Available online at www.sciencedirect.com
SCIENCE DIRECT
Pattern Recognition 39 (2006) 1522–1529
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

# Computing camera focal length by zooming a single point

N. Alberto Borghese[a,*], Franco M. Colombo[b], Alberto Alzati[b]

[a]AIS-Lab, Department of Computer Science, University of Milano, via Comelico 39/41, 20135 Milano, Italy
[b]Department of Mathematics, via Saldini 50, 20133 Milano, Italy

Received 26 July 2005; received in revised form 22 November 2005; accepted 24 January 2006

**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 those obtained 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 guarantees a reliable estimate of the parameters. We will focus here on camera calibration when zoom lenses are used.

There are two 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 to stop 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 are required to converge (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 simplifications of the calibration model can be accepted for real-time operation: the position of the principal point is fixed and known and the distortion field can be neglected [2,4]. A further simplification can be obtained

* Corresponding author.
*E-mail addresses:* borghese@dsi.unimi.it (N.A. Borghese), colombo@mat.unimi.it (F.M. Colombo), alzati@mat.unimi.it (A. Alzati)
*URL:* http://ais-lab.dsi.unimi.it (N.A. Borghese).

0031-3203/$30.00 
doi:10.1016/j.patcog.2006.01.011

considering that the introduction of zooming lenses has limited camera movements during filming: different scene frames being obtained simply by zooming in and out. 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, $\mathbf{P_C}(x_C, y_C)$, intersection of the optical axis, $a$, with the image plane, $\pi$, 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 these can be written in compact matrix notation as

$$\mathbf{p} = \mathbf{K}\,\mathbf{M}\,\mathbf{D}\,\mathbf{P}, \tag{1}$$

where

$$\mathbf{K} = \begin{bmatrix} -f & 0 & x_c \\ 0 & -f & y_c \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & -\mathbf{RT} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{2}$$

$\mathbf{R}$ and $\mathbf{T}$ are the exterior parameters and represent the orientation ($3 \times 3$ matrix) and the location ($3 \times 1$ vector) of the camera with respect to an external reference frame. For convenience, the exterior reference frame is positioned in $\mathbf{F}$, with the $X, Y$-axis 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:

$$\mathbf{p} = \mathbf{K}\,\mathbf{M}\,\mathbf{P} \tag{3}$$

which, for a point $\mathbf{p}(x_p, y_p)$ is

$$x_p - x_C = -f\frac{X_P}{Z_P},$$

$$y_p - y_C = -f\frac{Y_P}{Z_P}, \tag{4}$$

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 the focal length is changed, it is assumed implicitly that the image plane is 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 $\pi$ through the points $\mathbf{F_1}$, $\mathbf{F_2}$, and $\mathbf{F_3}$, which belong to the same line (the focal axis, $a$). We note 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:

$$[\mathbf{C}\,\mathbf{F_1}\,\mathbf{F_2}\,\mathbf{F_3}] = [\mathbf{C}\,\mathbf{p_1}\,\mathbf{p_2}\,\mathbf{p_3}]. \tag{5}$$

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 $\mathbf{p_j}$: $\mathbf{p'_j} = [x_{p_j} - x_C, y_{p_j} - y_C]$. If two focal lengths are known (for example $f_1$ and $f_3$) along with the principal point, $\mathbf{C}$, and the position on $\pi$ of the three points $\mathbf{p_1}$, $\mathbf{p_2}$, and $\mathbf{p_3}$, the unknown focal length, $f_2$, can be computed as

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

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 [2–4] or can be derived from factory specifications. We explicitly note that in this derivation the image plane is

