Imitation of Human Motion on a Humanoid Robot using Non-Linear Optimization

Martin Do, Pedram Azad, Tamim Asfour, Rüdiger Dillmann

University of Karlsruhe, Germany $do@ira.uka.de$, $azad@ira.uka.de$, $asfour@ira.uka.de$, $dillmann@ira.uka.de$

***Abstract*—In this paper, we present a system for the imitation of human motion on a humanoid robot, which is capable of incorporating both vision-based markerless and marker-based human motion capture techniques. Based on the so-called Master Motor Map, an interface for transferring motor knowledge between embodiments with different kinematics structure, the system is able to map human movement to a human-like movement on the humanoid while preserving the goal-directed characteristics of the movement. To attain an exact and goal-directed imitation of an observed movement, we introduce a reproduction module using non-linear optimization to maximize the similarity between the demonstrated human movement and the imitation by the robot. Experimental result using markerless and marker-based human motion capture data are given.**

I. Introduction

The interaction between robots and humans is one of the main goals in humanoid robotics research. A successful interaction depends on various factors like the acceptance of a humanoid robot by society, its capabilites to act in unconstrained human-centered environments and its communication skills. As a consequence, to raise its acceptance in society, a robot needs to adapt human characteristics to its actions and skills. Especially, human-like motion and gestures of a robot are main contributions to its appearance, which has a strong influence on a user. Hence, under these circumstances controlling the motion of a robot is a very challenging task and still a major topic in humanoid robotics research. The most intuitive solution for this problem lies in imitation, where the user adopts the role of a teacher by demonstrating how to perform a certain action, while the robot tries to repeat this action on the basis of the observation. The benefit of exploiting demonstration is clearly revealed in [1], where an anthropomorphic arm is capable of balancing a pole in the first trial after observing a human. The concept of imitation can be understood in many ways. In [2], imitation of humans in the field of robotics is divided into two categories: imitation learning and motion imitation.

Imitation learning sets the focus on the understanding of actions. Following this scheme, which underlies imitation learning methods, first, data is collected from multiple observations of a demonstrated action. From this data collection features are extracted allowing the robot to draw conclusions on the humans behaviour. Based on the learned behaviour, the robot should be able to reproduce a generalized version of the demonstrated action.

In [3], a neuroscientific inspired approach is presented, which solves imitation learning of cyclic motion with a set of basic motor primitives. These are learned by clustering and dimensionality reduction of visually acquired human motion data. For reproduction, a movement is classified into motor primitives which are played back sequentially.

In [4] and [5] methods are introduced, where Hidden Markov Models are trained with a collection of observations of a demonstrated movement. To reproduce a newly observed movement, the observation is recognized based on a set of trained models. With the complying model, a generalization of the recognized movement is generated.

Imitation learning approaches emphasize the learning and understanding of human behaviour by its interpretation by the humanoid. These methods require offline processing and due to the loss of accuracy as a result of generalization, they are often limited to simple movements.

The imitation of a complex motion requiring high precision and stability is addressed by approaches dealing with the pure imitation of motion. In contrast to imitation learning, the learning of any kind of behaviour is disregarded. Instead, the focus is on finding a trajectory, which corresponds exactly to the data, that a humanoid obtains from a human motion capture system. [6], [7], and [8] present methods for motion imitation, which make use of artificial markers on the humanoid robot as well as the demonstrator. For the reproduction of motion, corresponding marker positions between both subjects are minimized leading to similar postures. Instead of exploiting marker positions, [9] and [10] calculate the joint angles of a demonstrators posture, which are transferred to the robot for execution. Due to joint and velocity constraints, a scaling and transformation process must be performed to obtain a feasible joint angle configuration for the robot. In contrast to the mentioned motion imitation approaches mentioned above, a more natural way of imitation using the humanoid robots own stereo vision system to record human trajectories by exploiting color markers on the demonstrators clothing is presented in [11].

Each of these approaches is focused on a specific human motion capture technique. Since every technique has its advantages and drawbacks, in our approach, we propose a system for the imitation of motion within a framework that allows integration of various marker-based and markerless human motion capture systems and the reproduction on a robot. This compability leads to a high level of flexibility and versatility, which opens the system to a wide range of different applications from motion analysis to imitation of highly complex motions in real-time.

Concerning the reproduction of object manipulation actions, one desires a module that produces trajectories that keep the goal-directedness of the observed movement while keeping the human-like charactericstics of the motion. The term goal-directedness refers to the pose of the end effector relative to the object of interest. Since the pose of the object relative to the robot will always differ from the observed situation, one needs the possibility to incorporate the currently desired end effector pose which can be derived from the currently observed object pose into the transformation procedure.

However, due to severe constraints of mechanical systems and unknown environments, it becomes very difficult to satisfy all requirements. Inspired by the previous works, a reproduction module is developed based on a non-linear optimization problem, which incorporates the robots hand in the task space as well as the joint angles. Similar to the previous imitation solutions, we focus on the optimization of the humanoids posture in each frame.

The paper is organized as follows. Section II describes the proposed imitation system and the human robot used in the experiments. In Section III, an overview of markerless and marker-based human motion capture is given. The extension of the Master Motor Map is described in Section V. The generation of human-like movements from captured human motion using non-linear optimization techniques is presented in Section VI. Finally, experimental results are given in VII.

II. System Overview

As depicted in Fig. 1, the proposed system consists of three major components, which are coupled in consecutive processing stages: the acquisition of human motion, the Master Motor Map (MMM) interface [12], and the motion generation and reproduction.

