Finding the Collineation Between two Projective Reconstructions

Gabriella Csurka, Radu Horaud

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Finding the Collineation Between two Projective Reconstructions

Gabriella Csurka and Radu Horaud

No 3468

Août 1998

THÈME 3

ISSN 0249-6399

INRIA
RHÔNE-ALPES

Finding the Collineation Between two Projective Reconstructions

Gabriella Csurka and Radu Horaud

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Movi

Rapport de recherche n° 3468 — Août 1998 — 20 pages

Abstract: The problem of finding the collineation between two 3-D projective reconstructions has been proved to be useful for a variety of tasks such as calibration of a stereo rig and 3-D affine and/or Euclidean reconstruction. Moreover such a collineation may well be viewed as a point transfer method between two image pairs with applications to visually guided robot control. In spite of this potential, methods for properly estimating such a projective transformation have received little attention in the past. In this paper we describe linear, non-linear and robust methods for estimating this transformation. We test the numerical stability of these methods with respect to image noise and to the number of matched points. Finally we briefly describe three applications: stereo image transfer, Euclidean reconstruction, and self calibration of a stereoscopic camera pair.

Key-words: projective reconstruction, 3-D projective transformation, image transfer, projective to Euclidean mapping

(Résumé : tsvp)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Calcul d'une homographie entre deux reconstructions projectives

Résumé : Le calcul d'une homographie entre deux reconstructions projectives tri-dimensionnelles s'est avéré être utile dans un certain nombre de tâches telles que la calibration d'un système stéréo ainsi que pour al reconstruction affine et/ou euclidienne. De plus une telle homographie peut être utile pour transférer des points entre deux paires d'images. Ce transfert est utile pour l'asservissement visuel d'un robot. En dépit de ces possibilités, peu de travaux s'intéressent au calcul d'une telle homographie. Dans cet article nous décrivons des méthodes linéaire, non-linéaire et robuste pour estimer cette transformation. Nous testons la stabilité des calculs par rapport au bruit et au nombre de points mis en correspondance entre les deux paires d'images. Enfin, on décrit brièvement trois applications : transfert d'images, reconstruction euclidienne et auto-calibration d'une tête stéréo.

Mots-clé : reconstruction projective, transformation projective 3-D, transfert d'images, transformation projective/euclidienne

1 Introduction and motivation

Until very recently it was believed that visual tasks require some form of off-line camera calibration. For example, a moving calibrated camera provides a sequence of images from which 3-D Euclidean structure can in principle be recovered. More recently, it has been shown both theoretically and experimentally that an image sequence taken with an uncalibrated camera can provide 3-D Euclidean structure as well [9]. The basic paradigm consists of recovering projective structure first and then upgrading it to affine structure and then to Euclidean structure.

Another possibility that is being investigated by a number of researchers is to consider a sequence of image pairs gathered with a moving uncalibrated stereo rig [1, 14, 2, 6]. With such an uncalibrated image pair, matched points are reconstructed in projective space [3]. If the stereo pair undergoes a rigid motion, the two projective reconstructions (before and after the motion) are related by a 3-D projective transformation which is conveniently represented by a 4×4 regular matrix – a collineation. Therefore methods that attempt to recover projective, affine or Euclidean structure from a moving stereo pair need to estimate this collineation. Interesting enough, with the exception of [1] which briefly outlines a method for estimating this projective transformation, it appears to be no published paper in the computer vision literature describing and evaluating methods for estimating this collineation from stereoscopic data.

In this paper we address the problem of estimating such a collineation from two projective reconstructions obtained from two different image pairs. First, we describe two linear methods which minimize some algebraic distance defined in 3-D projective space and a non-linear method which minimizes an image-space Euclidean distance. Moreover, we outline a robust method that can deal with badly matched points between the two image pairs.

Second, we study the behavior of these methods (linear and non-linear) in the presence of image noise and as a function of the number of matched points. In the light of these experiments it appears that the non-linear methods performs slightly better than the linear methods. However for levels of noise below one pixel in magnitude, the linear methods performs as well as the non-linear ones and hence they should be preferred because they are less time consuming than the non-linear methods.

Third, we describe three possible applications: (i) stereo image transfer which shows that for such tasks as visual servoing, projective structure is useful in its own right, (ii) Euclidean upgrade using 3-D control points and an error function defined in 4-D vector space, and (iii) stereo self-calibration which takes advantage of a special parameterization of the 3-D projective transformation.

2 3-D projective collineations

Given a pair of images and the fundamental matrix $\mathbf{F}_x$ that describe their epipolar geometry, there exists a projective basis of the projective space $\mathcal{P}^3$ such that the projections from this

space onto the images are represented by[1] two 3×4 matrices [4, 10]:

$$\mathbf{P}_x = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{pmatrix} \qquad \text{and} \qquad \mathbf{P}'_x = \begin{pmatrix} \mathbf{G}'_x & \boldsymbol{e}'_x \end{pmatrix} \tag{1}$$

were $\mathbf{G}'_x$ is a 3×3 regular matrix and the epipole $\boldsymbol{e}'_x$ is the projection of the first optical center onto the second image.

It was shown in [3, 8] that we have a four degrees of freedom family of projective bases in which $\mathbf{P}_x$ has the form $\begin{pmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{pmatrix}$. However, when we fix $\mathbf{P}'_x$, the projective basis is uniquely determined.

If $\boldsymbol{x}, \boldsymbol{x}'$ are the images of a 3-dimensional point M, from $\rho_x \boldsymbol{x} = \mathbf{P}_x \boldsymbol{X}$, and $\rho'_x \boldsymbol{x}' = \mathbf{P}'_x \boldsymbol{X}$ results the projective coordinates $\boldsymbol{X}$ of point M in the projective basis $\mathcal{B}_x$ defined by $(\mathbf{P}_x, \mathbf{P}'_x)$ [10].

Consider an other pair of images of the same scene and note by $\mathbf{F}_y$ the corresponding fundamental matrix. The new projective matrixes are

$$\mathbf{P}_y = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{pmatrix} \qquad \text{and} \qquad \mathbf{P}'_y = \begin{pmatrix} \mathbf{G}'_y & \boldsymbol{e}'_y \end{pmatrix}$$

and if $\boldsymbol{y}, \boldsymbol{y}'$ are the new projections of the point M, its homogeneous representation in the basis $\mathcal{B}_y$ is $\boldsymbol{Y}$ that verifies $\rho_y \boldsymbol{y} = \mathbf{P}_y \boldsymbol{Y}$ and $\rho'_y \boldsymbol{y}' = \mathbf{P}'_y \boldsymbol{Y}$.

