Short Papers

Camera Self-Calibration from Unknown Planar Structures Enforcing the Multiview Constraints between Collineations

Ezio Malis and Roberto Cipolla

Abstract—In this paper, we describe an efficient method to impose the constraints existing between the collineations between images which can be computed from a sequence of views of a planar structure. These constraints are usually not taken into account by multiview techniques in order not to increase the computational complexity of the algorithms. However, imposing the constraints is very useful since it allows a reduction of geometric errors in the reprojected features and provides a consistent set of collineations which can be used for several applications such as mosaicing, reconstruction, and self-calibration. In order to show the validity of our approach, this paper focus on self-calibration from unknown planar structures proposing a method exploiting the consistent set of collineations. Our method can deal with an arbitrary number of views and an arbitrary number of planes and varying camera internal parameters. However, for simplicity, this papers will only discuss the case with one plane in several views. The results obtained with synthetic and real data are very accurate and stable even when using only few images.

Index Terms—Self-calibration, multiple views, planes, collineation, nonlinear constraints.

1 Introduction

The particular geometry of features lying on planes is often the reason for the inaccuracy of many computer vision applications (structure from motion, self-calibration) if it is not taken explicitly into account in the algorithms. Introducing some knowledge about the coplanarity of the features and about their structure (metric or topological) can improve the quality of the estimates [20]. However, the only prior geometric knowledge on the features that will be used here is their coplanarity. Two views of a plane are related by a collineation. Using multiple views of a plane, we obtain a set of collineations which are not independent. If there are multiple planes in the scene there will be a set of collineations for each plane and again some constraints between the different sets. In order to avoid solving nonlinear optimization problems, the constraints existing within a set of collineation and between sets have often been neglected. However, these multiview constraints can be used to improve the estimation of the collineations matrices as in [24], where multiple planes (≥ 2) are supposed to be viewed in the images. In this paper, we analyze the constraints existing between a set of collineations induced by a simple plane in the image but it is very easy to extend our analysis to the case of multiple planes. Imposing the constraint is useful since it allows the reduction of the geometric error in the reprojected features and provides a consistent set of collineations which can be used for several applications as mosaicing, reconstruction, and self-calibration. In this paper, we will focus on camera self-calibration. Camera self-calibration from views of a generic scene has been widely investigated [13], [22], [5], [17], [18], [8]. Depending on the a priori information provided the self-calibration algorithms can be classified as follows: Algorithms that use some knowledge of the observed scene: identifiable targets of known shape [14] and metric structure of planes [19], [25]. Algorithms that exploit particular camera motions: translating camera [16] or rotating camera [6]. Algorithms that use some a priori knowledge on the camera parameters: some fixed camera parameters (i.e., skew zero, unit ratio, ...), varying camera parameters [17], [15]. Camera self-calibration from planar scenes with known metric structure has been investigated in several papers. However, it is interesting to develop flexible techniques which do not need any a priori knowledge about the camera motion as in [6] or metric knowledge of the planar scene. A method for self-calibrating a camera from views of planar scenes without knowing their metric structure was proposed in [23]. In this work, Triggs developed a self-calibration technique based on some constraints involving the absolute quadric and the scene-plane to image-plane collineations. However, in practice, it is not possible to estimate these collineations without knowing the metric structure of the plane. Only the collineations with respect to a reference view (a key image) can be used to self-calibrate a camera with constant internal parameters. As noticed by Triggs, inaccurate measurements or poor conditioning in the key image contribute to all the collineations reducing the numerical accuracy or the stability of the method. In order to average the uncertainty over all collineations, Triggs proposed, in an extended version of [23], a one-step collineation factorization method analogous to factorization-based projective structure and motion [21], [9]. However, the factorization proposed by Triggs is not used in [23] as it turns out to give slightly worse results in practice. In this paper, we propose to impose the constraints between collineation using a different iterative method proposed in [10]. By imposing the constraints, we obtain accurate self-calibration from planar scenes with unknown metric structure. We do not use any key image but all the images are treated equally averaging the uncertainty over all of them. Furthermore, our method can be applied for the self-calibration of a camera with varying focal length [11].

2 Two-View Geometry

In this section, we describe the relationship between two views of a planar structure. Each camera performs a perspective projection of a point $\mathbf{x} \in \mathbb{P}^3$ (with homogeneous coordinates $\mathbf{x} = (X, Y, Z, 1)$) to an image point $\mathbf{p} \in \mathbb{P}^2$ (with coordinates $\mathbf{p} = (u, v, 1)$) measured in pixels: $\mathbf{p} \propto \mathbf{K}\,[\mathbf{R}\ \mathbf{t}]\,\mathbf{x}$, where $\mathbf{R}$ and $\mathbf{t}$ represent the displacement between frame $\mathcal{F}$ attached to the camera and an absolute coordinate frame $\mathcal{F}_0$, and $\mathbf{K}$ is a nonsingular (3×3) upper triangular matrix containing the intrinsic parameters of the camera.

