

Computing Camera Focal Length by Zooming a Single Point

N. Alberto Borghese^a, Franco M. Colombo^b and Alberto Alzati^b

a) AIS-Lab, Department of Computer Science, University of Milano, via Comelico
39/41, 20135 Milano, Italy, borghese@dsi.unimi.it.

b) Department of Mathematics, via Saldini 50, 20133 Milano, Italy, [colombo,
alzati]@mat.unimi.it.

Running head: Focal estimate

Number of pages of the manuscript: 19

Number of Figures: 5

Number of Tables: 1

Corresponding author:

N. Alberto Borghese, email: borghese@dsi.unimi.it.

AIS-Lab, Department of Computer Science, University of Milano, via Comelico
39/41, 20135 Milano, Italy, telephone: +39.02.503.16325.

<http://ais-lab.dsi.unimi.it>

Computing Camera Focal Length by Zooming a Single Point

N. Alberto Borghese^a, Franco M. Colombo^b and Alberto Alzati^b

a) AIS-Lab, Department of Computer Science, University of Milano, via Comelico 39/41, 20135 Milano, Italy, borghese@dsi.unimi.it.

b) Department of Mathematics, via Saldini 50, 20133 Milano, Italy, [colombo, alzati]@mat.unimi.it.

Abstract

In this paper we present a novel simple procedure to compute the focal length of a camera. The method is based on zooming in and out only a single point. The same approach allows computing the principal point when only two points are available on a pair of images surveyed with a different focal length. Experimental results show that the method is as accurate as classical full calibration methods. Moreover, its application to augmented reality produces more accurate results than when the simple pin-hole model is considered.

Keywords: Calibration; Cross-ratio; Focal length; Augmented reality; Zooming.

1. Introduction

Digital video applications, like sports video and TV shows, are more and more interested in compositing real video with digital models in real-time to add information content or increase the attractiveness through special effects [1]. To achieve this, the calibration parameters, which describe the projection of the 3D digital models on the 2D video images, have to be known. The procedure to determine these parameters is called camera calibration. The main difficulty to obtain on-the-field, real-time operation is related to the identification of good feature points and to the definition of an adequate model, which guarantee a reliable estimate of the parameters. We will focus here on camera calibration when zoom lenses are used.

Two are the main approaches to this problem. Classical calibration procedures adopt non-linear optimization and/or algebraic matrix manipulation, to compute the parameters altogether: both exterior (camera's position and orientation with respect to a given reference frame) and interior (focal length, principal point and distortion field) [1, 2, 4]. These procedures are based on surveying a 3D distribution of control points of known or partially known position, spread inside the calibrated volume [3]. However, such a

distribution is difficult to obtain in real situations and a host of calibration techniques, based on surveying simplified distributions such as linear, planar or circular have been developed [5, 6]. Methods of this category cannot be employed when camera parameters are changed often. In fact, they require stopping filming, setting-up the calibration object, surveying it and computing the parameters. Only afterwards, filming action can resume.

A different approach is based on computing the parameters from feature points automatically identified directly on the video image sequence. This approach belongs to the framework called “Structure from Motion” [7], and it has originated several calibration procedures [8, 9]. However, as a good distribution of feature points and a good initialization to converge are required (cf. Results Section), these techniques operate mainly on-line but not in real-time (cf. commercial products like Boujou and Realviz [10, 11]).

Two main simplification of the calibration model can be accepted for real-time operation: that the position of the principal point is fixed and known and that the distortion field can be neglected ([2, 4]). A further simplification can be obtained considering that the introduction of zooming lenses has limited camera movements during filming: different scene frames being obtained simply by zooming. In this condition, the only parameter required to project 3D digital models onto the scene, can be reduced to the actual focal length. This is changed often with zoom lenses, and each time it has to be determined. For this reason simple focal determination techniques have been derived. They are based on extracting lines from an image and evaluating the vanishing points: two vanishing points associated to two orthogonal directions, are sufficient to determine the actual focal length [9].

