Uncalibrated Euclidean Reconstruction: A Review

Andrea Fusiello

Dipartimento Scientifico e Tecnologico, Università di Verona
Ca' Vignal 2, Strada Le Grazie, I-37134 Verona, IT
Tel: +39 045 809 8088 Fax: +39 045 809 8928
E-mail: fusiello@sci.univr.it

Abstract

This paper provides a review on techniques for computing a three-dimensional model of a scene from a single moving camera, with unconstrained motion and unknown parameters. In the classical approach, called *autocalibration* or *self-calibration*, camera motion and parameters are recovered first, using rigidity; then structure is easily computed. Recently, new methods based on the idea of *stratification* have been proposed. They upgrade the *projective* structure, achievable from correspondences only, to the *Euclidean* structure, by exploiting all the available constraints.

Key words: 3D vision, Autocalibration, Euclidean reconstruction, Self-calibration, Uncalibrated cameras.

1 Introduction

The goal of Computer Vision (see [1] for an introduction) is to compute properties (mainly geometric) of the three-dimensional world from images. One of the challenging problems of Computer Vision is to *reconstruct* a three-dimensional model of the scene from a moving camera. Possible applications include: navigation of autonomous vehicles, object recognition, reverse engineering and synthesis of virtual environments.

Most of the earlier studies in the field assume that the intrinsic parameters of the camera (focal length, image center and aspect ratio) are known. Computing camera motion in this case is a well known problem in photogrammetry, called *relative orientation* [2, 3], for which several methods are available (see [4] for a review). Given all the parameters of the camera, reconstruction is straightforward.

However, there are situations wherein the intrinsic parameters are unknown or off-line calibration is impracticable. In these cases the only information one can exploit is contained in the video sequence itself.

Yet, some assumptions are necessary to make the problem tractable. We will focus on the classical case of a single camera with constant but unknown intrinsic parameters and unknown motion. Other approaches restrict the motion [5, 6, 7, 8] or assume a rigidly moving stereo rig [9].

The contribution of this paper is to give a critical, unified view of some of the most promising techniques. Such a comparative account sheds light on the relations between different methods, presented in different ways and formalisms in the original research articles.

In the next section some necessary notation and concepts will be introduced. Then (Sec. 3) the reconstruction problem will be formulated. In Sec. 4 the classical autocalibration approach will be briefly outlined. Stratification methods will be described in some details in Sec. 5. Applicability of the methods will be discussed in Sec. 6. Finally (Sec. 7), conclusions will be drawn.

2 Notation and basics

This section introduces the mathematical background on perspective projections necessary for our purposes. Our notation follows [10].

Figure 1 near here

A pinhole camera is modeled by its *optical center* C and its *retinal plane* (or *image plane*) $\mathcal{R}$. A 3-D point W is projected into an image point m given by the intersection of $\mathcal{R}$ with the line containing C and W.

Let $\mathbf{w} = (x, y, z)$ be the coordinates of W in the world reference frame (fixed arbitrarily) and $\mathbf{m}$ the pixel coordinates of m. In homogeneous (or projective) coordinates

$$\tilde{\mathbf{m}} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \qquad \tilde{\mathbf{w}} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

the transformation from $\tilde{\mathbf{w}}$ to $\tilde{\mathbf{m}}$ is given by the matrix $\tilde{\mathbf{P}}$:

$$\kappa \tilde{\mathbf{m}} = \tilde{\mathbf{P}} \tilde{\mathbf{w}}, \tag{2}$$

where κ is a scale factor called *projective depth.* If $\tilde{\mathbf{P}}$ is suitably normalized, κ becomes the true orthogonal distance of the point from the focal plane of the camera.

The camera is therefore modeled by its *perspective projection matrix* (henceforth simply *camera matrix*) $\tilde{\mathbf{P}}$, which can be decomposed, using the QR factorization, into the product

$$\tilde{\mathbf{P}} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]. \tag{3}$$

The matrix $\mathbf{A}$ depends on the *intrinsic parameters* only, and has the following form:

$$\mathbf{A} = \begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where $\alpha_u = -f k_u$, $\alpha_v = -f k_v$ are the focal lengths in horizontal and vertical pixels, respectively (f is the focal length in millimeters, k_u and k_v are the effective number of pixels per millimeter along the u and v axes), (u_0, v_0) are the coordinates of the *principal point*, given by the intersection of the optical axis with the retinal plane (Fig. 1), and γ is the *skew* factor.

The camera position and orientation (*extrinsic parameters*), are encoded by the 3×3 rotation matrix $\mathbf{R}$ and the translation $\mathbf{t}$, representing the rigid transformation that aligns the camera reference frame (Fig. 1) and the world reference frame.

2.1 Epipolar geometry

Let us consider the case of two cameras (see Fig. 2).

Figure 2 near here

If we take the first camera reference frame as the world reference frame, we can write the two following general camera matrices:

$$\tilde{\mathbf{P}} = \mathbf{A}[\mathbf{I}|\mathbf{0}] = [\mathbf{A}|\mathbf{0}] \tag{5}$$

$$\tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R}|\mathbf{t}]. \tag{6}$$

A three-dimensional point $\mathbf{w}$ is projected onto both image planes, to points $\tilde{\mathbf{m}} = \tilde{\mathbf{P}}\tilde{\mathbf{w}}$ and $\tilde{\mathbf{m}}' = \tilde{\mathbf{P}}'\tilde{\mathbf{w}}$, which constitute a *conjugate pair*. From the left camera we obtain:

$$\kappa'\tilde{\mathbf{m}}' = \mathbf{A}'[\mathbf{R}|\mathbf{t}]\tilde{\mathbf{w}} = \mathbf{A}'[\mathbf{R}|\mathbf{t}]\left(\begin{bmatrix} x \\ y \\ z \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right) = \mathbf{A}'\mathbf{R}\begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{A}'\mathbf{t}. \tag{7}$$

From the right camera we obtain: $\kappa\mathbf{A}^{-1}\tilde{\mathbf{m}} = [\mathbf{I}|\mathbf{0}]\ \tilde{\mathbf{w}} = [x\ y\ z]^\top$. Substituting the latter in (7) yields:

$$\kappa'\tilde{\mathbf{m}}' = \kappa\mathbf{A}'\mathbf{R}\mathbf{A}^{-1}\tilde{\mathbf{m}} + \mathbf{A}'\mathbf{t} = \kappa\mathbf{H}_\infty\tilde{\mathbf{m}} + \tilde{\mathbf{e}}' \tag{8}$$

where $\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}$ and $\tilde{\mathbf{e}}' = \mathbf{A}'\mathbf{t}$ (the reason for this notation will be manifest in the following).