Assume now, that we have a set of m points, and let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_m$ and $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_m$ be their homogeneous representations in the projective basis $\mathcal{B}_x$ respectively $\mathcal{B}_y$. Then, there exists a 4×4 collineation matrix $\mathbf{H}$ that maps the points $\boldsymbol{X}_i$ to the points $\boldsymbol{Y}_i$:

$$\mu_i \boldsymbol{Y}_i = \mathbf{H} \boldsymbol{X}_i \tag{2}$$

with μ_i arbitrary non-null scalars.

$\mathbf{H}$, defined up to a scale factor, has 15 degrees of freedom so to determine it, we need at least five point correspondences since each pair $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ gives us 3 constraints after eliminating μ_i.

Indeed, five points in the projective space generally define a projective basis of $\mathcal{P}^3$ such that the collineation between the standard projective basis and the set of points $\{\boldsymbol{X}_i\}_{i=1..5}$ is given by the matrix:

$$\mathbf{T}_1 = \begin{pmatrix} \mu_1 \boldsymbol{X}_1 & \mu_2 \boldsymbol{X}_2 & \mu_3 \boldsymbol{X}_3 & \mu_4 \boldsymbol{X}_4 \end{pmatrix}$$

where $(\mu_1, \mu_2, \mu_3, \mu_4)^\top = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4)^{-1} \boldsymbol{X}_5$. The collineation $\mathbf{T}_2$ which maps the standard reference points to the second set of points $\{\boldsymbol{Y}_i\}_{i=1..5}$ can be obtained similarly. Finally, the collineation between $\{\boldsymbol{X}_i\}_{i=1..5}$ and $\{\boldsymbol{Y}_i\}_{i=1..5}$ is then:

$$\mathbf{H} = \mathbf{T}_2 \mathbf{T}_1^{-1}.$$

This method is quite simple, but unstable in the presence of noise since we only use the minimum number of points to compute the collineation matrix $\mathbf{H}$. If we have $m > 5$ pair of points, it is better to use all of them and estimate $\mathbf{H}$ by some linear or non-linear methods.

[1] We denote by $\mathbf{I}_n$ the $n \times n$ matrix of identity and by $\mathbf{0}_n$ the n-vector containing n zeros.

2.1 Linear method 1

The classical way to estimate the entries of $\mathbf{H}$ is to eliminate the scale factors μ_i. A homogeneous linear system in the entries of $\mathbf{H}$ is thus obtained [12]:

$$\begin{aligned} Y_i^{(4)}V_i^{(1)} - Y_i^{(1)}V_i^{(4)} &= 0 \\ Y_i^{(4)}V_i^{(2)} - Y_i^{(2)}V_i^{(4)} &= 0 \\ Y_i^{(4)}V_i^{(3)} - Y_i^{(3)}V_i^{(4)} &= 0 \end{aligned} \tag{3}$$

where we used the notation $\boldsymbol{V}_i = (V_i^{(1)}, V_i^{(2)}, V_i^{(3)}, V_i^{(4)})^\top$ for the vector $\mathbf{H}\boldsymbol{X}_i$:

$$V_i^{(j)} = H_{j1}X_i^{(1)} + H_{j2}X_i^{(2)} + H_{j3}X_i^{(3)} + H_{j4}X_i^{(4)} \tag{4}$$

In the affine or Euclidean cases we have $Y_i^{(4)} = 1$ which is often at a different magnitude than $Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}$. This explains the choice of these three equations where $Y_i^{(4)}$ has the same role. In the projective case, we cannot say the same thing, so if we want that all four coordinates play the same role, we can add the three other possible equations (which are not independent from the previous ones):

$$\begin{aligned} Y_i^{(2)}V_i^{(1)} - Y_i^{(1)}V_i^{(2)} &= 0 \\ Y_i^{(3)}V_i^{(1)} - Y_i^{(1)}V_i^{(3)} &= 0 \\ Y_i^{(3)}V_i^{(2)} - Y_i^{(2)}V_i^{(3)} &= 0 \end{aligned} \tag{5}$$

This system can be solved when $m \geq 5$ point correspondences are available and with an additional constraint such as $\sum H_{ij}^2 = 1$.

If we denote $\boldsymbol{h} = (H_{11}, H_{12}, \ldots, H_{44})^\top$, the equations (3) and (5) can be written in the form[2]:

$$\mathbf{B}_i\, \boldsymbol{h} = \mathbf{0},$$

where $\mathbf{B}_i$ is a 6×16 matrix:

$$\mathbf{B}_i = \begin{pmatrix} Y_4^i\boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & Y_i^{(1)}\boldsymbol{X}_i^\top \\ \mathbf{0}_4^\top & Y_i^{(4)}\boldsymbol{X}_i^\top & \mathbf{0}_4^\top & Y_i^{(2)}\boldsymbol{X}_i^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & Y_i^{(4)}\boldsymbol{X}_i^\top & Y_i^{(3)}\boldsymbol{X}_i^\top \\ Y_i^{(2)}\boldsymbol{X}_i^\top & Y_i^{(1)}\boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ Y_i^{(3)}\boldsymbol{X}_i^\top & \mathbf{0}_4^\top & Y_i^{(1)}\boldsymbol{X}_i^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & Y_i^{(3)}\boldsymbol{X}_i^\top & Y_i^{(2)}\boldsymbol{X}_i^\top & \mathbf{0}_4^\top \end{pmatrix}.$$

[2] If we use only three equations one can remove the rows corresponding to the other equations and obtain $\mathbf{B}_i$ as a 3×16 matrix.

Given m correspondences $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, we are looking for $\boldsymbol{h}$ that minimizes the following error function:

$$\sum_{i=1}^{m} \|\mathbf{B}_i \boldsymbol{h}\|^2 = \boldsymbol{h}^\top \left(\sum_{i=1}^{m} \mathbf{B}_i^\top \mathbf{B}_i \right) \boldsymbol{h}. \tag{6}$$

The matrix $\mathbf{B} = \sum_{i=1}^{n} \mathbf{B}_i^\top \mathbf{B}_i$ being symmetric semi-definite and positive, the solution for $\boldsymbol{h}$ is the eigenvector of $\mathbf{B}$ corresponding to the smallest eigenvalue of $\mathbf{B}$. In practice, this eigenvector can be computed using the SVD (singular value decomposition) algorithm.

