Euclidean 3D reconstruction from stereo sequences with variable focal lengths

Marc Pollefeys[*], *Luc Van Gool, Theo Moons*[†]

Katholieke Universiteit Leuven, E.S.A.T. / MI2
Kard. Mercierlaan 94, B-3001 Leuven, BELGIUM
Marc.Pollefeys,Luc.VanGool,Theo.Moons@esat.kuleuven.ac.be

Abstract

A stereo rig can be calibrated using a calibration grid, but recent work demonstrated the possibility of autocalibration. There remain two important limitations, however. First, the focal lengths of the cameras should remain fixed, thereby excluding zooming or focusing. Second, the stereo rig must not purely translate, which however is the most natural type of motion. This also implies that these methods also collapse when the motion comes close to being a translation.

The paper extends the work to allow changes in focal lengths (these may be independent for both cameras) and purely translational motions of the stereo rig. First, the principal points of both cameras are retrieved. Changes in focal lengths are dealt with through weak calibration. Each position of the rig yields a projective reconstruction. The projective transformation between then allows to first retrieve affine structure which subsequently is upgraded to metric structure, following the general outline described in [12].

Rather than posing a problem to the method, rig translation allows further simplifications and is advantageous for robustness.

1 Introduction

Recently, methods to obtain the Euclidean calibration of a stereo rig have been proposed [12, 3]. These methods impose some restrictions. First, all intrinsic camera parameters are assumed fixed. This implies that e.g. the camera focal lengths are not allowed to change. This precludes useful adaptations to the scene such as zooming and focusing. Second, the rig is not allowed to purely translate. Unfortunately, translation is often preferable (e.g. shortest path between points). In practice, the methods only work well if the rotational motion component is sufficiently large.

2 camera model

The camera model used is the pinhole model, where the image is formed under central projection on a photo-sensitive plane perpendicular to the optical axis. Changes in focal length move the optical center along the axis, leaving the principal point[1] unchanged. This assumption is fulfilled to a sufficient degree with the cameras used in the experiments and this is typically the case [7]. The relation between image points and world points is given by

$$\lambda_{ijs} m_{ijs} = \mathbf{P}_{js} M_i \tag{1}$$

with P_{js} the 3x4 camera matrix [2], m_{ijs} and M_i are column vectors containing the homogeneous coordinates of the image points and world points resp., and λ_{ijs} expresses the equivalence up to a scale factor. If $\mathbf{P}_{js}$ represents a Euclidean camera, it can be put in the form [6]

$$\mathbf{P}_{js} = \mathbf{K}_{js} \left[\mathbf{R}_{js} | -\mathbf{R}_{js} t_{js} \right] \tag{2}$$

where $\mathbf{R}_{js}$ and t_{js} represent the Euclidean orientation and position of this camera with respect to a world frame, and $\mathbf{K}_{js}$ is the calibration matrix of the j^{th} camera:

$$\mathbf{K}_{js} = \begin{bmatrix} r_{xs}^{-1} & -r_{xs}^{-1} \cos\theta_s & f_{js}^{-1} u_{xs} \\ & r_{ys}^{-1} & f_{js}^{-1} u_{ys} \\ & & f_{js}^{-1} \end{bmatrix} \tag{3}$$

[*] IWT fellow (Flemish Institute for the Promotion of Scientific-Technological Research in Industry)

[†] Postdoctoraal researcher of the Belgian National Fund for Scientific Research

[1] The principal point is defined as the intersection point of the optical axis and the image plane

[2] $\mathbf{P}_{js}$ is the camera matrix for the j^{th} view, s stands for *left* or *right*.

In this equation r_{xs} and r_{ys} represent the pixel width and height, θ_s is the angle between the image axes, u_{xs} and u_{ys} are the coordinates of the principal point, and f_{js} is the focal length. Notice that the calibration matrix is only defined up to scale. In order to highlight the effect of changing the focal length the calibration matrix $\mathbf{K}_{js}$ will be decomposed in two parts:

$$\mathbf{K}_{js} = \begin{bmatrix} 1 & 0 & (f_{1s}/f_{js} - 1)u_{xs} \\ 0 & 1 & (f_{1s}/f_{js} - 1)u_{ys} \\ 0 & 0 & f_{1s}/f_{js} \end{bmatrix} \mathbf{K}_{1s} \qquad (4)$$

