Vicon Motion Capture and HD 1080 Standard Video Data Fusion Based on Minimized Markers Reprojection Error

Chapter · August 2011

DOI: 10.1007/978-3-642-23154-4_24

4 authors:

Karol Jędrasiak
The University of Dabrowa Gornicza
64 PUBLICATIONS **400** CITATIONS
SEE PROFILE

Łukasz Janik
Silesian University of Technology
8 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Andrzej Polanski
Silesian University of Technology
134 PUBLICATIONS **1,109** CITATIONS
SEE PROFILE

Konrad W. Wojciechowski
Silesian University of Technology
114 PUBLICATIONS **739** CITATIONS
SEE PROFILE

Vicon Motion Capture and HD 1080 Standard Video Data Fusion Based on Minimized Markers Reprojection Error

Karol Jędrasiak[1], Łukasz Janik[1,2],
Andrzej Polański[1,2], and Konrad Wojciechowski[1,2]

[1] Polish-Japanese Institute of Information Technology, Aleja Legionów 2, 41-902 Bytom, Poland
{kjedrasiak,apolanski,kwojciechowski}@pjwstk.edu.pl
[2] Silesian University of Technology, Akademicka 16, 41-100 Gliwice, Poland
{apolanski,kwojciechowski,lukasz.janik}@polsl.pl

Summary. We present an algorithm for quantity motion capture and multi camera HD 1080 standard reference video data fusion. It consists of initial calibration step which is based on some set of selected frames and final fusion for the rest of frames. Implemented data fusion algorithm can be used in case that it is possible to find a time interval when both devices were recording the same sequence of poses. It is worth to emphasise there are no special calibration patterns used during calibration. Advantage of the algorithm is that the required calibration step can be perfomed simultaneously with actor calibration from Vicon Blade system. It is also allowed that cameras locations can be changed during acquisition process as long as they observe known motion capture markers. After calibration and synchronization reprojection is possible in real time for VGA resolution or in reduced frequency for HD 1080 standard. Performed experiments determined that average projection error is about 1.45 pixel in the Full-HD 1920×1080 reference video and it is perceptualy acceptable. Practical usage for training video depersonification was presented.

1 Introduction

Gait disorder is a common phenomenon in the society. Orthopedist diagnose patients based on gait analysis. Important first part of analysis is a visual observation of gait. It is performed by a medical doctor by watching walking tests or acquired HD 1080 reference videos of them. During observation the medical doctor can notice gait anomalies for more accurate check. In Human Motion Lab (HML) PJIIT it is performed a research of possible ways to use the motion capture system to assist medical doctors in diagnosing walking abnormalities. Motion capture system supply preciese quantity data characterizing motion. Therefore there is a need for reference video and motion capture data fusion. Known existing systems suffer from serious limitations

R.S. Choraś (Ed.): Image Processing & Communications Challenges 3, AISC 102, pp. 209–216.
springerlink.com

in this field. Concurrent taking into account quantity and quality data takes the form of multiple windows. Each window separately shows its data.

Design and implementation of synchronous fusion of quantity video and quality motion capture data right into video stream can be acknowledged as a novelty of practical importance. Such defined problem requires solving two following partial subproblems:

1. fusion of video and motion capture data for single frame,
2. data streams synchronisation.

First problem requires design of algorithm which incorporates data from video frame and motion capture system frame. It is important to stress that both data has to be from the same time instant t, as well as that both data sources have to be calibrated into common coordinate system. As such common coordinate system selected camera coordinate system is assumed. Second problem is desynchronisation that often occurs in mocup systems. Because of difference in acquisition frequencies time offset between motion capture data and reference video data is highly mutable overtime and often changes in each frame. Therefore it is required to detect common timeline and use synchronisation methods. In this paper we propose a flexible method of solving video and motion capture data fusion. Different calibration and synchronisation methods will be tested, compared and experimental results from multiple tests will presented.