As mentioned before, the proposed system allows data input from different human motion capture systems. For applications requiring highly accurate data, marker-based motion capture systems are more suitable. In contrast, for online imitation in a natural way, markers cannot be used. For the experiments performed in the context of this paper, the Vicon system [13] was used for marker-based motion capture (see Section III-B) and the stereo-based markerless motion capture system presented in [14] (see Section III-A) for natural imitation.

In both cases, the acquired trajectories are first translated to the unifying MMM format. In order to enhance both, human-likeness and accuracy, the MMM joint angle configuration runs through an optimization procedure, which fits the configuration to the kinematical structure and constraints of the robot. By interpolation between the consecutive posture frames, a smooth imitated movement is generated. If communciation between the single modules becomes necessary, e.g. when using an external Vicon system, UDP is used to establish the connection.

A. ARMAR-IIIb

The humanoid robot ARMAR-IIIb, which serves as the experimental platform in this work, is a copy of the humanoid

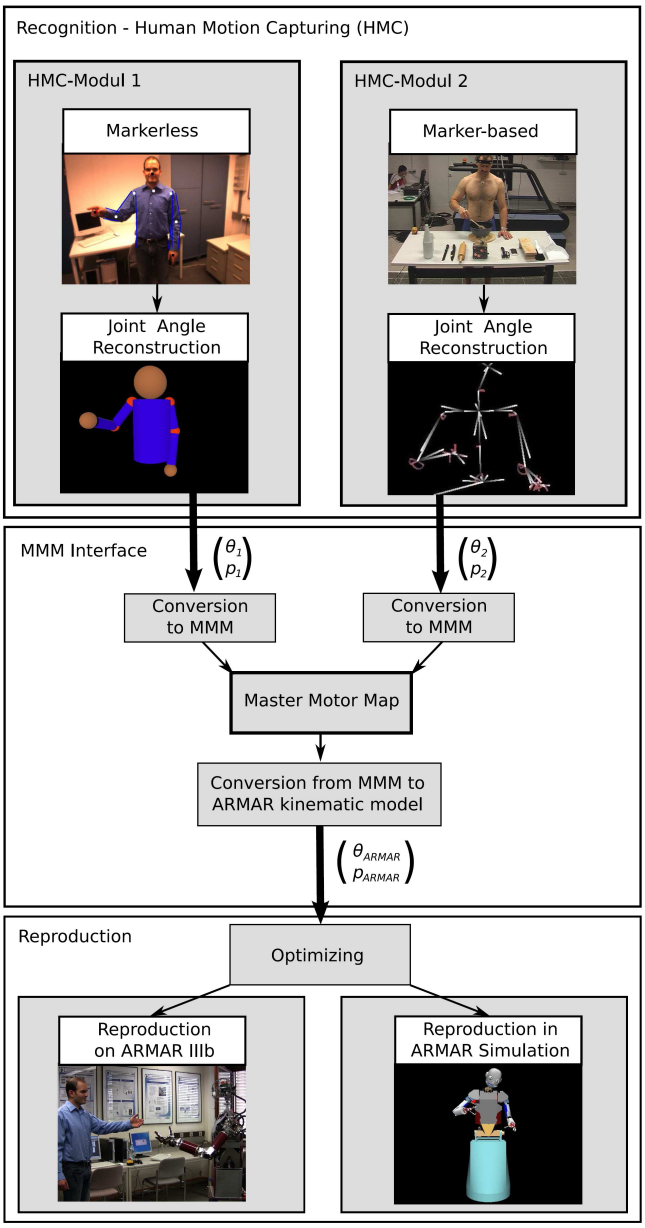

Fig. 1. Overview of the proposed system. θ_x denotes the joint angles, while p_x describe the hand position in the Cartesian space.

robot ARMAR-IIIa [15]. From the kinematics point of view, the robot consists of seven subsystems: head, left arm, right arm, left hand, right hand, torso, and a mobile platform. The head has seven DoF and is equipped with two eyes, which have a common tilt and can pan independently. Each eye is equipped with two digital color cameras, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision. The upper body of the robot provides 33 DoF: 2·7 DoF for the arms and three DoF for the torso. The arms are designed in an anthropomorphic way: three DoF for each shoulder, two DoF in each elbow and two DoF in each wrist. Each arm is equipped with a five-fingered hand with

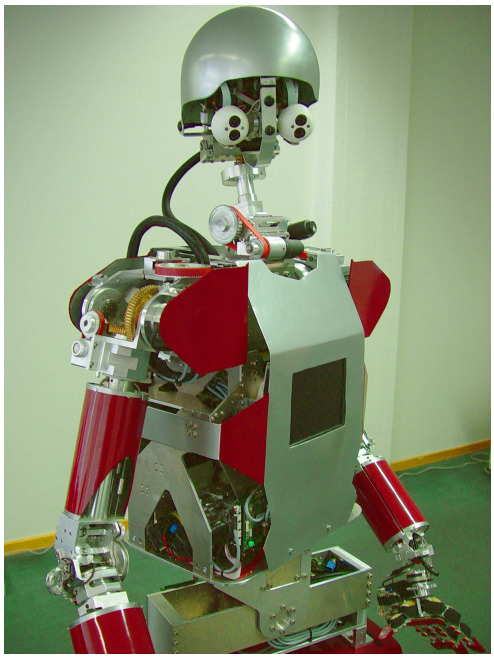

Fig. 2. The humanoid robot ARMAR-IIIb.

eight DoF. The locomotion of the robot is realized using a wheel-based holonomic platform.

