2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Accurate, Robust, and Real-time Estimation of Finger Pose with a Motion Capture System

Youngmok Yun, Priyanshu Agarwal, and Ashish D. Deshpande

Abstract— **Finger exoskeletons, haptic devices, and augmented reality applications demand an accurate, robust, and fast estimation of finger pose. We present a novel finger pose estimation method using a motion capture system. The method combines system identification and state estimation in a unified framework. The system identification stage investigates the accurate model of a finger, and the state estimation stage tracks the finger pose with the Extended Kalman Filter (EKF) algorithm based on the model obtained in the system identification stage. The algorithm is validated by simulation and experiment. The experimental results show that the method can robustly estimate the finger pose at a high frequency (greater than 1 Khz) in presence of measurement noise, occlusion of markers, and fast movement.**

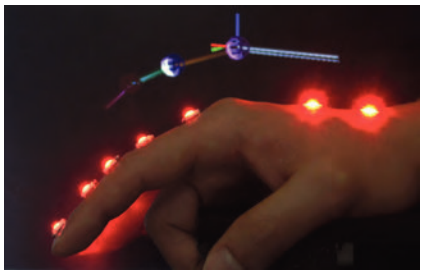


Fig. 1: Accurate, robust, and real-time estimation of finger pose with a motion capture system.

I. INTRODUCTION

Accurate, robust, and high-frequency estimation of finger pose is critical in many applications including finger exoskeletons [1][2], haptic devices[3][4], and augmented reality tools [5] used in medical operations, hazardous material handling, rehabilitation, and other applications. For example, kinematic loops of the hand rehabilitation robot inherently include a robot system and a human finger [6]. Therefore the kinematics modeling and estimation of the finger pose significantly affect the performance of the entire system.

A motion capture system is a powerful solution for measuring the pose of human finger [7]. Unlike mechanical systems, measuring the pose of the human finger is challenging, since it is hard to install mechanical elements such as encoders on a finger. In addition, installing mechanical components may significantly change the characteristics of finger dynamics. In this regard, the motion capture system is a desirable alternative because it only requires the attachment of a few markers on a human finger. Furthermore, modern motion capture systems can provide highly accurate and precise position information superior to the other contactless methods such as computer vision algorithms [8].

Nevertheless, the estimation of finger pose with a motion capture device has several challenges. The first challenge is associated with the *kinematic modeling*. An inaccurate kinematic model of the human body and markers degrades the accuracy of kinematic chain containing both the human body and robotic system. However, the modeling of a human finger and relation with markers is not easy because we cannot directly measure kinematic parameters such as bone lengths and markers' locations relative to finger joints. Moreover, mechanical joints cannot exactly describe human biological finger joints [9]. The second obstacle is related to *robust estimation*. The robust estimation is the foremost and essential condition for safe human-robot interaction. However, the estimation of finger pose is deteriorated by various factors such as occlusion of markers, deformation of human skin [10], limitation of modeling with idealized joints [11], and the noise in marker measurement data. The last obstacle is that the configuration of system may be different for different estimation trials. Different fingers have different kinematic parameters, and markers may not be attached at the same position for experiments. Therefore, finger model needs to be recalibrated before estimation. In other words, *both accurate system identification and robust estimation need to be performed in a unified framework.*

The most common method for the estimation of finger pose with a motion capture system is to first attach markers at predetermined positions (typically on the joints) and then calculate the finger's pose from the geometric relation of markers with a deterministic inverse kinematics method [12][13]. Several researchers, particularly in biomechanics, have developed more accurate kinematic modeling methods to investigate the dynamic biomechanical properties of the human finger. However, their estimation methods were not robust [11][14]. Many computer scientists have studied robust estimation algorithms of the human hand for the purpose of gesture recognition [16][17], but they do not include the accurate modeling of the finger. Their kinematic models were simplified and not subject-specific because they assumed incorrect geometric relation between the marker and joint

This work was supported, in part, by the National Science Foundation (NSF) grant #NSF-CPS-1135949.

