Modeling the Constraints of Human Hand Motion

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

Abstract

Hand motion capturing is one of the most important parts of gesture interfaces. Many current approaches to this task generally involve a formidable nonlinear optimization problem in a large search space. Motion capturing can be achieved more cost-efficiently when considering the motion constraints of a hand. Although some constraints can be represented as equalities or inequalities, there exist many constraints, which cannot be explicitly represented. In this paper, we propose a learning approach to model the hand configuration space directly. The redundancy of the configuration space can be eliminated by finding a lower-dimensional subspace of the original space. Finger motion is modeled in this subspace based on the linear behavior observed in the real motion data collected by a CyberGlove. Employing the constrained motion model, we are able to efficiently capture finger motion from video inputs. Several experiments show that our proposed model is helpful for capturing articulated motion.

1 Introduction

In recent years, there has been a significant effort devoted to gesture recognition and related work in body motion analysis due to interest in a more natural and immersive Human Computer Interaction (HCI). As the cost for more powerful computers decreases and PCs become more popular, a more natural interface is desired rather than the traditional input devices such as mouse and keyboard. Using gestures, as one of the most natural ways humans communicate with each other, thus becomes an apparent choice for a more natural interface. An effective recognition of hand gestures will provide major advantages not only in virtual environments and other HCI applications, but also in areas such as teleconferencing, surveillance, and human animation.

Recognizing hand gestures, however, involves capturing the motion of a highly articulated human hand with roughly 30 degrees of freedom (DoF). Hand motion capturing involves finding the global hand movement and local finger motion such that the hand posture can be recovered. One possible way to analyze hand motion is the appearance-based approach, which emphasizes the analysis of hand shapes in images [5]. However, local hand motion is very hard to estimate by this means. Another possible way is the model-based approach [3, 4, 6, 7, 9]. With a single calibrated camera, local hand motion parameters can be estimated by fitting a 3D hand model to the observation images.

One method of model-based approaches is to use gradient-based constrained nonlinear programming techniques to estimate the global and local hand motion simultaneously [6]. The drawback of this approach is that the optimization is often trapped in local minima. Another idea is to model the surface of the hand and estimate hand configurations using the ‘‘analysis-by-synthesis” approach [3]. Candidate 3D models are projected to the image plane and the best match is found with respect to some similarity measurement. Essentially, it is a search problem in a very high dimensional space that makes this method computational intensive. A decomposition method is also adopted to analyze articulated hand motion by separating hand motion into its global motion and local finger motions[9].

Although the 3D model-based approach makes motion capturing from monocular images possible, it also faces some challenging difficulties. Many current methods for hand posture estimation basically involve the problem of searching for the optimal hand posture in a huge hand configuration space, due to the high DoF in hand geometry. Such a search process is computationally expensive and the optimization is prone to local minima. At the same time, many current approaches suffer from self-occlusion.

However, although the human hand is a highly articulated object, it is also highly constrained. There are dependencies among fingers and joints. Applying the motion constraints among fingers and finger joints can greatly reduce the size or dimensions of the search space, which in turn makes the estimation of hand postures more

cost-efficient. Another major advantage of applying hand motion constraints is to be able to synthesize natural hand motion and produce realistic hand animation, which would be very useful to synthesize sign languages.

There has not been much done regarding the study of hand constraints other than the commonly used ones. Even though constraints would help reduce the size of the search space, too many or too complicated constraints would also add to computational complexity. Which constraints to adopt becomes an important issue. Some constraints have already been presented, studied, and used in many previous works [3, 4, 9]. The common ones include the constraints of joints within the same finger, constraints of joints between fingers, and the maximum range of finger motions. All these are presented as either equalities or inequalities. However, due to the large variation in finger motion, there are yet more constraints that cannot be explicitly represented by equations.

