Linear calibration of a rotating and zooming camera *

Lourdes de Agapito, Richard I. Hartley and Eric Hayman
Department of Engineering,
Oxford University
and
G.E. Corporate Research and Development
1 Research Circle,
Niskayuna, NY 12309

Abtract

A linear self-calibration method is given for computing the calibration of a stationary but rotating camera. The internal parameters of the camera are allowed to vary from image to image, allowing for zooming (change of focal length) and possible variation of the principal point of the camera. In order for calibration to be possible some constraints must be placed on the calibration of each image. The method works under the minimal assumption of zero-skew (rectangular pixels), or the more restrictive but reasonable conditions of square pixels, known pixel aspect ratio, and known principal point. Being linear, the algorithm is extremely rapid, and avoids the convergence problems characteristic of iterative algorithms.

1 Introduction

The subject of self-calibration of a camera has received considerable attention, following the groundbreaking paper of Maybank and Faugeras [7]. This idea opens the possibility of calibration of a camera in the field, without the aid of jigs, or knowledge of the position of world-points. A subsequent paper of Hartley ([4]) gave a method for the self-calibration of a rotating, but stationary camera, for which the theory of [7] does not apply. The algorithm of [4] had the advantage of being linear and hence very simple and rapid, unlike the Maybank-Faugeras method, which was somewhat complex.

The methods of these two papers required the calibration of the camera to be fixed over a sequence of images – no zooming was allowed. Subsequently interest in zooming cameras led to a method for self-calibration of cameras with changing internal parameters ([5]). These results were strengthened in [6, 8] to apply in case of a minimal assumption of zero-skew. These methods applied to moving cameras, and were iterative[1].

Recent papers ([1, 9]) have extended calibration capability by giving algorithms for rotating cameras, allowing changing internal parameters, similar in intent to the algorithms of [5, 6, 8]. The particular scenario of a rotating and zooming camera is common in practice. For instance a camera at a sports arena undergoes this sort of motion as it rotates and zooms to follow a game. Unlike the algorithm of Hartley ([4]), these methods use interation, based on the Levenberg-Marquardt method to minimize a non-linear cost function. The linear methods of ([4]) do not appear to be immediately extendable to the case of varying internal parameters.

In this paper however it is shown that a simple trick, first used in a different context in [11] transfers the basic calibration equation ((3) below) to one for which a simple linear method applies. The resulting linear method is extremely simple involving a least-squares solution of a set of homogeneous equations in 6 unknowns. Each image in the sequence leads to one to four equations, depending on the amount of assumed knowledge of the camera. It turns out that the linear method, apart from being quicker than the iterative methods gives results of comparable quality.

2 Rotating cameras

We consider a set of images taken with cameras all located at the same point in space, which will be taken to be the coordinate origin. As has been shown in (for instance) [4] one may analyse this situation by representing each of the cameras as a 3×3 matrix $\mathtt{P}_i$. A

*This work was sponsored by DARPA contract F33615-94-C-1549

[1]It may be noted that these papers were all predated by the paper [3] that gave a non-iterative algorithm for two-view calibration in the case of known aspect ratio and principal point.

point in the i-th image, represented by a homogeneous 3-vector $\mathbf{x}_i$ corresponds to a ray in space consisting of points of the form $\lambda \mathtt{P}_i^{-1}\mathbf{x}_i$. Points on this ray are mapped into the j-th image to a point $\mathbf{x}_j = \mathtt{P}_j\mathtt{P}_i^{-1}\mathbf{x}_i$. Denoting the transformation

$$\mathtt{H}_{ij} = \mathtt{P}_j\mathtt{P}_i^{-1} \tag{1}$$

one sees that the i-th and j-th images are related by a projective transformation $\mathtt{H}_{ij}$. In practice, one may compute these projective transformations between images by finding matching points in the two images and computing the projective transformation that relates the points. At least four matched points are necessary for computing the projective transformation between two images. Practical methods for computing these transformations are given in [4, 2].

