CHARACTER RECONSTRUCTION AND ANIMATION FROM MONOCULAR SEQUENCES OF IMAGES

Fabio Remondino
Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: fabio@geod.baug.ethz.ch

Commission V, ICWG V/III

KEY WORDS: Calibration, Orientation, Matching, Reconstruction, Body Modeling, Animation

ABSTRACT
In this paper we present different methods for the calibration and orientation of monocular image sequences and the 3D reconstruction of human characters. Three different situations are considered: a static character imaged with a moving camera, a moving character imaged with a fix camera and a moving character imaged with a moving camera. A self-acquired sequence is used in the first case while in the other cases we used existing sequences available on the Internet or digitized from old videotapes. Most of the image-based techniques use probabilistic approaches to model a character from monocular sequences; on the other hand we use a determinist approach, recovering characterís model and movement through a camera model. The recovered human models can be used for visualization purposes, to generate new virtual scenes of the analyzed sequence or for gait analysis.

1. INTRODUCTION

The realistic modeling of human characters from video sequences is a challenging problem that has been investigated a lot in the last decade. Recently the demand of 3D human models is drastically increased for applications like movies, video games, ergonomic, e-commerce, virtual environments and medicine. In this short introduction we consider only the passive image- and triangulation-based reconstruction methods, neglecting those techniques that do not use correspondences (e.g. shape from shading) or computer animation software. A complete human model consists of the 3D shape and the movements of the body (Table 1): most of the available systems consider these two modeling procedures as separate even if they are very closed. A standard approach to capture the *static 3D shape* (and colour) of an entire human body uses laser scanner technology: it is quite expensive but it can generate a whole body model in ca 20 seconds. On the other hand, precise information related to character *movements* is generally acquired with motion capture techniques: they involve a network of cameras and prove an effective and successfully mean to replicate human movements. In between, single- or multi-stations videogrammetry offers an attractive alternative technique, requiring cheap sensors, allowing markerless tracking and providing, at the same time, for 3D shapes and movements information. Model-based approaches are very common, in particular with monocular video streams, while deterministic approaches are almost neglected, often due to the difficulties in recovering the camera parameters. The analysis of existing videos can moreover allow the generation of 3D models of characters who may be long dead or unavailable for common modeling techniques.

3D Shape	Active Sensors	Single-station Videogrammetry *Howe [2000] Sidenbladh [2000] Sminchisescu [02] Remondino [02, 03]*	Multi-stations Videogrammetry *Gavrila [1996] Yamamoto[98] Vedula [1999] DiApuzzo [03]*
Movements	Motion Capture		

Table 1: Techniques for human shape and movements modeling.

In this paper we present the analysis of monocular sequences with the aim of (1) generating reliable procedures to calibrate and orient image sequences without typical photogrammetric information and (2) reconstruct 3D models of characters for visualization and animation purposes. The virtual characters can be used in areas like film production, entertainment, fashion design and augment reality. Moreover the recovered 3D positions can also serve as basis for the analysis of human movements or medical studies.

2. RECOVERING CAMERA PARAMETERS APPROXIMATIONS FROM EXISTING SEQUENCES

As we want to recover metric information from video sequences (3D characters, scene models or human movement information), we need some metric information about the camera (interior and exterior parameters) and the images (pixel size). The approximations of these parameters are also necessary in the photo-triangulation procedure (bundle adjustment), as we must solve a non-linear problem, based on the collinearity fundamental condition, to obtain a rigorous solution. We assume that we do not know the parameters of the used camera and that we can always define 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 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. MODELING A STATIC CHARACTER WITH AN IMAGE SEQUENCE

For the complete reconstruction of a static human model, a full 360 degree azimuth image coverage is required. A single camera, with a fix focal length, is used. The image acquisition can last 30 seconds and this could be considered a limit of the process, as no movement of the person is required. The 3D shape reconstruction from the uncalibrated image sequence is obtained with a process based on (1) calibration and orientation of the images, (2) matching process on the human body surface and (3) 3D point cloud generation and modeling.

