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 [3, 8]. 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 [4, 8]. However, local hand motion is very hard to estimate by this means. Another possible way is the model-based approach [1, 2, 6, 7, 10, 13, 15]. 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 [10]. 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 [6]. 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 proposed to analyze articulated hand motion by separating hand motion into its global motion and local finger motions [15].

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

0-7695-0939-8/00 $10.00 © 2000 IEEE

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 [1, 2, 6, 7]. 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 high flexibility 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.

2 Hand skeleton model

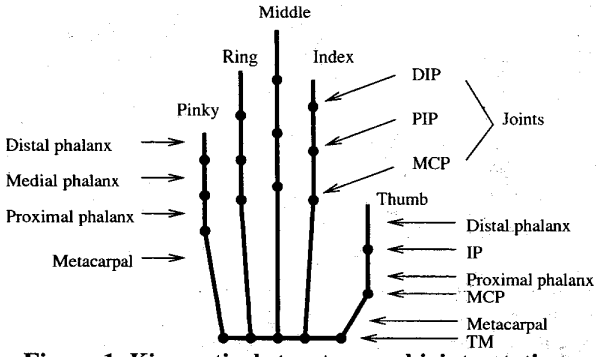


Figure 1: Kinematical structure and joint notations.

The human hand is highly articulated. To model the articulation of fingers, the kinematical structure of the hand should be modeled. In our research, the skeleton of a hand can be abstracted as a stick figure with each finger as a kinematical chain with base frame at the palm and each fingertip as the end-effector. Such a hand kinematical model is shown in Figure 1 with the names of each joint. This model has 27 Degrees of Freedom (DoF). There are 21 DoF contributed by the finger joints for the local motion and 6 DoF due to the global motion [7]. Since we will only focus on the estimation of the local finger motions rather than the global motion, these six parameters are not considered in our current study.

Articulated local hand motion, i.e. finger motion, can be represented by a set of joint angle values. 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. Although the goal of vision-based hand motion analysis is to be able to recognize hand configurations without the use of attached external devices, a glove-based device will help in collecting ground truth data, which enable the modeling and learning process in visual analysis.

In our study, we employ a right-handed CyberGlove, which provides 15 sensors for measuring 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. This is acceptable for gesture recognition; finger postures that are five degrees different would still appear to be the same posture.

3 Modeling the constraints

3.1 Constraints overview

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

Some motion constraints may have a closed form representation, and they are often employed in current research of animation and visual motion capturing [1, 2, 6, 7, 15]. However, many motion constraints are very difficult to express in closed forms. How to model such constraints still needs further investigation. Here we present three types of motion constraints and explain how we are able to represent hand motion with only 15 parameters instead of 21.

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 are usually referred to as static constraints. Type II constraints are the limits imposed on joints during motion, which are usually referred to as dynamic constraints in previous work. Type III constraints are applied in performing natural motion, and have 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 constraint is usually represented by the following inequalities:

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

where the subscript F denotes flexion and AA denotes abduction/adduction.

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

$$\theta_{MCP\_AA} = 0°. \quad (2)$$

This will reduce 1 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 four parameters instead of five.

Finally, the index, middle, ring, and little fingers are planar manipulators. In other words, the DIP, PIP and MCP joint of each finger move in one plane since the DIP and PIP joints only have 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 intrafinger and interfinger constraints. The intrafinger constraints are the constraints between joints of the same finger. A commonly used one based on hand anatomy states that for the index, middle, ring and little fingers, in order to bend the DIP joints, the corresponding PIP joints must also be bent. The relations can be approximated as follows:

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

By combining Eqs. (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.

Interfinger constraints are those imposed on joints between adjacent fingers. For instance, when an index MCP joint is bent, the middle MCP joint is forced to bend as well. Lee and Kunii [7] have performed measurements on several people and obtained a set of inequalities that approximates the limits of adjacent MCP joints. However, there are yet more constraints that cannot be explicitly represented in equations.