In this paper we propose a learning approach to model the constraints directly from sampled data in the hand configuration space (C-Space). Each point in this hand configuration space corresponds to a set of joint angles of a hand state, which is commonly estimated in model-based approaches. Rather than studying the global hand motion, we will focus only on the analysis of local finger motions and constraints with the help of a CyberGlove developed by Virtual Technologies Inc. Moreover, we will study the constraints of hand motions that are natural and feasible to everyone.

In section 2, a description of a commonly adopted hand kinematical model and the CyberGlove is given. Section 3 describes how we model the configuration space and the observations of this model. Section 4 shows the results of some preliminary examples of hand posture estimation taking advantage of this model. Section 5 concludes our work and discusses some future directions regarding modeling human motion constraints.

2 Hand skeleton model

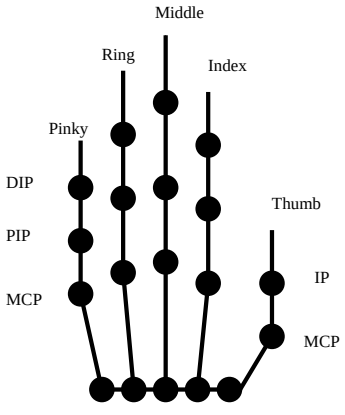


Figure 1: Kinematical structure and joint notations

The human hand is highly articulated. To model the articulation of fingers, the kinematical structure of hand should be modeled. In our research, the skeleton of a hand can be abstracted as stick figure with each finger as a kinematical chain with base frame at the palm and each fingertip as the end-effecter. Such a hand kinematical model is shown in Figure 1 with the names of each joint. This kinematical model has 27 Degrees of Freedom (DoF).

Each of the four fingers has four DoF. The distal interphalangeal (DIP) joint and proximal interphalangeal (PIP) joint each has one DoF and the metacarpo-phalangeal (MCP) joint has two DoF due to flexion and abduction. The thumb has a different structure from the other four fingers and has five degrees of freedom, one for the interphalangeal (IP) joint, and two for each of the thumb MCP joint and trapeziometacarpal (TM) joint both due to flexion and abduction. The fingers together have 21DoF. The remaining 6 degrees of freedom are from the rotational and translational motion of the palm with 3 DoF each. These 6 parameters are ignored since we will only focus on the estimation of the local finger motions rather than the global motion.

Articulated local hand motion, i.e. finger motion, can be represented by a set of joint angles θ, or the hand state. In order to capture the hand motion, glove-based devices have been developed to directly measure the joint angles and spatial positions by attaching a number of sensors to hand joints. CyberGlove is such a device.

The goal of vision-based analysis of hand gesture is to estimate the hand joint angles or hand states without using such physical devices but solely based on visual information. However, such glove-based device can help collecting ground truth data, which enable the modeling and learning processes in visual analysis.

In our study, we employ a right-handed CyberGlove. The glove has four sensors for the thumb, a MCP and a PIP sensor for the PIP (θ_{PIP}) and MCP (θ_{MCP-F}) flexion angles for each of the four fingers, three more abduction sensors for the abduction/adduction angles (θ_{MCP-AA}) between these four fingers. There are total of fifteen sensor readings of the finger joint angles; therefore we are able to characterize the local finger motion by 15 parameters. The glove can be calibrated to accurately measure the angle within 5 degrees, which is acceptable for gesture recognition. For finger postures that are five degrees different would still appear to be the same posture.

3 Modeling the constraints

Modeling motion constraints is crucial to effective and efficient motion capturing. A comprehensive study of hand/finger motion constraints and a learning approach to modeling the natural movement constraints are given in this section.

3.1 Constraints overview

Hand/finger motion is constrained so that hand cannot make arbitrary gestures. There are many examples of such constraints. For instance, fingers cannot bent backward too much and the pinky finger cannot be bend without bending the ring finger. The natural movements of human hands are implicitly caused by such motion constraints.

