A Model-based Image-matching Technique for 3D Motion Reconstruction from Uncalibrated 2D Video Sequences

Krosshaug T & Bahr R
Oslo Sports Trauma Research Center, Department of Sports Medicine
Norwegian School of Sport Sciences & Oslo Orthopaedic University Clinic, Oslo, Norway

Introduction

Biomechanical analyses of injury mechanisms are essential for understanding how to prevent injuries. As these situations cannot be reconstructed in the laboratory, particularly interest lies in utilizing video data from actual injuries optimally. At present the existing methods for analyzing injury situations from videotape are unsatisfactory (Fischer 1994). Even though visual inspection has been the common way to analyze video recorded injury situations, other possibilities exist. Methods for tracking and reconstruction of 3D human motion from uncalibrated video film, from one or more views, are common within the field of robot vision, the computer gaming industry, virtual reality, etc. (Aggarwal & Cai 1999). However, the main emphasis for these studies has been more on the ability to perform automatic computerized reconstruction rather than to achieve the most accurate kinematic estimates. It does not appear to be possible to combine automatic reconstruction and high precision. Thus, we wanted to develop and test a manual reconstruction method.

Aims

The aim of this study was to assess the accuracy of a new model-based image-matching technique for human motion reconstruction from one or more uncalibrated video sequences, using traditional motion analysis as a gold standard.

Methods

The model-based image-matching technique is based on the commercially available three-dimensional modeling software Poser® 4 and the Poser® Pro Pack (Curious Labs, Inc., Santa Cruz, California, USA). We used a skeleton model with 21 segments and 57 degrees of freedom, customized to match the anthropometry of the person in the video. The surroundings were modeled using straight lines, curved lines and boxes (Figure 1) The frame-by-frame matching of the surroundings enabled us to reconstruct camera motion for video footages shot from translating, rotating and zooming cameras. To minimize bias resulting from single-operator judgment, an expert panel of experienced clinicians gave their opinion on the goodness of the fit. When a sequence was satisfactory, the translation and joint angle time histories were read into Matlab with a customized script for further processing.

To test the validity of the method, a laboratory trial was conducted with one test subject performing jogging and side step cutting, while being filmed with three ordinary video cameras. The videotapes were digitized, enhanced, and de-interlaced to achieve an effective frame rate of 50Hz. In total, this provided three single camera matchings, three double camera matchings and one triple camera matching for each of the motions. The test subject wore 33 reflective skin markers and was filmed with a seven-camera, 240Hz motion analysis system (ProReflex, Qualisys Inc.) representing the gold standard. Two AMTI force platforms (AMTI LG6-4-1, Watertown, MA 02472, USA) measured ground reaction forces. Body segment parameters (BSP) for Center Of Mass (COM) were estimated using a slightly modified version of Yeadon's stadium solid method (Yeadon, 1990).

Results

Good agreement was found for the support leg flexion/extension angles in the hip and knee for all the matchings compared to the ProReflex measurements, with root mean square (RMS) differences ranging from 3 to 12°. Hip ad-/abduction RMS differences ranged from 12° to 14°, while varus/valgus angles of the knee were in the range of 3° to 5° (Figure 2). Rotation angles were clearly most variable in both the hip and knee, and RMS differences ranged from 6° to 16°. RMS velocity differences of the center of mass in all three directions ranged from 0.1 to 0.6 m/s (Table 1). Accelerations were only acceptable for the matchings that contained perpendicular views, with the triple camera matching as the best (RMS differences of 2.8-4.9 m/s^2). However, due to low frame rate (50 Hz for PAL videos), the high frequency acceleration peaks were not captured.

