# A Hybrid Approach towards Fully Automatic 3D Marker Tracking

Matthias Weber*
FGAN e.V., FKIE, Germany

## Abstract

Motion Capture is a powerful approach to track 3D position, usually utilizing markers. Especially passive markers do not hinder natural motion. Unfortunately, such systems do not provide any information about which anatomical landmark their markers belong to. Multiple manual actions are often required to guide the tracking process. This work presents a hybrid approach for nearly fully automatic identification and tracking of such markers. It encompasses three methods for identification, using PCA-based alignment or tree-based optimization, and tracking, using a neural network with self-organizing characteristics.

**CR Categories:** I.2.6 [Artificial Intelligence]: Learning—Connectionism and neural nets; I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Motion

**Keywords:** motion capture, hybrid tracking, neural networks

## 1 Introduction

Motion Capture (MoCap) provides a rapid way for acquiring realistic human motion data. Points are assigned to special landmarks on a participant's body with markers which are recorded and tracked subsequently. Wiring of markers as well as heavy-weighted markers should be prevented, to avoid hindering of natural motion. Therefore light-weight passive optical markers are used. Unfortunately, nearly all available passive MoCap systems require manual steps for identification and tracking of captured points. By adding intelligence it is possible to minimize this effort and enhance usability of such systems. To achieve this, this work describes a hybrid solution based on a combination of several algorithms. In an initial step markers have to be identified, either by using a PCA-based alignment of an initial skeleton model or a tree-based optimization procedure to identify markers. After that, the markers have to be tracked through consecutive frames. Neural networks, particularly self-organizing maps (SOMs) are intriguingly promising as they already represent a map of linked skeleton-like structures and have generalization features. Such a neural network can adapt easily to point clouds of tracked human motion data, making it possible to automatically follow markers through frames.

## 2 Related Work

Research on Motion Capture is in particular focused on robust, occlusion free marker tracking and skeleton fitting. For the former problem, [Dorfmüller-Ulhaas 2003] proposed an extended Kalman filter in conjunction with a motion model based on exponential maps. Skeleton fitting is often done by using least-squares methods [O'Brien et al. 2000]. [Hornung and Sar-Dessai 2005] present an integrated approach to robust marker tracking and skeleton fitting. For identifying indistinguishable markers, clique-based recognition techniques were used as well as inverse kinematics techniques, continuity assumptions and other heuristics. Another approach for identification of markers was introduced by Weber et al. [Weber et al. 2008]. They used the algorithms mentioned in this work separately, namely a tree-based optimization to identify markers as well as self-organizing maps to integrate a skeleton model into the tracking process.

To combine the advantages of several methods while avoiding their different disadvantages, hybrid methods encompass several algorithms. A hybrid approach, seems to be a unique investigation in the field of motion tracking. Most other work focuses on combining different tracking technologies to enhance tracking performance but does not combine different algorithms as shown in the following sections.

## 3 Aligning an Internal Skeleton Model

As a prerequisite, a skeleton-like structure is used as a basis to develop a description of the participant's body, in particular the respective anatomical landmarks. The provided skeleton does not need to precisely reproduce the human skeleton in size or proportions. Rather, it needs to specify an approximated link structure of the marker configuration on the subject's body, i.e. of the marker protocol. Figure 1 shows the skeleton model aka marker protocol used in this work.

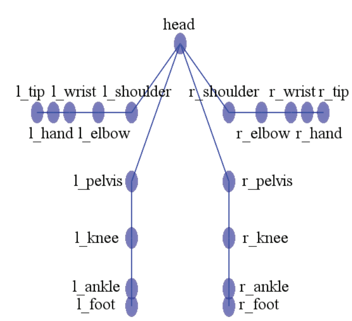


**Figure 1:** *The user-created skeleton model and marker protocol used in this work.*

After creating the skeleton model, acquiring a properly aligned internal model is a helpful first step to identify markers. To configure the internal model, a simple, distinguishable body posture out of the MoCap data is used. When querying an initial posture, the returned point cloud is arbitrarily positioned and oriented. To be able to fit the skeleton to this point cloud the two main axes of the body are calculated via principal components analysis (PCA). It is then projected onto the 2-dimensional space spanned by the two main axes. The skeleton model is scaled to the extensions on these two axes yielding a properly aligned skeleton model. It is retransformed into the normal 3-dimensional space and moved to the median point of the test subject's point cloud. As a result, the model is now aligned to the subject's initial posture. It is finally trained to the initial posture using the SOM neural network (see Section 5) to make identification possible.

*e-mail: m.weber@fgan.de


VRST'08, October 27-29, 2008, Bordeaux, France

## 4 Creating a Skeleton Model Using Tree-Based Optimization

An alternative to aligning an internal skeleton model to a simple posture is to use an optimization algorithm to fit the skeleton directly to the initial posture, using anthropometric data containing fundamental information about the participant's body characteristics. Again, a user-created skeleton is used which provides the link structure. Segment lengths in the skeleton model are computed out of the anthropometric data. Using this model a 3-dimensional tree is constructed. It contains all possible mappings from markers to joints. Discrepancies between reference segment lengths and their marker counterparts are annotated as costs at each created node.