Some motion constraints may have a closed form representation, such that they are often employed in current research of animation and visual motion capturing [3, 4, 9]. However, a large number of motion constraints are very difficult to be expressed in closed forms. How to model such constraints is still needs further investigation. Here we present some of the most commonly used motion constraints and justify the use of 15 parameters to represent the hand motion.

Hand constraints can be roughly divided into three types. Type I constraints are the limits of finger motions as a result of hand anatomy which is usually referred to as static constraints. Type II constraints are the limits imposed on joints during motion, which is usually referred to as dynamic constraints in previous work. Type III constraints are applied in performing natural motion, which has not yet been explored. Below we will describe each type in more detail.

Type I constraints. This type of constraint refers to the limits of the range of finger motions as a result of hand anatomy. We will only consider the range of motion of each finger that can be achieved without applying external forces such as bending fingers backward using the other hand. This type of constraints is usually represented using the following inequalities:

$$0° \le \theta_{MCP-F} \le 90°,$$
$$0° \le \theta_{PIP} \le 110°,$$
$$0° \le \theta_{DIP} \le 90°, \text{ and}$$
$$-15° \le \theta_{MCP-AA} \le 15°. \quad (1)$$

Another commonly adopted constraint states that middle finger displays little abduction/adduction motion and the following approximation is made for middle finger:

$$\theta_{MCP-AA} = 0. \quad (2)$$

This will reduce one DoF from the 21 DoF model.

Similarly, the TM joint also displays limited abduction motion and will be approximated by 0 as well.

$$\theta_{TM-AA} = 0. \quad (3)$$

As a result, the thumb motion will be characterized by 4 parameters instead of 5.

Finally, the index, middle, ring, and little fingers are planar manipulators. i.e. the DIP, PIP and MCP joint of each finger move in one plane since DIP and PIP joints only has 1 DoF for flexion.

Type II constraints. This type of constraint refers to the limits imposed on joints during finger motions. These constraints are often called dynamic constraints and can be subdivided into intra-finger constraints and inter-finger constraints. The intra-finger constraints are the constraints between joints of the same finger. A commonly used one based on hand anatomy states that in order to bend the DIP joints, the PIP joints must also be bend for the index, middle, ring and little fingers. The relations can be approximated as following:

$$\theta_{DIP} = \frac{2}{3}\theta_{PIP}. \quad (4)$$

By combining Eq 2-4, we are able to reduce the model with 21 DoF to one that is approximated by 15 DoF. Experiments in previous work have shown that postures can be estimated using these constraints without severe degradation in performance.

Inter-finger constraints refer to the ones imposed on joints between fingers. For instance, when one bends his index finger at MCP joint, he would naturally have to bend the middle MCP joint as well. Many of such Type II constraints and related equations can be found in [2, 4]. However, there are yet more constraints that can not be explicitly represented in equations.

Type III constraints. These constraints are imposed by the naturalness of hand motions and are more subtle to detect. Almost nothing has been done to account for these constraints in simulating a natural hand motion. Type III constraints differs from Type II in that they have nothing to do with limitations imposed by hand anatomy, but rather are a result of common and natural movements. Even though the naturalness of hand motions is different from person to person, it is similar for everybody. For instance, the most natural way for every person to make a fist from an open hand would be to curl all the fingers at the same time instead of curling one finger at a time. This type of constraint also can not be explicitly represented by equations.

3.2 Modeling the constraints in C-space

It is difficult to explicitly represent the constraints of natural hand motions in closed form. However, they can be learned from a large and representative set of training samples; therefore we propose to construct the configuration space (i.e., joint angle space) and learn the constraints directly from the empirical data using the approach described below. For notational convenience, let us denote the feasible C-space by $\Phi \subset \Re^{15}$ with each configuration denoted by $\phi = (\theta_1, \theta_2, \ldots \theta_{15})$.

