Capturing Articulated Human Hand Motion: A Divide-and-Conquer Approach

Ying Wu, Thomas S. Huang
Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{yingwu, huang}@ifp.uiuc.edu

Abstract

The use of human hand as a natural interface device serves as a motivating force for research in the modeling, analyzing and capturing of the motion of articulated hand. Model-based hand motion capturing can be formulated as a large nonlinear programming problem, but this approach is plagued by local minima. An alternative way is to use analysis-by-synthesis by searching a huge space, but the results are rough and the computation expensive. In this paper, articulated hand motion is decoupled, a new two-step iterative model-based algorithm is proposed to capture articulated human hand motion, and a proof of convergence of this iterative algorithm is also given. In our proposed work, the decoupled global hand motion and local finger motion are parameterized by 3D hand pose and the state of the hand respectively. Hand pose determination is formulated as a least median of squares (LMS) problem rather than the non-robust least squares (LS) problem, so that 3D hand pose can be reliably calculated even if there are outliers. Local finger motion is formulated as an inverse kinematics problem. A GA-based method is proposed to find a sub-optimal solution of inverse kinematics effectively. Our algorithm and the LS-based algorithm are compared in several experiments. Both algorithms converge when local finger motion between consecutive frames is small. When large finger motion is present, the LS-based method fails, but our algorithm can still estimate the global and local finger motion well.

1 Introduction

In current virtual environment (VE) applications, keyboards, mice, wands and joysticks are the most popular controlling and navigating devices. However, they are inconvenient and unnatural. The use of hand gestures has become an important part of HCI in recent years [1, 13]. In order to use the human hand as a natural interface, some alternatives, such as glove-based devices, are used to capture hand motion by attaching some sensors to the hand to measure the joint angles and the spatial positions of the hands directly. Unfortunately, such devices are expensive and unnatural.

Non-invasive vision-based techniques offer promising alternatives to capture human hand motion with affordable cameras. However, highly articulated human hand motion always presents complex rotation, translation and self- occlusion, which make it very hard to analyze.

Different methods have been proposed to analyze human hand motion. One possible way is to use physical hand shape models [5] which emphasis the deformation of hand shape under the action of various forces, with model motion governed by Newtonian dynamics. Another way to model shape deformation is the statistical hand shape model [5], which learns the deformation through a set of training examples. Mean shape and modes of variation are found using PCA. A hand shape is generated by adding a linear combination of some significant modes of variation to the mean shape. However, accurate estimates of hand poses are hard to obtain by these methods.

In the "analysis-by-synthesis" approach, geometrical hand shape models are used to estimate the hand configuration [8, 16]. Candidate 3D models are projected to the image plane and the best match is found with respect to certain similarity measurement. Essentially, it is a searching problem in very high dimensional space, which makes this approach computational intensive. If the 3D shape model is very precise, an accurate estimation can be obtained. However, those hand models are user- dependent. Rough models can only give an approximate estimation. Another approach is to take advantage of kinematical hand models [14, 10]. The hand can be modeled as a stick figure and hand postures can be calculated by deriving *kinematical Jacobian* [14]. The drawback of this approach is that the optimization is often trapped in local minima.

In this paper, a kinematical hand model is employed. Articulated hand motion is decoupled to its global hand motion and local finger motion, in which global motion is parameterized as the rotation and translation of the palm, and local motion is parameterized as the state of hand. They are captured by a robust pose determination algorithm and an inverse kinematics algorithm, respectively, in a two-step iteration. A convergence theorem is also given in this paper. Least squares (LS) fitting method is non-robust because single outliers can grossly bias estimates[15]. We use robust statistics to formulate hand pose determination problem as a least median of squares (LMS) problem instead of a LS problem. The advantage is that 3D hand pose can be reliably estimated even if there are many outliers. The local finger motion is formulated as an inverse kinematics problem. A GA-based method is proposed to find a sub-optimal solution of the inverse kinematics problem effectively.

