

Hand tracking with physiological constraints

Andreas Aristidou¹

Published online: 11 October 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Articulated hand tracking systems have been commonly used in virtual reality applications, including systems with human–computer interaction or interaction with game consoles; hand pose estimation has various other applications comprising sign language recognition and animation synthesis. The advanced technological achievements in motion capture over the last decade allow data acquisition with high accuracy and low cost. However, due to the high complexity of the human hand, it is still challenging to animate a hand model able to deal in details with the anatomical and physiological constraints of the hand. In this paper, we present a simple and efficient methodology for tracking and reconstructing 3D hand poses. Using an optical motion capture system, where markers are positioned at strategic points, we manage to acquire the movement of the hand and establish its orientation using a minimum number of markers. An Inverse Kinematics solver was then employed to control the postures of the hand, subject to physiological constraints that restrict the allowed movements to a feasible and natural set. The proposed methodology produces smooth and biomechanically correct movements, while the required processing time remains low, enabling an effective real-time hand motion tracking and reconstruction system.

Keywords Geometric Algebra · Hand Tracking · Inverse Kinematics · Motion Capture · Physiological Constraints

Electronic supplementary material The online version of this article (doi:[10.1007/s00371-016-1327-8](https://doi.org/10.1007/s00371-016-1327-8)) contains supplementary material, which is available to authorized users.

✉ Andreas Aristidou
a.aristidou@ieee.org

¹ Department of Computer Science, University of Cyprus, Nicosia 1678, Cyprus

1 Introduction

In recent years, there has been a growing demand for reliable hand motion tracking systems, which find applications in hand gesture and sign language recognition, to generate virtual figures for films or computer games, and for human–computer interaction (HCI), including interaction with game consoles. However, building a fast and effective hand pose tracker with full articulation remains challenging. The high dimensionality of the hand pose space, in our examples 31 Degrees of Freedom (DoF), its mechanical complexity, the ambiguities due to self-occlusions and the significant appearance variations due to shading, make efficient tracking difficult. To achieve a successful reconstruction and animation of the movement, the hand model is required to be structured taking into consideration the anatomical and physiological properties of the hand.

Many methods have been proposed to deal with hand tracking. Glove-based methods usually require a large number of markers to be attached on the hand and/or need of additional information in order to deal with marker occlusions [78]. The current trend is to use vision-based methods, such as infrared or RGB-Depth (RGB-D) cameras (e.g. [68, 70]). These methods can efficiently track the hand movement without requiring the user to wear a glove or attach markers. However, apart from their small capturing volume, it is challenging to automatically splice the hand captures with full-body acquisition to attain holistic performance capturing. Since it is not always possible to simultaneously capture fingers and full-body motions, hand movements are usually captured in separate sessions and later connected with full-body animation [44], or are manually created by animators; nevertheless, manual interventions may result in poor synchronization between the body and hand movements.

In this work, we introduce an efficient and highly structured method for accurate hand tracking and animation of virtual hands; it utilizes optical motion capture (mocap) data and can be used for performance capturing or precise interactive control in large capture volumes. Marker-based motion capture has been demonstrated in several interactive systems (including but not limited to hand reconstruction) producing results which are highly accurate and easily configurable. There are, however, instances where the number of markers that is possible to be attached on each limb segment is limited, complicating the reconstruction of the hand. To overcome these limitations and achieve real-time hand animation, a new hand model is presented; this model uses a reduced set of eight markers to capture the full articulation of the complex hand movement. A single labeled marker is attached and tracked on each finger, one marker at the wrist (root) and two additional markers at strategic positions to define the hand orientation. The problem is solved as an optimization process that includes Inverse Kinematics, to fit the joint positions to the hand model, and a marker prediction framework, combined with joint constraints, to maintain a continuous data-flow when occlusion of markers by other elements leads to missing data. The optimization process is then completed by integrating physiological constraints to eliminate the production of unnatural motion, ensuring that the reconstructed hand motion is within a feasible set. The results were visualized using a mesh deformation algorithm, demonstrating that the produced motion is smooth and without oscillations.

The proposed hand structural method is real-time implementable; it does not use pre-captured data, or search in large database to match a pose. It is not limited to low-accuracy hand interactive applications, and can find applications in hand gesture recognition, sign language, and high-quality performance animation. The main contribution of this paper is a two-step optimization process that consists:

- A simple framework that uses a reduced set of markers and Inverse Kinematics (IK) to track and reconstruct the full hand and fingers articulation; the hand data can be automatically spliced to full-body acquisition to obtain holistic performance capturing.
- Physiological constraints to ensure production of anatomically natural movements that are within a feasible set; in addition, they are used as inferred information to the marker prediction framework, recovering the occluded marker positions to maintain a continuous data-flow.

2 Literature review

Over the last few decades, many approaches have been presented for tracking and gesture configuring of the hand model. A great overview is given in Wheatland et al. [75]

work, where the authors present a state-of-the-art in hand and finger modeling and animation. The main research directions and industry applications are also illustrated, highlighting the advantages and disadvantages of each methodology.

In general, hand reconstruction and animation methods can be classified into two major classes: *glove based* and *vision based*. Glove-based methods are usually real time, they are, however, expensive and detect only a limited set of finger movements (e.g., P5 Data glove). The vision-based methods, on the other hand, are more accurate, but they have problems with occlusions, noise and spurious data. The following paragraphs briefs the most recent and popular technologies and methods for hand pose reconstruction.

Bend sensor gloves Bend sensor is a glove-based method that allows gesture reconstruction and can efficiently measure hand and finger joint angles in real time. There are many different types of data gloves available; for instance, CyberGlove [12] systems used sensors that convert the angles into voltages. Sensored gloves are favored for use in large spaces and outdoors, and since they do not face occlusions problems, they are popular for hand interaction applications. The main disadvantage of this technology is that they suffer from sensor cross-coupling, resulting in low joint angle accuracy [23]. Furthermore, the gloves usually have fewer sensors attached than hand DOFs, limiting the precision of the articulated hand reconstruction. Lin et al. [40] proposed a learning approach to model constraints in the hand configuration space using motion data collected by a CyberGlove. Typically, the hand posture is estimated by searching in high-dimensional configuration spaces; in contrary, the authors incorporate constraints to define a lower-dimensional subspace, thus generating more natural and feasible hand movements.

Marker-based motion capture Marker-based motion capture technology, such as PhaseSpace [53] and Vicon [72], offers high positional accuracy; it is commonly used for full-body motion and performance capture. These systems usually require a large number of markers to be attached on each limb segment to accurately reconstruct the motion. However, fingers are small and it is not possible to attach many markers for capturing the full articulation of the fingers [27]. In addition, missing data due to markers' occlusions by other elements add a substantial amount of post-processing time to clean the data. Lien and Huang [39] proposed a hand model together with a closed form IK solution for the finger fitting process. The 3D positions were obtained using color markers and stereo vision, and the finger poses were chosen using a search method. Park and Yoon, [52], also used a marker-based mocap system; a LED glove has been employed to produce interactions in multi-modal display environments, while the gestures were recognized and classified using hidden Markov chains. Aristidou and Lasenby [4] used optical mocap data and a reduced set of eight markers to track hand movements; however, each kinematic chain of their hand

model was structured and treated independently, eliminating the control of the movement. In addition, hand physiological constraints were not taken into consideration, resulting sometimes in unfeasible poses. Pollard and Zordan [54] presented a method that allows control for physically based grasping using mocap data; the motion can be further adjusted to external changes using Liu's optimization [42]. Recently, Schröder et al. [58] applied IK optimization in a subspace learned from prior hand movements, using a reduced marker set. The authors evaluated various reduced marker layouts to find the best configuration for capturing the hand articulation. Using Forward and Inverse Kinematics techniques, Hoyet et al. [21] has previously show that finger motions reconstructed using a set of eight markers per hand are perceived to be similar to the corresponding motions reconstructed using a full set of twenty markers.

Vision-based methods A cheaper alternative to the marker-based methods, that provides a more natural interaction experience to the user, are the vision-based methods. The vision-based methods can be separated into two major categories: the methods that use a monocular video camera to recover the hand posture, and the methods that utilize infrared and/or RGB-Depth (RGB-D) cameras. Both vision-based methods can be further divided into two classes, the methods that estimate the hand pose via *template-matching* [59,73], which reconstruct the hand pose from a single frame through classification or regression techniques, and the methods that utilize a *model-based search* [30,38,50], where the posture is approximated by projecting a 3D articulated hand model and aligning the projection with the observed image.

Monocular video cameras Monocular video cameras have been used for tracking and reconstruction of articulated motion due to its easy and quick setup. In this manner, Cerveri et al. [10] utilized a kinematic model using a multi-camera system and markers, which consists of a hierarchical chain and rigid body segments. Limitations relative to the joint rotational and orientational constraints were taken into consideration to restrict the motion to natural postures. In [17] and [11], vision-based hand shape estimation methods were introduced using shape features acquiring from camera images. The extracted features were then used to approximate the hand's state, and the local state of each finger was estimated using IK and physical hand constraints. On the contrary, Kaimakis and Lasenby [24] used a set of pre-calibrated cameras to extract the hand's silhouette as a visual cue. The 2D silhouette data is then modeled as a conic field and physiological constraints are imposed to improve the reliability of the hand tracking [25].