1. *Locating base states* ζ_i *in* Φ. We will directly locate the base states by fixing the hand in desired configurations and measure the 15 parameters associated with the corresponding state. Since the sensors are very sensitive to finger movements, little variations in finger postures will also be recorded and will be considered as

the same state. As a result, we will use the centroid from the set of training data $D_i = \{x_{ij}, j = 1 \ldots N\}$ as the location of the base state ζ_i. Another alternative would be to collect huge set of training samples x_i from predefined motions and apply a clustering algorithm in order to locate the base states. However, since we have full control of how a hand must be configured to form the base state, we do not need to apply clustering algorithms to locate the base states in C-space.

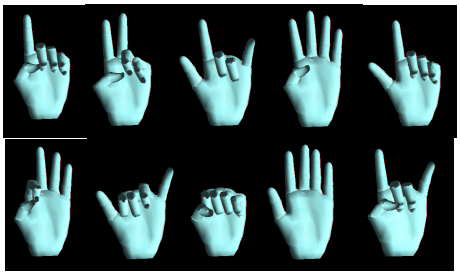


Figure 2a: Some base hand states

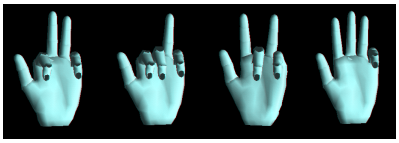


Figure 2b: Unfeasible configurations

In our model, the hand gestures are roughly classified into 32 discrete states by quantizing each finger into one of the two states: fully extended or curled. The reason for choosing these two states is that the entire motions of a finger falls roughly between these two states. Therefore the whole set of 32 states will roughly characterize the entire hand motion (Figure 2). However, since not everyone is able to bend the pinky without bending the ring finger or with the help of thumb to hold the pinky, four of the states will not be achievable by everyone without applying external forces. Therefore, these four states (Figure 2b) are not included in our set of base states in C-space modeling. Finally, the configurations that are similar are considered as the same state. For instance, the cases with five fingers opening wide apart and with all fingers straightened but closed together are considered the same.

2. *Motion modeling*. With the set of base states ζ_i established, we then collect motion data for state transitions in order to model the configurations during natural hand motion. A large number of sets of motion data are collected in order to observe the Type II and III constraints of natural hand motions. An example of motion between making and opening a fist is shown in Figure 3.

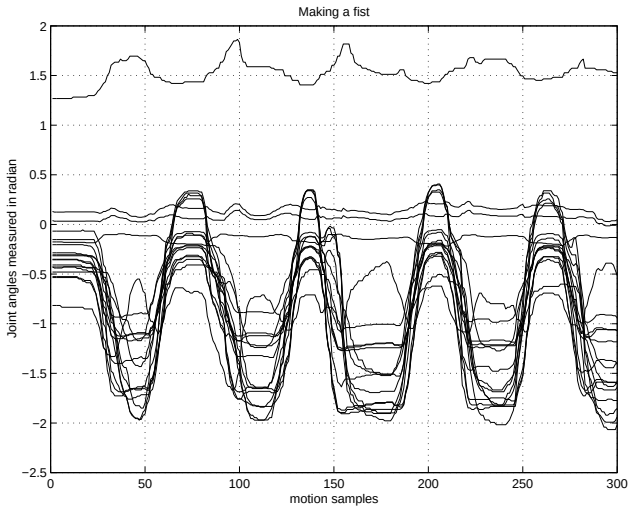


Figure 3: Joint angle measurements from the motion of making and opening a fist.

3. *Dimensionality reduction*. From Figure 3, we can clearly observe some correlations in the joint angle measurements. Therefore, together with the data collected from static states and the finger motions, we then perform Principal Component Analysis (PCA) to reduce the dimension of the model and thus reduce the search space while preserving the components with the highest energy. We note that 95% of the energy is contained in the 7 dimensions that have the largest eigenvalues. We thus perform the mapping $\Re^{15} \to \Re^7$ on Φ by projecting the original model onto a lower dimensional subspace $\Phi^c \subset \Re^7$ with principle directions associated with these 7 largest eigenvalues.