An even simpler approach is proposed here. It is shown that only one single point is sufficient to determine the actual focal length of a camera. This can be obtained through an innovative use of the projective transformation and requires that the image of the point is acquired with the actual focal length and two other known focal lengths, for instance the minimum and the maximum ones. Besides this we propose also an elegant solution for computing the principal point which requires only two points. Results on real images are reported and discussed.

2. Mathematical Background and Description of the Method

We will start with the pin-hole camera model (Fig. 1a), which is characterized by the position of the principal point, intersection of the optical axis with the image plane, $C(x_C, y_C)$, and by the focal length, f , which is the distance of the image plane from the

projection centre, \mathbf{F} . The pixel squareness is known: as it is independent from camera zoom, its value reported in the camera factory specifications can be used. The position of the principal point is supposed constant and known. Distortions are not considered.

The projection of a 3D point, $\mathbf{P}(X,Y,Z)$, on the image plane of the video camera, $\mathbf{p}(x_p, y_p)$, is described by the projective equations, which contain a factorization of the interior and exterior parameters in homogeneous notation:

$$\mathbf{p} = \mathbf{K} \mathbf{M} \mathbf{D} \mathbf{P} \quad (1)$$

where:

$$\mathbf{K} = \begin{bmatrix} -f & 0 & x_c \\ 0 & -f & y_c \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} \mathbf{R} & -\mathbf{RT} \\ \mathbf{0} & 1 \end{bmatrix} \quad (2)$$

\mathbf{R} and \mathbf{T} are the exterior parameters and represent the orientation (3x3 matrix) and the location of the camera with respect to an external reference frame. For convenience, the exterior reference frame is positioned in \mathbf{F} , with the X,Y axes parallel to those of the image plane. In this case, \mathbf{D} becomes equal to the identity matrix leading to the following simplified projection equation:

$$\mathbf{p} = \mathbf{K} \mathbf{M} \mathbf{P} \quad (3)$$

which, for x_p and y_p is:

$$x_p - x_c = -f \frac{X_p}{Z_p} \quad (4)$$

$$y_p - y_c = -f \frac{Y_p}{Z_p}$$

well known in the computer vision community. Eqs. (4) represent only an approximation of the real projective transformation as they are based on the implicit assumption that the centre of projection, \mathbf{F} , does not move; when focal length is changed, the image plane is implicitly moved further or closer to \mathbf{F} (cf. Fig. 1a).

2.1 Computation of the Focal Length

To compute the focal length taking into account the movement of the projection

centre we have developed a novel view of the perspective projection. Let us consider three images of one 3D point \mathbf{P} : $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ (Fig. 1b). Two of these are taken with known focal lengths, f_1 and f_3 and a third one with the focal length that we want to compute, f_2 (actual focal length). The point \mathbf{P} is projected on the image plane π through the points $\mathbf{F}_1, \mathbf{F}_2$, and \mathbf{F}_3 , which belong to the same line (the focal axis, a). We notice that:

- the points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ and the principal point \mathbf{C} are always collinear;
- the points $\mathbf{F}_1, \mathbf{F}_2$, and \mathbf{F}_3 and the principal point \mathbf{C} are always collinear;
- the points $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ and the principal point \mathbf{C} lie on the same plane.
- the points $\mathbf{C}, \mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ and $\mathbf{C}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ share the same cross-ratio:

$$[\mathbf{C} \mathbf{F}_1 \mathbf{F}_2 \mathbf{F}_3] = [\mathbf{C} \mathbf{p}_1 \mathbf{p}_2 \mathbf{p}_3] \quad (5)$$