Dewaele et al. [13] implemented a bare-hand tracking approach using edge data detection and silhouettes to identify the pose of the hand. Sudderth et al. [67] used a nonparametric belief propagation (NBP) algorithm for tracking the hand poses and kinematic constraints; in that way, they were able

to handle cases with self-interactions and self-collisions. De la Gorce et al. [29] proposed a 3D hand tracking approach from monocular video. A formulation for exploiting both shading and texture is presented, which is able to handle the problem of self-occlusions and time-varying illumination. However, due to the high dimensionality of the human hand, these methods have high computational cost, making it difficult to use them for real-time interactions in control applications. In contrast [34,57], achieve interactive speeds using bare-hand tracking systems at the cost of resolution and scope. More recently, Feng et al. [15] proposed a behavior-driven freehand tracking based on correlations among Local Motion Models and cognitive information, achieving real-time results with satisfactory reconstruction accuracy.

Wang and Popovic [73] and Fredriksson et al. [16] proposed methods for hand tracking using a single camera and an ordinary cloth glove that was imprinted with a custom pattern; the pose corresponding to the nearest database match was then retrieved. Furthermore, Wang and Neff [74] used data gloves, introducing a method that provides hand shape and fingertip reconstruction using a Linear Mean Composite Gaussian Process Regression (LMC-GPR) model. Although these methods offer a simple, computationally cheap and promising solution, a large database is required to correlate the resulting pose. Guan et al. [19] proposed an Isometric Self-Organizing Map for nonlinear dimensionality reduction, so as to organize the high-dimensional feasible hand pose space in a low dimension lattice structure. Taking into account that generally not all the required information is available, these methods suffer from erroneous pose predictions and oscillations. In addition, since the hand pose is obtained by search and template matching, the reconstructed hand postures are highly correlated with the training data, making these methods less reliable than optical mocap systems.

Infrared cameras and RGB-D cameras: Inferred and RGB-D cameras have been used for efficient and real-time human pose reconstruction, aiming to track the complex human motion for interactive applications [62]. Based on this technology, many companies manufactured products for real-time hand motion recovery and interaction. Nimble VR [49], introduces Nimble sense, a tiny depth sensor that is attached on virtual reality head-mounted display, that is specifically designed to capture the hand movements. Leap Motion [33] has produced a computer hardware sensor device for hand and finger motions control. Using two monochromatic infrared (IR) cameras and three infrared LEDs, the device synthesize 3D position data by comparing the 2D frames generated by the two cameras.

Many researchers over the last decade use RGB-D cameras to track a fully articulated hand [43,55,61,68,70]; the first attempts were almost a decade ago [47], where a real-time pose recognition was achieved, but due to the low-

resolution of the depth images at the time, the reconstruction accuracy was low. The resolution and quality of RGB-D cameras were later improved, allowing a more precise tracking and reconstruction. For instance, Krejov and Bowden [28], using RGB-D for 3D detection and tracking of fingertips, were capable of processing four hands simultaneously. Many papers in the literature use the RGB-D technology to fit a model in continuous pose space, achieving better accuracy compared to template matching [26,45,50,51]. In that way, Liang et al. [38] proposed a spatiotemporal feature, which enforces both spatial and temporal constraints in a unified framework for hand parsing and fingertip detection. Others, [18,76], utilized a regression forest (RF) to produce a number of votes for the hand pose for each individual input pixel; the votes are collected for all pixels, and are fused for the final estimation. RF proves to be fast, accurate, robust to partial occlusion, and very effective for articulated pose estimation. Later, Liang et al. [36,37] took advantage of the hand joint rigidity (bone lengths are fixed) and that the finger motion is highly constrained, proposing a multimodal prediction fusion, learning the joint correlations using PCA analysis in the training data. Finally, they seek the optimal joint parameters in this reduced dimensional space during testing; the depth image is parsed by per-pixel classification using a pre-trained classifier, aiming to obtain the hand parts. Zhao et al. [78] combine marker position data recorded using a mocap system with RGB-D information, to acquire high-fidelity hand motion; the inferred RGB-D information complements the marker-based mocap data when markers are occluded. An optimization technique is then introduced to estimate the pose that best matches with the observed data. Data-driven methods are also used to estimate the finger motion that satisfies spatiotemporal correlations with motion segments [22,44,77]. However, generating model features online is time-consuming, while searching for a matched pose of an input image to the high-dimensional space of the hand requires high computational cost. In addition, due to the high articulation of the hand, the fingers encounter self-occlusions in the projected image, making the full 3D reconstruction of the pose challenging.

Several papers [46,56,65,66] focus on statistical methods, such as an Unscented Kalman Filter and a Hierarchical Bayesian Filter, to track the hand motion. These statistical methods approximate the posterior by a single Gaussian and update these approximations via a liberalization of the measurement process. Shan et al. [60] employed a mean shift embedded particle filter for visual tracking; they incorporate the mean shift optimization into particle filtering to move the particles to local peaks in the likelihood. However, such methods are still far from real-time target, therefore limiting their use. More recently, Sridhar et al. [64] use a detection-guided optimization strategy to increase the robustness of the hand pose estimation; the pixels of the hand model shape are

classified as parts of the hand using random forests, while later are merged with a Gaussian mixture representation to estimate the pose that best fits the depth.

The human hand, as demonstrated in anatomy and anthropometry, is very intricate mechanical unit, consisted of many interrelated parts that cooperate to carry out a specific action. Many researchers in the past studied and incorporated physics-based models to imitate the hand movements based on muscle and skin movements [2,9,63]. Thereby, Albrecht et al. [1] employed a physically based muscle model to animate hands, including elastic skin and bones; in this way, the authors manage to produce anatomically correct poses. The authors also proposed a hybrid muscle model that comprises pseudo muscles for better control of the rotation of the bones based on anatomical and mechanical laws.

This paper introduces an efficient framework that does not employ any machine learning or other data-driven approaches, but carefully models the physical and anatomical constraints of the human hand. A reduced number of markers is attached on the hand to enable real-time tracking of the finger and palm positioning, while the hand articulation is reconstructed by utilizing an IK solver that incorporates physiologically based constraints. The principle of these constraints has been adopted from [25], where hand's physiology rules are used as integrated parameters of a Bayesian framework so as to establish a measure of validity and obtain the most natural pose that best conforms the silhouette data. In contrast, we use data captured using a marker-based mocap system, while the physiological constraints are employed within an IK framework to restrict the allowed pose only within a natural set. The use of marker-based mocap technology enables data acquisition in large capturing volumes, whereas the reconstructed hand can be utilized to attain holistic performance capturing. Finally, the problem of missing entries that most of the marker-based methods encounters has been solved using filtering and inferred information from neighboring visible markers.

3 The hand model

Human motion is typically represented as a series of different configurations of a rigid multibody mechanism consisting of a set of segments connected by joints. These joints are hierarchically ordered and have one or more DoFs. The DoFs describe the rotations relative to their parent joints up to the root joint, for which the position and orientation are represented with respect to a reference coordinate system. In order to construct the position and orientation of the hand, and re-establish its motion, optical motion capture systems use a number of markers attached over the body of the performer. It is important for these markers be located at strategic positions on the hand as they are more easily specified by an

animator and tracked by mocap systems. The placement of markers on the hand is very important, otherwise the system is vulnerable to false predictions; in addition, the accuracy of the system is sensitive to skin movements. Thus, this section presents the proposed hand configuration and layout that can be used for real-time hand reconstruction using a small number of markers attached.

The reconstructed hand pose \mathcal{H}^* can be expressed as a function of the positions of the markers that can be used to compute the positions and orientations of the hand's joints. The reconstruction process consist of two optimization processes executed simultaneously.

$$\mathcal{H}^* = \mathcal{K} + \mathcal{G} \quad (1)$$

where \mathcal{K} applies *Inverse Kinematics* to estimate the remaining joint positions, and \mathcal{G} is concerned with *physiological constraints*, aiming to keep the reconstructed hand pose within a feasible set. In the following we elaborate on each of these two processes.

3.1 Mathematical Background

The mathematical background used for hand reconstruction and animation is based on Geometric Algebra (GA) [20], which provides a convenient mathematical notation for representing orientations and rotations of objects in three dimensions. The conformal model of GA (CGA) is a mathematical framework that offers a compact and geometrically intuitive formulation of algorithms and an easy and immediate computation of rotors; it extends the usefulness of the 3D GA by expanding the class of rotors to include translations, dilations and inversions. Rotors are simpler to manipulate than Euler angles; they are more numerically stable and more efficient than rotation matrices, avoiding the problem of gimbal lock.¹ CGA also simplifies the mathematical model since basic entities, such as spheres, lines, planes and circles, are simply represented by algebraic objects. Thus, CGA gives us the ability to describe algorithms in a geometrically intuitive and compact manner, making it suitable for applications in engineering, computer vision and robotics. More detailed treatment of GA can be found in [14].

