Recovering metric information from old monocular video sequences

6th Conference on Optical 3D Measurement Techniques, Gruen/Kahmen editors, pp.214-222. September 2003, Zurich, Switzerland

RECOVERING METRIC INFORMATION FROM OLD MONOCULAR VIDEO SEQUENCES

Fabio Remondino
Institute of Geodesy and Photogrammetry - ETH Zurich - Switzerland
fabio@geod.baug.ethz.ch

KEY WORDS: Calibration, Orientation, 3D Reconstruction

ABSTRACT
In this paper we analysis monocular sport videos, digitized from old VHS tape, where players and camera are moving. It is very difficult to recover 3D information of the scene from these sequences, due to noisy measurements, low quality of the images, very small translation of the camera and almost no information concerning its parameters. Classic stereo algorithms are not always suitable and perspective models can easily fail, due to zooming effects and baseline absence. We calibrated a monocular sequence with perspective and projective approaches, comparing the methodologies and the results. Moreover, applying some geometric image invariants, metric measurements of the movements of the players are also recovered.

1. INTRODUCTION

Monocular video sequences are usually acquired with a moving camera (hand-held or on a small railroad) or with a stationary but freely rotating one (e.g. on a tripod). In particular, sports videos (football, basketball) are usually filmed far away from the scene and with a rotating camera. In the last years various techniques have been applied to sport sequences for automatic extraction of features, recovering of camera parameters and analysis of objects information from monocular (Kim et al., 1998) or multiple video sequences (Reid et al., 1996; Bebie, 2000; Pingali et al., 2000; Pera et al., 2001). Usually the problem is formulated within a projective framework because of the absence of camera and object information. Infect, when old videos are analyzed, it is very difficult to recover accurate 3D information of the scene and the camera parameters, mainly because of (1) low image quality, providing for noisy measurements, (2) almost no information concerning the camera parameters and (3) often absence of stereo-view. Classic stereo algorithms to retrieve the camera parameters are not really suitable and perspective models can easily fail due to the continuous changes of the internal parameters or the almost absence of baseline.
In the vision community many techniques have been presented to calibrate a stationary but freely rotating image sequences (Hartley, 1994; De Agapito et al., 1999; Seo et al., 1999): they rely on the homographies between the images and they retrieve the camera parameters with linear or iterative methods. Usually changes of the internal parameters (mainly zooming) are allowed but they often assume zero-skew or known pixel aspect ratio. All these approaches can also recover the position of the principal point in each frame, even if its determination could not be considered reliable due to the high correlation between the parameters.
Apart from camera parameters, other metric information, like persons heights or movements distances, can be recovered from single images or videos (Criminisi et al., 1998). Projective geometry provides different concepts that can be used to recover this knowledge from uncalibrated images (Semple et al., 1952) and it is a good basis for accurate estimation algorithms. Person's height is an important parameter for identification or scene analysis and different techniques have been proposed, in particular for forensic image analysis, to obtain this information from videos (Klasen et al., 1996; Bramble et al., 2001).

The aim of this paper is to extract metric information about the imaged scene and the used camera from old monocular videos of sport games, filmed with stationary but rotating cameras. After the digitizing process, the artifacts created by the interlaced display were removed. Then we retrieve the camera parameters, testing perspective and projective camera models; we are mainly interested in the interior parameters, as, because of very small baselines, the 3D information of the scene can not be recovered with forward rays intersection. Finally, an easy formulation and a reliable approach to measure heights of persons and lengths of their movements in the images are presented.

2. IMAGES ACQUISITION AND ANALYSIS

The images were acquired digitizing an old VHS videotape. The analog signal was imported in a PC through a video-recorder and a sequence of independent frames was obtained. Because of the interlaced display, interlace artifacts are created in the digitized frames and are disturbing the processing of the images as they unsharp the edges and do not allow the exact location of the features. But de-interlacing an image (or a video) is not a perfect process; there is no one right way to do it, even if some commercial software is available. Therefore several different techniques are implemented and tested (Table 1). At the end of our tests, the images obtained with the 'interpolating mode' are used for the processing.

