PHOTOGRAMMETRIC CALIBRATION OF IMAGE SEQUENCES ACQUIRED WITH A ROTATING CAMERA

Fabio Remondino[(1)], Niclas Börlin[(2)]

[(1)] Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: fabio@geod.baug.ethz.ch
Web: http://www.photogrammetry.ethz.ch
[(2)] Department of Computing Science, Umea University, Sweden,
E-mail: Niclas.Borlin@cs.umu.se
Web: http://www.cs.umu.se

KEY WORDS: Calibration, Orientation, Bundle Adjustment, Sequence Analysis, Panorama

ABSTRACT
This paper reports theory and examples about the calibration and orientation of fixed but freely rotating cameras with possible changes of the interior parameters. We consider cameras that are generally rotating, without any special adapter to remove the eccentricity between perspective center and rotation axis. That is the typical case of surveillance cameras or sport videos. Projective and perspective camera model are analyzed and between the reported examples, self-acquired images and old monocular videos of sports events are considered. We also show the possibility to achieve 3D object reconstruction using rotating cameras. Finally we will report the mosaic generation from images acquired with a rotating system.

1. INTRODUCTION

Camera calibration has always been a topic of great interest in the field of photogrammetry and machine vision and many methods and camera models have been proposed. In this work we consider cameras which are rotating and changing their interior parameters. This is the case of image streams acquired with a camera mounted on a tripod or rotated on the shoulder of a camera-man or like pan-tilt zooming surveillance cameras, panoramic acquisition with a single camera or sport events videos. Usually the problem is formulated within a projective framework because of the absence of camera and object information. In fact, when existing videos are analyzed, it is very difficult to recover accurate 3D scene information and the camera parameters, mainly because of (1) low image quality (interlaced video), (2) almost no information concerning the camera, (3) often absence of baseline and (4) possible variations of the internal parameters.

In the vision community many algorithms have been presented to calibrate image sequences acquired with a stationary but freely rotating camera [Hartley, 1994; De Agapito et al., 1998; Seo et al., 1999]: they rely on the homography (8-parameters projective transformation) between the images and they retrieve the camera parameters with linear or iterative methods. Usually changes of the internal parameters (mainly zooming) are also allowed but they often assume zero-skew or known pixel aspect ratio without a statistical check on the determinability of the parameters.

In the photogrammetric community, camera calibration by rotation (often called 'single station calibration') has been investigated by many authors [Wester-Ebbinghaus, 1982; Brown, 1985, reported in Fryer, 1996; Stein, 1995; Pöntinen, 2002], even if the process cannot be considered robust and accurate as conventional convergent self-calibration bundle adjustment.

In this paper we describe theory and examples concerning the calibration (and orientation) of cameras that are fixed in a location (e.g. tripod) but can freely rotate and maybe change their internal parameters (Figure 1). We will not use cocentric images as the cameras have no special adapter to minimize the eccentricity of the perspective center respect to the rotation axis

but, in same cases, the eccentricity will be neglected. Projective and perspective camera model are revisited while self-acquired sequences and existing videos are analyzed. The self-acquired images are taken always with the same camera, which is calibrated with a testfield to validate the results (Appendix 1). In the last section of the paper we also report about the generation of mosaics created stitching and registering together different images.
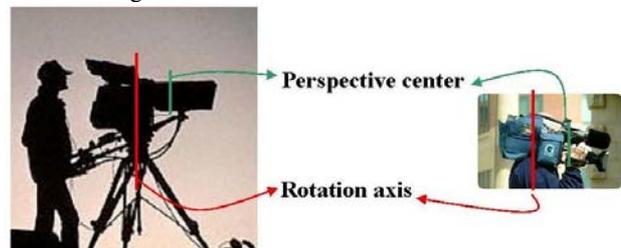
Figure 1: Videocamera mounted on a tripod (left) or rotated on a shoulder (right). The eccentricity between rotation axis and plane of the perspective center in same cases can be neglected.

2. CAMERA CALIBRATION AND ORIENTATION

All applications that deal with the extraction of precise 3D information from imagery require accurate calibration and orientation procedures as prerequisites of reliable results. The early theories and formulations of orientation procedures were developed in the first half of the 19th century and today a great number of procedures and algorithms are available. A fundamental criterion for grouping the orientation procedures is based on the used camera model (Table 1):

- perspective camera model: camera models based on perspective collineation have high stability, require a minimum of three corresponding points per image and a stable optics; they can include non-linear lens distortion function; they often contain non linear relations, requiring initial approximations of the unknowns.
- projective camera model: these approaches can handle variable focal lengths but need more parameters, a minimum of six correspondences and are quite instable (equations

often need normalization); they cannot easily deal with non-linear lens distortion, but contain often linear relationship.

Geometry	**Projective**	**Perspective**
Camera model	$p_i' = A \cdot p_i$ (3x1) (3x4) (4x1) p_i', p_i are projective coord.	$p_i' = \lambda_i \cdot R \cdot p_i + t$ (3x1) (3x3) (3x1) (3x1) p_i', p_i are cartesian coord. λ_i scale factor of p_i
Parameters	11 (relevant) in matrix A	6 for EO 3 for IO other correction functions
Relationship	Often Linear	Always Non-linear

Table 1: Orientation approaches and related parameters.

In [Wrobel, 2001] a good review of many orientation approaches is presented. The choice of the camera model is often related to the final application and the required accuracy. Photogrammetry deals with precise measurements from images and accurate sensor calibration is one of its major goals. Both camera models have been discussed and used in close-range photogrammetry but generally a sensor calibration is performed with a perspective geometrical model by means of the bundle method.

2.1 Approximations for the camera parameters

Most of the calibration and orientation solutions are based on non-linear algorithms, requiring initial approximations of the unknown parameters. Moreover we often need to recover metric results from image streams (e.g. 3D models, movement information), without any typical photogrammetric information. Therefore information about (1) the camera (interior and exterior parameters) and (2) the images (pixel size) are required to perform the adjustment. We assume that we can always define a scale factor or some control points, knowing the dimensions of some objects in the imaged scene.