3.2 The hand geometry

The first step toward an efficient and precise reconstruction of the hand motion is the structure definition of the hand layout; for implementation purposes, it is assumed that the hand geometry, meaning the initial joint configuration of the

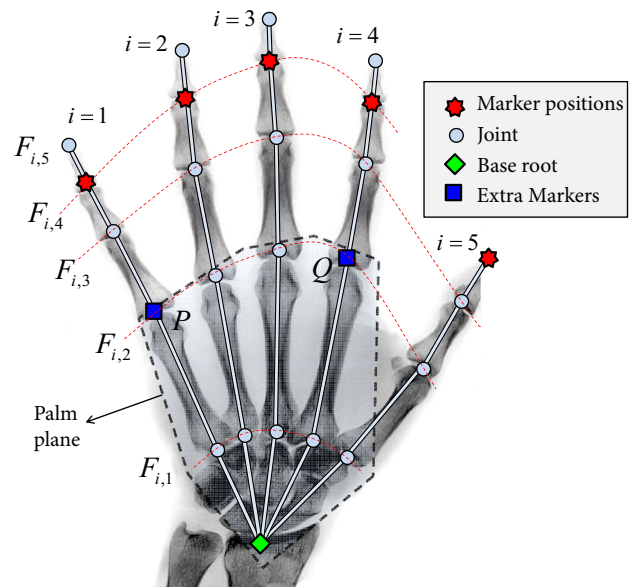


Fig. 1 The hand's model geometry used in this work

hand, is known a priori. An example of a hand model is illustrated in Fig. 1. The proposed hand model consists of 25 joints and has in total 31 DoFs (25 DoFs for the hand and 6DoFs for the wrist). Some others in the literature used different skeletal models with more or less DoFs [41,71,79]. The marker positions, which in this work are used as motion controllers, are captured using an optical motion capture system, such as the PhaseSpace Impulse X2 system, while the markers are labeled (e.g. in the PhaseSpace system, each LED marker is pulsed at a different frequency) so that it is known a priori on which finger each marker is placed. Markers are placed on the forth joint ($F_{i,4}$, for $i = 1, \dots, 4$), that is the joint connecting the distal phalanx and the middle phalanx. The reason we chose not to place the markers on the last joint is because is more likely to have markers occlusions when the hand closes (e.g., to form a fist). The marker, for the case of the thumb, is placed on the last joint ($F_{5,4}$). The orientation of the hand is also important so as to efficiently reconstruct the hand. This can be achieved by attaching 2 extra markers at specific positions, p and q , on the back of the hand (reverse palm). Assuming that the palm is always flat, we can find the plane describing the orientation of the hand using p , q and the position of the base root (r , which also lies on the palm plane). For simplicity, markers p and q can be placed at the joint positions $F_{1,2}$ and $F_{4,2}$, respectively, as shown in Fig. 1.

3.3 Inverse kinematics \mathcal{K}

The first process \mathcal{K} aims at estimating the hand joints, at each frame, using Inverse Kinematics and the inferred marker positions. However, before employing the IK solver, it is

¹ Gimbal lock is a common problem associated with Euler angles and occurs because two axes become aligned during rotational operations, producing unexpected behavior.

important to find the fingers' orientations, the chain roots and the end effectors for each chain; end effectors in this case are assumed to be the joints with the markers' attached. The target positions are considered to be known since they are tracked by the motion capture system. The procedure is simple. Firstly, we estimate the hand orientation; thereafter, we calculate the palm joints and the finger orientations at each time step. When each finger orientation is known, the finger joints at the previous time step are translated and rotated in such a way that all joints belong to the current finger plane. Finally, FABRIK [5], a simple IK solver, is incorporated to fit the joints of each finger. FABRIK is chosen due to its efficiency, simplicity and low computational cost.

The process initialized by calculating the hand orientation; hence, by accepting that the hand plane Φ_x is similar to the palm plane and that the markers p, q and r are lying on that plane, the hand orientation, meaning the plane Φ_x , can be estimated. Therefore,

$$\begin{aligned} P &= \frac{1}{2} (p^2n + 2p + -\tilde{n}) \\ Q &= \frac{1}{2} (q^2n + 2q + -\tilde{n}) \\ R &= \frac{1}{2} (r^2n + 2r + -\tilde{n}) \end{aligned} \tag{2}$$

where P, Q , and R are the 5D null vectors representing points p, q and r respectively, and n and \tilde{n} are the null vectors in CGA. The plane Φ_x is equal to

$$\Phi_x = P \wedge Q \wedge R \wedge n \tag{3}$$

where \wedge is the outer product.

Calculating the palm joints The next step is to incorporate constraints to obtain other palm joints. Thus, by assuming that the inter-joint distances (for the joints $F_{i,1}$ where $i = 1, \dots, 5$ and $F_{j,2}$ where $j = 1, \dots, 4$) are fixed over time and that all these joints lie on the palm plane, we can easily locate them using basic geometric entities such as planes, circles and spheres. An example of palm constraints is given in Fig. 2. As an example, the joint position we are looking for can be estimated by intersecting the spheres with centers being the marker positions p and q and radii being the distance between the marker and that joint position (taken from the model). Therefore, find the sphere with its center at the marker position P and radius equal to the distance between the marker P and the joint we are working on

$$\Sigma_p = \left(P - \frac{1}{2} \rho_1^2 n \right) I \tag{4}$$

where ρ is the sphere radii. Similarly, find the sphere with center at the marker position Q and radius equal to the distance between the marker Q and the joint we are working on

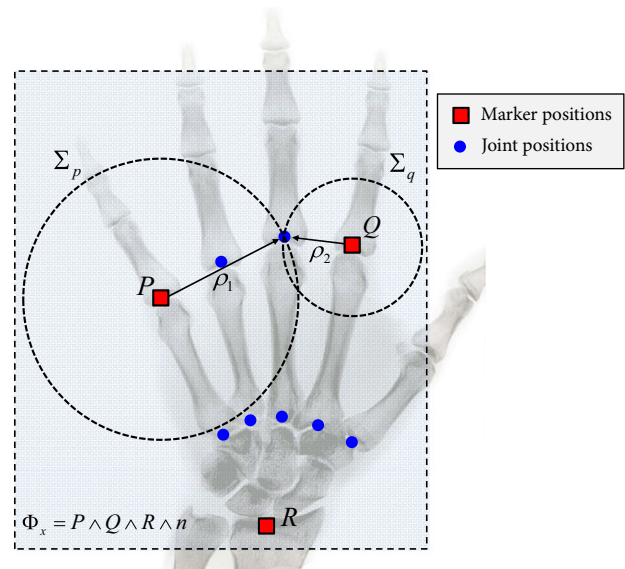


Fig. 2 The palm plane constraints: the hand plane Φ_x is calculated using the marker positions P, Q and R , accepting that the markers lie on Φ_x and that the hand and palm planes are similar. The rest of the palm joints are computed, assuming that their inter-joint distances remain constant, by intersecting the spheres Σ_p and Σ_q with centers at the positions P and Q and radii of the distance between their center and the joint position we are looking for

$$\Sigma_q = \left(Q - \frac{1}{2} \rho_2^2 n \right) I \tag{5}$$

The intersection of the two spheres gives a circle or a single point or no intersection. Thus, the meet between the two spheres equals to,

$$C = \Sigma_p \vee \Sigma_q \tag{6}$$

- If $C^2 > 0$, then C is a circle. In that case, the possible solutions are given by intersecting the circle C and the palm plane Φ_x

$$B = C \vee \Phi_x \tag{7}$$

- If $B^2 > 0$, the meet between C and Φ_x gives two points which can be extracted via projectors, as described in [31]. The new joint position is assigned as the point that is closer to the previous joint position (at time $k - 1$).
- If $B^2 = 0$, the intersection is a single point $X = BnB$.
- If $B^2 < 0$, the intersection does not exist. For that instance, the new joint position is then taken as the nearest point on circle, C , from the previous joint position (at time $k - 1$).
- If $C^2 = 0$, the intersection is a single point $X = CnC$.

- if $C^2 < 0$, the two spheres do not intersect. In that case, the final joint position is given by averaging the distance between the two markers $x = (p + q)/2$.

Calculating the finger joints To estimate the finger joints, we need to find the finger planes Φ_i , for $i = 1, \dots, 4$. Each Φ_i can be calculated using the known joint positions $F_{i,2}$, the marker positions $F_{i,4}$ and by assuming that they are perpendicular to the palm plane Φ_x (note that this does not hold for the thumb plane Φ_5). Since both points from each finger are known (the motion capture system tracks the joint positions $F_{i,4}$ and the finger roots $F_{i,2}$ lie on the palm plane with constant distance from the attached markers p and q , as explained in previous paragraphs), each finger plane can be estimated at the current time frame. The vector that is perpendicular to the hand plane Φ_x is given by

$$\hat{n} = \Phi_x^* - \frac{1}{2} (\Phi_x^* \cdot \vec{n}) n \tag{8}$$

as explained in [31]. The finger planes can then be calculated as

$$\Phi_i = F_{i,2} \wedge F_{i,4} \wedge \hat{n} \wedge n \quad \text{for } i = 1, \dots, 4 \tag{9}$$

The thumb orientation Φ_5 can be estimated using the marker position $F_{5,4}$, and the joint positions $F_{1,2}$ and $F_{5,2}$ that lie on the palm, assuming that when the thumb bends to the ventral side of the palm, it always points at the joint $F_{1,2}$ (approximately true in practice).