III. Human Motion Capture

In this section, a short outline of the integrated markerless and marker-based human motion capture methods is given. In addition to the brief descriptions of the techniques, the advantages as well as the drawbacks are discussed. Furthermore, possible applications are pointed out.

A. Markerless Human Motion Capture

In the following, our real-time stereo-based human motion capture system presented in [14] will be summarized briefly. The input to the system is a stereo color image sequence, captured with the built-in wide-angle stereo pair of the humanoid robot ARMAR-IIIb, which can be seen in Fig. 2. The input images are pre-processed, generating output for an edge cue and a so-called distance cue, as introduced in [16]. The image processing pipeline for this purpose is illustrated in Fig. 3. Based on the output of the image processing pipeline, a particle filter is used for tracking the movements in joint angle space. For tracking the movements, a 3D upper body model with 14 DoF (6 DoF for the base transformation, 2·3 for the shoulders, and 2·1 for the elbows) consisting of rigid body parts is used, which provides a simplified description of the kinematic structure of a human. The model configuration is determined by the body properties like the limbs length of the observed human subject. The core of the particle filter is the likelihood function that evaluates how well a given model configuration matches the current observations, i.e. stereo image pair. For this purpose, an edge cue compares the projected model contours to the edges in the image. On the basis of an additional 3D hand/head tracker, the distance cue evaluates the distance between the measured positions and the corresponding positions inferred by the forward kinematics of the model. Various extensions are necessary for robust real-time application such as a prioritized fusion method, adaptive shoulder positions, and the incorporation of the solutions of the redundant arm kinematics. The system is capable of online tracking of upper body movements with a frame rate of 15 Hz on a 3 GHz single core CPU. Details are given in [14].

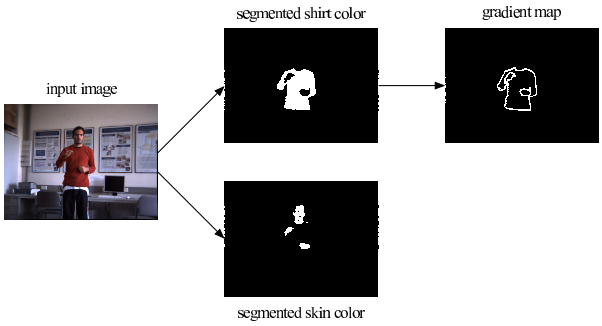

Fig. 3. Illustration of the image processing pipeline.

B. Marker-based Human Motion Capture

Marker-based human motion capture frameworks are widespread systems in the robotics research community as well as in the industry. One of the most popular commercially available systems is provided by Vicon [13]. The technique, which is used here, relies on infrared cameras and artificial reflective markers. The markers are placed on predefined body parts of a human subject. In a defined workspace, the subject is surrounded by a set of infrared cameras. Each camera is equipped with a infrared strobe, emitting a light signal, which is reflected by the markers. The reflected light, which distinguishes itself from the background, is registered by the cameras. The data from each camera consisting of 2D coordinates of each recognized marker position, is merged in a data station, which computes the 3D position by triangulation and the label of each visible marker. Besides the hardware, the system contains a comprehensive software package, which facilitates the calibration and handling of the system. Due to the high-speed and high-resolution properties of the cameras, the Vicon system provides an accurate method for capturing human motion at high frame rates. Furthermore, since the use of numerous markers allows capturing of barely visible motion of unobvious joints, complex kinematic models are applicable for the processing and representation of the motion data. The problem of occlusion of body parts is reduced to a minimum, since multiple cameras are used, which deliver multiple views of the same subject. However, the enormous equipment needs cause high costs. Furthermore, a time and space-consuming preparation is essential to provide the necessary setup for proper human motion capture. For our purpose, the joint angles are reconstructed by optimization of a human model based on the computed 3D marker positions. Details are given in [17].

IV. Extended Master Motor Map

Since each human motion capture system produces data in terms of its own specific model and format, respectively, one has to deal with a variation of different data formats. Likewise, for reproduction of movements, each robot system requires data in terms of its own kinematics. One possible solution could be the definition of an interface for each combination of a sensing system and a robot. However, doing so would restrict the robot and the utilization of the data. To overcome this difficulties, in this work, a standardized interface is established by using the MMM, which features a high level of flexibilty and combability. The MMM is introduced in [12] and provides a reference kinematic model by defining the maximum number of DoF, that can be used by a human motion capture module and a robot. A trajectory in the original MMM file format consists of 52-dimensional vectors, each vector describing a joint angle configuration with a floating point number for every single DoF. Since we are able to recognize the human finger movements using the Vicon system, the MMM is extended by three DoF for coupled finger flexion, thumb flexion and thumb abduction. The reference kinematic model of the extended MMM is illustrated in Fig. 4. As a result of the extension, one obtains a 58-dimensional vector for the description of a joint angle configuration of the model. Due to differences in the Euler conventions, active joint sets, which can be controlled, and the order of the joint angle values between the modules, a conversion module has to be implemented for each of the systems in order to provide a proper connection via the MMM. This conversion module transforms the module specific data into the MMM file format and vice versa. As depicted in Fig. 1, for the proposed system, one conversion module is implemented for each human motion capture system, converting the motion capture data to the MMM format. A third conversion module is implemented for mapping the MMM data to the kinematics of ARMAR-IIIb. The focus of this paper lies on this third module and is presented in the following. Further details on the MMM are given in [12].