Y. Yun and P. Agarwal are PhD students of Mechanical Engineering, The University of Texas at Austin, 3.130, ETC, Austin, Texas, USA {yunyoungmok}{mail2priyanshu} at utexas.edu

A.D. Deshpande is a faculty of Mechanical Engineering, The University of Texas at Austin, 3.130, ETC, Austin, Texas, USA Ashish at austin.utexas.edu

center, and each person has different shape of finger [14]. Till date, only Todorov has presented an augmented state estimation method for the simultaneous system identification and robust pose tracking of a generalized articulated multi-joint system [18]. However, the large dimension of the augmented state (61-dimension including joint angles and system parameters for two fingers) restricts the fast calculation of the state estimation, which is essential for the real-time control. Also, its tracking performance is highly dependent on the initial state due to the severe nonlinearity of the system.

To address the need for an accurate, robust, and fast estimation of finger pose, this paper presents a novel framework containing a system identification stage and a state estimation stage. The system identification stage investigates the optimal solution for the kinematic model parameters for the finger and markers. The estimation stage tracks the pose of a finger with Extended Kalman Filter (EKF) algorithm based on the optimal model obtained in the system identification stage.

This two-stage method realizes accurate, robust, and high-frequency estimation of finger pose with a motion capture system. The unified system configuration enables the use of an accurate optimal model in the estimation procedure. The separation into two stages allows for the real-time estimation at high sampling frequency (faster than 1Khz). The prediction model in the EKF, which is based on the jerk minimization theory [19][20], and the stochastic estimation algorithm estimate the finger pose robustly against the sensor noise, skin stretch, occlusion of markers, and the system modeling error. Furthermore, the EKF provides a quantitative confidence level for the estimates, which is important for the human-robot interaction.

II. System Identification Stage

Before the estimation stage, the system identification stage is carried out offline to determine the optimal model of the system consisting of the human finger and the markers of the motion capture system. Specifically, the model contains the kinematic model parameters of the finger and the location of markers relative to finger joints. After collecting a fixed number of marker data points while the finger moves freely, an optimization algorithm searches the optimal parameters minimizing the least square error between the actual marker positions and the estimated marker positions.

A. System Modeling

The human finger is modeled based on a previously presented method [21]. The metacarpophalangeal (MCP) joint is modeled with a saddle joint to represent two DOF motions of the abduction/adduction and flexion/extension. The proximal interphalangeal (PIP) joint and the distal interphalangeal (DIP) joint are modeled by a hinge joint to represent one DOF flexion/extension motions. A total of seven markers are attached to a finger. Two markers are attached on the metacarpal, proximal phalange, and intermediate phalange; one marker is attached on the distal phalange (Figure 2).

To develop a complete model of the above system, we need three types of parameter sets. The first set contains the translation (x,y,z) and rotation (roll, pitch, yaw) parameters to describe the relation between the world coordinate (O_{WOR}, or the coordinate of a motion capture system) and the local coordinate attached on the metacarpal O_{MET}. The second parameter set is for the lengths of the proximal and intermediate phalanges. The last set is three-dimensional locations of seven markers in the local coordinates attached on the joints and metacarpal (e.g., xyz positions of m3 in the coordinate O_{MCP}). The parameter sets are used to build a kinematic model of a finger and markers, and is indicated by p. Joint angles of MCP adduction, MCP flexion, PIP flexion, and DIP flexion are the state variables of this system, and are indicated by x. The three dimensional positions of seven markers in the world coordinates are estimated by a measurement function h with the model parameter p and the state variable x, and its relation is shown in (1).

$$\hat{z} = h(p, x) \tag{1}$$

where $\hat{z}$ is a vector indicating the estimated marker position in the world coordinate.