The next step is to estimate the rotation between the previous and the current frame of each finger plane. This can be done using rotors; the rotor R which expresses the rotation between the plane in the previous frame and the plane in the current frame, for each finger, can be found using the closed form expression given in [32]. Then each finger joint at time $k - 1$ is translated and rotated in such a way that all joints of a given finger lie on the plane of the current frame k , as demonstrated in Figs. 3 and 4. Hence,

$$\hat{F}_{i,j}^k = R F_{i,j}^{k-1} \tilde{R} \tag{10}$$

where $i = 1, \dots, 4$ and $j = 3, 4, 5$ (except for the thumb where $i = 5$ and $j = 2, 3, 4$).

All joints now lie on plane Φ_i^k . Lastly, the FABRIK Inverse kinematic solver is applied to each finger chain, assuming that the root of the chain is $F_{i,2}^k$, the end effector is the rotated point $\hat{F}_{i,4}^k$ and the target is the current marker position $F_{i,4}^k$, as shown in Fig. 4. The inter-joint distances are constant over time, thus, for computational efficiency, they can be calculated and stored at the first frame.

The resulting posture can be further improved in accuracy and naturalness by incorporating constraints subject to

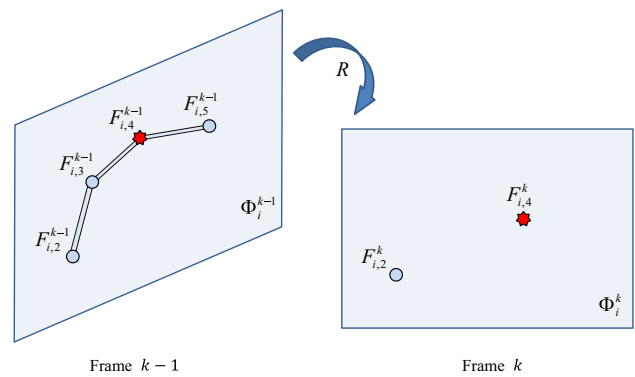


Fig. 3 The joints' positions at times $k - 1$ and k . Each finger joint at time $k - 1$ needs to be rotated by R in such a way that all joints of that finger lie on the plane of the current frame k

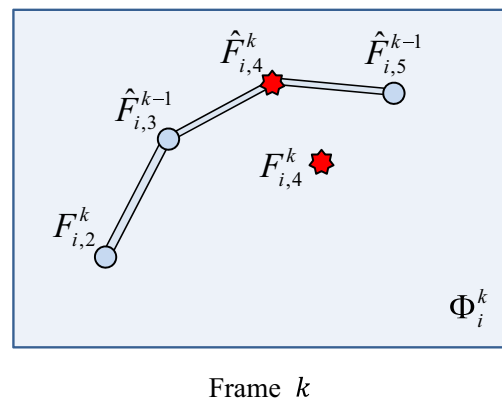


Fig. 4 The current joint positions, after rotating them to lie on the current finger plane Φ_i^k . The problem of orientation is, therefore, solved and FABRIK can then be utilized assuming that the root of the chain is $F_{i,2}^k$, the end effector is the point $\hat{F}_{i,4}^k$ and the target is the current marker position $F_{i,4}^k$

the physiological model of the hand, taking into account the hand, fingers, muscle, skin and individual joint properties.

3.4 Physiological constraints \mathcal{G}

Even though \mathcal{K} consists of soft constrains, a natural hand pose cannot be ensured since the physiological and anatomical restrictions of the hand are not guaranteed. To form a natural skeleton, we define \mathcal{G} that takes into consideration the physiological constraints of the hand movement [25], that is divided into six categories. The constraints are applied sequentially, in the order presented here.

Inertia The first physiological constraint is *inertia*, a limitation correlated with the dynamics of the articulated structure. For implementation, it is assumed that all moving parts of the hand skeleton have similar velocity and acceleration at different time periods. Obviously, the kinematic movement can be divided into two classes, the spatial velocity of the hand's root giving the translation of the hand, and

the local angular velocity of each bone. Since markers are tracked using an optical motion capture system, *inertia* constraints are useful when markers are not visible to the cameras due to self-occlusions (e.g. fingers overlay the markers) or occlusions from other elements in the scene. A state-of-the-art marker prediction mechanism is employed, such as [6], that uses a Variable Turn Model within an Unscented Kalman Filter (VTM-UKF); the trajectories of the omitted markers are predicted by assuming that the finger under consideration has similar direction, velocity and acceleration to that of the hand.

Transdigital correlation The fingers also share *transdigital correlations*; in particular, certain ligaments and muscles interact to cause an amount of flexion to be transmitted across neighboring fingers. An exception to the transdigital correlation is the thumb, which moves independently of other fingers since it is directly connected to the trapezium. An

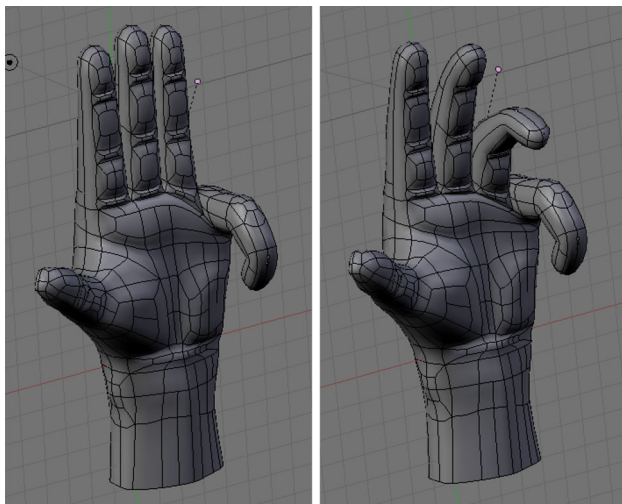


Fig. 5 An example explaining the transdigital correlation of the hand. **a** The finger flexes without affecting its neighboring fingers, breaching the transdigital correlation feature and producing an unnatural posture, **b** a realistic hand pose after taking into account the transdigital correlation between neighboring fingers

example showing violation of the transdigital correlation is given in Fig. 5a where even if both individual rotational and orientational constraints are satisfied, the resulting hand posture remains abnormal. Figure 5b shows the physiologically correct hand configuration, for this specific pose, where the linked pairs of bones share a transdigital correlation movement.

Friction is a constraint associated with the nature of the skin that restricts hand movements. For instance, when frictional forces are applied to a finger, they cause motion that is transmitted to other fingers. A clear example of the friction feature is given during the formation of the fist; the skin causes motion that is transferred from one finger to another when they are in contact.

Again, since each finger is tracked individually by labeled markers, the *transdigital correlation* and *friction* constraints are considered as solved. However, similar to *inertia*, a mechanism to detect violations of these constraints has been integrated to deal with cases of marker occlusion. In this direction, the spatiotemporal correlation between the positions of nearby markers are studied, stating a feasible candidate space for each marker with regards to the position of its neighboring markers. When a missing entry is detected, the VTM-UKF framework returns estimates of the occluded marker positions. Thus, our mechanism checks whether these marker estimates are within a feasible set by inferring information from other visible markers; if a violation is detected, our mechanism forces the estimates to remain within the physiological bounds. This can be achieved by selecting as the new marker position the state of the candidate space with the shortest Euclidean distance from the VTM-UKF marker estimate. Figure 6 indicates the linked pairs of bones which share a correlation movement.

Flexion The design and physical restrictions of the human hand mean that fingers can move to the ventral side, with very limited move in any other direction. The movements that a hand can undergo are, therefore, restricted in terms of *flexion* and *extension*.

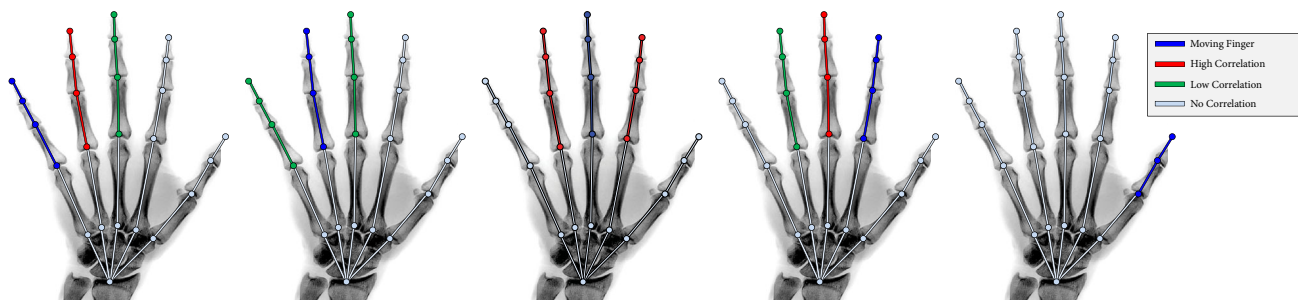


Fig. 6 The linked pairs of bones which share a transdigital correlation movement. **a** The little finger and the affected neighboring fingers, **b**, **c** and **d** the ring, middle and index fingers, respectively, with the effect of their flexion on the neighboring fingers, **e** the thumb. The thumb's movement is independent of the other fingers since it is directly con-

nected to the trapezium. The moving fingers are highlighted in *blue*, the highly correlated fingers in *red*, the fingers having *low* correlation to the moving finger in *green* and finally, the fingers with no correlation are colored in *light gray*

