Design and Evaluation of Reduced Marker Layouts for Hand Motion Capture

Matthias Schröder[1,2], Thomas Waltemate[1], Jonathan Maycock[2],
Tobias Röhlig[2], Helge Ritter[2], Mario Botsch[1]
[1]Computer Graphics & Geometry Processing Group
[2]Neuroinformatics Group
Bielefeld University
CITEC - Cognitive Interaction Technology
Inspiration 1, 33619 Bielefeld, Germany
Tel. +49 521 106 12107, Fax. +49 521 106 6011
Email: matthias.schroeder@uni-bielefeld.de

Abstract

We present a method for automatically generating reduced marker layouts for marker-based optical motion capture of human hands. The employed motion reconstruction method is based on subspace-constrained inverse kinematics, which allows for the recovery of realistic hand movements even from sparse input data. We additionally present a user-specific hand model calibration procedure that fits an articulated hand model to point cloud data of the user's hand. Our marker layout optimization is sensitive to the kinematic structure and the subspace representations of hand articulations utilized in the reconstruction method, in order to generate sparse marker configurations that are optimal for solving the constrained inverse kinematics problem. We propose specific quality criteria for reduced marker sets that combine numerical stability with geometric feasibility of the resulting layout. These criteria are combined in an objective function that is minimized using a specialized surface-constrained particle swarm optimization scheme, which generates marker layouts bound to the surface of an animated hand model. Our method provides a principled way for determining reduced marker layouts based on subspace representations of hand articulations. We demonstrate the effectiveness of our motion reconstruction and model calibration methods in a thorough evaluation.

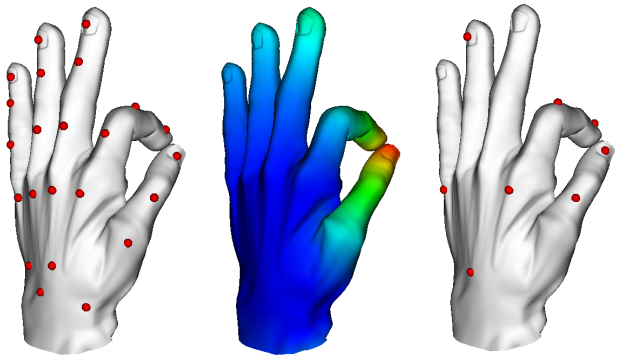

Figure 1: *Our method generates reduced marker layouts for optical motion capture of hands based on analyzing hand movements. Left: full marker set covering all joints of the hand. Center: qualitative illustration of regions that are static (blue) and in motion (red) during the analyzed precision grasp movement. Right: reduced marker set that is sufficient to reconstruct the observed motions using our method.*

Introduction

Marker-based optical motion capture, or mocap, is widely regarded as the standard method for acquiring motions of human performers in both research and industrial or entertainment contexts. Numerous commercial solutions [Vic 2015; Opt 2015; Pha 2015; Qua 2015] and considerable scientific literature exist on the topic. While there is a multitude of alternative solutions for motion tracking, such as markerless methods [Org 2015; Kin 2010] or systems using inertial sensors [Xse 2015; Bio 2015], they are not as widely deployed due to the reliability of marker-based systems. Marker-based optical mocap systems track the 3D positions of markers attached to a performer, which can then be used to infer the articulation of a skeletal model of the tracked subject. Such systems typically consist of 4 to 32 cameras that capture at 30 to 2000 Hz and acquire the marker locations with very high accuracy [Kitagawa and Windsor 2008].

However, despite the quality of marker-based mocap there are drawbacks and limitations to these systems. The captured data usually needs to be post-processed extensively, occlusions can cause gaps or mislabelings in the captured data, and any rotational information needs to be computed retrospectively. Some of these issues are amplified as the number of markers used for tracking increases. A common guideline for capturing articulated objects is to cover all major joints with markers [Guerra-filho 2005; Kitagawa and Windsor 2008]. In addition to making the marker attachment process tedious and error-prone, a high number of markers causes problems when capturing multiple subjects or tracking body movements and hand articulations simultaneously. Capturing hand articulations in detail typically requires a dense marker set consisting of 18–23 markers in a small capture volume. In a large capture volume that also allows for full body mocap the resolution of the optical tracking system and the required size of the markers prohibit the usage of a full marker set. Instead, reduced marker sets have been employed in large capture volumes – however, this strongly limits the expressiveness of the captured hand motions. Therefore, body and hand movements are sometimes captured in isolated sessions and combined in post-processing [Wheatland et al. 2015].

In this work, we present a method for automatically determining reduced marker layouts for inverse kinematics (IK) based motion reconstruction in optical mocap. The motion reconstruction method is based on performing the IK optimization in a subspace learned from prior hand movements, which allows for realistic recovery of hand articulations even from sparse input data. Our method for reduced marker set optimization is sensitive to this reconstruction method, particularly the employed subspace constraints, and thus produces layouts that are optimal for solving the subspace-constrained IK problem. We present an approach that minimizes an objective function that jointly optimizes numerical stability of the marker-IK problem and the geometric feasibility of the resulting layout. The optimization is done using a specialized surface-constrained particle swarm optimization (PSO) [Kennedy and Eberhart 1995; Kennedy and Eberhart 2001], which generates marker

layouts bound to the surface of an animated 3D hand model (see Figure 1).

We show that, rather than specifying one marker per joint of the articulated object, it is sufficient to specify one marker per degree of freedom (DoF) of the parameter space that represents particular hand articulations. Reduced marker layouts can therefore be determined by reducing the parameter space of hand postures based on prior knowledge. Furthermore, we show the principles by which a reduced marker layout that best corresponds to the subspace DoFs can be determined. We demonstrate marker layout results for various hand motions, in particular manual interaction movements based on the grasp taxonomy of [Cutkosky 1989], which distinguishes between different types of power grasps and precision grasps.

This paper is an extended version of the previous conference paper [Schröder et al. 2016]. In addition to the generic hand motion reconstruction and the marker placement optimization proposed in the conference version, this paper provides a detailed evaluation of the reduced marker layouts on real-world hand motion data. To enable these experiments, we extend the hand tracking method of [Schröder et al. 2016] by two aspects: First, we incorporate the automatic marker labeling technique of Maycock et al. [Maycock et al. 2015], which allows us to use the mocap marker data without manual preprocessing. Second, we generate user-calibrated hand models from 3D point clouds of the user's hand, which significantly improves the accuracy of our motion reconstruction. Our experiments clearly demonstrate that the reduced marker layouts can be used to robustly and accurately reconstruct hand motions even from sparse marker input.

Related work

There is a substantial amount of literature on optical motion capture, therefore we focus on the related work that is most relevant to ours, which includes the topics of motion reconstruction based on motion subspace priors, as well as optimized or reduced marker configurations.

Employing subspace representations of human motions has been shown to be effective for motion reconstruction from sparse input. In [Chai and Hodgins 2005; Liu et al. 2006] local linear models were used to represent full-body motions and recover skeletal articulations from sparse marker sets. While these methods are completely data-driven and can therefore limit the space of recovered articulations, our approach uses data-driven subspaces as a prior but also allows for articulation refinements that lie outside of the ground truth database using a layered inverse kinematics approach. Liu et al. [Liu et al. 2006] also target the problem of determining reduced marker configurations by finding a subset of an initial input marker set that can produce accurate predictions of the remaining markers. In contrast, we present a bottom-up approach for generating optimal reduced marker layouts for hands based on the kinematic DoFs of an articulated hand model. While previous methods usually determine reduced marker sets by subsampling a specific initial marker set, our method more generally prescribes properties that candidate marker regions on the surface of a hand model should exhibit, and automatically computes the optimal marker placement within these regions.

Other works deal with the optimal placement of markers, although not necessarily reduced marker layouts. Recently, Loper et al. [Loper et al. 2014] demonstrated an approach that is able to capture fine details of soft tissue deformations in addition to full-body skeletal motions without having to rely on very dense marker sets. To improve the accuracy of their motion and shape capture, they extend their initial sparse marker set in a greedy approach that iteratively adds the next best mesh vertex that minimizes an error metric. We show that, for the problem of finding good reduced hand marker layouts, such greedy approaches are outperformed by our PSO-based global search, as it is less prone to suboptimal local minima. Le et al. [Le et al. 2013] explore the problem of determining optimal marker layouts for facial performance capture using an approach that minimizes the reconstruction error for ground truth sequences of high-resolution facial meshes. While their approach is based on surface deformations of facial meshes, we find reduced marker layouts by purposefully exploiting the kinematic structure and correlations within an articulated hand model.

While a common guideline for marker placement on hands is to use one marker per joint [Guerra-filho 2005; Kitagawa and Windsor 2008], reduced marker layouts for hands have been frequently discussed. In [Kitagawa and Windsor 2008] an example for a reduced "mitten" layout was given, where only one marker was placed at the tip of a single finger. Given an estimation for the global location and orientation of the hand, the relative movement of this marker can be interpreted as the simultaneous bending of all fingers. Our work examines this concept more closely by considering how correlations and redundancies in hand articulations affect marker placement. Regarding the degree of realism of finger motions with reduced marker sets, Hoyet et al. [Hoyet et al. 2012] found that humans are not particularly sensitive to the subtle details of finger animations and the perceived quality of motions is not significantly affected by reduced marker sets. While they manually selected reduced marker configurations, we present an automatic approach based on subspace-constrained inverse kinematics. In contrast, Chang et al. [Chang et al. 2007] determine the most important markers in a reduced marker set for the purpose of grasp motion recognition by using supervised feature selection based on the prediction accuracy of grasp classifiers. In [Kang et al. 2012; Wheatland et al. 2013] a data-driven approach for hand motion reconstruction from sparse marker sets was used, where motions are synthesized by finding database postures that most resemble the low-dimensional input. Wheatland et al. [Wheatland et al. 2013] computed a subset of an initial full marker set by performing principal component analysis (PCA) on the marker trajectories and selecting the most influential ones. Our method differs from theirs in two significant aspects: first, our IK-based approach allows for the recovery of hand articulations that are not present in the prior database, and second, we determine reduced marker layouts in a bottom-up way based on the PCA of joint angles, which explicitly captures the correlations and redundancies present within hand kinematics, unlike positional marker trajectories.

