Camera Auto-Calibration from Articulated Motion

Paul Kuo
DIRC
Kingston University
London, UK

Jean-Christophe Nebel
DIRC
Kingston University
London, UK

Dimitrios Makris
DIRC
Kingston University
London, UK

Abstract

This paper presents a novel auto-calibration method from unconstrained human body motion. It relies on the underlying biomechanical constraints associated with human bipedal locomotion. By analysing positions of key points during a sequence, our technique is able to detect frames where the human body adopts a particular posture which ensures the coplanarity of those key points and therefore allows a successful camera calibration. Our technique includes a 3D model adaptation phase which removes the requirement for a precise geometrical 3D description of those points. Our method is validated using a variety of human bipedal motions and camera configurations.

1. Introduction

Geometric camera calibration is a valuable task because it reveals the relationship between the 3D space that is viewed by the camera and its projection on the image plane. Therefore, it is a key element for the interpretation of 3D articulated motions.

The common practice for calibrating cameras is to consider 2D views of rigid calibration patterns with known 3D structure [1]. Although such a task seems to be straightforward, it is not often practical e.g. the camera is not accessible or software deployment is planned for a large number of cameras.

As a consequence, many auto-calibration methods have been proposed that exploit either the ego-motion of the camera [2][3][4] or observations of moving objects [5][6] [7][8]. However, in both cases, the observed scene or objects are assumed rigid.

This paper presents a novel auto-calibration method from unconstrained human body motion which relies on the underlying biomechanical constraints associated with human bipedal locomotion. This general approach automatically detects frames within a sequence which are suitable for camera calibration and does not require a precise geometrical 3D description of the human body. This method was validated using a variety of human bipedal motions and camera configurations.

2. Previous work

Tsai [1] introduced camera calibration based on the knowledge of the coordinates of 3D points and their 2D image plane projections. He presented mathematical solutions for either 7 non-coplanar points or 5 coplanar points. However, these methods require that the camera views a calibration object of known geometry. As a consequence of this restriction, considerable research effort was invested in developing auto-calibration techniques.

Many auto-calibration methods exploit the ego-motion of the camera. For instance, Luong and Faugeras [1] and Pollefeys et al [3] used Kruppa equations to solve the calibration model. This assumes that intrinsic parameters are constant. However, there is no constraint regarding the viewed scene. There are two main drawbacks for those methods: first the Kruppa equations are highly complex, which is time consuming; secondly the algorithms usually require a large number of frames from different views. An alternative approach was suggested by Armstrong et al. [4]. They constrained the problem by assuming that the camera is under planar motion and also introduced the method of vanishing points.

However, the above methods cannot be used in the visual surveillance scenario, since the camera is usually fixed. Many researchers attempted to solve this issue by exploiting the observed activity of the scene and based their proposals on the assumption of constant human height and planar human walking. For instance, Renno et al [5] presented an auto-calibration method to estimate the relative position of the camera to the plane of motion. Lv et al. [6] obtained 3 orthogonal vanishing points from 3 different locations of the human in the sequence and used [4] to obtain camera intrinsic and extrinsic parameters. Krahnstoever and Mendonca [7][8] suggested a Bayesian extension of the previous method to improve the accuracy.

Approaches that exploit the observed activity of the scene were also proposed for the calibration of surveillance multiple camera systems, e.g Lee et al [9] and Black and Ellis [10] estimated homography transformations for pair of camera views, Stauffer and

Tieu [11] proposed a common planar model for groups of overlapped camera views , while Makris et al [12] tackled the problem of non-overlapped camera views.

While the methods described in [9][10][11] require coplanar motion of a large number of humans that are modeled by a rigid 1D model, our method only assumes 3D articulated motion of a single human.

3. Calibration using coplanar points

3.1.Camera model

We adopt the pin-hole camera model proposed by [1] in Figure 1.