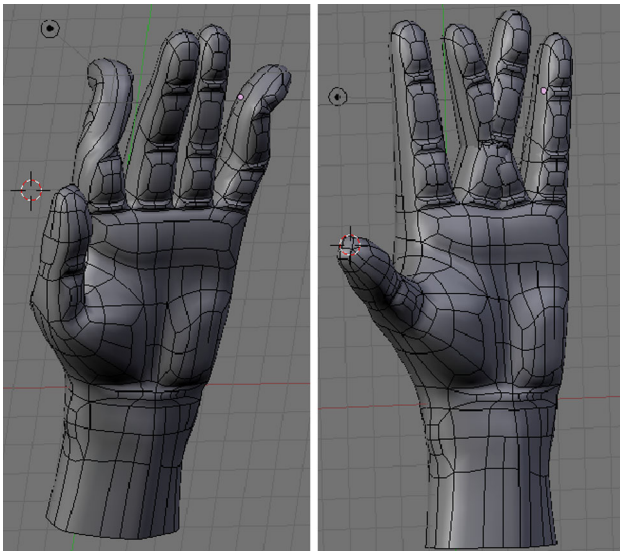


Fig. 7 Examples of unnatural hand postures due to violation of the **a** flexion and abduction, **b** rigidity properties of the finger

Abduction Another family of limitations, caused by hand physiology, are the *abduction* and *adduction* constraints. These constraints control and limit the amount of sideways motion. In this paper, it is assumed that finger orientation is highly correlated with that of the hand palm.

Figure 7a shows an example where the flexion and abduction constraints are not satisfied; the forefinger was erroneously rotated where the little finger was bent in an inappropriate direction and angle.

The hands posture constraints, related to *flexion* and *abduction*, can be incorporated directly into the FABRIK algorithm as rotational and orientational constraints, as described in [3]. In that manner, side rotations on fingers are eliminated, allowing motion with only 1 DoF. Taking into account that fingers do not twist, only rotational constraints are applied, locking the joint orientation to be identical to that of the palm (apart from the thumb).

FABRIK also guarantees that the performed rotation remains within the allowed range bound; the main idea is the re-positioning and re-orientation of the target to be within

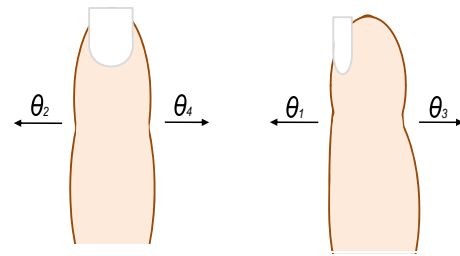


Fig. 8 Graphical representation of the angles $\theta_1, \dots, \theta_4$ which define the rotational constraints of each joint

the valid limits and to satisfy the model constraints. This can be accomplished by checking whether the target is within the valid bounds, at each step of FABRIK, and if it is not, to guarantee that it will be moved accordingly. The allowed range of motion is defined by the angles $\theta_1, \dots, \theta_4$, which represent the minimum and maximum allowed rotation of each joint about the x and y -axes, respectively. Table 1 lists the degrees of freedom for each joint as well as its rotational and orientational limits. Figure 8 presents the angles $\theta_1, \dots, \theta_4$ which define the rotational limits of each joint $F_{i,j}$.

Intradigital correlation Beyond flexion and abduction, several posture restrictions are caused by the muscles of the hand. For instance, the phalangeal flexion in particular fingers is influenced by tendinous synapses with more than one phalanx of that finger [8]. Therefore, it is clear that the muscle contraction and phalangeal flexion are not fully independent, but there is an inter-connection between them. The *intradigital correlation* constraint is responsible for the inter-finger connections caused by certain tendons. Similarly to finger's *transdigital correlation*, we studied the spatiotemporal correlation between the joints of each finger; in this way, parents and children joints are not treated independently. We integrate *pull weights* that links the rotation between nearby joints, distributing the rotation uniformly along the joints. The position of the last joints ($F_{i,5}$ for $i = 1, \dots, 4$) can be estimated by assuming that the rotation between the distal and middle phalanx is approximately 80 % of the rotation formed by the middle and proximal phalanx. *Pull weights* can be applied as follows: check whether the rotation between

Table 1 Hand joint configuration

	DoF	Rotational-x (°)		Rotational-y (°)		Orientational
		θ_1	θ_3	θ_2	θ_4	
$F_{i,1} \quad i = 1, \dots, 4$	1	–	–	–	–	No twist
$F_{i,1} \quad i = 5$	2	20	20	30	40	No twist
$F_{i,2} \quad i = 1, \dots, 4$	2	10	85	15	15	No twist
$F_{i,2} \quad i = 5$	2	5	30	10	10	No twist
$F_{i,3} \quad i = 1, \dots, 5$	1	10	95	–	–	No twist
$F_{i,4} \quad i = 1, \dots, 4$	1	10	90	–	–	No twist
Total	25					

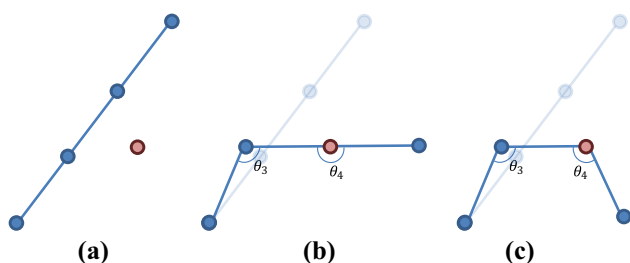


Fig. 9 Incorporating intradigital correlation constraints. **a** The initial configuration of the finger's chain; the joint positions are in *blue color*, the target is in *red*. **b** The FABRIK solution without taking into account the intradigital restrictions; the algorithm checks whether the rotations between joints (in this example θ_4 and θ_3) meet the intradigital restrictions. **c** The last joint is forced to flex so as θ_4 satisfies the intradigital correlation to θ_3

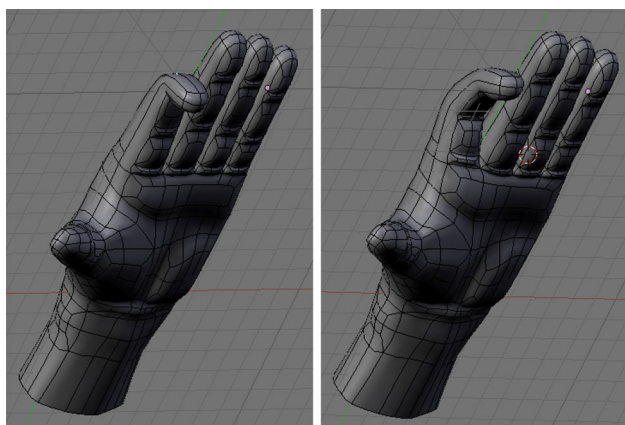


Fig. 10 An example showing the intradigital correlation feature. **a** The intradigital correlation constraint is violated; even if the rotational and orientational constraints are satisfied, the posture of the hand is not natural since it is impossible to bend the distal phalanges without flexing the intermediate and proximal phalanges. **b** the correct posture of the hand when the intradigital correlation of the finger has been taken into account

joints fulfils the intradigital correlation constraints. If they are not satisfied, bend the last joint in such a way that θ_4 satisfies the intradigital correlation to θ_3 . This procedure is illustrated in Fig. 9. Figure 10 shows an example where the intradigital correlation constraint is not satisfied; an unnatural pose of the hand is produced, even if the rotational and orientational constraints for individual joints are satisfied.

4 Results and discussion

Experiments were carried out using an 8 camera PhaseSpace Impulse X2 motion capture system. The implemented methodology was able to process up to 120 frames per second; runtimes were measured on an Intel Core i7 3.5 GHz personal computer. Our dataset comprises marker motion capture data, while data captured using RGB cameras are

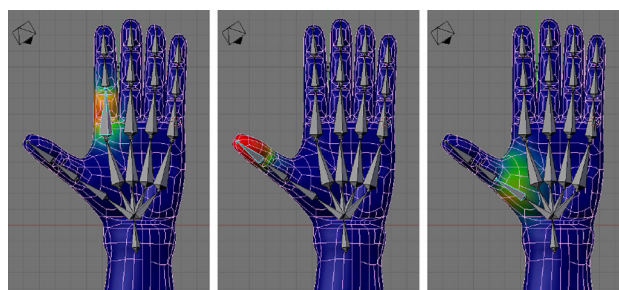


Fig. 11 Example of simple linear blend skinning scheme applied using the weights from bone-heat. The weight assigned to each vertex has been indicated using a gradation from *blue to red* to indicate the range [0, 1]

used to compare the reconstruction quality between the estimated and the true hand postures.

4.1 Mesh deformation

A mesh deformation algorithm is employed to visualize the movements of the underlying hand skeleton in order to compare the resulting animations with the true hand poses. Animating an articulated 3D character requires manual rigging to specify its internal skeletal structure and to define how the input motion deforms its surface. In this paper we used a mesh deformation algorithm driven by animation of a skeleton, named bone-heat [7]. The articulated hand is automatically assigned a per-vertex and per-bone weighting given only by the underlying skeleton. Figure 11 illustrates the mesh deformation algorithm, where the mesh and armature representing the hand are automatically associated using the bone-heat algorithm.