The second part $\mathbf{K}_{1s}$ is equal to the calibration matrix $\mathbf{K}_{1s}$ for that camera for view 1, whereas the first part, which will be called $\mathbf{K}_{f_{js}s}$ in the remainder of this text, models the effect of changes in focal length. From equation (4) it follows that once the principal point u is known, $\mathbf{K}_{f_{js}s}$ is known for any given value of f_{js}/f_{1s}. Therefore, finding the principal point is the first step of the reconstruction method. Then, if the change in focal length between two views can be retrieved, its effect is canceled by multiplying the image coordinates to the left by $\mathbf{K}_{f_{js}s}^{-1}$.

Retrieving the principal points u_s is relatively easy for cameras equipped with a zoom. Upon changing the focal length (without moving the camera or the scene), each image point will – according to the pinhole model – move on a line passing through the principal point. By taking two or more images with a different focal length and by fitting lines through the corresponding points, the principal point can be retrieved as the common intersection of all these lines. This is illustrated in figure 1. In practice the lines will not intersect precisely and a least squares approximation is used. This method has been used by others [7].

For the sake of simplicity we assume $\mathbf{R}_{1s} = \mathbf{I}$, $t_{1s} = 0$ and $f_{1s} = 1$ in the remainder of this paper. Because the reconstruction is up to scaled Euclidean (i.e. similarity) this is not a restriction. In this way the 7 degrees of freedom of a similarity transformation are fixed.

3 Retrieving focal length

It is clear that having a way of cancelling the effects of focal length changes would generalise the existing autocalibration methods [12, 3] to situations where such changes occur.

3.1 The effect on epipoles

The epipoles are two points associated with a pair of cameras. The epipole in one camera image is the projection of the other camera's center. Note that the epipoles of a fixed stereo rig stay put, independent of the rig's motion. If the focal lengths of the cameras change, however, then the epipoles will shift (fig. 2). These shifts suffice to derive the relative change in focal length. One can then also transform the images to what they would have been like without the change in focal lengths. It follows from the equations for the epipoles

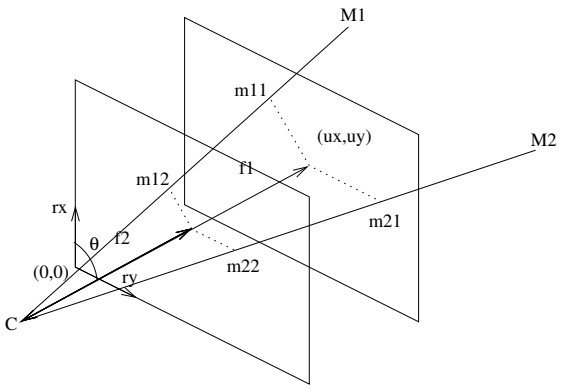

Figure 1: *Illustration of the camera and zoom model. The focal lengths f_1 and f_2 are different, the other parameters ($r_x, r_y, u_x, u_y, \theta$) are identical.*

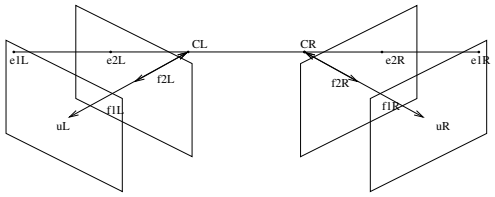

Figure 2: *Illustration of the displacement of the epipole for the left camera when changing the focal length. C_L and C_R are the camera centers, e_{0L} and e_{1L} the epipoles in the left camera for different focal lengths (f_{0L} and f_{1L}) and $u_{xL}, u_{yL}, r_{xL}, r_{yL}$ are the internal camera parameters*

$$\lambda_{e_{js}} e_{js} = \lambda_{e_s}(e_s + (f_{1s}/f_{js} - 1)u_s) \qquad (5)$$

that f_{js}/f_{1s} can be recovered in a linear way. In fact, the explicit calculation of focal lengths is not called for. It suffices to apply the transformation $\mathbf{K}_{f_{js}s}^{-1}$ to the image (see equation (4)).