3.1 Camera calibration and image orientation

The calibration and orientation of the images have to be performed in order to extract precise 3D information of the character. The process (Figure 1) is based on a photogrammetric bundle adjustment; the required image correspondences (section 3.1.1) are found with an improved version of a process already presented in [Remondino, 2002].

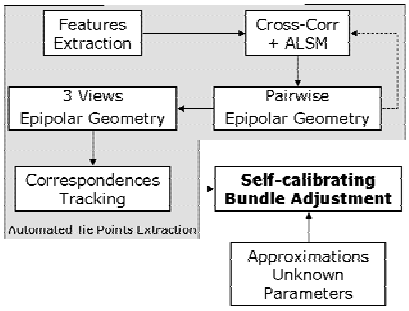

Figure 1: The camera calibration and image orientation pipeline.

3.1.1 Automatic tie points extraction

Most of the presented systems [e.g. Fitzgibbon et al., 1998; Pollefeys et al., 1998; Roth et al., 2000] developed for the orientation of image sequences with automatic extraction of corresponding points require very short baseline between the images (typically called ëshape-from-videoí). Few strategy instead can reliable deal with wide-baseline images [Tuytelaars et al., 2000; Remondino, 2002]. Our approach extracts automatically corresponding points with the following 6 steps:

1. *Interest points identification*. A set of interest points or corners in each image of the sequence is extracted using Foerstner operator or Harris corner detector with a threshold on the number of corners extracted based on the image size. A good point distribution is assured by subdividing the images in small patches (9x9 pixel on an image of 1200x1600) and keeping only the points with the highest interest value in those patches.
2. *Correspondences matching*. The extracted features between adjacent images are matched at first with cross-correlation and then the results are refined using adaptive least square matching (ALSM) [Gruen, 1985]. Cross-correlation alone cannot always guarantee the correct match while the ALSM, with template rotation and reshaping, provides for more accurate results. The point with biggest correlation coefficient is used as approximation for the template matching process. The process returns the best match in the second image for each interest point in the first image.
3. *Filtering false correspondences*. Because of the unguided matching process, the found matched pairs often contain outliers. Therefore a filtering of the incorrect matches is performed using the disparity gradient between the found correspondences. The smaller is the disparity gradient, the more the two correspondences are in agreement. The sum of all disparity gradients of each matched point relative to all other neighbourhood matches is computed. Those matches that have a disparity gradient sum greater than the median of the sums are removed. In case of big baselines or in presence (at the same time) of translation, rotation, shearing and scale between consecutive images, the algorithm can achieve incorrect results if applied on the whole image: therefore the filtering process has to be performed on small image regions.
4. *Epipolar geometry between image pairs*. A pairwise relative orientation and an outlier rejection using those matches that pass the filtering process are afterwards performed. Based on the coplanarity condition, the fundamental matrix is computed with the Least Median of the Squares (LMedS) method; LMedS estimators solve non-linear minimization problems and yield the smallest value for the median of the squared residuals computed for the data set. Therefore they are very robust in case of false matches or outliers due to false localisation. The computed epipolar geometry is then used to refine the matching process (step 3), which is now performed as guided matching along the epipolar line.
5. *Epipolar geometry between image triplets*. Not all the correspondences that support the pairwise relative orientation are necessarily correct. In fact a pair of correspondences can support the epipolar geometry by chance (e.g. a repeated pattern aligned with the epipolar line). These kinds of ambiguities and blunders are reduced considering the epipolar geometry between three consecutive images. A linear representation for the relative orientation of three frames is represented by the trifocal tensor T [Shashua, 1994]. T is represented by a set of three 3x3 matrices and is computed from at least 7 correspondences without knowledge of the motion or calibration of the cameras. In our process, the tensor is computed with a RANSAC algorithm [Fischler et al., 1981] using the correspondences that support two adjacent pair of images and their epipolar geometry. The RANSAC is a robust estimator, which fits a model (tensor T) to a data set (triplet of correspondences) starting from a minimal subset of the data. The found tensor T is used (1) to verify whether the image points are correct corresponding features between three views and (2) to compute the image coordinates of a point in a view, given the corresponding image positions in the other two images. This transfer is very useful when in one view are not found many correspondences. As result of this step, for each triplet of images, a set of corresponding points, supporting the related epipolar geometry is recovered.
6. *Tracking image correspondences through the sequence*. After the computation of a T tensor for every consecutive triplet of images, we consider all the overlapping tensors (e.g. T_{123}, T_{234}, T_{345}, ...) and we look for those correspondences which support consecutive tensors. That is, given two adjacent tensors T_{abc} and T_{bcd} with supporting points (x_a,y_a, x_b,y_b, x_c,y_c) and ($xí_b$,$yí_b$, $xí_c$,$yí_c$, $xí_d$,$yí_d$), if (x_b,y_b, x_c,y_c) in the first tensor T_{abc} is equal to ($xí_b$,$yí_b$, $xí_c$,$yí_c$) in the successive tensor T_{bcd}, this means that the point in images a, b, c and d is the same and therefore this point must have the same identifier. Each point is tracked as long as possible in the sequence and the obtained correspondences are used as tie points for the successive bundle-adjustment.