The projective transformation may be related to the calibration matrices of each of the cameras, as follows. Each camera matrix $\mathtt{P}_i$ may be decomposed into an upper-triangular calibration matrix $\mathtt{K}_i$ and a rotation matrix $\mathtt{R}_i$ representing the orientation of the camera : $\mathtt{P}_i = \mathtt{K}_i\mathtt{R}_i$. Substituting in (1) one obtains

$$\mathtt{H}_{ij} = \mathtt{K}_j\mathtt{R}_j\mathtt{R}_i^{-1}\mathtt{K}_i^{-1} = \mathtt{K}_j\mathtt{R}_{ij}\mathtt{K}_i^{-1} \quad . \tag{2}$$

Since $\mathtt{R}_{ij}$ is a rotation, $\mathtt{R}_{ij}{\mathtt{R}_{ij}}^\top = \mathtt{I}$. Straight-forward computation then shows that $\mathtt{H}_{ij}\mathtt{K}_i{\mathtt{K}_i}^\top{\mathtt{H}_{ij}}^\top = \mathtt{K}_j{\mathtt{K}_j}^\top$. This formula is often written as

$$\mathtt{H}_{ij}\omega_i^*{\mathtt{H}_{ij}}^\top = \omega_j^* \tag{3}$$

where $\omega_i^* = \mathtt{K}_i{\mathtt{K}_i}^\top$ is the dual image of the absolute conic ([4]) in the i-th image. This formula then represents the transformation of a conic under a projective transformation. In the case where $\mathtt{K}_i = \mathtt{K}_j$, this formula may be used to generate a set of linear equations in the entries of ω_i^* which may be used to solve for ω_i^*, and subsequently for $\mathtt{K}_i$ by Choleski factorization. This is briefly the calibration method described in [4].

The difficulty with extending this linear solution to the case of a camera with varying intrinsic parameters under minimal assumptions such as zero skew or known aspect ratio is that such constraints are not easily related to the entries of ω_i^*. Consider a calibration matrix

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \tag{4}$$

Three constraints are possible, of which the first two were tested in detail in this paper.

1. **The zero-skew constraint :** For each camera, $s = 0$.
2. **The square-pixel constraint :** For each camera, $s = 0$ and $\alpha_x = \alpha_y$.
3. **The known principal point constraint :** For each camera, $(x_0, y_0) = (0, 0)$.

If the pixels have a known aspect ratio other than 1, or the principal point is at a different know point other than the origin, then a simple change of image coordinates converts to one of the cases above.

Now, a constraint such as the zero-skew constraint is not reflected in any simple way in the entries of $\omega^* = \mathtt{K}\mathtt{K}^\top$. There seems to be no easy way to use (3) to develop an algorithm to compute the camera calibration, enforcing constraints of this type. Therefore, in [1, 9], iterative algorithms are proposed, using Levenberg-Marquardt iteration to find a non-linear least squares solution, parametrized directly by the entries of the $\mathtt{K}_i$. This may potentially cause problems with lack of convergence, or convergence to local minima, not to mention the greater complexity of coding.

A simple observation, however, leads to a linear algorithm. Taking the inverse of (3) gives

$${\mathtt{H}_{ij}}^{-\top}\omega_i\mathtt{H}_{ij}^{-1} = \omega_j \tag{5}$$

where $\omega_i = {\mathtt{K}_i}^{-\top}\mathtt{K}_i^{-1}$. Now, one may verify that with $\mathtt{K}$ of the form (4), with $s = 0$,

$$\begin{aligned} \omega &= \mathtt{K}^{-\top}\mathtt{K}^{-1} \\ &= \begin{bmatrix} 1/\alpha_x^2 & 0 & -x_0/\alpha_x^2 \\ 0 & 1/\alpha_y^2 & -y_0/\alpha_y^2 \\ -x_0/\alpha_x^2 & -y_0/\alpha_y^2 & 1 + x_0^2/\alpha_x^2 + y_0^2/\alpha_y^2 \end{bmatrix} \end{aligned} \tag{6}$$