If we suppose that the image plane reference system has the z axis parallel to the optical axis, a , we can write: $\mathbf{F}_j = [\mathbf{0}, f_j]$ and $\mathbf{p}_j = [\mathbf{p}'_j; 0]$, where \mathbf{p}'_j is the offset position of \mathbf{p}_j : $\mathbf{p}'_j = [x_{pj} - x_C; y_{pj} - y_C]$. If two focal lengths are known (for example f_1 and f_3) along with the principal point, \mathbf{C} , and the position on π of the three points $\mathbf{p}_1, \mathbf{p}_2$, and \mathbf{p}_3 , the unknown focal length, f_2 , can be computed as:

$$f_2 = \frac{f_1 f_3 \|\mathbf{p}'_2\| \|\mathbf{p}'_3 - \mathbf{p}'_1\|}{(f_1 - f_3) \|\mathbf{p}'_3\| \|\mathbf{p}'_2 - \mathbf{p}'_1\| + f_3 \|\mathbf{p}'_2\| \|\mathbf{p}'_3 - \mathbf{p}'_1\|} \quad (6)$$

Two focal lengths are particularly useful: the minimum and maximum ones. These can be estimated once, off-line, through a distribution of control points of high accuracy [2, 3, 4] are derived from factory specifications. We explicitly notice that in this derivation the image plane is considered constant (anchored to the camera body), while \mathbf{F} moves, translating on the optical axis.

2.2 Computation of the Principal Point

To compute the principal point, we start from the consideration that, given the image plane π , and two different points \mathbf{F}_1 and \mathbf{F}_2 on the focal axis a (Fig. 2a), we can define a projective transformation ω onto π , via a new plane π' , as follows. Let us consider a point \mathbf{P} on π' and let \mathbf{p}_1 and \mathbf{p}_2 be the projections of this point onto π , from the projection centres \mathbf{F}_1 and \mathbf{F}_2 . The projective transformation $\omega: \pi \rightarrow \pi$, can be defined as:

$$\omega(\mathbf{p}_1) = \mathbf{p}_2 \quad (7)$$

It is possible to show that:

- 1 given a point \mathbf{p}_j on the image plane π , the points \mathbf{p}_j , $\omega(\mathbf{p}_j)$, \mathbf{C} , \mathbf{F}_1 and \mathbf{F}_2 lay always on a unique plane;
- 2 the three points \mathbf{p}_j , $\omega(\mathbf{p}_j)$, and \mathbf{C} are always collinear;
- 3 every straight line s passing through \mathbf{C} and its image $\omega(s)$ satisfy the following constraint: $\omega(s) = s$;
- 4 and, at last, $\omega(\mathbf{C}) = \mathbf{C}$.

The last observation suggested to us the following algorithm. Take two images of at least two points, each with a different focal length, (e.g. by zooming); three points are shown in Fig. 2b for sake of clarity. Let us call $\mathbf{p}_j \mathbf{q}_j$ the pairs of points measured on the image plane π and r, s, t the lines through these pairs of points. The principal point will lie at the intersection of these lines (focus of expansion [9, 12]). An alternative method to compute the principal point is based on the determination of the eigenvalues and eigenvectors of the homology ω . We verified experimentally that this second solution, although elegant, is particularly sensitive to errors on point measurement and it was not pursued further.

2.3 Adding Digital Objects to the Images

The same novel view of the cross-ratio has suggested to us a technique to add digital objects to acquired images. Let us suppose that an object is present in the two images, I_1 and I_3 , acquired respectively with $f=f_1$ and $f=f_3$. Suppose that we want to insert the same object, or a digital object similar to that, into a third image, I_2 , acquired with $f=f_2$ (cf. Figs. 4).

In this situation the position of the (virtual) object in I_2 can be computed through the cross-ratio as:

$$x_2 = \frac{f_2(f_1 - f_3)(x_3 - x_C)x_1 + f_2f_3(x_3 - x_1)x_C - f_1f_3(x_3 - x_1)x_C}{f_2(f_1 - f_3)(x_3 - x_C) + f_2f_3(x_3 - x_1) - f_1f_3(x_3 - x_1)} \quad (8)$$

$$y_2 = \frac{f_2(f_1 - f_3)(y_3 - y_C)y_1 + f_2f_3(y_3 - y_1)y_C - f_1f_3(y_3 - y_1)y_C}{f_2(f_1 - f_3)(y_3 - y_C) + f_2f_3(y_3 - y_1) - f_1f_3(y_3 - y_1)}$$

Similar reasoning allows determining the area occupied on the image by an object surveyed with minimum and maximum focal length.