Section 2 of the paper describes the model employed. Section 3 decouples the articulated human hand motion. A two-step iterative approach for capturing articulated hand motion is proposed, and a convergence theorem is given next. Then, a robust pose determination algorithm and an efficient GA-based inverse kinematics algorithm are described. Section 4 presents the results of several experiments. Section 5 concludes the paper and outlines some future work.

2 Hand Modeling

As an approximation, the hand can be treated as a set of 16 rigid sub-objects. The structure of hand is defined by the dimension of each of the sub-object and the kinematical relations among them. The hand skeleton can be abstracted as a stick figure so that the dimensions of each sub-object reduces to its *link length*. Each finger can be modeled as a kinematical chain with the palm as its base reference frame. The name of each joint is also indicated in Figure 1.

Since each link is modeled in its local frame, coordinate transformations should be found to transform the model from its local frame to other frames, especially the base reference frame. Fingertips are like the end-effectors of the kinematical chains, which can be written as:

$$\mathbf{x}^b = \mathbf{H}_0^b(\theta_{MCP\_AA})\mathbf{H}_1^0(\theta_{MCP})\mathbf{H}_2^1(\theta_{PIP})\mathbf{H}_3^2(\theta_{DIP})\mathbf{x}^3 \quad (1)$$

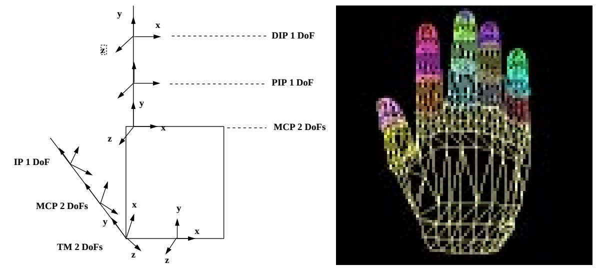

Figure 1: Fingers are modeled by kinematical chains (left). A generic free- form hand model used in this paper (right)

where $\mathbf{x}^3$ is the fingertip in the DIP frame, while $\mathbf{x}^b$ is the fingertip in the base frame. $\mathbf{H}_i^j$ is the coordinate transformation which transforms frame i to frame j. From equation 1, fingers present highly articulated local motion which can be written as:

$$\mathbf{M}_L = \mathbf{M}_L(\theta(t)) \quad (2)$$

where θ is the set of joint angles, and t represents time.

However, it is insufficient to treat the hand as a set of kinematical chains. Some constraints such as $0 \leq \theta_{MCP} \leq 90^0$, $-15^0 \leq \theta_{MCP\_AA} \leq 15^0$ and $\theta_{DIP} = \frac{2}{3}\theta_{PIP}$ are also integrated in our generic hand model. This generic model should be calibrated for different people to derive user-specific models. This step is done visually by warping the general hand model to different persons with two images of front view and side view.

3 Divide-and-Conquer Approach

Highly articulated human hand motion consists of the global hand motion and the local finger motion, which can be represented as $\mathbf{M} = \mathbf{M}(\mathbf{M}_G, \mathbf{M}_L)$, where $\mathbf{M}$ is the hand motion, $\mathbf{M}_G$ is the global motion and $\mathbf{M}_L$ is the local motion. The global hand motion that presents large rotation and translation can be written as $\mathbf{M}_G = \mathbf{M}_G(\mathbf{R}, \mathbf{T})$, where $\mathbf{R}$ and $\mathbf{T}$ are rotation and translation respectively. The local hand motion can be parameterized with a set of joint angles (or *state of the hand*), $\mathbf{M}_L = \mathbf{M}_L(\theta)$ where θ is the joint angle set. Thus, the hand motion can be represented as:

$$\mathbf{M} = \mathbf{M}(\mathbf{R}, \mathbf{T}, \theta) \quad (3)$$

Hand features under motion can be represented as:

$$\mathbf{x}^w = M(\mathbf{R}, \mathbf{T}, \theta)\mathbf{x}^m \quad (4)$$

where $\mathbf{x}^w$ is a model point in the world reference frame, and $\mathbf{x}^m$ is its coordinate in its local model frame. Combining this with the projection model, we have:

$$\mathbf{X} = \mathcal{F}(P(cam),\ M(\mathbf{R},\mathbf{T},\ \ \theta),\ \mathbf{x}^m) \qquad (5)$$

where $\mathbf{X}$ is the 2D projection of $\mathbf{x}^m$, and $\mathcal{F}$ describes the projection of the hand model under the articulated motion. Specifically, $\mathbf{x}^m$ can be fingertips. In our proposed work, the camera is calibrated so that P is known.

