Hand pose reconstruction using a low-cost three-camera stereo vision system

Eugen Richter, Norman Hendrich, Jianwei Zhang
Department of Informatics, Group TAMS, University of Hamburg
{erichter, hendrich, zhang}@informatik.uni-hamburg.de

Introduction

Human hand pose sequences, i.e. series of reconstructed hand model configurations, that represent articulated hand motion can be used for many purposes. Such sequences can be used as data sets for learning processes, grasp or manipulation task analysis as well as for remote operation or human-computer interaction purposes. Due to the high amount of degrees of freedom and the inherent complexity of extraction of a hand's internal articulated structure, the task of model reconstruction of an articulated hand based on visual information is a rather challenging one.
This work presents an approach to hand pose reconstruction based upon a low-cost optical motion capture system with three camera views using a glove equipped with differently colored marker objects. The developed system is used to record and process hand motion sequences in order to obtain sequences of complete configuration descriptions of a hand model that is fit onto the extracted three-dimensional data.

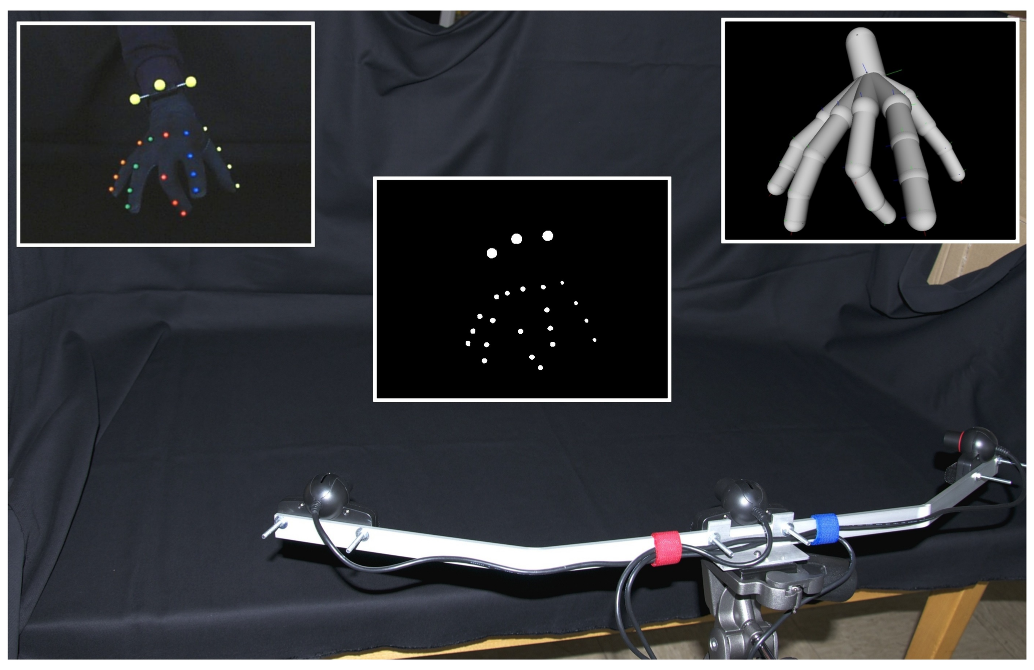

Figure 1: *Experimental stereo vision setup using three cameras with their optical axes loosely focusing on objects in the scene located at a depth of about 75cm. A glove equipped with colored marker objects in order to obtain a 3D model of the human hand pose.*

1. System characteristics

High cost of sophisticated commercial motion capture systems restrict general access to the field. Therefore the presented approach focuses on the low-cost aspect of the system in order to provide an easy possibility for reproduction of the obtained results. With all components being readily available off-the-shelf, such as the Sony PlayStation Eye camera used for the stereo vision setup, the approach offers a simple motion capture system with a reliable accuracy.
The selected camera offers a high image acquisition rate of 60Hz and allows to reduce motion blur associated with the high velocities the fingers can be moved at. Due to a dedicated, yet inexpensive USB controller all cameras can be operated with the full 60Hz without the need for hardware frame synchronization.