2 Calibration

Camera calibration originates from photogrammetry [1]. Since that time several methods in the field of computer vision were presented [2, 3, 4, 5]. Aim of the calibration is to establish reprojection matrices and distortion parameters using correspondences between point coordinates in 3D and 2D. Researched method use motion capture 3D marker coordinates in the Vicon motion capture coordinate system (Fig. 1a) saved in C3D files and 2D coordinates in the image coordinates system (Fig. 1b) from the reference video file.

Algorithm of establishing common coordinates system between 3D mocup data and 2D reference video based on the pinhole camera model consists of the following steps. First, a transformation matrix between mocup scene coordinate system and selected camera coordinate system has to be established. Marker cloud coordinates are read from the motion capture industrial standard C3D files. The binary files consists of three sections: header, parameters and data. The data section stores marker coordinates placed one after another in a X, Y, Z, R_m order, where R_m is a 2 bytes residual value. The first byte indicates how many cameras observed the marker. The second byte stores the average of the residuals for the marker measurements. When residual value equals to -1 it is interpreted that the marker coordinates are invalid.

Transformation matrix M from mocup coordinate system to selected camera coordinate system consists of 3×3 rotation matrix R and 3×1 translation vector T.

Second step is a perspective projection of a point P_c in camera coordinates system into reference camera's image plane. It is required to estimate intrinsic camera parameters for this step such as effective focal length f, pixel scale factor s and a point (u_0, v_0) of origin of the image plane coordinate system.

Real camera lenses suffer from distortions. Distortions are divided into radial and tangential distortions. Pinhole camera model does not include such deviations. Therefore it is usually extended to model also radial and tangential distortions. Model combining radial and tangential distortion approximation can be expressed by the formula:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u' + \delta u^{(r)} + \delta u^{(t)} \\ v' + \delta v^{(r)} + \delta v^{(t)} \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \tag{1}$$

where:

$\begin{bmatrix} \delta u^{(r)} \\ \delta v^{(r)} \end{bmatrix} = \begin{bmatrix} u' \left(k_1 r^2 + k_2 r^4\right) \\ v' \left(k_1 r^2 + k_2 r^4\right) \end{bmatrix}$ is radial distortion,

$\begin{bmatrix} \delta u^{(t)} \\ \delta v^{(t)} \end{bmatrix} = \begin{bmatrix} 2p_1 u' v' + p_2 \left(r^2 + 2u'^2\right) \\ v' \left(k_1 r^2 + k_2 r^4\right) \end{bmatrix}$ is tangential distortion,

u', v' - ideal image plane coordinates,

u, v- distorted image coordinates,

u_0,v_0- coordinates of the image center,

Scene, camera and image coordinate systems used during reprojection are presented in outlook Fig. 1. Distortions correction step was omitted for simplification of illustration.

During experiments it has been evaluated different methods of determining transformation matrix and their impact on the final reprojection result. Tsai [2] algorithm was used as an example of using separate matrices and Direct Linear Transformation (DLT) [5] for aggregated.

One of the most popular calibration methods is DLT. Method was revised in large number of publications [6, 7]. Standard pinhole camera model is used as a camera model. Algorithm uses N known points correspondences to compute the transformation matrix M.

The main disadvantage is that distortions are not taken into account during processing. Estimating distortions is an important part of calibration process. One of the most popular method for nonlinear camera calibration was designed by Tsai. Solution is able to approximate intrinsic, extrinsic parameters and distortions coefficients in the form of (2). As a starting point for nonlinear optimization the parameters acquired using DLT. Such optimization is usually done using modified Levenberg-Marquardt method [8].