Given a 3D hand model and the corresponding image features, capturing of articulated hand motion can be formulated as:

$$(\mathbf{R},\mathbf{T},\ \theta) = arg \ \min_{\{R,T,\ \theta\}} \mathcal{D}(\mathbf{X}, \mathcal{F}(M(\mathbf{R},\mathbf{T},\ \theta), \mathbf{x}^m)) \qquad (6)$$

where $\mathcal{D}$ is a similarity measurement between the observed image feature $\mathbf{X}$ and the projection of the 3D model $\mathbf{x}^m$ under articulated motion $\mathbf{M}(\mathbf{R},\mathbf{T},\ \theta)$.

The articulated hand motion can be treated as a whole, in which all motion parameters are estimated at the same time by nonlinear programming techniques [14] or greedy search and "analysis-by-synthesis" [8]. However, since the optimization space is huge, the search is easily trapped in local minima. We propose a divide-and-conquer approach, in which the global motion $\mathbf{R},\mathbf{T}$ is estimated by a robust pose determination algorithm, and the local motion is estimated based on the hand pose.

3.1 Two-step iterative algorithm

The human hand is a special articulated object, because the role of the fingers in motion is much different from the palm. Based on the observation that the global hand motion is represented by the position and orientation of the palm, we define the global hand motion as the exact pose of the palm. The local hand motion is largely determined by the motion of five fingers. Therefore, the problem of hand motion capturing can be specifically formulated as:

$$(\mathbf{R},\mathbf{T},\ \theta) = arg \ \min_{\{R,t,\ \theta\}} \mathcal{D}(\mathbf{X} - \mathbf{P} M_G(\mathbf{R},\mathbf{T}) M_L(\ \theta)\mathbf{x}^m) \qquad (7)$$

where $M_G(\mathbf{R},\mathbf{T})$ is the global motion, $M_L(\ \theta)$ is the local finger motion, and perspective projection is approximated by scaled orthographic projection. Then, it is possible to estimate M_G and M_L. The idea of our approach is to use a robust 3D hand pose determination algorithm to estimate M_G and to use a fast inverse kinematics algorithm to calculate M_L in an iterative manner.

Given a specific local motion, we can define an operation $\mathcal{G}(\ \theta)$ to estimate the global motion such that:

$$(\mathbf{R},\mathbf{T}) = \mathcal{G}(\ \theta) = arg \ \min_{\{R,T\}} e(\mathbf{R},\mathbf{T},\ \theta) \qquad (8)$$

Given a global motion, an operation $\mathcal{L}(\mathbf{R},\mathbf{T})$ is defined to estimate the local motion such that:

$$\theta = \mathcal{L}(\mathbf{R},\mathbf{T}) = arg \ \min_{\theta} e(\mathbf{R},\mathbf{T},\ \theta) \qquad (9)$$

where e is an error measurement defined as:

$$e = e(\mathbf{R},\mathbf{T},\ \theta) = \mathcal{D}(\mathbf{X} - \mathbf{P} M_G(\mathbf{R},\mathbf{T}) M_L(\ \theta)\mathbf{x}) \quad (10)$$

We propose a two-step iterative algorithm to estimate the global hand motion M_G and the local motion M_L for one time frame:

- Track 2D features $\mathbf{I}(t)$ at time t, using the previous hand model, $H(t) = H(t-1) = H(\ \theta(t-1))$. The features are detected and tracked by feature tracking algorithm based on prediction from time $t-1$.

- The iterative algorithm is applied to estimate the global motion $\mathbf{R}(t), \mathbf{T}(t)$ and the hand state $\theta(t)$ at time t. Here, x^k is used to represent $x^k(t)$ for short.