2. Model and approach

For the purpose of this work a hand model with 26 DOF was chosen disregarding the full complexity of the CMC joint of the thumb. In order to obtain an initial superimposed three-dimensional location of each joint with respect to the reference frame of the wrist joint, a glove with differently colored spherical marker objects placed roughly over the position of each joint as well as a three-marker wrist fixture were used.

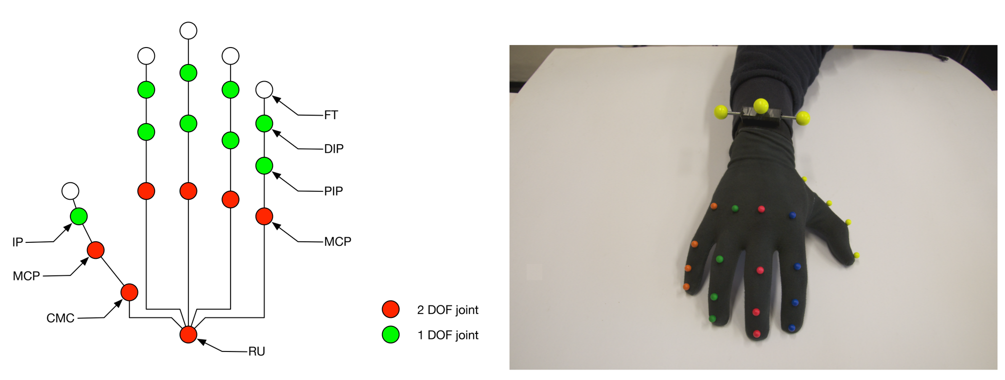

Figure 2: *Hand model (left) with 26 DOF based on the glove design (right) equipped with passive colored markers. Due to the camera's limited ability of color separation, the same color is used for markers of the wrist fixture and the thumb. Marker size was used as the differentiating criterion.*

Use of passive marker objects, while being the least invasive marker-based approach, allows to preserve most of the hands' dexterity. Moreover use of different marker colors enables an easy identification of each finger's kinematic chain reducing the overall computational effort.
Following initial noise reduction, color segmentation and contour search, resulting sets of extracted marker object shapes are evaluated regarding their eccentricity based on geometric moments:

$$\mu_{p,q} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} I(x,y)(x-\overline{x})^p(y-\overline{y})^q \qquad (1)$$

$$\epsilon = \frac{\mu_{2,0}+\mu_{0,2}+\sqrt{(\mu_{2,0}-\mu_{0,2})^2+4\mu_{1,1}^2}}{\mu_{2,0}+\mu_{0,2}-\sqrt{(\mu_{2,0}-\mu_{0,2})^2+4\mu_{1,1}^2}} \qquad (2)$$

Having obtained a set of valid marker object shapes and a set of reconstructed three-dimensional marker locations afterwards, a superimposed hand pose is determined using a distance-based approach. In order to obtain an approximation of the actual joint centers of rotation and therefore the actual hand pose, the approach outlined in figure 3 is used based on the anthropometric constraints determined in [1] and link measurements obtained for the initial pose.