2.2 Linear method 2

An alternative solution is to estimate simultaneously the entries of $\mathbf{H}$ and the scale factors $\mu_1, \ldots \mu_m$. The equation (2) can be decomposed into four distinct linear constraints and, for example, the first of these linear constraints can be written as:

$$H_{11} X_i^{(1)} + H_{12} X_i^{(2)} + H_{13} X_i^{(3)} + H_{14} X_i^{(4)} - \mu_i Y_i^{(1)} = 0$$

Without loss of generality, we fix one of the scale factors: $\mu_m = 1$. Therefore, we have 16 unknowns for the entries of $\mathbf{H}$ and $m-1$ unknown scale factors. The m equations (2) can be written as a linear system $\mathbf{C}\boldsymbol{s} = \boldsymbol{r}$ with $\boldsymbol{s} = (H_{11}, \ldots, H_{44}, \mu_1, \ldots, \mu_{m-1})^\top$, $\boldsymbol{r} = (\underbrace{0, \ldots, 0}_{4\times(m-1)}, Y_m^{(1)}, Y_m^{(2)}, Y_m^{(3)}, Y_m^{(4)})^\top$, and:

$$\mathbf{C} = \begin{pmatrix} \mathbf{E}_1 & -\boldsymbol{Y}_1 & \mathbf{0}_4 & \ldots & \mathbf{0}_4 \\ \mathbf{E}_2 & \mathbf{0}_4 & -\boldsymbol{Y}_2 & \ldots & \mathbf{0}_4 \\ \vdots & & & & \vdots \\ \mathbf{E}_{m-1} & \mathbf{0}_4 & \mathbf{0}_4 & \ldots & -\boldsymbol{Y}_{m-1} \\ \mathbf{E}_m & \mathbf{0}_4 & \mathbf{0}_4 & \ldots & \mathbf{0}_4 \end{pmatrix}$$

The 4×16 matrices $\mathbf{E}_i$ are defined by:

$$\mathbf{E}_i = \begin{pmatrix} \boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & \boldsymbol{X}_i^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \boldsymbol{X}_i^\top \end{pmatrix}$$

This linear system consists of $4m$ equations. Since there are $16+(m-1) = 15+m$ unknowns, we must have $m \geq 5$. This linear system can be solved using the standard pseudo-inverse technique:

$$\boldsymbol{s} = (\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{C}^\top \boldsymbol{r}$$

provided that the 3-D points are not coplanar.

2.3 Non-linear method

One way to assess the quality of the estimated collineation $\widehat{\mathbf{H}}$ is to compare the projections of $\widehat{\boldsymbol{Y}}_i = \widehat{\mathbf{H}}\boldsymbol{X}_i$ and of $\widehat{\boldsymbol{X}}_i = \widehat{\mathbf{H}}^{-1}\boldsymbol{Y}_i$ with the true image points. Let $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$ be the true image points (in the left and right images) from which the 3-D point $\boldsymbol{X}_i$ was reconstructed, and let $\mathbf{P}_x$ and $\mathbf{P}'_x$ be the corresponding projection matrixes (see (1)) The projections of $\widehat{\boldsymbol{X}}_i$ by $\mathbf{P}_x$ and $\mathbf{P}'_x$ are noted by $\widehat{\boldsymbol{x}}_i$ and $\widehat{\boldsymbol{x}}'_i$. Similarly we define $\boldsymbol{y}_i$, $\boldsymbol{y}'_i$, $\mathbf{P}_y$ and $\mathbf{P}'_y$, $\widehat{\boldsymbol{y}}_i$ and $\widehat{\boldsymbol{y}}'_i$:

$$\begin{aligned}
\widehat{\boldsymbol{x}}_i &= \mathbf{P}_x\widehat{\mathbf{H}}^{-1}\boldsymbol{Y}_i \\
\widehat{\boldsymbol{x}}'_i &= \mathbf{P}'_x\widehat{\mathbf{H}}^{-1}\boldsymbol{Y}_i \\
\widehat{\boldsymbol{y}}_i &= \mathbf{P}_y\widehat{\mathbf{H}}\boldsymbol{X}_i \\
\widehat{\boldsymbol{y}}'_i &= \mathbf{P}'_y\widehat{\mathbf{H}}\boldsymbol{X}_i
\end{aligned}$$

With the notation $\boldsymbol{y}^\top = (\overline{\boldsymbol{y}}^\top \ 1)$, let $d(\overline{\boldsymbol{y}}, \hat{\overline{\boldsymbol{y}}})$ denote the Euclidean distance between the image points $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$. The quality of the collineation is assessed by the following quadratic error function:

$$f(\widehat{\mathbf{H}}, \widehat{\mathbf{H}}^{-1}) = \frac{1}{4m}\sum_{i=1}^{m}\left(d(\overline{\boldsymbol{x}}_i, \widehat{\overline{\boldsymbol{x}}}_i)^2 + d(\overline{\boldsymbol{x}}'_i, \widehat{\overline{\boldsymbol{x}}}'_i)^2 + d(\overline{\boldsymbol{y}}_i, \widehat{\overline{\boldsymbol{y}}}_i)^2 + d(\overline{\boldsymbol{y}}'_i, \widehat{\overline{\boldsymbol{y}}}'_i)^2\right) \tag{7}$$

Finally the error function defined by eq. (7) can be used to estimate the collineation by minimizing the following non-linear criteria:

$$\min_{\mathbf{H},\mathbf{H}'}\left(f(\mathbf{H}, \mathbf{H}') + \eta\|\mathbf{H}\mathbf{H}' - \mathbf{I}_4\|^2\right) \tag{8}$$

where $\|\mathbf{H}\mathbf{H}' - \mathbf{I}_4\|^2$ is a penalty function and η is a real positive number. A high numerical value for η guarantees that $\mathbf{H}' = \mathbf{H}^{-1}$.

An alternative method is to minimize only the distances between $\widehat{\overline{\boldsymbol{y}}}_i, \widehat{\overline{\boldsymbol{y}}}'_i$ and the true image measurements $\overline{\boldsymbol{y}}_i, \overline{\boldsymbol{y}}'_i$ in the second pair of images. In this case we need no additional constraints during the minimization:

$$\min_{\mathbf{H}} f_1(\mathbf{H}) = \min_{\mathbf{H}} \frac{1}{2m}\sum_{i=1}^{m}\left(d(\overline{\boldsymbol{y}}_i, \widehat{\overline{\boldsymbol{y}}}_i)^2 + d(\overline{\boldsymbol{y}}'_i, \widehat{\overline{\boldsymbol{y}}}'_i)^2\right) \tag{9}$$

2.4 Experimental comparison with simulated data

We implemented the methods described above to estimate $\mathbf{H}$ and we carried out a large number of experiments in order to compare the quality of the results. The algorithms have been tested on synthetic images (41 points). Gaussian noise with varying standard deviation (from 0 to 2 pixels) was added to the image point locations. The back-projected errors given by (7) were considered for comparison.