Using PCA or other dimension reduction techniques for hand kinematics has found widespread success in hand tracking, animation and automation [Bernstein 1967; Wu et al. 2001; Kato et al. 2006; Mulatto et al. 2013; Schröder et al. 2014]. To reconstruct the kinematic parameters of an articulated hand model from positional marker data, we follow our previous subspace-constrained inverse kinematics approach [Schröder et al. 2014], where we showed that using subspace constraints the hand posture estimations remain realistic even when input data is missing. In this work, we reverse the problem and seek to find the minimal amount of marker input data necessary to reconstruct postures accurately using subspace priors. As in previous works on reduced marker sets for hand mocap [Chang et al. 2007; Kang et al. 2012; Wheatland et al. 2013; Hoyet et al. 2012], our marker layouts describe only the articulation of the hand, whereas the global position is given by markers placed on the forearm near the wrist.

In order to achieve accurate hand motion reconstructions even for strongly varying hand shapes, we employ a user-specific hand model generated from 3D scanner data. The generation of calibrated hand models has been discussed in a range of previous

works. For instance, Albrecht et al. [Albrecht et al. 2003] and Rhee et al. [Rhee et al. 2006] generate specific hand models by warping a general template model according to features extracted from 2D photographs of hands. For a more detailed geometric reconstruction based on sensor data, depth sensing devices can be used, which provide 3D point cloud information. Reconstruction of 3D geometry from point clouds has been addressed in registration techniques [Li et al. 2008; Li et al. 2009], in which a template model is deformed to fit to the input data by employing non-rigid deformation models in regularized energy optimization frameworks. However, these particular methods do not specifically account for the articulated, kinematic structures of hands. In contrast, Taylor et al. [Taylor et al. 2014] presented a method for generating user-specific hand models by simultaneously adapting a hand triangle mesh and its embedded skeleton to a sequence of depth images. Similarly, Zhu et al. [Zhu et al. 2015] created user-specific anatomically based models of upper and lower limbs by adapting the bone, skin and kinematic structure of an initial template model to depth data. Recently, Tan et al. [Tan et al. 2016] presented a fast method for hand personalization using a small set of depth images that minimizes an energy based on a sum of render-and-compare cost functions.

While recent advances in markerless free-hand tracking using commodity depth sensors [Tagliasacchi et al. 2015; Taylor et al. 2016] have shown increased viability for gesture-based interfaces, such methods can still struggle with capturing more than one subject or human-object interactions. Both of these methods also involve regularization with a statistical hand posture prior, which in the case of [Tagliasacchi et al. 2015] is derived from marker mocap data. Marker-based systems are still favored in scientific or industrial production contexts, where accuracy and reliability are paramount.

In the following we describe the employed motion reconstruction method before discussing the specific quality criteria for reduced marker layouts and presenting our layout optimization scheme. After this, we describe our method of user-specific hand model generation, and finally show and discuss some results of our marker layout optimization and motion reconstruction.

Motion reconstruction

Given a set of target marker positions from an optical mocap system, our motion reconstruction method estimates the hand posture, from which the observed positions originate, by fitting an articulated hand model to the data. Our hand model consists of 16 joints, which are driven by 26 kinematic parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{26})^{\mathrm{T}}$. Of those parameters, 6 describe the global pose of the hand: 3 for translation and 3 for rotation. The remaining 20 parameters describe the posture of the fingers, where each finger defines 4 joint angle parameters. The hand geometry is represented by a triangle mesh, which is animated using linear blend skinning [Jacka et al. 2007]. On the surface of this model, effector positions are defined, which correspond to the marker target positions in the input data. The associations between the target and effector positions are computed automatically by solving the assignment problem between the unlabeled mocap data and the model points, and robustly tracking these correspondences over time. Figure 2 shows the hand model with its underlying skeleton and some exemplary markers. The problem of finding the hand model parameters that move the effector positions to their corresponding targets is solved using inverse kinematics. We apply the subspace-constrained IK method of [Schröder et al. 2014] to the marker-based mocap problem.

Inverse kinematics

The positions of the k effectors on the surface of the hand model are represented as a stacked vector $\mathbf{x} \in \mathbb{R}^{3k}$ and move relative to the model articulation, and can therefore be expressed as a function of the kinematic parameters, $\mathbf{x} = \mathbf{x}(\boldsymbol{\theta})$. These effector positions are subject to move to their corresponding target positions $\mathbf{t} \in \mathbb{R}^{3k}$. The IK problem $\mathbf{t} = \mathbf{x}(\boldsymbol{\theta})$ is solved by finding an update to the kinematic parameter vector $\boldsymbol{\theta}$ that minimizes the objective function

$$E_{\mathrm{IK}}(\Delta\boldsymbol{\theta}) = \frac{1}{2} \left\| \mathbf{x}(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) - \mathbf{t} \right\|^2 + \frac{1}{2} \left\| \mathbf{D}\Delta\boldsymbol{\theta} \right\|^2 . \qquad (1)$$

In this objective function, the first term models the least squares error between the positions of the effector points $\mathbf{x}_i$ and the positions of their corresponding target points $\mathbf{t}_i$. The second term is a selective damping term for the parameter update $\Delta\boldsymbol{\theta}$ with a diagonal matrix $\mathbf{D}$. This damping stabilizes the solution and is used for joint limit avoidance [Schröder et al. 2014].

To find the parameter update $\Delta\boldsymbol{\theta}$, the objective function (1) is minimized with a Gauss-Newton approach, in which a linear system is solved in each iterations. The objective function leads to the linear system

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \mathbf{D} \right) \Delta\boldsymbol{\theta} = \mathbf{J}^{\mathrm{T}} \left(\mathbf{t} - \mathbf{x}(\boldsymbol{\theta}) \right) , \qquad (2)$$

where $\mathbf{J} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}$ is the $(3k \times 26)$ Jacobian matrix of the effector positions [Buss 2004]. After solving the linear system, the resulting update $\Delta\boldsymbol{\theta}$ is scaled using a line search in order to guarantee convergence. The process of solving the linear system (2) and updating the effector positions is iterated 5–10 times.

The result of this process is an update to the kinematic parameter vector $\boldsymbol{\theta}$ that moves the effector positions on the model to the marker target positions in the input data. Given a full marker set that specifies the articulation of every joint this produces accurate reconstructions of the input motion. However, when using reduced marker sets the input data is sparse and the motions of joints that are not constrained by marker positions cannot be recovered. For this reason, a subspace prior that captures the correlations of joint movements is employed in the inverse kinematics scheme.

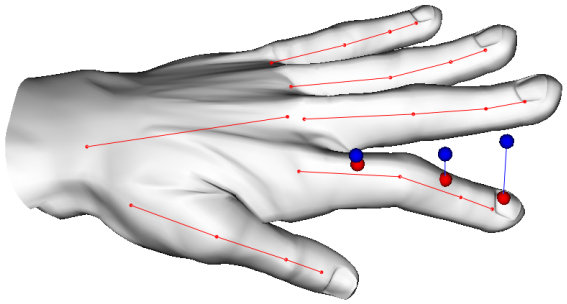

Figure 2: *Hand model and its underlying skeleton. Also shown are three exemplary markers on the hand model (red) that are constrained to move towards their target positions (blue) using inverse kinematics.*

Subspace prior

We use the subspace IK method proposeed in [Schröder et al. 2014], where we obtain a subspace representation of hand articulations from the publically available database of [Schröder et al. 2014], which contains a high variety of human hand motions. Performing PCA on this database of 20-dimensional hand posture data yields a set of eigenvectors and eigenvalues, which can be used to construct a $26 \times (6 + l)$ matrix of principal components $\mathbf{M}$, which maps between the full 20-dimensional posture space and a reduced

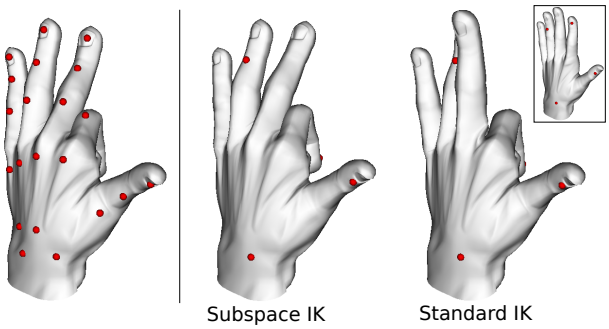

Figure 3: *Full and reduced marker sets and reconstructed hand postures with standard inverse kinematics optimizing for all joint angles and subspace inverse kinematics optimizing for reduced subspace parameters. While the standard approach cannot articulate the markerless fingers, the subspace approach captures the correlations between fingers and articulates them using the reduced marker set.*

l-dimensional subspace. The additional 6 dimensions encode the global pose of the hand, which is not captured in the PCA model. The number of subspace dimensions l determines the amount of variance in the input data covered by the subspace and can be seen as a control variable for the eventual number of markers k employed in a reduced marker layout. It was shown in [Schröder et al. 2014] that in order to represent 90% of given hand movements, 3–6 subspace dimensions are sufficient.

Given the PCA matrix $\mathbf{M}$, the full parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{26}$ can then be computed from the reduced subspace parameters $\boldsymbol{\alpha} \in \mathbb{R}^{6+l}$ as

$$\boldsymbol{\theta} = \mathbf{M}\boldsymbol{\alpha} + \boldsymbol{\mu}, \tag{3}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{26}$ is the mean of the database postures. This makes it possible to represent the forward kinematics of the effector points $\mathbf{x}$ subject to the subspace parameters: $\mathbf{x} = \mathbf{x}(\boldsymbol{\alpha}) = \mathbf{x}(\boldsymbol{\theta}(\boldsymbol{\alpha}))$. Based on this representation, the IK problem can be expressed in terms of the subspace parameters as well. Optimizing for the subspace parameters in (1) and (2) is possible using the subspace Jacobian

$$\mathbf{J}_{\text{PC}} := \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} \cdot \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\alpha}} = \mathbf{J} \cdot \mathbf{M}. \tag{4}$$

Substituting $\mathbf{J}_{\text{PC}}$ for $\mathbf{J}$ in the linear system (2) and analogously changing the damping matrix $\mathbf{D}$ yields the IK solution for the subspace parameters. This solution naturally constrains the reconstructed hand postures to linear combinations of the principal components of the posture database and allows joints to move in correlation to others even when they are not constrained by markers.