The tree construction algorithm starts from the root node which is the head node. Nodes are created in 3 dimensions, one dimension for the depth of a tree, one for joint names and one for the markers. First, joint names are considered, using the skeleton model. For one joint name tree nodes are created for all markers and linked to their parents. These steps are repeated for all joint names, but without markers that were already created. Additionally, costs are calculated for each node. Previously computed segment lengths are compared to distances between the nodes' and parent nodes' associated markers. These costs are accumulated, i.e. each node's costs are added with the accumulated costs of the node's parent.

After the tree has been built, the best leaf has to be found for each branch that corresponds to a subchain in the skeleton model, e.g. arms and legs. For each end joint name of the skeleton model the corresponding leaf with the least accumulated costs is searched by comparing all costs of leaves corresponding to this joint name.

## 5 Using a Neural Network for Fitting and Mapping Markers

When initialization is finished, the SOM is adapted to each frame of the MoCap data. First, the SOM has to be created. After the initial model was generated, the neurons are linked to the joints of this model. This reference model is used to keep the neural net consistent with regards to the skeleton structure. Then, for the current frame, the position of each tracked marker is presented to the SOM as input. Distances to all prototype vectors are computed, usually using Euclidean distance measure. The neuron with its prototype closest to the input vector is the winning neuron. The prototype vectors were set in the initialization step, when the SOM was created out of the initial skeleton. Each neuron is adapted towards the winning neuron with a certain learning rate and radius.

After adaptation is finished, the markers have to be labeled. This is done by using a nearest neighbors computation that uniquely maps nearest points of two sets. Then, the nearest marker of a neuron gets the name of the neuron. When this step has finished, the process has been completed for the current data frame. The model is now adapted to the markers and the markers are labeled according to the nearest neuron names. Figure 2 shows an example of a tracking process for several frames.

## 6 Evaluation

For capturing motion data, we used a passive optical tracking system with 8 cameras operating at 50Hz made by ART GmbH, Germany. Motion from 30 participants was captured. These were equipped with one 6DOF body above the head and 18 3DOF markers on the other anatomic body landmarks. Anthropometric dimensions were measured for each participant. Each participant performed several motions: It first moved to the initial pose, performed

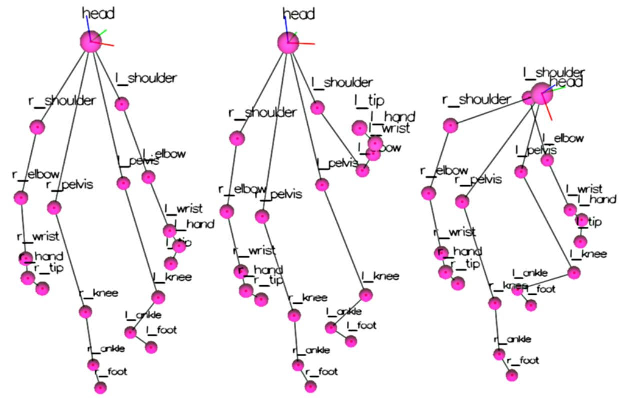


**Figure 2:** *Tracking of several frames.*

arm rotations and body movement, i.e. reaching to several points in space and standing on one foot. Additionally, the initial pose was performed on several different locations with different orientations.

The skeleton alignment algorithm was evaluated for the last movement case. For simple postures, like standing with arms reaching out, the algorithm works 100% correctly. For the tree-based optimization algorithm any standing postures out of the MoCap data were chosen (66 frames). The correct assignment of the recognized marker configuration was checked manually. 80% of the frames were recognized correctly. After correct initialization data was chosen that contained only few, 1-3, marker occlusions mostly on joints with few motion. Performance of the SOM was checked manually and it performed 100% correctly. Finally, the rest of the data, containing situations where even the whole right arm was missing for some frames, was used for adapting the SOM. Unfortunately, the SOM is not very well fitted for this sort of data making manual interaction unavoidable.

## 7 Conclusion and Future Work

This work presents a mostly automatic and robust approach to identify and track markers on a human subject. It encompasses skeleton alignment, tree-based optimization for initialization and a neural network for tracking. Even though such a hybrid approach achieves quite robust tracking results, it still has some manual actions left, namely setting parameters for the algorithms and search for a good start posture. It might be interesting to integrate other methods to further increase robustness and automatic processing.

## References

DORFMÜLLER-ULHAAS, K. 2003. Robust Optical User Motion Tracking Using a Kalman Filter. Tech. Rep. 2003-6, University of Augsburg, Institut für Informatik.

HORNUNG, A., AND SAR-DESSAI, S. 2005. Self-Calibrating Optical Motion Tracking for Articulated Bodies. In *Proc. IEEE Conf. on Virtual Reality*, IEEE Computer Society, 75–82.

O'BRIEN, J. F., BODENHEIMER, R., BROSTOW, G., AND HODGINS, J. K. 2000. Automatic Joint Parameter Estimation from Magnetic Motion Capture Data. In *Graphics Interface*, 53 – 60.

WEBER, M., BEN AMOR, H., AND ALEXANDER, T. 2008. Enhancing Motion Capture Performance by Means of an Internal Anthropometric Skeleton Model. In *SAE Digital Human Modeling Conference*.