Proc. of the XVIIIth Congress of the Int. Society of Biomechanics,
Zurich, Switzerland, 2001, CDROM paper #0106

Motion Capture from Multi Image Video Sequences

Nicola D'Apuzzo
Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, Switzerland
nicola@geod.baug.ethz.ch

Introduction

In this paper we present a method to extract 3-D motion information of the human body from multi image video sequences acquired with three synchronized CCD cameras without using markers. This work is part of a project aimed at developing a highly automated system to model most realistically human bodies from video sequences.

Methods

The process is composed of 5 steps: (1) acquisition of video sequences, (2) calibration of the system, (3) surface measurement of the human body for each frame, (4) 3-D surface tracking and filtering, (5) 3-D joints tracking. Photogrammetry and image processing techniques are used for these purposes.
Our image acquisition system is currently composed of three synchronized CCD cameras and a frame grabber which acquires a sequence of triplet images. Before stepping into the processing phase, the camera system has to be calibrated. Self calibration methods (Maas, 1998) are applied to gain exterior orientation of the cameras, the parameters of internal orientation and the parameters modeling the lens distortion (Brown, 1971).
From the video sequences, we extract two kinds of 3-D information: a three dimensional surface measurement of the visible parts of the body for each triplet and 3-D trajectories of points on the body. Our approach for surface measurement is based on multi-image matching, using the adaptive least squares method (Gruen, 1985). A semi automated matching process (D'Apuzzo, 1998) determines a dense set of corresponding points in the triplets, starting from few manually selected seed points. In case of poor natural texture, local contrast enhancement of the images is required for the least squares matching. The 3-D coordinates of the matched points are then computed by forward ray intersection using the orientation and calibration data of the cameras. A filter is applied to reduce the remaining noise in the 3-D data and get a more uniform density of the point cloud. Figure 1 shows an example of the surface measurement process using a pair of images.

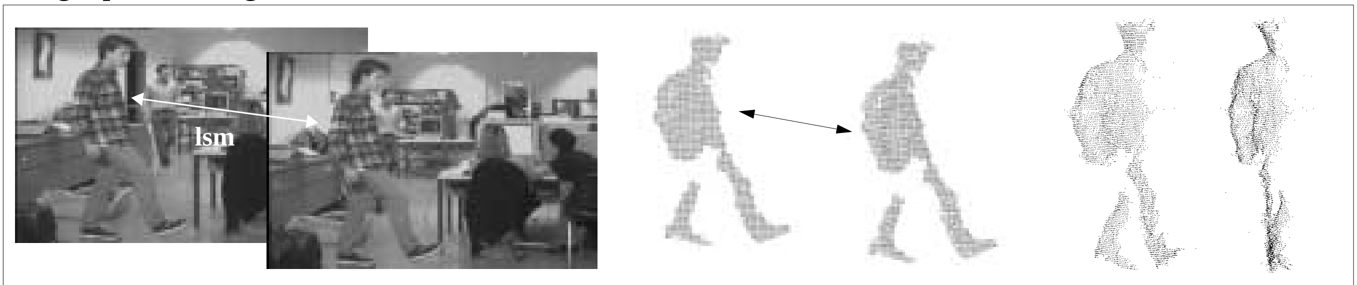

Figure 1: Image pair (left), matched points (centre) and 3-D point cloud

The tracking process is also based on least squares matching techniques. Its basic idea is to track triplets of corresponding points in the three images through the sequence and compute their 3-D trajectories. The spatial correspondences between the three images at the same time and the temporal correspondences between subsequent frames are determined with a least squares matching algorithm. The results of the tracking process are the coordinates of a point in the three images through the sequence, thus the 3-D trajectory is determined by computing the 3-D coordinates of the point at each time step by forward ray intersection. Velocities and accelerations are also computed. Figure 2 shows how the correspondences are determined by the least squares matching tracking algorithm and an example of computed 3-D trajectories.