Equation (8) means that $\tilde{\mathbf{m}}'$ lies on the line going trough $\tilde{\mathbf{e}}'$ and the point $\mathbf{H}_\infty\tilde{\mathbf{m}}$. In projective coordinates the collinearity of these three points can be expressed with the external product: $\tilde{\mathbf{m}}'^\top(\tilde{\mathbf{e}}' \wedge \mathbf{H}_\infty\tilde{\mathbf{m}}) = 0$, or

$$\tilde{\mathbf{m}}'^\top\mathbf{F}\tilde{\mathbf{m}} = 0, \tag{9}$$

where $\mathbf{F} = [\tilde{\mathbf{e}}']_\wedge\mathbf{H}_\infty$ is the *fundamental matrix*, relating conjugate points, and $[\tilde{\mathbf{e}}']_\wedge$ is a matrix such that $\tilde{\mathbf{e}}' \wedge \mathbf{x} = [\tilde{\mathbf{e}}']_\wedge \mathbf{x}$. From (9) we can see that $\tilde{\mathbf{m}}'$ belongs to the line $\mathbf{F}\tilde{\mathbf{m}}$ in the second image, which is called the *epipolar line* of $\tilde{\mathbf{m}}$. It's easy to see that $\tilde{\mathbf{e}}'^\top \mathbf{F} = \mathbf{0}$, meaning that all the epipolar lines contain the point $\tilde{\mathbf{e}}'$, which is called the *epipole* (Fig. 2). Since $\mathbf{F}\tilde{\mathbf{e}} = \mathbf{F}^\top \tilde{\mathbf{e}}' = \mathbf{0}$ the rank of $\mathbf{F}$ is in general two and, being defined up to a scale factor, it depends upon seven parameters. In the most general case, the only geometrical information that can be computed from pairs of images is the fundamental matrix. Its computation requires a minimum of eight point correspondences to obtain a unique solution [11, 12].

It can be seen that (9) is equivalent to

$$(\mathbf{A}'^{-1}\tilde{\mathbf{m}}')^\top [\mathbf{t}]_\wedge \mathbf{R}(\mathbf{A}^{-1}\tilde{\mathbf{m}}) = 0. \tag{10}$$

Changing to normalized coordinates, $\tilde{\mathbf{n}} = \mathbf{A}^{-1}\tilde{\mathbf{m}}$, one obtain the original formulation of the Longuet-Higgins [13] equation,

$$\tilde{\mathbf{n}}'^\top \mathbf{E}\tilde{\mathbf{n}} = 0 \tag{11}$$

involving the *essential matrix*

$$\mathbf{E} = [\mathbf{t}]_\wedge \mathbf{R}, \tag{12}$$

which can be obtained when intrinsic parameters are known. $\mathbf{E}$ depends upon five independent parameters (rotation and translation up to a scale factor). From (10) it is easy to see that

$$\mathbf{F} = \mathbf{A}'^{-\top}\mathbf{E}\mathbf{A}^{-1}. \tag{13}$$

2.2 Homography of a plane

Given two views of a scene, there is a linear projective transformation (an *homography*) relating the projection $\mathbf{m}$ of the point of a plane Π in the first view to its projection in the second view, $\mathbf{m}'$. This application is given by a 3×3 invertible matrix $\mathbf{H}_\Pi$ such that:

$$\tilde{\mathbf{m}}' = \mathbf{H}_\Pi \tilde{\mathbf{m}}. \tag{14}$$

It can be seen that, given the two projection matrices,

$$\tilde{\mathbf{P}} = \mathbf{A}[\mathbf{I} \mid \mathbf{0}], \qquad \tilde{\mathbf{P}}' = \mathbf{A}'[\mathbf{R} \mid \mathbf{t}] \tag{15}$$

(the world reference frame is fixed on the first camera) and a plane Π of equation $\mathbf{n}^\top \mathbf{x} = d$, the following holds [14]:

$$\mathbf{H}_\Pi = \mathbf{A}'(\mathbf{R} + \mathbf{t}\frac{\mathbf{n}^\top}{d})\mathbf{A}^{-1}. \tag{16}$$

$\mathbf{H}_\Pi$ is the homography matrix for the plane Π. If $d \to \infty$,

$$\mathbf{H}_\infty = \mathbf{A}'\mathbf{R}\mathbf{A}^{-1}. \tag{17}$$

This is the homography matrix for the *infinity plane*, which maps vanishing points to vanishing points and depends only on the rotational component of the rigid displacement. It can be easily seen that:

$$\mathbf{H}_\Pi = \mathbf{H}_\infty + \tilde{\mathbf{e}}'\frac{\mathbf{n}^\top}{d}\mathbf{A}^{-1} \tag{18}$$

where $\tilde{\mathbf{e}}' = \mathbf{A}'\mathbf{t}$.

3 The reconstruction problem

Consider a set of three-dimensional points viewed by N cameras with matrices $\{\tilde{\mathbf{P}}^i\}_{i=1\ldots N}$. Let $\tilde{\mathbf{m}}^i_j \simeq \tilde{\mathbf{P}}^i\tilde{\mathbf{w}}_j$ be the (homogeneous) coordinates of the projection of the j-th point onto the i-th camera. The *reconstruction problem* can be cast in the following way: given the set of pixel coordinates $\{\tilde{\mathbf{m}}^i_j\}$, find the set of camera matrices $\{\tilde{\mathbf{P}}^i\}$ and the scene structure $\{\tilde{\mathbf{w}}_j\}$ such that

$$\tilde{\mathbf{m}}^i_j \simeq \tilde{\mathbf{P}}^i\tilde{\mathbf{w}}_j. \tag{19}$$

Without further restrictions we will, in general, obtain a projective reconstruction [15, 16, 17] defined up to an arbitrary projective transformation. Indeed, if $\{\tilde{\mathbf{P}}^i\}$ and $\{\tilde{\mathbf{w}}_j\}$ satisfy (19), also $\{\tilde{\mathbf{P}}^i\tilde{\mathbf{T}}\}$ and $\{\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{w}}_j\}$ satisfy (19) for any 4×4 nonsingular matrix $\tilde{\mathbf{T}}$.

A projective reconstruction can be computed starting from points correspondences only, without any a-priori knowledge [18, 19, 20, 21, 22, 23, 24, 25]. Despite it conveys some useful in formations [26, 27], we would like to obtain a *Euclidean reconstruction*, a very special one that differs from the true reconstruction by an unknown similarity transformation. This is composed by a rigid displacement (due to the arbitrary choice of the world reference frame) plus a a uniform change of scale (due to the well-known depth-speed ambiguity: it is impossible to determine whether a given image motion is caused by a nearby object with slow relative motion or a distant object with fast relative motion).

Maybank and Faugeras [28, 29] proved that, if intrinsic parameters are constant, Euclidean reconstruction is achievable. The procedure is known as *autocalibration*.

In this approach, the internal unchanging parameters of the camera are computed from at least three views. Once the intrinsic parameters are known, the problem of computing the extrinsic parameters (motion) from point correspondences is the well-known relative orientation problem, for which a variety of methods have been developed [4, 30, 31]. In principle, from the set of correspondences $\{\tilde{\mathbf{m}}_i\}$ one can compute the fundamental matrix, from which the essential matrix is immediately obtained with (13). Motion parameters $\mathbf{R}$ and the direction of translation $\mathbf{t}$ are obtained directly from the factorization (12) of $\mathbf{E}$. In [32] direct and iterative methods are compared.