so ω represents a conic of the form

$$(x - x_0)^2/\alpha_x^2 + (y - y_0)^2/\alpha_y^2 + 1 = 0 \ ,$$

the image of the absolute conic. The important points to note here are

Proposition 2.1. *1. If $s = \mathtt{K}_{12} = 0$, then $\omega_{12} = 0$.*

2. If $s = 0$ and $\alpha_x = \alpha_y$, then $\omega_{11} = \omega_{22}$.

3. If $s = 0$ and $x_0 = 0$, then $\omega_{13} = 0$. Similarly if $y_0 = 0$ then $\omega_{23} = 0$.

3 Generation and solution of equations

Select one image as being a reference image, and let $\mathtt{H}_{0j}$ be the homography relating the reference image to the j-th image. For $j = 0$, one has $\mathtt{H}_{00} = \mathtt{I}$. Let the image of the absolute conic in the reference image be ω_0. Since it is symmetric, it may be

parametrized by its six diagonal and above-diagonal entries, which may be denoted in some designated order as $(a_1, a_2, \ldots, a_6)$. Since the homographies $\mathtt{H}_{0j}^{-1}$ are known, the entries of ω_j may be expressed linearly in terms of the entries a_i of ω_0. Now each of the equation types in Proposition 2.1 gives one linear equation in the entries of ω_j, hence a linear equation in the entries a_i of ω_0. For each image each condition gives one equation, and the set of all equations may be written as $\mathtt{E}\mathbf{a} = 0$, where $\mathbf{a} = (a_1, a_2, \ldots, a_6)^\top$, and each row of $\mathtt{E}$ represents one equation. Given at least five equations one can find a solution up to (non-essential) scale. With more than 5 equations, one finds a least-squares solution in the usual way ([4]), that is by finding $\mathbf{a}$ that minimizes $||\mathtt{E}\mathbf{a}||$ subject to $||\mathbf{a}|| = 1$.

In forming these equations, one should include equations for the reference image. Consider the zero-skew condition. With $\mathtt{H}_{00} = \mathtt{I}$, the corresponding equation becomes simply $a_2 = 0$, assuming that a_2 reprents the $(1,2)$ entry of ω_0. This equation should be included with the equations arising from the other images in the matrix $\mathtt{E}$ of all equations. A (possibly inferior) alternative would be to parametrize ω in terms of only 5 entries, setting the $(1,2)$ entry to zero. Then each image other than the reference image gives one equation. The difference is that in this latter case, the computed skew will be exactly zero in the reference image but non-zero (because the equations will not be satisfied exactly) in all other images. This approach would unreasonably single out the reference image for special treatment. The same remarks apply in the square-pixel and known-principal point cases as well.

Since one requires at least 5 equations in the 6 independent homogeneous entries of ω_0, one requires a total of 5 images to solve in the zero-skew case, 3 images in the square pixel case, and only 2 images if one also knows the principal point.

By solving for the vector $(a_0, a_1, \ldots, a_6)$ one determines the value of ω_0. The other ω_j may be computed from (5). From this one may compute the individual calibration matrices $\mathtt{K}_j$ by inversion and Choleski factorization of ω_j, or else directly from ω_j using (6). The algorithm will fail if the computed value of α_x^2 or α_y^2 is negative, which is equivalent to ω_j not being positive definite. This rarely occurs in practice except in unstable situations in which calibration is intrinsically infeasible, or with inaccurate or incorrect input data.

4 Unstable configurations

In the experimental evaluation of these algorithms, it was observed that the results were quite unstable if only the zero-skew constraint is used, *unless* the camera motion contains some component of rotation about the pricipal axis (Z-axis) of the camera. If some Z axis motion is included, however, then results are good. Thus, to obtain good results, one must either have some Z-rotations, or else include square-pixel constraints. It should be realized that this failure mode does not represent a weaknesses in the linear algorithm of this paper, but rather arises from a situation in which self-calibration is intrinsically unstable.