Method	*Action*	*Image size*	*Pro*	*Contra*
Single field mode	Use only even or odd lines	720x288	- very fast & sharp results - 100% de-interlaced	- half resolution lost - lose of sharpness
Resize mode	Reduce the size of the image (4:3 aspect ratio)	384x288	- very fast - sharp results	- lose of resolution
Adaptive de-interlace	Remove interlace artifacts only when present	720x576	- randomly sharp results	- not completely de-interlaced - slow and it eliminates wrong data - it introduces blur effects randomly
Duplicating mode	Remove one filed and duplicate the other	720x576	- almost sharp edges	- not completely de-interlaced - still blur results if there is motion
Interpolating mode	Remove one field and substitute it with the interpolation of the other	720x576	- almost sharp edges	- some smooth effects - slow method
Averaging mode	Each line is replaced with the average of the line above and below	720x576		- slow with smoothed results - not completely de-interlaced
Anisotropic mode	Non-linear noise filter based on anisotropic diffusion (Perona et al., 1990)	720x576	- it removes artefacts without removing lines - it preserves edges	- very slow (many iterations) - too many parameters

Table 1: Different techniques to remove interlace artifacts from digitized images (the original image size is 720x576).

3. CALIBRATION AND ORIENTATION OF MONOCULAR SEQUENCES

Generally calibration and orientation problems are formulated in terms of perspective or projective camera model. Camera models based on perspective collineation have high stability, require a minimum of three corresponding points per image and a stable optics. On the other hand, projective approaches can deal with variable focal length, but need more parameters, a minimum of six corresponding points and are quite instable (equations and coordinates need normalization).
Sport videos are usually acquired with zooming cameras, from relatively long distances, with small or absent translations and mainly with rotations (on a tripod or on the shoulder of the cameraman). In the next two sections we first analyze the case of videos acquired with pure rotations and then we try to recover the camera interior parameters in sequences acquired with a zooming camera undergoing rotations and very small translations.

3.1 Pure rotation case

If the camera undergoes a pure rotation **R** (around the projection center), we have no translation of the camera center and the collinearity perspective model can be approximated with:

$$x' = -c \cdot \frac{r_{11}X + r_{21}Y + r_{31}Z}{r_{13}X + r_{23}Y + r_{33}Z}$$
$$y' = -c \cdot \frac{r_{12}X + r_{22}Y + r_{32}Z}{r_{13}X + r_{23}Y + r_{33}Z} \quad (1)$$

where (x', y') are already corrected of the principal point displacement. A perspective projection can be represented with x=cX/Z and y=cY/Z, therefore the coordinates of a point undergoing pure rotations can be computed as:

$$x' = -c \cdot \frac{r_{11}x + r_{21}y + r_{31}c}{r_{13}x + r_{23}y + r_{33}c}$$
$$y' = -c \cdot \frac{r_{12}x + r_{22}y + r_{32}c}{r_{13}x + r_{23}y + r_{33}c} \quad (2)$$

i.e. the knowledge of 3D object coordinates is not required as the position of a point in one image, after a rotation, can be recovered only with the camera parameters and its position in the previous image. Equation (2) can also be seen as a projective transformation (8 parameters) between the two images and it can be extended to include some additional parameters (e.g. radial distortion, not unity of the aspect ratio or the skew parameter). If all the camera parameters and the rotations angles are know, we can compute directly the position of an image point after the camera rotation using (2). Instead, if we want to recover the correct camera parameters, we could solve:

$$\sum_{i,j} \left[(x_{i,CALC} - x_{i,CORR})^2 + (y_{i,CALC} - y_{i,CORR})^2 \right] \Rightarrow \text{MIN} \quad (3)$$

i.e., for each point i in each image j we minimize the image distance between calculated and measured points. The non-linear minimization can be solved over (c, x_0, y_0, R) differentiating (3) and setting the partial derivatives to zero. A simulation has been performed to test the camera model. A set of 3D points is back-projected onto 6 images that are only rotated between each other around the projection center. The focal length and principal point position are kept fixed. Then, starting from the second image, the coordinates of the points are computed with (2) and compared with the correct ones minimizing (3) over R. The results are shown in Figure 1: while the average RMS of all the points in all the images is 0.12 pixel in x direction and 0.30 pixel in y direction. Therefore, equations (2) could be used to model a camera that undergoes pure rotations around its projection center.