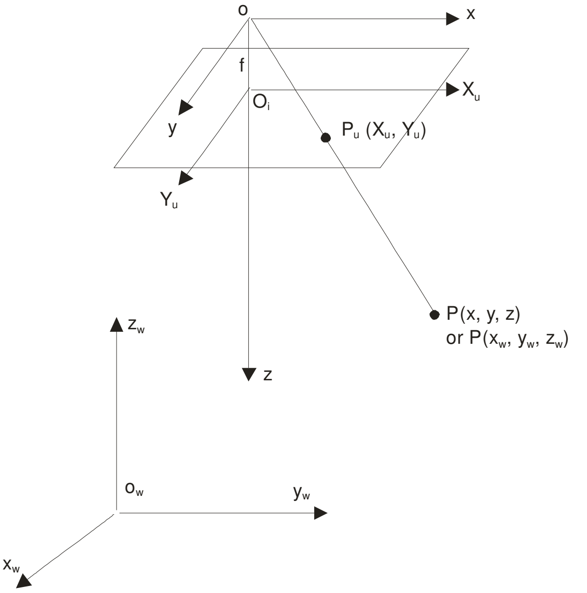

Figure 1: Camera Geometry and perspective projection

We define the world coordinate system ($x_w y_w z_w$), and the camera coordinate system (xyz). ($X_u Y_u$) defines the image plane which is a front plane perpendicular to the optical axis z. The focal length f is the distance between the image plane and the optical centre o. The rotation matrix R and the translation vector T are used to transform the object world coordinate system (x_w, y_w, z_w) to the camera 3D coordinate system (x, y, z), according to Eq. (1):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \qquad (1)$$

The rotation matrix R is defined by the rotation angles R_x, R_y, R_z the translation vector is defined by the translation components T_x, T_y, T_z around and along the three axes respectively.

In the context of this paper, we want to calibrate the six “extrinsic” parameters (R_x, R_y, R_z, T_x, T_y, T_z) and the “effective” focal length f_e measured in pixels/mm.

3.2. Camera parameters estimation

Coplanar calibration [1] relies on a set of known 3D coplanar points and their projected locations on the image plane. The calibration is divided into two stages. The first stage is to estimate the rotation parameters (R_x, R_y, R_z) and translations along x, and y axes (T_x, T_y) while the second stage is to resolve the ambiguity of the depth (T_z) and the effective focal length (f_e).

To simplify the calibration process and without loss of generality, the coplanar points are arranged to be on an arbitrary $x_w y_w$ plane of the world coordinate system where z_W=0.

In the stage 1, the parameters R_x, R_y, R_z, T_x and T_y are estimated by Eq. (2):

$$\begin{bmatrix} Y_{ui}x_{wi} & Y_{ui}y_{wi} & Y_{ui} & -X_{ui}x_{wi} & -X_{ui}y_{wi} \end{bmatrix} \begin{bmatrix} T_y^{-1}r_1 \\ T_y^{-1}r_2 \\ T_y^{-1}T_x \\ T_y^{-1}r_4 \\ T_y^{-1}r_5 \end{bmatrix} = X_{ui} \qquad (2)$$

where r_i, i=1..9, are the elements of the matrix R.

In stage 2, T_z and f_e are computed as follows:

$$\begin{bmatrix} y_i & -d_y Y_i \end{bmatrix} \begin{bmatrix} f_e \\ T_z \end{bmatrix} = w_i d_y Y_i \qquad (3)$$

where y_i and w_i are estimated by Eq. (4) and Eq. (5) respectively:

$$y_i = r_4 x_{wi} + r_5 y_{wi} + T_y \qquad (4)$$

$$w_i = r_7 x_{wi} + r_8 y_{wi} \qquad (5)$$

In section 4, we will exploit the fact that the linear system of Eq. (3) is over-determined (2 unknown variables, 5 equations) and therefore a set of solutions can be computed.

4. Human biomechanics and 3D model

4.1. Unconstrained human body motion

Although many calibration methods are based on observation of human activities, they rely on specific constraints such as constant velocity, linear or periodic motion. Study of human biomechanics, however, reveals that human motion itself provides some specific constraints. In this section, we show that some of those constraints can be utilised for auto-calibration purposes.