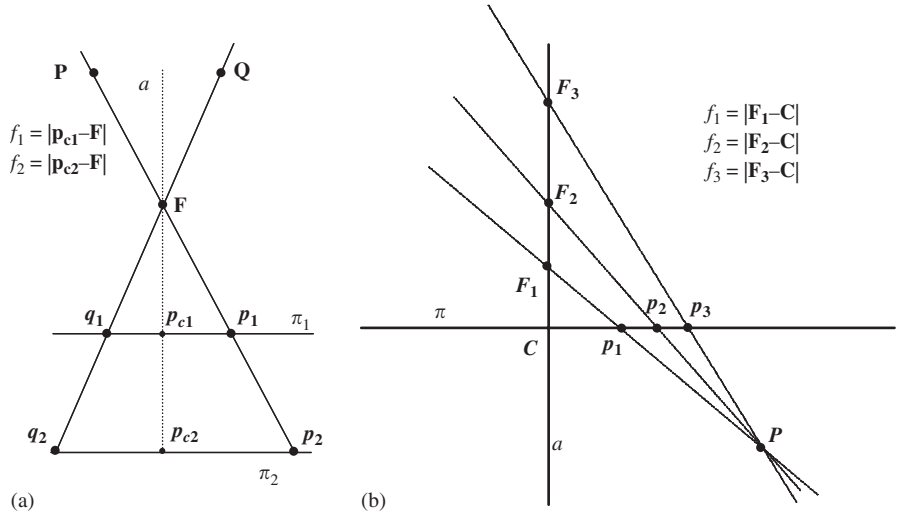


Fig. 1. The simple pinhole camera model usually adopted to describe the projective transformation is shown in panel (a). When zooming in, the image plane gets closer to the projection centre, **F**. **P** and **Q** are two 3D points which are projected with focal $f_j$ into $\mathbf{p_j}$ and $\mathbf{q_j}$ on the image plane. The plane is implicitly moved from the position $\pi_1$ to the position $\pi_2$. The projection model, which takes into account the movement of the projection centre when zooming, is reported in panel (b). In this case the image plane, ***P***, is fixed and the projection centre translates along the optical axis $a$. The novel way to look at the cross-ratio, introduced in this paper, is also shown in panel (b). $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ are the projection of the 3D point, **P**, on the image plane, $\pi$, through three different three projection centres $\mathbf{F_1}$, $\mathbf{F_2}$, and $\mathbf{F_3}$. A different focal length: $f_1$, $f_2$ and $f_3$, is associated to each of these points. These three points and the principal point, **C**, are collinear. The same collinearity condition holds for **C**, $\mathbf{F_1}$, $\mathbf{F_2}$, and $\mathbf{F_3}$ are. From the cross-ratio $[\mathbf{C}\,\mathbf{F_1}\,\mathbf{F_2}\,\mathbf{F_3}] = [\mathbf{C}\,\mathbf{p_1}\,\mathbf{p_2}\,\mathbf{p_3}]$ the actual focal length, $f_2$, can be computed. Only one point is sufficient.

considered fixed (anchored to the camera body), while **F** moves, translating along 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 $\pi$, and two different points $\mathbf{F_1}$ and $\mathbf{F_2}$ on the focal axis $a$ (Fig. 2a), we can define a projective transformation $\omega$ onto $\pi$, via a new plane $\pi'$, as follows (see also Ref. [9, Chapter 12]). Let us consider a point **P** on $\pi'$ and let $\mathbf{p_1}$ and $\mathbf{p_2}$ be the projections of this point onto $\pi$, from the projection centres $\mathbf{F_1}$ and $\mathbf{F_2}$. The projective transformation $\omega : \pi \to \pi$, can be defined as

$$\omega(\mathbf{p_1}) = \mathbf{p_2}. \tag{7}$$

It follows that the transformation of a point, $\mathbf{p_j}$, which lies on a straight line through **C** on $\pi$, performed through $\omega(.)$, produces a second point, $\mathbf{p_k} = \omega(\mathbf{p_j})$, which belongs to the same line. It is possible to show that:

1. given a point $\mathbf{p_j}$ on the image plane $\pi$, the points $\mathbf{p_j}$, $\omega(\mathbf{p_j})$, **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 **C** are always collinear;
3. every straight line $s$ passing through **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 $\pi$ 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 $\omega$. We verified experimentally that this second solution, although elegant, is particularly sensitive to error on points 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 $f = f_2$ (Fig. 3).