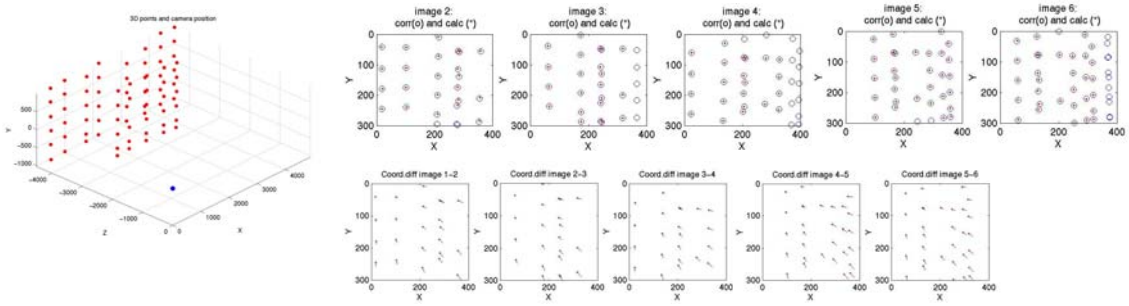

Figure 1: Left: The set of 3D points used in the simulation and the camera position in (0,0,0). Right upper: The image points back-projected onto the images (o) and the image points computed (•) with equations (2) and (3). Right lower: The (amplified) difference image vectors between corrected and computed coordinates.

But if the axis of rotation does not pass through the center of projection, the previous simplification is no more valid. Infect the location of the point in one image is no more independent of its depth and is also related to the camera translation.

3.2 General case: rotation and small translation

A short monocular sequence (Figure 2) is used to analyze this case and some control points, measured semi-automatically in the images with adaptive least square matching (Gruen, 1985), are

defined to recover the metric camera parameters. If the rotation arm is very small compared to the distance of the scene, equations (2) could be employed to calibrate the sequence. But this displacement is usually unknown and anyway too small to use stereo approaches.
We analyze the images with three different camera models. The results of every second frame are presented in Table 2. For each frame, we recovered focal length and aspect ratio as the determination of all the other camera (additional) parameters could lead to degradation (unreliability) of the results.

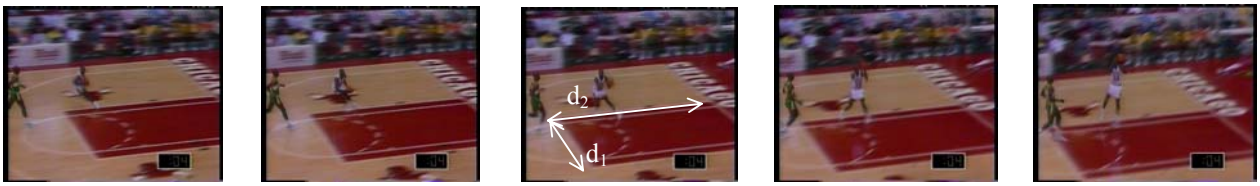

Figure 2: Every second frame of the analyzed video sequence (supposed pixel size of 0.025 mm). In the central image are visualized the two distances measured to control the increasing of the focal length of the camera.

3.2.1 Photogrammetric Space Resection. Space resection is usually applied to solve the orientation of a single image knowing the image coordinates of some control points. If enough control points are available, the interior parameters of the camera can also be recovered. It is a non-linear problem, based on the collinearity condition and it requires initial approximations of the unknowns. A resection is performed for each image. Then all the parameters of the camera are refined with a bundle adjustment (allowing single ray intersection), where (1) we fix the position of the first camera, (2) we allow very small translations and free rotations of the other cameras, (3) we fix the principal points all the cameras in the middle of the images, (4) we estimate only the focal lengths and pixel aspect ratio of the cameras. The results of the adjustment are presented in Table 2: they show a constant increasing of the focal length and an almost constant aspect ratio of the pixel. The correct increasing of the focal length can also be demonstrated by the increasing of the two segments d_1 and d_2 drawn in the central image of Figure 2.