3.1.2 Photo-triangulation with bundle-adjustment

A photogrammetric self-calibrating bundle-adjustment is performed using the found image correspondences and the approximations of the unknown camera parameters (section 2). Because of the network geometry and the lack of accurate

control information, usually not all the additional parameters (APs) are recovered.

3.2 Matching process

In order to recover the 3D shape of the static human figure, a dense set of corresponding image points is extracted with an automated matching process [DíApuzzo, 2003]. The matching establishes correspondences between triplet of images starting from some seed points selected manually and distributed on the region of interest. The epipolar geometry, recovered in the orientation process is also used to improve the quality of the results. The central image is used as template and the other two (left and right) are used as search images (slaves). The matcher searches the corresponding points in the two slaves independently and at the end of the process, the data sets are merged to become triplets of matched points. The matching can fail if lacks of natural texture are presents (e.g. uniform colour); therefore the performance of the process is improved with Wallis filter to enhance the low frequencies of the images.

3.3 3D reconstruction and modeling

The 3D coordinates of the 2D matched points are afterwards computed with forward intersection using the results of the orientation process. A spatial filter is also applied to reduce the noise in the 3D data (possible outliers) and to get a more uniform density of the point cloud. If the matching process fails, some holes could be present in the generated point cloud: therefore a semi-automatic closure of the gaps is performed, using the curvature and density of the surrounding points. Moreover, if small movements of the person are occurred during the acquisition, the point cloud of each single triplet could appear misalign respect to the others. Therefore a 3D conformal transformation is applied: one triplet is taken as reference and all the others are transformed according to the reference one.
Concerning the modeling of the recovered unorganized 3D point cloud, we can (1) generate a polygonal surface with reverse-engineer packages or (2) fit a predefined 3D human model to our 3D data [DíApuzzo et al., 1999; Ramsis].

3.4 Results of the modeling of a static character

The presented example shows the modeling of a standing person with a digital still video camera Sony F505 (Figure 2).

Figure 2: Four (out of 12) images (1200x1600 pixels) used for the 3D static human body reconstruction.

The automatic tie points identification (section 3.1.1) found more than 150 correspondences that were imported in the bundle as well as four control points (measured manually on the body) used for the space resection process and the datum definition. At the end of the adjustment, a camera constant of 8.4 mm was estimated while the position of the principal point was kept fix in the middle of the images (and compensated with the exterior orientation parameters) as no significative camera roll diversity was present. Concerning the distortion parameters, only the first parameter of radial distortion (K1) turned out to be significant while the others were not estimated, as an over-parameterization could lead to a degradation of the results. The final exterior orientation of the images as well as the 3D coordinates of the tie points are shown in Figure 3.

