Katholieke
Universiteit
Leuven

Departement Elektrotechniek
Afdeling ESAT/MI2
Kardinaal Mercierlaan 94
B-3001 Heverlee - Belgium

TECHNISCH RAPPORT - TECHNICAL REPORT

Euclidean 3D reconstruction from image sequences with variable focal lengths

Marc Pollefeys, Luc Van Gool and Marc Proesmans

October 1995

Nr. KUL/ESAT/MI2/9508

Euclidean 3D reconstruction from image sequences with variable focal lengths

Marc Pollefeys*, Luc Van Gool, Marc Proesmans
Katholieke Universiteit Leuven, E.S.A.T. / MI2
Kard. Mercierlaan 94, B-3001 Leuven, BELGIUM
Tel: +32 16 321064(MP) 321705(LVG) 321090(MPr)
Fax: +32 16 321986
Marc.Pollefeys, Luc.VanGool, Marc.Proesmans@esat.kuleuven.ac.be

Abstract

One of the main problems to obtain a Euclidean 3D reconstruction from multiple views is the calibration of the camera. A first possibility is explicit calibration. This is not always practical and has to be repeated regularly. Sometimes this approach is even impossible (i.e. for pictures taken by an unknown camera of an unknown scene). The second possibility is to do auto-calibration. Here the rigidity of the scene is used to obtain constraints on the camera parameters. Existing approaches of this second strand impose that the camera parameters stay exactly the same between different views. This can be very limiting since it excludes changing the focal length to zoom or focus. The paper describes a reconstruction method that allows to vary the focal length. Instead of using one camera one can also use a stereo rig following similar principles, and in which case also reconstruction from a moving rig becomes possible even for pure translation. Synthetic data were used to see how resistant the algorithm is to noise. The results are satisfactory. Also results for a real scene were convincing.

1 Introduction

Given a general set of images of the same scene one can only build a *projective* reconstruction [4, 6, 15]. Reconstruction up to a smaller transformation group (i.e. *affine* or *Euclidean*) requires additional constraints. Existing methods assume that all internal camera parameters stay exactly the same for the different views[1]. Hartley [7] proposed a method to obtain a Euclidean reconstruction from three images. The method needs a

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

[1]We assume that in general nothing is known about the scene to reconstruct.

non-linear optimisation step wich is not guaranteed to converge. Having an affine reconstruction eliminates this problem. Moons *et al* [11] described a method to obtain an affine reconstruction when the camera movement is a pure translation. Armstrong *et al* [1] combined both methods [11, 7] to obtain a Euclidean reconstruction from three images with a translation between the first two views.

In contrast to the 2D case, where viewpoint independent shape analysis and the use of uncalibrated cameras go hand in hand, the 3D case is more subtle. The precision of reconstruction depends on the level of calibration, be it in the form of information on camera or scene parameters. Thus, uncalibrated operation comes at a cost and it becomes important to carefully consider the pro's and con's of needing knowledge on the different internal and external camera parameters.

As an example, the state-of-the-art strategy to keep all internal camera parameters unknown but fixed, means that one is not allowed to zoom or adapt focus. This can be a serious limitation in practical situations. It stands to reason that the ability to keep the object of interest sharp and at an appropriate resolution would be advantageous. Also being allowed to zoom in on details that require a higher level of precision in the reconstruction can save much trouble. To the best of our knowledge no method using uncalibrated cameras for Euclidean reconstruction allows this (i.e. the focal length has to stay constant).

This paper describes a method to obtain a *Euclidean reconstruction* from images taken with an *uncalibrated camera* with a *variable focal length.* In fact, it is an adaptation of the methods of Hartley [7], Moons *et al* [11] and Armstrong *et al* [1], to which it adds an initial step to determine the position of the principal point. Thus, a mild degree of camera calibration is introduced in exchange for the freedom to change the focal length between views used for reconstruction. The very ability to change the focal length allows one to recover the principal point in a straight-forward way. From there, the method starts with an affine reconstruction from two views with a translation in between. A third view allows an upgrade to Euclidean structure. The focal length can be different for each of the three views. In addition the algorithm yields the relative changes in focal length between views.

Recently, methods for the Euclidean calibration of a fixed stereo rig from two views taken with the rig have been propounded [18, 3]. The stereo rig must rotate between views. It is shown here that also in this case the flexibility of variable focal length can be provided for and that reconstruction is also possible after pure translation, once the principal points of the cameras are determined.

2 camera model

In this paper a pinhole camera model will be used. Central projection forms an image on a light-sensitive plane, perpendicular to the optical axis. Changes in focal length move the optical center along the axis, leaving the principal point[2] unchanged. This assumption is fulfilled to a sufficient extent in practice [9]. The following equation expresses the relation between image points and world points.