4.2 Experimental results

The proposed method is simple and has low computational cost, meaning it is real-time implementable. It requires 8.3 ms per frame for tracking and fitting 25 joints, hence processing on average 120 frames per second. The rotational and orientational constraints ensure that each finger movement remains natural without showing asymmetries, or irregular bends and rotations. In addition, the physiological constraints restrict the results to anatomically correct postures, subject to the hand, muscle and skin properties.

The implemented system can smoothly track the hand movements, resulting in visually natural motion. The reconstruction quality can be checked visually by comparing the generated 3D hand animations with the data captured using an RGB camera, as demonstrated in the supplementary video and seen in Fig. 12; our system is sufficient, producing postures which are very close to the true hand poses. Figure 13 shows results when the physiological constraints model was disabled (third row), and it is compared against the true



Fig. 12 Four sequences showing examples of hand reconstruction using our methodology at a frame rate of 120 Hz. The first, third, fifth and seventh rows show the true hand posture as recorded using an RGB

camera, while the second, fourth, sixth and eighth show the 3D reconstructed virtual hand pose; it can be clearly observed that the resulting postures are visually natural and biomechanically correct

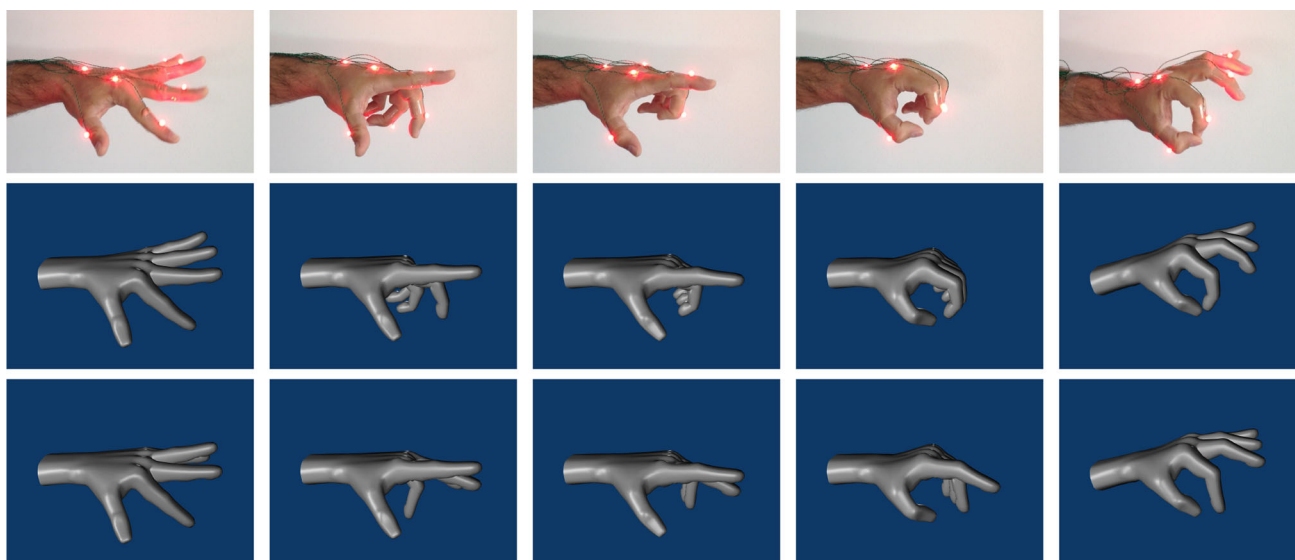


Fig. 13 Applying physiological constraints. The first row show the true hand posture, the second shows the 3D reconstructed hand pose when physiological constraints are applied, and the third row shows the same reconstruction but when the physiological model was disabled. It

can be observed that there is violation of various constraints, especially when markers are occluded, including the inertia, flexion, transdigital and intradigital correlation constraints



Fig. 14 An example of continuous hand pose tracking; in this example the hand flexes to its ventral side to form a fist

hand motion (first row) and our method (second row). More specifically, it can be observed violation of the flexion and abduction constraints (when markers are not visible, the system cannot track the occluded markers and reconstruct the hand posture), the intradigital correlation constraints (the last joints are not forced to bend), and the transdigital correlation constraints (nearby markers do not contribute on estimating the finger pose when markers are unobserved). Another example is illustrated in Fig. 14, where the hand flexes to its ventral side, to form a fist. The advantages of our method are its efficiency and ability to return natural and feasible motion, with low computational cost, while locking the non-permissible joint positions, thus eliminating the poses that do not satisfy the model constraints.

In addition, we experiment how the number of cameras used for the hand movement acquisition affects the performance of our method. By assuming that the hand pose that has been generated using data from 8 cameras is the ground truth, we evaluate the hand reconstruction quality when less cameras track the positions of the markers. We use Lee et al. [35] formula, that is a weighted sum of the difference in rotation between joints, to quantitatively compute the difference

between the two poses and assesses the precision of the reconstruction.

$$Dist^2 = \sum_{k=1}^m \left\| \log \left(q_{jk}^{-1} q_{ik} \right) \right\|^2, \quad (11)$$

where $m = 25$ is the number of joints of our hand model, $q_{ik}, q_{jk} \in \mathbb{S}^3$ are the complex forms of the quaternion for the k th joint for the two under investigation hand skeletons i and j , respectively. The log-norm term $\left\| \log \left(q_{jk}^{-1} q_{ik} \right) \right\|^2$ represents the geodesic norm in quaternion space, which delivers the distance between the joint q_{ik} to q_{jk} on \mathbb{S}^3 for each frame. Table 2a reports the error when information from six, four or two cameras was used; it is important to recall that each PhaseSpace Impulse X2 camera is equipped with two linear detectors, thus two cameras are enough to unambiguously establish the marker's 3d position. The error listed in Table 2 is the averaged, over all frames, summation of the distances between the joints. Results demonstrate that the use of only two or four cameras have long-term occlusions which result in large errors; six is the minimum recommended number of

Table 2 Performance evaluation when (a) a smaller number of cameras was used for the hand motion acquisition, and (b) at different frame rates

Number of cameras	Mean error (per frame)
(a)	
6 (six)	0.07 (max: 0.36)
4 (four)	0.49 (max: 1.12)
2 (two)	0.58 (max: 1.27)
Frame rate (Hz) (8 cameras)	Mean error (per frame)
(b)	
240	0.006 (max: 0.008)
120	0.007 (max: 0.011)
60	0.009 (max: 0.012)
30	0.010 (max: 0.015)

The error is reported against the corresponding hand pose that has been captured using eight cameras at 480 Hz

cameras needed for tracking and capturing hand movements without marker occlusions being problematic.

We also investigate the performance of the system under different capture rates. To evaluate the impact on the performance when capturing at different frame rates, we again assume that the reconstructed hand pose captured at 480Hz is the ground truth. Again, we use Lee et al. formula to measure the difference on joint rotations between hand postures. Looking at Table 2b, it can be observed that generally the reconstruction quality is slightly better when the capture rate becomes higher. This is mainly because the system is more vulnerable when low frame rates are combined with frequent marker occlusions. Nevertheless, even at low capture rates the performance of our methodology remains sufficient, delivering smooth and visually natural results. The time needed for the IK solver to fit the joints to the model also varies for data captured at different frame rates; by reducing the rate, the distance between the target and end effectors is increased, thus more computational time is required for the optimization process (mainly the IK solver) to track the target positions.

Our method, in contrary to others, can track and reconstruct the hand movement using a small number of optical markers, generating postures that are natural and biomechanically feasible, even in cases where data are missing due to self-occlusions or occlusions from other elements in the scene. For instance, Schroder et al. [58] performed an IK optimization in a subspace learned from prior hand movement that allows natural hand recovery using optical motion capture data. However, they do not consider the common phenomenon that markers are occluded or ways to overcome this problem. In contrast, our method introduces a physiological model that can deal with missing markers. Moreover, our configuration suggests to avoid placing the markers in the

edge of the fingers since it is more likely to face self-occlusion problems, especially when fingers bend on the inside of the hand.

4.3 Limitations

The effectiveness of the proposed hand tracking framework is depended on the accuracy of the capturing device. Long-term marker occlusions or marker swapping do affect the tracking performance. Nevertheless, an expensive motion capture system is not a prerequisite for an accurate reconstruction; our method works equally well with any capturing system as far as the 8 key positions are clearly captured and tracked.

In addition, the structuring of the hand relies on a number of assumptions (e.g., the palm is always flat, or the thumb bends to the ventral side of the palm, pointing at the joint $F_{1,2}$) that ultimately affect the precise reconstruction of the hand. Moreover, since markers are placed on the forth joint, that is the joint connecting the distal phalanx and the middle phalanx, there is no information available to infer the positions of the fingertips, except for intra-digital correlations between their joint angles. However, these intra-digital correlations may not be always true (e.g., when pushing fingers on a hard surface). For higher accuracy in the computation of the fingertips joints, a flex sensor (e.g., Tactilus[®] Flex sensors) can be employed to return the precise angle between the distal and middle phalanx.

Finally, the *rigidity* feature of the hand was not investigated in this work since data were captured using a marker-based optical motion capture system and the 3D animated hand did not automatically have a mesh that defines its external shape. As the hand is the most mobile part of the human body, we expect a considerable degree of interaction between fingers, despite the limitations already discussed. Rigidity concerns such interactions, where different fingers may self-intersect, thus causing unnatural postures. While some of the constraints already discussed will limit such self-intersections, there are instances where extra movement restrictions must be applied. This problem is also known as self-collision and has been tackled in several papers, such as [69].