2.1 The Collineation Matrix in Projective Space

Let $\mathcal{F}_i$ and $\mathcal{F}_j$ be two frames attached respectively to the image $\mathcal{I}_i$ and $\mathcal{I}_j$. The two views of a planar object are related by a collineation matrix in projective space. The image coordinates $\mathbf{p}_{ik}$ of the point $\mathcal{P}_k$ in the image $\mathcal{I}_i$ can be obtained from the coordinates $\mathbf{p}_{jk}$ of $\mathcal{P}_k$ in image $\mathcal{I}_j$:

$$\mathbf{p}_{ik} \propto \mathbf{G}_{ij}\mathbf{p}_{jk}, \qquad (1)$$

where the collineation matrix $\mathbf{G}_{ij}$ is a (3×3) matrix defined up to scalar factor which can be written as: $\mathbf{G}_{ij} \propto \mathbf{K}_i\mathbf{H}_{ij}\mathbf{K}_j^{-1}$, where $\mathbf{H}_{ij}$ is the corresponding collineation matrix in the Euclidean space. In this paper, we will use the term "E-collineation" to indicate a collineation expressed in Euclidean space.

- *E. Malis is with INRIA, 2004, route des Lucioles, 06902 Sophia Antipolis Cedex, France. E-mail: ezio.malis@sophia.inria.fr.*
- *R. Cipolla is with the University of Cambridge, Cambridge CB2 1PZ, UK.*

Manuscript received 22 Aug. 2000; revised 4 May 2001; accepted 2 Jan. 2002.
Recommended for acceptance by Z. Zhang.
For information on obtaining reprints of tpami@computer.org, and reference IEEECS Log Number 112751.

2.2 The Collineation Matrix in Euclidean Space

The E-collineation matrix can be written as a function of the camera displacement and the normal to the plane [2]: $\mathbf{H}_{ij} = \mathbf{R}_{ij} + \frac{\mathbf{t}_{ij}\mathbf{n}_j^\top}{d_j}$, where $\mathbf{R}_{ij}$ and $\mathbf{t}_{ij}$ are, respectively, the rotation and the translation between the frames $\mathcal{F}_i$ and $\mathcal{F}_j$, $\mathbf{n}_j$ is the normal to the plane π expressed in the frame $\mathcal{F}_j$ and d_j is the distance of the plane π from the origin of the frame $\mathcal{F}_j$. The matrix $\mathbf{H}_{ij}$ can be estimated from $\mathbf{G}_{ij}$ if we know the camera internal parameters of the two cameras: $\mathbf{H}_{ij} \propto \mathbf{K}_i^{-1}\mathbf{G}_{ij}\mathbf{K}_j$. Three important properties of the E-collineation matrix will be extended to the multiview geometry in the next section. E-collineation matrices are *not* defined up to a scale factor. If the E-collineation is multiplied by a scalar γ ($\mathbf{H}' = \gamma\mathbf{H}$), this scalar can be easily recovered. If $\mathrm{svd}(\mathbf{H}') = (\sigma_1, \sigma_2, \sigma_3)$ are the singular values of $\mathbf{H}'$ in decreasing order, $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$, then γ is the median singular value of $\mathbf{H}'$: $\gamma = \mathrm{median}(\mathrm{svd}(\mathbf{H}')) = \sigma_2$. Indeed, the matrix $\mathbf{H}$ has a unit singular value [26] and this property can be used to normalize the E-collineation matrix. It is easy to show [12] that the E-collineation matrix satisfies the following equation $\forall k > 0$ (where $[\mathbf{n}_i]_\times$ and $[\mathbf{n}_j]_\times$ are the skew symmetric matrices associated with vectors $\mathbf{n}_i$ and $\mathbf{n}_j$ which represent the normal to the plane expressed respectively in the image frame $\mathcal{F}_i$ and $\mathcal{F}_j$):

$$[\mathbf{n}_i]_\times^k \mathbf{H}_{ji}^T = \mathbf{H}_{ij}[\mathbf{n}_j]_\times^k. \tag{2}$$