Recently, new approaches based on the idea of *stratification* [14, 33] have been introduced. Starting from a projective reconstruction, which can be computed from the set of correspondences $\{\tilde{\mathbf{m}}_j^i\}$ only, the problem is computing the *proper* $\tilde{\mathbf{T}}$ that upgrades it to a Euclidean reconstruction, by exploiting all the available constraints. To this purpose the problem is stratified into different representations: depending on the amount of information and the constraints available, it can be analyzed at a projective, affine[1], or Euclidean level.

4 Autocalibration

In the case of two different cameras, the fact that for any fundamental matrix $\mathbf{F}$ there exist two intrinsic parameters matrix $\mathbf{A}$ and $\mathbf{A}'$ and a rigid motion represented by $\mathbf{t}$ and $\mathbf{R}$ such that $\mathbf{F} = \mathbf{A}'^{-\top}([\mathbf{t}]_\wedge \mathbf{R})\mathbf{A}^{-1}$ is called the *rigidity constraint*.

[1] An affine reconstruction differs from the true one by an affine transformation.

The seven parameters of the fundamental matrix are available to describe the geometric relationship between the two views; the five parameters of the essential matrix are needed to describe the rigid displacement, thus at most two independent constraint are available for the computation of the intrinsic parameters from the fundamental matrix. Indeed, Hartley [30] proposed an algorithm to factor the fundamental matrix that yields the five motion parameters and the two different focal lengths. He also noticed that no more information could be extracted from the fundamental matrix without making additional assumptions.

In the case of a moving camera with constant intrinsic parameters, it is possible to obtain a Euclidean reconstruction by cumulating constraints over different displacements. There are five unknown (the intrinsic parameters), each displacement yields two independent constraints, hence three views are sufficient (between three views there are three independent displacements: 1-2, 1-3 and 2-3).

4.1 Kruppa equations

With a minimum of three displacements, we can obtain the internal parameters of the camera using a system of polynomial equations due to Kruppa [34], which are derived from a geometric interpretation of the rigidity constraint [28, 35].

The unknown in the Kruppa equations is the matrix $\mathbf{K} = \mathbf{A}\mathbf{A}^\top$, called the *Kruppa coefficients matrix*, that represents the dual of the image of the *absolute conic* (see [10] for details). From $\mathbf{K}$ one can easily obtain the intrinsic parameters by means of Cholesky factorization ($\mathbf{K}$ is symmetric and positive definite), or in closed form:

$$\text{if} \quad \mathbf{K} = \begin{bmatrix} k_1 & k_2 & k_3 \\ k_2 & k_4 & k_5 \\ k_3 & k_5 & 1 \end{bmatrix} \quad \text{then} \quad \mathbf{A} = \begin{bmatrix} \sqrt{k_1 - {k_3}^2 - \frac{(k_2 - k_3 k_5]^2}{k_4 - {k_5}^2}} & \frac{k_2 - k_3 k_5}{\sqrt{k_4 - {k_5}^2}} & k_3 \\ 0 & \sqrt{k_4 - {k_5}^2} & k_5 \\ 0 & 0 & 1 \end{bmatrix}. \quad (20)$$

Kruppa equations were rediscovered and derived by Maybank and Faugeras [28]. Recently Hartley [36] provided a simpler form, based on the Singular Value Decomposition of the fundamental matrix. Let $\mathbf{F}$ be written as $\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ (with SVD), and

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^\top \\ \mathbf{u}_2^\top \\ \mathbf{u}_3^\top \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \mathbf{v}_3^\top \end{bmatrix} \quad \mathbf{D} = \mathrm{diag}(r, s, 0).$$

Then the Kruppa equations write (the derivation can be found in [36])

$$\frac{\mathbf{v}_2^\top \mathbf{K} \mathbf{v}_2}{r^2 \mathbf{u}_1^\top \mathbf{K} \mathbf{u}_1} = \frac{-\mathbf{v}_2^\top \mathbf{K} \mathbf{v}_1}{rs \mathbf{u}_1^\top \mathbf{K} \mathbf{u}_2} = \frac{\mathbf{v}_1^\top \mathbf{K} \mathbf{v}_1}{s^2 \mathbf{u}_2^\top \mathbf{K} \mathbf{u}_2}. \tag{21}$$

From (21) one obtains two independent quadratic equations in the five parameters of $\mathbf{K}$ for each fundamental matrix (i.e., for each displacement). Moreover, assuming that $\gamma = 0$, which is a good approximation for usual cameras, one has the additional constraint $k_3 k_5 = k_2$ [32]. There are basically two classes of methods for solving the resulting system of equations (assuming that more than three views are available) [32, 37]:

- Partition the equations set in groups of five and solve each group with a global convergent technique for systems of polynomial equations, like homotopy continuation methods [38, 39]. Each system will give a set of solutions and the solution common to all of them is chosen. This method – presented in [32] – has the great advantage of global convergence, but is computationally expensive. Moreover, the number of systems to be solved rapidly increases with the number of displacements.

- The over-constrained system of equation is solved with a non-linear least-squares technique (Levenberg-Marquardt [40], or Iterated Extended Kalman Filter [41]). The problem with non-linear least-squares is that a starting point close to the solution is needed. This can be obtained by applying globally convergent methods to subsets of equations (like in the previous case), or by making the additional assumption that (u_0, v_0) is in the center of the image, thereby obtaining (from just one fundamental matrix) two quadratic equations in two variables k_1, k_4, which can be solved analytically [36]. This technique is used in [37].

5 Stratification

Let us assume that a projective reconstruction is available, that is a sequence $\{\tilde{\mathbf{P}}^i_{\mathrm{proj}}\}$ of camera matrices such that:

$$\tilde{\mathbf{P}}^0_{\mathrm{proj}} = [\mathbf{I} \mid \mathbf{0}]; \qquad \tilde{\mathbf{P}}^i_{\mathrm{proj}} = [\mathbf{Q}^i \mid \mathbf{q}^i]. \tag{22}$$

We are looking for a Euclidean reconstruction, that is a 4×4 nonsingular matrix $\tilde{\mathbf{T}}$ that upgrades the projective reconstruction to Euclidean. If $\{\tilde{\mathbf{w}}_j\}$ is the sought Euclidean structure, $\tilde{\mathbf{T}}$ must be such that: $\tilde{\mathbf{m}}^i_j = \tilde{\mathbf{P}}^i_{\mathrm{proj}}\tilde{\mathbf{T}}\tilde{\mathbf{T}}^{-1}\tilde{\mathbf{w}}_j$, hence

$$\tilde{\mathbf{P}}^i_{\mathrm{eucl}} \simeq \tilde{\mathbf{P}}^i_{\mathrm{proj}}\tilde{\mathbf{T}} \ , \tag{23}$$

where the symbol $\simeq$ means "equal up to a scale factor."