B. Optimization of Kinematic Model Parameter

The goal of the system identification stage is to identify the best model parameter p for the system, and is achieved by an optimization method. For the optimization, first, several measurement sets are collected with a motion capture system. Then, the optimal model parameter is selected to minimize the error between the estimated and actual marker positions. The optimization problem can be expressed by (2):

$$\underset{p, x_1 \ldots x_N}{\operatorname{argmin}}\ J = (\Delta z^\top \Delta z) \tag{2}$$

$$\Delta z = \begin{bmatrix} \hat{z}_1 - z_1 \\ \vdots \\ \hat{z}_N - z_N \end{bmatrix} = \begin{bmatrix} h(p, x_1) - z_1 \\ \vdots \\ h(p, x_N) - z_N \end{bmatrix}$$

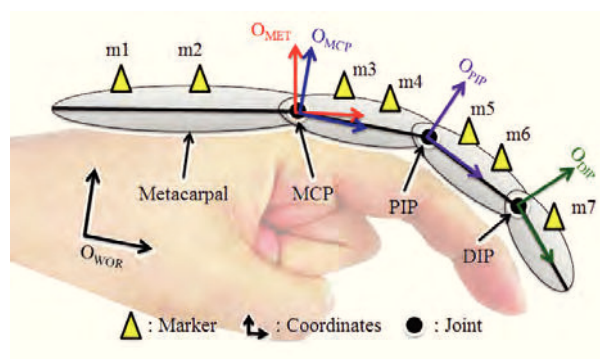


Fig. 2: Configuration of the kinematic modeling. MCP joint is modeled with a saddle joint, and PIP, DIP are modeled with a hinge joint. A total of seven markers are attached on metacarpal and phalanges. Each phalange and metacarpal have their local coordinates.

where N is the number of measurements for the optimization. z_i is a vector indicating the three dimensional positions of seven markers acquired at the i-th measurement from a motion capture system, and has 21-dimension. x_i is the state vector indicating joint angles at the i-th measurement. In normal cases, the above type of optimization problem is a challenging global optimization problem which has many local minimums and demands considerable amount of calculation time due to its large dimension and nonlinearity of the measurement function. However, the model parameters and states can be initialized based on the rough estimates obtained from the anthropometric constraints and the measurements. For example, the origin of coordinates of metacapal is located around the mid-point between the markers m2 and m3. The marker m4 probably is located on a phalange whose length is less than 10 cm. These initial guesses and constraints in the optimization problem dramatically reduce the calculation time and helps in avoiding inappropriate local minimum.

III. Estimation Stage

The goal of the state estimation stage is to estimate the state of finger robustly at high sampling rate under the presence of occlusion, noise of markers, and deformation of skin. The Extended Kalman Filter (EKF) algorithm is selected to achieve the above objectives. The EKF is a linearized version of Kalman Filter (another name is Linear Quadratic Estimator) and a probabilistic state tracking algorithm which has been widely verified through various research and industry applications [22] due to its robust and fast calculation performance. The EKF algorithm works in a two-step process. In the prediction step, it predicts the system state based on the system motion model, and the update step corrects the state based on the measurement data.

A. Prediction Step

The prediction step in the EKF predicts the state of the finger between the measurements or without measurements (during the occlusion of markers). Because the prediction step has to estimate the system state without measurement, it needs an information to "predict" the state at the next time step. Generally in Robotics, a control input and corresponding motion model are used as an information. However, in this application, we cannot acquire any control input such as subject's intention. Therefore, here we use the jerk minimization theory [19][20] which claims that human motion tends to minimize the jerk, defined as a derivative of acceleration with respect to time. It acts as a prediction model for the finger motion. Based on the jerk minimization theory, the jerk term is regarded as a small white noise. The state x representing the random variables of four joint angles is augmented to express the jerk $\Delta\ddot{x} = \ddot{x}_k - \ddot{x}_{k-1}$ as shown in (3):

$$y = \begin{bmatrix} x & \dot{x} & \ddot{x} \end{bmatrix}^\top \tag{3}$$

where y is the augmented state. The augmented state is predicted by the model in (4).