Type III constraints. These constraints are imposed by the naturalness of hand motions and are more subtle to detect and quantify. Almost nothing has been done to account for these constraints in simulating natural hand motion. Type III constraints differ from Type II in that they have nothing to do with limitations imposed by hand anatomy, but rather are results of common and natural movements. 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. Even though the naturalness of hand motions differs from person to person, it is broadly similar for everybody. This type of constraint also cannot 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 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 ϕ.

1. *Locating base states* ζ_i *in* Φ. We will directly locate the base states by fixing the hand in desired configurations and measuring 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 N training data $D_i = \{x_{ij}, j = 1...N\}$ as the location of the base state ζ_i. Another alternative would be to collect a huge set of training samples x_i from predefined motions and apply a clustering algorithm in order to locate the base states. This approach was taken in [11] for body posture estimation. 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.

In our model, the hand gestures are roughly classified into 32 discrete states by quantizing each finger into one of two states: fully extended or curled. The reason for choosing these two states is that the entire motion of a finger falls roughly between these two states. Therefore, the whole set of 32 states will roughly characterize the entire hand motion (Figure 2a and 2b). However, since not everyone is able to bend the pinky unless the ring finger is also bent or an external force is applied, 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 spreading apart and with all fingers straightened but closed together are considered the same.

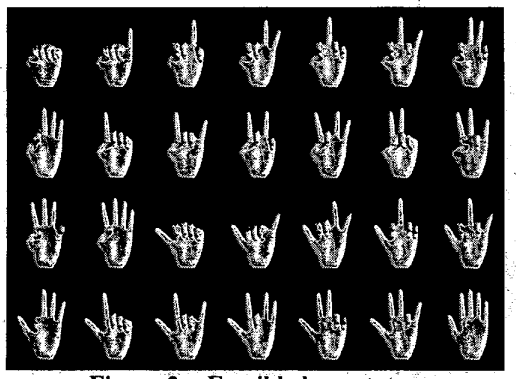


Figure 2a: Feasible base states.

Figure 2b: Infeasible base states.

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 motions. 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 measuring the motion of making and opening a fist is shown in Figure 3.

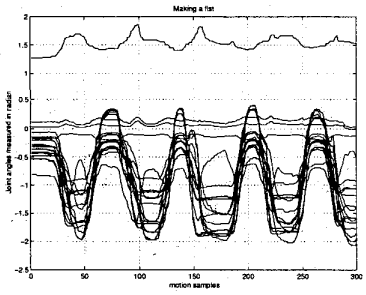


Figure 3: Joint angle measurements from the motion of making and opening a fist.

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 components 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 seven dimensions that have the largest eigenvalues. We thus perform the mapping $\Re^{15} \rightarrow \Re^{7}$ on Φ by projecting the original model onto a lower-dimensional subspace $\Phi^c \subset \Re^7$ with principle directions associated with these seven largest eigenvalues.

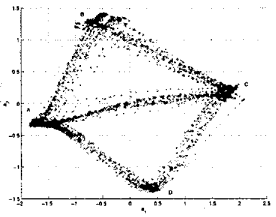


Figure 4a: Motion transitions between four states.

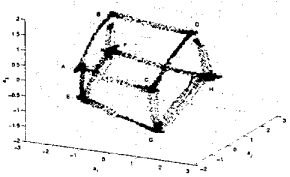


Figure 4b: Motion transitions between eight states.

4. *Interpolation in compressed C-space.* 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 in the movements of the index and middle fingers (Figure 4a). Since only a few joints are involved in making these movements, we are able to perform PCA and project the C-space into $\Re^2$ for observation without losing much information. The four corners are the locations of the four discrete base states. A linear transition is clearly demonstrated in Figure 4a. Another example is shown in Figure 4b with three-finger motions projected onto $\Re^3$. The eight base states are roughly located at the eight corners of a cube.