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

$$\lambda_{ij} m_{ij} = \mathbf{P}_j M_i \tag{1}$$

Here P_j is a 3x4 camera matrix, m_{ij} and M_i are column vectors containing the homogeneous coordinates of the image points resp. world points, λ_{ij} expresses the equivalence up to a scale factor. If $\mathbf{P}_j$ represents a Euclidean camera, it can be put in the following form [7]:

$$\mathbf{P}_j = \mathbf{K}_j \left[\mathbf{R}_j | -\mathbf{R}_j t_j\right] \tag{2}$$

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

$$\mathbf{K}_j = \begin{bmatrix} r_x^{-1} & -r_x^{-1}\cos\theta & f_j^{-1}u_x \\ & r_y^{-1} & f_j^{-1}u_y \\ & & f_j^{-1} \end{bmatrix} \tag{3}$$

In this equation r_x and r_y represent the pixel width and height, θ is the angle between the image axes, u_x and u_y are the coordinates of the principal point, and f_j 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 $\mathbf{K}_j$ will be decomposed in two parts:

$$\mathbf{K}_j = \mathbf{K}_{f_j}\mathbf{K} = \begin{bmatrix} 1 & 0 & (f_1/f_j - 1)u_x \\ 0 & 1 & (f_1/f_j - 1)u_y \\ 0 & 0 & f_1/f_j \end{bmatrix} . \begin{bmatrix} r_x^{-1} & -r_x^{-1}\cos\theta & f_1^{-1}u_x \\ 0 & r_y^{-1} & f_1^{-1}u_y \\ 0 & 0 & f_1^{-1} \end{bmatrix} \tag{4}$$

The second part $\mathbf{K}$ is equal to the calibration matrix $\mathbf{K}_1$ for view 1, whereas $\mathbf{K}_{f_j}$ models the effect of changes in focal length (i.e. zooming and focusing). From equation (4) it follows that once the principal point u is known, $\mathbf{K}_{f_j}$ is known for any given value of f_j/f_1. 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 $\mathbf{K}_{f_j}^{-1}$.

The first thing to do is to retrieve the principal point u. Fortunately this is easy for a camera equiped with a zoom. Upon changing the focal length (without moving the camera or the scene), each image point according to the pinhole camera model will 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. In practice these lines will not intersect precisely and a least squares approach is used to determine the principal point. This method has been used by others [17, 8, 9].

For the sake of simplicity we will assume $\mathbf{R}_1 = \mathbf{I}$,$t_1 = 0$ and $f_1 = 1$ 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 we fix the 7 degrees of freedom of a similarity transform.

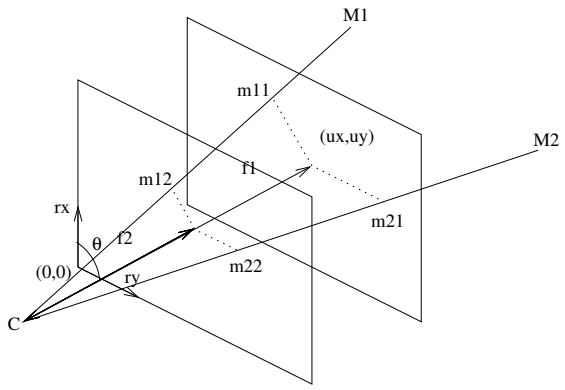

Figure 1: *Illustration of the camera and zoom model. The focal lengths f_1 and f_2 are different, the other parameters ($r_x, r_y, u_x, u_y, \theta$) are identical.*

3 affine structure from translation

It is possible to recover the affine structure of a scene from images taken by a translating camera [11]. This result can also be obtained when the focal length is not constant. Consider two perspective images with a camera translation and possibly a change in focal length in between.

3.1 recovering focal length for translation

There are different methods to recover the change in focal length. We first give a straightforward method based on the movement of the epipoles[3]. The performance of this method degrades very fast with noise on the image correspondences. Therefore an alternative, non-linear method was developed, that uses all available constraints. This method gives good results even in the presence of noise.

The first method is based on the fact that epipoles move on a line passing through the principal point u when the focal length is changed while the camera is translated and that this movement is related to the magnitude of this change. The following equation follows from the camera model:

$$\lambda_{e_{21}} e_{21} = -\lambda_{e_{12}} (e_{12} + (f_2^{-1} - 1)u) \tag{5}$$

where e_{21} is the epipole in the second image and e_{12} the epipole in the first image. e_{21}, e_{12} and u are column vectors of the form $[x\, y\, 1]^\top$. From equation (5) f_2^{-1} can be solved in a linear way. This method is suboptimal in the sense that it does not take advantage of the translational camera motion.

Determining the epipoles for two arbitrary images is a problem with 7 degrees of freedom. In the case of a translation (without changing the focal length) between two views,

[3] an epipole is the projection of the optical center of one camera in the image plane of the other camera. The epipoles can be retrieved from at least 7 point correspondences between the two images [5].

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 (2 degrees of freedom). 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 as outlined next.