Equation (2) provides useful constraints. If $k = 1$, the matrix $[\mathbf{n}_i]_\times \mathbf{H}_{ji}^T = [\mathbf{n}_i]_\times \mathbf{R}_{ij}$ has similar properties to the essential matrix (i.e., $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$). Indeed, this matrix has two equal singular values and one equal to zero. This means two constraints each E-collineation on the camera internal parameters [7] which can be used for the self-calibration as in [15]. If $k = 2$, knowing that $[\mathbf{n}]_x^2 = \mathbf{n}\mathbf{n}^T - \mathbf{I}$, (2) can be written $\mathbf{n}_i\mathbf{n}_i^T\mathbf{H}_{ji}^T - \mathbf{H}_{ij}\mathbf{n}_j\mathbf{n}_j^T = \mathbf{H}_{ji}^T - \mathbf{H}_{ij}$ and provides equations that will be used to compute $\mathbf{n}_i$ and $\mathbf{n}_j$. A very important equation can be obtained from (2) for $k = 1$ and will be used to compute $\mathbf{n}_i$ and $\mathbf{n}_j$: $[\mathbf{n}_i]_\times = \mathbf{H}_{ij}[\mathbf{n}_j]_\times \mathbf{H}_{ij}^T$. Indeed, since $\det(\mathbf{M})\mathbf{M}[\mathbf{v}]_\times \mathbf{M}^T = [\mathbf{M}^{-\top}\mathbf{v}]_\times$ then,

$$\mathbf{n}_i = \mathbf{Q}_{ij}\mathbf{n}_j, \tag{3}$$

where $\mathbf{Q}_{ij} = \det(\mathbf{H}_{ij})\ \mathbf{H}_{ij}^{-\top}$.

3 Multiview Geometry of Planes

In this section, we describe the relationships between several views of a planar structure. Since a super matrix of 2D collineations among m views has rank 3, we will show how to enforce the rank property in an iterative procedure. The properties of the corresponding super matrix of 2D collineations provide the necessary constraint for the self-calibration of the camera internal parameters. In what follows, we will describe the case when only one planar structure is used but the extension to more than one plane is straightforward.

3.1 The Supercollineation Matrix

If m images of an unknown planar structure are available, it is possible to compute $m(m-1)$ collineations (m collineations are always equal to the identity matrix). Let us define the supercollineation matrix as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \cdots & \mathbf{G}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{m1} & \cdots & \mathbf{G}_{mm} \end{bmatrix} \tag{4}$$

with $\dim(\mathbf{G}) = (3m, 3m)$ and $\mathrm{rank}(\mathbf{G}) = 3$. The rank of $\mathbf{G}$ cannot be less than three since $\mathbf{G}_{ii} = \mathbf{I}_3$ $i \in \{1, 2, 3, \ldots, m\}$ and cannot be more than three since each row of the matrix can be obtained from a linear combination of three others rows:

$$\mathbf{G}_{ij} = \mathbf{G}_{ik}\mathbf{G}_{kj}\ \forall i, j, k \in \{1, 2, 3, \ldots, m\}.$$

This is a very strong constraint which is generally never imposed. Indeed, bundle adjustment would require a complex nonlinear minimization algorithm over all the images. In order to impose the constraints, Triggs proposed, in an extended version of [23], a one-step collineation factorization method analogous to factorization-based projective structure and motion [21]. It consists in using the SVD decomposition in order to factorize the supercollineation matrix as follows:

$$\mathbf{G} = (\mathbf{G}_{10}^\top, \mathbf{G}_{20}^\top, \cdots, \mathbf{G}_{m0}^\top)^\top (\mathbf{G}_{01}, \mathbf{G}_{02}, \cdots, \mathbf{G}_{0m}),$$

where $\mathbf{G}_{k0}$ is the collineation matrix between image $\mathcal{I}_k$ and a reference frame $\mathcal{F}_0$ attached to the plane. The factorization proposed by Triggs has not been used in [23] since it gives worse results in practice. Thus, we propose to impose the constraints with an iterative but efficient algorithm. The constraints on all collineation matrices can be summarized by the following equation:

$$\mathbf{G}^2 = m\,\mathbf{G}. \tag{5}$$

Then, matrix $\mathbf{G}$ has three nonzero equal eigenvalues $\lambda_1 = \lambda_2 = \lambda_3 = m$ and $3(m-1)$ null eigenvalues $\lambda_4 = \lambda_5 = \ldots = \lambda_{3m} = 0$. If we can impose the constraint $\mathbf{G}^2 = m\,\mathbf{G}$ (with $\mathbf{G}_{ii} = \mathbf{I}_3$ $i = 1, 2, 3, \ldots, m$) then this is, in fact, equivalent to imposing the constraints $\mathbf{G}_{ij} = \mathbf{G}_{ik}\mathbf{G}_{kj}$. In projective space, collineations matrices are defined up to a scalar factor. Thus, $\tilde{\mathbf{G}}_{ij} = \lambda_{ij}\mathbf{G}_{ij}$ and $\mathbf{G}_{ij}$ represents the same collineation. As a consequence, supercollineation matrices are defined up to a diagonal similarity (i.e., $\tilde{\mathbf{G}} = \mathbf{D}\mathbf{G}\mathbf{D}^{-1}$ and $\mathbf{G}$ represents the same supercollineation, where $\mathbf{D}$ is a diagonal matrix). The supercollineation matrix can be normalized by choosing

$$\lambda_{ij} = \sqrt[3]{\det(\tilde{\mathbf{G}}_{ij})}.$$

As a consequence,

$$\det(\mathbf{G}_{ij}) = \det\left(\frac{1}{\lambda_{ij}}\tilde{\mathbf{G}}_{ij}\right) = 1.$$