Figure 1 shows the median error over 100 trials as a function of image noise. The following notations are used:

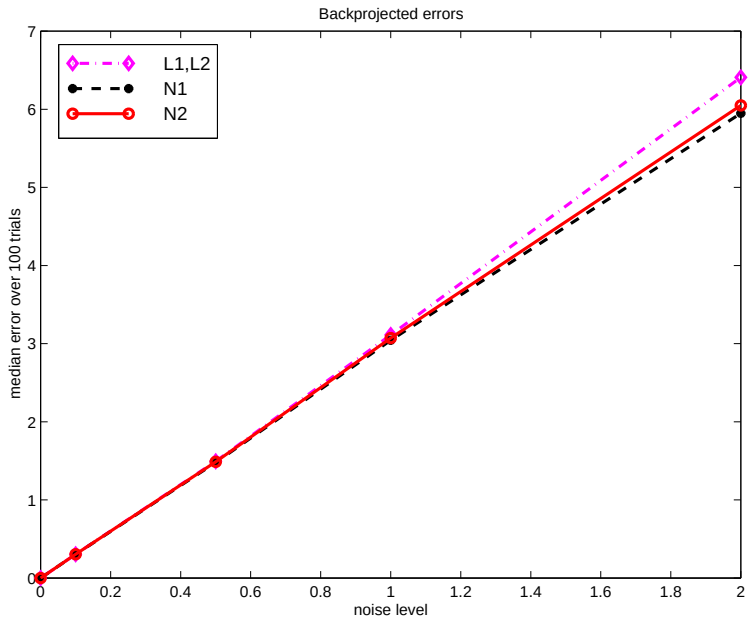

Figure 1: Comparison of different methods computing the projective collineation.

L1 - linear method 1;

L2 - linear method 2;

N1 - non-linear method, minimizing the back-projections in the two image pairs, and

N2 - non-linear method, minimizing the back-projections in the second image pair.

Note first, that the quality of the collineation linearly degrades in the presence of image noise. Moreover, the behavior of the four methods are very similar. The non-linear method (N1 and N2) slightly outperforms the linear methods which, in turn, have an identical behavior.

In the above experiments we used all the points ($m = 41$) in order to estimate the collineation. To study the sensitivity due to the number of points we applied each method to different subsets of reconstructed points. In Figure 2 we show the results obtained by the non linear method N2 using different number of points to estimate the collineation. The experiment was performed for three noise levels: $\sigma = 0.1, 0.5$ and 1.

We can deduce from Figure 2 that we need at least 15 points to compute the collineation $\mathbf{H}$ in order to have a stable back-projection of the points in the images.

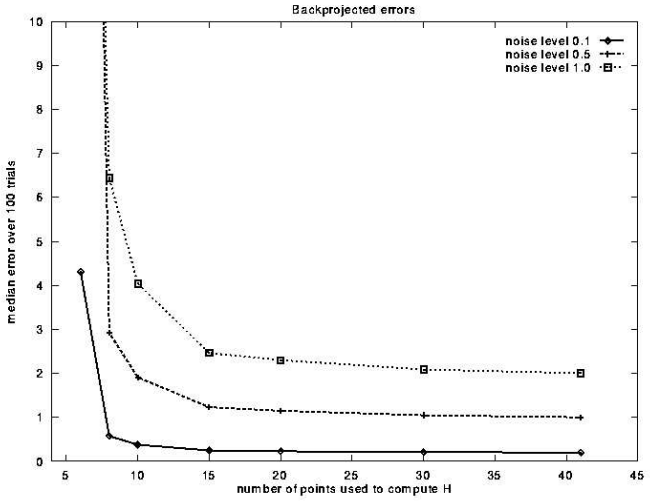

Figure 2: Testing the sensitivity due to the number of points used to estimate the projective collineation.

3 Robust method to compute H

The methods presented above work well if no mismatched points are present in the data. However, in practice mismatches or outliers may be present:

- The error in localization of most detected points of interest is small; however a few points may be inaccurately localized (over three pixels away from the correct location) and can therefore severely degrade the quality of the estimation;
- False matches may occur when points from the left image are matched against points from the right image, and
- False matches may occur when point matches from one image pair ($\boldsymbol{x}$ and $\boldsymbol{x}'$) are matched against point matches from the second image pair ($\boldsymbol{y}$ and $\boldsymbol{y}'$).

These outliers can severely affect the quality of the estimation of the collineation matrix when we apply least-squares techniques. Therefore, in this case, it is necessary to apply a robust technique.

A robust estimation method consists in the following steps (see [13] for a similar approach):

- For each i, $1 \leq i \leq m$, reconstruct $\boldsymbol{X}_i$ from $\boldsymbol{x}_i$, $\boldsymbol{x}'_i$, $\mathbf{P}_x$ and $\mathbf{P}'_x$ and $\boldsymbol{Y}_i$ from $\boldsymbol{y}_i$, $\boldsymbol{y}'_i$, $\mathbf{P}_y$ and $\mathbf{P}'_y$ [10].

- Consider n subsets of 3-D point correspondences $\boldsymbol{X}_i \leftrightarrow \boldsymbol{Y}_i$, each such subset containing k correspondences among the m available correspondences. The points associated with such a subset must be evenly distributed in the data. Such an even distribution can be achieved using a classical bucketing technique.
- For each such subset j, $1 \leq j \leq k$ a linear method to estimate a collineation $\mathbf{H}_j$ is used and a residual is determined for each trial j:

$$\begin{aligned} r_j^2 &= \frac{1}{m}\sum_{i=1}^{m} r_i^2 \\ &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{4}\Big(d(\overline{\boldsymbol{x}}_i,\widehat{\overline{\boldsymbol{x}}}_i)^2 + d(\overline{\boldsymbol{x}}'_i,\widehat{\overline{\boldsymbol{x}}}'_i)^2 + d(\overline{\boldsymbol{y}}_i,\widehat{\overline{\boldsymbol{y}}}_i)^2 + d(\overline{\boldsymbol{y}}'_i,\widehat{\overline{\boldsymbol{y}}}'_i)^2\Big) \end{aligned} \tag{10}$$

- Finally, retain the collineation $\widehat{\mathbf{H}}$ associated with the least residual.

This method can be used to reject the mismatched points (outliers) as follows. Consider the retained $\widehat{\mathbf{H}}$ and compute the back-projected error r_i for each point. Compute the mean (or median) error r^* and the standard deviation σ of the set $\{r_i\}_{i=1..n}$. Reject each matches $(\boldsymbol{x}_i, \boldsymbol{x}'_i, \boldsymbol{y}_i, \boldsymbol{y}'_i)$, for which

$$\mid r_i - r^* \mid \geq t\sigma$$

where t is a given threshold.

4 Applications

4.1 Stereo image transfer

A number of visual tasks may be carried out without using any Euclidean information about the 3-D layout. Among these tasks, visual servoing is a prominent example. Visual servoing is the task of controlling the motion of a robot manipulator using visual measurements. A sensor (either a single camera or a stereo rig) measures the error between a current robot position and a goal robot position. The robot is allowed to move until this error vanishes. The success of visual servoing depends primarily on the correctness with which the goal position has been predicted.