The following equations describe the projection from world point coordinates M_i to image projection coordinates $m_{i1,2}$ for both images

$$\begin{aligned} \lambda_{i1} m_{i1} &= \mathbf{K}[\,\mathbf{I}\,|\,0\,]M_i & (6)\\ \lambda_{i2} m_{i2} &= \mathbf{K}_{f_2}\mathbf{K}[\,\mathbf{I}\,|\,-t_2]M_i \\ &= \mathbf{K}_{f_2}\mathbf{K}[\,\mathbf{I}\,|\,0\,]M_i + \lambda_{e_{21}} e_{21} \\ &= \lambda_{i1}\left(m_{i1} + (f_2^{-1} - 1)u\right) + \lambda_{e_{21}} e_{21} & (7) \end{aligned}$$

where m_{i1}, m_{i2}, u and e_{21} are column vectors of the form $[x\,y\,1]^\top$. Equation (7) gives 3 constraints for every point. If f_2 is known this gives a linear set of equations in $2n+3$ unknowns ($\lambda_{i1}, \lambda_{i2}, \lambda_{e_{21}} e_{211}, \lambda_{e_{21}} e_{212}, \lambda_{e_{21}} e_{213}$). Because all unknowns of equation (7) (except f_2) comprise a scale factor, n must satisfy $3n \geq 2n + 3 - 1$ to have enough equations. To also solve for f_2 one needs at least one more equation, which means that at least 3 point correspondences are needed to find the relative focal length f_2 (remember that $f_1 = 1$).

For every value of f_2 one could try to solve the set of equations (7) by taking the singular value decomposition of the corresponding matrix. If f_2 has the correct value there will be a solution and the smallest singular value should be zero. With noisy data there will not be an exact solution anymore, but the value of f_2 which yields the smallest singular value will be the best solution in a least squares sense. For this paper the Decker-Brent algorithm was used to minimise the smallest singular value with respect to the relative focal length f_2. This gives very good results. Thanks to the fact that a non-linear optimisation algorithm in only one variable was used no convergence problems were encountered.

3.2 affine reconstruction

knowing the focal length, one can start the actual affine reconstruction. Notice that it follows from equation (6) that M_i is related to $\begin{bmatrix} \lambda_{i1} m_{i1} \\ 1 \end{bmatrix}$ by the affine transformation $\begin{bmatrix} \mathbf{K} & 0 \\ 0 & 1 \end{bmatrix}$. So it suffices to recover the λ_{i1} from equation (7) to have an affine reconstruction of the scene.

4 Euclidean structure from affine structure and supplementary camera motion

In this section the upgrade of the reconstruction from affine to Euclidean by using a supplementary image taken with a different orientation is discussed. Once an affine reconstruction of the scene is known the same constraints as in [7, 1, 18] can be used. Here they are less easy to use, because the focal length also appears in these constraints. Therefore one first has to find the relative change in focal length.

4.1 recovering focal length for a supplementary view

In this paragraph a method will be explained that allows to recover the relative focal length of a camera in any position and with any orientation (relative to the focal length of the camera at the beginning). This can be done by starting from an affine reconstruction. Choosing the first camera matrix to be $[\,\mathbf{I}\,|\,0\,]$ the second camera matrix associated to our affine reconstruction is uniquely defined up to scale [15]. In the following equations the relationship between the affine camera matrixes $\mathbf{P}_{1A}, \mathbf{P}_{3A}$ and the Euclidean ones is given:

$$\begin{aligned}
\mathbf{P}_{1A} &= [\,\mathbf{I}\,|\,0\,] = \mathbf{K}[\,\mathbf{I}\,|\,0\,] \begin{bmatrix} \mathbf{K}^{-1} & 0 \\ 0 & 1 \end{bmatrix} \\
\mathbf{P}_{3A} &\equiv [\tilde{\mathbf{P}}_{3A}|\,.\,] = \lambda_{\mathbf{P}_3}\mathbf{K}_{f_3}\mathbf{K}[\mathbf{R}_3|-\mathbf{R}_3 t_3] \begin{bmatrix} \mathbf{K}^{-1} & 0 \\ 0 & 1 \end{bmatrix} \\
&= \lambda_{\mathbf{P}_3}\mathbf{K}_{f_3}[\mathbf{K}\mathbf{R}_3\mathbf{K}^{-1}|\,.\,] \qquad (8)
\end{aligned}$$

By definition $\mathbf{K}\mathbf{R}_3\mathbf{K}^{-1}$ is conjugated to $\mathbf{R}_3$ and hence will have the same eigenvalues which for a rotation matrix all have modulus 1 (one of them is real and both others are complex conjugated or real). This will be called the *modulus* constraint in the remainder of this paper. From equation (8) it follows that $\tilde{\mathbf{P}}_{3A}$ is related to $\mathbf{K}\mathbf{R}_3\mathbf{K}^{-1}$ in the following way:

$$\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A} = \lambda_{\mathbf{P}_3}\mathbf{K}\mathbf{R}_3\mathbf{K}^{-1} \qquad (9)$$

with

$$\mathbf{K}_{f_3}^{-1} = \begin{bmatrix} 1 & 0 & (f_3-1)u_x \\ & 1 & (f_3-1)u_y \\ & & f_3 \end{bmatrix}$$