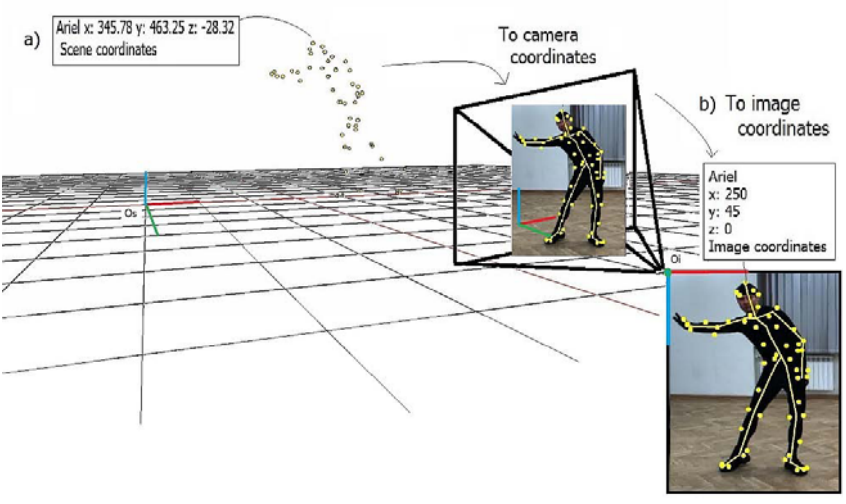


Fig. 1. Outlook illustration of 3D coordinates reprojection onto 2D image of video camera. a) 3D coordinate marker set in the mocup coordinate system, b) result of projection onto 2D image.

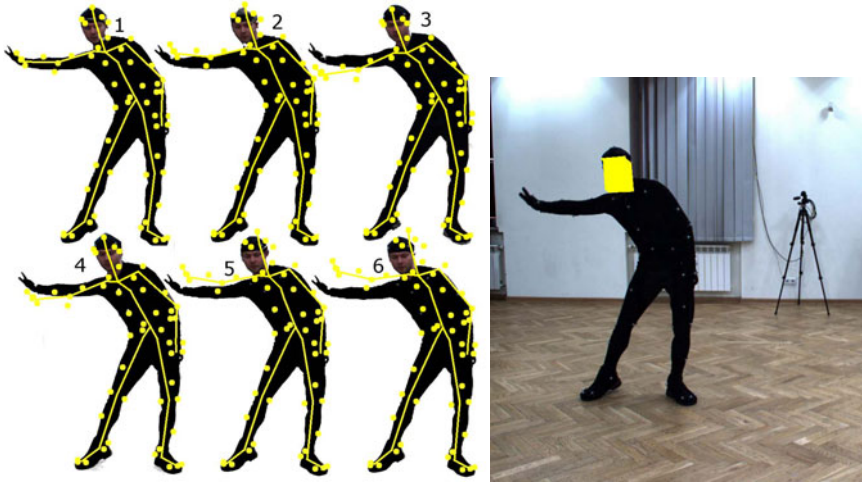
3 Synchronisation Problem

Mocup sequence data can be delayed or surpassed to video multiple times as presented in Fig 2a.

In order to build robust synchronous system for mocup and reference video recording multiple methods may be used. One of the most efficient is to run hardware synchronization, where cameras exposure, readout and transfer to computer are controlled by triggering signal and signal is triggered continuously for each frame. System provides triggering signal to each device engaged to recording.

4 Experimental Results

All results were acquired using prototype application for data fusion. The software allows to select points for calibration, calibrate and reproject 3D coordinates from input C3D files onto reference videos. Calibration points were selected with single pixel precision as the center of the calibration marker visible in the video sequence. For technique correcteness demonstration purpose only we have inserted into videos simplified skeleton. It was created by combining selected Vicon markers into segments. We acquired 65 test sequences. Test sequences were divided into sequences testing computing transform matrices, sequences for testing distortions effects and sequences for testing synchronization. In HML measurement volume is observed by 4 reference video cameras. To shorten the results section we present results only from single



(a) (b)

Fig. 2. (a) Features arm side movement of joint mocup and video data, where mocup and video data are shifted to each other of random time offset, (b) Example of successful video depersonification using motion capture and Full-HD reference video fusion.

camera. We account for the simplifying fact that process of calibrating other cameras was analogical.