3.2.2 Direct Linear Transformation (DLT). This camera model, developed in the 70's to solve the collinearity condition with a projective model (Abdel-Aziz et al, 1971), does not require initial approximations of the unknowns. DLT model is represented by 11 coefficients, which form a linear mapping between world and image coordinate system. The number of coefficients increases if we consider also the additional parameters to model the lens distortion. The solution of the coefficients is derived by means of least squares and from the recovered 11 coefficients of the projective transformation, the 9 parameters of the perspective camera model (X_0, Y_0, Z_0, ω,ϕ, κ, x_P, y_P, f) are then *sequentially* derived. DLT has mainly two limitations, i.e. the requirement for at least 6 control points and a lower accuracy of the solution compared to rigorous bundle adjustment. In our test, the DLT coefficients are computed for every frame and then refined minimizing the image distance between the reprojected points and the measured ones. DLT results (Table 2) differ a little from resection's results but confirm the continuos increasing of the focal lengths and the not unity of the affinity factor.

3.2.3 Decomposition of 3D Projective Camera Model. The projective camera model is mathematically written as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} P_1 & P_2 & P_3 & P_4 \\ P_5 & P_6 & P_7 & P_8 \\ P_9 & P_{10} & P_{11} & P_{12} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4)$$

with x, y the image coordinate, (X, Y, Z) the object coordinates and P_i are the coefficients of the projective matrix **P**.

Using some matrix factorization techniques, we can relate the projective model (4) to the DLT coefficients (Hartley et al., 2000; Seedahmed et al., 2002). Infect the DLT model can be seen as a non linear version of (4) and establishing a relationship between the projective and the perspective camera model, we can *simultaneously* derived the interior parameters of the camera in a matrix

form. The projective matrix **P** can be written as: $\mathbf{P} = \mathbf{KR}(\mathbf{I_{3x3}} \mid -\mathbf{X_0})$, with **R** the 3x3 rotation matrix, $\mathbf{X_0}$ the camera exterior orientation vector and **K** a calibration matrix, generally so defined:

$\mathbf{K} = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$, with f_x and f_y the focal length along x and y axis, s the skew factor, (x_0, y_0) the principal point position. Therefore we can write that $\mathbf{KR} = \mathbf{P_{13}}=(\mathbf{p_1}\ \mathbf{p_2}\ \mathbf{p_3})$, with $\mathbf{p_i}$ the i^{th} column of **P**. As **R** is an orthogonal matrix and $\mathbf{RR^T}=\mathbf{I}$, we can form a quadratic equation and write:

$$\mathbf{(KR)(KR)^T}=\mathbf{P_{13}(P_{13})^T} \quad (5)$$

$$\mathbf{KK^T}=\mathbf{P_{13}(P_{13})^T} \quad (6)$$

Now, knowing $\mathbf{P_{13}}$, we could use Cholesky factorization to recover the matrix **K**. But the direct factorization of the right term of (6) will not provide for the correct results as $\mathbf{P_{13}(P_{13})^T}$ should be a positive definite symmetric matrix to be uniquely decomposed as $\mathbf{KK^T}$, with **K** an upper-triangular matrix with positive diagonal elements. Therefore we have first to invert the right side to recover a correct factorization result. The factorization leads to an un-normalized **K** matrix and we have to divided it by *K(3,3)* to recover the correct calibration matrix.
This decomposition is performed on the coefficient matrix recovered with DLT approach. As shown in Table 2, the results are quite instable compared to the parameters sequentially recovered from the DLT model and with the space resection approach.

Images			*Space Resection*				*DLT*				*Decomposition*	
Nr.	Distances d_1,d_2		Focal L.	A. Ratio	RMS x	RMS y	Focal L.	A. Ratio	RMS x	RMS y	Focal L.	A. Ratio
66	190.0	486.3	91.4	1.13	68	46	89.3	1.17	90	54	79.5	1.03
68	192.1	500.6	92.1	1.13	59	40	91.1	1.16	65	50	87.4	1.09
70	195.0	504.8	93.9	1.13	55	47	93.7	1.17	60	52	85.3	1.08
72	200.6	515.2	95.4	1.14	73	61	95.4	1.17	73	66	89.8	1.06
74	204.5	519.4	98.1	1.14	72	69	96.1	1.18	78	63	90.1	1.10

Table 2: Results of the three calibration algorithms. Focal length and pixel aspect ratio are recovered. The RMS of the image residuals are given in μm.