Consider for example the task of piling a cube on top of two other objects using a robot manipulator and a stereo image pair. The visual servoing algorithm must be given the final cube position in a representation that is invariant with respect to intrinsic and extrinsic camera calibration parameters. One such a representation is a 3-D projective reconstruction [7].

Figure 3 (a) and (b) shows an image pair of the cube as it should finally lie on top of two other objects. This image pair is used to build a projective reconstruction of the

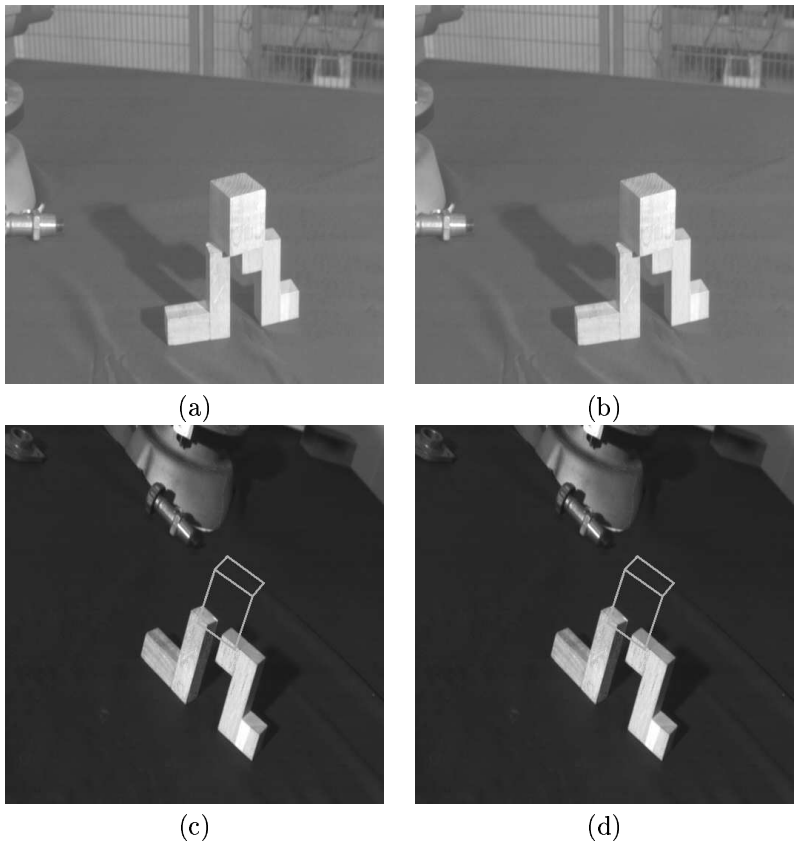

Figure 3: A stereo pair showing a cube which lies on top of two objects (a) and (b). The projective representation of the cube was "transferred" to another image pair (c) and (d).

cube in a projective basis that is attached to two other objects, the L-shaped object and the Z-shaped one – say the two "legs". This projective reconstruction consists of 24 leg points and 17 cube points. Figure 3 (c) and (d) shows an image pair of the two legs from a different viewpoint. The 24 leg points are reconstructed again and based on these points a collineation is computed between the projective reconstruction associated with the first image pair and the projective reconstruction associated with the second image pair. Based on this collineation the 17 cube points are "transferred" from the first image pair to the second image pair.

The advantage of using such a projective transfer from one image pair (learning) to another (execution) is that the two tasks (learning and execution) can be performed with two different stereo systems and that the object representation that is actually stored at

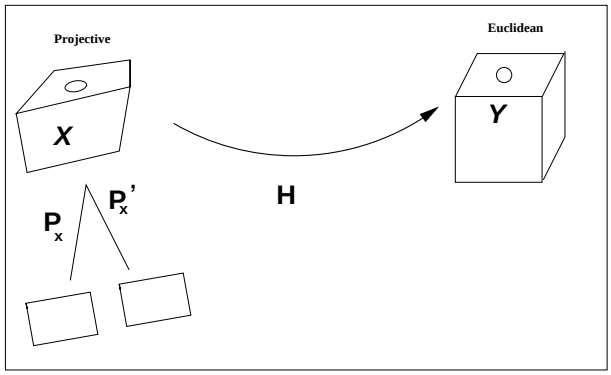

Figure 4: We are looking for **H**, that map projective reconstruction to a Euclidean one.

learning time is independent of camera calibration and of the relative position of the stereo image pair with respect to the scene.

4.2 From projective to Euclidean reconstruction

Given a projective reconstruction of a set of 3-D points, an important task is to upgrade it to Euclidean structure. In this case the collineation **H** maps the 3-D projective space onto its 3-D Euclidean subspace, Figure 4. It is assumed that the Euclidean coordinates $\overline{\boldsymbol{Y}}_i = (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)})^\top$ of a few "control" points are known in some Euclidean basis $\mathcal{B}_E$. Let $\boldsymbol{X}_i$ represent the 3-D projective coordinates of point i. $\boldsymbol{X}_i$ is a homogeneous 4-vector. Provided that there are at least 5 such control points in general position, one can compute the collineation **H** that maps $\boldsymbol{X}_i$ onto $\boldsymbol{Y}_i = (\overline{\boldsymbol{Y}}_i^\top \ 1)^\top$. The latter represents the homogeneous coordinates associated with the Euclidean coordinates of a control point i. Applying **H** to the other points of the projective reconstruction upgrades the whole projective structure to Euclidean structure.

Notice that this case is slightly different than the situation previously described because only one image pair is available. To estimate **H**, one possibility is to apply one of the two linear methods as described above. However, these linear methods minimize an algebraic distance. Here we would like to take full advantage of the known Euclidean structure associated with the control points and devise a linear method which minimizes an Euclidean rather than an algebraic error function.

Consider the 4-D vector space corresponding to the projective sub-space generated by the Euclidean basis $\mathcal{B}_E$ [11]. In this sub-space the hyper-plane $X^{(4)} = 1$ corresponds to the 3-D Euclidean space. One way to estimate **H** is to minimize the sum of squares of the

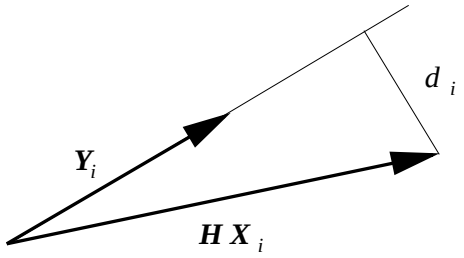

Figure 5: The Euclidean distance between vectors $\boldsymbol{Y_i}$ and $\mathbf{H}\boldsymbol{X_i}$ in 4-D vector space can be approximated by d_i.