V. Reproduction of Trajectories

Concerning the imitation of humanoid motion, the simplest and most desired way to reproduce a movement from given joint angles consists of a one-to-one mapping between an observed human subject and the robot. Unfortunately, due to the differences in the kinematic structures of a human and the robot e.g. differing joints and limb measurements, only in rare cases a one-to-one mapping shows acceptable performance regarding the functionality as well as the human-like appearance of the reproduced movement. In this work, we address this problem by applying a postprocessing procedure in joint angle space. In two stages, the joint angles, given in the MMM format, are optimized concerning the tool center point (TCP) position and the kinematic structure of the robot. First, a feasible solution is estimated, which serves as an intial solution for an optimization step in the second stage. Following this scheme, one obtains a human-like motion on the robot, while preserving its goal-directed characteristics.

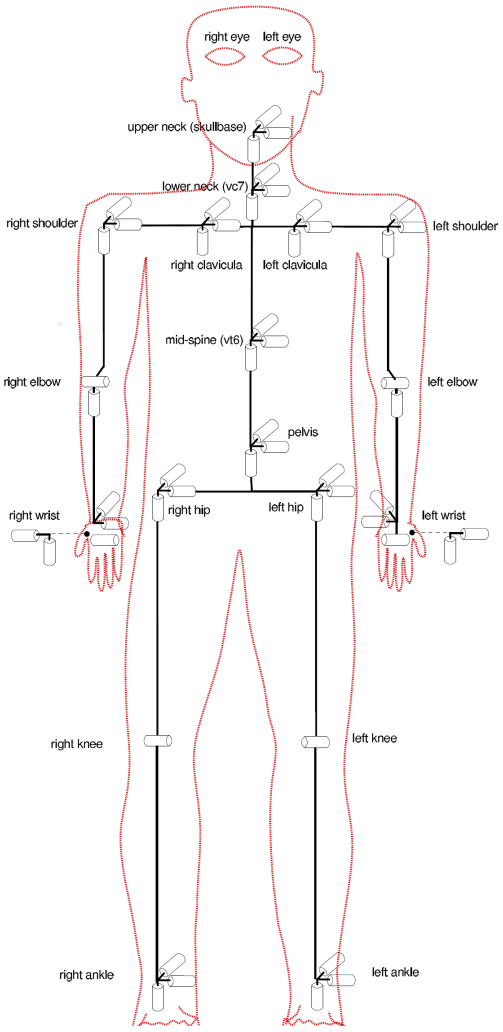

Fig. 4. Reference kinematic model of the Extended Master Motor Map.

A. Similarity Measure

One of the most crucial factors in the reproduction of human motion is the measure for rating the similarity between the imitated and the demonstrated movement. For the online reproduction of a human movement, one is more interested in comparing the current postures at the time t than in investigating a whole trajectory. In [7], it is proposed to determine the distance between the postures of the robot and the human by exploiting point correspondences between specified points on both bodies. To infer useful statements concerning the similarity, accurate localization and identification of the limbs are required, which makes the use of physical markers inevitable. In [9], a similarity measure is introduced, which only considers the joint angle relations. However, it disregards structural differences between human and robot like differing limb lengths, which one has to take into account in order to preserve the goal of a movement when mapped on the robot. Combining both, the joint angle configuration and key point correspondences, for a joint angle configuration $\boldsymbol{\sigma} \in \mathbb{R}^n$ with n joints, we define the similarity measure as follows:

$$S(\boldsymbol{\sigma}) = 2 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma_i}^t - \sigma_i\right)^2}{\pi^2} - \frac{\frac{1}{3}\sum_{i=1}^{3}\left(\hat{p_k}^t - p_k\right)^2}{\left(2 \cdot l_{arm}\right)^2} \quad (1)$$

with $\sigma_i, \hat{\sigma_i}^t \in [0, \pi]$ and $p_k, \hat{p_k}^t \in [-l_{arm}, l_{arm}]$, whereas l_{arm} describes the robots arm length. The reference joint angle configuration is denoted by $\hat{\boldsymbol{\sigma}} \in \mathbb{R}^n$, while $\hat{\boldsymbol{p}} \in \mathbb{R}^3$ stands for the desired TCP position. The current TCP postion $\boldsymbol{p}$ can be determined by applying the forward kinematics of the robot to the joint angle configuration $\boldsymbol{\sigma}$.

B. Estimation of an Initial Solution

To obtain a posture, which bears a high resemblance to the one of the demonstrator and at the same time meets all the mechanical constraints of the robot, the original joint angle configuration is optimized regarding the similarity measure as specified in Eq. 1. An optimal solution is found by applying a numerical optimization algorithm, namely the Levenberg-Marquardt (LM). However, the efficiency of most of the numerical optimization algorithms strongly depends on the initial estimation of the parameters to be optimized. An initial estimation within the neighbourhood of the optimal solution leads to a high chance that the algorithm converges fast directly towards the optimum without being trapped in local extrema. In this work, an initial estimation is determined from a preselection of candidate initial joint angle configurations, which are generated and evaluated by means of the similarity measure. To generate a candidate initial estimation $\boldsymbol{\sigma}^j$, the reference joint angle configuration $\hat{\boldsymbol{\sigma}}^t$ computed at time t is mapped into the robot joint angle space and projected on the bound constraints:

$$\hat{\sigma}_i^t = \begin{cases} C_{i_{min}} & \text{if } \hat{\sigma}_i^t \leq C_{i_{min}} \\ \hat{\sigma}_i^t & \text{if } C_{i_{min}} \leq \hat{\sigma}_i^t \leq C_{i_{max}} \\ C_{i_{max}} & \text{if } \hat{\sigma}_i^t \geq C_{i_{max}} \end{cases} \quad (2)$$