5.1 Using additional information

Projective reconstruction differs from Euclidean by an unknown projective transformation in the 3-D projective space, which can be seen as a suitable change of basis. Thanks to the fundamental theorem of projective geometry [42], a collineation in space is determined by five points, hence the knowledge of the true (Euclidean) position of five points allows to compute the unknown 4×4 matrix $\tilde{\mathbf{T}}$ that transform the Euclidean frame into the projective frame. An application of this is reported in [43]. Moreover, if intrinsic parameters $\mathbf{A}$ are known, then $\tilde{\mathbf{T}}$ can be computed by solving a linear system of equations (see (52) in Sec. 5.2.5).

5.2 Euclidean reconstruction from constant intrinsic parameters

The challenging problem is to recover $\tilde{\mathbf{T}}$ without additional information, using only the *hypothesis of constant intrinsic parameters.* The works by Hartley [18], Pollefeys and Van Gool [44], Heyden and Åström [45], Triggs [46] and Bougnoux [47] will be reviewed, but first we will make some remarks that are common to most of the methods.

We can choose the first Euclidean-calibrated camera to be $\tilde{\mathbf{P}}^0_{\mathrm{eucl}} = \mathbf{A}[\mathbf{I} \mid \mathbf{0}]$, thereby fixing arbitrarily the rigid transformation:

$$\tilde{\mathbf{P}}^0_{\text{eucl}} = \mathbf{A}[\mathbf{I} \mid \mathbf{0}] \qquad \tilde{\mathbf{P}}^i_{\text{eucl}} = \mathbf{A}[\mathbf{R}^i \mid \mathbf{t}^i]. \tag{24}$$

With this choice, it is easy to see that $\tilde{\mathbf{P}}^0_{\text{eucl}} = \tilde{\mathbf{P}}^0_{\text{proj}}\tilde{\mathbf{T}}$ implies

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{r}^\top & s \end{bmatrix} \tag{25}$$

where $\mathbf{r}^\top$ is an arbitrary vector of three elements $[r_1\ r_2\ r_3]$. Under this parameterization $\tilde{\mathbf{T}}$ is clearly non singular, and being defined up to a scale factor, it depends on eight parameters (let $s = 1$).

Substituting (22) in (23) one obtains

$$\tilde{\mathbf{P}}^i_{\text{eucl}} \simeq \tilde{\mathbf{P}}^i_{\text{proj}}\tilde{\mathbf{T}} = [\mathbf{Q}^i\mathbf{A} + \mathbf{q}^i\mathbf{r}^\top \mid \mathbf{q}^i], \tag{26}$$

and from (24)

$$\tilde{\mathbf{P}}^i_{\text{eucl}} = \mathbf{A}[\mathbf{R}^i \mid \mathbf{t}^i] = [\mathbf{A}\mathbf{R}^i \mid \mathbf{A}\mathbf{t}^i], \tag{27}$$

hence

$$\mathbf{Q}^i\mathbf{A} + \mathbf{q}^i\mathbf{r}^\top \simeq \mathbf{A}\mathbf{R}^i. \tag{28}$$

This is the basic equation, relating the unknowns $\mathbf{A}$ (five parameters) and $\mathbf{r}$ (three parameters) to the available data $\mathbf{Q}^i$ and $\mathbf{q}^i$. $\mathbf{R}$ is unknown, but must be a rotation matrix.

Affine reconstruction. Equation (28) can be rewritten as

$$\mathbf{Q}^i + \mathbf{q}^i\mathbf{r}^\top\mathbf{A}^{-1} \simeq \mathbf{A}\mathbf{R}^i\mathbf{A}^{-1} = \mathbf{H}^i_\infty, \tag{29}$$

relating the unknown vector $\mathbf{a}^\top = \mathbf{r}^\top \mathbf{A}^{-1}$ to the homography of the infinity plane (compare (29) with (18)). It can be seen that $\tilde{\mathbf{T}}$ factorizes as follows

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{a}^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}. \tag{30}$$

The right-hand matrix is an *affine transformation*, not moving the infinity plane, whereas the left-hand one is a transformation moving the infinity plane.
Substituting the latter into (23) we obtain:

$$\tilde{\mathbf{P}}^i_{\text{eucl}} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \tilde{\mathbf{P}}^i_{\text{affi}} \simeq \tilde{\mathbf{P}}^i_{\text{proj}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{a}^\top & 1 \end{bmatrix} = [\mathbf{H}^i_\infty | \mathbf{q}^i] \tag{31}$$

Therefore, the knowledge of the homography of the infinity plane (given by $\mathbf{a}$) allows to compute the Euclidean structure up to an affine transformation, that is an *affine reconstruction.*

From affine to Euclidean. Another useful observation is, if $\mathbf{H}_\infty$ is known and the intrinsic parameters are constant, the intrinsic parameters matrix $\mathbf{A}$ can easily be computed [8, 18, 14, 48].
Let us consider the case of two cameras. If $\mathbf{A}' = \mathbf{A}$, then $\mathbf{H}_\infty$ is exactly known (with the right scale), since

$$\det(\mathbf{H}_\infty) = \det(\mathbf{A}\mathbf{R}\mathbf{A}^{-1}) = 1. \tag{32}$$

From (17) we obtain $\mathbf{R} = \mathbf{A}'^{-1}\mathbf{H}_\infty\mathbf{A}$, and, since $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$, it is easy to obtain:

$$\mathbf{H}_\infty \mathbf{K} \mathbf{H}_\infty^\top = \mathbf{K} \tag{33}$$

where $\mathbf{K} = \mathbf{A}\mathbf{A}^\top$ is the Kruppa coefficients matrix. As (33) is an equality between 3×3 symmetric matrices, we obtain a linear system of six equations in the five unknown k_1, k_2, k_3, k_4, k_5 . In fact, only four equations are independent [14, 48], hence at least three views (with constant intrinsic parameters) are required to obtain an over-constrained linear system, which can be easily solved with a linear least-squares technique.

Note that two views would be sufficient under the usual assumption that the image reference frame is orthogonal ($\gamma = 0$), which gives the additional constraint $k_3 k_5 = k_2$.
If points at infinity (in practice, sufficiently far from the camera) are in the scene, $\mathbf{H}_\infty$ can be computed from point correspondences, like any ordinary plane homography [48]. Moreover, with additional knowledge, it can be estimated from vanishing points or parallelism [33, 49], or constrained motion [8].
In the rest of the section, some of the most promising stratification techniques will be reviewed.