  1. Take motion parameters of time $t-1$ as initial values, i.e., $\mathbf{R}^0 = \mathbf{R}(t{-}1)$, $\mathbf{T}^0 = \mathbf{T}(t{-}1)$, $\theta^0 = \ \theta(t-1)$.

  2. Estimate the global motion parameters at the 2k-th iteration. $(\mathbf{R}^{2k}, \mathbf{t}^{2k}) = \mathcal{G}(\ \theta^{2k-1})$, i.e., keep the local motion of previous iteration θ^{2k-1}.

  3. Estimate the local motion parameters $\theta^{2k+1} = \mathcal{L}(\mathbf{R}^{2k}, \mathbf{T}^{2k})$, i.e., hold the global motion $\mathbf{R}^{2k}, \mathbf{T}^{2k}$ of the 2k-th iteration for the (2k+1)-th iteration.

  4. Terminate the iteration when the change in error falls below a preset threshold. Then, the estimation of $\mathbf{R}(t)$, $\mathbf{T}(t)$, and $\theta(t)$ are obtained at time t.

- Update the 3D model $H(t) = H(\theta(t))$, then process the next time frame.

A proof of convergence is given below. The ideas of the two-step iterative algorithm are that the operation $\mathcal{G}(\ \theta)$ finds the best global motion given known local motion, and the operation $\mathcal{L}(\mathbf{R},\mathbf{T})$ also find the best hand state given known global motion.

Convergence Theorem: The two-step iterative algorithm converges monotonically to a limit point with respect to certain nonnegative error measurement such as the mean square error.

Proof: We can suppose the error measurement to be the mean square error w.l.g.:

$$e^k = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{X}_i - \mathbf{P} M_G(\mathbf{R}^k, \mathbf{T}^k) M_L(\ \theta^k)\mathbf{x}_i^m||^2 \quad (11)$$

Since $\theta^{2k} = \ \theta^{2k-1}$, apply the operation $\mathcal{G}$ to estimate global motion at the 2k-th iteration.

$$(\mathbf{R}^{2k}, \mathbf{T}^{2k}) = \mathcal{G}(\ \theta^{2k-1}) = arg \ \min_{\{R,T\}} e(\mathbf{R}, \mathbf{T},\ \theta^{2k-1}) \quad (12)$$

so, the error of the 2k-th iteration is:

$$e^{2k} = e(\mathbf{R}^{2k}, \mathbf{T}^{2k},\ \theta^{2k-1}) = \min_{\{R,T\}} e(\mathbf{R}, \mathbf{T},\ \theta^{2k-1}) \quad (13)$$

Obviously, $e^{2k} \leq e^{2k-1}$. Then, the operation $\mathcal{L}$ is applied to estimate local motion at the (2k+1)-th iteration:

$$\theta^{2k+1} = \mathcal{L}(\mathbf{R}^{2k}, \mathbf{T}^{2k}) = arg \ \min_{\{\ \theta\}} e(\mathbf{R}^{2k}, \mathbf{T}^{2k},\ \theta) \quad (14)$$

Since we keep the global motion $(\mathbf{R}^{2k+1}, \mathbf{T}^{2k+1}) = (\mathbf{R}^{2k}, \mathbf{T}^{2k})$, the error of the (2k+1)-th iteration is:

$$e^{2k+1} = e(\mathbf{R}^{2k}, \mathbf{T}^{2k},\ \theta^{2k+1}) = \min_{\{\ \theta\}} e(\mathbf{R}^{2k}, \mathbf{T}^{2k},\ \theta) \quad (15)$$

Obviously, $e^{2k1} \leq e^{2k}$. So, we have:

$$0 \leq e^{2k+1} \leq e^{2k} \leq e^{2k-1} \ \forall k \quad (16)$$

Since the error measurement cannot be negative, the lower bound occurs. Because the error sequence is non-increasing and bounded below, this two-step iterative algorithm should converge to a limit point. Furthermore, it can be shown that the algorithm converges to a stationary point. Q.E.D.