4. *Interpolation in compressed C-space*. Once a set of base states ζ_i have been determined, the whole feasible configuration space Φ can be approximated by these base states ζ_i and an interpolation scheme. Our approach takes a linear interpolation in the lower-dimensional configuration subspace Φ^c. For each configuration ϕ^c in Φ^c we will represent its parameters using a polynomial interpolation, i.e.,

$$\phi^c = \sum_{i=1}^{28} \alpha_i \zeta_i^c , \qquad (5)$$

in which ζ_i is the location of base state i and α_i is the parameters for ϕ^c.

3.3 Model characteristics

Our model has three main advantages that will help reduce the search space in gesture recognition. First, the

model is compact due to the dimensionality reduction using PCA. Second, the motion constraints are automatically incorporated into the model. Third, a linear behavior is observed in the state transitions in C-space.
The reason that motion constraints are incorporated into this model is because we sample directly from natural hand motions. Configurations that are outside of permissible range limited by hand anatomy will not be achievable in natural hand motions. Consequently, the inequalities and equalities including the intra-finger constraints [3, 4], such as $\theta_{MCP-F} = k\theta_{PIP}$ with $k \geq 0$, and inter-finger constraints [4] are automatically covered in this model.

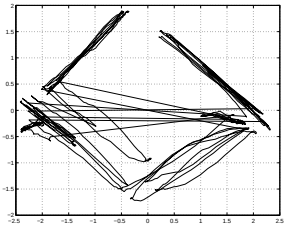


Figure 4: Motion transistions between four states between.

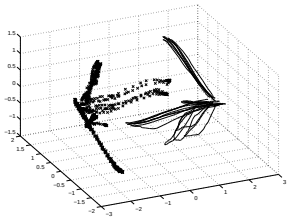


Figure 5: Motion transistions between eight states.

An interesting phenomenon regarding the Type III motion constraints is observed from the motion data. We observe a nearly linear transition between states in C-space. An example is shown for the case of transitioning between four states from the result of moving only index and middle fingers (Figure 4). We have projected the C-space into $\Re^2$ for observation in this case. The four corners are the locations of the four discrete base states. A linear transition is clearly observed from Figure 4. The middle lines are the path resulting from curling and extending both fingers together. This result reflects the high correlation between fingers when performing natural movements. Although state transitions does not necessarily need to be performed in this manner and there exists infinitely many ways to move from one configuration to another, when the fingers are moving in their most natural way, it will take a nearly straight line path in C-space. This observation will justify Eq 5 in estimating the hand configurations. Another example is shown with three-finger motions projected in $\Re^3$. The eight base states are roughly located at the eight corners of a cube (Figure 5).

4 Experiments

In order to evaluate the validity of this model, we perform some experiments using low-level visual features and estimate the postures constituted by a subset of the 28 base states. The input images are assumed to be segmented.

4.1 General approach

Using the result we observe from the linear behavior, we are able to approximate a configuration by taking the following steps:

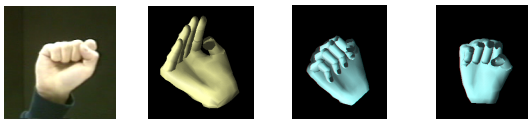
1. In the training stage, first associate each base state ζ_i^c with a feature vector ψ_i.
2. Extract features ψ_{input} from the input 2D image, such as edge, area, centroid, etc.
3. Compute $\alpha_i = h(\psi_i, \psi_{input})$, where $h(\psi_i, \psi_{input})$ measures the closeness of ψ_{input} to ψ_i.
4. Based on the observation made from Type III motion constraints, linearly interpolate the estimated configuration in compressed space Φ^c:

$$\phi_{estimate}^c = \sum_{i=1}^{28} \alpha_i \zeta_i^c \qquad (6)$$

5. Reconstruct the estimated configuration state $\phi_{estimate} \subset \Re^{15}$ from $\phi_{estimate}^c$.

4.2 Experimental results

Figure 6: Configuration estimations.