The *pixel size* is mainly a scale factor for the camera focal length. Its value can be recovered from a set of corresponding object and image coordinates distributed on a plane.

The camera *interior parameters* can be recovered with an approach based on vanishing point and line segments clustering [Caprile et al., 1990; Remondino, 2002] or with orthogonality conditions on line measurements [Krauss, 1996; Van den Heuvel, 1999]. If the image quality does not allow the extraction of lines, the decomposition of the 3x4 matrix of the projective camera model can simultaneously derive the interior parameters given at least 6 control points [Hartley et al., 2000; Remondino, 2003].

Concerning the *exterior parameters*, an approximate solution can be achieved with a closed form space resection [Zeng et al., 1992] or the classical non-linear space resection based on collinearity, given more than 4 control points. The DLT method can sequentially recover all the 9 camera parameters given at least 6 control points [Abdel-Aziz et al., 1971]. DLT contains 11 parameters, where two mainly account for film deformation: if no film deformation is present, two constraints can be add to solve the singularity of the redundant parameters [Bopp et al., 1978]. Other approaches are also described in [Slama, 1980; Criminisi, 1999; Foerstner, 2000; Wolf et al., 2000].

3. THE PROJECTIVE CAMERA MODEL

Projective geometry is widely used, in particular in the vision community, to recover camera and scene information from images. As shown in Table 1, a general projective camera maps an object point $\mathbf{X}$ to an image point $\mathbf{x}$ according to $\mathbf{x} = P\ \mathbf{X}$,

with P a 3x4 matrix which can be decomposed as P = K [R | **t**], where:

$$\mathbf{K} = \begin{pmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

is an upper triangular matrix with the interior parameters of the camera (f_x and f_y are the focal length along x and y axis, s the skew factor, (x_0, y_0) the principal point position); R is a rotation matrix and **t** is translation vector.

If the camera is fix and undergoes only rotations (cocentric images or negligeble eccentricity), we can eliminate the vector **t** and express the mapping of $\mathbf{X}$ onto $\mathbf{x}$ as [Hartley, 1994]:

$$\mathbf{x} = K\ R\ \mathbf{X} \tag{1}$$

as P = K [R | 0], i.e., P = K R.

Given 2 images, the projection of $\mathbf{X}$ onto them will be given by: $\mathbf{x}_i = K_i R_i \mathbf{X}$ and $\mathbf{x}_j = K_j R_j \mathbf{X}$. Therefore, eliminating $\mathbf{X}$, we get:

$$\mathbf{x}_j = H_{ij}\ \mathbf{x}_i \tag{2}$$

with:

$$H_{ij} = K_j\ R_j\ R_i^{-1}\ K_i^{-1} = K_j\ R_{ij}\ K_i^{-1} \tag{3}$$

or, if the camera parameters are constant:

$$H_{ij} = K\ R_{ij}\ K^{-1} \tag{4}$$

where H_{ij} is the inter-image homography containing the element of the 8-parameters projective transformation. Given n>4 image correspondences, we can recover the H matrix with a least squares solution. H can be multiplied with an arbitrary scale factor without altering the projective transformation result.

Thus, constructing the homography H from image correspondences is an easy job. However, unpacking K and R from H is more elaborate. From (3), considering only one camera, we get:

$$H_{ij}\ K = K\ R_{ij} \tag{5}$$

and postmultiplying the two sides by their transposes yields

$$H_{ij}K(H_{ij}K)^T = KR_{ij}\ (KR_{ij})^T \tag{6}$$

$$H_{ij}KK^TH_{ij}{}^T = KR_{ij}{}^TR_{ij}K^T = KK^T \tag{7}$$

where the last simplification is due to R being orthogonal. Using the substitution $A = KK^T$, with

$$A = \begin{pmatrix} f_x^2 + s^2 + x_0^2 & f_y s + x_0 y_0 & x_0 \\ f_y s + x_0 y_0 & f_y^2 + y_0^2 & y_0 \\ x_0 & y_0 & 1 \end{pmatrix} \tag{8}$$

equation (6) becomes

$$H_{ij}\ A\ H_{ij}{}^T = X \tag{9}$$

or

$$H_{ij}\ A - A\ H_{ij}{}^{-T} = 0 \tag{10}$$

which is a linear homogeneous Sylvester equation in the 9 entries of A [Bartels et al., 1972]. A is symmetric and can only be determined up to a constant factor. To solve for its entries, we need N (>2) images and solve a system of equations G**a** = 0, where G is a 9Nx6 matrix and **a** is a vector containing the entries of A. The solution is the eigenvector corresponding to the least eigenvalue of the Sylvester matrix A.

Then, we can derive the value of the calibration matrix K from A applying the Cholesky decomposition, if A it positive-definite.

On the other hand, if we consider images acquired with a rotating camera, which is changing its interior parameters, equation (9) becomes

$$H_{ij}\ A_i\ H_{ij}{}^T = A_j \tag{11}$$

and a solution for the A (or K) entries can be found using a non-linear least squares algorithm and minimizing the cost function:

$$\frac{}{\displaystyle\sum_{j}^{n_{img}-1}} \parallel K_j K_j^T - H_{ij} K_i K_i^T H_{ij}^T \parallel \qquad (12)$$

over the parameters of each calibration matrix K_j and taking image i as the reference image [De Agapito et al., 1998].