5.2.1 Hartley

Hartley [18] pioneered this kind of approach. Starting from (28), we can write

$$(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top)\mathbf{A} \simeq \mathbf{A}\mathbf{R}^i. \tag{34}$$

By taking the QR decomposition of the left-hand side we obtain an upper triangular matrix $\mathbf{B}^i$ such that $(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top)\mathbf{A} = \mathbf{B}^i \mathbf{R}^i$, so (34) rewrites $\mathbf{B}^i \mathbf{R}^i = \lambda^i \mathbf{A}\mathbf{R}^i$ or

$$\frac{1}{\lambda^i}\mathbf{A}^{-1}\mathbf{B}^i = \mathbf{I}. \tag{35}$$

The scale factor $1/\lambda^i$ can be chosen so that the sum of the squares of the diagonal entries of $(1/\lambda^i)\mathbf{A}^{-1}\mathbf{B}^i$ equals three. We seek $\mathbf{A}$ and $\mathbf{a}$ that minimizes

$$\sum_{i>0} \left\| \frac{1}{\lambda^i}\mathbf{A}^{-1}\mathbf{B}^i - \mathbf{I} \right\|^2. \tag{36}$$

Each camera excluding the first, gives six constraints in eight unknowns, so three cameras are sufficient. In practice there are more than three cameras, and the non-linear least squares problem can be solved with Levenberg-Marquardt minimization algorithm [40]. As noticed in the case of Kruppa equations, a good initial guess for the unknowns $\mathbf{A}$ and $\mathbf{a}$ is needed in order for the algorithm to converge to the solution.
Given that from $\mathbf{H}^i_\infty$ the computation of $\mathbf{A}$ is straightforward, a guess for $\mathbf{a}$ (that determines $\mathbf{H}^i_\infty$) is sufficient. The *cheirality constraints* [50] are exploited by Hartley to estimate the infinity plane homography, thereby obtaining an approximate affine (or *quasi-affine*)

reconstruction.

5.2.2 Pollefeys and Van Gool

In this approach [44], a projective reconstruction is first updated to affine reconstruction by the use of the *modulus constraint* [14, 51]: since the left-hand part of (29) is conjugated to a (scaled) rotation matrix, all eigenvalues must have equal moduli. Note that this holds if and only if intrinsic parameters are constant. To make the constraint explicit we write the characteristic polynomial:

$$\det(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top - \lambda \mathbf{I}) = l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0. \tag{37}$$

The equality of the roots of the characteristic polynomial is not easy to impose, but a simple necessary condition holds:

$$l_3 l_1^3 = l_2^3 l_0. \tag{38}$$

This yields a fourth order polynomial equation in the unknown $\mathbf{a}$ for each camera except the first, so a finite number of solutions can be found for four cameras. Some solutions will be discarded using the modulus constraint, that is more stringent than (38).
As discussed previously, autocalibration is achievable with only three views. It is sufficient to note that, given three cameras, for every plane homography, the following holds [14]:

$$\mathbf{H}^{1,3} = \mathbf{H}^{2,3}\mathbf{H}^{1,2}. \tag{39}$$

In particular it holds for the infinity plane homography, so

$$\mathbf{H}^{i,j}_\infty = \mathbf{H}^j_\infty {\mathbf{H}^i_\infty}^{-1} \simeq (\mathbf{Q}^j + \mathbf{q}^j \mathbf{a}^\top)(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top)^{-1}. \tag{40}$$

In this way we obtain a constraint on the plane at infinity for each pair of views. Let us write the characteristic polynomial:

$$\det((\mathbf{Q}^j + \mathbf{q}^j \mathbf{a}^\top)(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top)^{-1} - \lambda \mathbf{I}) = 0 \quad \Longleftrightarrow \tag{41}$$

$$\det((\mathbf{Q}^j + \mathbf{q}^j \mathbf{a}^\top) - \lambda(\mathbf{Q}^i + \mathbf{q}^i \mathbf{a}^\top)) = 0 \tag{42}$$

Writing the constraint (38) for the three views, a system of three polynomial of degree four in three unknowns is obtained. Here, like in the solution of Kruppa equations, homotopy continuation methods could be applied to compute all the $4^3 = 64$ solutions.
In practice more than three views are available, and we must solve a non-linear least-squares problem: Levenberg-Marquardt minimization is used by the author.

5.2.3 Heyden and Åström

The method proposed by Heyden and Åström [45] is again based on (28), which can be rewritten as

$$\tilde{\mathbf{P}}^i_{\text{proj}} \begin{bmatrix} \mathbf{A} \\ \mathbf{r}^\top \end{bmatrix} \simeq \mathbf{A}\mathbf{R}^i. \tag{43}$$

Since $\mathbf{R}^i {\mathbf{R}^i}^\top = \mathbf{I}$ it follows that:

$$\tilde{\mathbf{P}}^i_{\text{proj}} \begin{bmatrix} \mathbf{A} \\ \mathbf{r}^\top \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{r}^\top \end{bmatrix}^\top {\tilde{\mathbf{P}}^i}{}^\top_{\text{proj}} = \tilde{\mathbf{P}}^i_{\text{proj}} \begin{bmatrix} \mathbf{A}\mathbf{A}^\top & \mathbf{A}\mathbf{r} \\ \mathbf{r}^\top\mathbf{A}^\top & \mathbf{r}^\top\mathbf{r} \end{bmatrix} {\tilde{\mathbf{P}}^i}{}^\top_{\text{proj}} \simeq \mathbf{A}\mathbf{R}^i {\mathbf{R}^i}^\top \mathbf{A}^\top = \mathbf{A}\mathbf{A}^\top. \tag{44}$$

The constraints expressed by (44) are called the Kruppa constraints [45]. Note that (44) contains five equations, because the matrices of both members are symmetric, and the homogeneity reduces the number of equations with one. Hence, each camera matrix, apart from the first one, gives five equations in the eight unknowns $\alpha_u, \alpha_v, \gamma, u_0, v_0, r_1, r_2, r_3$. A unique solution is obtained when three cameras are available. If the unknown scale factor is introduced explicitly, (44) rewrites:

$$0 = f_i(\mathbf{A}, \mathbf{r}, \lambda_i) = \lambda_i^2 \mathbf{A}\mathbf{A}^\top - \tilde{\mathbf{P}}^i_{\text{proj}} \begin{bmatrix} \mathbf{A}\mathbf{A}^\top & \mathbf{A}\mathbf{r} \\ \mathbf{r}^\top\mathbf{A}^\top & \mathbf{r}^\top\mathbf{r} \end{bmatrix} {\tilde{\mathbf{P}}^i}{}^\top_{\text{proj}}. \tag{45}$$

Therefore, 3 cameras yield 10 equations in 8 unknowns.

5.2.4 Triggs

Triggs [46] proposed a method based on the *absolute quadric* and, independently from Heyden and Åström, he derived an equation closely related to (44). The absolute quadric

$\mathbf{\Omega}$ consists of planes tangent to the absolute conic [10], and in a Euclidean frame, is represented by the matrix

$$\mathbf{\Omega}_{\text{euc}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{46}$$

If $\tilde{\mathbf{T}}$ is a projective transformation acting as in (23), then it can be verified [46] that it transforms $\mathbf{\Omega}_{\text{euc}}$ into $\mathbf{\Omega} = \tilde{\mathbf{T}}\mathbf{\Omega}_{\text{euc}}\tilde{\mathbf{T}}^{\top}$. Since the projection of the absolute quadric yields the dual image of the absolute conic [46], one obtain

$$\tilde{\mathbf{P}}^i_{\text{proj}}\mathbf{\Omega}\tilde{\mathbf{P}}^{i\top}_{\text{proj}} \simeq \mathbf{K} \tag{47}$$

from which, using (25), (44) follows immediately. Triggs, however, does not assume any particular form for $\tilde{\mathbf{T}}$, hence the unknown are $\mathbf{K}$ and $\mathbf{\Omega}$. Note that both these matrix are symmetric and defined up to a scale factor.

