32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Tracking whole hand kinematics using extended Kalman filter

Qiushi Fu, Marco Santello

Abstract—**This paper describes the general procedure, model construction, and experimental results of tracking whole hand kinematics using extended Kalman filter (EKF) based on data recorded from active surface markers. We used a hand model with 29 degrees of freedom that consists of hand global posture, wrist, and digits. The marker protocol had 4 markers on the distal forearm and 20 markers on the dorsal surface of the joints of the digits. To reduce computational load, we divided the state space into four sub-spaces, each of which were estimated with an EKF in a specific order. We tested our framework and found reasonably accurate results (2-4 mm tip position error) when sampling tip to tip pinch at 120 Hz.**

I. Introduction

The human hand is a complex structure and the neural control of its movements is a subject of active investigation. Many research areas, such as biomechanics [1], sensorimotor control [2], teleoperation [3], brain-machine interfaces, and human-computer interaction, all requires reliable and efficient approaches to acquire kinematic data during grasping and manipulation tasks.

There are three major methods for tracking whole hand kinematics: data gloves [3], computer vision based estimation [4], and surface marker based estimation [1], [5]. Data gloves are electro-mechanical sensing devices worn on hand and are easy use. However, there are two main drawbacks: the lack of customization for individual subjects' hands, and the obstruction of tactile sensing from the palmer surface of the hand. Vision based estimation causes the least interference to hand movements and usually only uses 1 or 2 cameras, but the sampling rate is very low because it is computationally expensive and sensitive to environmental conditions.

Surface marker based estimation for tracking hand kinematics has been used extensively [5], as is relatively easy to implement, reliable, and because it does not interfere normal movement and sensing as much as data gloves. However, most of the existing approaches rely on geometrical calculations, or optimizations within single frame capture. Therefore they do not make full use of established non-linear estimation algorithms for error minimization and robustness.

Although Kalman filter has been implemented in vision-based tracking [6], it has not been investigated for marker based real time hand tracking. This paper proposes a surface maker based high frequency hand kinematic tracking framework using extended Kalman filter (EKF). We tested the EKF framework with a 29 degrees of freedom (DoF) hand model and a 24 marker set. This framework has been implemented in our lab for the research on sensorimotor control of hand.

Manuscript received June 27, 2010. This work was supported in part by a National Science Foundation Grant IIS 0904504.

Q. Fu is with School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287 USA. (phone: 623-262-0769; e-mail: qiushifu@asu.edu).

M. Santello is with School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287 USA. (e-mail: marco.santello@asu.edu).

II. Kinematic Model

A. The hand model

The hand model we used consisted of articulations of rigid links. It has 21 DoF on joints of five digits, 2 DoF on wrist joint, and 6 DoF on hand global posture (Fig. 1a). Note that other similar hand models from the literature [4] could be used to derive the kinematic structures depending on the desired anatomical accuracy. The hand global posture, or root frame of reference (RT) was defined by the special location and orientation with respect to an arbitrary global frame of reference. In the actual experimental setup, RT is defined with four markers attached to the wrist rigidly (see below). The origin of the RT was located at the approximate center of rotation (CoR) of the wrist.

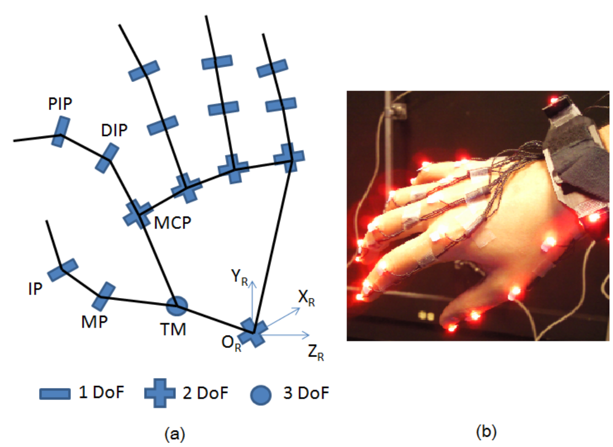

Fig. 1. (a) Kinematic structure of the hand model, and (b) marker protocol.