3.1 Obtaining the rotation angle from the projective transformation

Equation (2) (which is an 8-parameters projective transformation) relates two image planes that undergo a rotation and are intersected by the same image ray. The 8 parameters of the H matrix hide the 9 coefficient of the orthogonal rotation matrix R. The eigenvalues of an orthogonal matrix (whose product gives the determinant of the matrix and whose magnitude is always 1) must satisfy one of the following conditions:
1. all eigenvalues are 1.
2. one eigenvalue is 1 and the other two are -1.
3. one eigenvalue is 1 and the other two are complex conjugates
$\{1, e^{i\vartheta}, e^{-i\vartheta}\}$

The rotation matrix is uniquely defined by its rotation angle ϑ and rotation axis a. ϑ can be computed from the eigenvalues of R while a is afterwards derived from ϑ.

The matrix H has the same eigenvalue of R, up to a scale factor. Therefore, knowing H, we can estimate the rotation between two images i and j up to a sign. If we have multiple frames and the rotation is continuous in one direction, we may use the positive sign for all of them and estimate each consecutive rotation from the composite one.

4. SIMPLIFIED PERSPECTIVE CAMERA MODEL

If the camera undergoes a rotation on a tripod (cocentric images, small eccentricity or camera far away from the scene), the image correspondences are related only with a rotation and the collinearity model can be approximated with:

$$x' - x_0 = -c \cdot \frac{r_{11} X + r_{21} Y + r_{31} Z}{r_{13} X + r_{23} Y + r_{33} Z}$$

$$y' - y_0 = -c \cdot \frac{r_{12} X + r_{22} Y + r_{32} Z}{r_{13} X + r_{23} Y + r_{33} Z} \qquad (13)$$

On the other hand, a perspective projection can also be represented with x=cX/Z and y=cY/Z, therefore the coordinates of an image point (x',y') that undergoes a rotation can be computed as:

$$x' - x_0 = -c \cdot \frac{r_{11}(x - x_0) + r_{21}(y - y_0) + r_{31} c}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33} c}$$

$$y' - y_0 = -c \cdot \frac{r_{12}(x - x_0) + r_{22}(y - y_0) + r_{32} c}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33} c} \qquad (14)$$

given the coordinates of the corresponding point (x,y) in the previous image. Equation (14) can also be seen as a projective transformation

$$x' = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1}$$

$$y' = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1} \qquad (15)$$

as already formulated in [Wang, 1990].

If we have multiple images, we may want to estimate the camera parameters using (14) and solving the minimization:

$$\frac{}{\displaystyle\sum_{j=1}^{n_{img}} \sum_{i=1}^{n_{pts}}} \left[(\hat{x}_{ij} - x_{i,j})^2 + (\hat{y}_{i,j} - y_{i,j})^2 \right] \heartsuit \ MIN \qquad (16)$$

where:
$(\hat{x}_{ij}, \hat{y}_{ij})$ are the estimated coordinates and (x_{ij}, y_{ij}) are the measured coordinates.

If we have n images acquired with the same camera and we want to use fix interior parameters, the following functional constraints are used:
$c_n = c = $ const;
$x_{0,n} = x_0 = $const;
$y_{0,n} = y_0 = $const.
Equations (14) can also be extended to include some additional functions (e.g. Brown's model) to model the systematic image errors and recover lens distortion parameters.

The advantage of this approach is that we do not require any 3D information as we assumed that the image correspondences depend only on the rotation matrix and the camera interior parameters. But, on the other hand, we cannot recover the position of the camera's perspective center.

A similar mathematical model, describing the relationship between corresponding image points only with a rotation matrix is presented in [Pöntinen, 2002].

4.1 Determinability of camera's parameters with the simplified camera model

In the following examples, the camera interior parameters are recovered with the minimization function (16). Camera parameters, residuals in image space and adjustment's standard deviation are reported to check the accuracy of the model.

4.1.1 Self-acquired images
A sequence of a small testfield (Figure 2) is acquired rotating a Leica Digilux 1 on a tripod without any special adapter.

Figure 2: The 'testfield sequence' (1600x1200 pixels).

The rotation is prominent in the vertical direction. The image correspondences as extracted automatically using LSM and the coded target information. The results of the adjustment are presented in Table 2. The RMS of the residuals in image space are 1.7 µm in both directions.

Parameter	Value
Camera constant	7.326 mm
PP x dir.	0.117 mm
PP y dir.	-0.234 mm
K1	-3.076e-3
K2	5.423e-5
$\sigma_{0,priori} = 1$ pixel	$\sigma_{0,post} = 0.6$ pixel

Table 2: Recovered camera's parameters.

4.1.2 Existing videos
The sequence presented in Figure 3 was digitized from an old videotape with a Matrox DigiSuite frame grabber. The image size is 720x576 pixels and between all the grabbed frames, only 6 images are analyzed. From a quick analysis of the images, we can deduce that a rotation is mainly occurring during the video acquisition while no zooming effects are presents. Because of

the low image quality (interlaced video) the image measurements are performed manually.

Figure 3: A rotating camera set on the shoulder of a camera-man ('walking sequence').

The recovered camera parameters are reported in Table 3. The focal length is kept fix while principal point and distortion parameters resulted highly correlated and not determinable.

Parameter	Value
Focal length	23.4 mm
RMS_x	24.6 μm
RMS_y	24.5 μm
$\sigma_{0,post}$	1.11 pixel

Table 3: Recovered parameters of the 'walking sequence'.

These results look a bit better respect to the values obtained in section 5.2.2 with the standard bundle method. But, mainly due to the very low image quality, no additional parameter could be recovered.

5. CONVENTIONAL PERSPECTIVE CAMERA MODEL

An image acquired with a traditional camera is in principle a central projection of the scene. If more images from different viewpoints are available, the bundle method allows the computation of all camera parameters and 3D object coordinates as well as the compensation of the systematic errors. But if the baseline between adjacent images is very small, the bundle adjustment can easily fail as the normal equations may be ill-conditioned or the rays cannot correctly intersect. Moreover, if the images are acquired with a stationary but freely rotating camera, the conventional bundle method cannot solve the adjustment, unless we introduce constraints and we treat all the system's parameters as observed values.