5 Conclusions and future work

This paper presents a system that tracks a human hand of 31 DoFs, relying on a reduced set optical motion capture data; one labeled marker is attached on each finger, treated as end effector, and three more markers are placed at strategic positions on the hand reverse palm to help in identifying the root and orientation of the hand. A physically based hand model is employed, as part of a two step optimization process that involves Inverse Kinematics, to fit the joint positions, subject

to physiological and anatomical restrictions. Without utilizing any machine learning or other data driven approaches, it converts marker positions into a feasible movement of the hand skeleton. Finally, bone-heat, a mesh deformation algorithm, is used to visualize the results for evaluation and comparison.

The proposed methodology produces smooth and natural hand postures, taking into consideration the anatomical and physiological limitations of the hand. The target for real-time implementation is satisfied, processing up to 120 frames per second, and enabling an effective interaction and reconstruction of the hand movement. Even with a low capture frame rate, the proposed methodology can animate the hand smoothly, without oscillations or discontinuities and with high reconstruction quality. Our method is able to operate in rooms with large capturing volume, and it can be used for simultaneous acquisition of the hand with other parts of the human body, attaining an integrated performance capturing.

In future work, a more sophisticated model will be implemented which takes into consideration, in addition to the proposed physiological constraints, muscle dynamics and properties. In this way, bio-potential sensors can be incorporated, such as muscle gesture sensors (e.g., MYO gesture control armbands [48]), which can sense finger and wrist motions and recognize gestures patterns in the muscle triggering.

Acknowledgements This work has received funding from the Office of Naval Research (ONR) Global (N62909-13-1-V090).

References

- Albrecht, I., Haber, J., Seidel, H.P.: Construction and animation of anatomically based human hand models. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '03, pp. 98–109. Eurographics Association (2003)
- An, K., Ueba, Y., Chao, E., Cooney, W., Linscheid, R.: Tendon excursion and moment arm of index finger muscles. *J. Biomech.* **16**(6), 419–425 (1992)
- Aristidou, A., Chrysanthou, Y., Lasenby, J.: Extending FABRIK with model constraints. *Comp. Anim. Virt. Worlds* **27**(1), 3557 (2016)
- Aristidou, A., Lasenby, J.: Motion capture with constrained inverse kinematics for real-time hand tracking. In: Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, ISCCSP '10, pp. 1–5. IEEE, Limassol, Cyprus (2010)
- Aristidou, A., Lasenby, J.: FABRIK: A fast, iterative solver for the inverse kinematics problem. *Graph. Models* **73**(5), 243–260 (2011)
- Aristidou, A., Lasenby, J.: Real-time marker prediction and Co R estimation in optical motion capture. *Visual Comp.* **29**(1), 7–26 (2013)
- Baran, I., Popović, J.: Automatic rigging and animation of 3d characters. *ACM Trans. Graph.* **26**(3) (2007)
- Brüser, P., Gilbert, A.: Finger bone and joint injuries. Taylor & Francis, UK (1999)
- Buchholz, B., Armstrong, T.J.: A kinematic model of the human hand to evaluate its prehensile capabilities. *J. Biomech.* **25**(2), 149–162 (1992)
- Cerveri, P., De Momi, E., Lopomo, N., Baud-Bovy, G., Barros, R.M.L., Ferrigno, G.: Finger kinematic modeling and real-time hand motion estimation. *Ann. Biomed. Eng.* **35**(11), 1989–2002 (2007)
- Chen, W., Fujiki, R., Arita, D., Taniguchi, R.i.: Real-time 3d hand shape estimation based on image feature analysis and inverse kinematics. In: Proceedings of the International Conference on Image Analysis and Processing, ICIAP '07, pp. 247–252. IEEE Computer Society (2007)
- CyberGlove Systems. <http://www.cyberglovesystems.com/> (2016). Accessed: 2016-09
- Dewaele, G., Devernay, F., Horaud, R.P., Forbes, F.: The alignment between 3-d data and articulated shapes with bending surfaces. In: Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, LNCS, vol. III, pp. 578–591. Springer (2006)
- Doran, C., Lasenby, A.: *Geomet. Algeb. Phys.* Cambridge University Press, Cambridge UK (2003)
- Feng, Z., Yang, B., Li, Y., Zheng, Y., Zhao, X., Yin, J., Meng, Q.: Real-time oriented behavior-driven 3d freehand tracking for direct interaction. *Patt. Recogn.* **46**(2), 590–608 (2013)
- Fredriksson, J., Ryen, S.B., Fjeld, M.: Real-time 3d hand-computer interaction: optimization and complexity reduction. In: Proceedings of the Nordic conference on Human-computer interaction, pp. 133–141. ACM (2008)
- Fujiki, R., Arita, D., Taniguchi, R.i.: Real-time 3d hand shape estimation based on inverse kinematics and physical constraints. In: Proceedings of the International Conference on Image Analysis and Processing, ICIAP '05, vol. 3617, pp. 850–858. Springer (2005)
- Girshick, R., Shotton, J., Kohli, P., Criminisi, A., Fitzgibbon, A.: Efficient regression of general-activity human poses from depth images. In: Proceedings of the International Conference on Computer Vision, ICCV '11, pp. 415–422. IEEE Computer Society (2011)
- Guan, H., Feris, R., Turk, M.: The isometric self-organizing map for 3d hand pose estimation. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition, FGR '06, pp. 263–268 (2006)
- Hestens, D., Sobczyk, G.: Clifford algebra to geometric calculus: a unified language for mathematics and physics. D. Reidel, Dordrecht (1984)
- Hoyet, L., Ryall, K., McDonnell, R., O'Sullivan, C.: Sleight of hand: Perception of finger motion from reduced marker sets. In: Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. I3D '12, pp. 79–86. ACM, New York, NY, USA (2012)
- Jörg, S., Hodgins, J., Safonova, A.: Data-driven finger motion synthesis for gesturing characters. *ACM Trans. Graph.* **31**(6), 189:1–189:7 (2012)
- Kahlesz, F., Zachmann, G., Klein, R.: 'visual-fidelity' dataglove calibration. In: Proceedings of the Computer Graphics International, CGI '04, pp. 403–410. IEEE Comp. Soc. (2004)
- Kaimakis, P., Lasenby, J.: Gradient-based hand tracking using silhouette data. In: Proceeding of the 3rd International Symposium on Visual Computing, vol. 1, pp. 24–35. Lake Tahoe, NV/CA, USA (2007)
- Kaimakis, P., Lasenby, J.: Physiological modelling for improved reliability in silhouette-driven gradient-based hand tracking. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition. CVPR '09, pp. 19–26. FL, USA, Miami (2009)
- Kerdvibulvech, C., Saito, H.: Model-based hand tracking by chamber distance and adaptive color learning using particle filter. *J. Image Video Process.* **2009**, 6:2–6:12 (2009)

