3D Hand and Fingers Reconstruction from Monocular View

1. Research Team

Project Leader: Prof. Isaac Cohen, *Computer Science*

Graduate Students: Sung Uk Lee

2. Statement of Project Goals

The needs for an accurate hand motion tracking have been increasing in HCI and applications requiring an understanding of user motions. Hand gestures can be used as a convenient user interface for various computer applications in which the state and the action of the users are automatically inferred from a set of video cameras. The objective is to extend the current mouse-keyboard interaction techniques in order to allow the user to behave naturally in the immersed environment, as the system perceives and responds appropriately to user actions. HCI related applications remains the most frequent however; we have focused our effort on the accurate detection and tracking of un-instrumented hands for assessing user performance in accomplishing a specific task.

3. Project Role in Support of IMSC Strategic Plan

Multimodal communications rely on speech, haptics and visual sensing of the user in order to characterize the user's emotional and/or physical state. The proposed research focuses on augmenting the interaction capabilities of a multimodal system by providing an automatic tracking of hands motion and recognizing the corresponding hands gestures. The objective here is to augment or replace the mouse and keyboard paradigm with functionalities relying on natural hands motion for driving a specific application.

4. Discussion of Methodology Used

Inferring hand motion from monocular view is challenging, as the detected 2D silhouette is not sufficient to depict a complicated hand pose. Therefore, several studies introduce respective hand model and model-based approach. Although these methods can track the articulated hand motion from monocular view, many of them are computationally too expensive and have difficulties in addressing hand's self-occlusions. We revised the articulated hand model proposed previous year to deal with natural hand motion constraints, and it allows tracking detected hand efficiently in the presence of self-occlusions and has a low computational cost. After the segmentation of the hand silhouette from the video stream, an articulated hand model is fitted onto the 2D mask of a hand. For initialization, 3D hand model configuration such as joint offset and size are refined by the favorable stretched-out pose. Section 4.2 describes the initialization process in detail. First joints connecting palm and each finger form a rigid shape. Once a 3D model is refined, global hand pose is computed using this rigid form constraint and then estimation of each finger motion is performed. To reconstruct accurate pose, 3D model is updated by re-projecting it on 2D detected silhouette in the tracking process. An overview of the proposed approach is shown in figure 1.

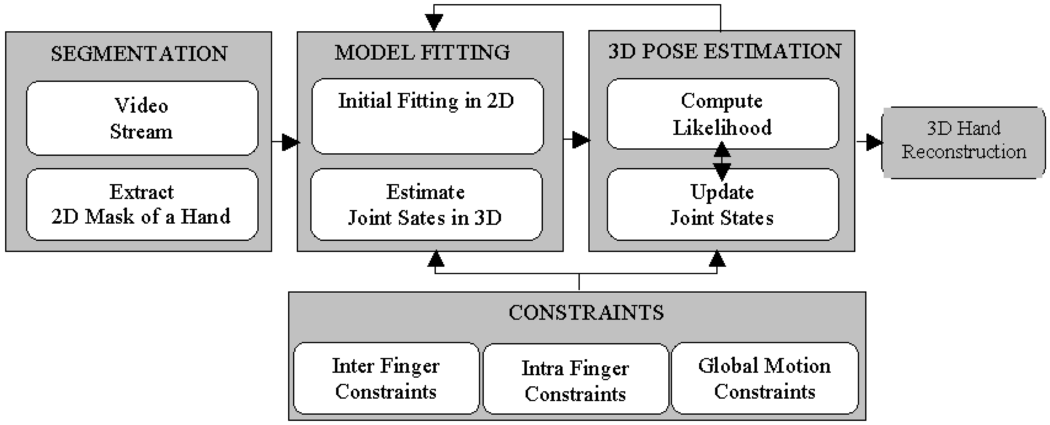

Figure 1. Overview of the proposed approach.