5.1 Compensation of systematic error

The bundle method and its statistical model are usually extended by additional parameters functions that are supposed to model the systematic image errors (self-calibrating bundle adjustment) [Brown, 1971; Beyer, 1992]. In the adjustment, the additional parameters (APs) can be imported as:
1. block-invariant: one set of APs is used for all the images; it is the most common approach, in particular for laboratory calibration.
2. frame-invariant (or focal-invariant): a set of APs is used for each image (or camera); this approach is necessary e.g. in multi-cameras applications (robotics or machine vision inspections) or if zooming effects are presents. This solution can create over-parameterisation problems.
3. a combination of focal- and block-invariant: the APs are divided in (1) a group which is supposed to be block-invariant (e.g. principal point, affinity and shear factor) and

(2) a group which is related to a specific focal length (e.g. lens distortion parameters). The new equations expressing the functional dependence between the APs must be incorporated in the mathematical model of the self-calibrating bundle adjustment, 'bordering' the normal equations with the new geometric conditions [Fraser, 1980].

The procedure of self-calibration using APs introduces new observations and unknowns in the least squares estimation, extending the bundle model and arising problems concerning the quality of the model. The improper use of the APs can deteriorate the results as the parameters weaken the condition of the system of the normal equations or lead to a singular system. Therefore an 'additional parameter testing' is always required [Gruen, 1981], in particular when the network geometry is not optimal for system calibration.

5.1.1 Analysis of determinability

The undeterminably of APs can lead the normal matrix to singularity. In the literature, different approaches have been proposed, in particular:
- analysis of the correlation between the parameters: it is the most and widely used approach. The APs are usually strongly correlated with one another or with the camera parameters. The correlation coefficient between two parameters i and j can be computed from the cofactor matrix Q_{xx} (i.e. the inverse the of the normal equation matrix). Generally correlations bigger than 0.95 should be eliminated.
- Jacobsen's method [Jacobsen, 1982]: a measure of the APs' determinability is given by the relation:

$$B = I - [(diagN)*(diag(N^{-1}))]^{-1} \qquad (17)$$

with
B = diagonal matrix for the measure of the determinability;
I = identity matrix;
diagN = diagonal element of the normal equation matrix
The elements B_{ii} vary between 0 (diagonal normal equation matrix) and 1 (singular normal equation) and for the APs they should not be larger than 0.85.
- Gruen's method [Gruen, 1985]: it is a stepwise procedure that should be performed at the different stages of the least squares adjustment. It is based on the trace check of the covariance matrix to detect and delete those APs that are in the critical range between poorly determined and sufficiently well determined parameters.

5.1.2 Significance tests

Insignificant APs do not affect the solution of the adjustment but may weaken the covariance matrix without positive contributions to the functional model. Their rejection is usually performed with statistical tests, like the Student's test: we compare the null-hypothesis "the AP x is not significant" against the alternative hypothesis "the AP x is significant" using a variable *t*:

$$t = \frac{x_{EST} - x'}{\sigma_{0,POST} \cdot q_{xx}} \qquad (18)$$

with:
x_{EST} = estimated value of the AP x;
x' = value approximated or at the previous step of the iterations;
$\sigma_{0,POST}$ = sigma a posteriori of the adjustment;
q_{xx} = variance of the AP x.
The values of *t*, for a particular significance level and with a certain DOF, may be found in statistical books.
Another approach, called spectral analysis, is described in [Foerstner, 1980].

5.2 Determinability of camera's parameters with the conventional bundle method

In the following we will show, using some examples, how individual or sets of APs can be determined under very unfair network geometry conditions, like sequences of images acquired with a camera undergoing a rotation on a tripod. The adjustment is mainly a space resection adjustment, where all the object points and camera parameters are treated as observed values and not as free unknowns. The accuracy of the adjustment will be check with the 'a posteriori' standard deviation and the RMS of the residuals in image space.

5.2.1 Self-acquired images

A *first test* is performed using the 'testfield sequence' presented in section 4.1.1 (Figure 2). A set of ca 60 control points is available for the calibration. All the camera parameters and the object points are treated as observed values.

The results of the adjustment, presented in Table 4, show that almost all the camera parameters can be reliable recovered (compare the results with the calibration values of Appendix 1). In spite of a bad network geometry, a good number of control points and the treatment of all unknown parameters as stochastic variables helped in the determination of the camera parameters. The P_i parameters turned out to be highly correlated with the principal point position ($\Delta x_p \leftrightarrow P_1 \cong 0.98$; $\Delta y_p \leftrightarrow P_2 \cong 0.97$) while the other APs were not significant. The RMS of the residuals in image space are 0.24 μm in x and 0.21 μm in y.

Parameter	Value	Std. Dev.
Camera constant	7.319 mm	9.86e-03
PP x dir.	0.118 mm	5.43e-03
PP y dir.	-0.212 mm	1.69e-03
K1	-3.792e-03	2.31e-05
K2	6.475e-05	1.45e-06
$\sigma_{0,priori}$ = 1 pixel	$\sigma_{0,post}$ = 0.08 pixel	

Table 4: The recovered parameters after the adjustment.

During the experiments, we observed that if the camera is rotating only along its horizontal axis (i.e. vertical rotation), the position of the principal point in x direction cannot be recovered, because of high correlations (>0.994) with the angles around the horizontal axis and other parameters. On the other hand, rotations only around the vertical axis, do not allow the correct determination of the principal point in y directions.

A *second test* is done with 6 images, acquired from two different stations (Figure 4), with a Leica Digilux 1 mounted on a tripod. No control point information is available and the baseline between the stations is ca 2.5 m. Three distances are imported in the adjustment for the datum definition while the tie points are measured semi-automatically with LSM. The calibration results are reported in Table 5.

Parameter	Value	Std. Dev.
Camera constant	7.380 mm	5.91e-2
PP in x dir	0.102 mm	5.10e-2
K1	-3.81e-3	2.51e-4
K2	9.580e-5	1.30e-6

Table 5: Recovered camera parameters of the 'shelf sequence'.