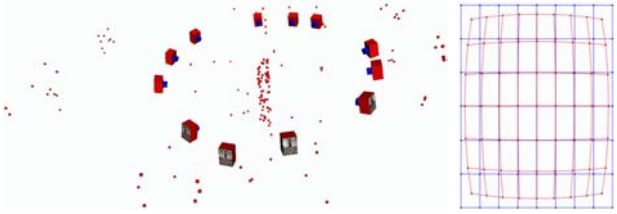

Figure 3: Recovered camera poses and 3D coordinates of the tie points (left). The influence of the APs on the image grid, 3 times amplified (right).

Afterwards, the matching process between 4 triplets of images produced ca 36 000 2D correspondences that have been converted and filtered in a point cloud of ca 34 000 points (Figure 4). The recovered 3D point cloud of the person is computed with a mean accuracy in x-y of 2.3 mm and in z direction of 3.3 mm. The 3D data can then easily be imported in commercial packages for modeling, visualization and animation purposes or e.g. used for diet management.

Figure 4: 3D point cloud of the human body imaged in Figure 2 pre and after the filtering (left). Visualization of the recovered point cloud with pixel intensity (right).

4. MODELING A MOVING CHARACTER WITH A FIX CAMERA

Nowadays it is very common to find image streams acquired with a fix camera, like in forensic surveillance, movies and sport events. Due to the complicate shape of the human body, a fix camera that images a moving character cannot correctly model the whole shape, unless we consider small part of the body (e.g. head, arm or torso). In particular, in the movies, we can often see a static camera filming a rotating head. Face modeling and animation has been investigated since 20 years in the graphic community. Due to the symmetric forms and geometric properties of the human head, the modeling requires very precise measurements. A part from laser scanner, most of the single-camera approaches are model-based (requiring fitting and minimization problems) while few methods recover the 3D shape through a camera model. Our solution tries to model the head regarding the camera as moving around it. Therefore it can be considered as a particular problem of the previous case. We have only to assume that the head is not deforming during the movement.
An example is presented in Figure 5. The image sequence, found on the Internet and with a resolution of 256x256 pixels, shows a person who is rotating the head. No camera and scene information is available and, for the processing, we consider the images as acquired by a moving camera around a fix head.

Figure 5: A fix camera imaging a rotating head (16 frames).

Due to the small region of interest (the head) and the very short baseline, the corresponding points for the image orientation are selected manually in the first frame: then they are matched automatically in all the other images using a tracking algorithm based on least squares template matching. For the datum definition and the space resection algorithm, 4 points extracted from a face laser scanner data set are used. Afterwards, with a bundle-adjustment we recovered the camera parameters: no additional parameters (APs) were used and only the focal length was computed. The recovered epipolar geometry is displayed in Figure 6.

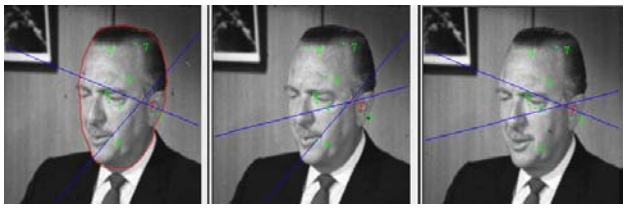

Figure 6: Recovered epipolar geometry between an image triplet.

Finally we applied the matching process described in section 3.2 on 3 triplets of images to get the 3D point cloud of the head. The results, with related pixel intensity, are shown in Figure 7.

Figure 7: 3D model of the moving head.