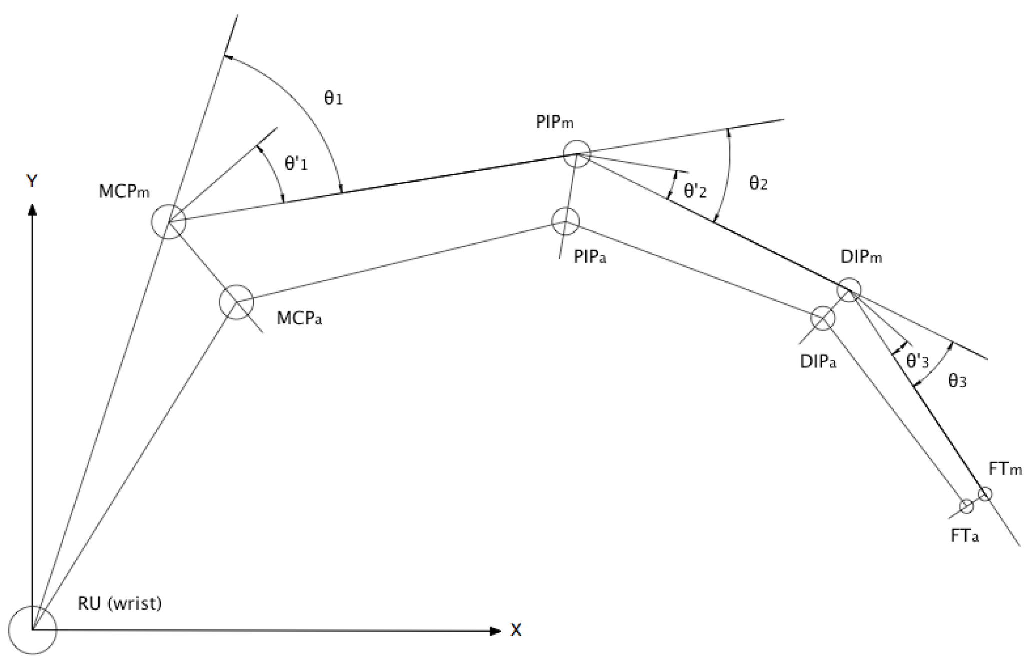

Figure 3: *Approximation of actual joint centers of rotation based on sequential traversal and adjustment of a finger's kinematic chain beginning at the wrist joint.*

3. Experimental results

Among others the tip-to-tip precision grasp shown in figure 4 using the thumb and the index finger was used as a benchmark for repeatability measurement. Very good repeatability was obtained with quite stable joint angle trajectories as shown in figures 5 and 6.

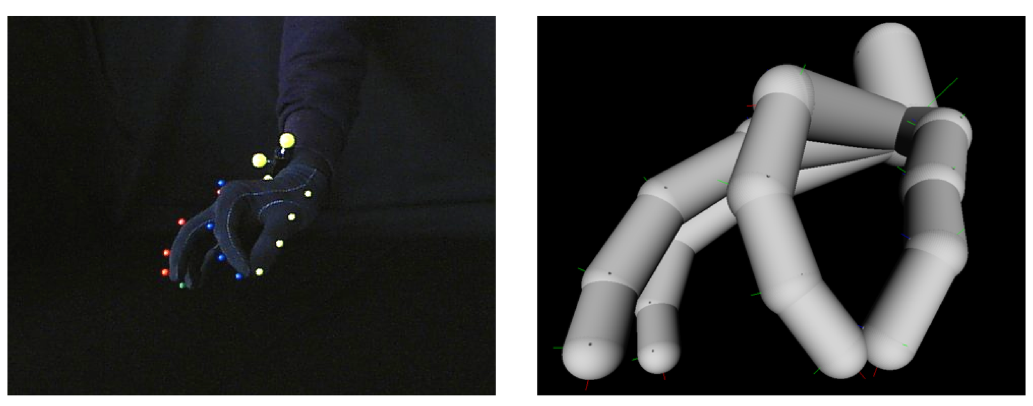

Figure 4: *Repeatability evaluation using a tip-to-tip precision grasp hand motion sequence.*