27. Kitagawa, M., Windsor, B.: *MoCap for artists: workflow and tech. for motion capture*. Focal Press, UK (2008)
28. Krejov, P., Bowden, R.: Multi-touchless: Real-time fingertip detection and tracking using geodesic maxima. In: *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG '13*, pp. 1–7 (2013)
29. de La Gorce, M., Fleet, D.J., Paragios, N.: Model-based 3d hand pose estimation from monocular video. *IEEE Trans. Patt. Anal. Mach. Intel.* **33**(9), 1793–1805 (2011)
30. de La Gorce, M., Paragios, N., Fleet, D.J.: Model-based hand tracking with texture, shading and self-occlusions. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008)
31. Lasenby, A.N., Lasenby, J., Wareham, R.: A covariant approach to geometry using geometric algebra. Tech. Rep. F-INFENG/TR-483, Cambridge University Engineering Department (2004)
32. Lasenby, J., Fitzgerald, W.J., Lasenby, A.N., Doran, C.J.L.: New geometric methods for computer vision: An application to structure and motion estimation. *Int. J. Comp. Vision* **26**(3), 191–213 (1998)
33. Leap Motion Inc. <http://www.leapmotion.com/> (2016). Accessed: 2016-09
34. Lee, C., Ghyme, S., Park, C., Wohn, K.: The control of avatar motion using hand gesture. In: *Proceedings of the Symposium on Virtual Reality Software and Technology. VRST '98*, pp. 59–65. ACM, Taipei, Taiwan (1998)
35. Lee, J., Chai, J., Reitsma, P.S.A., Hodgins, J.K., Pollard, N.S.: Interactive control of avatars animated with human motion data. *ACM Trans. Graph.* **21**(3), 491–500 (2002)
36. Liang, H., Yuan, J., Thalmann, D.: Parsing the hand in depth images. *IEEE Trans. Multi.* **16**(5), 1241–1253 (2014)
37. Liang, H., Yuan, J., Thalmann, D.: Resolving ambiguous hand pose predictions by exploiting part correlations. *IEEE Trans. Circ. Syst. Video Technol.* **99**, 1–14 (2014)
38. Liang, H., Yuan, J., Thalmann, D., Zhang, Z.: Model-based hand pose estimation via spatial-temporal hand parsing and 3d fingertip localization. *Visual Comp.* **29**(6–8), 837–848 (2013)
39. Lien, C.C., Huang, C.L.: Model-based articulated hand motion tracking for gesture recognition. *Image Vision Comp.* **16**(2), 121–134 (1998)
40. Lin, J., Wu, Y., Huang, T.S.: Modeling the constraints of human hand motion. In: *Proceedings of the Workshop on Human Motion. HUMO '00*, pp. 121–127. IEEE Computer Society, Washington, DC, USA (2000)
41. Liu, C.K.: Synthesis of interactive hand manipulation. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '08*, pp. 163–171. Eurographics Association (2008)
42. Liu, C.K.: Dextrous manipulation from a grasping pose. *ACM Trans. Graph.* **28**(3), 59:1–59:6 (2009)
43. Ma, Z., Wu, E.: Real-time and robust hand tracking with a single depth camera. *Visual Comp.* **30**(10), 1133–1144 (2014)
44. Majkowska, A., Zordan, V.B., Faloutsos, P.: Automatic splicing for hand and body animations. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '06*, pp. 309–316. Eurographics Association, Aire-la-Ville, Switzerland (2006)
45. Melax, S., Keselman, L., Orsten, S.: Dynamics based 3d skeletal hand tracking. In: *Proceedings of Graphics Interface, GI '13*, pp. 63–70. Canadian Information Processing Society, Toronto, Ont., Canada (2013)
46. Mikić, I., Trivedi, M., Hunter, E., Cosman, P.: Human body model acquisition and tracking using voxel data. *Int. J. Comp. Vision* **53**(3), 199–223 (2003)
47. Mo, Z., Neumann, U.: Real-time hand pose recognition using low-resolution depth images. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition - Vol. 2, CVPR '06*, pp. 1499–1505. IEEE Computer Society (2006)
48. Thalmic labs: MYO gesture control armbands. <https://www.thalmic.com/myo/> (2016). Accessed: 2016-09
49. Nimble VR. <http://nimblevr.com/> (2016). Accessed: 2016-09
50. Oikonomidis, I., Kyriazis, N., Argyros, A.: Efficient model-based 3d tracking of hand articulations using kinect. In: *Proceedings of the British Machine Vision Conference, BMVA '11*, pp. 101.1–101.11. BMVA Press (2011)
51. Oikonomidis, I., Kyriazis, N., Argyros, A.: Tracking the articulated motion of two strongly interacting hands. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR '12*, pp. 1862–1869 (2012)
52. Park, J., Yoon, Y.L.: LED-glove based interactions in multi-modal displays for teleconferencing. In: *Proceedings of the International Conference on Artificial Reality and Telexistence, ICAT '06*, pp. 395–399 (2006)
53. PhaseSpace Inc.: *Motion Capture Systems*. <http://www.phasespace.com> (2016). Accessed: 2016-09
54. Pollard, N.S., Zordan, V.B.: Physically based grasping control from example. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '05*, pp. 311–318. ACM, New York, NY, USA (2005)
55. Qian, C., Sun, X., Wei, Y., Tang, X., Sun, J.: Realtime and robust hand tracking from depth. In: *Proceedings of the Computer Vision and Pattern Recognition conference, CVPR '14* (2014)
56. Rehg, J., Kanade, T.: *Digiteyes: Vision-based hand tracking for human-computer interaction*. In: *Proceedings of the workshop on Motion of Non-Rigid and Articulated Bodies*, pp. 16–22. IEEE Computer Society Press (1994)
57. Schlattman, M., Klein, R.: Simultaneous 4 gestures 6 dof real-time two-hand tracking without any markers. In: *Proceedings of the Symposium on Virtual Reality Software and Technology. VRST '07*, pp. 39–42. California, USA (2007)
58. Schröder, M., Maycock, J., Botsch, M.: Reduced marker layouts for optical motion capture of hands. In: *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games. MIG '15*, pp. 7–16. ACM, New York, NY, USA (2015)
59. Shakhnarovich, G., Viola, P., Darrell, T.: Fast pose estimation with parameter-sensitive hashing. In: *Proceedings of the International Conference on Computer Vision - Vol. 2, ICCV '03*, pp. 750–758. IEEE Computer Society (2003)
60. Shan, C., Tan, T., Wei, Y.: Real-time hand tracking using a mean shift embedded particle filter. *Patt. Recogn.* **40**(7), 1958–1970 (2007)
61. Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., Freedman, D., Kohli, P., Krupka, E., Fitzgibbon, A., Izadi, S.: Accurate, robust, and flexible real-time hand tracking. In: *Proceedings of the Conference on Human Factors in Computing Systems, CHI '15*, pp. 3633–3642. ACM (2015)
62. Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., Blake, A.: Efficient human pose estimation from single depth images. *Trans. Patt. Anal. Mach. Intell.* **35**(12) (2013)
63. Sibille, L., Teschner, M., Srivastava, S., Latombe, J.C.: Interactive simulation of the human hand. In: Lemke, H., Inamura, K., Doi, K., Vannier, M., Farman, A., Reiber, J. (eds.) *Computer assisted radiology and surgery (CARS 2002)*, pp. 7–12. Springer, Berlin Heidelberg (2002)
64. Sridhar, S., Mueller, F., Oulasvirta, A., Theobalt, C.: Fast and robust hand tracking using detection-guided optimization. In: *Proceedings of Computer Vision and Pattern Recognition, CVPR '15* (2015)
65. Stenger, B., Mendonça, P.R.S., Cipolla, R.: Model-based 3d tracking of an articulated hand. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 310–315. IEEE Computer Society (2001)

66. Stenger, B., Thayananthan, A., Torr, P.H., Cipolla, R.: Model-based hand tracking using a hierarchical bayesian filter. *IEEE Trans. Patt. Anal. Mach. Intell.* **28**(9), 1372–1384 (2006)
67. Sudderth, E.B., Mandel, M.I., Freeman, W.T., Willsky, A.S.: Distributed occlusion reasoning for tracking with nonparametric belief propagation. In: *Neural Information Processing Systems, NIPS '04*, pp. 1369–1376. MIT Press (2004)
68. Tagliasacchi, A., Schröder, M., Tkach, A., Bouaziz, S., Botsch, M., Pauly, M.: Robust articulated-ICP for real-time hand tracking. *Comp. Graph. Forum* **34**(5), 101–114 (2015)
69. Tao, S., Yang, Y.: Collision-free motion planning of a virtual arm based on the FABRIK algorithm, pp. 1–20. *Robotica* (2016)
70. Tompson, J., Stein, M., Lecun, Y., Perlin, K.: Real-time continuous pose recovery of human hands using convolutional networks. *ACM Trans. Graph.* **33**(5), 169:1–169:10 (2014)
71. Tsang, W., Singh, K., Fiume, E.: Helping hand: An anatomically accurate inverse dynamics solution for unconstrained hand motion. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '05*, pp. 319–328. ACM (2005)
72. Vicon Inc.: Motion capture systems. <http://www.vicon.com> (2016). Accessed: 2016-09
73. Wang, R.Y., Popović, J.: Real-time hand-tracking with a color glove. *ACM Trans. Graph.* **28**(3), 63:1–63:8 (2009)
74. Wang, Y., Neff, M.: Data-driven glove calibration for hand motion capture. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '13*, pp. 15–24. ACM (2013)
75. Wheatland, N., Wang, Y., Song, H., Neff, M., Zordan, V., Joerg, S.: State of the art in hand, finger modeling and animation. *Comp. Graph. Forum* **34**(2) (2015)
76. Xu, C., Cheng, L.: Efficient hand pose estimation from a single depth image. In: *IEEE International Conference on Computer Vision, ICCV '13*, pp. 3456–3462. IEEE Computer Society (2013)
77. Ye, Y., Liu, C.K.: Synthesis of detailed hand manipulations using contact sampling. *ACM Trans. Graph.* **31**(4), 41:1–41:10 (2012)
78. Zhao, W., Chai, J., Xu, Y.Q.: Combining marker-based mocap and RGB-D camera for acquiring high-fidelity hand motion data. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '12*, pp. 33–42. Eurographics Association, Aire-la-Ville, Switzerland (2012)
79. Zhao, W., Zhang, J., Min, J., Chai, J.: Robust realtime physics-based motion control for human grasping. *ACM Trans. Graph.* **32**(6), 207:1–207:12 (2013)



Andreas Aristidou is a Post-Doc researcher at the Efi Arazi School of Computer Science, The Interdisciplinary Center Herzliya (Israel), and the Graphics & Virtual Reality Lab, University of Cyprus. He had been a Cambridge European Trust fellow, at the Department of Engineering, University of Cambridge, where he obtained his PhD. Andreas has a BSc in Informatics and Telecommunications from the National and Kapodistrian University of Athens, and he is an honor MSc graduate of Kings

College London. He has been awarded a number of prestigious awards, including the $\Delta\text{I}\Delta\text{AKT}\Omega\text{P}$ fellowship by the Cyprus Research Promotion Foundation and the Office of Naval Research Visiting Award. Andreas collaborates with PhaseSpace Inc. and he has established his own motion capture laboratory at the Graphics & Virtual Reality Lab. His research interests lie on motion analysis and classification, motion synthesis, and involve motion capture and Inverse Kinematics.