where $C_{i_{min}}$ and $C_{i_{max}}$ denote the lower and upper joint angle bounds of joint i. If the value of $\hat{\sigma}_i^t$ exceeds the given bounds, the joint i is fixed at the closest of the two boundaries. A candidate is obtained by altering each non-fixed joint angle of the mapped configuration by means of a vector $\boldsymbol{\delta}^t \in \mathbb{R}^n$ with $\delta_i^t = \hat{\sigma}_i^t - \hat{\sigma}_i^{t-1}$. Thus, $\boldsymbol{\delta}^t$ describes the changes between two consecutive frames. As a result, a candidate initial estimation can be described as:

$$\sigma_i^j = \hat{\sigma}_i^t + \alpha_i \beta_i \quad (3)$$

with

$$\alpha_i = \begin{cases} 1 & \text{if } C_{i_{min}} \leq \hat{\sigma}_i^t \leq C_{i_{max}} \\ 0 & \text{else} \end{cases} \quad (4)$$

$$\beta_i \in \{-\delta_i^t, 0, \delta_i^t\} \quad (5)$$

Given n joints to control, in the worst case $M = 3^n$ candidates need to be calculated and evaluated. The best initial estimation satisfies the following equation:

$$\boldsymbol{\sigma}_{init} = \underset{j=1,\ldots,M}{\arg\max}\ S(\boldsymbol{\sigma}^j) - \|\hat{\boldsymbol{\sigma}}^t - \boldsymbol{\sigma}^j\| \quad (6)$$

Finding the best initial estimation causes some overhead regarding processing time, but it is necessary to ensure that the LM algorithm will provide an optimal solution.

C. Optimization Problem

For optimization of a reference joint angle configuration regarding the similarity measure, one can use the Levenberg-Marquardt algorithm. The algorithm, which was first introduced in [18], provides a standard technique for solving non-linear least squares problems by iteratively converging to a minimum of function expressed as sum of squares. Combinating the Gauss-Newton and the steepest descent method, the algorithm unites the advantages of both methods. Hence, using the LM method, a more robust convergence behaviour is achieved at points far from a local minimum, while a faster convergence is gained close at a minimum. Due to its numerical stability, the LM method has also become a popular tool for solving inverse kinematics problems as demonstrated in [19]. For our problem, where, given the reference joint angle configuration $\hat{\boldsymbol{\sigma}}^t$, we seek a $\boldsymbol{\sigma}^t$, which maximizes Eq. 1. To interpret Eq. 1 as a function of sum of squares to be minimized, we define a function $s(\boldsymbol{\sigma}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $n < m$ as follows:

$$s(\boldsymbol{\sigma}) = \begin{pmatrix} \frac{1}{\sqrt{3}\cdot 2\cdot l_{arm}}\ p_1 \\ \frac{1}{\sqrt{3}\cdot 2\cdot l_{arm}}\ p_1 \\ \frac{1}{\sqrt{3}\cdot 2\cdot l_{arm}}\ p_1 \\ \frac{1}{\sqrt{n}\cdot \pi}\ \sigma_1 \\ \vdots \\ \frac{1}{\sqrt{n}\cdot \pi}\ \sigma_n \end{pmatrix} \quad (7)$$

The corresponding optimization problem can be written in the following form:

$$\min S'(\boldsymbol{\sigma}) = 2 - S(\boldsymbol{\sigma}) \quad (8)$$

$$\text{subject to} \quad C_{i_{min}} \leq \hat{\sigma}_i \leq C_{i_{max}} \quad (9)$$

which is equivalent to the maximization of Eq. 1. Similar to the Gauss-Newton method, in the LM method a Taylor expansion of s is performed around $\boldsymbol{\sigma}$. For a small $\boldsymbol{\rho}$, s can be approximated by the following equation:

$$s(\boldsymbol{\sigma} + \boldsymbol{\rho}) \approx s(\boldsymbol{\sigma}) + J_s \boldsymbol{\rho} \quad (10)$$

where J_s denotes the Jacobian of s. Based on the initial guess $\boldsymbol{\sigma}_{init}$, a sequence of estimations $\boldsymbol{\sigma} + \boldsymbol{\rho}$ is calculated that converges to a solution of Eq. 9. Therefore, in each iteration, the optimization problem is reduced to finding a $\boldsymbol{\rho}$, that minimizes $\|s(\hat{\boldsymbol{\sigma}}^t) - s(\boldsymbol{\sigma}) + J_s\boldsymbol{\rho}\|$. For an adequate $\boldsymbol{\rho}$ the following condition must hold true:

$$\left(s(\hat{\boldsymbol{\sigma}}^t) - s(\boldsymbol{\sigma}) + J_s\boldsymbol{\rho}\right) J^T = 0 \quad (11)$$

Solving the least squares problem of Eq. 11 yields the sought $\boldsymbol{\rho}$. Based on Eq. 11, the LM algorithm solves following slightly modified equation:

$$I\mu + J_s^T J_s \boldsymbol{\rho} = J_s^T s(\boldsymbol{\sigma} + \boldsymbol{\rho}) \quad (12)$$

which includes a dampening term μ. If reduction of S' concerning $\boldsymbol{\rho}$ can be accomplished, then for the next iteration $\boldsymbol{\sigma} := \boldsymbol{\sigma} + \boldsymbol{\rho}$ holds and a smaller value is assigned to μ to achieve faster convergence. If reduction fails, μ is set to a higher value, which slows down the convergence. Furthermore, μ prevents meeting singularites in the Jacobian. To obtain a feasible joint angle configuration, after each iteration, $\boldsymbol{\sigma}$ is projected onto the bound constraints according to Eq. 2. The algorithm terminates if $S'(\boldsymbol{\sigma}) < \epsilon_1$ or $\|\boldsymbol{\rho}\| < \epsilon_2$, and one can set $\boldsymbol{\sigma}^t = \boldsymbol{\sigma}$. More practical details on the algorithm can be found in [20].