5. MODELING A MOVING CHARACTER WITH A MOVING CAMERA

A moving character imaged with a moving camera represents the most difficult case for the character reconstruction problem from image sequences. The camera can be (1) moved with a mechanical arm or on a small railroad or can be (2) stationary but freely rotating on a tripod or on the shoulder of a cameraman. In particular, sport videos are usually acquired with a ërotating systemí and often far away from the scene. In these cases, the baseline between the frames is very short and because of the movements of the character, a standard perspective approach cannot be used, in particular for the 3D modeling of the character. Nevertheless, we recover camera parameters and 3D object information through a camera model, without any model-based adjustment.

5.1 Image acquisition and orientation

A sequence of 60 images has been digitized from an old videotape, using a Matrox DigiSuite grabber, with a resolution of 720x576 pixels (Figure 8). For the orientation and reconstruction only 9 images are used (those images where the moving character has both feet on the ground). 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. A right-hand coordinate system with the origin in the left corner of 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 were performed manually.

Figure 8: Moving character filmed with a moving camera.

All the measurements are 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, using a block-invariant APs set. We allowed free rotations and very small translation of the camera, weighting the parameters and applying significance tests to analysis the determinability of the APs. The adjustment results ($\sigma_{0,post}$=1.3 pixel) show a focal length value of 22.7 mm and a pixel aspect ratio of 1.12. The non-unity of the aspect ratio can come from the old video camera or because of the frame grabber used in the digitization process. Concerning the lens distortion, only K1 (radial distortion) turned out to be significant while the others parameters could not be reliable determined. The principal point was kept fix in the middle of the images and compensated with the exterior orientation parameters. Figure 9 shows the global distortion effect on the image grid (3 times amplified) as well as the recovered camera positions.

5.2 3D reconstruction and modeling

For man-made objects (e.g. buildings), geometric constraints on the object (e.g. perpendicularity and orthogonality) can be used to solve the ill-posed problem of the 3D reconstruction from a monocular image. In case of free-form objects (e.g. the human body), we could use a probabilistic approach [Sidenbladh, 2000; Sminchisescu, 2002] or other assumptions must be provided [Remondino et al., 2003]:

1. the perspective collinearity model is simplified into a scaled orthographic projection;
2. the human body is represented in a skeleton form, with a series of joints and connected segments of known relative lengths;
3. further constraints on joints depth and segmentís perpendicularity are applied to obtained more accurate and reliable 3D models.

This reconstruction algorithm is applied to every frame of the sequence, given the image coordinates of some joints of the human body and the relative lengths of the skeleton segments. For each image, a 3D human model is generated but, because of the orthographic projection, the models are no more in the same reference system. Therefore a 3D conformal transformation is applied using, as common points, the 2 feet (which are always the ground) and the head of the character (known height). The object position of the feet is recovered with a 2D projective

invariance between two planes (basketball court and its image) that undergo a perspective projection [Semple et al., 1952]: the relationship between the two planes is specified if the coordinates of at least 4 corresponding points in each of the two projectively related planes are given. On the other hand, the object position of the head is computed as the middle point between the 2 feet. The invariance property and the conformal transformation are applied to each ëorthographicí model of the sequence and the obtained 3D coordinates are then refined using the camera parameters recovered in the orientation process. The final result is presented in Figure 9, together with the reconstructed scene.

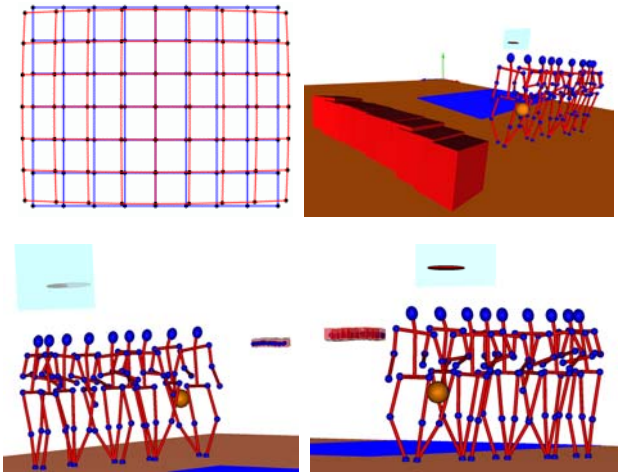

Figure 9: Influence of APs for the analyzed camera (upper left). The camera poses as well as the 3D reconstruction of the basketball court and the moving character (other images).