X-axis rotation. Consider the case where the rotation is about the X axis of the camera and the cameras have zero skew. In addition, suppose for the present that the coordinate origin is at the principal point, so that the principal point $(x_0, y_0) = (0, 0)$. In this case the calibration matrix of each camera is diagonal, $\mathtt{K} = \mathrm{diag}(\alpha_x, \alpha_y, 1)$, and the transformation matrix $\mathtt{H}_{0j}^{-1} = \mathtt{K}_0 {\mathtt{R}_X}^\top \mathtt{K}_j$ is of the form

$$\mathtt{H}_{0j}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix}$$

where the x values represent non-zero entries, not all the same. Now (5) may be written as

$$\omega_j = {\mathtt{H}_{0j}}^{-\top} \omega_0 \mathtt{H}_{0j}^{-1} \tag{7}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_4 & a_5 \\ a_3 & a_5 & a_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix}. \tag{8}$$

From this one easily verifies that the expression for the $(1,2)$ entry of ω_j involves only the entries a_2 and a_3 of ω_0. Assuming that $a_2 = 0$ (the skew is zero) and referring to the expression (6) for ω_0, one deduces that the zero-skew constraint imposes a constraint only on x_0, allowing one to deduce that $x_0 = 0$. There is no constraint on y_0, α_x or α_y, which may vary freely. Thus, although $y_0 = 0$ is assumed in this analysis, there is no way this may be deduced using the zero-skew constraint.

One may verify that the choice of coordinate systems used to compute the homographies $\mathtt{H}_{0j}$ does not materially alter the constraints imposed on the calibration by a rotation about the X-axis, as analyzed above.

Y-axis rotation. In a similar way, if the rotation is about the Y axis, then the expression for $(\omega_j)_{12}$ involves only a_2 and a_5, as may be deduced in a similar manner. Thus, x_0, α_x and α_y may vary freely. If all rotations R_{0j} are about the X or Y axis, then there is

no way to determine the parameters α_x and α_y of the camera. If on the other hand, one includes a square-pixel constraints, or rotations about the Z-axis, then it turns out that the stability problems disappear.

Pan-tilt motion. A further important situation that leads to failure is that of a camera mounted on an alt-azimuth mounting, allowing panning about a vertical axis, and tilting about a horizontal elevation axis. In this case, the rotation of the camera with respect to a horizonal reference position may be written in the form $\mathtt{R}_x\mathtt{R}_y$, (pan followed by tilt). One easily verifies that this rotation matrix has the distinguishing property that the $(1,2)$ element is zero. Mimicking the analysis for X-axis rotation above, and mutiplying out as in (7) reveals that the $(1,2)$ element of ω_j does not depend on the element $a_4 = (\omega_0)_{22}$. According to (6), this implies that α_y is not determined. Furthermore, although element a_5 of ω_0 may be determined, it represents the quotient y_0/α_y^2. Hence y_0 depends on α_y, and so can not be determined either.

5 Experiments

In this section we present experimental results obtained using the linear algorithm on both synthetic and real image sequences.

5.1 Synthetic data

Experiments were first carried out with synthetic data to evaluate the performance of the linear algorithm using the *zero-skew* and the *square-pixel* constraints described in section 2. The data was created to simulate a camera with a zoom lense providing a total focal length range of 12.5 mm to 35 mm. A cloud of 250 points was randomly generated within a confined spheric space of diameter 2 m lying in front of the rotating camera at a distance of 5 m. The points were then projected onto each of the image planes arising from the different orientations of the camera and the location of each image point was then perturbed in a random direction by a distance governed by a Gaussian distribution with zero mean and standard deviation σ measured in pixels. The size of the image planes was 384×288 pixels. The skew of the image axes was taken to be zero, the aspect ratio of the image pixels equal to one and the principal point was asumed to be located at the centre of the image. The camera motion was such that the principal ray described a circular trajectory of radius $\theta = 5^\circ$ measured from the positive Z axis. To avoid the degenerate configuration described at the end of section 4 a small amount of cyclorotation about the principal axis was then added to each camera position. The focal length of the camera increased linearly throughout the sequence. Figure 1 shows the results of this experiment. They show that for typical image noise levels of σ smaller than 0.5 pixels the algorithm performs well. The plot for the principal point shows a large variation when $\sigma = 1$ pixel. However, it is well known that the principal point is a poorly constrained parameter which easily fits its value to the noise. Note also that the small deviation of the camera (5°) represents a very demanding test for the algorithm. With wider excursions of the camera, better results would be achieved ([4]).