VI. Experimental Results

In this section, results of experiments of the imitation system with the two human motion capture systems introduced in Section III are demonstrated. The approach was evaluated by comparison to an inverse kinematics method based on the Jacobian transpose and a one-to-one mapping of the captured joint angles onto the robot. The results were generated with the humanoid robot platform ARMAR-IIIb in real-world as well as in simulation.

A. Marker-based Motion Capture Data

The hardware setup which was used to capture the human motion consists of ten Vicon cameras. Since using a marker-based approach allows to capture a large set of degrees of freedom, the number of active joint angle adds up to 24 DoF, ten for each arm, three DoF for the head and one DoF for the hip rotation. Concerning the arm, three DoF are assigned to the shoulder rotation, two for the elbow, two for the wrist and three DoF describe the finger movements. The experiments focused on the reproduction of actions in a kitchen scenario. The data was generated within the work of [21]. The kitchen actions included movements like stirring, cutting with a knife, sweeping, grinding coffee beans, grating, and pouring. Fig. 8 shows screenshots of cutting sequence, which was reproduced on ARMAR in simulation. The results using the optimization as proposed in this work on marker-based captured motion data are illustrated in Fig. 5. The left plot of Fig. 5 shows the joint angle error of a reproduced joint angle configuration on the robot and the reference configuration. Due to redundancy, the inverse kinematics method produces results with higher error, while a one-to-one mapping naturally leads to a minimal error. In the center plot of Fig. 5, by the right arm TCP, the deviation of the TCP positioning is illustrated. Here, given a TCP destination, the inverse kinematics method leads to an exact positioning of the TCP, while using the one-to-one mapping the destinated position is not reached. In both plots, it is shown, that the application of the optimization procedure as proposed in Section V, a tradeoff is attained, which results in a quite accurate TCP positioning with an maximum error of 25 mm and an acceptable mean joint angle error of 2.0 degrees for each DoF. One of the most crucial joints which has a huge impact on the style of a trajectory is the shoulder joint. Therefore, the right plot of Fig. 5 shows the joint angle error for this joint in particular.

B. Markerless Motion Capture Data

For the online reproduction and imitation of the observed human motion, the stereo camera system of ARMAR-IIIb was used to capture the upper body movements with the method described in Section III-A. Using the onboard cameras allows to perform a more natural way of imitation, but limits the number of DoF, which can be measured, since the system is more sensitive to noise and occlusion. A total number of eight DoF is used, four for each arm, three DoF for the shoulder joint and one for elbow flexion. The under arm rotation, the wrist, and finger movements cannot be recognized with this system. The online reproduction was tested with simple movements like reaching, waiving, and approaching certain postures. Some sample images showing the online imitation of human motion can be seen Fig. 7. Similar to the results achieved with the Vicon system, applying the proposed motion imitation system leads to a tradeoff between the accuracy of the TCP position and the joint angle error. However, due to the reduced number of measured joints, one obtains results with a mean joint angle error of 2.7 degrees for each DoF, as shown in the left plot of Fig. 6, and a maximum deviation of 65 mm in the TCP position of the right arm, as shown in the center plot of Fig. 6. The reason for the relatively large deviation is that the utilized vision-based motion capture system is not yet capable of measuring the torso rotation. This lack information leads to a decreased flexibility throughout the reproduction, assuming the hip joint angles to be fixed. One solution would be to incorporate the hip rotation into the optimization procedure in order to allow for the missing flexibility even if the torso rotation cannot be measured.

VII. Conclusions

In this work, we have presented a system for motion imitation with the goal of attaining a human-like motion, but without loss of functionality. Based on the Master Motor Map, a system was developed, which is capable of incorporating various human motion capture techniques. In particular, it was dealt with the marker-based Vicon system and a markerless vision-based approach. Their output is transformed to the structure of the robot platform ARMAR-IIIb by using a non-linear optimization technique in form of the LM algorithm. A natural way of motion imitation is demonstrated by applying the system successfully for the online reproduction of observed motion. Furthermore, with the system, reproduction of complex kitchen actions was achieved based on high-resolution Vicon data. In the near future, the involvement of objects is planned to enable imitation of manipulation tasks. The proposed system provides a solid basis for further studies towards human motion analysis. By incorporating machine learning methods, the system can be extended imitation learning tasks.

VIII. Acknowledgment

The work described in this paper was partially conducted within the EU Cognitive Systems projects GRASP (FP7-215821) and PACO-PLUS (FP6-027657) funded by the European Commission.