However, as there can be variations between the movements contained in the database and the ones observed in the mocap data, we only use this subspace estimate as an initialization for a subsequent refinement of the full posture parameters. By removing the subspace constraints after the initialization of the subspace parameters $\boldsymbol{\alpha}$ and refining the estimate by solving the IK problem again for the full parameter vector $\boldsymbol{\theta}$, the joints with markers are allowed to move more closely to the observed marker positions. This layered IK scheme makes it possible to obtain hand motion reconstructions that are both realistic, due to the subspace initialization, and accurate, due to the full kinematic refinement. Figure 3 shows a comparison of standard IK with the subspace approach we employ.

Fully automatic tracking

Our system tracks the user's hands as well as any objects with markers by fitting models to the captured data. The models are geometric representations of the tracked objects and include preset marker layouts (generated automatically for hands). The input to our tracking system is a sequence of unlabeled mocap data. We determine correspondences between the unlabeled marker point cloud and the tracking models using a fully automatic approach that requires no user intervention [Maycock et al. 2015], which we briefly review in the following.

In the initial frame, the input point cloud is partitioned using Euclidean clustering [Rusu 2009] in order to distinguish between hand and object markers. The tracking models are initialized to these clusters by matching each possible model (cylinder, sphere, hand, etc.) to each cluster in a brute force manner and choosing the configuration with the lowest mean Euclidean distance between the model positions $\mathbf{x}$ and the cluster positions $\mathbf{t}$. The model positions $\mathbf{x}$ are the model's preset marker positions after fitting to the cluster. The point-to-point correspondences within each model-cluster pair are optimized using the Hungarian method [Kuhn 1955; Edmonds and Karp 1972], which solves the assignment problem

$$\min_{\{m_{i,j}\}} \sum_{i=1}^{k} \sum_{j=1}^{k} m_{i,j} \left\| \mathbf{x}_i - \mathbf{t}_j \right\|, \tag{5}$$

where $\{m_{i,j}\}$, $1 \leq i,j \leq k$, denotes a complete assignment of effector positions $\mathbf{x}_i$ to target positions $\mathbf{t}_j$, with $m_{i,j} = 1$ if $\mathbf{x}_i$ is matched to $\mathbf{t}_j$, and 0 otherwise. Equation (5) measures the cost of a bijective assignment $m_{i,j}$ based on Euclidean distances. The Hungarian method optimizes the problem using a square $k \times k$ distance matrix, which is padded with ∞ values if the number of points in $\mathbf{x}$ and $\mathbf{t}$ differ. Based on these assignments $(\mathbf{x}_i \leftrightarrow \mathbf{t}_j)$, the inverse kinematics solution is computed as described above.

The marker assignment optimization is computed not only for the initialization, but also in each frame of the input sequence. This makes our system highly robust against data artifacts such as ghost markers, which are spurious data points that can temporarily pop up due to sensor noise or reflections (see the accompanying video). Another data artifact that our system explicitly handles is markers temporarily disappearing, e.g., due to occlusions. Disappearing markers are detected during the assignment optimization. Reappearing markers are distinguished from ghost markers based on an adaptive distance threshold [Maycock et al. 2015]. As there is no information to animate the respective segments of the hand model between disappearance and reappearance of a marker, our system smoothly interpolates the affected joint angles during the frames in this animation gap. Without interpolation, the sudden reappearance of a missing marker can cause a jump in the movement of the joints affected by this marker. This includes joints that are affected by the marker directly as part of the marker's kinematic chain, or ones that are affected indirectly via the subspace prior and are not constrained by any other markers. Interpolation of all affected joint angles produces a smooth movement for the whole hand.

If a marker is visible in frame t, disappears for T frames, and reappears in frame $t+T+1$, we determine the missing joint angle values $\boldsymbol{\theta}_{t+1}, \ldots, \boldsymbol{\theta}_{t+T}$ for all involved joints by smoothly interpolating the states between time t and $t+T+1$. A simple linear interpolation of the boundary values $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_{t+T+1}$ would lead to discontinuities in the angular velocity $\dot{\boldsymbol{\theta}}(t)$. We avoid this and additionally minimize unnecessary oscillations by finding a joint angle

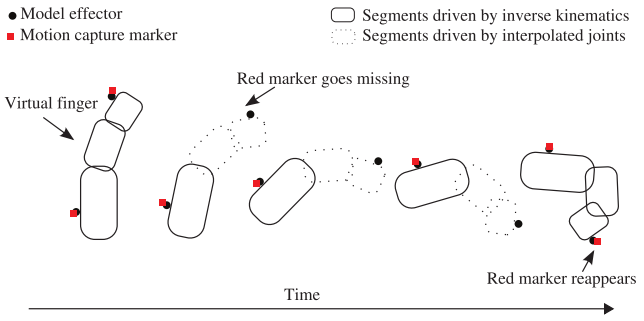

Figure 4: *A virtual finger is shown as it moves through space and is bent. In the second rendering the motion capture marker is no longer visible. Once the marker reappears, inverse kinematics are computed in order to verify that the posture is possible and then the intervening frames in which the marker was missing can be updated using interpolation of the affected joints.*

function $\boldsymbol{\theta}(t)$ that interpolates the C^1 boundary constraints $\boldsymbol{\theta}$ and $\dot{\boldsymbol{\theta}}$ at times t and $t+T+1$ while minimizing angular acceleration:

$$\min_{\boldsymbol{\theta}(t)} \int_t^{t+T+1} \left\| \ddot{\boldsymbol{\theta}}(t) \right\|^2 \mathrm{d}t. \tag{6}$$

Because of the uniform time steps of the tracking system we can safely discretize temporal derivatives by recursive finite differences, such that finding the missing joint angle values for the T time steps leads to a simple $T \times T$ linear system to be solved (for each missing marker individually). Figure 4 shows an example in which a marker disappears while the affected finger is changing its position and posture simultaneously. Our interpolation method automatically produces smooth animations in spite of intermittent noise and missing data, but if markers remain missing for long periods (e.g. due to significant occlusions), any complex motions taking place during that time will be simplified to an interpolation between the states before and after the gap. However, in practice such extreme cases rarely occur using our acquisition setup. For a detailed evaluation and discussion of the fully automatic tracking and gap interpolation we refer the reader to [Maycock et al. 2015].

Reduced marker layouts

Subspace-constrained inverse kinematics makes it possible to fully articulate a hand model based on a sparse set of marker points. However, the choice of marker placement is not arbitrary, and to find the optimal marker layout necessitates a method that can assess the quality of a given layout in relation to others. In the following, we discuss the general considerations taken into account and the specific quality metrics employed in our marker layout optimization.

Given a ground truth trajectory of hand motions from a database, the most straightforward way to evaluate the quality of a given marker set is to compare the ground truth trajectory with one reconstructed using a reduced marker set. The specific metric we consider here is the positional reconstruction error, which measures the deviation of the reconstructed trajectories of the model vertices $\mathcal{V}$ from the ground truth trajectories. While this is an intuitive measurement for the deviations in the results of the motion reconstruction (see, e.g., Figure 12), it is not convenient as a metric for choosing an optimal marker layout. Its computation is prohibitively inefficient and it does not generalize beyond the specific input trajectory. Instead, we use metrics that effectively incorporate the IK

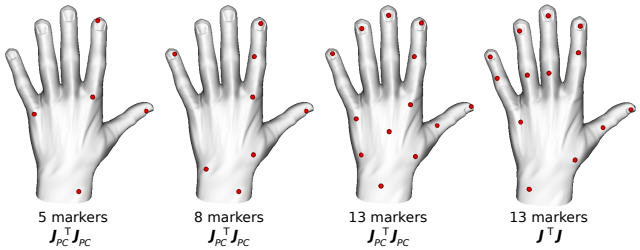

Figure 5: *Marker layouts of different sizes for a precision grasp movement involving the index finger and thumb. The rightmost layout with 13 markers was computed using the full Jacobian* $\mathbf{J}$ *for the condition number metric, whereas the others were computed using the reduced Jacobian* $\mathbf{J}_{PC}$.

problem setup, the subspace DoFs and generic geometric considerations.

A reduced marker set must be configured in such a way that the subspace IK can produce the most accurate results. Additionally the layout must be designed such that it is well suited for practical use, which means that it should be unobtrusive, easy to apply, and should obviate occlusions and self-contact. In the following, we break these requirements down into two categories: numerical stability and geometric feasibility.

Numerical stability

Our IK hand motion reconstruction is based on solving the linear system (2). The numerical stability of the IK problem is measured by the invertibility of the left-hand-side matrix $\left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \mathbf{D}\right)$, the key component of which is the Jacobian $\mathbf{J}$ (or $\mathbf{J}_{\mathrm{PC}}$), which is the derivative of the marker positions with respect to the kinematic (or subspace) parameters. Different marker layouts define different Jacobians, each marker defines three rows in the Jacobian matrix. Therefore we denote the Jacobian matrix produced by a specific marker layout $\mathcal{M}$ as $\mathbf{J}_{\mathcal{M}}$. Each kinematic (or subspace) DoF corresponds to a column in the Jacobian. As we are only interested in the minimal layout necessary for accurate posture estimation (joint angles), we omit the three columns in the Jacobian that correspond to translational DoFs, which means that ${\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}$ is a 23×23 matrix for the full parameter space and a $(3+l) \times (3+l)$ matrix for the reduced parameter space.

A criterion for the invertibility of a matrix is its condition number, which is low when the problem is well-conditioned and high when it is ill-conditioned. As we are interested in the most numerically stable marker layout, we omit the damping matrix $\mathbf{D}$, which is not impacted by the markers, and only regard the condition number of the matrix ${\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}$. We compute the condition number of the matrix ${\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}$ using its singular values as

$$\kappa\left({\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}\right) = \left| \frac{\sigma_{\max}\left({\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}\right)}{\sigma_{\min}\left({\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}\right)} \right|, \tag{7}$$

where $\sigma_{\max}(\mathbf{A})$ and $\sigma_{\min}(\mathbf{A})$ denote the maximum and minumum singular values of matrix $\mathbf{A}$, respectively.

Optimizing the marker layout $\mathcal{M}$ for the condition number $\kappa\left({\mathbf{J}_{\mathcal{M}}}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}\right)$ produces marker layouts whose IK solutions are numerically stable by covering the kinematic DoFs of the hand. Taking into account the subspace prior in the IK system by using the subspace Jacobian (4), the marker positions tend to positions that optimally cover the subspace DoFs. Figure 5 illustrates this concept. Note that the number of markers needed to specify the IK

problem is determined by the number of DoFs representing the posture. The full posture space therefore cannot be used to produce sparse marker sets (less than 8 markers), since the IK problem would be underspecified. Employing a subspace representation facilitates reduced marker sets.