In figure 1 we also show the performance of the linear algorithms versus the iterative non-linear algorithm described in [1]. The non-linear algorithm was run imposing equivalent constraints of *zero skew* and *square pixels* on the intrinsic parameters to minimize a non-linear cost function using a Levenberg-Marquardt method. For clarity we have only shown the result when the image noise level of $\sigma = 0.5$ pixels which shows that the results obtained with the linear algorithm are comparable with those from the non-linear iterative method.

5.2 Real image sequences

The image sequences used in our experiments were taken using a camera with a zoom lens mounted on a Yorick stereo head/eye platform [10]. The camera was rotated using one of the two independent vergence axes to pan the camera, and the common elevation axis to tilt it. Note therefore that the motion selected for these experiments falls in the case of the pan-tilt degeneracy described in section 4. The mechanics of our head do not permit rotations about the Z axis. This experiment will therefore illustrate how the *skew zero* constraint does not resolve this ambiguity and the *square pixels* constraint must be imposed.

Two image sequences were taken. In the first sequence, the focal length of the camera remained fixed, while the pan and the tilt of the camera were varied to perform a circular trajectory. In the second sequence, the focal length of the camera was set to increase linearly, using the controlled zoom lens, while the camera performed a similar circular motion. The encoders of the head/eye platform provided ground truth values for the pan and tilt angles of the camera which are accurate to 0.01 of a degree. The servo control of the zoom lens provided ground truth values of the position of the zoom lens for each frame in the image sequence. The camera was then calibrated, using an accurately machined calibration grid and classical calibration algorithm, to obtain ground truth values for the internal parameters at each of the different positions of the zoom lens. Radial lens distortion was modelled using a one parameter model and the images were appropri-