Many human activities rely on some form of walking (e.g. running and dancing). Therefore, the underlying mechanical constraint associated with bipedal locomotion is worth investigating. When walking, each leg alternately undergoes two phases –"support" and "swing" phases [13]. The hip bone is rotated by being pushed and pulled by the leg and the shoulder bone is rotated in a reverse direction to compensate the angular momentum generated by the legs. The joints on the arms, shoulders and legs also move in order to maintain the balance of the body. As the hip and shoulder bones are rotated oppositely, it is clear that at some stage the joints on these bones are coplanar. This stage is defined as "mid-stance" phase [14], which corresponds to one leg swinging across the other (supporting) leg. At this specific instant, 5 points are coplanar: left shoulder, right shoulder, left hip, right hip, and mid-hip (see Figure 2). Therefore, if this instant can be detected, these points are sufficient for coplanar auto-calibration (see Section 3.2).

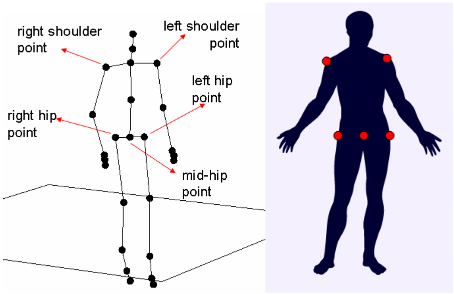

Figure 2: a) Articulated human body model and b) position of the five coplanar points during mid-stance position highlighted on a human silhouette.

We performed further studies using motion capture data to check if other groups of 5 coplanar points could be used for auto-calibration. They revealed that other coplanar configurations exist involving also either the neck or the top of the head. In the rest of this paper, we will limit our experiment to the 5 points corresponding to the "mid-stance" phase.

4.2. 3D human body model

Coplanar calibration relies on a set of 3D coplanar points whose 3D coordinates are known and their 2D image plane projections. Since human bodies vary with every individual and are not rigid objects, a single 3D model cannot capture the relative positions of the 3D coplanar points on a plane of the world coordinate system. Therefore, we designed a methodology which allows deformation of an initial coarse 3D model to produce 3D models tailored to specific individuals and postures. Since camera calibration is scale dependent, in this work we fix the width of the hip.

The starting point of the process is a 3D human model produced by Leonardo Da Vinci as a result of his study of the human body [11]. This model is used as an initial seed to generate an initial model search space, K_0, containing all allowed human configurations. Then models of K_0 are used to calibrate the camera for each frame. A calibration accuracy criterion defined in Section 5 is used to select the 3D models which have provided the best estimates of the camera focal length (f_e). A new seed model can then be calculated using the selected models and a higher resolution model search space is generated. This process is iterated until a 3D human model allows the accurate calculation of the camera focal length.

5. Auto-calibration method

Our novel auto-calibration method is based on the coplanar calibration process described in section 3.2. It takes a sequence of images as an input and generates the intrinsic parameters of the camera and the extrinsic parameters between the camera and the person at "mid-stance" frames. In the process, it detects the best frame showing the "mid-stance" posture and generates a 3D model of the 5 coplanar points.

Our method is much more general than traditional auto-calibration techniques since it does not need a precise 3D description of the required 5 coplanar points (or key points) and is able to detect automatically which frame of a sequence, if any, can provide these coplanar points. In this piece of work, we assume that the image processing task of extracting the key points from each frame of the sequence is solved.

The method relies on the fact that the linear system of Eq. (3), which is used to compute T_z and f_e, is over-determined. Hence, it can provide 10 estimates of f_e and T_z: f_e^i and T_z^i with i=1..10. If the key points are coplanar and a perfect 3D description of their positions (i.e. 3D model) is available, those estimates are all identical. Otherwise their variability reflects errors regarding key points coplanarity and 3D model. We propose to use the standard deviation of the f_e^i to select the frames of a sequence where the key points are coplanar and optimise the 3D model which fits those points.