3.2 Imposing the Constraints

In order to impose the constraints, we exploit the properties of the supercollineation matrix. Let $\mathbf{p}_{ij}$ be the jth point ($j = \{1, 2, 3, \ldots, n\}$) of the ith image ($i = \{1, 2, 3, \ldots, m\}$). The jth point in all the images can be represented by the vector of dimension $(3m, 1)$ (which we will call a superpoint): $\mathbf{p}_j = (\mathbf{p}_{1j}, \mathbf{p}_{2j}, \cdots, \mathbf{p}_{mj})$. Generalizing (1), we obtain:

$$\boldsymbol{\Gamma}_j\mathbf{p}_j = \mathbf{G}\mathbf{p}_j, \tag{6}$$

where $\boldsymbol{\Gamma}_j = \mathrm{diag}(\gamma_{1j}\mathbf{I}_3, \gamma_{2j}\mathbf{I}_3, \ldots, \gamma_{mj}\mathbf{I}_3)$ is a diagonal matrix relative to the set of points j. Then, multiplying both sides of (6) by $\mathbf{G}$, we have: $\mathbf{G}\boldsymbol{\Gamma}_j\mathbf{p}_j = \mathbf{G}^2\mathbf{p}_j = m\mathbf{G}\mathbf{p}_j = m\boldsymbol{\Gamma}_j\mathbf{p}_j$. The vector $\boldsymbol{\Gamma}_j\mathbf{p}_j$ (representing the homogeneous coordinates of the point j in all the images) is an eigenvector of $\mathbf{G}$ corresponding to the eigenvalue m. As a consequence, any vector $\boldsymbol{\Gamma}_j\mathbf{p}_j$ can be obtained as a linear combination of the eigenvectors of $\mathbf{G}$ corresponding to the eigenvalue $\lambda = m$: $\boldsymbol{\Gamma}_j\mathbf{p}_j = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3$. The matrix $\mathbf{G}$ can always be diagonalized and, thus, three linearly independent eigenvectors always exist, i.e., $\exists\mathbf{X} : \mathbf{X}^{-1}\mathbf{G}\mathbf{X} = diag(\lambda_1, \lambda_2, \ldots, \lambda_{3m})$. The columns of the matrix $\mathbf{X}$ are, in fact, eigenvectors of $\mathbf{G}$. Since $\mathbf{X}$ is nonsingular, the eigenvectors of $\mathbf{G}$ are linearly independent and span the space $\mathbb{R}^{3m}$. That means that an initial estimation $\widehat{\mathbf{p}} \in \mathbb{R}^{3m}$ of the superpoint $\mathbf{p}$ can be written as $\boldsymbol{\Gamma}\widehat{\mathbf{p}} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3 + \ldots + \alpha_{3m}\mathbf{x}_{3m}$. The real vector $\mathbf{\Gamma p}$ is an eigenvector of $\mathbf{G}$ corresponding to the largest eigenvalue $\lambda = m$. If we have a perfect estimate of $\mathbf{G}$, we can use a well-known algorithm to find an eigenvector of $\mathbf{G}$ starting from $\mathbf{\Gamma}\widehat{\mathbf{p}}$. In practice, the real supercollineation matrix $\mathbf{G}$ is unknown and we must use an approximation $\widehat{\mathbf{G}}$ estimated from the points measured in the images. Thus, the algorithm we use is similar to the previous one but the approximation $\widehat{\mathbf{G}}$ is updated at each iteration. We start with a set of n points $\widehat{\mathbf{p}}_j$ $(j = 1, 2, 3, \ldots, n)$ and compute the supercollineation matrix $\widehat{\mathbf{G}}$ solving independently the linear problem of estimating each block $\widehat{\mathbf{G}}_{ij}$ from (1). It is not necessary that all the points are visible in all the images. Then, we compute a new set of superpoints which is a better estimate of the true image points thanks to reprojections averaging. The better is the initial estimate of $\mathbf{G}$ the faster the algorithm converges and the more accurate is the result. Even if we do not formally prove here that the algorithm do not diverge, we did not observe any large drift in the experiments. Furthermore, we compared our method with bundle adjustment. Experiments on simulated data showed that the results given by our algorithm are also a solution of the bundle adjustment problem. Indeed, the algorithm produces consistent points and homographies exactly as a bundle adjustement. A detailed desccription of the algorithm and of the comparison with bundle adjustment can be found in [12].