4.1 Modeling Hand Constraints

Without any constraints, it is a very exhaustive task to analyze human hand motion. However, joints in a hand are interrelated and their motions are related by physical constraints. Huang et al [4] showed in detail that human hand motion is highly constrained. A learning approach was proposed to model the hand configuration space directly sensed from a CyberGlove. In this project, three types of constraints are defined for natural hand motion. Figure 2 shows the articulated hand model proposed to model hand motion constraints.

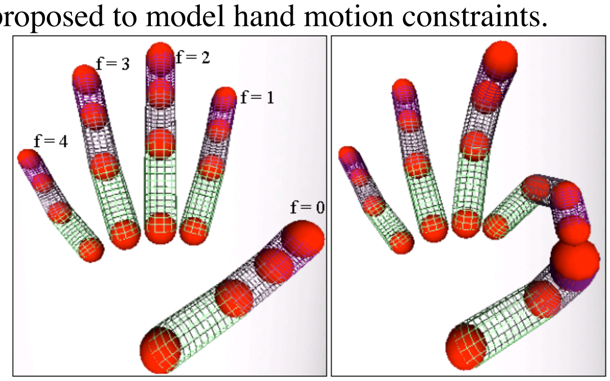

Figure 2. The articulated hand model. Each joint j_i^f is a revolute joint where *f* is the index of fingers enumerated from thumb to little finger (*f = 0 ~ 4*). All the first joints (where i = 0) connecting palm and each finger have *2* degrees of freedom and other joints have a single rotational axis. Therefore, the suggested model has 20 DOF.

Intra-Finger Constraints. In this category, the main constraint considered is of finger joints during a bending. The number of rotational axes in a joint is constrained. By using the planar 3-joint-arm finger model, DOF of a finger is reduced to 4. j_0^f, the first joint of a finger, has 2 rotation axes – one for bending motion and another is for abduction/adduction motion – but other joints have single rotation axis. Table 1 shows the range of each joint angle for bending motion.

Table 1. Joint angle constraints for bending motion

j_i^f	*i = 0*	*i = 1*	*i = 2*
f = 0	-10° ~ +45°	+0° ~ +90°	-5° ~ +90°
f = 1~4	-15° ~ +90°	-5° ~ +90°	-5° ~ +90°

Each contiguous joint has the following constraint:

$$\Delta j_i^f = \alpha_i \left| \Delta j_{i-1}^f \right| \quad (1)$$

where finger index $f = 1$ to 4 and joint index $i = 1, 2$. λ_i is defined by (limit angle of j_i^f) / (limit angle of j_{i+1}^f). This type of constraint is not strictly applied, but is an efficient guide of joint offset estimation.

Inter-Finger Constraints. Proposed model has fixed geometric proportion among j_0^f (offsets of all first joint of a finger). No joint can overlap each other in 3D coordinates. Interrelated motion constraints for abduction/adduction motion between adjacent fingers are defined in this category. Adjacent finger tends to follow the dominant articulated motion.

Global Motion Constraints. This category includes some restrictions on target motion. Tracking objective is natural hand motion, begins with favorable pose for initialization. No external torque on a joint is expected. Motion should be smooth. By these constraints, we address the problem of tracking articulated hand motion with self-occlusion, rotation and translation, but the proposed approach does not address the scaling and depth estimation from monocular view. Some recent studies introduce similar constraints to the above mentioned ones. However, they are hardly modeled as an explicit form of equations.

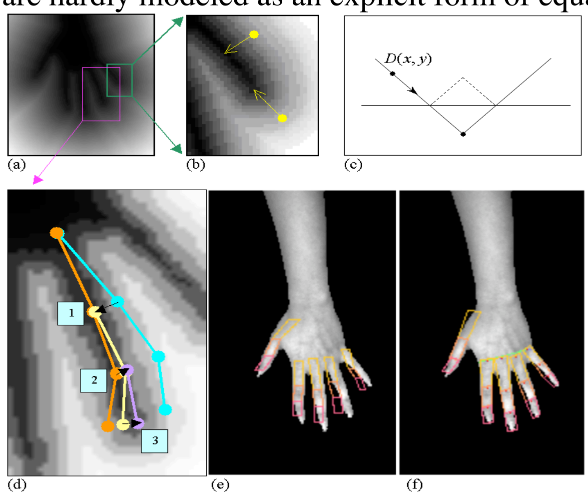

Figure 3. Model initialization. (a) Distance map. (b), (c) Approaching median axis (minimum distance) along the gradient of distance map. (d) Joint position refinement. (e) Coarse 2D model with ellipsoid fitting. (f) An example after refinement.