The recovered poses of the moving human can be used for gait analysis or for the animation of virtual characters in the movie production.

5.3 Other example

Another sequence, presented in Figure 10, is analyzed. 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, performed with a self-calibrating bundle adjustment with frame-invariant APs sets, recovered a constant increasing of the camera focal length and, again, a non-unity of the pixel aspect ratio ($1.10 \pm 4.5e\text{-}^{3}$).

Figure 10: Some frames of a video sequence of a basketball action. The camera is rotating and zooming.

Because of the low precision of the image measurements ($\sigma_{,priori}$ = 2 pixel) and the unfair network geometry, the principal point of the camera and the other terms used to model the lens distortion are not computed as very poorly determinable. The final standard deviation resulted 1.7 pixels while the RMS of image coordinates residuals are 38.45 μm in x direction and 29.08 μm in y direction.

The 3D reconstruction of the moving character is afterwards performed as described in section 5.2. In this case, the orthographic models of each frame could not be transformed into the camera reference system with a conformal transformation. Nevertheless the recovered 3D models are imported in Maya to animate the reconstructed character and generate new virtual scenes of the analyzed sequence (Figure 11).

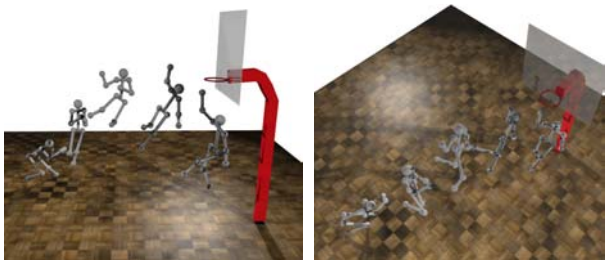

Figure 11: 3D models of the moving character visualized and animated with Maya.

To improve the visual quality and the realism of the reconstructed 3D human skeleton, we fitted a laser scanner human body model [Cyberware] to our data (Figure 12). The modeling and animation features of Maya software allow a semi-automatic fitting of the laser-data polygonal mesh to the skeleton model. The inverse kinematics method and a skinning process are respectively used to animate the model and bind the polygonal mesh with the skeleton [Learning Maya, 2003; Remondino et al., 2003].

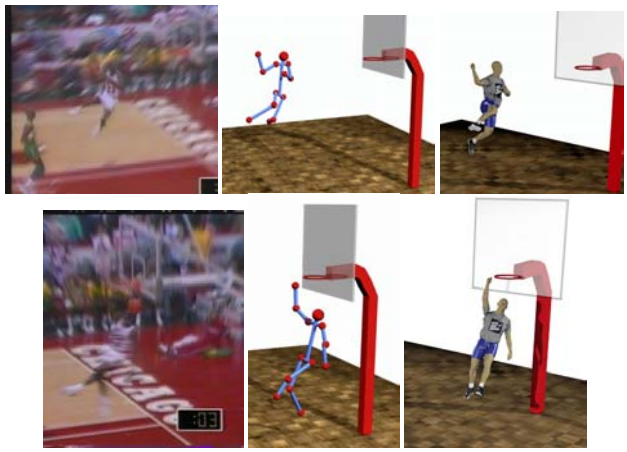

Figure 12: Two examples showing the results of the 3D reconstruction and the modeling process. Original frame of the sequence (left), reconstructed 3D human skeleton (middle) and fitting result, from a slightly different point of view (right).

6. CONCLUSION

The photogrammetric analysis of monocular video sequences and the generation of 3D human models were presented.