Our divide-and-conquer approach has a lot of advantages over other methods treating hand motion as a whole in which the optimization problem is always hard to handle. In our decoupled problem, we have many alternatives to deal with each sub-problem. In this paper, a robust pose determination method is proposed to act as the operation $\mathcal{G}(\ \theta)$ to estimate $\mathbf{R}$ and $\mathbf{T}$; and a simple but effective GA-based method acts as the operation $\mathcal{L}(\mathbf{R}, \mathbf{T})$ to estimate the hand state θ by solving the inverse kinematics problem.

3.2 Robust Pose Determination

From our motion capturing formulation, estimating global motion M_G is similar to the fundamental 2D-3D pose determination problem which is usually formulated as a least squares (LS) optimization problem [7, 4, 6], and pose parameters $\mathbf{R}$ and $\mathbf{t}$ are estimated by:

$$\underset{\{\mathbf{R},\mathbf{t},\ \theta\}}{minimize} \sum_{i=1}^{N} ||\mathbf{X}_i - \mathbf{P}(\mathbf{R}\mathbf{x}_i + \mathbf{T})||^2 \quad (17)$$

However, LS is known not to be a robust method, because it may break down even under only one outlier [15]. And LS is only optimal under the assumption of Gaussian noise. So, LS pose algorithms are not suitable for some practical applications because outliers are hard to avoid and the Gaussian noise assumption is hard to satisfy. In hand motion tracking, there are two sources of outliers. First, detection of accurate features cannot be guaranteed. Second, there are always some features that do not correspond to the delayed model. These features should be treated as outliers. If LS pose algorithms are used, the estimation risks a greater chance of failure.

Two typical robust regression methods from robust statistics [15, 3] are receiving attention in the computer vision community [9, 17, 11]. One type is M- estimators based on maximum likelihood, and the other is least median of squares (LMS). Different outlier detection methods are associated with these robust methods. However, M-estimators essentially are not a robust approach [15]. So, we use LMS to estimate the hand pose, i.e., $\mathbf{R}$ and $\mathbf{t}$ are estimated by:

$$\underset{\{\mathbf{R},\mathbf{t},\ \theta\}}{minimize} \ \underset{i}{median} ||\mathbf{X}_i - \mathbf{P}(\mathbf{R}\mathbf{x}_i + \mathbf{T})||^2 \quad (18)$$

By this means, pose $(\mathbf{R}, \mathbf{T})$ can be estimated robustly even if up to half of the data are outliers. This method can still give an accurate estimation when the observed image features are noisy. We use N feature points distributed around the fingertips, joints and the palm. Suppose 3D pose can be estimated with a minimum of p points. The least median is approximated using a trial estimate by repeatedly drawing sub-sample sets of p different observations. Our robust pose algorithm is outlined below:

- Repeat drawing sub-sample sets of p different observation points in 2D and model points in 3D correspondingly. This part can be done in a parallel way.
- For each sub-sample set, estimate $\mathbf{R}$ and $\mathbf{T}$ by the least square pose algorithm using only those p correspondences, then calculate the error for all N features, and record the least one. LS pose algorithm is described below:
  - Back project each observed feature point to the image plane in camera frame $\{\mathbf{v}_i\}$, $i = 1, \ldots, p$.

- Choose an initial depth $\{d_i^0\}\ \forall i$
- Fit two 3D point sets $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\} = \{d_n^k \mathbf{v}_i\}$ by SVD to obtain $\mathbf{R}^k$ and $\mathbf{T}^k$ for the $k-th$ iteration, where $\{\mathbf{x}_i\}$ is the model point set.
- Update $\{d_i^k\}$ by $d_n^{k+1} = \frac{(\mathbf{R}^k \mathbf{x}_n + \mathbf{T}^k)'\mathbf{v}_n}{||\mathbf{v}_n||^2}$ [4].
- Terminate the iteration when the change of error falls below a preset threshold.

- Find the sub-sample set of p observations with the least median error, and identify outliers [15] based on the estimation from those p points.
- Estimate $\mathbf{R}$ and $\mathbf{T}$ again using LS pose algorithm without considering outliers.