3. Experimental Results

The algorithm has been extensively tested on different sets of natural images taken with different cameras.

To evaluate the quality and reliability of the estimate of the actual focal length a planar black and white 7 x 9 chessboard with 30mm square length, was used (cf. Fig. 3). This pattern is widely adopted [5, 8] as it allows a reliable identification of the chessboard square corners, which serve as feature points. The results presented here were obtained from images acquired with a Sony DSC-S50 digital camera, with a target of 640 x 480 pixels. Minimum and maximum focal length, reported by the factory, was 6.1mm and 18.3mm respectively (zoom 3x).

Chessboard square corners were identified semi-automatically with Matlab[®] software made publicly available by [13]. The parameters estimated with the method presented here were evaluated comparing them with those obtained with the classical technique proposed by Zhang [8] and implemented in the same software package [13]. Typical results are reported in Table I. They are referred to ten different calibration session, where in each session camera was calibrated with six different focal lengths: the minimum and the maximum and four intermediate ones. For each focal length, the chessboard was acquired in at least 25 different positions and orientations for Zhang's algorithm, while it was acquired in the same single position for all the focal lengths for the method presented here. A subset of ten of the forty-eight chessboard corners were extracted randomly and used to determine the focal length with the method presented here.

We first checked that the principal point does not move significantly with zooming. The obtained mean and standard deviation of the principal point position obtained with Zhang's algorithm averaged over the six calibrations, each with a different focal length,

was: $(322.62; 220.28) \pm (0.94; 1.84)$ pixels (Table I). This is well in accordance with the data reported in the literature [4, 12, 14]. Moreover, for this particular digital camera the pixel form factor was one up to the sixth decimal digit, as reported in the factory specifications. The principal point value obtained with the method presented here is slightly less stable than in Zhang's calibration.

The focal length value was similar for both methods, with generally a smaller standard deviation for the method presented here. Moreover, we explicitly notice that for long focal lengths the estimate becomes critical for the limited amplitude of the field of view. In this situation, Zhang's algorithm requested up to 150 images to get a reasonable estimate of the focal length (standard deviation below 10 pixels). Similar problems were encountered when using Bundle adjustment in combination of control points of known position [3]. The method proposed here does not suffer from this problem: only one single point is sufficient also for long focal lengths as shown with the next data set.

Typical results obtained compositing digital objects and real images are presented in the soccer field images of Figs. 4. Here a portion of the field was surveyed with a Fuji, FinePix S602 digital camera, 2048 x 1536 pixels. We trusted the minimum ($f_1 = 8\text{mm}$) and maximum ($f_3 = 48\text{mm}$) focal lengths declared in factory specification.

The image taken with f_1 (wide angle, I_1) is shown in panel (a), while the image taken with f_3 (macro zoom, I_3) is shown in panel (b). The zooming effect here is very large, being a 6x. In these two images a soccer ball was present close to the corner flag. A third image (Fig. 4c) was taken with an intermediate focal length (image I_2);

We first notice that there is only one point which can be determined with a high reliability, which is the mast base of the corner flag. This point is used to compute the focal length with the method described in Section 2.2 (Eq. (6)). The value obtained was 24.4mm, which was congruent with the position of the zoom slider on the camera.

We then wanted to add the soccer ball also in this image, in the same position and with adequate size. To the scope, Eqs. (7) were applied to the six vertexes of the central white hexagon of the soccer ball pattern to compute their position in the image I_2 . The soccer ball image in I_1 was then shrank to fit these points and added to the image taken with the intermediate focal length, I_2 . The error in the ball position (measured in the hexagon vertexes) was of $[-2.083 \pm 0.68; 6.42 \pm 0.82]$ pixels on the x and y directions respectively. The difference in ball size was less than 1 pixel. The difference between the true ball and the ball added on the image is difficult to be appreciated also in the zoom circle at the bottom

right of Fig. 4c.