3.3 The Super-E-Collineation Matrix

Let us define the super-E-collineation matrix in the Euclidean space as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{m1} & \cdots & \mathbf{H}_{mm} \end{bmatrix} \tag{7}$$

with $\dim(\mathbf{H}) = (3m, 3m)$ and $\mathrm{rank}(\mathbf{H}) = 3$. The super-E-collineation matrix can be obtained from the supercollineation matrix and the camera parameters: $\mathbf{H} = \mathbf{K}^{-1}\mathbf{G}\mathbf{K}$, where $(\dim(\mathbf{K}) = (3m, 3m)$ and $\mathrm{rank}(\mathbf{K}) = 3m)$ and $\mathbf{K} = \mathrm{diag}(\mathbf{K}_1, \mathbf{K}_2, \cdots, \mathbf{K}_n)$ is the block diagonal matrix containing the internal parameters of all the cameras. It should be noticed that if the constraint $\mathbf{G}^2 = m\mathbf{G}$ was imposed, then the constraint $\mathbf{H}^2 = m\mathbf{H}$ is automatically imposed which means that the following constraints are satisfied: $\mathbf{H}_{ij} = \mathbf{H}_{ik}\mathbf{H}_{kj}$. Unlike the supercollineation matrix, the super-E-collineation matrix is *not* defined up to a diagonal similarity. Indeed, each E-collineation matrix must have the median singular value equal to 1. If σ_{ij} (which is generally different from 1) denotes the median singular value of the estimated matrix $\tilde{\mathbf{H}}_{ij}$ we can build the following matrix which contain all the coefficients of normalization: $\mathbf{D} = \mathrm{diag}(\sigma_{11}\mathbf{I}_3, \sigma_{12}\mathbf{I}_3, \ldots, \sigma_{1m}\mathbf{I}_3)$. The super-E-collineation matrix is thus normalized as follows: $\mathbf{H} = \mathbf{D}\tilde{\mathbf{H}}\mathbf{D}^{-1}$. After this, normalization all the E-collineation matrices $\mathbf{H}_{ij}$ in $\mathbf{H}$ have their median singular value equal to 1. From this equation, we can easily see that the constraint $\mathbf{H}^2 = m\mathbf{H}$ holds. Even in the presence of noise, normalizing $\mathbf{H}$ will conserve the rank constraint of the matrix since it is a similarity transformation.

3.4 Decomposition of the Super-E-Collineation Matrix

To our knowledge, existing methods for the decomposition of E-collineations (calibrated homographies) are given in [2], [26] for image pairs. In our case, the super-E-collineation matrix cannot be decomposed using these methods. Thus, we propose a method which take into account multiple views of a plane without using a key image. After normalization, the E-collineation matrix can be decomposed as: $\mathbf{H} = \mathbf{R} + \mathbf{T}\mathbf{n}^T$, where $\mathbf{R}$ is a $(3m, 3m)$ symmetric matrix with $\mathrm{rank}(\mathbf{R}) = 3$ and such that $\mathbf{R}^2 = m\mathbf{R}$. As a consequence, not only are the three largest eigenvalues $\lambda_1 = \lambda_2 = \lambda_3 = m$, but also the three largest singular values are $\sigma_1 = \sigma_2 = \sigma_3 = m$. The $(3m, m)$ matrix $\mathbf{t}$ contains the translations, while the $(3m, m)$ matrix $\mathbf{N}$ contains the normals to the plane. As already mentioned, in [2] and [26] are presented two different methods for decomposing the E-collineation matrix, computed from two views of a planar structure. In general, there are two possible solutions but the ambiguity can be resolved by adding more images. In our case, the normals to the plane can be extracted from the super-E-collineation matrix. Setting $\mathbf{Q} = \mathbf{W}\mathbf{H}^T\mathbf{W}^{-1}$, where

$$\mathbf{W} = \mathrm{diag}(\mathbf{I}_3, \det(\mathbf{H}_{21})\mathbf{I}_3, \ldots, \det(\mathbf{H}_{m1})\mathbf{I}_3)$$

and $\dim(\mathbf{Q}) = (3m, 3m)$ and $\mathrm{rank}(\mathbf{Q}) = 3$. Matrix $\mathbf{Q}$ has similar properties to the matrix $\mathbf{H}$, for example, it has an eigenvalue $\lambda = m$ of multiplicity three. The vector $\mathbf{n}$ is an eigenvector of $\mathbf{Q}$ corresponding to the eigenvalue $\lambda = m$:

$$\mathbf{Q}\mathbf{n} = m\mathbf{n}, \tag{8}$$

where $\mathbf{n} = [\mathbf{n}_1^T \mathbf{n}_2^T \ldots \mathbf{n}_m^T]^T$. The vector can be written as a linear combination of the eigenvectors $\mathbf{n} = x\,\mathbf{v}_1 + y\,\mathbf{v}_2 + z\,\mathbf{v}_3 = \mathbf{V}\mathbf{w}$, where $\mathbf{w} = (x, y, z)$ is a vector containing three unknowns and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is a known matrix. Imposing the constraint $\|\mathbf{n}_k\| = 1$, we obtain: $\mathbf{V}_i\mathbf{w}\mathbf{w}^T\mathbf{V}_i^T\mathbf{H}_{ji}^T - \mathbf{H}_{ij}\mathbf{V}_j\mathbf{w}\mathbf{w}^T\mathbf{V}_j^T = \mathbf{H}_{ji}^T - \mathbf{H}_{ij}$ from which it is possible to compute the unknown matrix $\mathbf{w}\mathbf{w}^T$ and then, by singular values decomposition, the original unknown which is $\mathbf{w}$. Once $\mathbf{w}$ has been found, the normals to the plane are extracted from $\mathbf{H}$ and knowing that $\mathbf{R}\mathbf{N} = \mathbf{N}\mathbf{U}_m$ (where $\mathbf{U}_m$ is a $m \times m$ matrix full of ones), we find: $\mathbf{T} = \mathbf{H}\mathbf{N} - \mathbf{U}_{3m}\mathbf{N}$ and $\mathbf{R} = \mathbf{H}(\mathbf{N}\mathbf{N}^T - \mathbf{I}_{3m}) + \mathbf{N}^T\mathbf{U}_{3m}\mathbf{N}$.

4 Camera Self-Calibration

The super-E-collineation can of course be used in many applications. In this section, we use the properties of the set of E-collineation matrices to self-calibrate the cameras. It should be noticed that we avoid the use of a bundle adjustment technique to impose the rank 3 constraint on the super-E-collineation (as explained in Section 3) and, thus, we considerably simplify the algorithm. In this case, the only unknowns are the camera internal parameters. Each independent E-collineation will provide us two constraints on the parameters according to (2). Indeed, if σ_{ij}^I and σ_{ij}^{II} are the two nonzero singular values of the matrix $[\mathbf{n}_i]_\times\mathbf{H}_{ji}^T$ our self-calibration method is based on the minimization of the following cost function proposed in [4] and [15]:

$$\mathcal{C} = \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\sigma_{ij}^I - \sigma_{ij}^{II}}{\sigma_{ij}^I}. \tag{9}$$

Even if our cost funtion is based on the equality of two singular values, it must be noticed that the singular values used in [4] and [15] come from the essential matrix while the singular values we use to compute $\mathcal{C}$ come from matrix $[\mathbf{n}_i]_\times\mathbf{H}_{ji}^T$. It has been proven in [4] that the constraint $\sigma_{ij}^I = \sigma_{ij}^{II}$ is equivalent to impose two constraints on the camera internal parameters. Thus, each independent E-collineation can be used to compute the matrix $[\mathbf{n}_i]_\times\mathbf{H}_{ji}^T$ and will provide us two constraints on the parameters. Then, with constant camera parameters, we need a minimum of: three independent E-collineations (four images) to recover the focal length and the principal point supposing $r = k_u/k_v = 1$ and $\theta = \pi/2$; four independent E-collineations (five images) to recover all the parameters. If the camera parameters are varying (we fix $\theta = \pi/2$), we need a minimum of: three independent E-collineations (four images) to recover the four different focal lengths (supposing the ratio $r = k_u/k_v$ and the principal point approximatively known); four independent E-collineations (five images) to recover the five different focal lengths and the fixed ratio (with the principal point approximatively known); six independent E-collineation matrices (seven images) to recover the seven different focal lengths, the ratio, and the the principal point. Our self-calibration algorithm is the following:

TABLE 1
Results Using Digital Images of the Grid (Statistics on 50 Tests)