The characteristic equation of $\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A}$ is as follows:

$$\det\left(\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A} - \lambda\mathbf{I}\right) = a\lambda^3 + b\lambda^2 + c\lambda + d = 0 \qquad (10)$$

For a detailed elaboration the reader is referred to Appendix A.1. The *modulus* constraint imposes $|\lambda_1| = |\lambda_2| = |\lambda_3|\,(= \lambda_{\mathbf{P}_3})$. From this one gets the following constraint (see Appendix A.2):

$$ac^3 = b^3 d \qquad (11)$$

Substituting the left hand side of equation (9) in equation (10), yields first order polynomials in f_3 for a, b, c, d. Substituting these in equation (11), one obtains a 4^{th} order polynomial in f_3.

$$a_4 f_3^4 + a_3 f_3^3 + a_2 f_3^2 + a_1 f_3 + a_0 = 0 \tag{12}$$

This gives 4 possible solutions. One can see that if f_3 is a real solution, then $-f_3$ must also be a solution[4]. Filling this in in equation (12) one gets the following result (see Appendix A.3).

$$f_3 = \pm\sqrt{\frac{a_1}{a_3}} \tag{13}$$

where the sign is dependent on camera geometry and is known[5]. One can conclude this paragraph by stating that the relative focal length f of any view with respect to a reference view can be recovered for any Euclidean motion.

4.2 Euclidean reconstruction

To upgrade the reconstruction to Euclidean the camera calibration matrix $\mathbf{K}$ is needed. This is equivalent to knowing the image $\mathbf{B}$ of the dual of the absolute conic for the first camera, since $\mathbf{B} = \mathbf{K}\mathbf{K}^\top$. Images are constrained in the following way:

$$\kappa_{13}\mathbf{B}_3 = \mathbf{H}_{13\infty}\mathbf{B}\mathbf{H}_{13\infty}^\top \tag{14}$$

with $\mathbf{B}_3 = \mathbf{K}_3\mathbf{K}_3^\top$ the inverse of the image of the absolute conic in the third image and $\mathbf{H}_{13\infty}$ the infinity homography[6] between the two images. This would be a set of linear equations in the coefficients of $\mathbf{B}$ if κ_{13} was known. This can be achieved by imposing equal determinants for the left and right hand side of equation (14). But before doing this it is interesting to decompose $\mathbf{B}_3$:

$$\mathbf{B}_3 = \mathbf{K}_3\mathbf{K}_3^\top = \mathbf{K}_{f_3}\mathbf{K}\mathbf{K}^\top\mathbf{K}_{f_3}^\top = \mathbf{K}_{f_3}\mathbf{B}\mathbf{K}_{f_3}^\top \tag{15}$$

From equation (15) one finds an equation for the determinant of $\mathbf{B}_3$ and by imposing the equality with the determinant of the right hand side of equation (14), the following equation is obtained.

$$\det\mathbf{B}_3 = (\det\mathbf{H}_{13\infty})^2\det\mathbf{B} = (\det\mathbf{K}_{f_3})^2\det\mathbf{B} \tag{16}$$

Equation (16) will hold if the following equation holds:

$$\det\mathbf{H}_{13\infty} = \det\mathbf{K}_{f_3} \equiv f_3^{-1}, \tag{17}$$

[4]This is because the only constraint imposed is the *modulus* constraint (same modulus for all eigenvalues). If the real part of λ_2 and λ_3 have opposite sign then $\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A}$ does not represent a rotation but a rotation and a mirroring. Changing the sign of f_3 has the same effect.

[5]for a non-mirrored image the sign must be positive.

[6]The infinity homography, which is a plane projective transformation, maps vanishing points from one image to the corresponding points in another image.

when f_3 has been obtained following the principles outlined in section 4.1. This constraint can easily be imposed because $\mathbf{H}_{13\infty}$ is only determined up to scale. The following equations (derived from equations (14) and (15)) together with the knowledge of u and f_3 then allows to calculate $\mathbf{B}$ (and $\mathbf{K}$ by cholesky factorisation).

$$\mathbf{K}_{f_3}\mathbf{B}\mathbf{K}_{f_3}^{\top} = \mathbf{H}_{13\infty}\mathbf{B}\mathbf{H}_{13\infty}^{\top} \tag{18}$$

This approach could be simplified by assuming that the camera rows and columns are perpendicular ($\theta = 90^o$)[18]. In that case equation (18) boils down to an overdetermined system of linear equations in r_x^{-2} and r_y^{-2} which gives more stable results. r_x and r_y being the only unknowns left, one will also have $\mathbf{K}$. Finally the affine reconstruction can be upgraded to Euclidean by applying the following transformation

$$\mathbf{T}_{AE} = \begin{bmatrix} \mathbf{K}^{-1} & 0 \\ 0 & 1 \end{bmatrix} \tag{19}$$

5 Euclidean calibration of a fixed stereo rig

The auto-calibration techniques proposed by Zisserman [18] and Devernay [3] for two views taken with a rotating fixed stereo rig can also be generalised to allow changes in focal lengths for both cameras independently and purely translational motions. In fact the method is easier than for a single camera.