Let $\mathbf{k}$ be the matrix composed by the the six elements of the lower triangle of $\mathbf{K}$, and $\boldsymbol{\omega}$ be the matrix composed by the six elements of the lower triangle of $\tilde{\mathbf{P}}^i_{\text{proj}}\mathbf{\Omega}\tilde{\mathbf{P}}^{i\top}_{\text{proj}}$, then (47) is tantamount to saying that the two vectors are equal up to a scale, hence

$$\mathbf{k} \wedge \boldsymbol{\omega} = \mathbf{0} \tag{48}$$

in which the unknown scale factor is eliminated. For each camera this amounts to 15 bilinear equations in $9 + 5$ unknowns, since both $\mathbf{k}$ and $\boldsymbol{\omega}$ are defined up to a scale factor. Since only five of them are linearly independent, at least three images are required for a unique solution.

Triggs uses two methods for solving the non-linear least-squares problem: sequential quadratic programming [40] on $N \geq 3$ cameras, and a quasi-linear method with SVD factorization on $N \geq 4$ cameras. He recommend to use data standardization [52] and to enforce $\det(\mathbf{\Omega}) = 3$. The sought transformation $\tilde{\mathbf{T}}$ is computed by taking the eigen-decomposition of $\mathbf{\Omega}$.

5.2.5 Bougnoux

This methods [47] is different from the previous ones, because it does not require constant intrinsic parameters and because it achieves only an approximate Euclidean reconstruction, without obtaining meaningful camera parameters as a by-product.
Let us write (23) in the following form:

$$\tilde{\mathbf{P}}^i_{\text{eucl}} = \left[\begin{array}{c|c} \mathbf{q}_1^{i\top} & \\ \mathbf{q}_2^{i\top} & \mathbf{q}^i \\ \mathbf{q}_3^{i\top} & \end{array} \right] \simeq \tilde{\mathbf{P}}^i_{\text{proj}} \tilde{\mathbf{T}} \tag{49}$$

where $\mathbf{q}_1^{i\top}, \mathbf{q}_2^{i\top}, \mathbf{q}_3^{i\top}$ are the rows of $\tilde{\mathbf{P}}^i_{\text{eucl}}$. The usual assumptions $\gamma = 0$ and $\alpha_u = \alpha_v$, are used to constraint the Euclidean camera matrices:

$$\gamma = 0 \quad \Longleftrightarrow \quad (\mathbf{q}_1^i \wedge \mathbf{q}_3^i)^\top (\mathbf{q}_2^i \wedge \mathbf{q}_3^i) = 0 \tag{50}$$

$$\alpha_u = \alpha_v \quad \Longleftrightarrow \quad ||\mathbf{q}_1^i \wedge \mathbf{q}_3^i|| = ||\mathbf{q}_2^i \wedge \mathbf{q}_3^i||. \tag{51}$$

Thus each camera, excluding the first, gives two constraints of degree four. Since we have six unknown, at least four cameras are required to compute $\tilde{\mathbf{T}}$. If the principal point (u_0, v_0) is forced to the image center, the unknowns reduce to four and only three cameras are needed.
The non-linear minimization required to solve the resulting system is rather unstable and must be started close to the solution: an estimate of the focal length and $\mathbf{r}$ is needed. Assuming known principal point, no skew, and unit aspect ratio, the focal length α_u can be computed from the Kruppa equations in closed form [47]. Then, given the intrinsic parameters $\mathbf{A}$, an estimate of $\mathbf{r}$ can be computed by solving a *linear* least-squares problem. From (44) the following is obtained:

$$\mathbf{Q}^i \mathbf{A} \mathbf{A}^\top \mathbf{Q}^{i^\top} + \mathbf{Q}^i \mathbf{A} \mathbf{r} \mathbf{q}^{i^\top} + (\mathbf{Q}^i \mathbf{A} \mathbf{r} \mathbf{q}^{i^\top})^\top + ||\mathbf{r}||^2 \mathbf{q}^i \mathbf{q}^{i^\top} = \lambda^2 \mathbf{A} \mathbf{A}^\top. \tag{52}$$

Since $[\mathbf{A}\mathbf{A}^\top]_{3,3} = \mathbf{K}_{3,3} = 1$, then λ is fixed. After some algebraic manipulation [47], one ends up with four linear equations in $\mathbf{Ar}$. This method works also with varying intrinsic parameters, although, in practice, only the focal length is allowed to vary, since principal point is forced to the image center and no skew and unit aspect ratio are assumed. The estimation of the camera parameters is inaccurate, nevertheless Bougnoux proves that the reconstruction is correct up to an anisotropic homotethy, which he claims to be enough for the reconstructed model to be usable.

6 Discussion

The applicability of autocalibration techniques in the real world depends on two issues: sensitivity to noise and initialization. The challenge is to devise a method that exhibits graceful degradation as noise increases and needs only an approximate initialization. Several attempt have been made, as reported in this survey, but the problem is is far from being solved yet.

As for the Kruppa equations, in [32] the authors compare three solving methods: the homotopy continuation method, Levenberg-Marquardt and the Iterated Extended Kalman Filter. From the simulations reported, it appears that all the methods give comparable results. However, the homotopy continuation method is suitable for the case of few displacements, as it would be difficult to use all the constraints provided by a long sequence, and its computational cost would be too high. Iterative approaches (Levenberg-Marquardt and Iterated Extended Kalman Filter) are well suited to the case where more displacements are available. The main limitation of all these methods is the sensitivity to the noise in the localization of points.

Methods based on stratification have appeared only recently, and only preliminary and partial results are available. In many cases they show a graceful degradation as noise increases, but the issue of initialization is not always addressed.

Hartley's algorithm leads to a minimization problem that requires a good initial guess; this is computed using a quite complicated method, involving the cheirality constraints. Pollefeys-VanGool's algorithm leads to an easier minimization, and this justify the claim that convergence toward a global minimum is relatively easily obtained. It is unclear, however, how the initial guess has to be chosen. The method proposed by Heyden and Åström was evaluated only on one example, and was initialized close to the ground-truth. Experiments on synthetic data reported by Triggs, suggest that his non-linear algorithm is stable and requires only approximate initialization (the author reports that initial calibration may be wrong up to 50%).

Bougnoux's algorithm is quite different form the others, since its goal is not to obtain an accurate Euclidean reconstruction. Assessment of reconstruction quality is only visual.

7 Conclusions

This paper presented a review of recent techniques for Euclidean reconstruction from a single moving camera, with unconstrained motion and unknown *constant parameters*. Such unified, comparative discussion, which has not yet been presented in the literature, sheds light on the relations between different methods. Indeed, even though formulations may be different, to all the methods reviewed, much of the underlying mathematics is common. However, since problems are inherently non-linear, proper formulation is very important to avoid difficulties created by the numerical computation of the solutions.