Consistent with most hand models found in the literature, the palm was modeled as a rigid segment, although limited palm curving may occur [4]. The palm frame of reference was located as the CoR of the wrist with the y-axis pointing to the CoR of the metacarpophalangeal (MCP) joint of the middle finger and the z-axis pointing to the dorsal direction. The palm segment was connected with the root with the 2 DoF wrist joint. The AoR of wrist flexion/extension was oriented from the radial to the ulnar styloid process and the

AoR of radial/ulnar deviation was orthogonal to the flexion/extension axis.

Each of the four fingers had 4 DoF with the MCP joint as a 2 DoF saddle joint capable of abduction/adduction, the proximal interphalangeal (PIP) joint and the distal interphalangeal (DIP) joint as 1 DoF hinge joints. We defined the digit plane as the medial plane perpendicular to the flexion/extension axis.

Modeling of the thumb is difficult due to the complex anatomical features of the trapezio-metacarpal (TM) joint. Biomechanical studies have shown that the TM joint has two non-orthogonal and non-intersecting axes of rotation (AoR) and its kinematics is sensitive to the parameters of the model [7]. In this work, we were not able to estimate the two AoRs due to our marker arrangement. Instead, we used a 3 DoF spherical TM joint to preserve the complete information of the thumb orientation. The order of the 3 AoR of this spherical joint was arranged such that the last AoR was parallel to the AoR of the 1 DoF MCP joint interphalangeal (IP) joint. This allowed modeling all flexion/extension of the thumb in a digit plane defined the same as for other digits.

B. Marker Protocol

A marker set that consisted of 24 markers distributed on digits and wrist (Fig. 1b) was used. Specifically, we adopted the bracelet used by [5]. It was attached to the wrist with four markers fixed at the vertices of a known rectangle to provide the RT. The markers were arranged such that the center of the bracelet is approximately aligned with the CoR of the wrist as shown in Fig. 2a, and the long side of the rectangle was aligned with the flexion/extension axis of the wrist.

Each of the four fingers was marked with three markers located on the dorsal surface of MCP, PIP, and DIP joints, as well as the fourth marker on the finger nails (finger tip, FT). They were carefully adjusted to lie within each digit plane. The thumb also had four markers and they were placed on the dorsal surface of the TM, MCP, and IP joints as well as the thumb nail within the thumb plane. Each non-FT marker on the hand is assumed to move along a spherical surface centered at the CoR of the corresponding joint with the radius d_i measured as the joint thickness.

A clear correlation between marker motion around the joint AoR and the actual flexion/extension of the joint has been shown by [1]. This correlation is caused by the skin stretch around the joint CoR during the joint motion. Therefore our model assumes a linear relationship $\alpha_i = \theta_i / 2$ between the movement of the markers and the motion of the joints within the digit plane as shown in Fig. 2b, where α_i is the rotation of marker along the surface of the MCP, PIP or DIP joint within digit plane and θ_i is the actual rotation of the these joints. FT markers were assumed to follow the motion of distal phalanges.

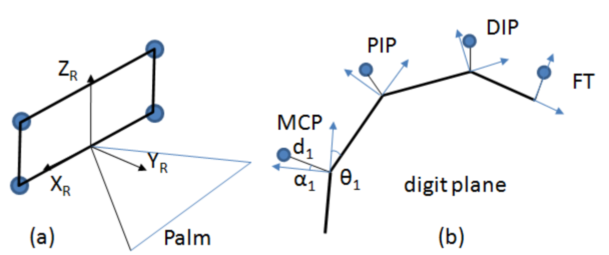

Fig. 2. Illustration of marker model: (a) markers placed on bracelet for tracking hand global posture, and (b) markers placed on the joint surface for tracking of digit movements.