4. RETRIVING METRIC MEASUREMENTS FROM THE SEQUENCE

Projective geometry is strongly used in this section, with the goal of recovering metric distances and dimensions of people in the images. Object and image points are represented with homogeneous vector (a_1, a_2, a_3); a line can be represented with the cross product of two homogeneous points while a point is given by the cross product of two lines (i.e. their intersection).

4.1 Vanishing points and lines

Assuming perfect projection, a set of parallel lines in the scene is projected into a set of lines in the image that meet in a common point, the *vanishing point*. The three vanishing points can be real (finite) or ideal (at infinite) and a pair of them that lie on the same plane in the scene define a *vanishing line* in the image space. Different approaches have been presented to detect the vanishing points of an image (Collins, 1993; Van den Heuvel, 1998, Liebowitz, 2001, Remondino, 2003). The majority of the methods rely on straight-line segments extracted from the image and the use of a parameter space to identify the three orthogonal directions. If low quality images are used, line detectors do not provide reliable results and a manual identification of the end points of a line is required. In general, once that a set of lines l_i is identified, we can estimate the associated vanishing point v minimizing the sum of the squares of the perpendicular distances of the line l_i from v:

$$\sum_i d_\perp^2(l_i, v) \Rightarrow MIN \quad (7)$$

where the minimization is over v. If only two lines are involved, the minimization to find the closest point near the lines is reduced to the cross product of the lines. In the general case, we can find the

solution setting the partial derivatives of (7) to zero, forming the normal equations and solving for *v* by means of least squares.

4.2 The Cross-ratio

Given 4 points on a line, with $d(P_i, P_j)$ the Euclidean distance between P_i and P_j, their *cross-ratio*

$$CR(P_1,P_2,P_3,P_4) = \frac{d(P_3,P_1)\cdot d(P_4,P_2)}{d(P_3,P_2)\cdot d(P_4,P_1)} \tag{8}$$

is preserved under any projective transformation (Figure 3-A). As projective geometry does not preserve distances or ratios of distances but ratio of ratio of distances, it follows that:

$$CR(P_1,P_2,P_3,P_4) = CR(P'_1,P'_2,P'_3,P'_4) \tag{9}$$

This concept goes back to (Semple et al., 1952) and other permutations of the points in (8) will also lead to an image invariant. There are 24 possible permutations of the four points P_i, but only 6 distinct values of the CR within the permutations. These different values could create confusions in using CR as index for measurements, because the order of the points along a line can change after a projective transformation (even if the 6 values can be obtained from any one of them).

4.3 Measurements between parallel planes using vanishing points and cross-ratio invariant

The cross-ratio invariant can also be applied to points lying on (parallel) planes. Consider Figure 3-B, where point T and B, lying respectively on plane P' and P, are at a distance H and perpendicular to a reference direction V_3. In image space they are specified respectively by the corresponding image points t and b that are on two planes defined by the two vanishing points v_1 and v_2 (Figure 3-C). The image point c is defined as the intersection of the line joining the corresponding points with the vanishing line l_{v1v2}. The image point c (representing the camera center C in Figure 3-D) lies on a plane at distance H_C from the reference plane P. With this configuration, the four points b, t, c and v_3 are aligned (along the vertical reference direction) and they define a cross-ratio. At the same time, in object space, the points B, T, C' and V_3 define the same cross-ratio, therefore, from (9):

$$CR = \frac{d(b,c)\cdot d(t,v_3)}{d(t,c)\cdot d(b,v_3)} = \frac{d(B,C)\cdot d(T,\infty)}{d(T,C)\cdot d(B,\infty)} \tag{10}$$

as V_3 in object space is at infinity. The right side of (10) becomes $H_C/(H_C\text{-}H)$ and we obtain:

$$\frac{H}{H_c} = 1 - \frac{d(t,c)\cdot d(b,v_3)}{d(b,c)\cdot d(t,v_3)} \tag{11}$$

If a *reference distance* H is known, we can compute the distance of the camera H_C and then any other distance between two planes perpendicular to the reference direction. The reference direction does not need to be the vertical one; moreover, if the camera position C is between the points B and T, the cross ratio is still valid, but equation (11) is slightly different (because of the different order of the points).