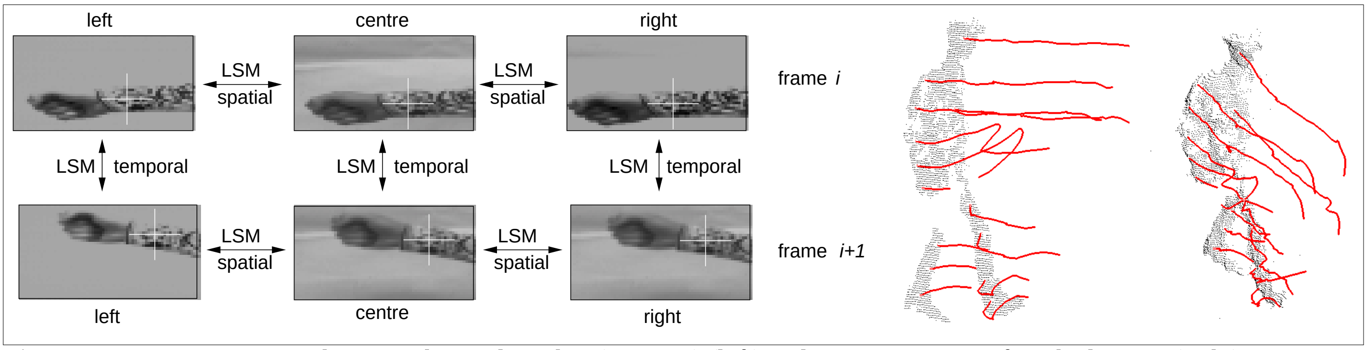

Figure 2: Least Squares Matching Tracking Algorithm (LSMTA) (left) and 3-D trajectories of tracked points (right)

The advantage of this tracking process is twofold: it can track natural points, without using markers; and it can track local surfaces on the human body. In the last case, the tracking process is applied to all the points matched in the region of interest. The result can be seen as a vector field of trajectories (position, velocity and acceleration). This way of tracking may produce false trajectories, which can be removed checking for consistency and local uniformity of the movement (figure 3).

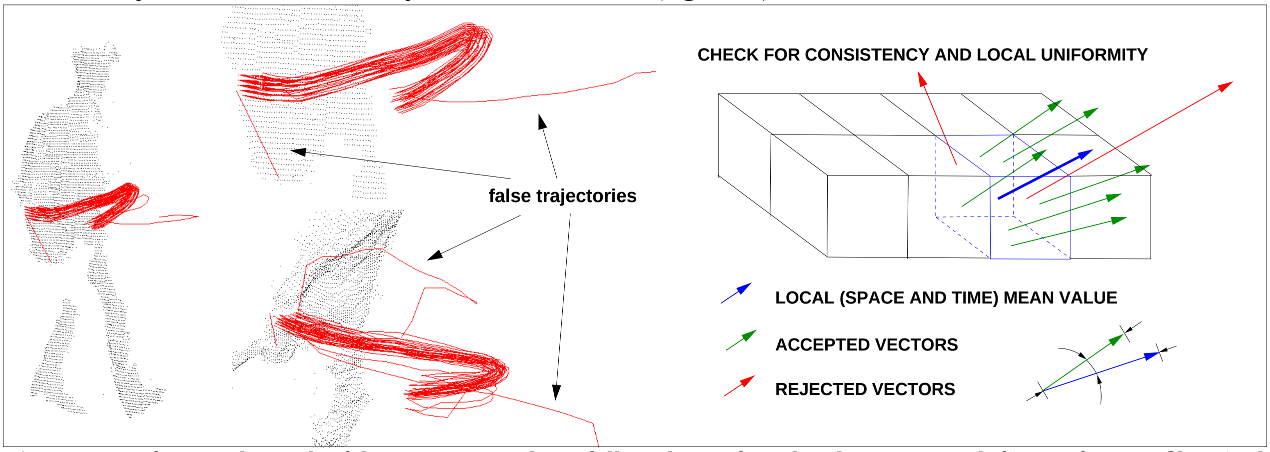

Figure 3: Surface tracking: the false trajectories don't follow the uniform local movement (left). Uniformity filter (right)

The last step of the process is the definition of joints or key points of the human body. Joints are groups of trajectories: 3-D regions whose size can vary. They are interactively selected in a graphical enhanced user interface, where they can be easily placed and moved in 3-D space. The joints are tracked in a simple way: the position at the next time step is established by the mean value of the displacement of all the trajectories inside that region. As we are interested in the capture of the movement of the human body, we have therefore selected key joints which define the motion. Depending on the complexity of the movement we want to capture, a minimal number of key joints have to be defined. For this test sequence, we have chosen a small set of joints: feet, knees, hips, hands, elbows, shoulders, neck, head, bust. Figure 4 shows the selected joints. Since we use only 3 CCD cameras which acquires video sequences frontally, we cannot track the complete motion performed by the person. Other cameras would be required to get informations at the back side.