$$\bar{y}_k = Ay_{k-1} + \mathcal{N}(\mathbf{0}, Q) \tag{4}$$
$$A = \begin{bmatrix} I_3 & \Delta t & 0 \\ 0 & I_3 & \Delta t \\ 0 & 0 & I_3 \end{bmatrix}, Q = \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_a^2 & 0 \\ 0 & 0 & \sigma_j^2 \end{bmatrix}$$

where the bar $\bar{(\cdot)}$ indicates a predicted state before the update step, $\bar{y}_k$ is the predicted state at the kth time step, A is a motion model for the prediction, Q is a process noise matrix, I_3 is 3×3 identity matrix, and $\Delta t, \sigma_v^2, \sigma_a^2$, and σ_j^2 are three dimensional diagonal matrices whose elements are the sampling time, and model noises for velocity, acceleration and jerk, respectively. The mean and covariance are predicted using (5) and (6).

$$\bar{\mu}_k = A\mu_{k-1} \tag{5}$$
$$\bar{P}_k = FP_{k-1}F^\top + Q \tag{6}$$

where μ and P are the mean and covariance of y, respectively. F is the Jacobian of A with respect to the y. The predicted state $\bar{x}_k = \mathcal{N}(\bar{\mu}, \bar{P})$ is corrected by the update step, but if there is no measurement in the k-th step, it becomes the current state, $x_k = \bar{x}_k$.

B. Update Step

The update step in the EKF corrects the predicted state by the measurements, which are obtained from the motion capture device. The updated mean μ_k and covariance P_k are calculated using (7) and (8).

$$\mu_k = \bar{\mu}_k + K_k(z_k - \hat{z}_k) \tag{7}$$
$$P_k = (I - K_kH_k)\bar{P}_k \tag{8}$$
$$K_k = \bar{P}_kH_k^\top\Psi_k^{-1}, \quad \Psi = H_k\bar{P}_kH_k^\top + R_k$$

where z and $\hat{z}$ are the measurements and estimated measurements which are calculated by using (1), K is the Kalman gain, H is the Jacobian of measurement function h with respect to y, Ψ is the innovation matrix, and R is the measurement noise matrix.

The measurement noise matrix R is regulated to handle the uncertainty of measurement and the occlusion of markers. In the normal case when a motion capture system successfully provides the position of markers, the elements of R are filled with the square of average error of marker positions containing device errors and skin stretch. However, when it fails, in most cases due to occlusion, the elements of R corresponding to the occluded marker are filled with an infinite number, (practically a very large number). The infinite size of variance physically means the absence of observed information, thus the occluded marker data is not used to correct the predicted state using (7) and (8). The stochastic and selective update by Kalman gain in (7) and (8) makes the algorithm robust to the noise of the measurement.

IV. Results

The system identification and estimation methods were validated by simulation and an experiment with a subject. Because it is difficult to know ground-truth values for the finger pose without an X-ray device, a virtual experiment was conducted with a computer simulation. Then, an experiment was also performed, in which the tracking performance and practical effectiveness are validated with a real-time GUI.

A. Simulation

We developed a virtual finger and a virtual motion capture system to evaluate the performance of algorithms by comparing with the ground-truth values. The virtual finger moved along a predetermined trajectory and the virtual motion capture device collected the marker positions. The measurement noise was modeled by a white noise whose standard deviation is 3 mm.

1) System Identification: For the system identification of the virtual system, 20 virtual measurement sets were collected while the finger moved. The optimization was conducted with a line search algorithm [23], a gradient-based and constrained optimization algorithm, in MATLAB. The optimization time was less than 60 seconds with a computer which has Intel I5 2.5 Ghz CPU and 8 GB RAM. The performance of the system identification stage was evaluated by repeating the above procedure 100 times. Fig. 3 shows the average error compared to the ground-truth parameters.

2) State Estimation: The state estimation algorithm was validated by moving the virtual finger along the specific trajectory. The prediction step was executed with 1kHz, and the virtual marker data were acquired with 480Hz. With the same trajectory, the state estimation was performed twice with 10% marker occlusion and 50% marker occlusion. The occlusion was simulated by eliminating marker data randomly with the given ratio.

