{s}_{ij}$ are provided. Such a reconstruction determines shape and motion up to a 3-D affine transformation. Indeed, for any 3×3 invertible matrix T we have:

$$A_j\mathbf{P}_i = A_jTT^{-1}\mathbf{P}_i$$

In order to convert affine shape and motion into Euclidian shape and motion, one needs to consider some Euclidian constraints associated either with the motion of the camera or with the shape being viewed by the camera. Since we deal here with a calibrated camera, we may well use rigid motion constraints in conjunction with weak or para perspective [16], [11]. Therefore, step 3 of the algorithm provides both Euclidian shape ($\mathbf{P}_1 ... \mathbf{P}_n$) and Euclidian motion, i.e., k matrices of the form:

$$\begin{pmatrix} \mathbf{i}_j^T & t_{x_j} \\ \mathbf{j}_j^T & t_{y_j} \\ \mathbf{k}_j^T & t_{z_j} \\ 0\ 0\ 0 & 1 \end{pmatrix}$$

Based on the parameters of the Euclidian shape and motion thus computed one can estimate ϵ_{ij} for all i and for all j using eq. (29) – step 4.

The first iteration of the algorithm performs a 3-D reconstruction using the initial image measurements and a weak or para perspective camera model. This first reconstruction allows an estimation of values for the ϵ_{ij}'s which in turn allow the image vectors $\mathbf{s}_{ij}$ to be *modified* (step 2 of the algorithm). The $\mathbf{s}_{ij}$'s are modified according to eq. (27) (for weak perspective) or according to eq. (28) (for para perspective) such that they better fit the approximated camera model being used.

The next iterations of the algorithm perform a 3-D reconstruction using (i) image vectors that are incrementally modified and (ii) a weak (or para) perspective camera model.

At convergence, the equations (26) are equivalent with the perspective equations (20), (21) or (22), (23). In other terms, this algorithm solves for Euclidian reconstruction with a perspective camera by iterations of an Euclidian reconstruction method with an affine camera. Therefore, before we