Only K1 and K2 were reliable determinable while the principal point in y direction had to be fixed because the camera was rotated only along the horizontal direction. The final standard deviation of unit weight a posteriori resulted 0.005 mm (ca 1.1 pixel).

Figure 4: The 'shelf sequence': 6 images (1600x1200) acquired from two stations.

As two stations are used, we could also recover the object points coordinates with good accuracy (Table 6). The camera poses and the 3D object coordinates are shown in Figure 5.

RMS_X=0.0022 m	RMS_Y=0.0018 m	RMS_Z=0.0025 m

Table 6: Mean RMS of the computed object coordinates.
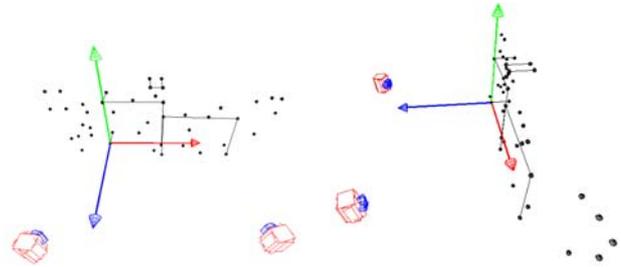
Figure 5: The camera poses of the 2 stations and the 3D scene.

This example and the modelling results show that the 3D object reconstruction from multi-stations rotating cameras is possible with good accuracy and a not expensive camera system.

5.2.2 Existing videos

Nine frames of the 'walking sequence' presented in Figure 3 are analyzed. A right-hand coordinate system with the origin in the left corner of the court (and the XZ plane parallel to the court) is set and some control points are defined knowing the dimensions of the basketball court. Because of the low image quality (interlaced video) the image measurements are performed manually. All the measurements are then imported as weighted observations and used as tie points in the adjustment. At first, for each single frame, DLT and space resection are used to get an approximation of the camera parameters. Afterwards a bundle adjustment is applied to recover all the parameters ($\sigma_{0,priori}$ = 1.5 pixel) with 2 different computational versions:

1. *bundle with frame-invariant APs set*s (Table 7): this version (9 sets of APs) recovered a constant value for the affinity factor ($1.11 \pm 1.6e^{-3}$) and a mean focal length value of 22.4 mm ± 0.26 mm, even if the oscillations reported in Figure 6 suggested a block-invariant configuration. The other APs were not significant in all the images, probably because of the low image measurements' quality and an over-parameterised system.

Mean Focal length	22.4 mm
Mean Affinity factor	1.1105
RMS_x	24.8 μm
RMS_y	18.6 μm
$\sigma_{0,post}$	1.24 pixel

Table 7: Results of the adjustment with frame-invariant APs sets.
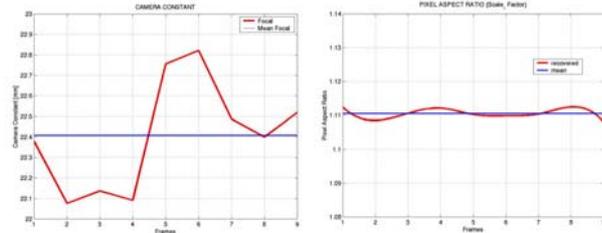
Figure 6: Left: The behaviour of the focal length during the analyzed frames. Right: The recovered affinity factor.

The behaviour of the recovered EO parameters is also consistent with the images, as shown in Figure 7. X_0 is increasing while Y_0 and Z_0 are constant.
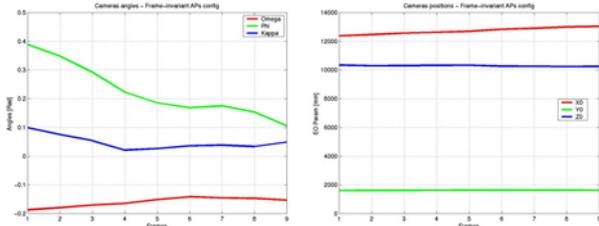
Figure 7: The motion of the camera (left: angles, right: positions) recovered with a frame-invariant APs set.

2. *bundle with block-invariant APs set* (Table 8): this version recovered very similar results compared to the frame-invariant version. Moreover, the K_1 parameter could also be determined.
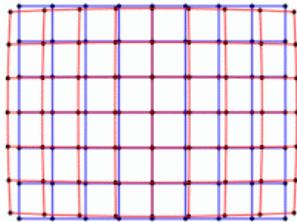
Figure 8: Influence of the APs on the image grid (3 times amplified).

$\sigma_{0,post}$	1.29 pixel
Focal length	22.71 mm
Affinity factor	1.1192
K1	-4.36e-04
RMS_x	29.7 µm
RMS_y	22.1 µm

Table 8: Results of the bundle adjustment with block-invariant APs set.
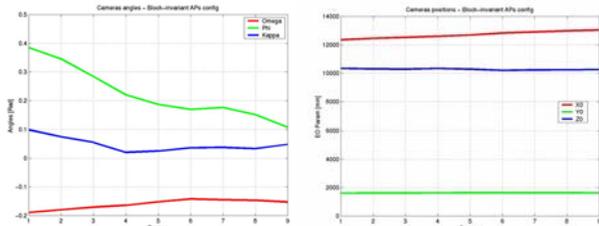
Figure 9: The motion of the camera (left: angles, right: positions) recovered with a block-invariant APs set.

The non-unity of the pixel aspect-ratio can come from the old videocamera or because of the used frame grabber.
In Figure 10 are shown the camera positions together with the reconstructed 3D scene.
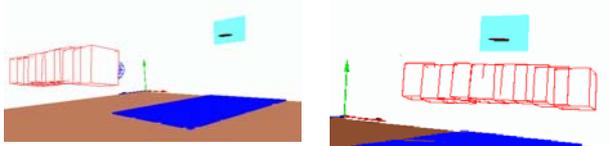
Figure 10: 3D scene and camera poses recovered after the bundle adjustment.