Fig. 4 shows the simulation results for the state estimation compared with the ground-truth trajectories. The estimation algorithm was able to track the finger pose quickly while keeping the errors low. Also, the result shows that the estimation error of DIP flexion is larger than that of MCP flexion. This is because MCP flexion is updated by five markers' positions, but DIP flexion is affected by only one markers' position. Regarding the occlusion, although in both cases of 10% and 50%, the estimated joint angles converged to the ground-truth trajectories, the occlusion increased the size of error and uncertainty, and reduced the convergence time as shown in Fig. 4(e) and (f).

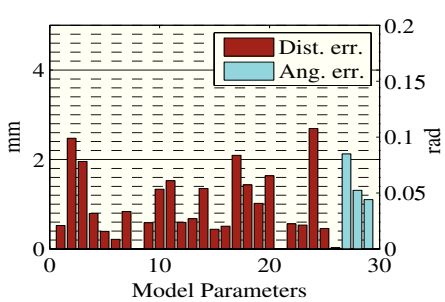


Fig. 3: Results of system identification in the simulation. The height of bars indicates the average error of 100 optimized model parameter sets compared with the ground-truth values.

B. Experiment

An experiment with a subject was conducted to confirm the practical effectiveness of the algorithm. A motion capture device by Phasespace Inc [7] was used, and its seven active LED markers are attached on an index finger. The system identification stage and the estimation stage were applied to the motion capture data (see the video [24] and Fig. 5).

1) System identification: While the subject moves his finger freely, 20 measurement sets with no occlusion were

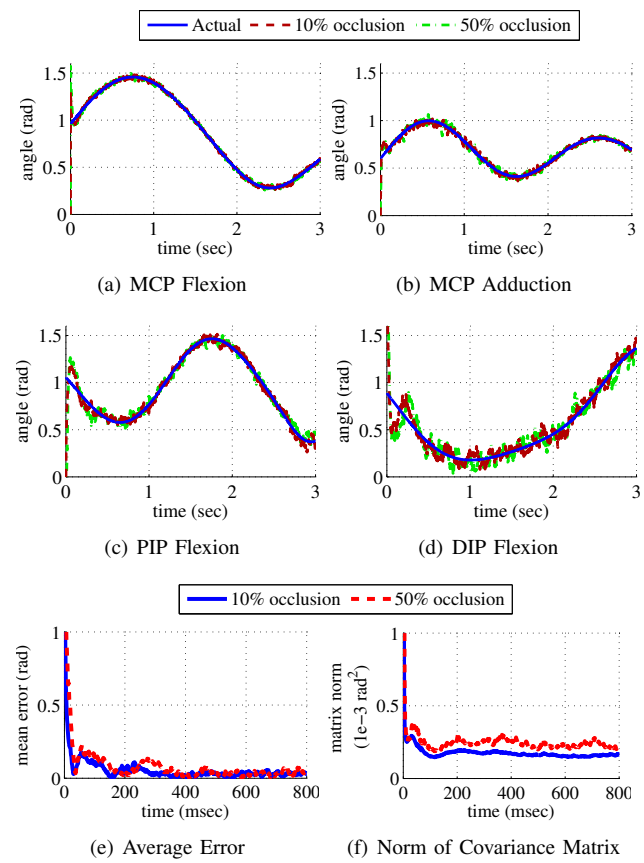


Fig. 4: Results of state estimation in the simulation. A virtual finger moved along a predetermined arbitrary trajectory, and a virtual motion capture system provided the positions of markers with white noise whose standard deviation is 3 mm. The estimation was performed twice with 10 % occlusion and then 50 % occlusion cases. (a)-(d) show the tracking results for the four joint motions. (e) demonstrates the tracking error averaged over four joint poses. (f) shows the size of covariance matrix via the second norm.