In this situation the position of the (virtual) object in $I_2$ can be computed through the cross-ratio (Eq. (6)) as

$$x_2 = \frac{f_2(f_1 - f_3)(x_3 - x_C)x_1 + f_2 f_3(x_3 - x_1)x_C - f_1 f_3(x_3 - x_1)x_C}{f_2(f_1 - f_3)(x_3 - x_C) + f_2 f_3(x_3 - x_1) - f_1 f_3(x_3 - x_1)},$$

$$y_2 = \frac{f_2(f_1 - f_3)(y_3 - y_C)y_1 + f_2 f_3(y_3 - y_1)y_C - f_1 f_3(y_3 - y_1)y_C}{f_2(f_1 - f_3)(y_3 - y_C) + f_2 f_3(y_3 - y_1) - f_1 f_3(y_3 - y_1)}. \tag{8}$$

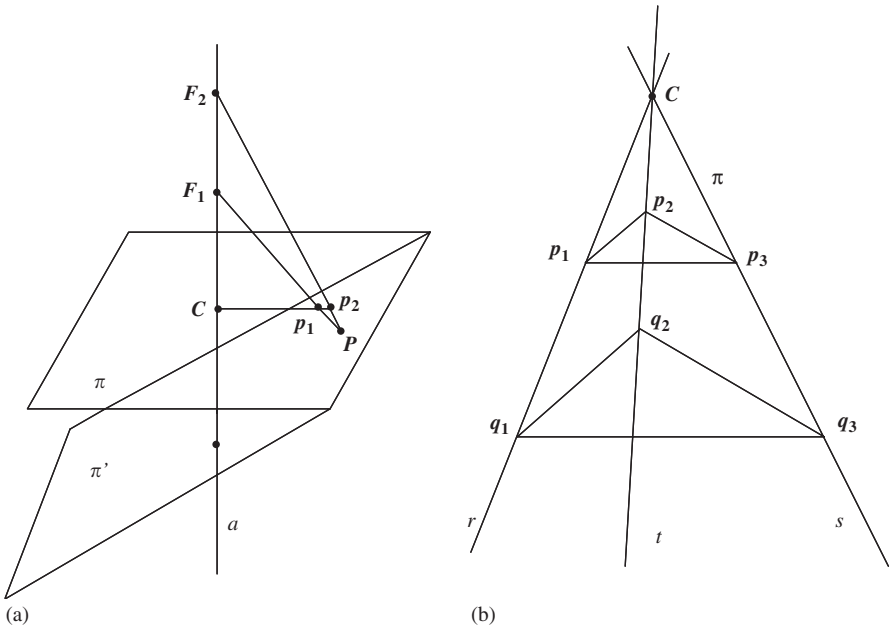


Fig. 2. The projective transformation model used to compute the principal point is depicted in panel (a). $\mathbf{p_1}$ and $\mathbf{p_2}$ are the projection onto the image plane $\pi$ of a point $\mathbf{P}$ belonging to $\pi'$, through the points $\mathbf{F_1}$ and $\mathbf{F_2}$, respectively. $\mathbf{C}$ is the principal point. In panel (b) the projection on the image plane, $\pi$, of three points, $\mathbf{P_1}$, $\mathbf{P_2}$, and $\mathbf{P_3}$, positioned at two different distances from $\mathbf{C}$ (two different focal lengths) is shown. These points are $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ and $\mathbf{q_1}$, $\mathbf{q_2}$, $\mathbf{q_3}$, respectively. The three lines, $s, r, t$, from the pair of points, $\mathbf{p_j}$, $\mathbf{q_j}$, intersect in $\mathbf{C}$. Two of such lines are sufficient to determine the principal point or focus of expansion.