calibration method	f	r	θ	u	v
DLT linear	685 ± 3	1.0005 ± 0.0033	90.00 ± 0.14	322 ± 5	229 ± 4
Faugeras-Toscani	685 ± 3	1.0003 ± 0.0022	90.00 ± 0.16	322 ± 5	229 ± 4
right plane (10 im)	680 ± 8	0.9976 ± 0.0088	89.23 ± 0.59	318 ± 8	230 ± 8
left plane (10 im)	680 ± 6	0.9943 ± 0.0058	89.89 ± 0.30	320 ± 7	232 ± 4
right plane (8 im)	681 ± 12	0.9950 ± 0.0105	89.21 ± 0.80	315 ± 11	232 ± 9
left plane (8 im)	678 ± 12	0.9969 ± 0.0075	90.06 ± 0.24	327 ± 14	233 ± 3
right plane (6 im)	686 ± 13	0.9891 ± 0.0126	89.63 ± 0.69	312 ± 18	231 ± 11
left plane (6 im)	685 ± 10	0.9886 ± 0.0147	89.80 ± 0.59	339 ± 18	232 ± 7
Faugeras-Toscani	685 ± 3	1 ± 0	90 ± 0	322 ± 6	229 ± 4
right plane (10 im)	679 ± 6	1 ± 0	90 ± 0	318 ± 5	224 ± 8
left plane (10 im)	675 ± 6	1 ± 0	90 ± 0	325 ± 4	232 ± 4
right plane (8 im)	687 ± 6	1 ± 0	90 ± 0	323 ± 3	231 ± 6
left plane (8 im)	676 ± 4	1 ± 0	90 ± 0	343 ± 27	231 ± 12
right plane (6 im)	676 ± 8	1 ± 0	90 ± 0	314 ± 9	227 ± 5
left plane (6 im)	677 ± 18	1 ± 0	90 ± 0	327 ± 34	230 ± 31

1. Match corresponding points in m images of a planar structure.
2. Compute the supercollineation imposing the rank 3 constraint using the algorithm described in Section 3.1.
3. Using an initial guess of the camera parameters compute the normalized super-E-collineation matrix as described in Section 3.2.
4. Decompose the super-E-collineation matrix and find the normal to the plane as described in Section 3.3.
5. Compute a new set of camera parameters which minimize the cost function given in Section 3.4 and go to Step 3.

5 EXPERIMENTS

The self-calibration algorithm has been tested on real images. The results obtained with a calibration grid were compared with the standard Faugeras-Toscani method [3]. A detailed desccription of the experiments can be found in [12].

5.1 Sequence with Constant Camera Parameters

A sequence (26 images of dimension (640×480)) of a calibration grid was taken using a Fuji MX700 camera with a 7 mm lens. Table 1 gives the results for the following experiments:

- *Nonplanar calibration.* The mean and the standard deviation on 26 images of the grid calibrated with the standard Faugeras-Toscani method initialized with the DLT linear method [1].
- *Planar self-calibration*. The mean and the standard deviation on 50 tests using m images (m = 6, 8, 10) randomly chosen between the 26 images of the grid.

The same tests are repeated using the right plane alone, the left plane alone and then again with $r = k_v/k_u$ and θ fixed to nominal values. The results are very good and agree with the simulations presented in [10]. The angle of rotation between the images of the sequence can be greater than 60 degrees which has, in general, the effect to improve the results. However, this is not always true since the planes can be very close to the optical center of the camera and, in this case, the estimation of the collineations is not accurate. The calibration obtained using the right plane is very similar to the calibration obtained using the left plane. As we expected, the accuracy decreases as we decrease the number of images but the worst result (obtained using only six images of the grid) is only an error of 2 percent on the focal length. The number of iteration necessary to converge obviously depends on the initialization of the algorithm. The starting point was $f(0) = 1,000$, $u_0(0) = 250$, and $v_0(0) = 250$. The constraints on the supercollineation matrix are imposed after four iterations and the maximal number of iterations for the minimization of the cost function was set to 300.

5.2 Sequence with Variable Camera Parameters

In order to test our self-calibration technique with varying camera parameters, the ratio k_u/k_v is fixed to one and the principal point is supposed to be in the center of the image. Thus, the unknowns are the focal lengths. A new set of 10 images of the grid was taken with

TABLE 2
Self-Calibration of the Focal Lengths with a Zooming Camera

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
Faugeras-Toscani	1407	1835	1195	1492	1337	1158	985	1534	1845	1839
proposed method	1491	1950	1212	1472	1393	1233	1012	1609	1929	1905
relative error	-6 %	-6 %	-1 %	1 %	-4 %	-6 %	-3 %	-5 %	-5 %	-4 %

a zooming camera. The camera was calibrated (in order to have a ground truth) with the standard Faugeras-Toscani method using the 10 images of the sequence of the calibration grid. The obtained focal lengths are given in Table 2. The results obtained by our self-calibration algorithm using the left plane of the calibration grid (similar results have been obtained using the right plane) are summarized in Table 2. The starting focal length was $f_0 = 1,000$ for all the unknown f_i. Considering that, in our self-calibration algorithm, the principal point was supposed to be in the center of the image, the results are satisfactory, and the 3D reconstruction of the grid can be done with sufficient accuracy.

6 Conclusion

In this paper, we presented an efficient technique to impose the constraints existing within a set of collineation matrices computed from multiple views of a planar structure. The obtained set of collineations can be used for several applications such mosaicing, reconstruction and self-calibration from planes. In this paper, we focused on self-calibration proposing a new method which does not need any a priori knowledge of the metric structure of the plane. The method was tested with real images and the obtained results are very good. The method can be improved by imposing further constraints in order to obtain not only a consistent set of collineations but also a consistent set of E-collineation matrices. The method can also be improved using a probabilistic model of noise.