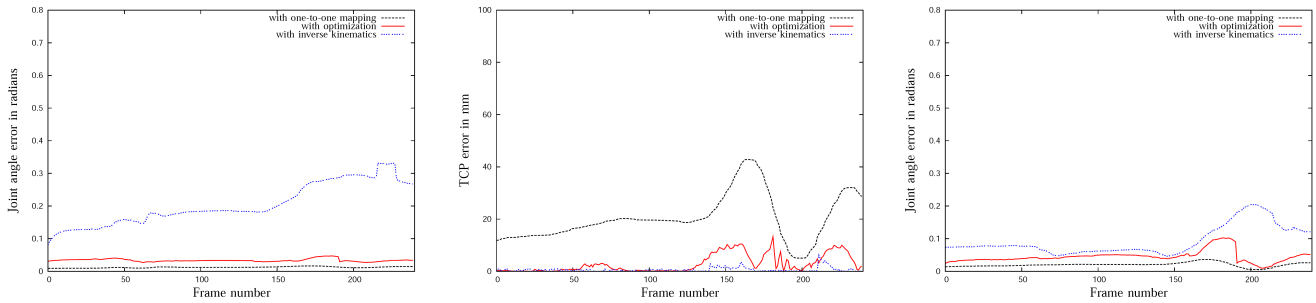

Fig. 5. Evaluation results for the reproduction of motion captured by the Vicon system. Left: Mean joint angle error over all active joints in radians. Center: Deviation of right arm TCP of the robot and a predefined destination in mm. Right: Mean joint angle error over the shoulder joint in radians.

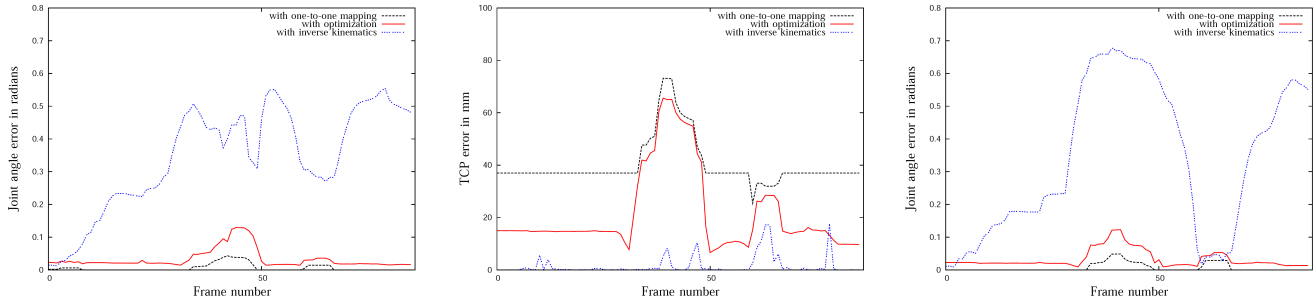

Fig. 6. Evaluation results for the reporduction of vision-based captured motion. Left: Mean joint angle error over all active joints in radians. Center: Deviation of right arm TCP of the robot and a predefined destination in mm. Right: Mean joint angle error over the shoulder joint in radians.

REFERENCES

[1] S. Schaal, "Learning from Demonstration," in *Advances in Neural Information Processing Systems 9*, Denver, USA, December 1997, pp. 1040–1046.

[2] P. Bakker and Y. Kuniyoshi, "Robot see, Robot do: An Overview of Robot Imitation," in *AISB96 Workshop: Learning in Robots and Animals*, Brighton, UK, 1996.

[3] M. J. Mataric, "Getting Humanoids to Move and Imitate," *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 18–24, 2000.

[4] S. Calinon and A. Billard, "Learning of Gestures by Imitation in a Humanoid Robot," in *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge University Press, 2007, pp. 153–177.

[5] T. Asfour, F. Gyarfas, P. Azad, and R. Dillmann, "Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots," in *IEEE-RAS International Conference on Humanoid Robots*, Genova, Italy, December 2006, pp. 40–47.

[6] C. Kim, D. Kim, and Y. Oh, "Adaption of Human Motion Capture Data to Humanoid Robots for Motion Imitation using Optimization," *Integrated Computer-Aided Engineering*, vol. 13, no. 4, pp. 377–389, 2006.

[7] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, "Generating Natural Motion in an Android by Mapping Human Motion," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Alberta, Canada, August.

[8] A. Ude, C. Atkeson, and M. Riley, "Programming Full-Body Movements for Humanoid Robots by Observation," *Robotics and Autonomous Systems*, vol. 47, no. 2-3, pp. 93–108, 2004.

[9] X. Zhao, Q. Huang, Z. Peng, and K. Li, "Humanoid Kinematics Mapping and Similiarity Evaluation based on Human Motion Capture," in *IEEE International Conference on Information Acqusition*, Hefei, China, June 2004, pp. 426–431.

[10] N. Pollard, J. Hodgins, M. Riley, and C. Atkeson, "Adapting Human Motion for the Control of a Humanoid Robot," in *IEEE International Conference on Robotics and Automation*, Washington, DC, USA, May 2002, pp. 1390–1397.

[11] M. Riley, A. Ude, K. Wade, and C. Atkeson, "Enabling Real-Time Full-Body Imitation: A Natural Way of Transferring Human Movement to Humanoids," in *IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003, pp. 2368–2374.

[12] P. Azad, T. Asfour, and R. Dillmann, "Toward an Unified Representation for Imitation of Human Motion on Humanoids," in *IEEE International Conference on Robotics and Automation*, Rome, Italy, April 2007.

[13] "Vicon Peak," Website, available online at http://www.vicon.com.

[14] P. Azad, T. Asfour, and R. Dillmann, "Robust Real-time Stereo-based Markerless Human Motion Capture," in *submitted to International Conference on Humanoid Robots*, Daejeon, Korea, December 2008.

[15] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control," in *IEEE/RAS International Conference on Humanoid Robots*, 2006.

[16] P. Azad, A. Ude, T. Asfour, G. Cheng, and R. Dillmann, "Image-based Markerless 3D Human Motion Capture using Multiple Cues," in *International Workshop on Vision Based Human-Robot Interaction*, Palermo, Italy, 2006.