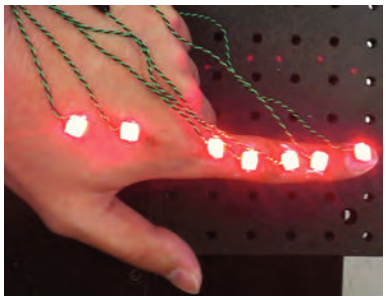


Fig. 5: We attached seven active markers on a subject's index finger and a motion capture system (Phasespace Inc. [7]) provides the three dimensional positions of markers at 480 Hz.

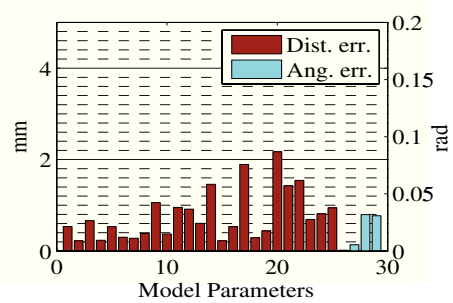


Fig. 6: Results of system identification in the experiment. The size of bars indicates the standard deviation of optimized kinematic model parameters.

collected. Then, the optimization algorithm investigated the solution. Because we do not know the ground-truth value of parameters, we repeated the system identification stage 100 times and calculated the standard deviation of the optimized solution. Fig. 6 demonstrates the consistency of the optimized solutions with the small size of standard deviations.

2) State Estimation: The states were estimated based on the kinematic model parameters obtained in the system identification stage. The estimation algorithm was implemented in C++ (RTAI Linux) to ensure hard-real-time performance with a GUI rendering the finger pose in real-time. The computational time for the execution of two prediction steps and one update step was less than 48.2 micro-seconds. It verifies that the algorithm computational speed is sufficiently fast for the control of a human-robot interaction device. Additionally, we built a real-time GUI with QT library and OpenGL to see the estimated pose of the finger and the real finger motion simultaneously as shown in Fig. 1.

In the experiment, first the subject moved his finger at a nominal speed, then after a while a researcher covered markers, and lastly the subject moved his finger very fast. Fig. 7 illustrates the detailed procedure of the experiment. The real-time estimation algorithm successfully tracks the subject's finger pose even though a researcher intentionally disturbed the measurement and the subject generated unpredictable fast motion. Furthermore, when the measurement condition is not reliable, the algorithm increased the size of covariance, thus providing the quantitative level of confidence for the estimation.

V. CONCLUSIONS

We developed an accurate, robust and high-frequency estimation method for the finger pose with a motion capture device. The goal of 'accuracy' was achieved by finding an optimized kinematic model of the system before the estimation. 'Robustness' was obtained by introducing the jerk minimization model as a predictor and by correcting the system states stochastically. 'High-frequency' was accomplished by separating two procedures (system identification and state estimation) and by selecting a fast estimation algorithm, namely EKF. The simulation showed that the methods can identify the accurate system model, and estimate the finger state robustly. An experiment with the human finger verified that the method is practically effective and runs at a high frequency.

ACKNOWLEDGMENT

The authors thank Dongyang Chen and other research colleagues in the Rehabilitation and Neuromuscular Robotics (ReNeu) Lab at The University of Texas at Austin for their assistance with the experiments.

REFERENCES

[1] A. Wege and G. Hommel, "Development and control of a hand exoskeleton for rehabilitation of hand injuries," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3046–3051, 2005.

[2] T. Worsnopp, M. Peshkin, J. Colgate, and D. Kamper, "An actuated finger exoskeleton for hand rehabilitation following stroke," in *IEEE International Conference on Rehabilitation Robotics*, pp. 896–901, 2007.

[3] P. Stergiopoulos, P. Fuchs, and C. Laurgeau, "Design of a 2-finger hand exoskeleton for vr grasping simulation," *Eurohaptics, Dublin, Ireland*, pp. 80–93, 2003.

[4] M. J. Lelieveld and T. Maeno, "Design and development of a 4 DOF portable haptic interface with multi-point passive force feedback for the index finger," in *IEEE International Conference on Robotics and Automation*, pp. 3134–3139, 2006.