4 Affine and Euclidean calibration

The proposed work is a generalisation of the elegant method proposed by Zisserman [12]. This method first retrieves the affine calibration based on the eigenvector

structure of the transformation $\mathbf{T}$ between the projective reconstructions from the two positions of the rig. Once the infinite homography is known we can use the constraints on the camera calibration matrix described in [6, 8]. In the case of a translating stereo rig this is not enough [12]. This means that for any movement close to translation the problem is ill-conditioned. Hence, one has to strike a balance between the ease of setting up the system (the less calibration the better) and the flexibility it has to offer in its use (e.g. being able to perform any kind of motion and to dynamically zoom and focus). Hence, this paper gives in with respect to completely calibration-free operation. The principal points are extracted, which is not too difficult through the very application of changes in the focal lengths.

4.1 Affine calibration

One view with a stereo rig is enough to get a projective reconstruction [5, 4]. Having two views yields two reconstructions, M_{iP1} and M_{iP2} say (these are vectors of homogenous coordinates for all the scene points). The two reconstructions are related by a projective transformation $\mathbf{T}$:

$$\lambda_{M_{iP1}} M_{iP1} = \lambda_{M_{iP2}} \mathbf{T} M_{iP2} \tag{6}$$

If one uses the same camera matrices for both reconstructions (i.e. $\mathbf{P}_L = [\mathbf{I}|0]$ and $\mathbf{P}_R = [[e_R]_\times \mathbf{F}|e_R]$ with $\mathbf{F}$ the fundamental matrix and e_R the epipole of the right camera), $\mathbf{T}$ can be written as

$$\mathbf{T} = \mathbf{T}_{PE}^{-1} \mathbf{T}_E \mathbf{T}_{PE} \ , \tag{7}$$

a conjugation of the euclidean transformation $\mathbf{T}_E$, the motion of the rig between the two views, with $\mathbf{T}_{PE}$, the transformation between the reconstructions and the Euclidean world. This observation is key to the following analysis proposed in [12]. The eigenvectors of $\mathbf{T}$ are related to these of $\mathbf{T}_E$ by the same transformation $\mathbf{T}_{PE}$. Because the eigenvectors of $\mathbf{T}_E$ are all in the plane at infinity π_∞, the eigenvectors of $\mathbf{T}$ indicate the position of π_∞ in the projective frame. Once π_∞ is known it is easy to get $\mathbf{H}_{\infty LR}$. One can project the eigenvectors of $\mathbf{T}$ using the projection matrices $\mathbf{P}_L$ and $\mathbf{P}_R$. The homography between these two projections and between the two epipoles is $\mathbf{H}_{\infty LR}$. To obtain an affine reconstruction one can use the camera matrices

$$\mathbf{P}_L = [\mathbf{I}|0] \qquad \mathbf{P}_R = [\mathbf{H}_{\infty LR}|e_R] \tag{8}$$

4.2 Affine calibration for translation

If the motion of the stereo rig is restricted to a translation, there is an easier and more robust method to recover the affine structure of the scene, however [9], which is first generalised to changing focal lengths.

In the case of a translation (without changing the focal length) between two views, the epipolar geometry is the same for both images and the image points lie on their own epipolar lines. This means that the epipolar geometry is completely determined by knowing the position of the unique epipole. Adding changes in focal length between the images adds one degree of freedom when the principal point is known.

Given three points in the two views, one know that a scaling equal to the focal length ratio should bring them in position such that the lines through corresponding points intersect in the epipole. This immediately yields a quadratic equation in the focal length ratio. The epipole follows as the resulting intersection. In practice the data will be noisy, however, and it is better to consider information from several points. The following equations describe the relation between the image coordinates for both images

$$\lambda_{i2s} m_{i2s} = \lambda_{i1s} \left(m_{i1s} + (f_{1s}/f_{2s} - 1) u_s\right) + \lambda_{e_{21s}} e_{21s} \tag{9}$$

where m_{i1s}, m_{i2s}, u_s and e_{21s} are column vectors of the form $[x\, y\, 1]^\top$. Equation (9) gives 3 constraints for every point and was used to form an overdetermined system, yielding among other things f_{2s}/f_{1s} and λ_{i1s}. This leads to a system of nonlinear equations, which can be solved robustly [10].