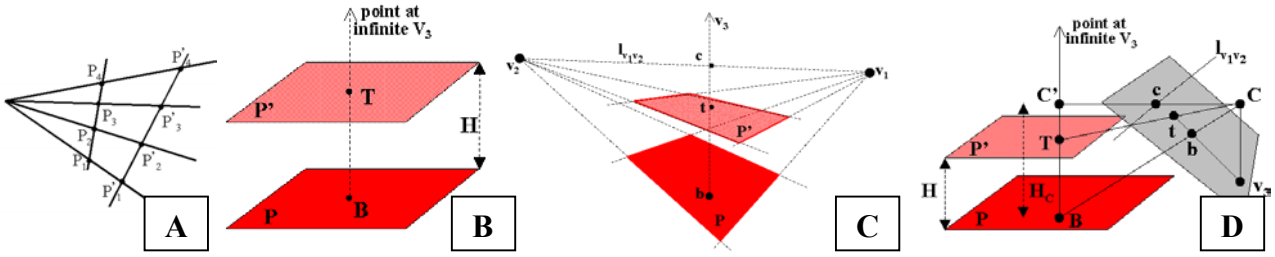

Figure 3: Four points aligned defining the cross-ratio invariant (A). Two planes, in object (B) and image (C) space, perpendicular to a reference direction. The cross-ratio defined by the distances between the parallel planes (D).

Similar results, with an algebraic representation of (11) and an uncertainty analysis, are presented in (Criminisi, 1999), involving the 3x4 projection matrix of the camera and complicate formulas but avoiding possible problems with the order of the cross-ratio.

4.4 Accuracy of the measurements

The covariance estimation of the measurements can be handled with the error propagation law. If we consider our equations as a continuously differentiable function $\mathbf{y}=f(\mathbf{x})$, with $\mathbf{\Sigma_{xx}}$ the covariance matrix of the data $\mathbf{x}$, we can find the covariance matrix $\mathbf{\Sigma_{yy}}$ of $\mathbf{y}$ as

$$\mathbf{\Sigma}_{yy}=\nabla f\, \mathbf{\Sigma_{xx}}\, \nabla f^{T} \tag{12}$$

with ∇ operator representing the Jacobian of $\mathbf{y}$ function, i.e. $\partial\mathbf{y}/\partial\mathbf{x}$.

The accuracy of the measurements obtained with equation (11) depends on the accuracy of the measured distances d_i between the points, the variance of the vanishing points and the accuracy of the reference distance. The precision of the points defining the distances is determined by the (manual) measurement or by the cross product used to find them. The variance of a cross product $\mathbf{c}=\mathbf{x}\times\mathbf{y}$ can be computed for each single component of the resulting homogeneous vector using (12):

$$\begin{aligned}
\sigma^2_{c_1} &= y_3^2\sigma^2_{x_2} + x_2^2\sigma^2_{y_3} + y_2^2\sigma^2_{x_3} + x_3^2\sigma^2_{y_2} \\
\sigma^2_{c_2} &= y_1^2\sigma^2_{x_3} + x_3^2\sigma^2_{y_1} + y_3^2\sigma^2_{x_1} + x_1^2\sigma^2_{y_3} \\
\sigma^2_{c_3} &= y_2^2\sigma^2_{x_1} + x_1^2\sigma^2_{y_2} + y_1^2\sigma^2_{x_2} + x_2^2\sigma^2_{y_1}
\end{aligned} \tag{13}$$

The precision of the vanishing point determined with (7) is related to the residuals of the least squares and is controlled by its covariance matrix. The accuracy of the euclidean distance between two points is instead given by:

$$\sigma^2_d = \left(\frac{\partial d}{\partial x_i}\right)^2 \sigma^2_{x_i} + \left(\frac{\partial d}{\partial y_i}\right)^2 \sigma^2_{y_i} + \left(\frac{\partial d}{\partial x_j}\right)^2 \sigma^2_{x_j} + \left(\frac{\partial d}{\partial y_j}\right)^2 \sigma^2_{y_j} \tag{14}$$

Finally, we can compute the variance of the estimated measurement H (11) as:

$$\sigma^2_H = \left(\frac{\partial H}{\partial H_C}\right)^2 \sigma^2_{H_C} + \sum_i \left(\frac{\partial H}{\partial d_i}\right)^2 \sigma^2_{d_i} \tag{15}$$

and the same when we have to estimate H_C.