For a fixed stereo rig the epipoles are fixed as long as one doesn't change the focal length. In this case the movement of the epipole in one camera is in direct relation with the change of its focal length. This is illustrated in figure 2. Knowing the relative change in focal length and the principal points allows to remove the effect of this change from the images. From then on the techniques of Zisserman [18] or Devernay [3] can be applied.

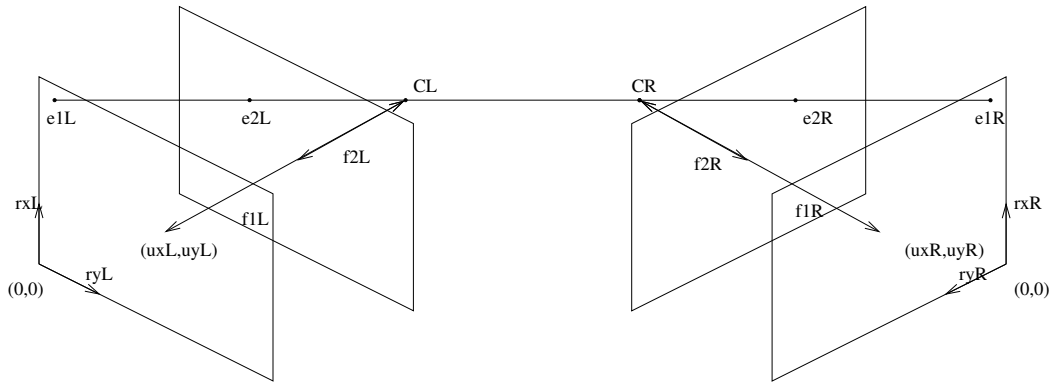

Figure 2: *this figure illustrates how the epipoles will move in function of a change in focal length.*

By first extracting the principal points -i.e. mildly calibrating the camera - one can then also get a Euclidean reconstruction even for a translating stereo rig, which was not

possible with earlier methods [18, 3]. Between any pair of cameras i and j we have the following constraints:

$$\kappa_{ij}\mathbf{B}_j = \mathbf{H}_{ij\infty}\mathbf{B}_i\mathbf{H}_{ij\infty}^{\top} \tag{20}$$

For two views with a fixed stereo rig there are 3 different constraints of the type of equation (20): for the left camera (between view 1 and 2), for the right camera (between view 1 and 2) and between the left and the right camera. For a translation $\mathbf{H}_{12\infty} = \mathbf{I}$ which means that the two first constraints become trivial. The constraint between the left and the right camera in general gives 6 independent equations[7]. This is not enough to solve for $r_{Lx}, r_{Ly}, u_{Lx}, u_{Ly}, \theta_L, r_{Rx}, r_{Ry}, u_{Rx}, u_{Ry}$ and θ_R. Knowing the principal points restricts the number of unknowns to 6, which could be solved from the available constraints. Assuming perpendicular images axes [18] one can solve for the 4 remaining unknowns in a linear way (see appendix). In practical cases this is very important because with the earlier techniques any movement close to translation gives unstable results which isn't the case anymore for this technique.

It is also useful to note that in the case of a translational motion of the rig, the epipolar geometry can be obtained with as few as 3 points seen in all 4 views. Superimposing their projections in the focal length corrected second views onto the first, it is as if one observes two translated copies of the points. Choosing two of the three points, one obtains four coplanar points from the two copies (coplanarity derives from the fact that the rig translates). Together with projections of the third point, this suffices to apply the algorithm propound in [2]. Needing as few as 3 points clearly is advantagous to detect e.g. independent motions using RANSAC strategies [16]

6 Results

In this section some results obtained with the single camera algortithm are presented. First an analysis of the noise resistance of the algorithm is given based on synthetic data with variable levels of noise. A reconstruction of a real scene is given as well.

6.1 synthetic data

Altough synthetic data were used to perform the experiments in this paragraph, due attention has been paid to mimic real data. A simple house shape was chosen as scene and the "camera" was given realistic parameter values (see figure 3). From this sequence 320x320 disparity maps were generated. These maps were altered with different amounts of noise to see how robust the method is. In figure 4 the results of the experiments are plotted. The error on the different camera parameters as calculated by the algorithm are plotted against the standard deviation of the noise on the correspondences (in pixels). One can see that the focal length change between the first two images (translation) can be recovered very accurately. The non-linear method was used to obtain f_2/f_1. The focal length for the

[7] the cameras of the stereo rig should not have the same orientation.