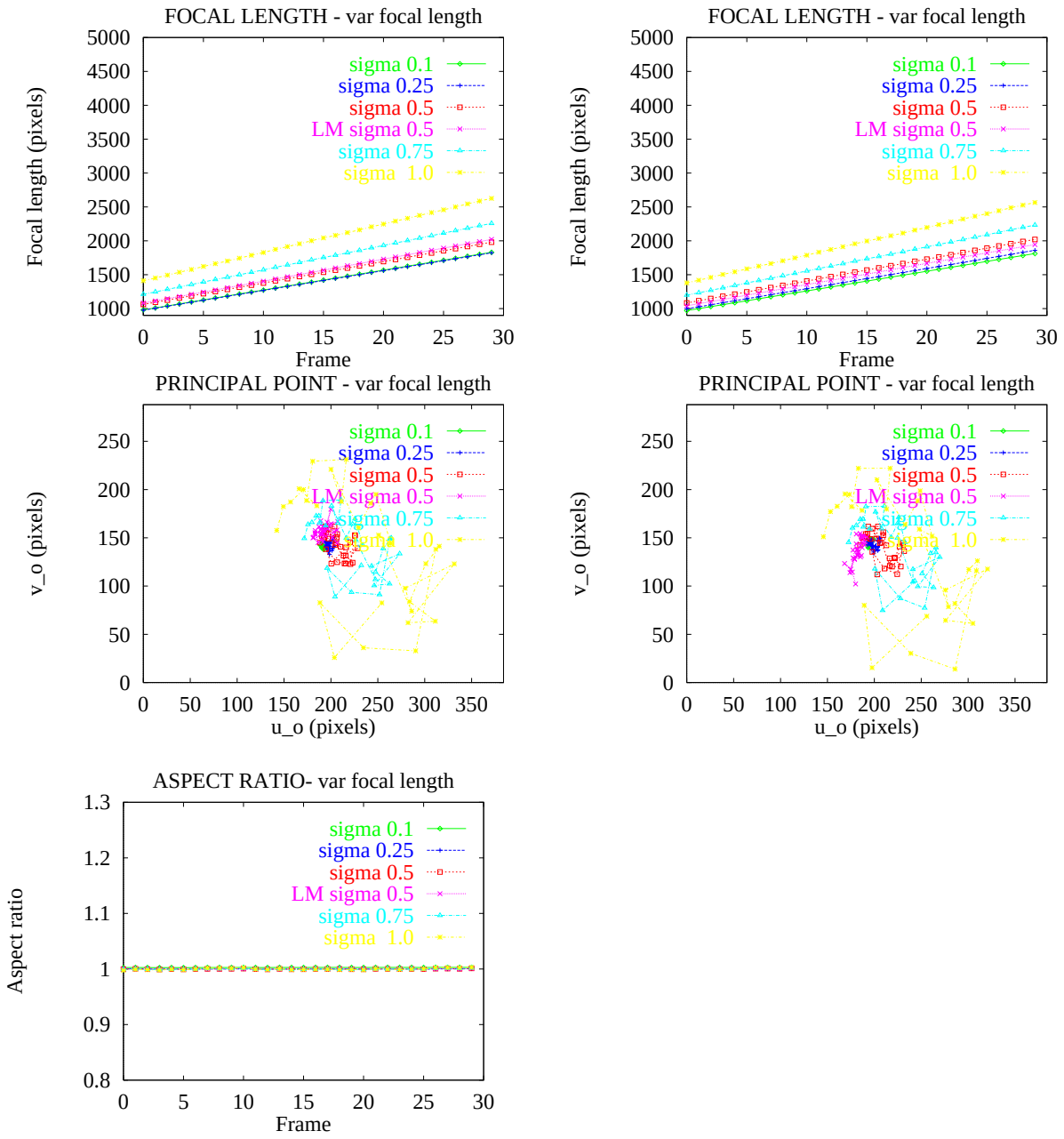

Figure 1: Calibration results with synthetic data in presence of various degrees of image noise, with one run at each noise level. Computed values for the focal length (top), the location of the principal point (middle) and the aspect ratio (bottom) imposing the *zero skew* (left) and *square-pixels* (right) constraints. Results obtained using the non-linear iterative method of [1] imposing the *zero skew* and the *square-pixels* constraint are shown for $\sigma = 0.5$. The aspect ratio was set to one by the algorithm when the *square-pixels* constraint was used

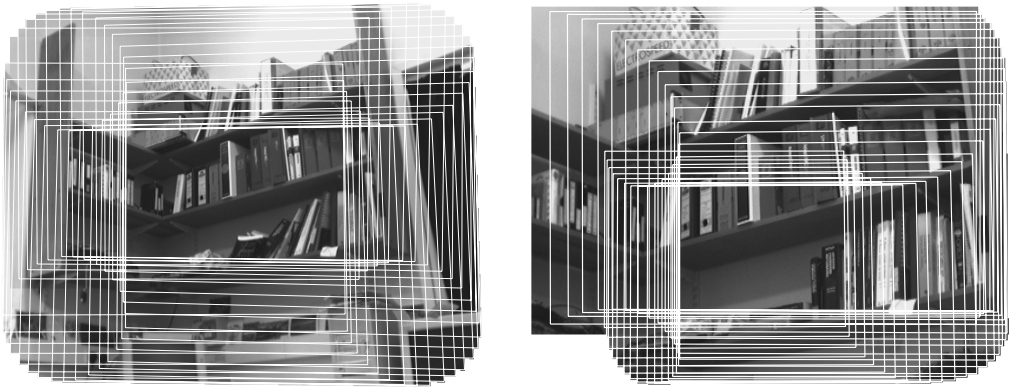

Figure 2: Mosaics constructed from the two bookshelf sequences during which the camera panned and tilted while the focal length remained fixed (left) and was varied (right).

ately warped to correct for this factor.

The homographies that relate corresponding points between views were computed in two stages. First, the inter-image homographies were computed from corresponding corners and second, they were refined by minimizing the global reprojection image error using a bundle-adjustment technique [4, 2]. Figure 2 shows the mosaics constructed by registering both image sequences.