Geometric feasibility

Optimizing only for the condition number of the system matrix produces numerically stable and kinematically meaningful marker layouts, however they can be unsuitable for practical use by placing markers at positions that are obstructive for the mocap performer or are sensitive to occlusions and self-contact. Therefore, we consider geometric feasibility in addition to numerical stability in order to produce well-conditioned marker layouts that are also good in practice. We do this in part by limiting the areas where markers can be placed. While this could be done by manually predefining allowed regions, this would cause the need for user intervention. Instead, we define some generic properties that the model vertices should exhibit to select feasible ones automatically. Additionally, we need to model geometric properties that cannot be accounted for by preselecting vertices, as they change during hand motions (e.g. self-contact).

The first set of geometric feasibility properties is the potential areas for positioning the markers on the surface of the hand model. As the hand naturally bends inwards and can come in contact with objects in the front, markers should generally not be placed on the front side, but rather on the back. Similarly, the markers should be prevented from touching the other fingers during motion and therefore markers should not be placed towards the sides of the fingers. We therefore define feasible regions on the surface of the hand model based on the vertex normals. Only vertex positions $\mathbf{p}_i \in \mathcal{V}$ whose normals $\mathbf{n}_i$ satisfy the condition $\mathbf{n}_i \cdot \mathbf{h} > 0.9$, where $\mathbf{h}$ is the hand model's back-facing vector, are eligible as marker positions.

The second set of geometric feasibility properties taken into account is marker movement. In practice, markers placed near the joint pivot can move non-rigidly along with the joint rotation due to stretching and sliding of the skin. To prevent this, we identify regions on the skinned mesh that move rigidly relative to joints by considering the hand model vertices' convex skinning weights and only using vertices with weight 1 for one joint. Another movement-related issue is when markers can come in contact with each other during motions, which is especially important even with reduced marker layouts when using large markers. To prevent marker contact from occurring, we maximize the minimum distance between markers across multiple keyframes in the input trajectory. For a single frame, the minimum distance between two markers in a marker set $\mathcal{M}$ is

$$\delta(\mathcal{M}) = \min_{\mathbf{a}\in\mathcal{M}} \left\{ \min_{\mathbf{b}\in\mathcal{M}\setminus\{\mathbf{a}\}} \left\{ \|\mathbf{a}-\mathbf{b}\|^2 \right\} \right\}. \tag{8}$$

Maximizing this metric over all frames causes markers to spatially disperse as far from each other as possible, particularly when finger movements cause otherwise spatially distant markers to approach each other more closely.

The combination of these criteria serves as a geometric regularization to the kinematic constraints imposed on the marker set. As a result, the markers are placed in geometrically feasible hand regions during the optimization. The layouts shown in Figure 5 combine the numerical and geometric criteria. In the following, the combination of the discussed metrics and their respective influences are discussed.

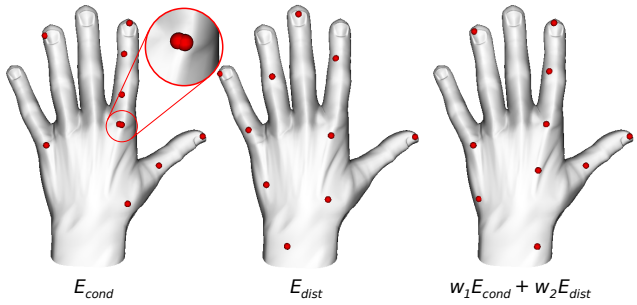

Figure 6: *Example layouts with 10 markers for the objective function terms. The input data is a precision grasp, where mostly the index finger and thumb are in motion. Left: when optimizing only for the numerical stability term E_{cond} markers can be placed in close proximity, which is geometrically impractical. Center: optimizing for the geometric distance term E_{dist} results in spatially distant markers, but the layout does not capture the analyzed hand articulations. Right: a weighted combination of the two terms results in a layout that is both numerically stable and geometrically feasible.*

Layout optimization

We now combine the quality measures for reduced marker layouts in an energy minimization scheme, in which the marker set $\mathcal{M}$ that minimizes an objective function $E(\mathcal{M})$ is found using stochastic optimization. To this end, we employ a specialized surface-constrained particle swarm optimization (PSO) scheme, which confines the solution domain to the vertices $\mathcal{V}$ of an animated hand model. In addition to the vertices, the input to this optimization includes the vertex normals and skinning weights, as well as a training set of example hand motions. The marker set quality properties are evaluated on the model's vertex positions. A distinction can be made between static properties, which are invariant to hand motion and relative marker placements, and dynamic properties, which vary with different motions and marker layouts.

Static aspects of marker layout quality are those that prevent negative effects of skin sliding, by penalizing the vertices' skinning weights, and obstructiveness, by penalizing the vertices' normal angles. These properties can be incorporated by preselecting only vertices that satisfy them. This yields a set of preselected vertices $\mathcal{V}' \subset \mathcal{V}$ on the hand model surface that are eligible as potential marker positions. Ultimately, the optimized marker layout will be a subset $\mathcal{M} \subset \mathcal{V}'$ of this preselection.

In contrast, dynamic aspects of marker layout quality cannot be evaluated as isolated vertex properties, as they vary with changes in hand articulation and placement of the remaining markers within the layout. These include the numerical stability measured by the condition number of the IK system matrix $\mathbf{J}_{\mathcal{M}}{}^{\mathrm{T}}\mathbf{J}_{\mathcal{M}}$ and the minimum marker distance. To account for these changes with respect to different hand articulations, we evaluate and accumulate these metrics over a set $\mathcal{F} = \{f_1, \ldots, f_F\}$ of representative keyframes of a given input hand motion trajectory, which can be automatically computed using farthest point optimization [Schlömer et al. 2011] in the hand posture domain. These dynamic properties of marker set $\mathcal{M}$ are modeled in the objective function $E(\mathcal{M})$, whose definition and optimization are discussed in the following.

Objective function

The objective function that is minimized in the marker layout optimization is a weighted combination of energy terms with respect to marker set $\mathcal{M}$

$$E(\mathcal{M}) = w_1 \cdot E_{\text{cond}}(\mathcal{M}) + w_2 \cdot E_{\text{dist}}(\mathcal{M}), \tag{9}$$

where $E_{\text{cond}}(\mathcal{M})$ penalizes the condition number of the IK system matrix induced by the marker layout Jacobian, and $E_{\text{dist}}(\mathcal{M})$ penalizes the minimum distance between any two marker positions in the layout. Both of these terms are evaluated over a set $\mathcal{F}$ of key-frames in a hand motion trajectory that are representative of the movements that should be captured in the reduced marker set. We denote the marker configuration of layout $\mathcal{M}$ in frame $f \in \mathcal{F}$ as $\mathcal{M}^{(f)}$ and its Jacobian as $\mathbf{J}_{(f)}$.

Based on (7), the energy term penalizing the condition numbers of the induced system matrices is defined as

$$E_{\text{cond}}(\mathcal{M}) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \kappa\Big(\mathbf{J}_{(f)}{}^{\mathrm{T}} \mathbf{J}_{(f)}\Big). \tag{10}$$

This term minimizes the average condition number across all frames $\mathcal{F}$. Since the considered marker layout is a subset of the preselected vertices $\mathcal{M} \subset \mathcal{V}'$, we can precompute the vertex Jacobian $\mathbf{J}_{\mathcal{V}'}$ for all frames $\mathcal{F}$ and construct the respective marker Jacobians by selecting the corresponding rows in this matrix.

Based on (8), the energy term penalizing the minimum distance between two marker positions across all keyframes is defined as

$$E_{\text{dist}}(\mathcal{M}) = -\frac{1}{L} \min_{f \in \mathcal{F}} \Big\{ \delta\Big(\mathcal{M}^{(f)}\Big) \Big\}, \tag{11}$$

where L is the length of the hand model, making the term scale invariant. As we want to maximize the minimum distance between two markers, this term aims to minimize the negative of the overall minimum distance over all frames $\mathcal{F}$.

Combining these two energy terms integrates the desired numerical stability and geometric feasibility properties of the marker layout in a single objective function. The results of minimizing the two energy terms and their weighted sum is illustrated in Figure 6. In this particular example, the condition energy places two markers close to each other, because the linear system for the subspace parameters is overspecified by the number of markers, which means that close-by markers do not corrupt the matrix conditioning. Combining the two energies improves the resulting layout. We use weights $w_1 = 0.1$ and $w_2 = 100$ in all our experiments. In the following, the optimization of the objective function (9) is detailed.

Marker PSO

We find reduced marker layouts by minimizing the objective function (9) using particle swarm optimization (PSO). PSO is a stochastic meta-heuristic for finding global optima of arbitrary objective functions without the need for prior knowledge or assumptions about the optimized problem. The method has recently found widespread application and success in the context of visual hand tracking [Oikonomidis et al. 2011; Qian et al. 2014; Sharp et al. 2015]. Our use of PSO for marker placement optimization aims to overcome the issues of suboptimal local minima often associated with non-global or greedy approaches.

In the PSO method, an optimal solution to a given problem is found by iteratively updating and evaluating candidate solutions, or solution hypotheses. A large set of such hypotheses is managed as a swarm or population of particles, each of which has an associated position $\mathbf{x}_t$ and velocity $\mathbf{v}_t$ in the solution domain of the objective function at iteration t. Each particle keeps track of its local previous best position $\bar{\mathbf{x}}_{\text{par}}$ in the solution domain and the population keeps track of the global optimum $\bar{\mathbf{x}}_{\text{pop}}$ across all particles. In each iteration of the PSO process, the velocity of every particle is updated such that the particle is attracted to the local and global optima in addition to moving along its own inertia. The local and global optima are updated after each particle movement by evaluating the objective function at the new particle position. Finally, the solution of the PSO process is the global optimum achieved after a given number of iterations or after convergence of the optimum value.