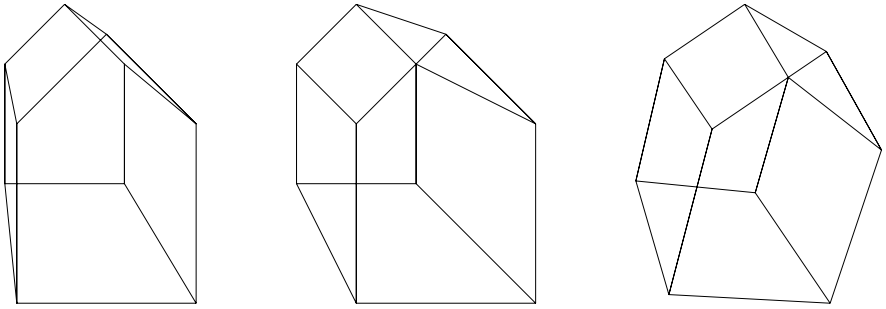

Figure 3: *three views that were used for the simulations*

third image is much more sensitive to noise, but this doesn't seem to influence very much the calculation of r_x^{-1} or r_y^{-1}. This is probably due to the fact that the set of equations (18) gives us 6 independent equations for only 2 unknowns.

The influence of a bad localisation of the principal point u was also analysed. The errors on the estimated parameters f_2, f_3, r_x^{-1} and r_y^{-1} came out to be of the same order as the error on u, which in practice was small when determined from zooming.

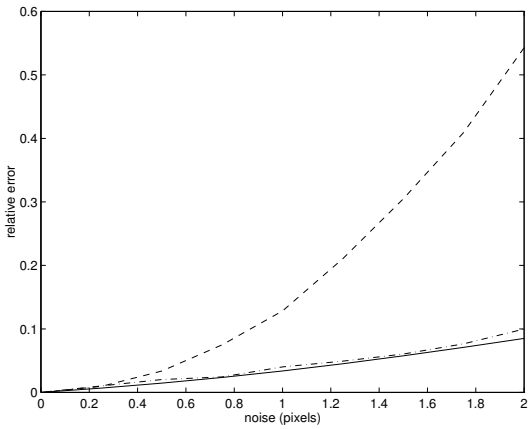

Figure 4: *Results of the tests on synthetic data: relative errors plotted against noise level for different camera parameters* (solid line: f_2 , dashed line: f_3 , dash-dotted line: r_x^{-1} and r_y^{-1})

From these experiments one sees that the Euclidean calibration of the camera and hence also the reconstruction degrades gracefully in the presence of noise. This indicates that the presented method is usable in practical circumstances. This will be illustrated in the next paragraph.

6.2 real images

Here some results obtained from a real scene are presented. The scene consisted of a cornflakes boxes, a lego box and a cup. The images that were used can be seen in figure 5. The scene was chosen to allow a good qualitative evaluation of the Euclidean reconstruction. The boxes have right angles and the cup is cylindrical. These characteristics must be preserved by a *Euclidean* reconstruction, but will in general not be preserved by an *affine* or *projective* reconstruction. To build a reconstruction one first needs correspondences

Figure 5: *The 3 images that were used to build a Euclidean reconstruction. The camera was translated between the first two views (the zoom was used to keep the size more or less constant). For the third image the camera was also rotated.*

between the images. INRIA's public corner matcher was used to extract these correspondences. A total of 99 correspondences were obtained between the first two images. From these the affine calibration of the camera was computed. From the first up to the third image 34 correspondences could be tracked. The corresponding scene points were reconstructed from the first two images which allowed to find an affine calibration for the camera in the third position. From this the method described in section 5 to find the Euclidean calibration of the camera was used. Subsequently, the output of an algorithm to compute dense point correspondences was used to generate a more complete reconstruction. This algorithm yields a pointwise correspondence and confidence level. Only points with a confidence level above a fixed threshold were used for the reconstruction.

Figure 6 shows two views of the reconstructed scene[8]. The left image is a front view while the right image is a top view. Note , especially from the top view, that 90^o angles are preserved and that the cup keeps its cylindrical form which is an indication of the quality of the *Euclidean* reconstruction. Figure 7 shows a further view, both shaded and texture mapped to indicate the consistency with the original image texture.

[8]We prefered shading to the projection of the original texture on the model because this gives a better impression of the 3D structure.