When the *zero skew* constraint was imposed very poor results were obtained for the calibration parameters. The fundamental ambiguity described in section 4 was confirmed by observing that the two smallest singular values of the equation matrix `E` were very close to zero, implying that there is a one parameter family of possible solutions to the calibration. Imposing the *zero skew* constraint also failed to provide a solution when used in the non-linear iterative minimization.

Figure 3 shows the results obtained with the linear algorithm imposing the *square-pixels* constraint. The results confirm the good performance of the linear method, which in these particular experiments give better estimates than the iterative algorithm.

6 Conclusion

The key idea of the paper is to focus on the image of the absolute conic rather than its dual. This leads to a linear algorithm for the constrained calibration problem, rather than the iterative algorithms previously reported. The linear algorithm is extremely simple to implement and performs very well compared with iterative algorithms, ofter giving better results. The method fails in the case where the computed image of the absolute conic is not positive-definite. However, this did not occur in our experiments, except in the case of critical rotation sequences for which the calibration problem is inherently unstable. This serves as a warning that the data used does not support a useful estimate of the cameras' calibration parameters. It is probable that in these cases a calibration estimate given by any other algorithm, such as the previous iterative algorithms, would be virtually useless.

The linear methods do not apply in some cases for which the iterative algorithms may be used, such as fixed, but unknown aspect ratio, or fixed, but unknown principal point. The common cases of zero skew and known aspect ratio are covered, however. In practice, skew is almost always zero, and the aspect ratio is usually known to be one, or is available from a spec-sheet for the camera. In any case the aspect ratio is essentially invariant, and could be determined off-line. Further experiments (not explained in detail here) show that the linear algorithm can be used in a linear search over a range of feasible aspect ratios to determine the aspect ratio that gives the best fit to the data.

References

[1] L. de Agapito, E. Hayman, and I. D. Reid Self-calibration of a rotating camera with varying intrinsic parameters. In *Proc. British Machine Vision Conference*, pages 105–114, 1998.

[2] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 885–891, June 1998.

[3] R. I. Hartley Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conference on Computer Vision*, LNCS 588, pages 579–587. Springer-Verlag, 1992.