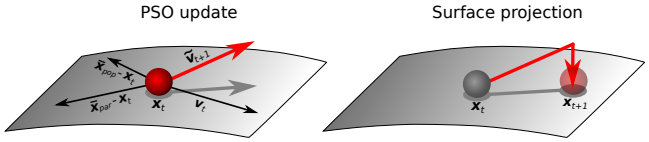

Figure 7: *Illustration of a PSO update for one marker position. First, the new velocity $\widetilde{\mathbf{v}}_{t+1}$ of the marker is computed as a weighted linear combination of the vectors towards the particle's local optimum $\bar{\mathbf{x}}_{par} - \mathbf{x}_t$, the population's global optimum $\bar{\mathbf{x}}_{pop} - \mathbf{x}_t$ and the particle's current velocity vector $\mathbf{v}_t$. This update can send the marker off the surface of the hand model due to the curvature of the model surface. Therefore, in a second step, the new position $\mathbf{x}_{t+1}$ is computed by projecting back onto the surface. The velocity $\mathbf{v}_{t+1}$ is then recomputed accordingly.*

In our application, the solution domain of the objective function is the domain of marker layouts $\mathcal{M}$. To map this to the PSO scheme, we define a particle at iteration t as the stacked vector of k marker positions $\mathbf{x}_t \in \mathbb{R}^{3k}$ of the candidate solution. We further modify the generic PSO scheme such that the 3D positions within each particle are constrained to the surface of the hand model. Specifically, after every particle update we project each marker position in $\mathbf{x}_t$ onto its spatially closest vertex in the set $\mathcal{V}'$ of preselected feasible positions on the hand model. The new position $\mathbf{x}_{t+1}$ of a particle is determined by computing its new velocity $\mathbf{v}_{t+1}$ and translating along this vector. To this end, we first compute the standard PSO velocity update as

$$\widetilde{\mathbf{v}}_{t+1} = w \cdot (\mathbf{v}_t + c_1 \cdot r_1 \cdot (\bar{\mathbf{x}}_{\text{par}} - \mathbf{x}_t) + c_2 \cdot r_2 \cdot (\bar{\mathbf{x}}_{\text{pop}} - \mathbf{x}_t)), \tag{12}$$

where w is a weight determining the overall step length of the update, c_1 and c_2 are importance weights for the local and global attractors respectively, and r_1 and r_2 are uniformly distributed random numbers in $[0, 1]$. Due to the curvature of the hand model surface, applying this linear update to the current particle position can cause the markers to stray from the surface. To counteract this, we project the updated marker positions back onto the permissible regions defined by vertices $\mathcal{V}'$, which we denote by a projection operator $\Pi_{\mathcal{V}'}$. The final particle position update is therefore

$$\mathbf{x}_{t+1} = \Pi_{\mathcal{V}'}(\mathbf{x}_t + \widetilde{\mathbf{v}}_{t+1}). \tag{13}$$

After this, the new particle velocity is computed as $\mathbf{v}_{t+1} = \mathbf{x}_{t+1} - \mathbf{x}_t$. Figure 7 illustrates the surface-constrained PSO update.

Similar to [Oikonomidis et al. 2011], we perturb one randomly chosen marker position in 50% of the particles once in every third iteration, and use the weights $c_1 = 2.8$, $c_2 = 1.3$, and $w = 2/|2 - \psi - \sqrt{\psi^2 - 4\psi}|$ with $\psi = c_1 + c_2$. We use a total of 1000 particles, perform 100 PSO iterations and use between 3 and 10 keyframes depending on the input hand motion trajectory. Using this method, we can find reduced marker layouts that optimize the objective function (9) and as a result are numerically stable and geometrically feasible.

Figure 8: *We fit a template model (a) to the point cloud (b) by using landmarks (red and blue dots). For initial alignment and adaption of the posture we perform rigid ICP (c) alternating with inverse kinematic (d). Afterwards we fit the shape of the template model based on landmarks (e) and closest point correspondences (f).*

Hand model generation

Our hand tracking approach fits an articulated virtual hand model to the marker positions obtained from an optical motion tracking system in order to determine pose and posture of the user's hand. While the marker layout optimization, as described in the previous section, is independent of the particular hand geometry and can therefore be performed on a generic hand template, the hand tracking itself requires the virtual hand model to closely match the user's hand proportions to obtain accurate results. We therefore developed a framework for generating user-specific hand models from 3D scanner data of the user's hand, which we describe below. In order to stress-test our model generation and hand tracking approach, we performed experiments with the two participants that featured the largest and smallest hands we could find in our lab environment.

Our 3D Scanner consists of eight simultaneously triggered DSLR cameras. From the resulting images we compute a dense point cloud using the commercial multi-view stereo reconstruction software Agisoft Photoscan. We denote these n points by $\mathcal{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$. Due to the scanner setup, each point is equipped with a normal $\mathbf{n}_i$ and a color $\mathbf{c}_i$ as well. As shown in Figure 16, the resulting point clouds suffer from noise, missing data, and outliers. To create a clean and complete user-specific hand model from these point clouds, we employ nonrigid registration to fit a generic hand template to the user's point cloud [Allen et al. 2003; Hasler et al. 2009; Achenbach et al. 2015]. Our template model is a triangle mesh consisting of $m \approx 12k$ vertices, whose positions we denote by $\mathcal{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$. It is fully rigged and can be animated by its skeleton (see Figure 8(a)).

As a preprocessing step we remove most of the outliers from the point cloud by discarding points that (because of the non-skin color) do not belong to the hand [Kovac et al. 2003]. To bootstrap the template fitting procedure we then manually select a set $\mathcal{L}$ of 20 landmarks on the template model $\{\mathbf{x}_l\}_{l\in\mathcal{L}}$ and on the point cloud $\{\mathbf{p}_l\}_{l\in\mathcal{L}}$. Based on these landmarks and closest point correspondences we alternatingly optimize the model's pose (translation, rotation, scaling) and posture (joint angles) by rigid iterative closest point (ICP) [Besl and McKay 1992; Horn 1987] and inverse kinematics (IK), respectively (see Figure 8(c) and Figure 8(d)).

Now that we have a good initial alignment of the template model and the scanned point cloud, we start adjusting the geometric shape of the template model to the point cloud data. To this end, we adapt the nonrigid face registration of Achenbach et al. [Achenbach et al. 2015] to our problem by minimizing the composed energy

$$E(\mathcal{X}) \;=\; \lambda_{\mathrm{lm}} E_{\mathrm{lm}}(\mathcal{X}) + \lambda_{\mathrm{fit}} E_{\mathrm{fit}}(\mathcal{X}) + \lambda_{\mathrm{reg}} E_{\mathrm{reg}}\big(\mathcal{X}, \bar{\mathcal{X}}\big)\,, \quad (14)$$

where the three energy terms are explained below.

The *landmark term* E_{lm} penalizes the (squared) distance between the 20 landmark points $\mathbf{x}_l$, $l \in \mathcal{L}$, on the template model and their landmark points $\mathbf{p}_l$ in the point cloud:

$$E_{\mathrm{lm}}(\mathcal{X}) = \frac{1}{|\mathcal{L}|} \sum_{l\in\mathcal{L}} \|\mathbf{x}_l - \mathbf{p}_l\|^2 \,. \quad (15)$$

The *fitting term* E_{fit} similarly measures the (squared) distance between corresponding point pairs $(\mathbf{x}_c, \mathbf{p}_c)$, $c \in \mathcal{C}$, where $\mathcal{C}$ denotes the set of closest-point correspondences:

$$E_{\mathrm{fit}}(\mathcal{X}) = \frac{1}{|\mathcal{C}|} \sum_{c\in\mathcal{C}} \|\mathbf{x}_c - \mathbf{p}_c\|^2 \,. \quad (16)$$

The *regularization term* E_{reg} penalizes the geometric distortion from the undeformed template model $\bar{\mathcal{X}}$ to the deformed state $\mathcal{X}$, measured by the norm of the per-vertex deformation Laplacian

$$E_{\mathrm{reg}}\big(\mathcal{X}, \bar{\mathcal{X}}\big) = \frac{1}{|\mathcal{X}|} \sum_{v\in\mathcal{V}} \|\Delta\mathbf{x}_v - \mathbf{R}_v \cdot \Delta\bar{\mathbf{x}}_v\|^2 \,, \quad (17)$$

where $\mathbf{R}_v$ are per-vertex rotations to best-fit deformed and undeformed Laplacians (see, e.g., [Achenbach et al. 2015] for details).

The three coefficients λ_{lm}, λ_{fit}, and λ_{reg} are used to guide the iterative fitting procedure, where the surface stiffness is controlled by λ_{reg}. In the begining, only the manually specified (hence quite reliable) landmarks are taken into account, using $\lambda_{\mathrm{reg}} = 1$, $\lambda_{\mathrm{lm}} = 1$ and $\lambda_{\mathrm{fit}} = 0$. We then decrease λ_{reg} gradually after each iteration until $\lambda_{\mathrm{reg}} = 10^{-6}$. Figure 8(e) depicts the result of this step. After these iterations, the models are sufficiently well aligned to rely on closest-point constraints. We therefore continue with $\lambda_{\mathrm{reg}} = 10^{-6}$ and $\lambda_{\mathrm{lm}} = 1$, but additionally set $\lambda_{\mathrm{fit}} = 1$ to also consider E_{fit}. λ_{reg} is again gradually decreased until $\lambda_{\mathrm{reg}} = 10^{-8}$. Figure 8(f) depicts the result of this step.

Due to this shape deformation the template's joints are not at the correct position anymore. We correct this by calculating the new joint positions with respect to the new vertex positions by exploiting mean value coordinates [Floater et al. 2005], which are precomputed in the initial undeformed state.

The final result is is a clean, complete, and ready-to-animate hand model that closely matches the shape and proportions of the user's hand. This hand model calibration will be shown in the following section to yield noticably more accurate tracking results.