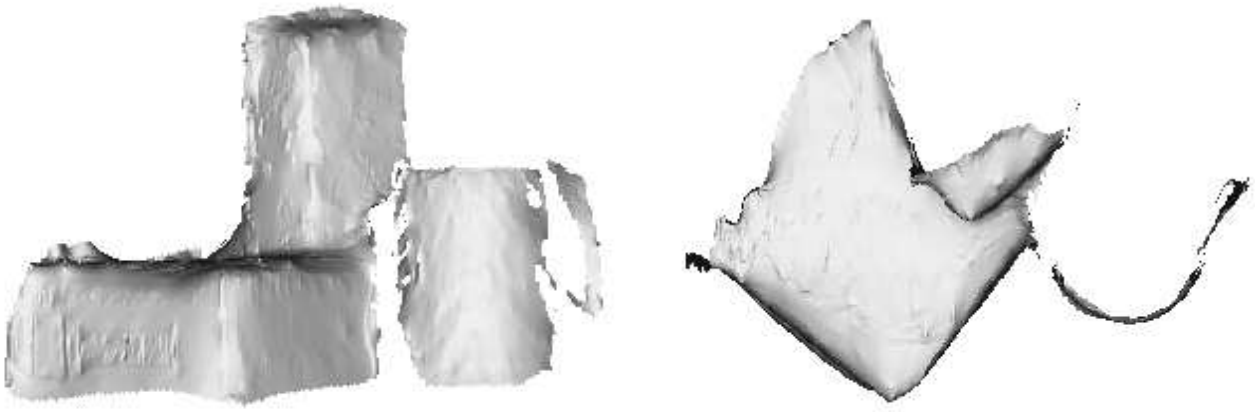

Figure 6: *front and top view of the reconstruction.*

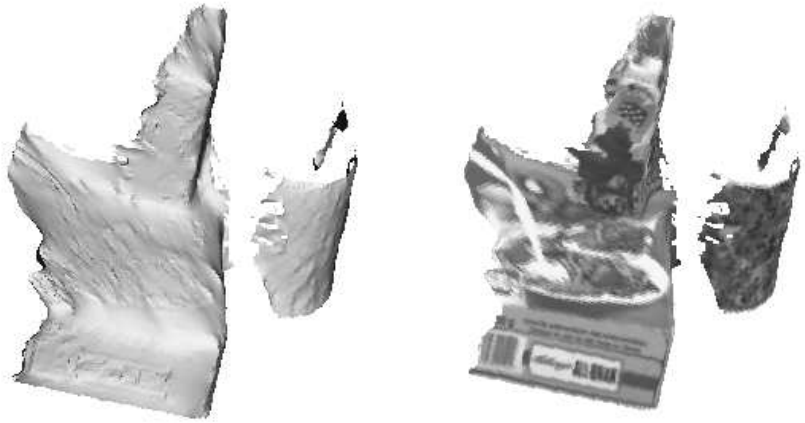

Figure 7: *side views of the reconstructed scene (with shading and with texture).*

7 conclusions and future work

In this paper the possibility to obtain the auto-calibration of both a single moving camera and a moving stereo-rig was demonstrated and this without the need of keeping all internal parameters constant. The complete method for a single camera was described. From the experiments one can conclude that this method is relatively stable in the presence of noise. This makes it suitable for practical use. Also a method for auto-calibration of a fixed stereo rig with independently zooming cameras was briefly presented. An additional advantage of this method is that it is even suitable for a *purely translating* stereo rig, whereas previous methods required a rotational motion component.

We plan to enhance the implementations of both the single camera and the stereo rig calibration algorithm. The input of more correspondences in the auto-calibration stage would certainly yield better results. We will also look at other possibilities of the *modulus*

constraint (see section 4.1) which must hold for a camera in any position in an affine camera reference frame[9]. We could use this constraint to calculate any unknown parameter of a camera. Using different views one could solve for more parameters, like for example the *affine* calibration of the matrix itself. This could be intresting because it allows to work in a 3 parameter space in stead of the 8 parameters that Hartley [7] had to solve for at once.

A determining the focal length for a general motion

A.1 equation for eigenvalues from $\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A}$

Here equation (10) is elaborated in detail:

$$\det\left(\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A} - \lambda I\right) = a\lambda^3 + b\lambda^2 + c\lambda + d = 0$$

yields with p_{ij} as the coefficients of $\tilde{\mathbf{P}}_{3A}$

$$\begin{aligned}
a &= -1 \\
b &= (u_x p_{31} + u_y p_{32} + p_{33})f + p_{11} + p_{22} - u_x p_{31} - u_y p_{32} \\
c &= (u_x(p_{21}p_{32} - p_{31}p_{22}) + u_y(p_{31}p_{12} - p_{11}p_{32}) + p_{31}p_{13} + p_{32}p_{23} - p_{11}p_{33} - p_{22}p_{33})f^2 \\
&\quad + u_x(p_{31}p_{22} - p_{21}p_{32}) + u_y(p_{11}p_{32} - p_{31}p_{12}) + p_{21}p_{12} - p_{11}p_{22} \\
d &= \det\tilde{\mathbf{P}}_{3A} f
\end{aligned} \tag{21}$$

A.2 constraint on a, b, c, d

The roots of equation (10) must obey $|\lambda_1| = |\lambda_2| = |\lambda_3|$. A general condition on a, b, c, d for this to hold is derived next. Afterwards we can fill in the equations of Appendix A.1. A general third order polynomial looks as follow.

$$a\lambda^3 + b\lambda^2 + c\lambda + d = a(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \tag{22}$$

From equation (22) the following relations follow:

$$\lambda_1 + \lambda_2 + \lambda_3 = -\frac{b}{a} \tag{23}$$

$$\lambda_1(\lambda_2 + \lambda_3) + \lambda_2\lambda_3 = \frac{c}{a} \tag{24}$$

$$\lambda_1\lambda_2\lambda_3 = -\frac{d}{a} \tag{25}$$

We want to derive a necessary condition for $|\lambda_1| = |\lambda_2| = |\lambda_3|$. If we choose λ_1 to be real (λ_2 and λ_3 can be either real or complex), the following equivalence must be true.

$$\lambda_1^2 = \lambda_2\lambda_3 \tag{26}$$

[9] it must be the same affinely calibrated camera for all views and the camera matrix must be $[\mathbf{I}|0]$ for the first view.