If we define j as a frame of the sequence J and k_r as a model of the model search space K_r at resolution r, the focal length estimate is calculated using the mean value of the best focal length estimates: $f_e = (\mu_{f_e})_{j,k_r}$, where j and k_r are estimated by minimising the standard deviation σ_{f_e}.

Inputs to our system are a set of frames where the key points have been extracted and an initial coarse 3D model, s_o (or seed model) representing the positions of these points. Then, we follow an iterative process of increasing model resolution, r. The process is organised in four consecutive steps:

- generation of 3D model search space (K_r) using model seed (s_r)
- estimation of camera parameters for each frame ($j\in J$) for each model ($k_r\in K_r$) using Eq. (2) and (3)
- selection of frames and models (j, k_r) according to standard deviations of estimated focal lengths (see next paragraph for details)
- and 3D model adaptation: generation of a new seed (s_{r+1}) based on selected k_r

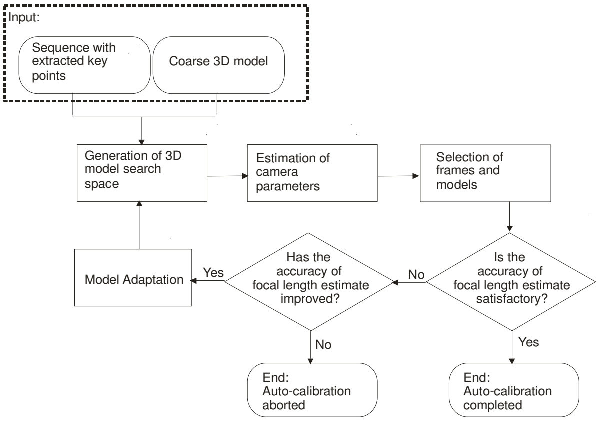
Input:
Sequence with extracted key points
Coarse 3D model
Generation of 3D model search space
Estimation of camera parameters
Selection of frames and models
Is the accuracy of focal length estimate satisfactory?
No
Has the accuracy of focal length estimate improved?
Yes
Model Adaptation
No
Yes
End: Auto-calibration aborted
End: Auto-calibration completed

Figure 3: Auto-calibration method description

The process stops once either the standard deviation of the estimated focal length reaches a certain minimum threshold or it starts diverging. The later situation happens if key points are never coplanar in the sequence.

Since selection of frames and models is key to our method, it is further described here. First for each frame and model configuration (j, k_r), $(\sigma_{f_e})^i{}_{j,k_r}$ is computed using $(f_e)^i{}_{j,k_r}$. Configurations with large standard deviations are discarded. Then, for each of the remaining configurations, a focal length is calculated $(\mu_{f_e})^i{}_{j',k'_r}$. Finally $(\sigma_{f_e})_{j',k'_r}$ is calculated for all these estimated focal lengths. Configuration within one standard deviation from $(\mu_{f_e})_{j',k'_r}$ are used for generating the next seed model.

We have described an auto-calibration method which relies on minimising the standard deviation of estimated focal lengths. It is general since it can be applied to any sequence of moving objects where 5 points are expected to be coplanar at some instant during the sequence. In the following section, our methodology is validated using different motions of the human articulated body.

6. Experimental results

In this section, three experiments are conducted to demonstrate the validity of our method. We use motion capture data so that the ground truth regarding the 3D positions of the points of interest is known. The points of interest (i.e. left shoulder, right shoulder, left hip, right hip, and mid-hip) from the articulated human were projected with set camera models and their 2D coordinates were used as input to our method. In the first experiment where we assume the 3D model is known, we validate our assumption regarding the use of standard deviation as an indicator of focal length accuracy (see Section 4.2). In the second experiment, the algorithm is tested using three different camera focal lengths on four different sequences. Finally, in the last experiment we exhaustively test our algorithm for all possible camera angles for a given focal length. Since the 3D model is unknown in the last two experiments, they validate our model customisation procedure.