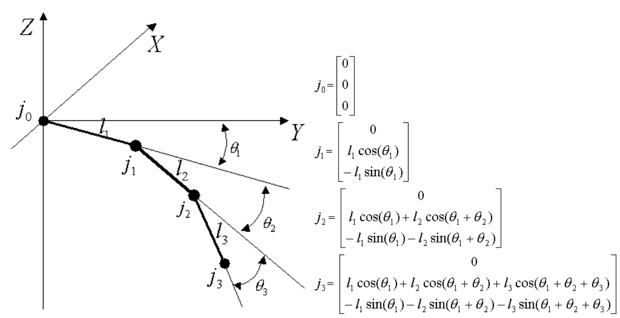

Figure 4. Kinematics finger model

4.2 Hand Model Fitting with Hand Constraints

A sequence of hand motion begins with a favorable initial pose as shown in figure 3. By using an ellipsoid fitting [12] on top of the segmented hand silhouette, brief scale and orientation of a hand are computed. Borgefors' two-pass algorithm [12] is used to compute the distance map *d(x,y)* of a hand image for the model refinement.

$$D(x,y) = \begin{cases} -d(x,y), & (x,y) \in \textit{Silhouette} \\ d(x,y), & \textit{Otherwise} \end{cases} \quad (2)$$

By minimizing the distance *D(x,y)* along the gradient $\nabla D(x,y)$, each joint offset can be refined. Each joint width (thickness) can be also estimated with *D(x,y)*. Figure 3 shows an example of model initialization in 2D. Based on this information, the initial configuration of 3D hand model is refined.

Using inter-finger constraints and global motion constraints, a hand motion can be divided into global pose and individual finger motion. At time *t*, first joints of j_0^0, j_0^1, and j_0^4 form a palm-triangle *PT(t)* which has three sides of w_t^0, w_t^1 w_t^2. In the two consecutive palm-triangle *PT(t)* and *PT(t+1),* the proportion of corresponding sides of each palm-triangle provides the global pose. By the global motion constraint, j_0^0 is the origin of a hand configuration. Thus, the translation vector T_g can be estimated by two consecutive 2D position of j_0^0 detected from hand silhouette. By the rigid motion constraint, w_t^p / w_{t+1}^p gives the angle for rotation vector R_g. In Figure 5, a palm-triangle of bright violet shows the estimation results of global pose.

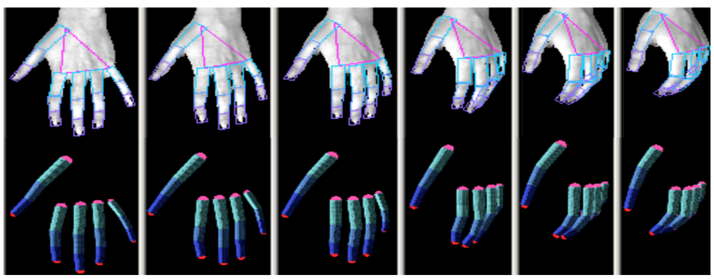

Figure 5. 3D hand reconstruction 1: 1st row show the result of the global pose estimation with palm-triangle. Re-projected hand model 2nd row shows the reconstructed 3D hand model.