Euclidean distances between the 3-vectors $\overline{\boldsymbol{Y}}_i$ and $\overline{\boldsymbol{V}}_i$ with:

$$\overline{\boldsymbol{V}}_i = \left(\frac{(\mathbf{H}\boldsymbol{X}_i)^{(1)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}}, \frac{(\mathbf{H}\boldsymbol{X}_i)^{(2)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}}, \frac{(\mathbf{H}\boldsymbol{X}_i)^{(3)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}} \right)^\top$$

Hence, the error function to be minimized is:

$$\min_{\mathbf{H}} \sum_{i=1}^{n} \varepsilon_i^2$$

with:

$$\varepsilon_i^2 = \left(\frac{(\mathbf{H}\boldsymbol{X}_i)^{(1)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}} - \boldsymbol{Y}_i^{(1)} \right)^2 + \left(\frac{(\mathbf{H}\boldsymbol{X}_i)^{(2)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}} - \boldsymbol{Y}_i^{(2)} \right)^2 + \left(\frac{(\mathbf{H}\boldsymbol{X}_i)^{(3)}}{(\mathbf{H}\boldsymbol{X}_i)^{(4)}} - \boldsymbol{Y}_i^{(3)} \right)^2$$

which is a non-linear function in the entries of $\mathbf{H}$. To minimize this function, non-linear optimization methods are necessary.

Alternatively, it is possible to define an Euclidean distance in the 4-D vector space associated with the Euclidean space when the latter is a sub-space of the projective space. One way to define an Euclidean distance between $\boldsymbol{Y_i}$ and $\mathbf{H}\boldsymbol{X_i}$ is to arbitrarily fix the scale factor associated with the homogeneous vector $\mathbf{H}\boldsymbol{X_i}$ and to compute the distance d_i between this vector and its projection onto the direction of $\boldsymbol{Y_i}$ (see Figure 5):

$$\begin{aligned}
d_i^2 &= (\mathbf{H}\boldsymbol{X}_i)^\top (\mathbf{H}\boldsymbol{X}_i) - \frac{(\boldsymbol{Y}_i^\top \mathbf{H}\boldsymbol{X}_i)^2}{\boldsymbol{Y}_i^\top \boldsymbol{Y}_i} \\
&= \boldsymbol{X}_i^\top \mathbf{H}^\top \mathbf{I}_4 \, \mathbf{H}\boldsymbol{X}_i - \left(\frac{1}{\boldsymbol{Y}_i^\top \boldsymbol{Y}_i} \right) \boldsymbol{X}_i^\top \mathbf{H}^\top \boldsymbol{Y}_i \boldsymbol{Y}_i^\top \mathbf{H}\boldsymbol{X}_i \\
&= \boldsymbol{X}_i^\top \mathbf{H}^\top \underbrace{\left(\mathbf{I}_4 - \frac{\boldsymbol{Y}_i \boldsymbol{Y}_i^\top}{\boldsymbol{Y}_i^\top \boldsymbol{Y}_i} \right)}_{\mathbf{A}_i} \mathbf{H}\boldsymbol{X}_i \\
&= \boldsymbol{X}_i^\top \mathbf{H}^\top \mathbf{A}_i \mathbf{H}\boldsymbol{X}_i
\end{aligned}$$

It is straightforward to check that matrix $\mathbf{A}_i$ is symmetric and since its eigenvalues are $(1, 1, 1, 0)$ it is semi-definite and positive. With the notations already used in sections 2.1 and 2.2:

$$\mathbf{H}\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \boldsymbol{X}_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & \boldsymbol{X}_i^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \boldsymbol{X}_i^\top \end{pmatrix} \begin{pmatrix} H_{11} \\ \vdots \\ H_{44} \end{pmatrix} = \mathbf{E}_i \boldsymbol{h}$$

we obtain the following error function:

$$d^2 = \sum_i d_i^2 = \boldsymbol{h}^\top \sum_i \left(\mathbf{E}_i^\top \mathbf{A}_i \mathbf{E}_i \right) \boldsymbol{h} = \boldsymbol{h}^\top \mathbf{A} \boldsymbol{h} \tag{11}$$

The solution that minimizes this error function is the eigenvector corresponding to the smallest eigenvalue of the positive semi-definite symmetric matrix $\mathbf{A}$.

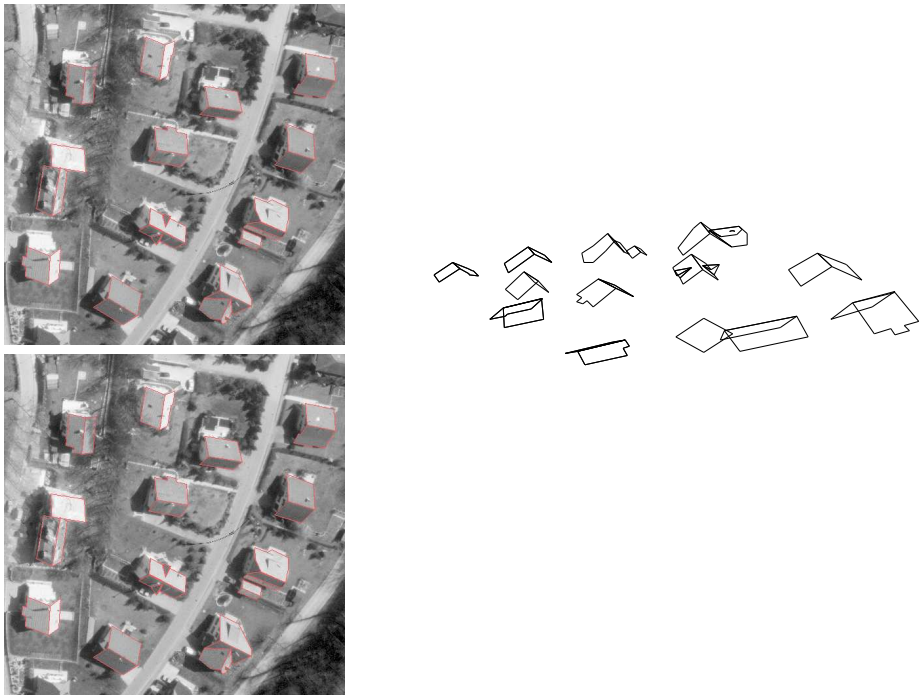

Figure 6: A pair of real images (left) and the Euclidean reconstruction of the scene (right).