Despite this problem is far from being completely solved, the more general one in which intrinsic parameters are varying is gaining the attention of researchers. In fact, Bougnoux's method already copes with varying parameters. Heyden and Åström [53] proposed a method that works with varying and unknown focal length and principal point. Later, they proved [54] that it is sufficient to know any of the five intrinsic parameters to make Euclidean reconstruction, even if all other parameters are unknown and varying. A similar method that can work with different types of of constraints has been recently presented in [55].

Acknowledgments

The help of Arrigo Benedetti and Emanuele Trucco who read preliminary drafts of this paper is gratefully acknowledged.

References

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

[2] C.C. Slama, editor. *Manual of Photogrammetry.* American Society of Photogrammetry, 4th edition, 1980.

[3] K.B. Atkinson. *Close Range Photogrammetry and Machine Vision.* Whittles Publisher, 1996.

[4] T. S. Huang and A. N. Netravali. Motion and structure from feature correspondences: A review. *Proceedings of IEEE*, 82(2):252–267, 1994.

[5] M. Armstong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In *Proceedings of the European Conference on Computer Vision*, pages 5–16, Cambridge, UK, 1996.

[6] R. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings of the European Conference on Computer Vision*, pages 471–478, Stockholm, 1994.

[7] T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12(4):227–237, 1994.

[8] M. Armstong, A. Zisserman, and P. Beardsley. Euclidean structure from uncalibrated images. In *British Machine Vision Conference*, pages 509–518, 1994.

[9] Z. Zhang, Q.-T. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. *IEEE Transactions on Robotics and Automation*, 12(1):103–113, February 1996.

[10] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* The MIT Press, Cambridge, MA, 1993.

[11] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17:43–75, 1996.

[12] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, March/April 1998.

[13] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(10):133–135, September 1981.

[14] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996.

[15] R. Mohr and E. Arbogast. It can be done without camera calibration. *Pattern Recognition Letters*, 12:39–43, 1991.

[16] G. Sparr. An algebraic-analytic method for reconstruction from image correspondences. In *Scandinavian Conference on Image Analysis*, pages 274–281, 1991.

[17] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the European Conference on Computer Vision*, pages 563–578, Santa Margherita L., 1992.

[18] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In Joseph Mundy and Andrew Zisserman, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 237–256. Springer-Verlag, 1993.

[19] A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1058–1063, 1995.

[20] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proceedings of the European Conference on Computer Vision*, pages 709–720, Cambridge, UK, 1996.

[21] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proceedings of the International Conference on Pattern Recognition*, 1996.

[22] A. Heyden. Projective structure and motion from image sequences using subspace methods. In *Scandinavian Conference on Image Analysis*, 1997.

[23] P. Beardsley, A. Zisserman, and D. Murray. Sequential update of projective and affine structure from motion. *International Journal of Computer Vision*, 23(3):235–259, 1997.

[24] S. Avidan and A. Shashua. Threading fundamental matrices. In *Proceedings of the European Conference on Computer Vision*, pages 124–140, University of Freiburg, Germany, 1998.

[25] T. Ueshiba and F. Tomita. A factorization method for projective and Euclidean reconstruction from multiple perspective views via iterative depth estimation. In *Proceedings of the European Conference on Computer Vision*, pages 296–310, University of Freiburg, Germany, 1998.

[26] C. Rothwell, A. Zisserman, J. Mundy, and D. Forsyth. Efficient model library access by projectively invariant indexing functions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1992.

[27] L. Robert and O. Faugeras. Relative 3-D positioning and 3-D convex hull computation from a weakly calibrated stereo pair. *Image and Vision Computing*, 13(3):189–197, 1995.

[28] S. J. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[29] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: Theory and experiments. In *Proceedings of the European Conference on Computer Vision*, pages 321–334, Santa Margherita L., 1992.

[30] R. I. Hartley. Estimation of relative camera position for uncalibrated cameras. In *Proceedings of the European Conference on Computer Vision*, pages 579–587, Santa Margherita L., 1992.

[31] B.K.P. Horn. Relative orientation revisited. *Journal of the Optical Society of America A*, 8(10):1630–1638, October 1991.

[32] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.

[33] O. Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, 12(3):465–484, 1994.

[34] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Kl., Abt. IIa.*, 122:1939–1948, 1913.

[35] O. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 4(3):225–246, June 1990.

[36] R. I. Hartley. Kruppa's equations derived from the fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):133–135, February 1997.

[37] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. Research Report 2793, INRIA, February 1996.

[38] A. P. Morgan, A. J. Sommese, and L. T. Watson. Finding all isolated solutions to polynomial systems using hompack. *ACM Trans. Math. Software*, 15:93–122, 1989.

[39] A. J. Sommese and C. W. Wampler. Numerical algebraic geometry. In J. Renegar, M. Shub, and S. Smale, editors, *Mathematics of Numerical Analysis: Real Number Algorithms*, volume 32 of *Lectures in Applied Mathematics*, pages 749–763. Park City, Utah, 1996.

[40] P. Gill, W. Murray, and M. Wright. *Practical Optimization.* Academic Press, 1981.

[41] P. S. Maybeck. *Stochastic Models, Estimation and Control*, volume 1. Academic Press, New York, NY, 1979.

[42] J. G. Semple and G. T. Kneebone. *Algebraic projective geometry.* Oxford University Press, 1952.

[43] K. Plakas, E. Trucco, and A. Fusiello. Uncalibrated vision for 3-D underwater applications. In *Proceedings of the OCEANS'98 Conference*, pages 272–276, Nice, France, September 1998. IEEE/OS.

[44] M. Pollefeys and L. Van Gool. A stratified approach to metric self-calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 407–412, Puerto Rico, 1997.

[45] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proceedings of the International Conference on Pattern Recognition*, pages 339–343, Vienna, 1996.

[46] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 609–614, Puerto Rico, 1997.

[47] S. Bougnoux. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 790–796, Bombay, 1998.

[48] T. Viéville, C. Zeller, and L. Robert. Using collineations to compute motion and structure in an uncalibrated image sequence. *International Journal of Computer Vision*, 20(3):213–242, 1996.

[49] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, and I. Zoghlami. 3-d reconstruction of urban scenes from image sequences. *Computer Vision and Image Understanding*, 69(3):292–309, March 1998.

[50] R. I. Hartley. Cheirality invariants. In *Proceedings of the Image Understanding Workshop*, pages 745–753, Washington, DC, April 1993. ARPA, Morgan Kaufmann.

[51] M. Pollefeys, L. Van Gool, and A. Oosterlinck. The modulus constraint: a new constraint for self-calibration. In *Proceedings of the International Conference on Pattern Recognition*, pages 349–353, Vienna, 1996.

[52] R. I. Hartley. In defence of the 8-point algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*, 1995.

[53] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 438–443, Puerto Rico, 1997.

[54] A. Heyden and K. Åström. Minimal conditions on intrinsic parameters for Euclidean reconstruction. In *Proceedings of the Asian Conference on Computer Vision*, Hong Kong, 1998.