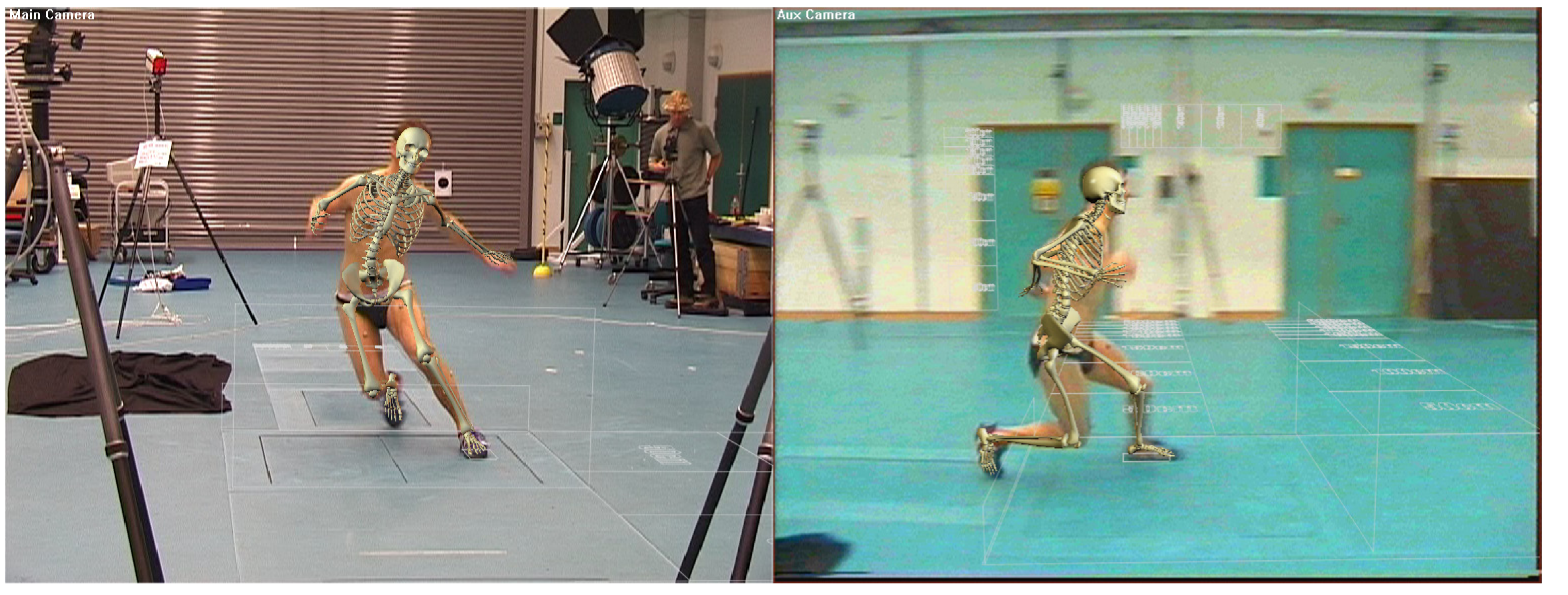

Figure 1. The images show the Poser model superimposed onto the original videotapes. Notice how the white "calibration"-lines on the floor matches patterns on the real floor.