Figures 4 and 5 show the results of the first experiment. It reveals the relationship between focal length, point coplanarity and standard deviation of focal length for a standard walking sequence and a sequence containing suspicious movements (sneaky walking). The camera is set on the floor at a distance of 5m from the subject and has a focal length f=50.8 mm.

Figures 4 a) and 5 a) show the sum of distance of the 5 points to their “least-square-fit plane” (i.e. coplanar error) for all frames of both sequences. On Figure 4, there are 7 “coplanar instants” (e.g., the frame 17, 44, 67...etc) which correspond to the periodic mid-stance positions during a standard walking. Although sneaky walking is not periodic, 8 coplanar instants can be detected nevertheless.

As it can be seen on Figures 4 b) and 5 b), the more coplanar the points are, the more accurate the estimation of focal length is (the horizontal purple line is the expected value: f=50.8). Figures 4 c) and 5 c), also reveal that the standard deviation of the estimated focal length is small when the points are coplanar. This result justifies our choice to use the standard deviation as criterion of coplanarity. Therefore, we can trust estimations whose standard deviation is minimal. Here focal lengths are predicted with an accuracy of 0.06% - frame 44 - (respectively 0.94% - frame 134) for the standard (respectively sneaky) walking.

In the second experiment, the algorithm is tested by using three different camera focal lengths (30, 50 and 100 mm) on four different sequences- standard walking, sneaky walking, female cat walking and running. The camera is set to simulate a street surveillance scenario: the camera is 5m above and 6m away from the subject with a look-down angle of 45 degrees.

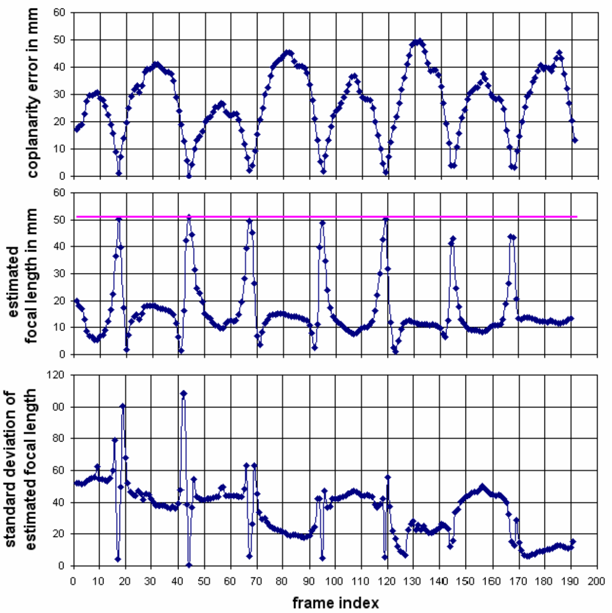

Figure 4: Relationship between a) point coplanarity, b) focal length and c) its standard deviation for standard walking.

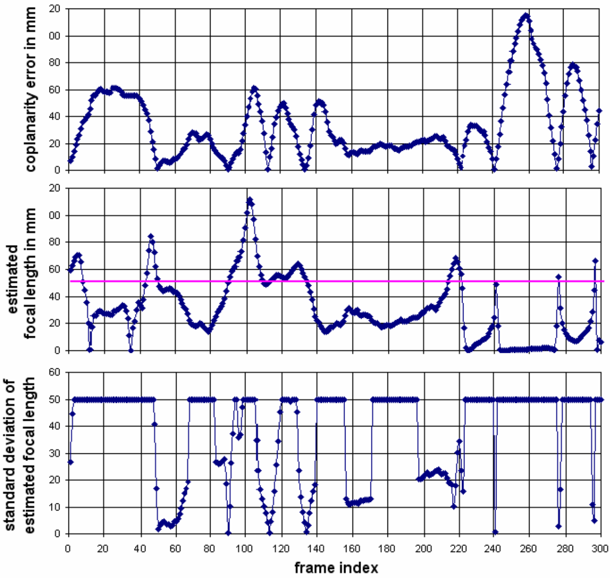

Figure 5: Relationship between a) point coplanarity, b) focal length and c) its standard deviation (values are capped at 50) for sneaky walking.