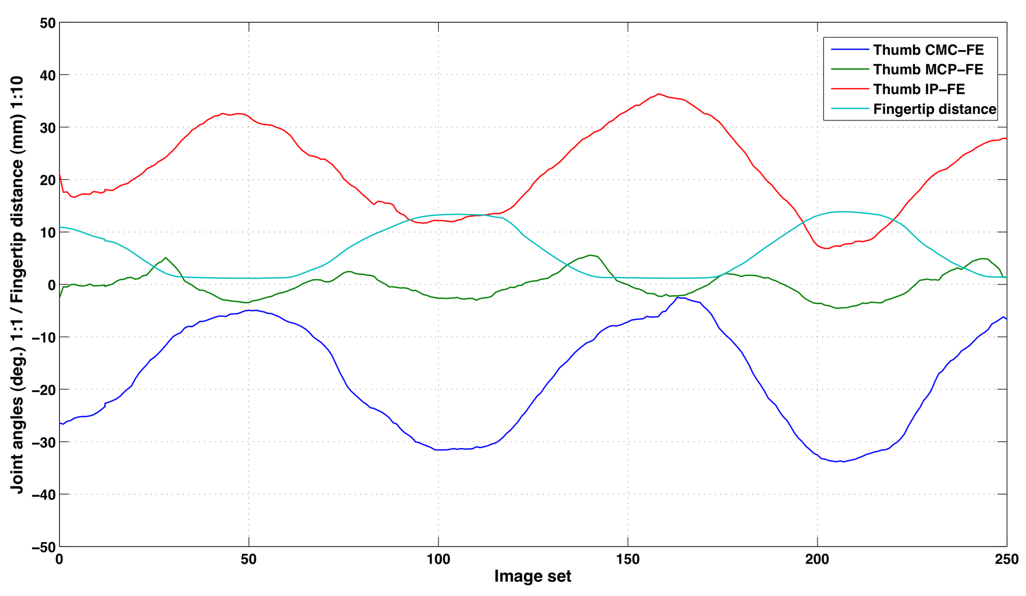

Figure 5: *Joint angle trajectories of the thumb during repeated execution of the tip-to-tip precision grasp.*

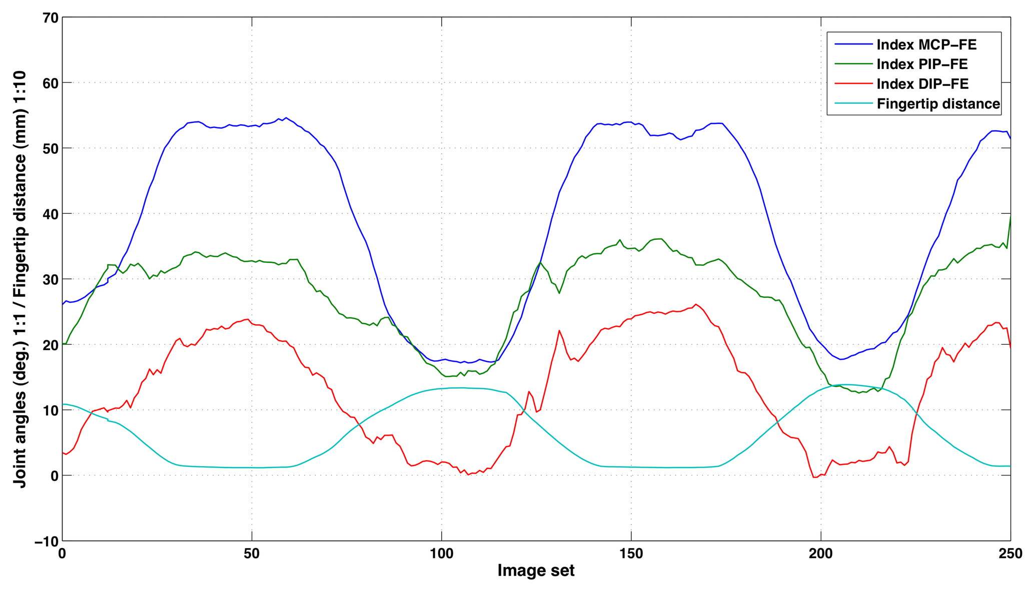

Figure 6: *Joint angle trajectories of the index finger during repeated execution of the tip-to-tip precision grasp.*

Furthermore it was possible to verify the dynamic intra-finger constraint described in [2],

$$\theta_{DIP} = \frac{2}{3}\theta_{PIP} \qquad (3)$$

with the mean angle obtained for the DIP joint being 0.5723 and the standard deviation of ± 0.1513.