[5] H. Koike, Y. Sato, and Y. Kobayashi, "Integrating paper and digital information on enhanceddesk: a method for realtime finger tracking on an augmented desk system," *ACM Transactions on Computer-Human Interaction*, vol. 8, no. 4, pp. 307–322, 2001.

[6] P. Agarwal, P.-H. Kuo, R. R. Neptune, and A. D. Deshpande, "A novel framework for virtual prototyping of rehabilitation exoskeletons," *IEEE International Conference on Rehabilitation Robotics*, 2013.

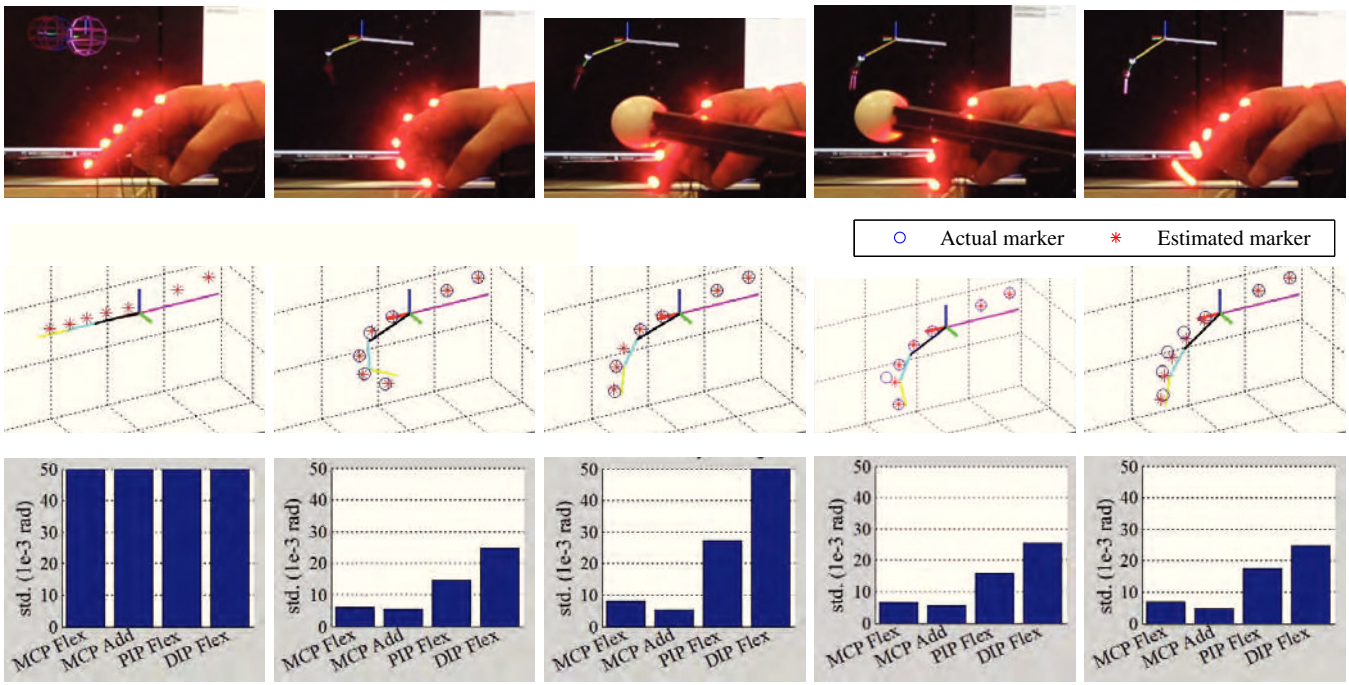
[7] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[8] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.

[9] J. Landsmeer, "The coordination of finger-joint motions," *The Journal of Bone & Joint Surgery*, vol. 45, no. 8, pp. 1654–1662, 1963.