Estimated focal lengths are presented as percentage error. We also show the Root-Mean-Square (RMS) error between the reconstructed 3D points and their positions in the motion capture data. The results are fairly accurate, as shown in Table 1. As it can be seen, auto-calibration on the standard and cat walking generally performs better than the other two sequences where there is no instant where key points have good coplanarity.

Table 1: *Results for different focal lengths and sequences*

Walk type	Fe = 30		Fe = 50		Fe = 100	
	Error	RMS	Error	RMS	Error	RMS
Standard	0.7%	5	1.2%	18	0.5%	17
Sneaky	0.3%	87	2.0%	185	2.1%	195
Catwalk	0.3%	25	0.8%	45	0.9%	65
Run	0.3%	31	2.8%	101	2.4%	121

In the last experiment, we exhaustively test our algorithm on the walking sequence for a full range of camera angle settings: $R_x \in [0, 360]$ and $R_y \in [0, 360]$, while we keep constant the focal length and position (f=50, T_x=0, T_y=0, T_z=5000). Since changes in R_z imply rotating the image plane around its perpendicular axis, which has no influence on the results, R_z was also kept constant. Figure 6 shows the accuracy of the estimated focal length for those settings. Apart from angles where the image and model planes are close to being parallel, the focal length can be estimated with an error below 4% for most settings.

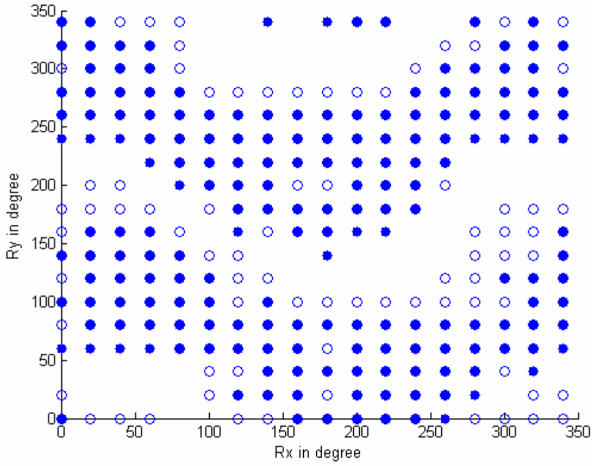

Figure 6: Estimation of focal length for a range of R_x and R_y. Dots represent estimated values within ±4% (filled dots represent estimated values within ±2%).

Focal lengths can be estimated accurately, only when the adapted planar models converge towards the expected model of the best coplanar frame in the sequence. This is illustrated on Figure 7. For each mid-stance frame of the standard walking sequence (see Figure 4), we display distances from the actual key point positions to the initial coarse model and then distances from the key points to the best adapted models. As expected, adapted models converge best at frame 44, when the key points are the most coplanar (see Figure 4).

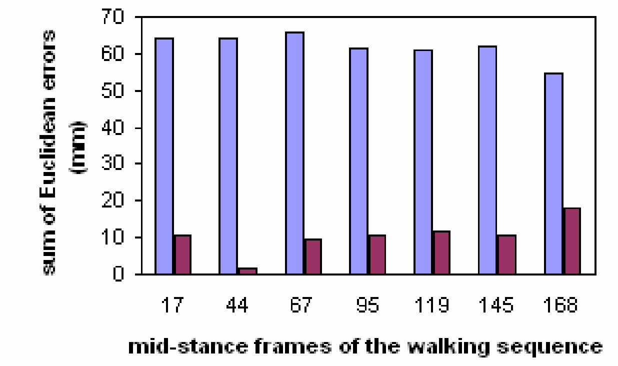

Figure 7: Model errors before (blue) and after model adaptation (magenta).