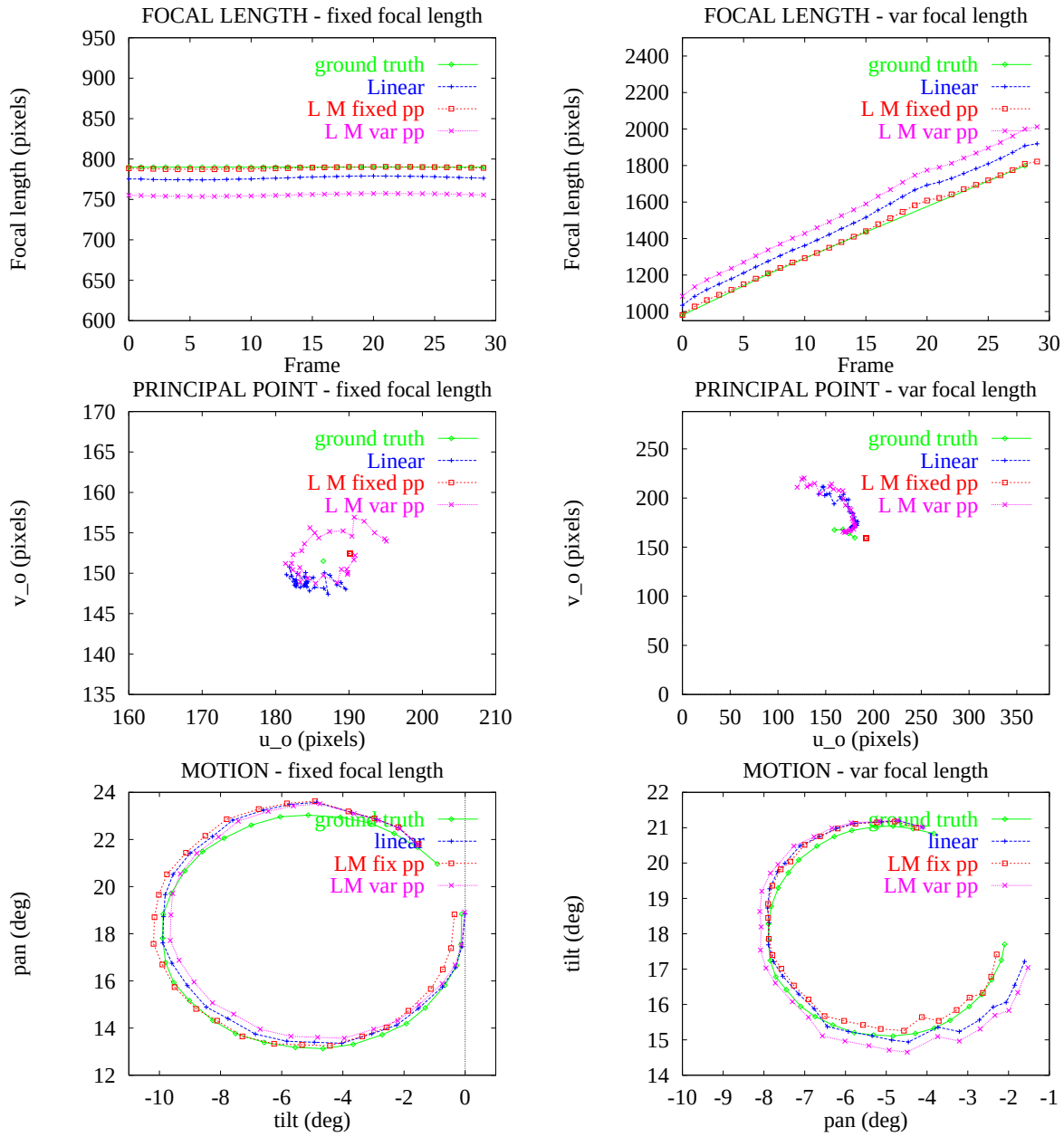

Figure 3: Ground truth and computed values for the focal length (top), the location of the principal point (middle) and the motion of the camera (bottom) for the fixed focal length (left) and the variable focal length (right) bookshelf sequences. Results are shown for (i) the linear algorithm imposing the *square-pixels* constraint (ii) the Levenberg-Marquardt algorithm imposing the *square-pixels* constraint and (iii) the Levenberg-Marquardt algorithm imposing both the *square-pixels* and the *fixed principal point* constraints. For visualization purposes, the motion was represented by plotting pan versus elevation angles. Also for clarity, only the central section of the image is shown in the plot depicting the principal point.

Note that the value of the focal length obtained from the linear algorithm is more accurate than the LM algorithm run using the same calibration assumptions (LM-var-pp). However, in both sequences, focal length estimation is improved by using the linear estimate as a starting estimate for iteration fixing the principal point (LM-fixed-pp). The principal point (in reality invariant during the fixed focal length sequence, and varying only slightly during the variable focal length sequence) appears to be better estimated by the linear algorithm.

[4] R. I. Hartley. Self-calibration of Stationary Cameras. *International Journal of Computer Vision*, volume 22, number 1, pages 5–23, February, 1997.

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

[6] A. Heyden and K. Åström. Minimal conditions on intrinsic parameters for euclidean reconstruction. In *Proc. Asian Conference on Computer Vision*, Hong Kong, 1998.

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

[8] M. Pollefeys, R. Koch, and L. Van Gool. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th International Conf. on Computer Vision, Bombay*, pages 90–96, 1998.

[9] Yongduek Seo and Ki Sang Hong, Auto-calibration of a Rotating and Zooming Camera, In *Proc. of IAPR workshop on Machine Vision Applications (MVA'98)*, Makuhari, Chiba, Japan, Nov. 17 – 19, 1998.

[10] P. M. Sharkey, D. W. Murray, S. Vandevelde, I. D. Reid, and P. F. McLauchlan. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4):517–535, 1993.

[11] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. Philosophical Transactions of the Royal Society of London. Vol 356(1740):1193–1211, 1998.