At this stage the affine reconstruction is trivial to obtain. From equation (1) it followse that $[\lambda_{i1s} m_{i1s}^\top\ 1]^\top$ is related to M_i by an affine transformation.

In the next paragraph the infinity homographies will be needed. For translational motions $\mathbf{H}_{\infty 12L}$ and $\mathbf{H}_{\infty 12R}$ will be equal to $\mathbf{K}_{f_{2L}L}$ and $\mathbf{K}_{f_{2R}R}$ respectively. $\mathbf{H}_{\infty 1LR}$ can be extracted as the 3x3 upper-left submatrix of the affine transformation relating the affine reconstructions obtained from the left and the right camera respectively[3].

4.3 Euclidean calibration

To upgrade the reconstruction to Euclidean structure, the camera calibration matrix $\mathbf{K}_{1L}$ (or $\mathbf{K}_{1R}$) has to be known. This is equivalent to knowing the image $\mathbf{B}_{1L}$ of the dual of the absolute conic for the left camera, since $\mathbf{B}_{1L} = \mathbf{K}_{1L}\mathbf{K}_{1L}^\top$. The images are constrained in the following way [6, 8, 12]:

$$\kappa_{1LR} \mathbf{B}_{1R} = \mathbf{H}_{\infty 1LR} \mathbf{B}_{1L} \mathbf{H}_{\infty 1LR}^\top \tag{10}$$

and for each camera:

$$\kappa_{12L} \mathbf{B}_{2L} = \mathbf{H}_{\infty 12L} \mathbf{B}_{1L} \mathbf{H}_{\infty 12L}^\top \tag{11}$$

[3] These reconstructions must be built with camera centered reference frames

$$\kappa_{12R}\mathbf{B}_{2R} = \mathbf{H}_{\infty 12R}\mathbf{B}_{1R}\mathbf{H}_{\infty 12R}^{\top} \quad (12)$$

Eqs. (11) and (12) are easier to use because κ_{12L} and κ_{12R} can be forced to 1 by taking $\det \mathbf{H}_{\infty 12L} = 1$ which gives a set of linear equations. The problem with pure translation is that eqs. (11) and (12) become trivial.

The knowledge of u_L and u_R is called to the rescue. Assuming that the camera axes are orthogonal avoid having to solve a non-linear set of equations. Take a closer look at $\mathbf{B}_{1L}$ (or $\mathbf{B}_{1R}$ for that matter).

$$\mathbf{B}_{1L} = \begin{bmatrix} f_{1L}^2 r_{Lx}^{-2} + u_{Lx}^2 & u_{Lx}u_{Ly} & u_{Lx} \\ u_{Lx}u_{Ly} & f_{1L}^2 r_{Ly}^{-2} + u_{Ly}^2 & u_{Ly} \\ u_{Lx} & u_{Ly} & 1 \end{bmatrix} \quad (13)$$

Combining eqs. (10) and (13) gives 4 linear equations in 3 unknowns (equations for $\mathbf{b}_{R12}, \mathbf{b}_{R13}, \mathbf{b}_{R23}$ and $\mathbf{b}_{R33}$):

$$u_{Rx}u_{Ry}\kappa_{\mathbf{LR}} = a_{12}\mathbf{r_{Lx}^{-2}} + b_{12}\mathbf{r_{Ly}^{-2}} + c_{12} \quad (14)$$

$$u_{Rx}\kappa_{\mathbf{LR}} = a_{13}\mathbf{r_{Lx}^{-2}} + b_{13}\mathbf{r_{Ly}^{-2}} + c_{13} \quad (15)$$

$$u_{Ry}\kappa_{\mathbf{LR}} = a_{23}\mathbf{r_{Lx}^{-2}} + b_{23}\mathbf{r_{Ly}^{-2}} + c_{23} \quad (16)$$

$$\kappa_{\mathbf{LR}} = a_{33}\mathbf{r_{Lx}^{-2}} + b_{33}\mathbf{r_{Ly}^{-2}} + c_{33} \quad (17)$$

where the unknowns are in bold and a_{ij}, b_{ij} and c_{ij} only depend on $\mathbf{H}_{\infty RL}$ and u_L (which are both known). Once we know κ_{LR}, r_{Lx} and r_{Ly} we could solve the following equations for $f_R^2 r_{Rx}^{-2}$ and $f_R^2 r_{Ry}^{-2}$.