3.3 Inverse Kinematics

From our motion capturing formulation, the estimation of local finger motion $\mathbf{M_L}(\theta)$ is approached by solving inverse kinematics [12], which is one way to recover the joint angles given some position of 2D or 3D points on the kinematical chain. The difficulty of inverse kinematics is that the solution may not be unique. If we want to capture motion from monocular image sequences, multiple points on the kinematical chain in image should be specified. If motion is captured from multiple cameras, we may only specify fingertips which are the end-effectors of the kinematical chains, and the 3D locations of these fingertips can be calculated from multiple cameras. However, the solution may still not be unique. We have to add a set of constraints to narrow the search space.

One possible approach to this problem is the gradient-based method. The drawback of this method is that the optimization is prone to local minima. Another method is the Genetic Algorithm which may reach the global minimum. We take a GA-based method in this paper. In our algorithm, a potential solution, i.e., θ is coded as a binary string with each joint angle in the kinematical chain described by a binary code. A fitness value is associated with each potential solution. Solutions with less error have larger fitness values. A pool of candidate solutions evolves according to their fitness. A crossover operator and a mutation operator are designed to apply to two selected individuals to produce descendants. Individuals with higher fitness values have a greater probability to be selected for reproduction. After several rounds of evolution, the individual with the largest fitness value is taken as the near optimal solution to the inverse kinematics problem. The solution of the inverse kinematics need not be accurate since exact joint angles are not necessary for hand gestures recognition. So, we find a near optimal solution with less computational cost.

4 Experiments

Comparison experiments are given in this section. The observed feature points in the images include unknown noise, and the 3D model is "out-of- date", i.e., the set of joint angles θ of the hand model was estimated at a previous time frame so that it does not correspond to the current image, in which the hand has performed large local motions with flexion and abduction of several fingers. A polygon mesh hand model is driven by the estimated motion parameters to show the results. The experiments show that the LS method fails to estimate the motion, while our algorithm works well, see Figure 2 and Figure3. The thumb is omitted for simplicity.

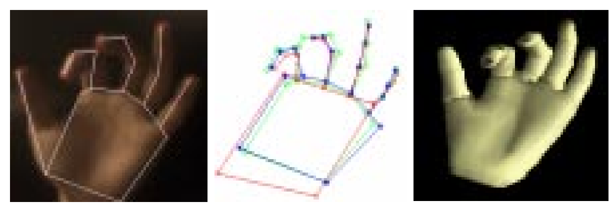

Figure 2: Hand with flexion of several fingers, but all joint angles of the initial model are at 0 degrees. Image with detected features (left). 2D projection of estimated motion, 'o' detected feature, '*' LMS, 'x' LS. (middle). Recovered hand state using the estimated parameters by LMS (right).

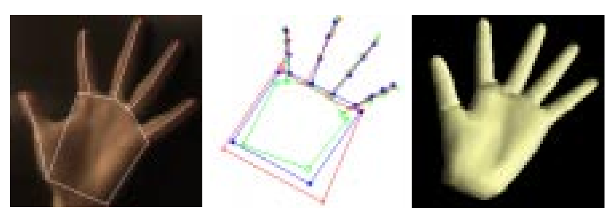

Figure 3: Hand with abduction of several fingers

5 Conclusion and Future work

A new 3D model-based approach is proposed to capture articulated human hand motion. Articulated human hand motion is decoupled to its global motion and its local finger motion. The global motion is formulated as a pose determination problem, and a robust pose algorithm based on least median of squares is developed to accurately estimate the global motion. The local finger motion is formulated as an inverse kinematics problem, and an effective genetic-based method is employed to search a high dimensional space to get a sub-optimal solution. The motion capturing algorithm at each time frame is an iterative procedure, and a convergence theorem is also given in this paper.

Several experiments show that the proposed approach can effectively estimate articulated hand motion. Comparison between a LMS-based method and a LS-based method is given in the paper. If the local hand motion between two consecutive frames is small, both algorithms converge. However, when large local finger motion is present, the LS-based method fails, but our algorithm can still estimate the global and local hand motion.