4. Discussion

Zoom lenses are more and more used in digital videos as they allow capturing the smallest details and to get a panoramic view of the same scene with the same camera. This reduces the need of camera motion and simplifies the procedure of camera calibration. In this paper, thanks to a different view of the cross-ratio in the perspective projection, the focal length can be computed using only one single point. This greatly simplifies the problems related to identification and matching multiple points over different images.

The method relies on an accurate estimate of two focal lengths: typically the minimum and maximum ones. These focal lengths should be such that they can be set again in the field. They can be estimated ones off the field with accurate markers, or we can rely on the factory data. Once these two focal lengths have been measured, the method allows a reliable estimate of the focal lengths in the entire range, and in particular of the longest ones. These are particularly difficult because of the geometrical set-up: classical solutions require many images to produce good results. Other focal lengths, between the maximum and the minimum one, could be identified on the camera and measured. This would improve the accuracy of the method.

The projection model used requires two assumptions. First, the principal point is known and is in a constant position. This is a reasonable assumption with modern CCD sensors: in many cameras, its displacement can be compared with the localization error of the features and it can be neglected [2, 4] (cf. Table I). Therefore it can be estimated once, off the field along with the minimum and maximum focal lengths. The second assumption is the neglect of the distortion field. This can offset image measurements, the offset increasing mainly in radial direction from the image centre and therefore having more impact on points close to the image borders. The distortion field is usually taken into account when very precise measurements are required and/or high distortion field is present [2, 9, 14, 15]. However, when the distortion field has moderate entity, and selected feature points are not close to the border, it can be left out from the calibration model [10].

The power of the method presented here is well exemplified by the very small error in the position computed in images taken with an arbitrary focal length, I_2 , of the same points measured on the images taken with reference focal lengths. Typical mean offset was of a few pixels, which has to be considered extremely small given the resolution of the cameras. In the soccer sequence the error in the object's position amounts to less than 0.1% in the horizontal direction and less than 0.4% in the vertical one (and it was not the smallest error

obtained in all the sequences that we have taken). The error was usually mainly along the y direction. This asymmetry in the error may be ascribed to the experimental conditions. The camera was mounted on a tripod and no remote control was given to the user to simulate the most frequent operating conditions.

4.1 Comparison with simple projective model

Projective transformation through zoom lenses is correctly represented by the thick-lens-model, which is based on the definition of an inner and outer principal point [15, 16] (Fig. 5a). The distance between these two points increases during zooming in. The corresponding geometry is represented in Fig. 1b and the relationship between a point in 3D space and its projection on the image plane taken with a different zoom, is described by Eqs. (6).

In computer vision, very often, this model is simplified by assuming that the image plane translates and the principal point is fixed. Under this condition (Fig. 1a) Eqs (4) are derived. To compute an unknown focal length, f_2 (the actual focal length) we have to measure two projections of the same 3D point on the image plane, each taken with a different focal length: \mathbf{p}_2 is obtained when the unknown focal length f_2 was used and \mathbf{p}_k with the known f_k , (e.g. the minimum or the maximum focal length). From \mathbf{p}_k , \mathbf{p}_2 and f_k , f_2 , can be computed through Eqs. (4) as:

$$f_2 = -f_k \frac{|\mathbf{p}_2 - \mathbf{C}|}{|\mathbf{p}_k - \mathbf{C}|} \quad (9)$$

Once f_2 has been determined, the position of any point measured with focal length f_k , $\mathbf{p}_k(x_k, y_k)$, can be identified over the image taken with focal length f_2 simply as:

$$\begin{aligned} x_2 &= x_C + (x_k - x_C) \frac{f_2}{f_k} \\ y_2 &= y_C + (y_k - y_C) \frac{f_2}{f_k} \end{aligned} \quad (10)$$