$$\kappa_{LR}(\mathbf{f_R^2 r_{Rx}^{-2}} + u_{Rx}^2) = \ldots \quad (18)$$

$$\kappa_{LR}(\mathbf{f_R^2 r_{Ry}^{-2}} + u_{Ry}^2) = \ldots \quad (19)$$

In fact r_{Rx} and r_{Ry} are not required. The Euclidean calibration of the left camera suffices to obtain a Euclidean reconstruction. We can upgrade the affine reconstruction (obtained by the methods described in the previous paragraphs) to Euclidean by applying the transformation

$$\mathbf{T}_{AE} = \begin{bmatrix} \mathbf{K}_{1L}^{-1} & 0 \\ 0 & 1 \end{bmatrix} \quad (20)$$

5 Results

The algorithm was applied to synthetic images as well as real images. From tests with synthetic data one can conclude that restricting the motion to translation gives more stable results. For a report on these results, see [10].

Next some results obtained from a real scene are presented. The scene consists of a box and a cylindrical object on a textured background. Images were acquired with a translating stereo rig, they can be seen in figure 3. Figure 4 shows the reconstruction results. Notice that angles are well preserved (e.g.the top and the front view differ by 90^o, the box and the floor have right angles in the reconstruction. The inaccuracies in the reconstruction (like the dent in the cylindrical object) are mainly due to the rendering process which uses triangulation between matched points and are not related to the accuracy of the calibration.

Figure 3: *Images of a scene taken with a translating stereo rig*

6 conclusion

The possibility to obtain the auto-calibration of a moving stereo rig with variable focal lengths was demonstrated. Only very mild forms of camera calibration had to be introduced in return. Moreover, it was shown that the method generalizes to cases of pure translation, which was not only impossible with existing methods but could also be implemented with increased robustness. The method was illustrated with a real scene which. The results are -for the Euclidean auto-calibration aspects- convincing.

Further plans includes the integration of the methods into an implementation that detects the degenerated cases (i.e. translation) by itself. Also the application of more robust techniques for the recovery of the projective structure is under investigation. Another interesting path of research is to investigate the possibility of dealing with variations in other parameters than focal length.

acknowledgement

Research financed by a specialisation grant from the Flemish Institute for Scietific Research in Industry (IWT) and European ACTS project VANGUARD.

Figure 4: *different views of the reconstruction (*top: *top view* ,middle: *general view*,bottom: *front view)*

References

[1] M. Armstrong, A. Zisserman, and P. Beardsley, Euclidean structure from uncalibrated images, *Proc. 5th BMVC*, 1994.

[2] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. *Proc.ECCV'94*, pp. 567–576, Springer-Verlag, 1994.

[3] F. Devernay and O. Faugeras, From Projective to Euclidean Reconstruction, *INSIGHT meeting Leuven*, 1995.

[4] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, *Proc.ECCV'92*, pp.321-334, 1992.

[5] R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc.ECCV'92*, pp.579-587, 1992.

[6] R. Hartley, Euclidean reconstruction from uncalibrated views, in: J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of invariance in Computer Vision*, Lecture Notes in Computer Science **825**, pp. 237–256, Springer, 1994.

[7] M. Li, Camera Calibration of a Head-Eye System for Active Vision, *Proc. ECCV'94*, pp. 543–554, Springer-Verlag, 1994.

[8] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *Proc. ECCV'94*, pp. 589–597. Springer-Verlag, 1994.

[9] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, Affine reconstruction from perspective image pairs, in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science **825**, pp. 297–316, Springer, 1994.

[10] M. Pollefeys, L. Van Gool, and M. Proesmans, Euclidean 3D reconstruction from image sequences with variable focal lengths, Technical Report K.U.Leuven, E.S.A.T./MI2, 1995.

[11] C. Rothwell, G. Csurka, and O.D. Faugeras, A comparison of projective reconstruction methods for pairs of views, *Proc. ICCV'95*,pp. 932-937, 1995

[12] A. Zisserman, P.A.Beardsley, and I.D. Reid, Metric calibration of a stereo rig. In *Proc. Workshop on Visual Scene Representation*, Boston, MA, June 1995