4.5 Mosaic of the images

The result (11) can be applied to single image metrology (Criminisi, 1999) or to sequence of images, where camera and person are moving. But due to the camera movements, the scenario can change and it is not always possible to determine the required *reference distance* or the entire movement of the person. Therefore, in these cases, a mosaic of the sequence has to be performed.

Figure 4 shows three frames of a basketball match sequence where we want to recover the height of the jump and its length. It is clear that is not possible to perform measurements on the single images, as e.g. in the central frame, we do not see the floor of the field; therefore a mosaic of the images has to be created.

Figure 4: The three images used to create a mosaic of the player's movement and the resulted mosaic (right).

The mosaic is realized with a projective transformation (8 parameters) between the central image (reference) and the other two. The corresponding points are measured semi-automatically with Adaptive LSM (Gruen, 1985) and the transformation parameters are computed with a least squares adjustment. Then the transformed images are merged together automatically, but no radiometric correction is performed. The resulted mosaic is presented in Figure 4, right.

4.6 Recovering metric measurements: distances and heights

Using the created mosaic, we can now derive e.g. the length of the jump and its height. At first the 3 vanishing points of the image are recovered. Due to the low image quality, a line detector did not produce accurate results; therefore with manual measurements, the end points of the segments representing the convergent lines are identified in each direction and then equation (7) is applied. Afterward some reference distances (Figure 5-A) are required and must be measured in the image:

- the height of the basket (Hr=3.05 m): the base point b is identified as the intersection between two lines: (1) the line through the top point t on the basket and the vertical vanishing point v_3 and (2) the line through the vanishing point v_1 and the point i (middle point of the upper red area);
- the distance between the baseline and the free-throw line (H'r=5.8 m);
- the width of the red area (H"r=4.9 m).

Then we use this knowledge and equation (11) to recover the lengths presented in Table 3. To compute the length of the jump, we suppose that the player is moving in a vertical plane defined by his starting and ending position. To check the metrology technique, its reliability and repeatability, each measure is repeated 3 times: the presented measures are an average of the results while in the last column of Table 3 is reported the standard deviations of the measures.

Kind of length/distance	*Length (m)*	*Accuracy (cm)*	*measures STD*
height of the player at the beginning of the jump	1.71	± 2.7	0.033
height of the player at the end of the jump	1.52	± 2.3	0.028
length of the jump (distance b_1b_3)	4.94	± 3.7	0.093
height of the jump (ball)	3.28	± 3.0	0.030
height of the jump (waist)	2.02	± 2.8	0.033
height of the second player	1.97	± 2.8	0.031

Table 2: Results of the sequence metrology and related accuracy (average of 3 measures).

The correct height of the (standing) jumping player is 1.98 m while the height of the second player is 2.01 m.

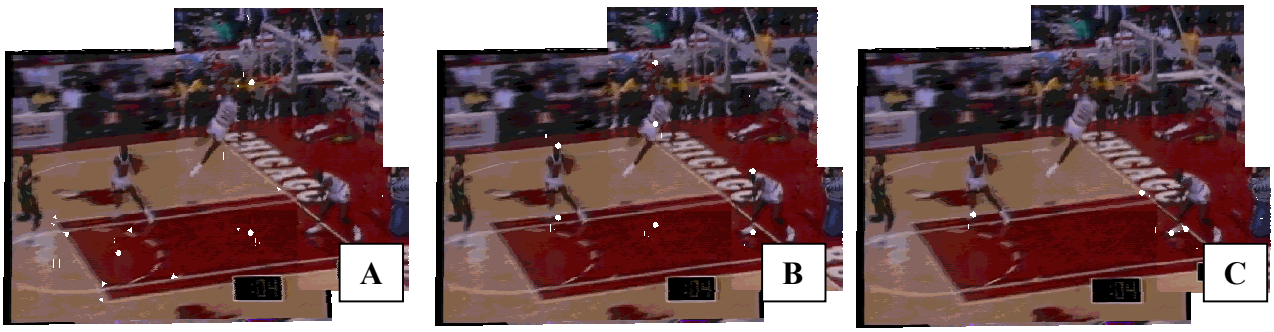

Figure 5: Distances between parallel planes with respect to a reference plane. The reference distances on the basketball field (A), the distances measured to recover the length and height of the jump (B and C).