The image orientation and calibration was successfully achieved with a perspective bundle adjustment, weighting all the parameters and analysing their determinability with statistical tests. The human modeling, in particular from old videos, showed the capability of videogrammetry to provide for virtual characters useful for augmented reality applications, persons identification and to generate new scenes involving models of characters who are dead or unavailable for common modeling

systems. Finally, the markerless motion measurements and the recovered 3D positions of the human joints can also be used for gait analysis and sport medicine.

REFERENCES

Abdel-Aziz, Y., Karara, H., 1971: Direct linear transformation from comparator coordinates into object-space coordinates. Close range Photogrammetry, pp.1-18, ASPRS, Falls Church, Virginia.

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

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

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

Cyberware: http://www.cyberware.com [January 2004]

D'Apuzzo, N., Pl%akers, R., Fua, P., Gruen, A., Thalmann, D., 1999: Modeling human bodies from video sequences. In El-Hakim/Gruen (Eds.), Videometrics VI, Proc. of SPIE, Vol. 3461, pp. 36-47

DíApuzzo, N., 2003: Surface Measurement and Tracking of Human Body Parts from Multi Station Video Sequences Ph.D. Dissertation, ETH Zurich, Nr. 15271

Fischler, M, Bolles, R, 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. Assoc. Comp. Mach., Vol. 24(6), pp. 381-395

Fitzgibbon, A, Zisserman, A., 1998: Automatic 3D model acquisition and generation of new images from video sequences. Proceedings of European Signal Processing Conference, pp. 1261-1269

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

Gavrila, D.M., Davis, L., 1996: 3D model-based tracking of humans in action: a multi-view approach. IEEE CVPR Proc. pp. 73-80

Gruen, A., 1985: Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, Vol. 14(3), pp.175-187

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

Howe, N., Leventon, M., Freeman, W., 2000: Bayesian reconstruction of 3D human motion from single-camera video. Advances in Neural Information Processing System, Vol. 12, pp. 820-826, MIT Press

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

Learning Maya 5 - Foundation, 2003. Alias Wavefront http://www.aliaswavefront.com [January 2004]

Pollefeys, M., Koch, R., Van Gool, L., 1998: Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters, IEEE. ICCV Proc., pp.90-96

Ramsis: http://www.ramsis.de [January 2004]

Remondino, F., 2002: Image Sequence Analysis for Human Body Reconstruction. Int. Arch. of Photogrammetry and Remote Sensing, Vol. 34(5), pp. 590-595

Remondino, F., 2003: Recovering Metric Information from Old Monocular Video Sequences. In Gruen/Kahmen (Eds.). VI Conference on Optical 3D Measurement Techniques, Vol.2, pp 214-222

Remondino, F., Roditakis, A., 2003: Human Figures Reconstruction and Modeling from Single images or Monocular Video Sequences. 4th IEEE International 3DIM Conference

Roth, G., Whitehead, A., 2000: Using projective vision to find camera positions in an image sequence. 13th Vision Interface Conference

Semple, J.G., Kneebone, G.T., 1952: *Algebraic Projective Geometry*, Oxford Press

Shashua, A., 1994: Trilinearity in visual recognition by aligment. ECCV, Lectures Notes in Computer Science, Vol. 800, Springer-Verlag, pp.479-484

Sidenbladh, H., Black, M., Fleet, D., 2000: Stochastic Tracking of 3D Human Figures Using 2D Image Motion. European Conference on Computer Vision, D. Vernon (Ed.), Springer Verlag, LNCS 1843, pp. 702-718

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

Sminchisescu, C., 2002: Three Dimensional Human Modeling and Motion Reconstruction in Monocular Video Sequences Ph.D. Dissertation, INRIA Grenoble

Tuytelaars T., Van Gool, L., 2000: Wide baseline stereo matching based on local, affinely invariant regions, Proc. British Machine Vision Conference, Vol. 2, pp. 412-425

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

Yamamoto, M., Sato, A., Kawada, S., Kondo T., Osaki, Y., 1998: Incremental Tracking of Human Actions from Multiple Views. IEEE CVPR Proceedings

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

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