These equations derive from the simplification in the projection equations described in this Section and lead to the large errors as it can be clearly seen in the soccer sequence (Fig. 4d). Here the actual focal length, f_2 , was computed through Eq. 9 using the same point

at the mast base of the corner flag, used to calibrate with the method introduced here. Two different values were obtained: $f_2 = 12.3\text{mm}$ when the image taken with the maximum focal length was considered ($f_k = f_3$ and $\mathbf{p}_k = \mathbf{p}_3$), and $f_2 = 40.2\text{mm}$ when $f_k = f_1$ and $\mathbf{p}_k = \mathbf{p}_1$. We have then determined the position of the soccer ball and its size in the intermediate image, assuming $f_2 = 12.3\text{mm}$, which has given the best result with this solution (Fig. 4d). The mean offset of the same six points identified in Fig. 4c (Section 3) was $[-40.2 \pm 0.99 \ 25.4 \pm 7.09]$ pixels in the horizontal and vertical direction respectively. Moreover, the ball diameter was 19 pixels larger than in the real image.

Form this experiment it is clear that the displacement of the projection centre when zooming has to be taken into account if large errors have to be avoided.

5. Conclusion

In this paper we present a novel simple procedure to compute the focal length of a zooming camera. The method is based on zooming in and out only one single point. Experimental results show that the method produces results comparable to classical calibration methods. In particular, in comparison with Zhang's method, the estimate of the focal length is reliable also at large focal lengths, where Zhang's algorithm requires much more images. Experimental results support the need of an accurate projective model of zooming, which is able to take into account the translation of the projection centre. Large errors are originated by models which fail to achieve this.

References

- [1] F. Sparacino, G. Davenport, and A. Pentland, Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits, *IBM Systems Journal*, 39(3&4) (2000), pp. 479-509.
- [2] R.Y. Tsai. A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using off-the-shelf TV Cameras and Lenses. *IEEE J. Robotics and Autom.*, Ra-3(4) (1987), pp. 3-21.
- [3] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon. Bundle Adjustment —A Modern Synthesis. *Vision Algorithms: Theory & Practice*, B. Triggs, A. Zisserman & R. Szeliski (Eds.), Springer-Verlag LNCS 1883, 2000.

- [4] J. Weng, P. Cohen, M Henriou, Camera Calibration with Distortion Models and Accuracy Evaluation – IEEE Trans. Patt. Analysis and Machine Intell, 14(10) (1992), pp. 965-979.
- [5] N.A. Borghese and P. Cerveri (2000) Calibrating a video camera pair with a rigid bar. Pattern Recognition, 33(1) (2000), pp. 81-95.
- [6] P.Gurdjos, A.Grouzil and R.Payrissat, “Another Way of Looking at Plane-Based Calibration: the Centere Circle Constraint”, ECCV, 2002.
- [7] H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections. Nature, 293 (1981), pp. 133-135.
- [8] Z. Zhang. A Flexible new Technique for Camera Calibration, IEEE Trans. PAMI, 22(11) (2000), pp. 1330-1334.
- [9] R. Hartely and A. Zisserman – Multiple View Geometry – Cambridge University Press, 2003.
- [10] <http://www.2d3.com/jsp/index.jsp>.
- [11] <http://www.eagle.gr/products/Products/RealViz/realviz.htm>
- [12] M. Li and J. Lavest, Some Aspects of Zoom Lens Camera Calibration, IEEE Trans. Patt. Analysis and Machien Intell., 18(11) 1996, pp. 1105-1110.
- [13] <http://www.vision.caltech.edu/bouguet/calib/>
- [14] R.K. Lenz and R.Y. Tsai, Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology, IEEE Trans. Patt. Anal. Machine Intell., 10(5) (1988), pp. 713-720.
- [15] P.R. Wolf, Elements of Photogrammetry, McGraw Hill, 1980.
- [16] J. Lavest, G. Rives, M. Dhome, Three Dimensional Reconstruction by Zooming, IEEE Trans. On Robotics and Automation, 9(2) (1993), pp. 19-24.