Calibration test sequences allowed us to compute the following parameters. For DLT the matrix M:

$$M = \begin{bmatrix} -958.37 & 1086.96 & -127.68 & 3.0773e+6 \\ -396.29 & -1.52 & -1173.62 & 2.9751e+6 \\ -0.99 & -0.0087 & -0.14 & 3.5735e+3 \end{bmatrix} \tag{2}$$

Tsai algorithm computed the following intrinsic, extrinsic and distortion coefficients:

$$M_R = \begin{bmatrix} 0.0121 & 0.999 & 0.0046 & 381.1 \\ 0.1327 & 0.0030 & -0.9911 & 916.8 \\ -0.9911 & 0.0126 & -0.1327 & 3461.9 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Radial distortions were computed as $\delta^{(r)2} = -0.1577$, $\delta^{(r)4} = 0.08576$. Tangential distortions and effective focal length f were computed as $\delta_1^{(t)} = -6.8e-4$,$\delta_2^{(t)} = 8.0e-5$, $f = \begin{bmatrix} 1.111606e+3 & 1.111387e+3 \end{bmatrix}$.

$c_c = \begin{bmatrix} 978.7653 & 548.3601 \end{bmatrix}$.

Coordinates of 3D points reprojection using DLT and Tsai matrices were computed using data from 997 frame. The Euclidean distance between ground truth points and their reprojections is labeled as reprojection error.

Tsai algorithm proved to be efficient enough for real life use. High values of error DLT method makes it good only as a starting point for further nonlinear optimization techniques. First column named ID displays numbers of markers as in Vicon Blade marker set. GT stands for ground truth selected manually with single pixel precision. X or Y after name means adequately first or second point component. Acquired results were presented for upper (RFHD-(right front head), ... , RFSH-(right front shoulder)) and medium (CLAV-(clavicle), ... , LFWT-(left front waist)) human body segments(Table 1).

Table 1. Collation of reprojection coordinates and Euclidean errors. Results were acquired using frame 997. Results show only upper and medium body marker values.

ID	Name	GT X	GT Y	DLT X	DLT Y	DLT Error	Tsai X	Tsai Y	Tsai Error
UP.B.									
1	ARIEL	1034	258	1033.52	264.54	6.56	1035.63	259.29	2.08
4	RFHD	1010	274	1009.19	279.9	5.96	1010.65	274.32	0.72
2	LFHD	1054	275	1052.1	280.87	6.17	1054.94	275.66	1.15
20	RFSH	987	360	987.3	366.24	6.25	988.08	361.6	1.93
10	LFSH	1074	361	1070.96	366.95	6.68	1074.74	362.45	1.63
AVG						6.32			1.50
M.B.									
8	CLAV	1032	383	1030.49	388.25	5.46	1033.03	384.19	1.57
9	STRN	1029	433	1027.57	437.05	4.3	1030.16	434.07	1.58
33	RFWT	986	499	985.66	501.81	2.83	986.54	500.62	1.71
31	LMWT	1028	493	1026.58	495.37	2.76	1029.23	494.03	1.6
30	LFWT	1066	506	1062.84	507.91	3.69	1066.91	507.01	1.36
AVG						3.81			1.56

Euclidean error values for DLT are about 6 pixels for Full-HD resolution. This value is not suitable for medical application but can be accepted for most standard visual systems where only approximation of location is enough. Tsai algorithm result is more than 4.21 times better with error of value 1.50 pixel. Markers were placed on the front side of the head, chest and loins. It can be seen that DLT is in the range 4 - 6 pixels. It is in unison with pressumption that distortions are the smallest in the center of the image.

To test the impact factor of the distortion in the image we repeat the experiment for selected frames 1, 308, 690, 809 and 911. To reduce the ammount of space we present in Table 2 only the average values of the collation tables acquired. It can be seen that distortions significantly arose and made impractical DLT algorithm. Tsai algorithm proved to be reliable even in face

of heavy radial and tangential distortions. DLT and Tsai algorithms are comparable only if person is standing in the middle of the image where distortions are the smallest. It can be seen that distortions in the right side of the image are stronger than in the left side.