C. Model Calibration

Since we cannot directly measure the link length, joint CoR and AoR, calibration is necessary before implementing any model based estimation. The wrist CoR is crucial for good estimation and we used a least square method [8] to estimate it based on data from a calibration trial in which subjects performed wrist movement across the range of motion. For the other joints, we adopted the geometrical approach described in [5] for our calibration since we used a similar marker protocol and hand model except for the thumb TM joint. The required measurements were joint thickness of MCP, PIP, DIP, IP, and TM joints, as well as the thickness of the FT. Then the other parameters of the model could be derived geometrically by collecting data from calibration trials. For each frame:

1. Calculate the root frame of reference from the wrist markers;
2. Fit the digit plane using 4 markers of each digit and project the markers into corresponding digit planes;
3. Calculate the orientation of the articulations based on the new marker positions to find joint AoR and displace the maker position by half of the joint thickness to find the joint CoR;
4. Calculate link length using joint CoR.

The model parameters can be computed by taking the mean of the parameters derived for each frame. The repeatability and accuracy of the calibration procedure were tested and were consistent with [5].

III. Implementation of Extended Kalman Filter

A. Extended Kalman Filter

Estimating hand postures with the marker position information is essentially a non-linear estimation problem in which the joint angles (as well as the root position) of the hand is the state and the 3D marker positions are observations. Extended Kalman filter is the non-linear version of Kalman filter and has been successfully implemented in many situations [9]. Let the nonlinear estimation problem be

$$\begin{aligned} &\text{State transition model:} \quad \mathrm{x}_k = f(\mathrm{x}_{k-1}, \mathrm{u}_{k-1}) + \mathrm{w}_{k-1} \\ &\text{Observation model:} \quad \mathrm{z}_k = h(\mathrm{x}_{k-1}) + \mathrm{v}_k \end{aligned} \tag{1}$$

where w_k and v_k are process and observation zero mean Gaussian noise with covariance matrices Q_k and R_k

respectively. The EKF linearizes about the current mean and covariance and recursively estimate the states while minimize *a posteriori* estimation error covariance $P_{k|k}$ with the following structure

$$
\begin{aligned}
\text{Predict:}\quad & \hat{x}_{k|k-1} = f(\hat{x}_{k|k-1}, u_{k|k-1}) \\
& P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^{T} + Q_{k-1} \\
\text{Update:}\quad & \tilde{y}_k = z_k - h(\hat{x}_{k|k-1}) \\
& K_k = P_{k|k-1}H_k^{T}(H_k P_{k|k-1} H_k^{T} + R_k)^{-1} \\
& \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \\
& \hat{P}_{k|k} = (I - K_k H_k) P_{k|k-1}
\end{aligned}
\tag{2}
$$

where P is the estimation error covariance, K_k is the Kalman gain, F and H are Jacobians defined as

$$
F_{k-1} = \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_{k-1|k-1}, u_{k-1}}, \quad H_k = \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_{k-1|k}} \tag{3}
$$

B. EKF tracking procedure

We divided the state space into four subspaces and estimated them using four separate EKFs to reduce the computational load of large (29 DoF) state space. The four subspaces are RT space (x_R^k, 6 states), index-middle-wrist space (x_{IMW}^k, 10 states), thumb space (x_T^k, 5 states), and ring-little space (x_{RL}^k, 8 states). We also assumed a static state transition model in which the posture of the hand at time *k* is the posture of the hand at time *k*-1 superimposed by a random small motion that is modeled with zero mean Gaussian noise. Here we present the RT tracking as an example. Let the root state vector be x_R^k and the marker position vector be z_R^k, the state transition model, and observation model in eq. (2) become

$$
\begin{aligned}
x_R^k &= x_R^{k-1} + w_{k-1} \\
z_R^k &= h(x_R^{k-1}) + v_k
\end{aligned}
\tag{4}
$$

where h is nonlinear with trigonometric functions that transform joint angles into Cartesian space positions. w_k has a diagonal covariance matrix with the diagonal elements derived from the assumption that the state vector undergoes a independent velocity with normal distribution of zero mean and standard deviation vector σ_R. σ_R could be tuned according to the estimated speed of hand movement during the upcoming task. The covariance matrix of v_k was hand tuned to represent the marker placement error and skin stretch.