[10] J. H. Ryu, N. Miyata, M. Kouchi, M. Mochimaru, and K. H. Lee, "Analysis of skin movement with respect to flexional bone motion using mr images of a hand," *Journal of biomechanics*, vol. 39, no. 5, pp. 844–852, 2006.

[11] P. Kuo and A. D. Deshpande, "Muscle-tendon units provide limited contributions to the passive stiffness of the index finger metacarpophalangeal joint," *Journal of Biomechanics*, vol. 45, pp. 2531–2538, 2012.



(a) Initial Status (b) Convergence (c) Marker Occlusion (d) Marker Noise (e) Fast Movement

Fig. 7: Estimation results of the experiment. The first row demonstrates the experimental environment. A subject moved his finger freely, and an OpenGL program showed the estimated finger pose behind the subject's motion in real-time. The second row illustrates the estimated marker positions and the actual marker positions to show the accuracy of estimation. The third row demonstrates the uncertainty of the estimated finger joints via standard deviation. (a) At the start, the initial joint angles are zero and their covariance values are very large. (b) Within 0.1 sec, all estimated joint angles converged, and the comparison with actual motion and actual marker position indicates that the algorithm tracks accurately. (c) A researcher occluded some of markers intentionally to test for its robustness. The marker occlusion increased the uncertainty of the estimation, but still the algorithm tracked its actual motion through the predictor. (d) The disturbance made by the researcher produced unrealistic measurement (noise), but the stochastic Kalman gain selectively corrected its state. (e) Lastly, the subject moved his finger very fast. Although the uncertainty of estimation slightly increased, the method tracked the finger pose. (see the attached video [24])

[12] P. Braido and X. Zhang, "Quantitative analysis of finger motion coordination in hand manipulative and gestic acts," *Human Movement Science*, vol. 22, no. 6, pp. 661 – 678, 2004.

[13] N. A. Baker, R. Cham, E. H. Cidboy, J. Cook, and M. S. Redfern, "Kinematics of the fingers and hands during computer keyboard use," *Clinical Biomechanics*, vol. 22, no. 1, pp. 34 – 43, 2007.

[14] X. Zhang, S.-W. Lee, and P. Braido, "Determining finger segmental centers of rotation in flexion–extension based on surface marker measurement," *Journal of biomechanics*, vol. 36, no. 8, pp. 1097–1102, 2003.

[15] P. Cerveri, E. De Momi, N. Lopomo, G. Baud-Bovy, R. M. Barros, and G. Ferrigno, "Finger kinematic modeling and real-time hand motion estimation," *Annals of biomedical engineering*, vol. 35, no. 11, pp. 1989–2002, 2007.

[16] J. Maycock, J. Steffen, R. Haschke, and H. Ritter, "Robust tracking of human hand postures for robot teaching," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2947–2952, 2011.

[17] Q. Fu and M. Santello, "Tracking whole hand kinematics using extended kalman filter," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 4606–4609, 2010.

[18] E. Todorov, "Probabilistic inference of multijoint movements, skeletal parameters and marker attachments from diverse motion capture data," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 11, pp. 1927–1939, 2007.

[19] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *The journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[20] E. Todorov and M. I. Jordan, "Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements," *Journal of Neurophysiology*, vol. 80, no. 2, pp. 696–714, 1998.

[21] K.-N. An, E. Y. Chao, W. Cooney, and R. L. Linscheid, "Normative model of human hand for biomechanical analysis," *Journal of Biomechanics*, vol. 12, no. 10, pp. 775–788, 1979.

[22] Y. Yun, B. Park, and W. K. Chung, "Odometry calibration using home positioning function for mobile robot," in *IEEE International Conference on Robotics and Automation*, pp. 2116–2121, 2008.

[23] M. Celis, J. Dennis, and R. Tapia, "A trust region strategy for nonlinear equality constrained optimization," *Numerical Optimization*, vol. 1984, pp. 71–82, 1985.

[24] Y. Yun, P. Agarwal, and A. D. Deshpande, "Experimental result video," *http://youtu.be/0XcAL-xo3S4*, 2013.