Results

In the following we present some results produced by our system as well as quantitative evaluations of the accuracy of our motion

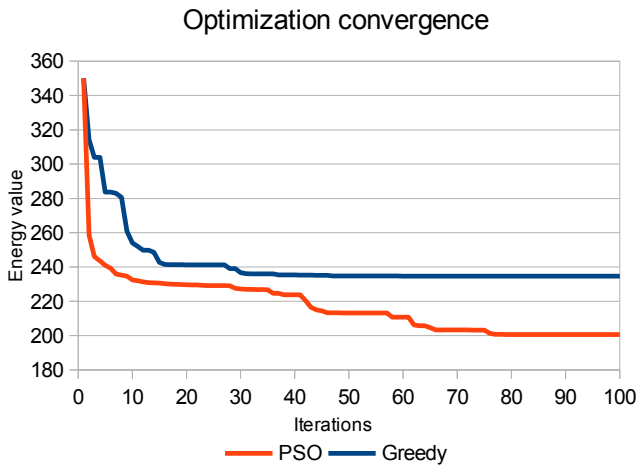

Figure 9: *Example for the energy minimization convergence of our PSO method compared to a greedy approach with identical initialization. While the greedy approach converges to a suboptimal local minimum after about 50 iterations, our stochastic optimization minimizes the energy faster and achieves a better result.*

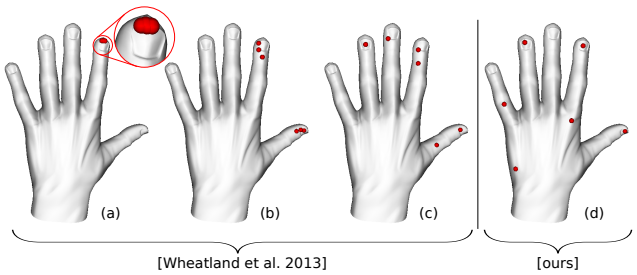

Figure 10: *Six marker layout generation for precision grasp movements using the method of Wheatland et al. [Wheatland et al. 2013] and our approach. In (a) the complete set of preselected vertices $\mathcal{V}'$ is used as the base marker set, which causes the selected markers to cluster at the index fingertip, as it exhibits the most movement. In (b) a random subset with 5% of $\mathcal{V}'$ is used as the base marker set, which leaves 3 candidate positions per joint. In this case the markers cluster around the index and thumb tips. In (c) 1% of $\mathcal{V}'$ is used, which leaves one candidate position per joint. The resulting marker set is distributed among the most active joints in the input motion. In (d) our approach generates a marker layout from the complete set $\mathcal{V}'$ based on the DoFs of our subspace model.*

reconstruction method with respect to the employed subspace IK approach, the auto-generated reduced marker layouts, as well as the user-specific hand model calibrations.

Layout optimization

We evaluated the convergence properties of our marker optimization in a varied set of evaluation trials. The hand movements involved in these trials included a variety of grasping and other manual interaction movements, as well as generic finger movements and gestures.

We analyze the convergence properties of our PSO-based marker layout optimization by comparing it to a more straightforward greedy approach. For this, we adapted the farthest point optimization scheme of Schlomer et al. [Schlömer et al. 2011] to find the marker subset of the initial vertex set $\mathcal{V}'$ that minimizes the objective function (9). Briefly stated, this method first iteratively selects the next best vertex as a marker position that reduces the objective value until the desired number of markers has been placed. Then, this greedy process is repeated such that each selected marker position is replaced by the next better remaining vertex position, until no more substitutions can be done to improve the objective value. This is already a more sophisticated approach than the greedy methods for constraint selection used in [Loper et al. 2014; Thiery et al. 2012] and can therefore serve as an upper bound for the effectiveness of such methods. Figure 9 compares this greedy approach with our PSO-based one with identical initialization and shows that our method converges faster and achieves better objective values. The runtime for our PSO method varies between 5 and 10 seconds for 100 iterations, depending on the number of keyframes (up to 10). For the same problem setup, the greedy approach takes between 45 seconds and 3 minutes to converge.

A comparison of our marker layout optimization method with the marker subset selection approach of Wheatland et al. [Wheatland et al. 2013] is shown in Figure 10. A crucial aspect to note regarding this comparison is that the two methods are based on different marker layout generation paradigms. While Wheatland and colleagues select the most influential markers in an initial base marker set, our method generates marker layouts more freely within the dense set of preselected vertices $\mathcal{V}'$. Figure 10 shows that the results of the subset selection method are strongly influenced by the choice of the base marker layout. As the method of Wheatland [Wheatland et al. 2013] is based on computing an importance ranking for the base markers based on their positional trajectories, the selected marker layouts are clustered around the areas of the hand that move the most in the considered hand motion. In contrast, our method is sensitive to the hand kinematics and the subspace model employed in our approach, which produces layouts that are well-suited for subspace-constrained IK, as opposed to the data-driven regression approach of Wheatland [Wheatland et al. 2013].

Motion reconstruction

We evaluated the motion reconstruction accuracy of the reduced marker layouts based on several trials including a large variety of hand movements. In the performed trials, we measured runtime statistics and average per-vertex errors of the reconstructed hand motions compared to the ground truth input. For proper evaluation of the accuracy of our approach, the input motions being reconstructed were not contained in the database used to generate the subspace model. As our reduced marker sets are optimized to represent only rotational DoFs of the hand articulation, an initial estimate for the global position of the hand is given by a fixed anchor marker on the forearm.

To assess the suitability of the marker layouts generated by our method for motion reconstruction, we compare the reconstruction error of differently obtained marker layouts in Figure 11. The testbed of this evaluation is a set of grasping motions based on the grasp taxonomy of [Cutkosky 1989]. We generated two different types of marker sets with varying sizes – a specific type based on grasping input motions, and a generic type based on general gestures and hand articulations. Additionally, we compare with the manually selected marker layouts of [Hoyet et al. 2012, Figures 4 and 8], who also performed motion reconstruction based on constrained IK. These manually selected layouts produce similar results to our automatically generated generic layouts. The reconstruction error is lower when using the grasp-specific marker layouts than generic ones. In particular, to achieve a reconstruction error below 2 mm, a specific layout generated by our method only

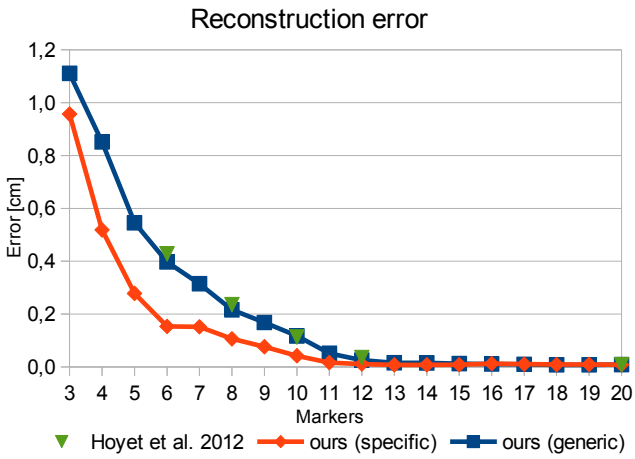

Figure 11: *Average reconstruction error using marker sets of varying sizes for grasping motions. Using marker layouts specifically generated based on grasping input motions the reconstruction error is lower than when using a layout based on generic motions. The manually selected marker layouts of Hoyet et al. [Hoyet et al. 2012] produce similar results to our automatically generated generic layouts.*

Method	Average error	Maximum error
Subspace IK	0.89 cm	2.1 cm
Standard IK	1.79 cm	7.9 cm

Table 1: *Average and maximum reconstruction error using subspace-constrained IK and standard IK with a 4-marker layout generated for a variety of manual interaction motions. While the standard method can deviate by almost 8 cm, the subspace method achieves adequate results consistently.*

requires 6 markers, whereas generic layouts require 9 markers or more. Examples for our generated layouts are given in Figure 12.

We verify the general accuracy and generalization capability of the subspace-constrained IK motion reconstruction based on marker layouts produced by our method by comparing its average reconstruction error to the error when using standard IK. Table 1 shows the average and maximum errors for a variety of manual interaction motions using standard IK and subspace IK and a reduced layout with four markers. The improvement of the subspace method over the standard method ranges between 9 mm to almost 6 cm. The overall error produced by our layered IK scheme using the subspace prior for initialization produces more accurate results for sparse marker layouts than the standard IK approach.

Figure 12 shows some examples for reduced marker layouts computed by our approach for various different movements. The results show that markers are preferentially placed in areas that have the most involvement in the considered hand motion. If the motions contain more varied articulations for specific fingers over others, these fingers will receive more markers, as the low-frequency details of the remaining markers are not influenced by as many subspace DoFs. In the third row of Figure 12, the input motion involves all fingers and the reduced marker layout accordingly distributes markers across all of them.

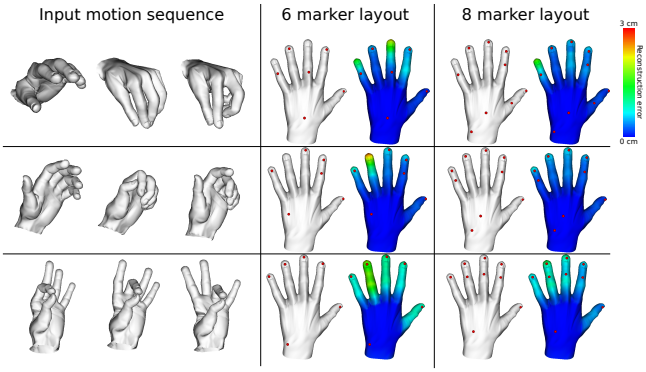

Figure 12: *Reduced marker layouts for some example motions. First row: precision grasp motion involving multiple fingers. Second row: power grasp motion of a small object. Third row: sequence in which the thumb touches all the other fingers. The marker layouts are optimized to allow for accurate reconstruction of the input motion. Markerless fingers tend to have slightly larger reconstruction errors, however they still move in correlation to the marker-constrained fingers due to the subspace approach.*

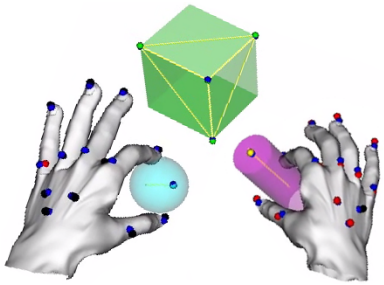

Figure 13: *Synchronous tracking of two hands manipulating a number of rigid objects. Our system automatically clusters the raw marker point cloud and fits generic model templates to the data.*

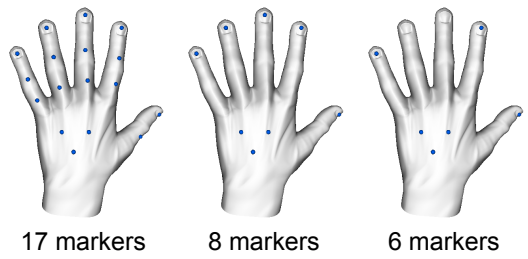

Figure 14: *Marker layouts used in our experiments. The reduced layouts (middle, right) have been computed as subsets of the full marker layout (left).*

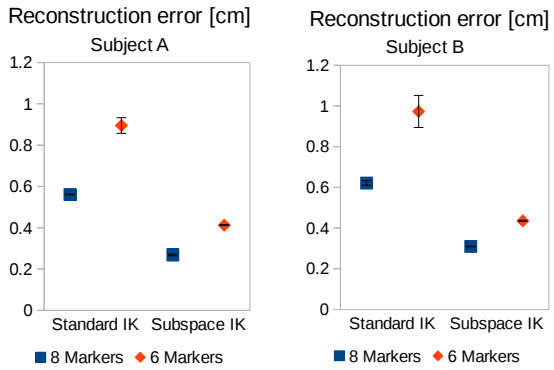

Figure 15: *Comparison between the reconstruction error for reduced marker layouts using standard IK and subspace IK. Reduced marker sets perform within 5 millimeters accuracy using our subspace IK approach. The slightly larger errors for Subject B are due to larger hands.*

Our method is also able to accurately track scenes containing multiple objects in concert. Figure 13 and the accompanying video show an example of tracking two hands interacting with multiple objects.