Another image sequence, obtained digitizing a videotape from 1989, is considered. A total of 40 frames are grabbed and 21 images are used for the analysis (Figure 11).

Figure 11: Some frames of a basketball video ('dunking sequence'). The camera is rotating and zooming.

The camera is far away from the scene and is rotating (probably on a tripod) and zooming to follow the moving character. The calibration and orientation process is again performed with a self-calibrating bundle adjustment. Because of the zooming effect, a frame-invariant APs set is used.
The diagram of the recovered focal length (Figure 12, right) shows the visible zooming-in effect of the camera, except for the last 3 frames (not displayed in Figure 11). The affinity factor resulted in $1.11 \pm 4.5e^{-3}$ (Figure 12, right). Because of the low precision of the image measurements ($\sigma_{0,priori}$ = 2 pixel) and the network geometry, the principal point and the lens distortion terms cannot be computed as very unreliable determinable. The final standard deviation of unit weight a posteriori resulted 1.4 pixels while the RMS of image coordinates residuals are 38.45 µm in x and 29.08 µm in y.
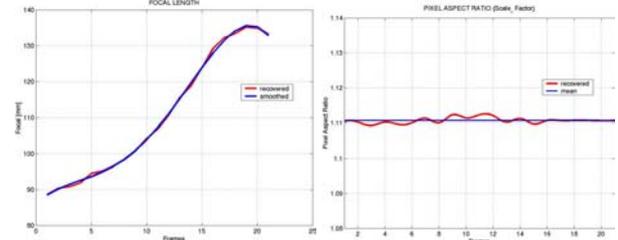
Figure 12: The focal length values in the 21 frames (left) and the affinity factor (right).

6. MOSAIC FROM AN IMAGE SEQUENCE

There mainly three approaches to generate a panoramic view (or image mosaic) and they are based on:
- single images: it is the traditional method, in particular in the vision and graphic community. The panoramic view is generated stitching and registering together different images.
- mirror techniques: they use single or double mirror, providing for high capturing rate but low resolutions.
- rotating linear array CCD devices: these are panoramic cameras that capture 360 degrees view in one scan.

6.1 Panorama from single images

This is a low cost technique, but usually time consuming as non-linear optimization methods are used for the registration and blending procedure. The images are acquired using simple tripods or just rotating manually the camera. More than 30 commercial software is available on the market to produce such panoramas (see Table 9). They distinguish in the extent of

automation and in the requirements for the input data. The source format is usually JPEG, while the output file can have different formats (MOV, IVR, PAN, etc.)

3DVista Studio	Cool 360	PhotoVista
Image Assembler	REALVIZ stitcher	Panorama Tools
Panorama Factory	PhotoShop Elements	Reality Studio
PanEdit	PanoStitcher	QTVR

Table 9: Some software for panoramic images generation.

After the stitching, the panoramic image is usually warped applying a particular projection for better visualization. There are mainly 4 types of projections (planar, cylindrical, spherical and cubic) and they distinguish in the CPU requests and distortion correction [http://www.panoguide.com].
The panoramic image can then be visualized with special viewers (Table 10) that allow interactive navigation.

3DVista Java applet	360 WorldView	glPanorama
PanaView	Zoom Viewer	PT Viewer
JSphere	iPIX	Janorama

Table 10: Viewers for panoramic images.

Panoramic views created with this approach do not allow precise reconstruction and are very useful for virtual tour [http://www.world-heritage-tour.org], virtual reality and e-commerce.

6.2 Examples

A program was developed with the goal of align (or warp) different views and blend them together into a larger image. It is based on the projective transformation (15) and works according to the following steps [Szeliski, R., 1996]:
1. selection of corresponding points (x, y) and (x', y') between image pairs I and I';
2. compute the projective transformation (15) between the two images recovering the 8 parameters iteratively by minimize the sum of the squared intensity errors E:

$$E = \frac{}{i} \left[I'(x', y') - I(x, y) \right]^2 \qquad (19)$$

over all the corresponding pixels pairs i. The minimization is performed using a Levenberg-Marquardt algorithm.
3. blend the resampled image with the reference image I using a bilinear weighting function (weighted average) and project the new image on a plane.

A similar approach, based on Gauss-Markov least squares minimization, has been presented in [Pöntinen, 1999].
All the approaches based on a projective transformation between successive images implicitly assume that the imaged scene is approximately planar or that the images are cocentric or that the scene is very far away from the camera.

6.2.1 Results from the 'testfield sequence'
After the calibration procedure, we can generate the distortion-free images of Figure 2, using the recovered APs to remove the distortion effects. Then we apply the algorithm of section 6.2 to produce a larger image of the testfield (Figure 13).

6.2.2 The 'shelf sequence'
A panoramic view of the whole shelf can be generated using the three distortion-free images acquired e.g. in the right station (Figure 14).

6.2.3 The 'basketball dunking sequence'
A mosaic of a sequence can be produced to retrieve scene information. Considering Figure 11, because of the movement of the character, only 3 frames are considered. The tie points measured for the photo-triangulation are used as correspondences for the computation of the projective transformation between adjacent images. The approach described in section 6.2 produced a 'ghost effect' in the final results (Figure 15, upper image) because of the automated blending procedure. Therefore a manual stitching, after the projective alignment, is also performed (Figure 15, lower). The mosaic can finally be used to derive metric measurements of the character movements, like length or height of the jump [Remondino, 2003].
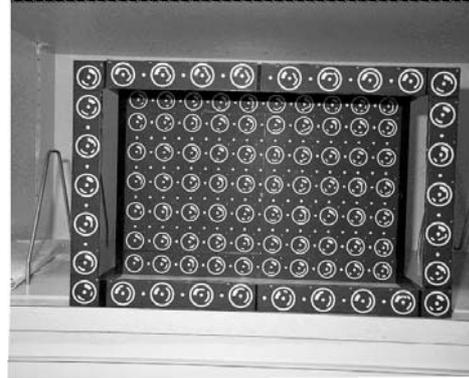
Figure 13: A mosaic of the 'testfield sequence'.