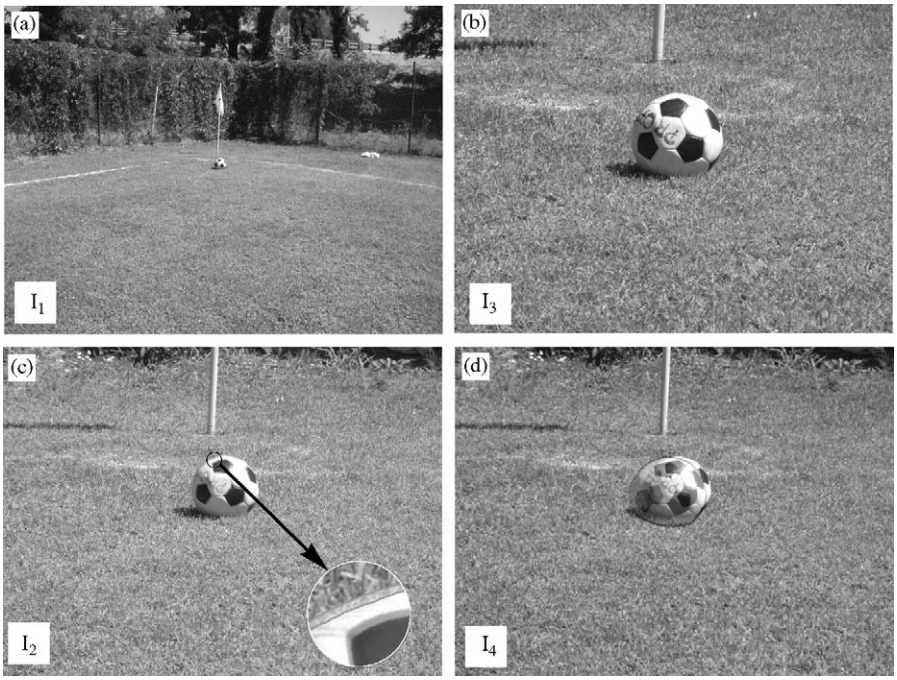


Fig. 3. The portion of a soccer field close to one of the four corners was taken with a Fuji FinePix 602S camera with a target size of $2048 \times 1536$ pixels and a $6\times$ zoom capability. The image taken with the minimum focal length ($f_1 = 8$ mm) is shown in panel (a) while that taken with the maximum focal length ($f_3 = 48$ mm) is shown in panel (b). The camera was positioned approximately at 12 m from the corner. Note that the only point which can be reliably computed is the mast basis of the corner flag. From the position of this point measured in $I_1$ and $I_3$, the actual focal length used to take the image $I_2$ (panel (c)) was computed and resulted of 24.4 mm. The image of the ball synthesized from images (a) and (b) as described in Section 2.3, is shown superimposed as a transparency to the real position in panel (c). The difference is hard to distinguish also inside the zoomed image shown inside the circle in the bottom right corner, being the ball center offset of 2 pixels in horizontal and 6 pixels in vertical (0.1% and 0.4% of the image size, respectively). Ball image diameter was computed within less than 1 pixel of error. In panel (d) the same situation is shown when a simple ratio is adopted to describe the projective transformation (Eqs. (9)–(10) in Section 4.1). Here, both the displacement and the dimension of the ball exhibit large errors (see text).

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 $7 \times 9$ chessboard with 30 mm square length was used (cf. Fig. 4). 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 $640 \times 480$ pixels. Minimum and maximum focal length, reported by the factory, was 6.1 and 18.3 mm, respectively (zoom $3\times$).

Chessboard square corners were identified semi-automatically with Matlab$^{TM}$ software made publicly available from Ref. [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]. Ten different calibration sessions were carried out; in each session the camera was calibrated with six different focal lengths: the minimum, the maximum and four intermediate ones. The mean and standard deviation of the estimated values are reported in Table 1. 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 fixed position for the method presented here. A subset of 10 of the 63 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 1). 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 note 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 with 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 Fig. 3. Here, a portion of the field was surveyed with a Fuji, FinePix S602 digital camera, with a target size of $2048 \times 1536$ pixels. We trusted the minimum ($f_1 = 8$ mm) and maximum ($f_3 = 48$ mm) focal lengths declared in the factory specifications.

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 $6\times$. In these two images, a soccer ball was present close to the corner flag. A third image (Fig. 3c) was taken with an intermediate focal length (image $I_2$).