Acknowledgments

This work was supported by an EC (ESPRIT) grant no. LTR26247 (VIGOR). The authors would like to thank Paulo Mendonca and Kenneth Wong for their help.

References

[1] Y.I. Abdel-Aziz and H.M. Karara, "Direct Linear Transformation into Object Space Coordinates in Close-Range Photogrametry," *Proc. Symp. Close-Range Photogrametry*, pp. 1-18, 1971.

[2] O. Faugeras and F. Lustman, "Motion and Structure from Motion in a Piecewise Planar Environment," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 2, no. 3, pp. 485-508, 1988.

[3] O. Faugeras and G. Toscani, "The Calibration Problem for Stereo," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 15-20, June 1986.

[4] R. Hartley, "Estimation of Relative Camera Positions for Uncalibrated Cameras," *Proc. European Conf. Computer Vision*, G. Sandini, ed., vol. 588, pp. 579-587, May 1992.

[5] R. Hartley, "Euclidean Reconstruction from Uncalibrated Views," *Application of Invariance in Computer Vision*, D.F.J.L. Mundy and A. Zissermann, eds., vol. 825, pp. 237-256, 1994.

[6] R. Hartley, "Self-Calibration from Multiple Views with a Rotating Camera," *Proc. European Conf. Computer Vision*, pp. 471-478, May 1994.

[7] R. Hartley, "Minimising Algebraic Error in Geometric Estimation Problem," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 469-476, 1998.

[8] A. Heyden and K. Astrom, "Euclidean Reconstruction from Constant Intrinsic Parameters," *Proc. 13th Int'l Conf. Pattern Recognition*, pp. 339-343, 1996.

[9] A. Heyden, R. Berthilsson, and G. Sparr, "An Iterative Factorization Method for Projective Structure and Motion from Image Sequences," *Image and Vision Computing*, vol. 17, no. 13, pp. 981-991, 1999

[10] E. Malis and R. Cipolla, "Multiview Constraints between Collineations: Application to Self-Calibration from Unknown Planar Structures," *Proc. European Conf. Computer Vision*, vol. 2, pp. 610-624, 2000.

[11] E. Malis and R. Cipolla, "Self-Calibration of Zooming Cameras Observing an Unknown Planar Structure," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 85-88, Sept. 2000.

[12] E. Malis and R. Cipolla, "Camera Self-Calibration from Unknown Planar Structures Enforcing Multi-View Constraints between Collineations," technical report, Univ. of Cambridge, Dec. 2001.

[13] S. Maybank and O. Faugeras, "A Theory of Self-Calibration of a Moving Camera," *Int'l J. Computer Vision*, vol. 8, no. 2, pp. 123-151, 1992.

[14] J. Mendelsohn and K. Daniilidis, "Constrained Self-Calibration," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 581-587, 1999.

[15] P. Mendonca and R. Cipolla, "A Simple Techinique for Self-Calibration," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 500-505, 1999.

[16] T. Moons, L. Van Gool, M. Proesmans, and E. Pauwels, "Affine Reconstruction from Perspective Image Pairs with a Relative Object-Camera Translation in Between," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 77-83, Jan. 1996.

[17] M. Pollefeys, R. Koch, and L. VanGool, "Self-Calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters," *Int'l J. Computer Vision*, vol. 32, no. 1, pp. 7-25, Aug. 1999.

[18] M. Pollefeys and L. VanGool, "Stratified Self-Calibration with the Modulus Constraint," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 707-724, Aug. 1999.

[19] P. Sturm and S. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 432-437, 1999.

[20] R. Szeliski and P. Torr, "Geometrically Constrained Structure from Motion: Points on Planes," *Proc. European Workshop 3D Structure from Multiple Images of Large-Scale Environments (SMILE)*, pp. 171-186, June 1998.

[21] B. Triggs, "Factorisation Methods for Projective Structure and Motion," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 845-851, 1996.

[22] B. Triggs, "Autocalibration and the Absolute Quadric," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 609-614, 1997.

[23] B. Triggs, "Autocalibration from Planar Scenes," *Proc. European Conf. Computer Vision*, pp. 89-105, 1998.

[24] L. Zelnik-Manor and M. Irani, "Multiview Subspace Constraints on Homographies," *Proc. IEEE Int'l Conf. Computer Vision*, vol. 1, pp. 710-715, Sept. 1999.

[25] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.

[26] Z. Zhang and A.R. Hanson, "Scaled Euclidean 3D Reconstruction Based on Externally Uncalibrated Cameras," *Proc. IEEE Symp. Computer Vision*, pp. 37-42, 1995.