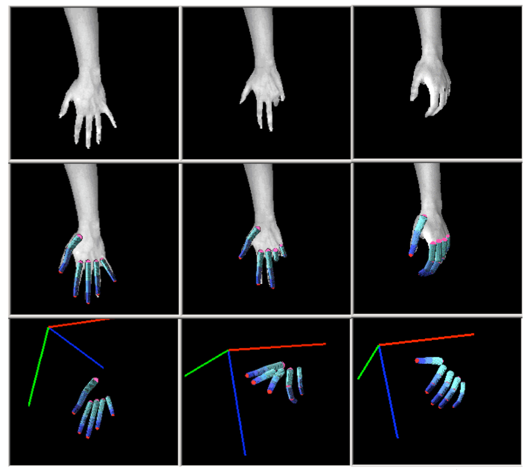

Figure 6. 3D hand reconstruction 2: 1st Row shows the original hand pose. 2nd row represents the reconstructed pose. 3rd row is different angle view.

After estimating the global rotation and translation [R_g, T_g], each joint angle of a finger can be estimated from the equations in Figure 4. Each length of links in a finger was fixed in the initialization. Therefore, joint angle θ_i for the bending motion is explicated. The proposed tracking technique is based on '*tracking by synthesis*'. To maximize the likelihood, possible sets of configuration is tested. The adjustment of a hand model is guided by the motion constraints and the configuration of previous frame. Disparity vector of the closest boundary between the model segment and the detected silhouette is computed for estimating the position of a joint angle. The adjusted 3D hand configuration is projected back onto the 2D image and updated from 2D image measurements back and forth. Figure 5 shows some results of motion tracking. Reconstructed 3D hand model and its re-projected 2D model are correctly matched with the detected silhouette. Various hand poses including self-occlusion and the corresponding results are shown in Figure 6.

5. Short Description of Achievements in Previous Years

The 3D shape of the hands is reconstructed from 2D hands' silhouettes extracted from two calibrated cameras using GC-based approach. Hand pose information can be computed by matching these synthesized shapes to pre-defined hand model or learning hands' postures using a 3D shape description. The previous framework we have developed for 3D body reconstruction

[14] has been modified for 3D hands reconstruction. The detected silhouettes are segmented into separate fingers: we compute edge information in the segmented hand region (inside the silhouettes) to reconstruct each finger as a cylindrical form. This local edge analysis allows us to reconstruct the 3D hand shape accurately when the fingers are splayed or when they are not.

5a. Detail of Accomplishments During the Past Year

The goal of this project is to understand and reconstruct human hand motion with natural articulation and occlusion in real time. The objective of this research effort is to augment or replace the mouse and keyboard paradigm with functionalities relying on natural hands motion for driving a specific application. During this year we have focused on inferring accurate hand motion from monocular view. An articulated model is fitted accurately by the proposed model refinement technique and global pose of the hand can be computed by characterizing the orientation of the palm. Individual finger motion is modeled and tracked as a planar arm. We have mainly addressed the problem of tracking the simultaneous motion of rotation, translation, and self-occlusion in monocular view. Proposed approach is easy to extend for real-time and multi-view framework, although the target motion is somewhat restrictive.

6. Other Relevant Work Being Conducted and How this Project is Different

A large number of studies have been proposed for capturing human hand motion. Some of them have focused on very specific domain and thus have a large number of restrictive assumptions. In [9.10.11], trajectories of hand motion are considered for classifying human gestures and are sufficient for the studied domain. Several frameworks have been proposed we can distinguish two types: appearance-based and model-based methods. In [7] an appearance-based method is proposed for characterizing the various hand configuration and finger pose. This relies on a learning-based approach and requires training the algorithm on a large data set that depicts all possible configurations. Similarly in [8], articulated pose of a hand is formulated as a database indexing problem. From a large set of synthetic hand images, the closest matches of an input hand image are retrieved as an estimated pose.

Model based approaches rely on the physical properties of the hand and finger motion [1.2.3] in order to infer it from a single or multiple views. Huang et al. [5] proposed a model-based approach that integrates constraints of natural hand motion. They also proposed a learning approach to model the kinematics constraints directly from a real input device such as the CyberGlove [4]. They focused on the analysis of local finger motion and constraints from observed data. For articulated motion tracking, a Sequential Monte Carlo algorithm based on importance sampling is employed. In a fixed viewpoint, this produces good results; however, global hand position and motion are not dealt with.