From experiment 1, we can confirm that standard deviation can be used as an indicator of focal length accuracy. Experiment 2 shows our method can cope with various types of human body motions without constraints of either periodicity or linearity. The last experiment shows our method can work with most of camera settings provided the image plane is not parallel to the model plane.

7. Conclusions - future work

We presented a method for auto-camera calibration which relies on the underlying biomechanical constraints associated to human bipedal locomotion. Our method was validated using a variety of human bipedal motions and camera configurations. Based on the "mid-stance" position where five joints of the human body (left/right shoulder, left/right hip and mid-hip) become coplanar, our technique was able to detect frames where the human body adopt that posture which ensures a successful camera calibration. Moreover since our method includes a 3D adaptation phase, a precise geometrical 3D description of that posture is not required.

We plan to use feature detectors and trackers to localise the joints on real video sequence to test further our method. We are optimistic that our method can deal with uncertainty coming with feature locations. Firstly, we will be able to select the five points from a variety of joints (i.e. shoulders, hips, head and neck) that can be coplanar. Secondly, our method can deal with noise, as we showed with the usage of a coarse 3D model.

Acknowledgements

The authors acknowledge financial support for the MEDUSA project from the Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/E001025/1.

References

[1] R.Y. Tsai, An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986

[2] Q. T. Luong and O. Faugeras, "Self-Calibration of a Moving Camera from Point correspondences and Fundamental Matrices," International Journal of Computer Vision 22(3), pp 261-289, 1997.

[3] M. Pollefeys, L. Van Gool "Stratified Self-Calibration with the Modulus Constraint", IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(8) pp707-724, 1999.

[4] M. Armstrong, A. Zisserman and R. Hartley, "Euclidean Reconstructing from Image Triplets", IEEE European Conferenceo on Computer Vision, Lecture Notes in Compute science, Vol 1064 pp 3-16, 1996

[5] J.R Renno, P. Remagnino, G. A. Jones. "Learning Surveillance Tracking Models fro the Self-Calibrated Ground Plane" in 'Acta Automatica Sinica', Special Issue on Visual Surveillance of Dynamic Sc 29(3) pp. 381-392, 2003

[6] F. Lv, T. Zhao, R. Nevatia, "Self-Calibration of a Camera from Video of a Walking Human", Proc. of International Conference on Pattern Recognition, 2002

[7] N. Krahnstoever and P. Mendonca, "Bayesian autocalibration for surveillance", Proc. IEEE International Conference on Computer Vision (ICCV05), Beijing, China, 2005

[8] N. Krahnstoever and P. Mendonca, "Autocalibration from Tracks of Walking People", British Machine Vision Conference, Edinburgh, UK, 2006.

[9] L. Lee, R. Romano, and G. Stein, "Monitoring activities from multiple video streams: Establishing a common coordinate frame," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 22, pp. 758–767, Aug. 2000.

[10] J. Black, T.J. Ellis, "Multi Camera Image Tracking", Proceedings of the Second International Workshop on Performance Evaluation of Tracking and Surveillance, December, Kauai, Hawaii, USA, 2001

[11] Chris Stauffer, Kinh Tieu: Automated multi-camera planar tracking correspondence modeling. IEEE Conference on Computer Vision and Pattern Recognition, pp.259-266, 2003

[12] D. Makris, T.J. Ellis, J. Black, "Bridging the Gaps between Cameras", IEEE Conference on Computer Vision and Pattern Recognition CVPR 2004, June, Washington DC, USA, pp. 205-210, 2004.

[13] Sven Carlsoo, How Man Moves, Kinesiological Methods and Studies., Heinemann, London , 1972.

[14] P. M. Galley and A. L. Forster, Human Movement-An Introductory Text for Physiotherapy Students. Churchill Livingstone 1987.

[15] L. Da Vinci, Description of "Vitruvian Man", 1492