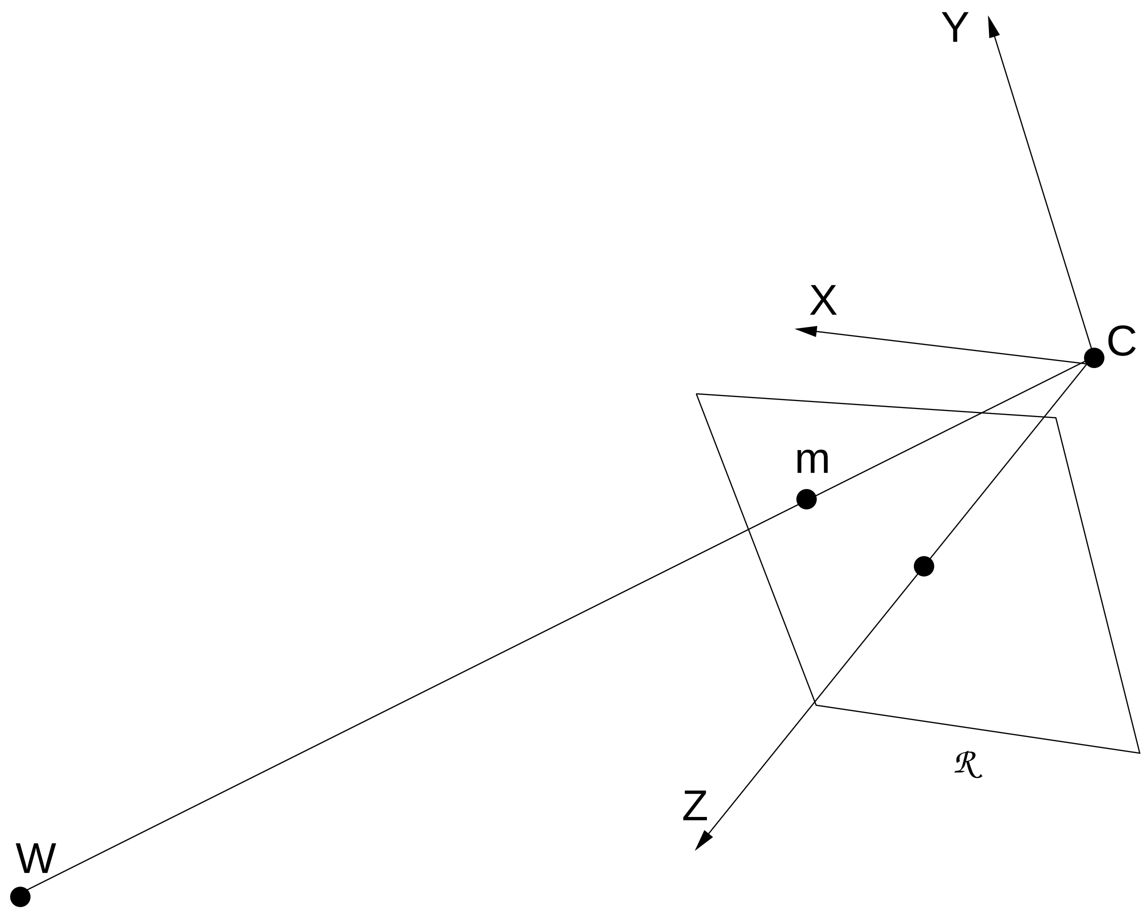
[55] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 90–95, Bombay, 1998.

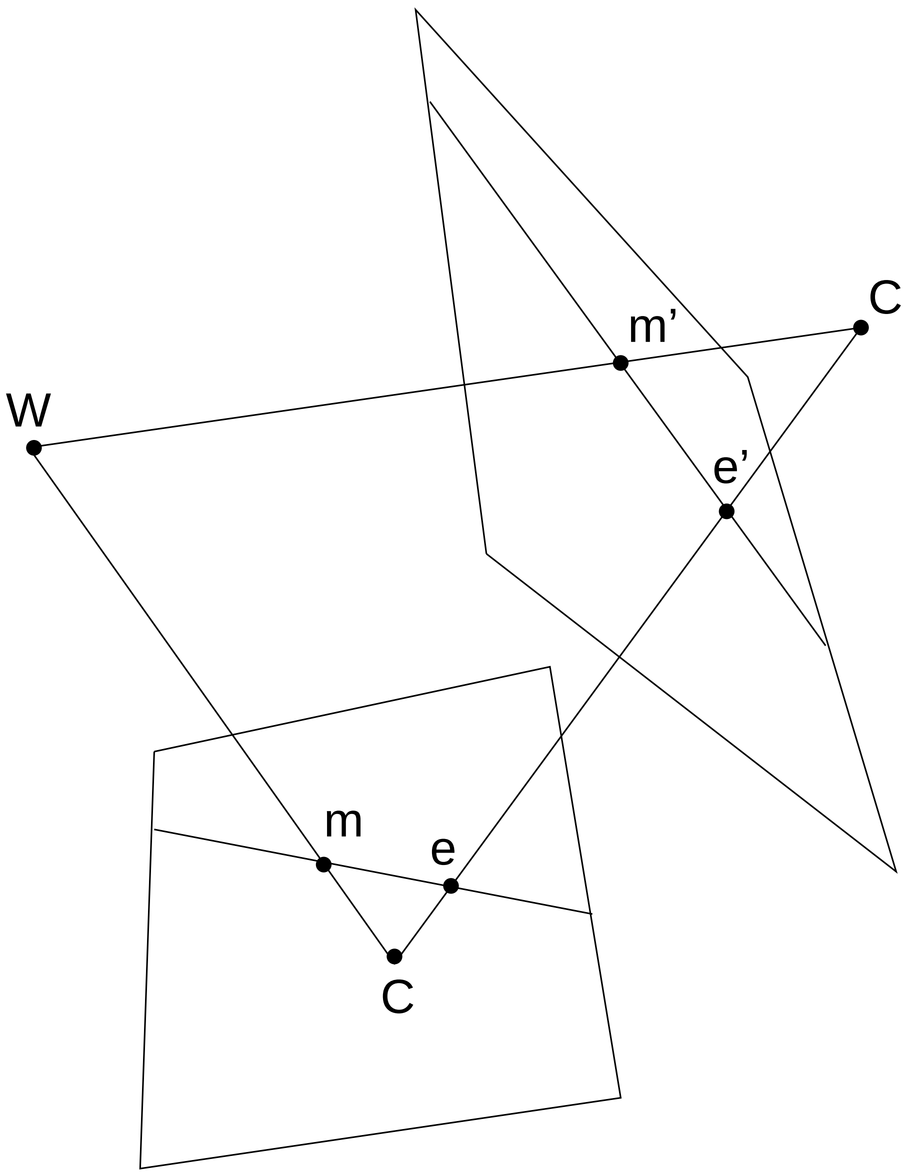
Image and Vision Computing - A. Fusiello - Captions to illustrations

Figure 1: The pinhole camera model, with the camera reference frame (X,Y,Z) depicted. Z is also called the *optical axis*.

Figure 2: Epipolar geometry. The epipole of the first camera e is the projection of the optical center C' of the second camera (and vice versa).

Image and Vision Computing - A. Fusiello - Figure 1




C'
m'
W
e'
m
e
C