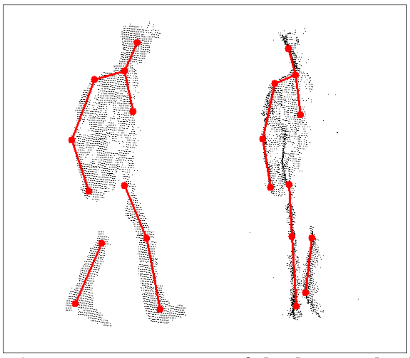

Figure 4: Key joints of the human body

Results & Discussion

The tracked key joints lead to a final result comparable to the conventional motion capture systems: 3-D trajectories of key joints which can be afterwords analyzed and used for animation or medical purposes. An example of the results of the tracking process is shown in figure 5. As it can be seen the results are satisfactory.

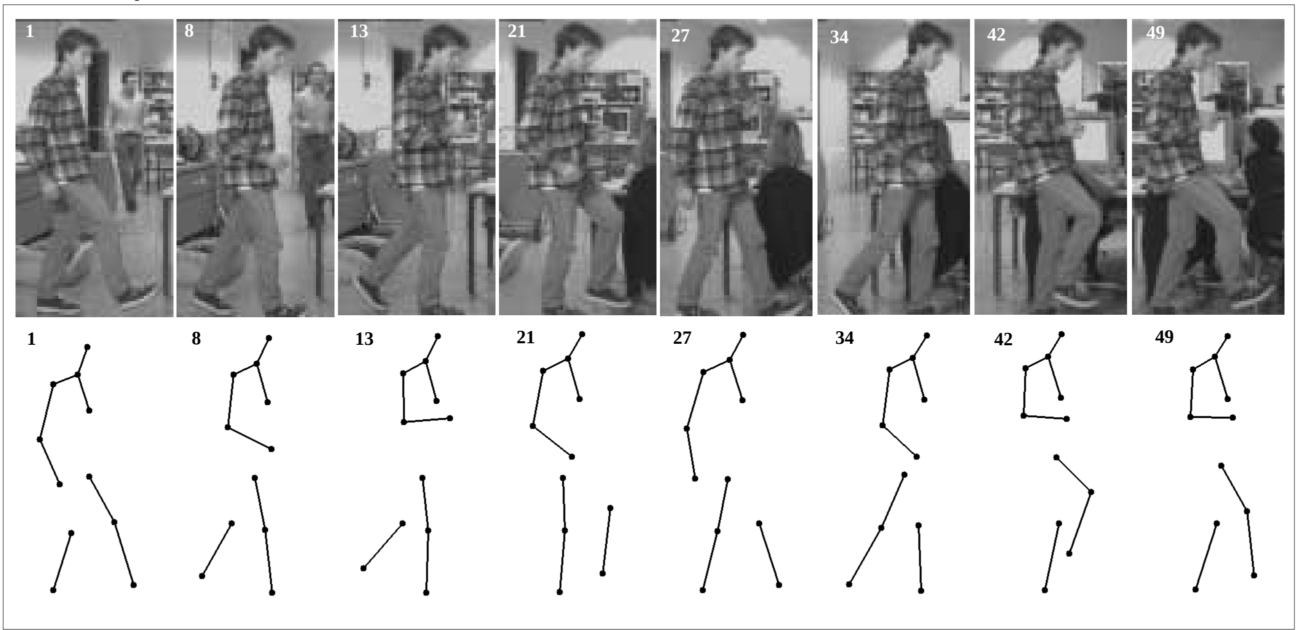

Figure 5: Key joints tracking results. Top: original image sequence, bottom: tracked 3-D key joints

This method is still under development and a lot of work remains for the future in order to improve the quality and accuracy of the extracted 3-D data. In addition, the gain in robustness and level of automation should be considered too, since the final goal of the project is the development of a fully automated and robust process.

References

Brown D.C. Close-Range Camera Calibration, *Photogrammetric Engineering and Remote Sensing*, **37(8)**, 855-866, 1971

D'Apuzzo N. Automated photogrammetric measurement of human faces, *Int. Archives of Photogrammetry and Remote Sensing*, **32(B5)**, Hakodate, Japan, 402-407, 1998

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

Maas H.-G. Image sequence based automatic multi-camera system calibration techniques. *Int. Archives of Photogrammetry and Remote Sensing*, **32(B5)**, Hakodate, Japan, 763-768, 1998

Acknowledgements

The work reported here was funded in part by the Swiss National Science Foundation.