(a) (b) (c) (d)

Figure 7: Comparison of different technique. (a) original image. (b) estimation without Type II & III constraints. (c) estimation without Type III constraints. (d) estimation with Type I, II & III constraints

The results of the experiments are shown in Figure 6. The first row shows some input images and the second

row shows the reconstructed 3D hand model based on estimation by our approach. The results are visually agreeable. Such preliminary experiments show that the motion constraints play an important role in hand posture estimation. More accurate and cost-efficient estimation can be obtained when a better motion constraint model can be applied. Better result can be obtained with better feature extraction methods, which will be implemented in the future research.

A comparison of estimations using different types of constraints is also shown in Figure 7. In Figure 7(b), estimation without applying Type II and III constraints result in a feasible, yet unnatural configuration. In Figure 7(b), a closer approximation is obtained without applying Type III constraints. The DIP and PIP joints should bend more to approximate a fist. Finally, by applying all three types of constraints together produce the better result with a more natural approximation in Figure 7(c).

5 Conclusion/Future Development

A posture estimation problem generally involves a search in high dimensional C-space. Useful hand constraints have been demonstrated to be able to greatly reduce the search space, and thus improve gesture recognition results. Many constraints can be represented in simple closed forms while many more can not and have not been found.

In this paper, we presented a novel approach to model the hand constraints. Our model has three characteristics. First, it is compact by utilizing PCA technique. Second, it incorporates constraints that can and cannot be represented by equations. Third, it displays a linear behavior in state transitioning as a result of natural motion. These properties together simplify configuration estimation in C-space as shown in Eq 5 by a simple interpolation with linear polynomials. Some preliminary gesture estimation experiments are shown, taking advantage of this model.

However, there is still much to be done to improve this model. For instance, more states can be included to further refine the model. Deciding which states to choose will require more analysis of the C-space. Furthermore, other constraints might exist in the C-space that haev not yet been observed. Finally, even though a nearly linear behavior is observed in state transition, it is not exactly linear. A more detailed study can better approximate the trajectories, which in turn would help improve the configuration estimation. Nevertheless, such modeling provides a different interpretation of hand motions and the current results look promising.

Acknowledgement

This work was supported in part by National Science Foundation Alliance Program and Grant CDA 96-24396.

References

[1] C. Chang, W. Tsai, “Model-Based Analysis of Hand Gestures From Single Images Without Using Marked Gloves Or Attaching Marks on Hands”, *ACCV2000*, pp.923-930, 2000

[2] C.S. Chua, H. Y. Guan and Y. K. Ho, “Model-based Finger Posture Estimation”, *ACCV2000*, pp.43-48, 2000.

[3] J. Kuch and Thomas S. Huang, “Vision-Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration”, *ICCV95*, pp.666-671, 1995.

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

[5] V. Pavlovic, R. Sharma, Thomas S. Huang, “Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review,”, *IEEE PAMI*, Vol. 19, No. 7, July, pp.677-695, 1997

[6] J. Rheg, T. Kanade, “Model-Based Tracking of Self-Occluding Articulated Objects”, *IEEE Int’l Conf. Computer Vision*, pp.612-617. 1995.

[7] N. Shimada, et al., “Hand Gesture Estimation and Model Refinement Using Monocular Camera-Ambiguity Limitatio by Inequalty Constraints”, *Proc. Of the 3rd Conf. On Face and Gesture Recognition*, 1998.

[8] Y. Wu, Thomas S. Huang, “Human Hand Modeling, Analysis and Animation in the Context of HCI”, *ICIP99*, Japan, Oct., 1999.

[9] Y. Wu, Thomas S. Huang, “Capturing Human Hand Motion: A Divide-and-Conquer Approach”, *IEEE Int’l Conf. Computer Vision*, Greece, 1999.