Our work is based on the assumption that all the fingertips can be reliably detected and tracked. If one of the fingertips is occluded, our algorithm will fail to find the joint angles of that finger. It is necessary to investigate other robust pose determination methods without accurate feature point correspondences and even missing feature correspondences. Pose from line correspondences should also be investigated in future work. Research on the inverse kinematics problem will be extended in future work.

6 Acknowledgment

This work was supported in part by National Science Foundation Grant IRI- 9634618 and Grant CDA-9624396, and in part by Army Research Laboratory Cooperative Agreement No. DAAL01-96-2-0003. The authors would like to appreciate Dr. Kentaro Toyama for the proofreading. They would also like to thank the anonymous reviewers for their comments.

References

[1] J..K.Aggarwal, Q.Cai, "Human Motion Analysis: A Review, " *IEEE proc. Nonrigid and Articulated Motion Workshop'97*, pp90-102, 1997.

[2] James Davis, Mubarak Shah, "Determining 3-D Hand Motion", *IEEE Asilomar Conference on Signals, Systems and Computers*, 1994.

[3] F.R.Hampel, et.al. "Robust Statictics: The Approach Based on Influence Functions", *Wiley*, 1986.

[4] R.M.Haralick, et al. "Pose Estimation from Corresponding Point Data", *IEEE SMC*, Vol19, No.6, Nov/Dec, 1989.

[5] T.Heap, D.Hogg, "Towards 3D Hand Tracking Using a Deformable Model", *Proc.Int'l Conf.Automatic Face and Gesture Recognition*, Killington, Vt., pp.140-145, Oct.1996

[6] R.Holt, A.N.Netravali, T.S.Huang, R.Qian, "Determining Articulated Motion from Perspective Views: A Decomposition Approach", *IEEE MNAO'94*, pp.126-137, 1994.

[7] T.S.Huang, A.N.Netravali, "Motion and Structure from Feature Correspondences: A Review", *Proc. IEEE*, Vol.82, No.2, Feb, 252-268, 1994.

[8] J.J.Kuch, T.S.Huang "Vision-Based Hand Modeling and Gesture Recognition for Human Computer Interaction" *Master thesis, Univ.of Illinois at Urbana-Champaign*, 1994.

[9] Rakesh Kumar, Allen R. Hanson, "Robust Methods for Estimating Pose and a Sensitivity Analysis", *CVGIP:Image Understanding*, Vol.60, No.3, Nov.pp313-342, 1994.

[10] J.Lee, T. Kunii, "Model-based Analysis of Hand Posture", *IEEE Computer Graphics and Applications*, Sept, pp.77-86,1995.

[11] Peter Meer, Doron Mintz, Azriel Rosenfeld, Dong Yoon Kim, "Robust Regression Methods for Computer Vision: A Review", *IJCV*, 6:1, 59-70, 1991.

[12] Richard P.Paul, "Robot Manipulators: Mathematics, Programming, and Control", *The MIT Press*, 1982.

[13] Vladimir I. Pavlovic, R.Sharma, T.S.Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE PAMI*, Vol.19, No.7, July, pp.677-695, 1997.

[14] J.Rehg, T.Kanade, "Model-Based Tracking of Self-Occluding Articulated Objects" *Proc. IEEE Int'l Conf. Computer Vision*, Cambridge, Mass. pp.612-617, June 20-23 1995.

[15] P.J.Rousseeuw, A.M.Leroy, "Robust Regression and Outlier Detection", *Wiley-Interscience*, New York, 1987.

[16] N.Shimada, et al. "Hand Gesture Recognition Using Computer Vision Based on Model- matching Method", *Proc.of 6th Int'l Conf. on HCI*, 1995.

[17] Charles V. Stewart, "MINPRAN: A New Robust Estimation for Computer Vision", *IEEE PAMI*, Vol.17, No.10, Oct. pp925-938, 1995.

[18] Ying Wu, Thomas S. Huang, "Human Hand Modeling, Analysis and Animation is the Context of HCI", *Proc. of the IEEE Int'l Conf. on Image Processing*, Japan, 1999