For the other three sub state spaces, the state transition model was formulated in the same way. However, we used transformed marker positions as the observation z_k instead of raw marker positions. The four EKF worked in particular order at each time step, as shown in Fig. 3. We first estimated the RT states x_R^k, then constructed the homogeneous transformation T_R^k such that all digit markers can be expressed in the root frame of reference. Then, the EKF for state space of index, middle fingers and the wrist x_{IMW}^k were always working in the root frame of reference. Lastly, the markers of thumb, ring, and little fingers were further expressed in the palm frame of reference using homogeneous transformation T_W^k constructed from wrist state, such that the kinematic structure of thumb, ring, and little fingers can be estimated using EKF.

All model parameters were calculated in the calibration trials. At the first frame, the initial values of the states were estimated using the same geometrical technique described in the model calibration section.

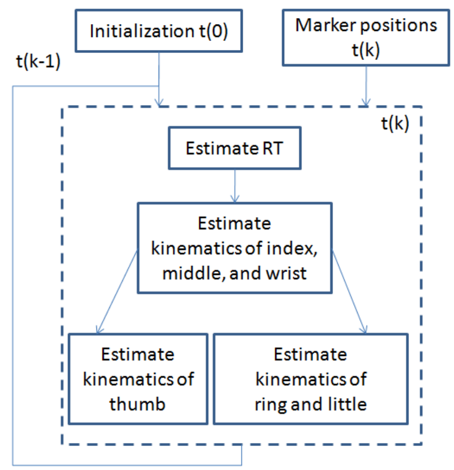

Fig. 3. Flowchart of the EKF tracking procedure.

C. Missing markers

Hand motion may occasionally produce maker occlusions. In this work, we took care of missing marker data at time step *k* by using the state estimations made at time step *k*-1 and *k*-2, assuming the states changed the same amount from time *k*-1 to time *k* as from time *k*-2 to *k*-1. Other methods could also be used. For instance, we can change the *h* function in eq. (4), thus the rank of the corresponding Jacobian matrix.

D. Experimental Setup

In our actual implementation, we used a motion tracking system with active markers (PhaseSpace Inc., San Leandro, CA, U.S.). The marker positions were sampled at 120Hz and automatically labeled by the system. The joint thickness was measured using a digital caliper. The EKF scheme was implemented in MATLAB as post processing. However, we recorded the computation time for each step to show its capability of real time implementation. The PC we used in this study was Intel Core 2 Duo 2.26GHz, and 3GB RAM.

Three subjects participated the experiment. Each subject performed 1 calibration session, and 2 different tasks. All the trials started with the hand wrapping and resting naturally on a spherical handle. In the calibration session, subjects were instructed to move hand as well as all the joints at self selected speed for 30 seconds. In the first task, subjects were instructed to do tip to tip pinch, that is, using tip of the thumb to touch lightly on each of the tips of the other digits for about 1 second (Fig. 4). This task was repeated 5 times for each finger (therefore 25 contact events for each subject).

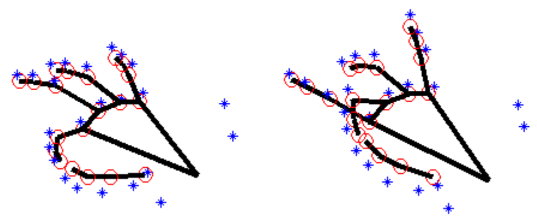

Fig. 4. Snapshot of subject performing tip to tip pinch, blue stars and red circles represent markers and joint CoRs, respectively.

In the second task, the subjects were instructed to reach and grasp 5 different objects (scissor, jar lid, key, pen, and bottle). Each of the objects was grasped 5 times (25 grasps for each subject).