Calibrated hand models

We evaluated our system in general, and the model calibration in particular, by capturing a series of hand motions from two different performers. In the following the performers will be referred to as Subject A and Subject B. Subject A is a female performer with small hands, whereas Subject B is a male performer with large hands. Both subjects were asked to perform a variety of hand movements, including grasping and handling objects of different sizes as well as touching the finger tips with the thumb. This provided a varied real-world data set for evaluating the quality of our calibration and motion reconstruction system.

For the purpose of evaluating reconstruction accuracy, a configuration of 17 markers was attached to the subjects, which inherently covers most of the hand's kinematic DoFs. This fairly dense marker set is well-suited for evaluating the quality of our calibration, as mismatches between the model and the data are more easily highlighted than when using sparse marker sets. In order to still evaluate the reconstruction accuracy of reduced marker layouts, we generated reduced 6- and 8-marker layouts as subsets of the 17-marker ground truth layout. Figure 14 shows the marker layouts used in our experiments.

The reconstruction accuracy was evaluated by comparing the solution using the reduced marker layouts with that using the full marker layout. When using the reduced layouts, the additional markers contained in the captured data are ignored. It is worth noting that our automatic tracking procedure performs very robustly in the presence of spurious data. Figure 15 shows that the mean reconstruction error of the reduced 6- and 8-marker layouts lies below 5 mm across all our experiments for both subjects.

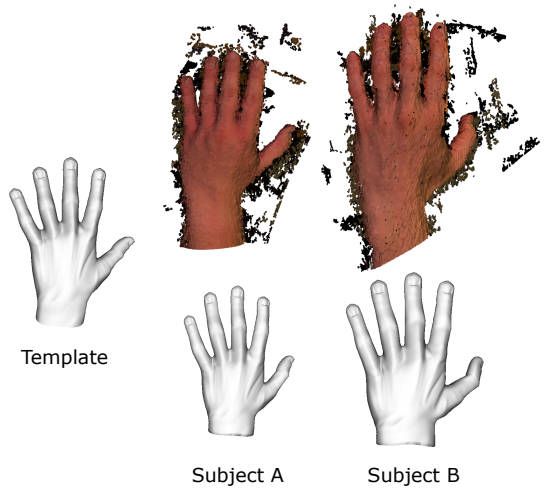

Figure 16: *Hand models used in our experiments with their corresponding point clouds. Left: uncalibrated template model. Middle: calibrated model and point cloud for Subject A (female, small hands). Right: calibrated model and point cloud for Subject B (male, large hands).*

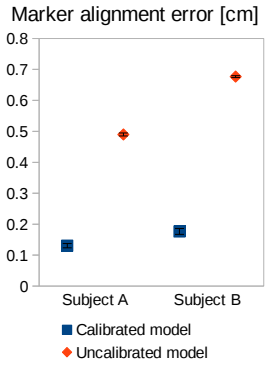

Figure 17: *Average marker alignment error using calibrated and uncalibrated models for two subjects across all captured data. The error is significantly improved by our user-specific model calibration. The slightly larger error for Subject B is due to larger hands.*

The hand models used in our experiments are shown in Figure 16. The subjects' hands were scanned and subject-specific hand models were generated using our approach. To ascertain the accuracy improvements of our calibrated models over the uncalibrated model, we measured the alignment between the real marker positions in the mocap data and the respective model's proposed marker positions after convergence. Figure 17 compares this marker alignment error for the uncalibrated and the calibrated models and shows improvements of almost half a centimeter.

Figure 18 shows the trajectory of the marker alignment error along a single grasping movement sequence by Subject B. The error of the calibrated model is initially close to zero as the subject's hand is in the same neutral position that was used for calibration. The

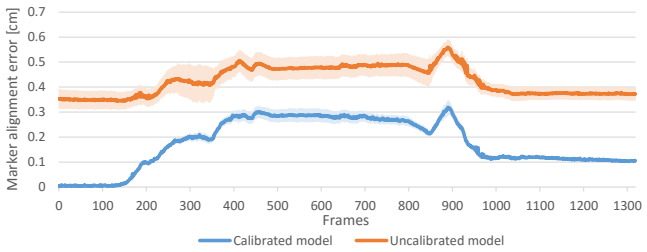

Figure 18: *Marker alignment error using calibrated and uncalibrated models for one movement sequence by Subject B. The graph shows both the mean marker error and the variance across all markers for each frame. Our calibration improves upon both the mean and the variance of the marker alignment error.*

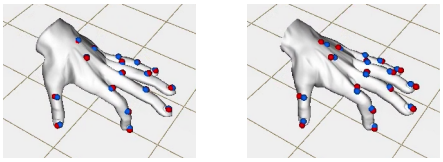

Figure 19: *The alignment between the model's virtual markers (blue) and the captured marker data (red) is more accurate with our user-calibrated model (left) than the uncalibrated template model (right). Small discrepancies with the calibrated model are due to differences between real hand deformations and our simple LBS deformation model.*

subsequent error increase is due to the fact that our straightforward deformation model (linear blend skinning) cannot reconstruct some details of the real hand's deformation with high accuracy. Nonetheless, the calibration provides significant accuracy improvements.

A qualitative impression of the difference in marker alignment between the calibrated and the uncalibrated model is given in Figure 19. In Figure 20 a qualitative comparison between the real hand of Subject B and the generated hand posture reconstruction is shown.

Discussion

We have presented a method that automatically computes reduced marker layouts for optical motion capture using subspace-constrained inverse kinematics motion reconstruction. Our marker layout optimization method minimizes an objective function that jointly measures the numerical stability and geometric feasibility of the reduced marker configuration. The objective function is minimized using a specialized surface-constrained particle swarm optimization scheme, which stochastically explores the solution space of feasible marker configurations on the surface of an animated hand model. We showed that the resulting marker layouts are suitable for solving the subspace-constrained inverse kinematics problem for motion reconstruction from sparse input. Additionally, the optimized marker layouts are specific to the type of hand motions that should be expressed with and recovered from the sparse marker data. The marker sets are well-suited for practical use as they are intuitive and scale well with reduced resolution of the mocap system. We presented a hand model calibration procedure that fits a geometric model to point cloud data of the user's hand and we showed that this provided significant improvements for the overall motion reconstruction accuracy.

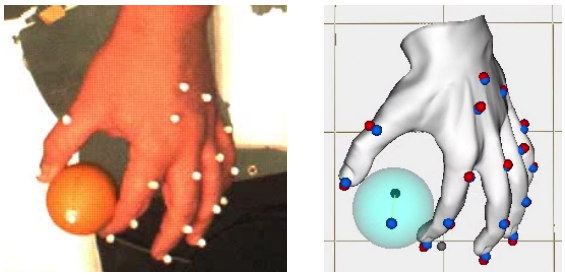

Figure 20: *Comparison of the real human subject performing a grasping motion (left) and the animation of the virtual hand model reconstructed by our system (right). The animation is best appreciated in the accompanying video.*

Our method makes it possible to generate marker layouts that are fine-tuned to the parameters of a given mocap setup. If there is a limitation to the number of markers that can be used in the mocap setup, our method computes the optimal placements for the given number of markers that allows for realistic motion reconstruction that is also rich in expressiveness. An insight provided by our work is that it is sufficient for high quality motion reconstruction to place individual markers on the hand that correspond to low-dimensional control parameters of hand articulations. For instance, to track grasping motions with high quality using our method, it is sufficient to only place one marker on the thumb, index finger, pinky finger and wrist. The subspace based reconstruction will plausibly interpolate the movements of joints that are not immediately constrained by markers.

Limitations of our approach include the stochastic nature of the particle swarm optimization and the need for parameter tweaking. Another drawback of our subspace-oriented method is that while it produces good results for specific hand movements, it does not necessarily provide a general-purpose marker layout result that can be used for all types of motions and produce high-quality results. The marker placement as well as the motion reconstruction are limited by the subspace priors employed. However, given prior knowledge of the motions intended to be tracked, our method produces accurate and robust results. Beyond marker placement, our approach could be used generally to identify salient regions in articulated bodies, which could be of interest for different avenues of motion detection and reconstruction.

Acknowledgments

This work was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

ACHENBACH, J., ZELL, E., AND BOTSCH, M. 2015. Accurate Face Reconstruction through Anisotropic Fitting and Eye Correction. In *Vision, Modeling & Visualization*, 1–8.

ALBRECHT, I., HABER, J., AND SEIDEL, H.-P. 2003. Construction and animation of anatomically based human hand models. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 98–109.

ALLEN, B., CURLESS, B., AND POPOVIĆ, Z. 2003. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics 22*, 3, 587–594.

BERNSTEIN, N. 1967. *The Co-ordination and Regulation of Movements*. Pergamon Press Ltd.

BESL, P. J., AND MCKAY, N. D. 1992. A method for registration of 3-D shapes. *IEEE Trans. on Pattern Anal. Mach. Intell. 14*, 2, 239–256.

2015. Biosyn. http://www.biosynsystems.net/.

BUSS, S. R. 2004. Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods. *IEEE Journal of Robotics and Automation 17*, 1-19, 16.

CHAI, J., AND HODGINS, J. K. 2005. Performance animation from low-dimensional control signals. *ACM Transactions on Graphics 24*, 3, 686–696.

CHANG, L., POLLARD, N., MITCHELL, T., AND XING, E. 2007. Feature selection for grasp recognition from optical markers. In *International Conference on Intelligent Robots and Systems. IEEE/RSJ*, 2944–2950.

CUTKOSKY, M. 1989. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation 5*, 3, 269–279.