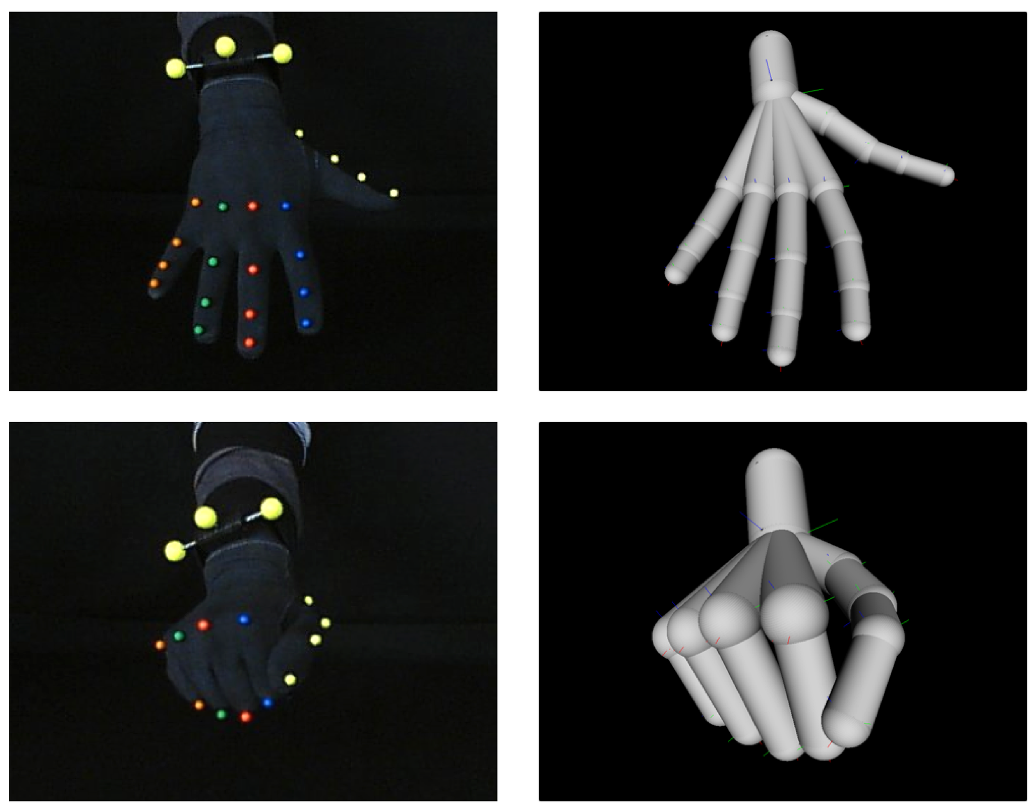

Figure 7: *Further examples of reconstructed hand poses. View of the middle camera (left) and visualization of the obtained full (right/top) or degraded (right/bottom) hand pose.*

4. Conclusion and future work

Despite its simplicity and low-cost components the experimental stereo vision setup has proven to deliver reliable results for the purpose of hand pose reconstruction. It allows to record sequences of simple grasping skills of the human hand in form of series of parameters of the 26 DOF hand model.
Due to time constraints imposed by the combination of image processing steps the approach was split into two stages in order to still be able to benefit from the camera's acquisition frequency: 1. Sequence recording at full frame rate and 2. Offline processing of the recorded sequences.
Planned future work includes usage of Kalman or particle filter techniques for further stabilization of pose reconstruction and reduction of jitter due to image noise. Incorporation of further cameras will be investigated in order to reduce occlusion. Workload distribution, e.g. using the CPU and GPU, in order to achieve real-time performance will be carried out as well.

References

[1] B. Buchholz, T. J. Armstrong, and S. A. Goldstein, *Anthropometric data for describing the kinematics of the human hand*, Ergonomics, vol. 35, 261–273, 1992

[2] J. M. Rehg and T. Kanade, *Visual tracking of high DOF articulated structures: an application to human hand tracking*, Proceedings of the third European conference on Computer Vision, vol. 2, 35–46, 1994

Acknowledgements This work is partially supported by the European FP7 project HANDLE ICT-236410, www.handleproject.eu.