5. CONCLUSION

In this paper we discussed the possibility of recovering metric information from old monocular sequences of sport events. The techniques used to find the camera parameters produced accurate and fairly similar results, even if the network geometry was not adequate, in particular for the perspective camera model. On the other hand, the presented metrology approach showed to be reliable and applicable to other class of problems (traffic accident, video surveillance) to provide useful information about the course of events or the size of items and persons.

ACKNOWLEDGEMENT

The author would like to thank Devrim Akca of the Institute of Geodesy and Photogrammetry for his help with the mosaic program.

REFERENCES

1. Abdel-Aziz, Y.I., Karara, H.M., 1971: Direct Linear Transformation for comparator coordinates into object space coordinates in close-range photogrammetry. Proc. of ASP Symposium on Close-Range Photogrammetry, pp. 1-18
2. Bebie T., 2000: A Video-Based 3D-Reconstruction of Soccer games. Proc. of EuroGraphics
3. Bramble, S., Compton, D., Klasen L., 2001: Forensic image analysis. 13th Interpol Forensic Science Symposium
4. Collins, R.T., 1993: Model acquisition using stochastic projective geometry. Ph.D. thesis, Computer Science Dep., University of Massachusetts
5. Criminisi, A., Zisserman, A., Van Gool, L., Bramble, S., Compton, D., 1998: A new approach to obtain height measurements from video. Proc. of SPIE, Vol. 3576, pp. 227-238
6. Criminisi, A, 1999: Accurate Visual Metrology from Single and Multiple Uncalibrated Images. Ph.D. Dissertation, Oxford University, Department Eng. Science, Robotics Research Group
7. De Agapito, L., Hartley, R., Hayman, E., 1999: Linear calibration of a rotating and zooming camera. Proc. of 14th CVPR Conference
8. Gruen, A., 1985: Adaptive least squares correlation: a powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, 14(3), pp.175-187
9. Hartley, R., 1994: Self-calibration from multiple views with a rotating camera. Proc. 3rd ECCV Conference, Vol. 1, pp. 471-478
10. Hartley. R., Zisserman, A., 2000: Multiview geometry in computer vision. Cambridge Press.
11. Kim, T., Seo, Y., Hong K.S., 1998: Physics-based 3D position analysis of a soccer ball from monocular images sequences. Proc. of ICCV Conference, pp. 721-726
12. Klasen, L., Fahlander, O., 1996: Using videogrammetry and 3D image reconstruction to identify crime suspects. Proc. of SPIE, Vol. 2942, pp. 162-169
13. Liebowitz, D, 2001: Camera Calibration and Reconstruction of Geometry from Images. Ph.D. Dissertation, Oxford University, Department Eng. Science, Robotics Research Group
14. Pera, J., Kovacic, S., 2001: Tracking People in Sport: Making Use of Partially Controlled Environment. 9th CAIP Conf., Lecture notes in computer science, 2124, Springer, pp.374-382
15. Perona, P., Malik, J., 1990: Scale-space and edge detection using anisotropic diffusion. IEEE Transaction on PAMI, Vol. 12, Nr. 7, pp.629-639
16. Pingali, G., Jean, Y., Opalach A., 2000: Ball Tracking and Virtual Replays for Innovative Tennis Broadcasts. Proc. of 15th CVPR Conference, Vol. 4, pp. 152-156
17. Reid, I., Zisserman, A., 1996: Goal-directed video metrology. Proc. of 5th ECCV Conference
18. Remondino, F., 2003: 3D Reconstruction of Static Human Body with a Digital Camera. Proc. of Videometrics VII, SPIE Electronic Imaging, Vol. 5013, pp. 38-45
19. Seedahmed, G., Schenk, T., 2002: Retrieval of the calibration matrix from the 3D projective camera model. Proc. of the 15th International Conference on Vision Interface, Calgary, Canada
20. Semple, J.G., Kneebone, G.T., 1952: Algebraic Projective Geometry, Oxford Press
21. Seo, Y., Hong, K.S., 1999: About the self-calibration of a rotating and zooming camera: theory and practise. Proc. of ICCV Conference, pp. 183-189
22. Van den Heuvel, F.A., 1998: Vanishing point detection for architectural photogrammetry. Intern. Arch. of Photogrammetry and Remote Sensing, Vol. XXXII part 5, pp.652-659