EDMONDS, J., AND KARP, R. M. 1972. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of ACM (JACM) 19*, 2, 248–264.

FLOATER, M. S., KÓS, G., AND REIMERS, M. 2005. Mean value coordinates in 3D. *Computer Aided Geometric Design 22*, 7, 623–631.

GUERRA-FILHO, G. B. 2005. Optical motion capture: Theory and implementation. *Journal of Theoretical and Applied Informatics (RITA) 12*, 61–89.

HASLER, N., STOLL, C., SUNKEL, M., ROSENHAHN, B., AND SEIDEL, H.-P. 2009. A statistical model of human pose and body shape. *Computer Graphics Forum 28*, 2, 337–346.

HORN, B. K. P. 1987. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A 4*, 4, 629–642.

HOYET, L., RYALL, K., MCDONNELL, R., AND O'SULLIVAN, C. 2012. Sleight of hand: Perception of finger motion from reduced marker sets. In *Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 79–86.

JACKA, D., REID, A., MERRY, B., AND GAIN, J. 2007. A comparison of linear skinning techniques for character animation. In *Proceedings of the 5th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, ACM, 177–186.

KANG, C., WHEATLAND, N., NEFF, M., AND ZORDAN, V. B. 2012. Automatic hand-over animation for free-hand motions from low resolution input. In *Proceedings of ACM Motion in Games*, Springer, vol. 7660 of *Lecture Notes in Computer Science*, 244–253.

KATO, M., CHEN, Y.-W., AND XU, G. 2006. Articulated hand motion tracking using ICA-based motion analysis and particle filtering. *Journal of Multimedia 1*, 52–60.

KENNEDY, J., AND EBERHART, R. 1995. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, vol. 4, 1942–1948.

KENNEDY, J., AND EBERHART, R. C. 2001. *Swarm Intelligence*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

2010. Microsoft kinect.

KITAGAWA, M., AND WINDSOR, B. 2008. *MoCap for Artists: Workflow and Techniques for Motion Capture*. Focal Press.

KOVAC, J., PEER, P., AND SOLINA, F. 2003. Human skin color clustering for face detection. In *Proceedings of EUROCON 2003. Computer as a Tool. The IEEE Region 8*, vol. 2, 144–148.

KUHN, H. W. 1955. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly 2*, 1–2, 83–97.

LE, B. H., ZHU, M., AND DENG, Z. 2013. Marker optimization for facial motion acquisition and deformation. *IEEE Transactions on Visualization and Computer Graphics 19*, 11, 1859–1871.

LI, H., SUMNER, R. W., AND PAULY, M. 2008. Global correspondence optimization for non-rigid registration of depth scans. In *Computer Graphics Forum*, 1421–1430.

LI, H., ADAMS, B., GUIBAS, L. J., AND PAULY, M. 2009. Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics 28*, 5, 175:1–175:10.

LIU, G., ZHANG, J., WANG, W., AND MCMILLAN, L. 2006. Human motion estimation from a reduced marker set. In *Proceedings of ACM Symposium on Interactive 3D Graphics and Games*, 35–42.

LOPER, M. M., MAHMOOD, N., AND BLACK, M. J. 2014. MoSh: Motion and shape capture from sparse markers. *ACM Transactions on Graphics 33*, 6, 220:1–220:13.

MAYCOCK, J., RÖHLIG, T., SCHRÖDER, M., BOTSCH, M., AND RITTER, H. 2015. Fully automatic optical motion tracking using an inverse kinematics approach. In *IEEE-RAS International Conference on Humanoid Robots*, 461–466.

MULATTO, S., FORMAGLIO, A., MALVEZZI, M., AND PRATTICHIZZO, D. 2013. Using postural synergies to animate a low-dimensional hand avatar in haptic simulation. *IEEE Transactions on Haptics 6*, 5, 106–116.

OIKONOMIDIS, I., KYRIAZIS, N., AND ARGYROS, A. A. 2011. Efficient model-based 3D tracking of hand articulation using kinect. In *Proceedings of British Machine Vision Conference*, vol. 1, 3.

2015. Optitrack. http://www.optitrack.com/.

2015. Organic motion. http://www.organicmotion.com/.

2015. Phasespace. http://www.phasespace.com/.

QIAN, C., SUN, X., WEI, Y., TANG, X., AND SUN, J. 2014. Realtime and robust hand tracking from depth. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 1106–1113.

2015. Qualisys. http://www.qualisys.com/.

RHEE, T., NEUMANN, U., AND LEWIS, J. P. 2006. Human hand modeling from surface anatomy. In *Proceedings of ACM Symposium on Interactive 3D Graphics and Games*, 27–34.

RUSU, R. B. 2009. *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. PhD thesis, Technische Universität München.

SCHLÖMER, T., HECK, D., AND DEUSSEN, O. 2011. Farthest-point optimized point sets with maximized minimum distance. In *Proceedings of ACM SIGGRAPH Symposium on High Performance Graphics*, 135–142.

SCHRÖDER, M., MAYCOCK, J., RITTER, H., AND BOTSCH, M. 2014. Real-time hand tracking using synergistic inverse kinematics. In *IEEE International Conference on Robotics and Automation*, 5447–5454.

SCHRÖDER, M., MAYCOCK, J., AND BOTSCH, M. 2016. Reduced marker layouts for optical motion capture of hands. In *Proceedings of ACM Motion in Games*, 7–16.

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., AND IZADI, S. 2015. Accurate, robust, and flexible real-time hand tracking. In *Proceedings of ACM Conference on Human Factors in Computing Systems*, 3633–3642.

TAGLIASACCHI, A., SCHRÖDER, M., TKACH, A., BOUAZIZ, S., BOTSCH, M., AND PAULY, M. 2015. Robust articulated-icp for real-time hand tracking. *Computer Graphics Forum 34*, 5.

TAN, D. J., CASHMAN, T., TAYLOR, J., FITZGIBBON, A., TARLOW, D., KHAMIS, S., IZADI, S., AND SHOTTON, J. 2016. Fits like a glove: Rapid and reliable hand shape personalization. In *IEEE Conference on Computer Vision and Pattern Recognition*.

TAYLOR, J., STEBBING, R., RAMAKRISHNA, V., KESKIN, C., SHOTTON, J., IZADI, S., HERTZMANN, A., AND FITZGIBBON, A. 2014. User-specific hand modeling from monocular depth sequences. In *IEEE Conference on Computer Vision and Pattern Recognition*, 644–651.

TAYLOR, J., BORDEAUX, L., CASHMAN, T., CORISH, B., KESKIN, C., SHARP, T., SOTO, E., SWEENEY, D., VALENTIN, J., LUFF, B., TOPALIAN, A., WOOD, E., KHAMIS, S., KOHLI, P., IZADI, S., BANKS, R., FITZGIBBON, A., AND SHOTTON, J. 2016. Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences. *ACM Transactions on Graphics 35*, 4, 143:1–143:12.

THIERY, J.-M., TIERNY, J., AND BOUBEKEUR, T. 2012. CageR: Cage-based reverse engineering of animated 3D shapes. *Computer Graphics Forum 31*, 8, 2303–2316.

2015. Vicon. `http://www.vicon.com/`.

WHEATLAND, N., JÖRG, S., AND ZORDAN, V. 2013. Automatic hand-over animation using principle component analysis. In *Proceedings of ACM Motion in Games*, 175:197–175:202.

WHEATLAND, N., WANG, Y., SONG, H., NEFF, M., ZORDAN, V., AND JÖRG, S. 2015. State of the art in hand and finger modeling and animation. *Computer Graphics Forum 34*, 2, 735–760.

WU, Y., LIN, J. Y., AND HUANG, T. S. 2001. Capturing natural hand articulation. In *IEEE International Conference on Computer Vision (ICCV)*, 426–432.

2015. Xsens. `https://www.xsens.com/`.

ZHU, L., HU, X., AND KAVAN, L. 2015. Adaptable anatomical models for realistic bone motion reconstruction. *Computer Graphics Forum 34*, 2, 459–471.

Authors' Biographies

Matthias Schröder received his M.Sc. in Computer Science from Bielefeld University in 2011. He received his Ph.D. in Computer Science in 2015 after working in the Computer Graphics & Geometry Processing Group at Bielefeld University, where he worked on real-time hand tracking. He started as a postdoctoral researcher in the Neuroinformatics Group at Bielefeld University in 2016 and his current research focus is on computer vision, machine learning and data science.

Thomas Waltemate received his M.Sc. in Computer Science from Bielefeld University in 2012. Since 2013 he is working in the Computer Graphics & Geometry Processing Group at Bielefeld University as a Ph.D. student. His main research is on creation, animation and visualization of virtual characters.

Jonathan Maycock graduated in 2000 with a B.Sc. in Computer Science from the National University of Ireland Maynooth and undertook a Ph.D. in November 2004. In the interim he worked as a software engineer with Anam Wireless Internet Solutions, Marconi Communications and Scottish Equitable International. Jonathan started working in the Neuroinformatics Group of Bielefeld University in March 2008.

Tobias Röhlig received his M.Sc. in Computer Science from Bielefeld University in 2014. Following this, he began working as a Ph.D. student in the Neuroinformatics Group at Bielefeld University, where his main research focus was on robotic grasping and manual intelligence.

Helge J. Ritter studied physics and mathematics at the Universities of Bayreuth, Heidelberg and Munich and received a Ph.D. in Physics from the Technical University of Munich in 1988. Since 1990, he is professor at the Department of Computer Science of Bielefeld University. His main interests are principles of neural computation, in particular self-organizing and learning systems, and their application to robot cognition, data analysis and interactive man-machine interfaces. In 1999, Helge Ritter was awarded the SEL Alcatel Research Prize and in 2001 the Leibniz Prize of the German Research Foundation DFG. Helge Ritter is co-founder and one of the directors of the Bielefeld Institute of Cognition and Robotics (CoR-Lab) and coordinator of the excellence cluster "Cognitive Interaction Technology".

Mario Botsch received his M.Sc. in Mathematics and Computer Science from the University of Erlangen-Nürnberg in 1999, and his Ph.D. in Computer Science from RWTH Aachen in 2005. From 2005 to 2008 he was a senior researcher and lecturer at the Computer Graphics Laboratory at ETH Zurich. He has been a full professor in the Computer Science Department at Bielefeld University since May 2008 and is the head of the Computer Graphics & Geometry Processing Group.