To test this method on real data, consider the pair of images of Figure 6. Point correspondences were obtained interactively, and the fundamental matrix was estimated using the

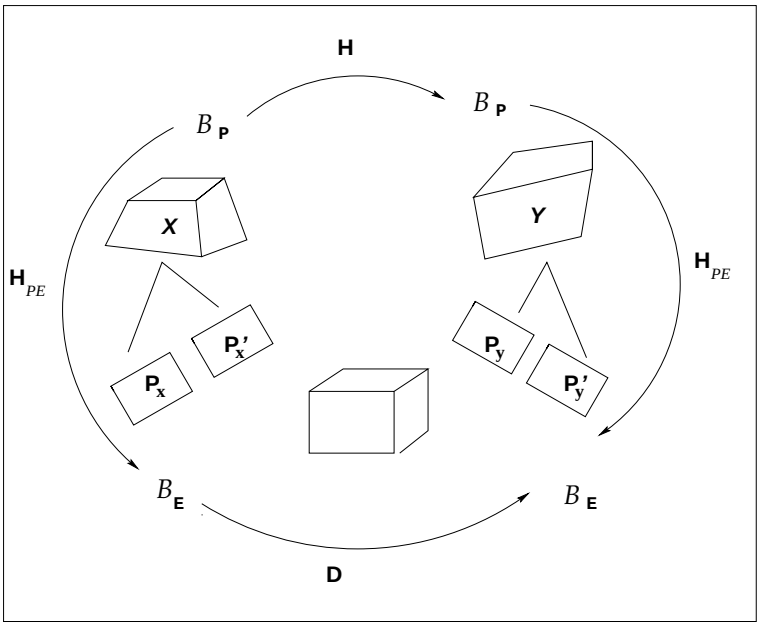

Figure 7: Rigid motion of a stereo rig.

method described in [13]. Euclidean 3-D coordinates of 26 points were provided with the data and we used them to compute the collineation **H**. Finally, this collineation was applied to all the projectively reconstructed points. Figure 6 shows the Euclidean reconstruction thus obtained.

4.3 Self-calibration and Euclidean reconstruction

We consider now an uncalibrated stereo rig that observes an unknown 3-D scene while it performs a set of rigid motions. 3-D Euclidean coordinates of the scene points are desired. In the general case, 3-D structure can be recovered only up to a 3-D projective transformation. However, if the stereo rig undergoes rigid motion and for unchanging intrinsic camera parameters, the projective ambiguity can be reduced to affine or to Euclidean [14, 2, 6]. These methods make use of some of the algebraic properties of the collineation matrix **H** between two projective reconstructions. Indeed, as outlined below, the projection transformation is conjugated, in this special case, to a rigid transformation.

Indeed, if the two pairs of images were obtained by the same stereo rig which was moved rigidly, the fundamental matrixes $\mathbf{F}_x$ and $\mathbf{F}_y$, and hence $\mathbf{P}'_x = \mathbf{P}'_y$ are the same. Therefore, $(\mathbf{P}_x, \mathbf{P}'_x)$ and $(\mathbf{P}_y, \mathbf{P}'_y)$ define the same the projective basis $\mathcal{B}_P$. Assume that $\mathcal{B}_E$

is the Euclidean basis associated with the stereo rig before the motion and let $\mathbf{H}_{PE}$ be the collineation between the bases $\mathcal{B}_P$ and $\mathcal{B}_E$ (see Figure 7).

In this case, the collineation that maps the reconstructed points before the motion $\boldsymbol{X}_i$ to the reconstructed points after the motion $\boldsymbol{Y}_i$ is given by

$$\mathbf{H} \simeq \mathbf{H}_{PE}^{-1}\mathbf{D}\mathbf{H}_{PE} \tag{12}$$

where $\mathbf{D}$ is the displacement of the stereo rig and $\mathbf{H}_{PE}$ is the collineation which update the projective points $\boldsymbol{X}_i$ (or equivalently $\boldsymbol{Y}_i$) to a Euclidean representation. It has been shown in [6] that

$$\mathbf{H}_{PE} = \begin{pmatrix} \mathbf{K}^{-1} & \mathbf{0}_3 \\ \boldsymbol{a}^\top & \lambda \end{pmatrix} \tag{13}$$

where $\mathbf{K}$ is the matrix of intrinsic parameters of the left camera and $\begin{pmatrix} \boldsymbol{a}^\top & \lambda \end{pmatrix}$ is the equation of the plane at infinity in the projective basis $\mathcal{B}_P$.

Expanding (12), after a normalization of $\mathbf{H}$ such that $det(\mathbf{H}) = 1$ and $trace(\mathbf{H}) > 0$, gives

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} \mathbf{K} & \mathbf{0}_3 \\ -\frac{1}{\lambda}\boldsymbol{a}^\top\mathbf{K} & \frac{1}{\lambda} \end{pmatrix} \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0}_3^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{K}^{-1} & \mathbf{0}_3 \\ \boldsymbol{a}^\top & \lambda \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{K}\mathbf{R}\mathbf{K}^{-1} + \mathbf{K}\boldsymbol{t}\boldsymbol{a}^\top & \lambda\mathbf{K}\boldsymbol{t} \\ \frac{1}{\lambda}(-\boldsymbol{a}^\top\mathbf{K}\mathbf{R}\mathbf{K}^{-1} - \boldsymbol{a}^\top\mathbf{K}\boldsymbol{t}\boldsymbol{a}^\top + \boldsymbol{a}^\top) & -\boldsymbol{a}^\top\mathbf{K}\boldsymbol{t} + 1 \end{pmatrix} \end{aligned} \tag{14}$$

It was shown in [6], that if $\mathbf{D}$ is a general displacement, one can compute $(\boldsymbol{a}^\top, \lambda)^\top$ up to a scale from:

$$\mathbf{H}^\top \begin{pmatrix} \boldsymbol{a} \\ \lambda \end{pmatrix} = \mathbf{0}_4$$

Furthermore, from (14) results

$$\mathbf{G}_\infty = \mathbf{K}\mathbf{R}\mathbf{K}^{-1} = \mathbf{H}_{33} - \frac{1}{\lambda}\boldsymbol{h}_1\boldsymbol{a}$$

where the following notation was considered

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{33} & \boldsymbol{h}_1 \\ \boldsymbol{h}_2^\top & h_{44} \end{pmatrix}$$

Note that $\mathbf{G}_\infty$ is the infinite homography between the images of the left camera, before and after the rigid motion. Using the orthogonality of rotation matrices one can easily obtain the following relationship:

$$\mathbf{G}_\infty^\top\mathbf{K}^{-\top}\mathbf{K}\mathbf{G}_\infty = \mathbf{K}^{-\top}\mathbf{K}$$