We first note 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 24.4 mm, 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 vertices 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 $[-2.083 \pm 0.68;$

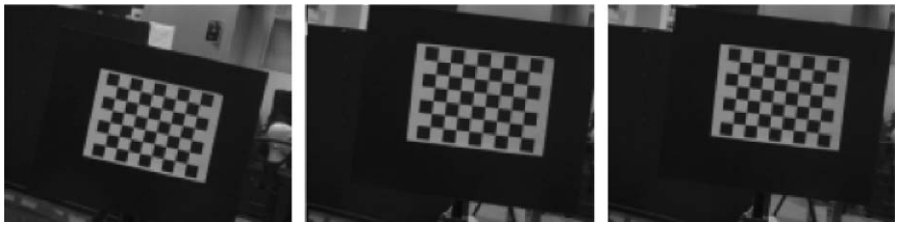

Fig. 4. Three images of the chessboard used to calibrate with Zhang's algorithm, taken with different position and attitude. The sequence was taken by a Sony DSC-S50 digital camera (target size of $640 \times 480$ pixels) with its minimum focal length ($f = 6.1$ mm).

Table 1
The position of the principal point, **C**, and the focal length, $f$, estimated by the method proposed here and by Zhang's algorithm

| Computation of **C** (proposed method) | Computation of **C** (Zhang's calibration) |
|---|---|
| (316.59; 219.97) ± (4.53; 4.42) | (321.65; 220.66) ± (4.41; 6.36) |
| (333.01; 217.36) ± (6,61; 6.48) | (321.08; 221.90) ± (1.20; 8.73) |
| (316.20; 219.47) ± (9.07; 8.39) | (320.47; 220.67) ± (5.32; 8.93) |
| (318.36; 236.83) ± (57.08; 45.52) | (324.52; 222.01) ± (16.68; 16.17) |
| **Computation of $f$ (proposed method)** | **Computation of $f$ (Zhang's calibration)** |
| 1079.88 ± 4.15 | 1081.97 ± 9.92 |
| 1221.42 ± 5.12 | 1198.86 ± 5.11 |
| 1395.98 ± 5.32 | 1392.00 ± 6.58 |
| 1773.70 ± 12.10 | 1779.93 ± 9.50 |

A Sony DSC-S50 digital camera with a target size of 640 × 480 pixels and a 3× zooming capability, was used to survey a 7 × 9 chessboard. For the method presented here, the camera stayed still and only zooming was performed to calibrate; for each zoom level a set of 30 calibrations was carried out; in each of them a sub-set of 10 vertexes was randomly extracted from the 63 vertexes identified on the image, and used to compute the parameters. The statistics was obtained by analysing the parameters obtained in the 30 different calibrations. For Zhang's method, the same chessboard was placed in at least 25 different positions and orientations, for each focal length. Statistics reported is that provided from Ref. [13]. Units are pixels in both cases.

$6.42 \pm 0.82$] 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. 3c.

## 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; these focal lengths should be such that they can be reproduced on the field, and they are typically the minimum and maximum one. These focal lengths can be estimated once 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 assumption is that 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 1). 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 without degrading the accuracy significantly [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 the 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