Table 2. Comparison of only average errors values for chosen frames. It can be seen that Tsai algorithm error values are stable and in range of 1-2 pixels for Full-HD reference videos.

Name	AVG DLT Err	AVG TSAI Err
Frame 1	8.99	1.38
Frame 308	17.45	1.89
Frame 690	9.06	1.31
Frame 809	11.09	1.01
Frame 911	7.56	1.52
Frame 997	4.73	1.58
AVG	9.81	1.45

One of possible practical implementations of the described data fusion method is simple video depersonification. Known locations of markers placed on the head of the subject allow to determine head orientation. It is further used to assign cuboid's vertices which covers actor face in the world coordinates. Those points reprojections completely hide actor's head which is useful for various training videos (Fig. 2b). Head markers used are LFHD, RFHD, LBHD and RBHD. We assign the surface determined by their coordinates as upper base of the cuboid. The figure's height is the length of the vector LFHD-RFHD multiplicated by a factor. Factor value which covers the whole head of our actors was experimentally measured as 1.4. Vector pointing in the floor direction is computed as a cross product of the upper surface's vectors. Future implementations will use human anatomical data to determine characteristic head features for more advanced body parts cloaking.

5 Conclusions

The purpose of the work was to design and implement an algorithm for quantity motion capture and quality multi camera HD 1080 reference video data fusion. The aim of the work was achived. Performed experiments of the suggested solution determined that average projection error is about 1.45 pixel in the Full-HD 1920x1080 reference video and it is perceptualy acceptable. It is required to take into account camera lenses distortion factors during reprojection only when reprojection takes place near the borders of the image where the distortion effect are accumulated. Fusion time sequences show

that after short-term desynchronisation period system is able to recount synchronisation factor using time stamps. Advantage of the algorithm is that the required calibration step can be perfomed simultaneously with actor calibration from Vicon system. After calibration step reprojection is possible in real time for VGA resolution or in reduced frequency for HD 1080 standard. Implemented data fusion algorithm can be used in case that it is possible to find a time interval when both devices were recording the same action. It is worth to emphasise there are no special calibration patterns used during calibration. It is also allowed that cameras locations can be changed during acquisition process as long as they observe knowm motion capture markers. Performing data fusion of video stream with kinematic skeleton acquired from motion capture data is considered future use as well as possibility of more precisely configuring quantity data.

Acknowledgement

This paper has been supported by the project „System with a library of modules for advanced analysis and an interactive synthesis of human motion" co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1. Research and development of modern technologies, measure 1.3.1 Development projects.

References

1. Brown, D.C.: Close-range camera calibration. Photogrammetric Engineering 37(8), 855–866 (1971)
2. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. IEEE Journal of Robotics and Automation 3(4), 323–344 (1987)
3. Weng, J., Cohen, P., Herniou, M.: Camera calibration with distortion models and accuracy evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(10), 965–980 (1992)
4. Zhang, Z.: A Flexible New Technique for Camera Calibration. Technical Report MSRTR- 98-71, Microsoft Research, December
5. Faugeras, O.D., Toscani, G.: Camera calibration for 3D computer vision. In: Proc. International Workshop on Industrial Applications of Machine Vision and Machine Intelligence, Silken, Japan, pp. 240–247 (1987)
6. Melen, T.: Geometrical modelling and calibration of video cameras for underwater navigation. PhD thesis, Norges tekniske hogskole, Institutt for teknisk kybernetikk (1994)
7. Shih, S.W., Hung, Y.P., Lin, W.S.: Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem. Optical Engineering 32(1), 138–149 (1993)
8. More, J.: The Levenberg-Marquardt algorithm, implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis. Lecture Notes in Mathematics, vol. 630. Springer, Heidelberg (1977)
9. Webpage of PJWSTK Human Motion Group, `http://hm.pjwstk.edu.pl`
10. Webpage of the 3D Biomechanics Data Standard, `http://www.c3d.org`