Rewriting (24) using (23) and (26) yields

$$\lambda_1(-\frac{b}{a} - \lambda_1) + \lambda_1^2 = \frac{c}{a} \tag{27}$$

or

$$\lambda_1 = -\frac{c}{b} \tag{28}$$

substituting (26) in (25) implies

$$\lambda_1^3 = -\frac{d}{a} \tag{29}$$

Eliminating λ_1 from the equations (28) and (29) gives a necessary condition that is only depending on a, b, c, d.

$$ac^3 = b^3 d \tag{30}$$

In equation (30) we can substitute the equations for a, b, c, d derived as Appendix A.1, yielding a 4^{th} order polynomial in f.

A.3 analysis of the roots of $a_4 f_3^4 + a_3 f_3^3 + a_2 f_3^2 + a_1 f_3 + a_0 = 0$ (Eq. 12)

The constraint $ac^3 = b^3 d$ (Eq. (11)) was obtained by imposing equal moduli to the eigenvalues of $\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A}$ (the *modulus* constraint). If f_3 is a real solution then $-f_3$ will also be a solution. Changing the sign of the focal length is equivalent to a point reflection of the image around the principal point, which means that the moduli of the eigenvalues of $\mathbf{K}_{f_3}^{-1}\tilde{\mathbf{P}}_{3A}$ will stay the same (only signs can change). What does this mean for the coefficients of equation (12)? We choose λ_1 and $-\lambda_1$ to be the real roots[10].

$$a_4(f_3^2 - \lambda_1^2)(f_3^2 + bf_3 + c) = 0 \tag{31}$$

$$a_4(f_3^4 + bf_3^3 + (c - \lambda_1^2)f_3^2 - \lambda_1^2 bf_3 - \lambda_1^2 c) = 0 \tag{32}$$

From equation (32) one can easily obtain λ_1 which is the desired solution for f_3.

$$f_3 = \pm\sqrt{\frac{a_1}{a_3}} \tag{33}$$

Here a_1 and a_3 are the coefficients of the first order resp. third order term of equation (32). These coefficients are obtained by filling in a, b, c, d from equation (21) in equation (11).

References

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

[10] a different real root λ_2 would imply $-\lambda_2$ to be a solution too. This would lead to $b = 0$ and thus also $a_3 = 0$ and $a_1 = 0$ in eq.(12). In practice we only encountered 4 real roots for pure translation. Three were identical and one had opposite sign.

[2] B.Boufama, R. Mohr, Epipole and fundamental matrix estimation using virtual parallax, *Proc. ICCV'95*, pp.1030-1036, 1995

[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. 2nd European Conf. Computer Vision*, pp.321-334, 1992.

[5] R. Deriche, Z. Zhang, Q.-T. Luong, O. Faugeras, Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig *Computer Vision - ECCV'94*, Lecture Notes in Computer Science **800**, pp. 567–576, Springer-Verlag, Berlin / Heidelberg, 1994.

[6] R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc. 2nd European Conf. Computer Vision*, Santa Margherita, Italy, pp.579-587, 1992.

[7] 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, Berlin / Heidelberg / New York, 1994.

[8] J.M. Lavest, G. Rives, and M. Dhome. 3D reconstruction by zooming. *IEEE Robotics and Automation*, 1993

[9] M. Li, Camera Calibration of a Head-Eye System for Active Vision *Computer Vision - ECCV'94*, Lecture Notes in Computer Science **800**, pp. 543–554, Springer-Verlag, Berlin / Heidelberg, 1994.

[10] Q.T. Luong and T. Vieville. Canonic representations for the geometries os multiple projective views. In *Proc. 3rd European Conference on Computer Vision*, pages 589-597. Springer-Verlag, 1994.

[11] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, Affine reconstruction from perspective image pairs, Proc. Workshop on Applications of Invariance in Computer Vision II, pp.249-266, 1993

[12] M. Pollefeys, L. Van Gool and M. Proesmans, Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths, *ECCV'96*, Lecture Notes in Computer Science, Springer-Verlag, 1996.

[13] M. Pollefeys, L. Van Gool, and T. Moons, Euclidean 3D Reconstruction from Stereo Sequences with Variable Focal Lengths *Recent Developments in Computer Vision*, Lecture Notes in Computer Science, pp.405-414, Springer-Verlag, 1996.

[14] M. Proesmans, L. Van Gool and A. Oosterlinck, Determination of optical flow and its discontinuities using non-linear diffusion, *Proc. ECCV'94*, pp. 295-304, 1994.

[15] 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

[16] P.H.S. Torr, Motion Segmentation and Outlier Detection, *Ph.D.Thesis*, Oxford 1995

[17] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323-331, August 1987.

[18] 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