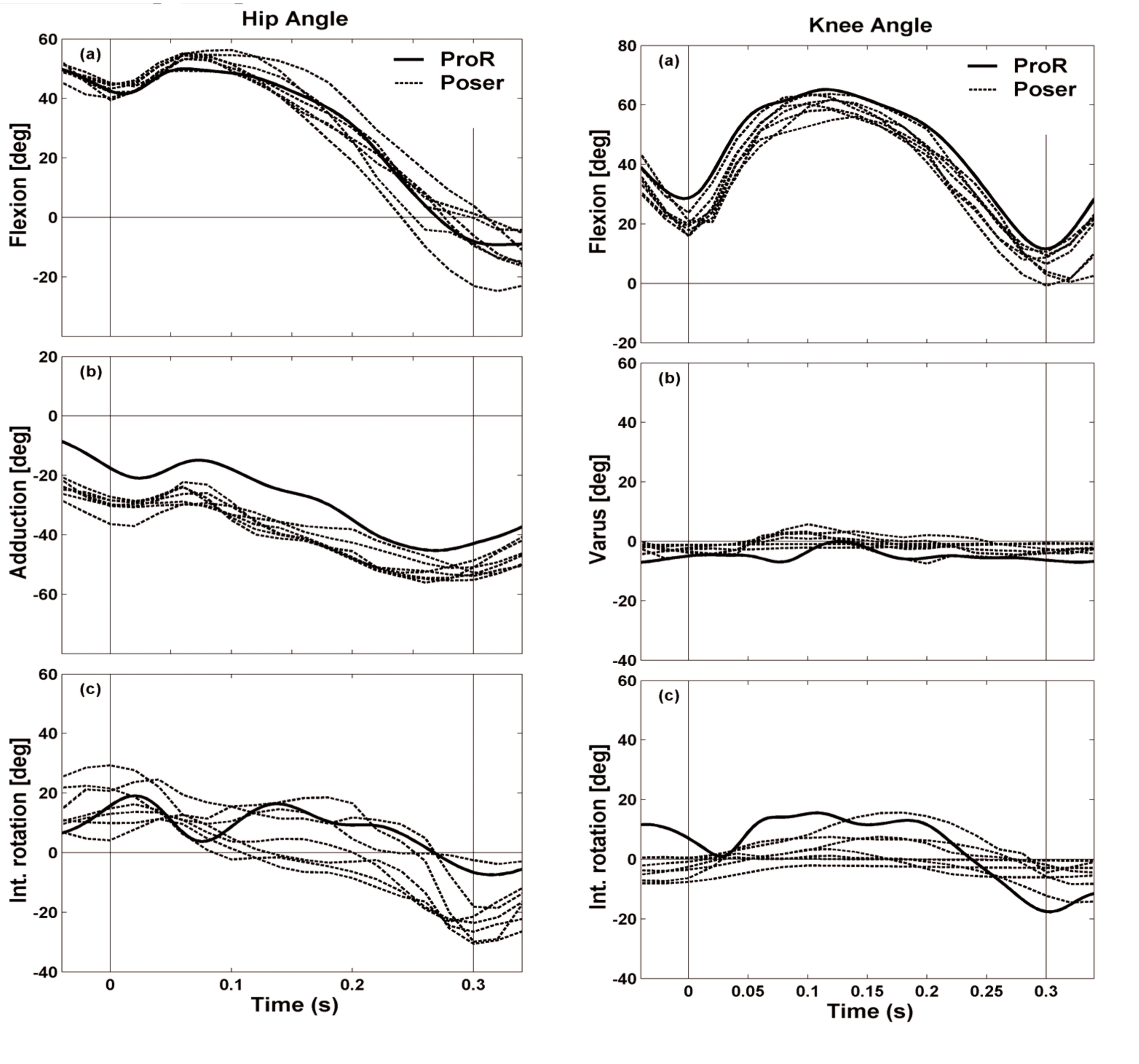

Figure 2. Hip and knee joint angles (°) of the support leg, calculated with the Pro Reflex system (solid lines) and the model-matching technique for each of the seven matchings (dotted lines). The vertical lines indicate initial ground contact of the support leg and toe-off.

Discussion

Analyses of the individual segments showed that error in the hip joint angles mainly originated from the pelvis matching. This is not surprising, since the shape of the pelvis makes interpretation of its attitude difficult. The matching of the shank was, on the other hand, very good, which implies the knee joint center will probably be estimated more accurately than the hip joint center. The best COM velocity estimates were obtained with the side/front and rear/side/front cameras. This is probably due to the fact that the nearly perpendicular camera positions provided complimentary information, as well as the relatively higher resolution of the test subject in the front and side views.

Although the acceleration estimate from the triple camera matching was reasonably good, we were not able to capture the high-frequency dynamics of the impact. We found, by down-sampling the original Pro Reflex recording from 240 to 50 Hz, that the acceleration error increased considerably, resulting in an estimate very similar to what was produced with the model-matching technique. The initial vertical impact force peak from the heel strike was no longer detectable and the estimated maximal force peak was seen about 60 ms after the true peak, indicating that the frame rate rather than the method itself is the main cause of errors in acceleration estimates.

Table 1. Root mean square and maximal difference in velocity (m/s) for the Center of Mass (COM) between the Pro Reflex recordings and the estimates from our model-based matching-technique for each of the seven matchings of the plant and cut motion. The maximal differences are shown in parentheses.

		Antero-posterior vel.	Medio-lateral vel.	Vertical vel.
Camera views	Rear	0.38 (0.70)	0.34 (0.69)	0.10 (0.22)
	Side	0.17 (0.39)	0.62 (1.06)	0.19 (0.44)
	Front	0.31 (0.78)	0.15 (0.26)	0.14 (0.33)
	Rear/side	0.10 (0.18)	0.19 (0.32)	0.13 (0.36)
	Rear/front	0.31 (0.78)	0.12 (0.22)	0.15 (0.33)
	Side/front	0.11 (0.28)	0.10 (0.22)	0.14 (0.36)
	Rear/side/front	0.09 (0.19)	0.11 (0.22)	0.13 (0.39)

Conclusions

A new model-based image matching technique has been developed capable of producing good kinematic estimates from uncalibrated video recordings, provided the video quality is good and at least two camera views are available. This method can potentially be used to arrive at more precise descriptions of the mechanisms of sports injuries, e.g. for knee injuries.

References

1. Fischer, J. F., Leyvraz, P. F. and Bally, A.
A dynamic analysis of knee ligament injuries in alpine skiing
Acta Orthopedica Belgica, 1994: 60(2): 194-203.
2. Aggarwal, J. K. & Cai, Q. Human motion analysis: a review.
Computer Vision and Image Understanding 1999. 73: 428-440.
3. Yeadon,M.R. (1990). The simulation of aerial movement--II. A mathematical inertia model of the human body. J.Biomech. 23, 67-74.

Acknowledgements: The Oslo Sports Trauma Research Center has as been established through generous grants from the Eastern Norway Regional Health Authority, the Royal Norwegian Ministry of Culture, the Norwegian Olympic Committee & Confederation of Sport, Norsk Tipping AS, and Pfizer AS

Oslo Sports Trauma RESEARCH CENTER