[17] H. Köhler, M. Pruzinec, T. Feldmann, and A.Wörner, "Automatic Human Model Parametrization from 3D Marker Data for Motion Recognition," in *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, February 2008.

[18] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *The Quarterly of Applied Mathematics*, vol. 2, pp. 164–168, 1944.

[19] C. W. Wampler, "Manipulator Inverse Kinematic Solutions based on Vector Formulations and Damped Least Squares Methods," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, pp. 93–101, 1986.

[20] M. L. A. Lourakis and A. A. Argyros, "Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment ?" in *IEEE International Conference on Computer Vision*, vol. 2, Beijing, China, October 2005.

[21] T. Stein, A. Fischer, I. Boesnach, H. Köhler, D. Gehrig, and H. Schwameder, *Kinematische Analyse menschlicher Alltagsbewegungen für die Mensch-Maschine-Interaktion*. V. Aachen: Shaker, 2007, in press.

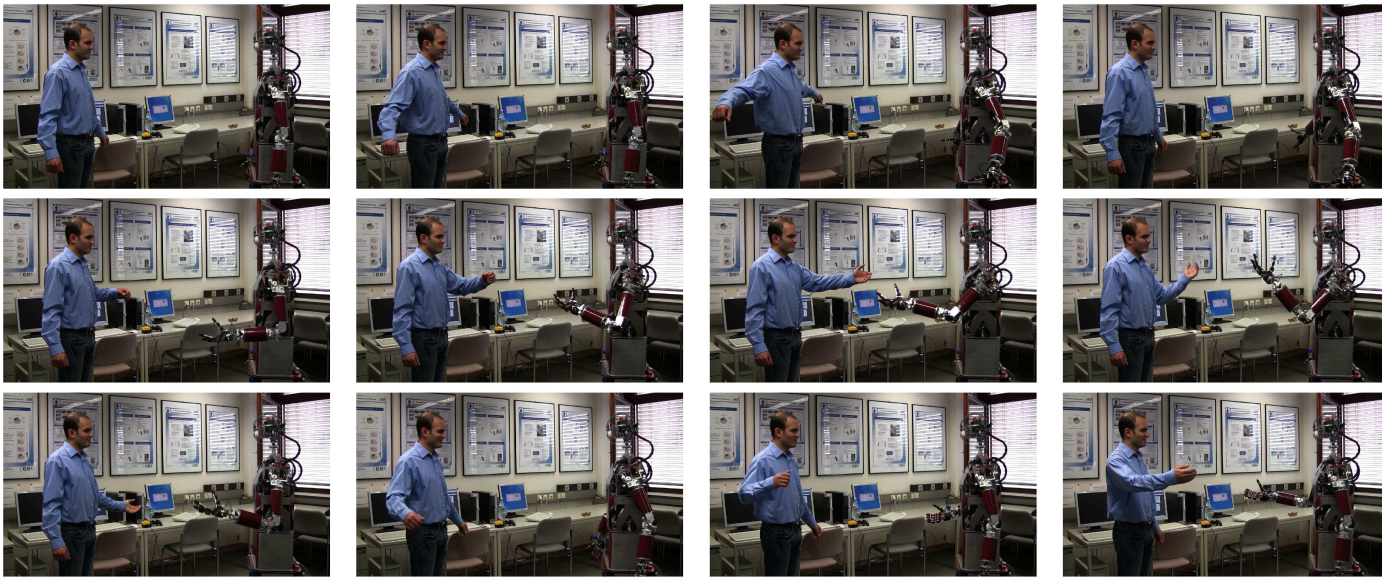

Fig. 7. Image samples of the online imitation of human motion by the humanoid ARMAR-IIIb.

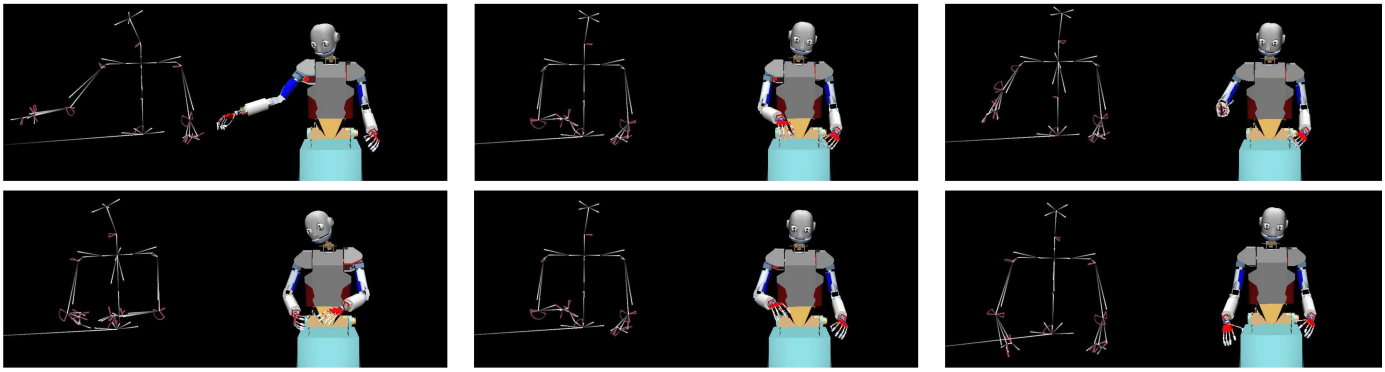

Fig. 8. Image sequence of a cutting trajectory captured by the Vicon system and the reproduction in the ARMAR III simulation.