Figure 14: The 'shelf sequence'. Panoramic view created using 3 images (upper image); zoomed details of difficult areas near the images' borders.

7. CONCLUSIONS

The presented work was realised to show how the perspective camera model based on the conventional bundle method can be employed to calibrate rotating cameras that do not generate cocentric images. Otherwise a simplified camera model, that relates image correspondences only with a rotation matrix, can be used. The significance tests on the APs and the treatment of all the unknowns as observed values helped to reduce the ill condition of the normal equations and recover the camera parameters. The results obtained from the existing video do not

respect the usual photogrammetric accuracy, mainly because of the very low image quality. Probably the results could be improved using simultaneously block- and image-invariant APs sets in the adjustment; but, on the other hand, it can be very difficult to select which parameters are block-invariant (i.e. are common videocameras free of lens barrel misalignment?).

We also showed how multi-stations rotating cameras can be employed for efficient and low cost 3D reconstruction of object with photogrammetric precision ('shelf sequence').

As future work we would like to perform other tests with the simplified camera model and extend it with a deeper stochastic analysis of the system's parameters.
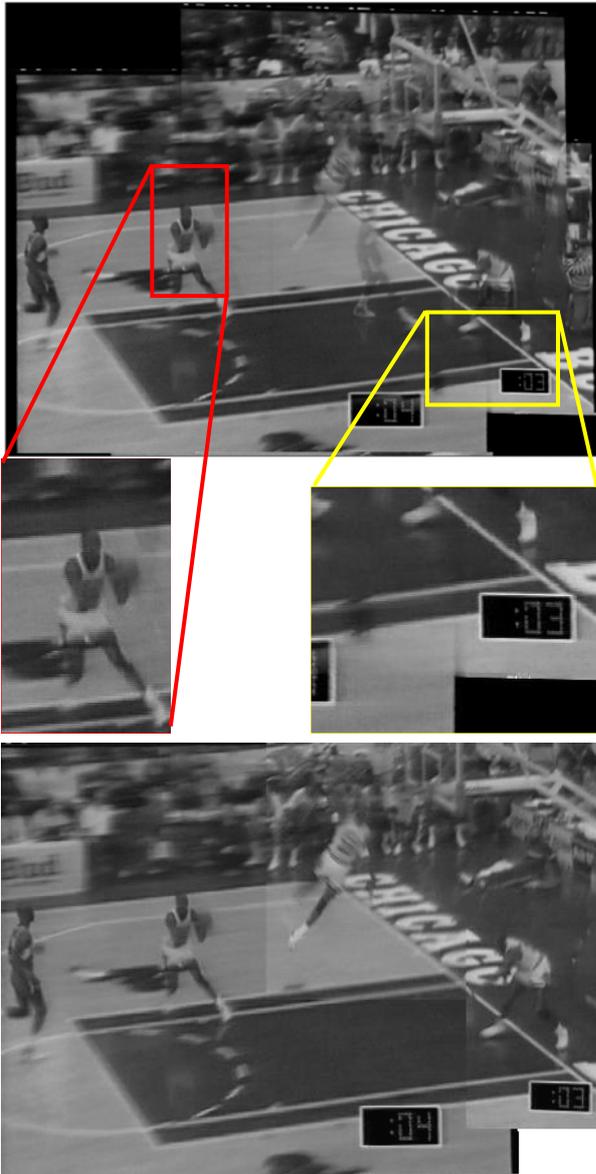
Figure 15: The 'dunking sequence'. In the upper image, the automatic blending procedure generated the 'ghost' effect visible in the middle. In the central images, two details in correspondence of image blending are enlarged. The lower image shows the same mosaic but with manual stitching, after the projective transformation of the images.

APPENDIX 1

The digital camera *Leica Digilux 1* is used for some of the presented experiments (sections 4.1.1 and 5.2.1). In the following we report its main characteristics (Table 11), the results of a testfield calibration performed at wide-angle focal length with 12 stations (Table 12) and a comparison of the recovered APs' influence.

Sensor size	1/1.76", ca 7.2x5.4 mm (4 Mil. Pixel)
Image size	2240x1680, 1600x1200, 1120x840, 640x480
Pixel Size	3.21 x 3.21 micron

Table 11: Some features of the Leica Digilux 1.

Parameter	Value	Std. Dev.
Camera constant	7.309 mm	8.68e-04
PP x dir.	0.111 mm	1.48e-03
PP y dir.	-0.239 mm	1.46e-03
K1	-3.748e-03	2.05e-05
K2	3.744e-05	2.94e-06
P1	-1.505e-04	1.07e-05
P2	-1.817e-05	1.11e-06
Affinity factor (*)	0.9986	1.65e-05
$\sigma_{0,priori}$ = 1 pixel	$\sigma_{0,post}$ = 0.05 pixel	

Table 12: The calibration parameters for the Leica Digilux 1 recovered using images with 1600x1200 pixels.
(*) Affine factor = pixel_size_x/pixel_size_y.

In the next table we compared the effect of the recovered APs on an image point with coordinate x=3, y=2 mm according to the different procedure and imaged object (testfield or shelf).