IV. Experimental Results

A. *Tracking performance*

To evaluate the performance of the EKF tracking approach, we present the results of the two tasks. In the tip to tip pinch task, we computed the distance between the two tips D_e, and compare it with the half of the sum of the FT thickness D_a for each contact event. This is under the assumption that FT has a spherical surface. The results (Mean ± SD) from each subject are shown in Table I. It can be seen that the computed errors were only 2-4 mm. Furthermore, we observed an increasing trend of error from the index to the little finger. This can be explained by our hand model which did not account for the palm curving.

In the grasping task, we show the grasping posture from one representative subject in Fig. 5. The results reasonably reconstructed the actual posture, therefore suggesting the application of EKF based tracking in hand kinematic

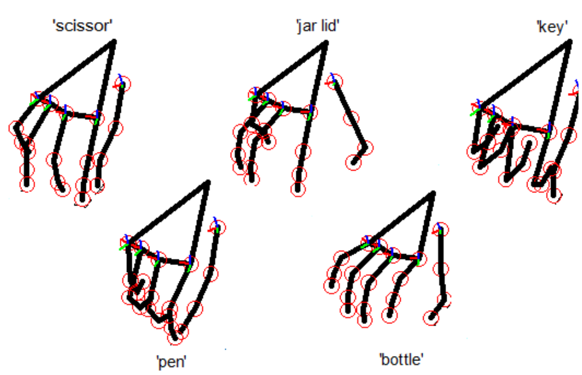

Fig. 5. Snapshots of subject performing different grasps.

analysis. Overall, the consistency of both small tip to tip error and postural tracking across subjects indicated that the performance of the EKF based whole hand tracking is quite well.

B. *Computation time*

The mean computation time for updating one time step is 4.36 ms. This indicates that the proposed EKF estimation can be safely implemented in a real time system that samples the hand posture at up to 200 Hz.

TABLE I
Errors of Finger Tip Estimation (Mean ± SD, mm)

Finger in contact	Subject 1	Subject 2	Subject 3
index	1.9±0.2	2.0±0.1	2.3±0.1
middle	2.1±0.1	1.8±0.2	2.0±0.1
ring	3.6±0.1	3.5±0.2	2.9±0.3
little	4.2±0.2	3.8±0.3	3.5±0.2

V. Conclusion and Future work

We presented a framework of tracking whole hand kinematics in real time with high sampling rate using extended Kalman filter. The hand model and marker protocol was described, as well as the calibration and computational structure. Future work includes testing and comparison of different hand models and marker protocols, optimization of the computational structures, and investigation of other non-linear estimation techniques.

References

[1] X. Zhang, S. Lee, and P. Braido, "Determining finger segmental centers of rotation in flexion–extension based on surface marker measurement," *J. Biomech.*, vol. 36, pp. 1097-1102.

[2] J. Lukos, C. Ansuini, and M. Santello, "Choice of contact points during multidigit grasping: effect of predictability of object center of mass location.," *J. Neurosci.*, vol. 27, pp. 3894-903.

[3] W. Griffin, R. Findley, M. Turner, and M. Cutkosky, "Calibration and mapping of a human hand for dexterous telemanipulation," *ASME IMECE Symp. on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, 2000.

[4] 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, pp. 52-73.

[5] 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 Biomed. Eng.*, vol. 35, pp. 1989-2002.

[6] B. Stenger, P. Mendonça, and R. Cipolla, "Model-based hand tracking using an unscented kalman filter," *Proc. British Machine Vision Conf.*, 2001, pp. 63–72.

[7] V. Santos and F. Valero-Cuevas, "Reported anatomical variability naturally leads to multimodal distributions of Denavit-Hartenberg parameters for the human thumb," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 155–163.

[8] L. Y. Chang, and , N. S. Pollard, "Constrained least-squares optimization for robust estimation of center of rotation," *J. Biomech.*, 40(6), 1392-400.

[9] H. Su, L.Dickstein-Fischer, K. Harrington, Q. Fu, W. Lu, H. Huang, G. Cole and G.S. Fischer, "Cable-Driven Elastic Parallel Humanoid Head with Face Tracking for Autism Spectrum Disorder Interventions", In Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina, 2010