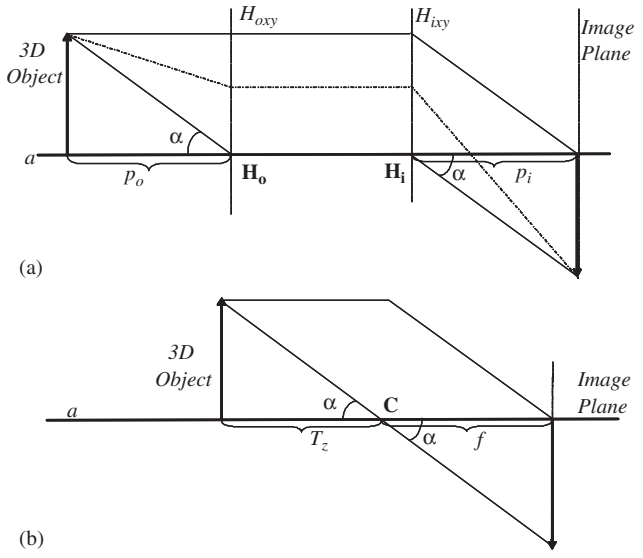


Fig. 5. The thick-lens model is plotted in panel (a). The planes $H_{oxy}$ and $H_{ixy}$ are called principal planes. They are orthogonal to the optical axis, $a$, and are positioned such that the optical rays travel parallel from $H_{oxy}$ to $H_{ixy}$. $\mathbf{H_o}$ and $\mathbf{H_i}$ are the intersection of the optical axis with $H_{oxy}$ and $H_{ixy}$. The distance of the object, $p_o$, is measured with respect to $H_{oxy}$ while the image plane distance, $p_i$, is measured with respect to $H_{ixy}$. During zooming in, the distance between $H_{oxy}$ and $H_{ixy}$ decreases. The thick-lens model can be transformed into the pin-hole model (panel b) by making $\mathbf{H_i}$ and $\mathbf{H_o}$ coincident: $\mathbf{H_i} = \mathbf{H_o} = \mathbf{C}$. $\mathbf{C}$ is the centre of projection in the pin-hole model. $p_o = T_z$ (distance of the object from the centre of projection) and $p_i = f$ (distance of the centre of projection from the Image plane).

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 along the optical axis and the principal point is fixed. Under this condition (Fig. 1a), Eqs. (4) are derived. In this model, we have to measure only two projections on the image plane of the same 3D point to compute an unknown focal length, $f_2$ (the actual focal length). Each projection has to be taken with a different focal length: $\mathbf{p_2}$ is obtained with the unknown focal length, $f_2$, and $\mathbf{p_k}$ with the known one, $f_k$, (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

$$f_2 = f_k \frac{|\mathbf{p_2} - \mathbf{C}|}{|\mathbf{p_k} - \mathbf{C}|}. \tag{9}$$

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

$$x_2 = x_C + (x_k - x_C)\frac{f_2}{f_k},$$
$$y_2 = y_C + (y_k - y_C)\frac{f_2}{f_k}. \tag{10}$$

These equations derive from the simplification in the projection equations (8) and leaded to the large errors as it can be clearly seen in the soccer sequence (Fig. 3d). With the simplified model, 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: $f_2 = 12.3$ mm when the image taken with the maximum focal length was considered ($f_k = f_3$ and $\mathbf{p_k} = \mathbf{p_3}$), and $f_2 = 40.2$ mm with the minimum focal length ($f_k = f_1$ 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 $f_2 = 12.3$ mm, which has given the best result with this solution. Nevertheless, the mean offset of the same six points identified in Fig. 3c (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.

From 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, A. Pentland, Media in performance: interactive spaces for dance, theater, circus, and museum exhibits, IBM Syst. J. 39 (3&4) (2000) 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 Autom. Ra-3 (4) (1987) 3–21.
[3] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment—a modern synthesis, in: B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory & Practice, Lecture Notes in Computer Science, vol. 1883, Springer, Berlin, 2000.
[4] J. Weng, P. Cohen, M. Henriou, Camera calibration with distortion models and accuracy evaluation, IEEE Trans. Pattern Anal. Mach. Intell. 14 (10) (1992) 965–979.
[5] N.A. Borghese, P. Cerveri, Calibrating a video camera pair with a rigid bar, Pattern Recognition 33 (1) (2000) 81–95.
[6] P. Gurdjos, A. Grouzil, R. Payrissat, Another way of looking at plane-based calibration: the Center Circle Constraint, ECCV, 2002.
[7] H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, Nature 293 (1981) 133–135.
[8] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. PAMI 22 (11) (2000) 1330–1334.

[9] R. Hartely, A. Zisserman, Multiple View Geometry, Cambridge University Press, Cambridge, 2003.
[10] ⟨http://www.2d3.com/jsp/index.jsp⟩.
[11] ⟨http://www.eagle.gr/products/Products/RealViz/realviz.htm⟩.
[12] M. Li, J. Lavest, Some aspects of zoom lens camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 18 (11) (1996) 1105–1110.
[13] ⟨http://www.vision.caltech.edu/bouguet/calib/⟩.
[14] R.K. Lenz, R.Y. Tsai, Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology, IEEE Trans. Pattern Anal. Mach. Intell. 10 (5) (1988) 713–720.
[15] P.R. Wolf, Elements of Photogrammetry, McGraw Hill, New York, 1980.
[16] J. Lavest, G. Rives, M. Dhome, Three dimensional reconstruction by zooming, IEEE Trans. Robotics Autom. 9 (2) (1993) 19–24.