Calibration Method and Object	Infl_x	Infl_y	Vector
Testfield Conventional Calibration (*)	-120	-94	153
Testfield Rotation Bundle Model	-111	-85	14
Shelf Rotation Bundle Model	-102	-71	124
Testfield Rotation Simplified Model	-96	-64	115

Table 13: Influence (in µm) of the APs on the image grid.
(*) = reference data

In Figure 16 another comparison of the influence of the APs on the image grid of the Leica Digilux is presented. The APs are recovered with the different calibration procedures described in the paper's sections.
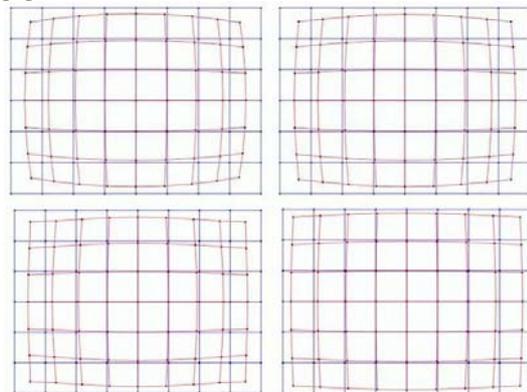
Figure 16: The influence of the recovered APs on the Leica Digilux image grid (3 times amplified). Conventional self-calibrating bundle solution (upper left). Perspective camera model on the rotating testfield sequence (upper right). Perspective model on the rotating shelf sequence (lower left). Simplified perspective camera model on the rotating testfield sequence (lower right).

REFERENCES

Abdel-Aziz, Y., Karara, H., 1971: Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry. Proc. ASP/UI Symposium on Closed-Range Photogrammetry, pp.1-18

Bartels, R., Stewart, G.W., 1972: Solution of the equation AX+XB=C. Communications of the ACM, Vol.15, pp. 820-856

Beyer, H., 1992: Geometric and radiometric analysis of a CCD camera based photogrammetric close-range system. PhD Diss., No 9701, ETH Zurich

Bopp, H., Krauss, H., 1978: An orientation and calibration method for non-topographic applications. PE&RS, 44(9)

Brown, D., 1971: Close-range camera calibration. Photogrammetric Engineering, 37(8), pp. 855-866

Caprile B., Torre, V., 1990: Using vanishing point for camera calibration. International Journal of Computer Vision, 4(2), pp. 127-139

Criminisi, A, 1999: Accurate Visual Metrology from Single and Multiple Uncalibrated Images. Ph.D. Diss., Oxford University

De Agapito, L., Hayman, E., Reid, I., 1998: Slef-calibration of a rotating camera with varying intrinsic parameters. Proc. 9th British Machine Vision Conference, pp. 105-114

Jacobsen, K., 1982: Programmgesteuerte Auswahl zusaetzlicher Parameter. Bildmessung und Luftbildwesen, 50, pp. 213-217

Foerstner, W., 1980: Zur Pruefung zusaetzlicher Parameter in Ausgleichungen. Zeitschrift fuer Vermessungswesen, 105, pp. 510-519

Foerstner, W., 2000: New orientation procedure. Int. Arch. of Photogrammetry and Remote Sensing, 33(3), pp. 297-304

Fraser, C., 1980: Multiple focal setting self-calibration of close-range metric cameras. PE&RS, 46(9), pp. 1161-1171

Fryer, J.G., 1996: Single station self-calibration techniques. Int. Arch. of Photogrammetry and Remote Sensing, 31(5), pp. 178-181

Gruen, A., 1981: Precision and reliability aspects in close-range photogrammetry. The Photogrammetric Journal of Finland, 8(2), pp. 117-132

Gruen, A., 1985: Data processing methods for amateur photographs. Photogrammetric Record, 11(65), pp.576-579

Hartley, R., 1994: Self-calibration from multiple views with a rotating camera. Proc. 3rd ECCV, Vol. 1, pp. 471-478

Hartley. R., Zisserman, A., 2000: Multiview geometry in computer vision. Cambridge Press

Krauss, K., 1996: Photogrammetry, Vol.2. Duemmler Verlag, Bonn

Pöntinen, P., 1999: On the creation of panoramic images from image sequence. The Photogrammetric Journal of Finland, 16(2), pp. 43-67

Pöntinen, P., 2002: Camera calibration by rotation. Int. Arch. of Photogrammetry and Remote Sensing, 34(5), pp.585-589

Remondino, F., 2002: Image sequence analysis for human body reconstruction. Int. Arch. of Photogrammetry and Remote Sensing, 34(5), pp.590-595

Remondino, F., 2003: Recovering metric information from old monocular video sequences. In Gruen, Kahmen (Eds.): VI Optical 3-D Measurement Techniques Conference, Vol. 2, pp. 214-222

Seo, Y., Hong, K.S., 1999: About the self-calibration of a rotating and zooming camera: theory and practise. Proc. of 7th ICCV, pp. 183-189

Slama, C., 1980: Manual of Photogrammetry. ASPRS, Falls Church, Virginia

Stein, G.P., 1995: Accurate camera calibration using rotation with analysis of source error. Proc. of 5th ICCV, pp.230-236

Szeliski, R., 1996: Video Mosaics for Virtual Environments. IEEE Computer Graphics and Applications, 16(2), pp. 22-30

Van den Heuvel, F.A., 1999: Estimation of interior parameters from constraints on line measurements in a single image. Int. Arch. of Photogrammetry and Remote Sensing, 32(5), pp.81-88

Wang, Z., 1990: Principle of Photogrammetry (with Remote Sensing). Press of TU of Surveying and Mapping, Publishing House of Surveying and Mapping, Beeijing

Wester-Ebbinghaus, W., 1982: Single station self-calibration: mathematical formulation and first experiences. Int. Arch. of Photogrammetry and Remote Sensing, 24(5/2), pp. 533-550

Wolf, P., Dewitt, B., 2000: Elements of Photogrammetry. McGraw Hill, New York

Wrobel, B.P., 2001: Minimal solutions for orientation. In: Grün, Huang (Eds.): Calibration and Orientation of Cameras in Computer Vision, Springer 34, pp. 7-62

Zheng, Z., Wang, X, 1992: A general solution of a closed-form space resection. PE&RS, 58(3), pp. 327-338