Recently, some studies of integrating various cues such as motion, shading and edges were proposed. Metaxas et al [6] computes the model driving forces using gradient-based optical flow constraint and edges. A forward recursive dynamic model is introduced to track the motion in response to 3D data derived forces applied to the model. They assumed that global hand motion is rigid motion. No simultaneous articulated motion is expected. When the fingers are in motion of significant relative motion with occlusions, they hardly drive the hand model.

7. Plan for the Next Year

We plan to extent the capabilities of the current system to multi-view framework. We will also focus the development of an efficient joint angle tracking technique for both of the real-time finger motion tracking and the hand localization in 3D.

8. Expected Milestones and Deliverables

For the next five-years, we will focus on expanding current method for recognition of hands gestures, as well as tracking hands motion of a person manipulating objects.

2004-2005

- Articulated model fitting and tracking
- Efficient tracking technique for real time performances

2005-2006

- Gesture recognition from the inferred articulated model
- Computer interaction using basic hand gestures.

2006-2008

- Multimodal interaction using hand motion

9. Member Company Benefits

We are investigating two potential applications: manipulation of 3D objects displayed on a 3D LCD and evaluation of handwriting deficiencies. We initiated a collaboration with occupational therapy specialists that identify this technology to be very important for characterizing handwriting deficiencies among persons afflicted with ADD syndrome. A potential use of the technology for monitoring operators performing maintenance tasks on expensive machinery.

10. References

[1] J.M. Rehg and T. Kanade, "Model-Based Tracking of Self-Occluding Articulated Objects", *ICCV95*

[2] Q. Delamarre and O. Faugeras, ”Finding pose of hand in video images: a stereo-based approach”, *In IEEE International Conference on Automatic Face and Gesture Recognition*, Japan, 1998

[3] B. Stenger, P. R. S. Mendonça, and R. Cipolla, “Model-Based 3D Tracking of an Articulated Hand”, *In CVPR*, Volume II, pages 310-315, Kauai, USA, December 2001.

[4] J. Lin, Y. Wu and Thomas S. Huang, "Modeling Human Hand Constraints", *In. Workshop on Human Motion* , Austin, Dec., 2000.

[5] Y. Wu, J. Lin and T. S. Huang, "Capturing Natural Hand Articulation", *in ICCV’01*, pp.426-432.

[6] S. Lu, G. Huang, D. Samaras and D. Metaxas, “Model-based Integration of Visual Cues for Hand Tracking” In Proc. *WMVC 2002*

[7] R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff, "3D Hand Pose Reconstruction Using Specialized Mappings" *IEEE ICCV*. Canada. Jul. 2001.

[8] V. Athitsos and S. Sclaroff, "Estimating 3D Hand Pose from a Cluttered Image", *In CVPR 2003*.

[9] R. Cutler and M. Turk, “View-based interpretation of realtime optical flow for gesture recognition”, *In Proc. Face and Gesture Recognition*, pp. 416–421, 1998.

[10] A. Nishikawa, A. Ohnishi, and F. Miyazaki. “Description and recognition of human gestures based on the transition of curvature from motion images”, *In Proc. Face and Gesture Recognition*, pp. 552–557, 1998.
[11] Y. Xiong and F. Quek, “Gestural Hand Motion Oscillation and Symmetries for Multimodal Discourse: Detection and Analysis”, *In Proc. Workshop on Computer Vision and Pattern Recognition for HCI*. June, 2003.
[12] Intel Corp. “Open Source Computer Vision Library Reference Manual”, 1999
[13] L. Sciavicco, B. Siciliano, “Modeling and control of robot manipulators” MacGraw-Hill, NY, 1996.
[14] Isaac Cohen and Mun Wai Lee, "3D Body Reconstruction for Immersive Interaction", Second International Workshop on Articulated Motion and Deformable Objects Palma de Mallorca, Spain, 21-23 November, 2002.