Based on this observation of linear behavior, once a set of base states ζ_i has 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 $\phi^c \subset \Phi^c$ we will represent its parameters by a polynomial interpolation, i.e.,

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

in which ζ_i^c is the projected location of base state ζ_i and

$$\sum_{i=1}^{28} \alpha_i = 1 . \qquad (6)$$

3.3 Model characteristics

Our model has three main characteristics that will help reduce the search space in gesture recognition. First, the model is compact due to the dimensionality reduction by PCA. This property also helps to compactly encode gesture representations. To obtain the data in original C-space only requires linear computations with low complexity.

Second, the motion constraints are automatically incorporated in the model. The reason for incorporating

motion constraints is that we sample directly from natural hand motions. Type I constraints are represented as the boundary in the C-space, since configurations that are outside of the range permitted by hand anatomy will not be achievable in natural hand motions. Type II constraints are shown through the direction of the paths during motion. Type III constraints are observed as the straight lines from state transition paths involving multiple fingers moving together.

The third characteristic of the model is the linear behavior observed in the state transitions in the C-space. As stated before, this is the result of the Type III constraints. This observation allows us to justify the representation of all feasible configurations using linear interpolations as in Eq. (5). Furthermore, we are able to produce synthetic hand motions that replicate real hand motions with simple computations by knowing the trajectories of the state transitions. Although many current techniques exist that strive to generate lifelike hand motions [5, 9, 12, 14], many of them suffer from great computational complexity. Finally, by including time domain knowledge of the hand configuration with this linear behavior, we will be able to better predict the new configurations.

4 Posture estimations

Using the result we observe from the linear behavior, we are able to utilize the model for applications in posture estimation by taking the general approach as follows:

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, and α_i are normalized as in Eq. (6).
4. Based on the observation made from Type III motion constraints, linearly interpolate the estimated configuration in the compressed space Φ^c:

$$\phi_{estimate}^c = \sum_{i=1}^{28} \alpha_i \zeta_i^c \qquad (7)$$

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

5 Experiments

In order to evaluate the validity of this model, we perform some experiments in synthesizing realistic finger motions and estimating the postures constituted by a subset of the 28 base states. The input images are assumed to be segmented. In our current experiment, we manually identify the 2D locations of the fingertips relative to the center of the base of the palm as the features ψ_{input} for each input image.

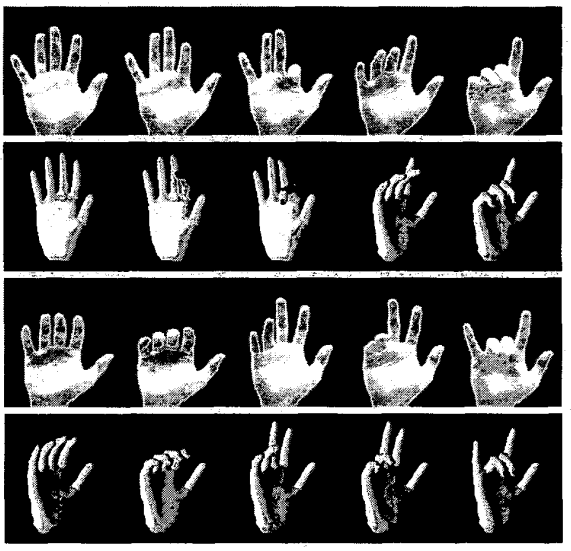
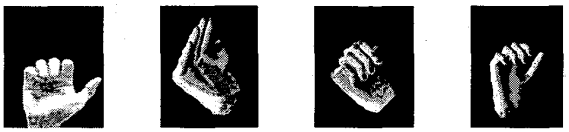


Figure 5: Configuration estimations.



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

Figure 6: Comparison of different techniques. (a) original image. (b) estimation with Type I constraints only. (c) estimation with Type I & II constraints only. (d) estimation with Type I, II & III constraints.