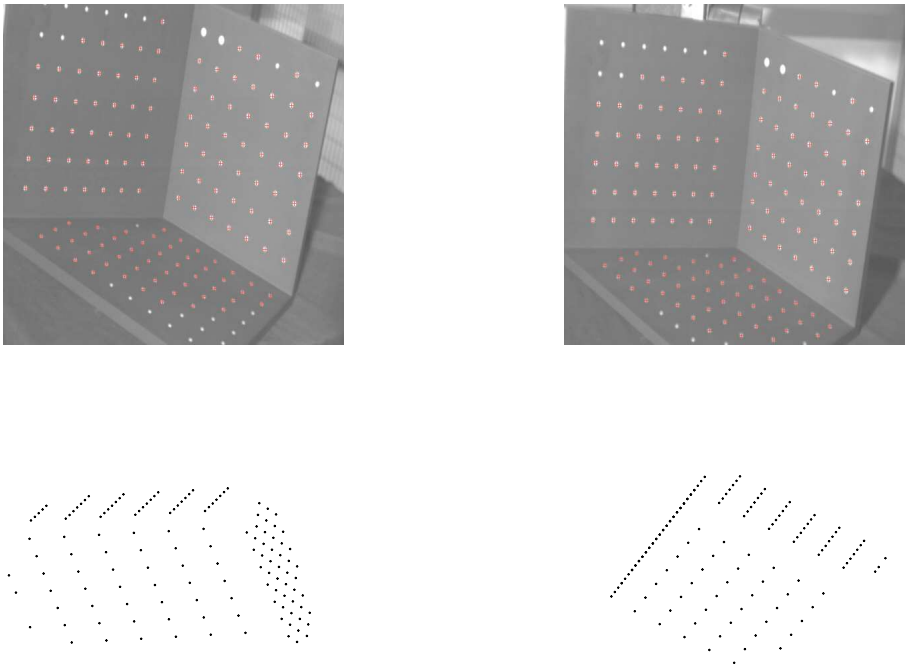

Figure 8: A pair of images of the calibration grid (up) and the Euclidean reconstruction of the scene (down).

The matrix $\mathbf{K}^{-\top}\mathbf{K}$ is known as the image of the absolute conic. Therefore, one can compute the image of the absolute conic, and hence the camera intrinsic parameters, if matrix $\mathbf{G}_\infty$ is known [5].

From $\mathbf{G}_\infty$ and $\mathbf{K}$ results $\mathbf{R} = \mathbf{K}^{-1}\mathbf{G}_\infty\mathbf{K}$ and from $\boldsymbol{h}_1 = \lambda\mathbf{K}\boldsymbol{t}$ the translation up to a scale is given by:

$$\boldsymbol{t} = \frac{1}{\lambda}\mathbf{K}^{-1}\boldsymbol{h}_1$$

To summarize, the collineation $\mathbf{H}$ allows us to estimate the plane at infinity and the internal camera parameters and hence, to convert projective structure to affine and to Euclidean structure.

To give an example of the usefulness of $\mathbf{H}$ in this case (one can find further experimental results in [14, 2, 6]) we gathered two image pairs of the calibration grid (Figure 8) with

the same stereo rig and we computed the collineation $\mathbf{H}$. From $\mathbf{H}$ we obtained the plane at infinity $\begin{pmatrix} \boldsymbol{a}^\top & \lambda \end{pmatrix}$ and the intrinsic parameters $\mathbf{K}$. Furthermore, we updated the projective points $\boldsymbol{X}_i$ using $\mathbf{H}_{PE}$ and thus obtaining a Euclidean reconstruction. Two different views of the 3-D sets of points thus obtained are shown on Figure 8.

5 Conclusion

In this paper we described various methods to estimate the collineation between two projective reconstructions. We described two linear methods which minimize an algebraic distance in 3-D projective space, two non-linear methods which minimize a metric distance in 2-D image space, and a robust method which combines any of the linear methods with an outlier rejection strategy. Moreover, for the special case where Euclidean coordinates of a few control points are available, we devised a linear method based on a metric distance in the 4-D vector space associated with homogeneous coordinates.

In order to evaluate these methods and assess their respective merits, we carried out a large number of experiments with both synthetic and real data. The synthetic data allowed us to add Gaussian noise to image data and to study the numerical stability of the various implemented methods. In the light of these experiments it appears that all the methods have an almost identical behavior and that the non-linear methods outperform the linear ones. However for levels of noise below one pixel in magnitude it is not possible to distinguish between the performances of the four methods. Another set of experiments allowed us to test the performance as a function of the number of points being considered. The conclusion of this set of experiments is, not surprisingly, that (i) the more point matches are available the better it is and (ii) below 15 to 20 point matches one should expect unreliable results.

For the special case where projective coordinates of a few control points are matched against Euclidean coordinates, we have been able to devise a linear method which is based on a metric distance in 4-D space – the vector space associated with homogeneous coordinates of the 3-D projective space and its Euclidean sub-space. Such a linear error function could probably be extended to other estimation problems such camera calibration and bundle adjustment.

One interesting application associated with the computation of such a 3-D projective transformation is the transfer of 2-D points from one stereo image pair to another stereo image pair. This transfer technique is useful whenever a set of 3-D points is observed with one stereo pair and one wants to display the same 3-D set as it would be observed by another stereo pair. This stereo transfer technique has been applied to a visual grasping method developed elsewhere [7].

Finally we described the relationship between the 3-D projective transformation and the self-calibration of a stereo rig. This is probably one of the most promising outcome because it allows the upgrading of a 3-D projective reconstruction to affine and to Euclidean based on linear algebra.

Acknowledgments

The authors would like to thank Zhengyou Zhang for previous discussions and suggestions.

References

[1] P. A. Beardsley, I. D. Reid, , A. Zisserman, and D. W. Murray. Active visual navigation using non-metric structure. In *Proceedings Fifth International Conference on Computer Vision*, pages 58–64, Cambridge, Mass, June 1995. IEEE Computer Society Press, Los Alamitos, CA.

[2] F. Devernay and O. Faugeras. From projective to Euclidean reconstruction. In *Proceedings Computer Vision and Pattern Recognition Conference*, pages 264–269, San Francisco, CA., June 1996.

[3] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[4] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[5] R.I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 471–478. Springer-Verlag, May 1994.

[6] R. Horaud and G. Csurka. Self-calibration and Euclidean reconstruction using motions of a stereo rig. In *Proceedings Sixth International Conference on Computer Vision*, pages 96–103, Bombay, India, January 1998. IEEE Computer Society Press, Los Alamitos, Ca.

[7] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 14(4), August 1998.

[8] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 589–599, May 1994.

[9] 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 6th International Conference on Computer Vision, Bombay, India*, pages 90–95, January 1998.

[10] C. Rothwell, O. Faugeras, and G. Csurka. A comparison of projective reconstruction methods for pairs of views. *Computer Vision and Image Understanding*, 68(1):36–58, October 1997.

[11] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry.* Oxford Science Publication, 1952.

[12] Z. Zhang. A stereovision system for a planetary rover: Calibration, correlation, registration and fusion. In *Proceedings of the* IEEE *Workshop on Planetary Rover Technology and Systems*, Minneapolis, Minnesota, USA, April 1996.

[13] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.

[14] A. Zisserman, P.A. Beardsley, and I.D. Reid. Metric calibration of a stereo rig. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 93–100, June 1995.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399