The results of the experiments are shown in Figure 5. The first and third rows are the input images and the second and fourth rows are the corresponding reconstructed 3D hand models based on the 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 is applied. Moreover, better results 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 6(a)-(d). In Figure 6(b), estimation with only Type I constraints results in a feasible, yet unnatural configuration. In Figure 6(c), a closer approximation is obtained by applying Type I & II constraints. Some additional adjustments are required in order to approximate the configurations correctly. Finally, applying all three types of constraints together produces the better result with a more natural approximation in Figure 6(d).

Another application is hand motion synthesis by reconstructing the sequences of configurations along the lines that approximate the state transitions (Figure 7). Since the lines are the approximations of the original real motion data, the reconstructed sequences also incorporate the constraints, which make the motion realistic.

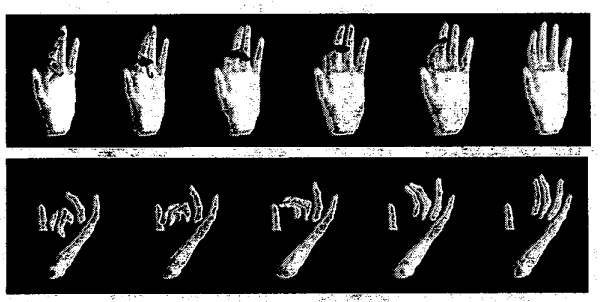


Figure 7: Sequences of synthesized finger motions.

6 Conclusion

A posture estimation problem generally involves a search in a high dimensional C-space. Useful hand constraints have been demonstrated to greatly reduce the search space, and thus improve gesture recognition results. Many constraints can be represented in simple closed forms while many more cannot 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. 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 the 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 have 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, our constraints modeling provides a different interpretation of hand motions and the current results look promising.

Acknowledgement

This work was supported in part by NSF CDA-96-24396, NSF IRI-9634618, and ARL Cooperative Agreement DAAL01-96-2-0003.

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*, 2000, pp. 923-930.

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

[3] R. Cipolla and A. Pentland (Editors), *Computer Vision for Human-Machine Interaction.* Cambridge: Cambridge University Press, 1998.

[4] T. Heap and D. Hogg, "Wormholes in shape space: tracking through discontinuous changes in shape," in *Proc. of IEEE ICCV98*, 1998, pp. 344-349.

[5] P. Kalra, N. Magnenat-Thalmann, L. Moccozet, G. Sannier, A. Aubel, and D. Thalmann, "Real-time animation of realistic virtual humans," *IEEE Computer Graphics and Applications,* vol. 18, issue:5, pp. 42-56, Sept-Oct, 1998.

[6] J. Kuch and T. S. Huang, "Vision-Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration", *ICCV95*, 1995, pp.666-671.

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

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

[9] Z. Popović and A. Witkin, "Physically based motion transformation," *SIGGRAPH99,* 1999, pp. 11-20.

[10] J. Rheg, T. Kanade, "Model-Based Tracking of Self-Occluding Articulated Objects", *Proc. of IEEE ICCV95*, 1995, pp. 612-617.

[11] R. Rosales and S. Sclaroff, "Inferring body pose without tracking body parts," in *Proc. of IEEE CVPR*, 2000, vol. 2, pp. 721-727.

[12] C. Rose, B. Guenter, B. Bodenheimer, and M. F. Cohen, "Efficient generation of motion transitions using spacetime constraints," in *Proc. of SIGGRAPH 96*, 1996, pp. 147-154.

[13] 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, pp. 268-273.

[14] A. Witkin and M. Kass, "Spacetime constraints," in *Proc. of SIGGRAPH 98*, 1998, pp. 159-168.

[15] Y. Wu, T. S. Huang, "Capturing Human Hand Motion: A Divide-and-Conquer Approach", *Proc. of IEEE ICCV99*, 1999, vol. 1, pp. 606-611.