

Hand Modeling and Tracking from Voxel Data: An Integrated Framework with Automatic Initialization

Cuong Tran Mohan M. Trivedi

Computer Vision and Robotics Research Laboratory (CVRR)

University of California, San Diego, CA 92037

{cutran,mtrivedi}@ucsd.edu

Abstract

We propose an integrated framework for automated hand model initialization and tracking using voxel data. Starting with an initial specific hand pose, the Laplacian Eigenspace (LE) based segmentation method [7] is applied to segment hand voxel into different parts. This segmentation result is then used to extend the Kinematically Constrained Gaussian Mixture Model (KC-GMM) method for articulated body pose inference [2] with an automated hand model initialization. Our experiment with both synthesized hand voxel and real hand voxel captured from multi-perspective thermal cameras show that by combining the two methods, we have a more powerful system than using each one solitarily.

1. Introduction

Vision-based pose estimation of articulated body, e.g. full body or hand, has many potential applications including surveillance, advance HCI (Human Computer Interaction), 3D animation, intelligent environment, robot control, etc. This is however a challenging task and one main reason is the very high dimensionality of the pose configuration space. Some reviews of several techniques for articulated body pose estimation can be found in [3, 5, 8]. In the past few years, a lot of proposed methods make use of voxel data reconstructed from multiple camera views [1, 2, 4, 7, 9]. Compare to monocular approaches, voxel data can help to avoid issues of self-occlusion, image scale and provide more information to make the pose estimation task easier. Furthermore, efficient techniques for voxel reconstruction like shape-from-silhouette are also existed [6]. A common way for body pose estimation is the model-based approach, which consists of a model of the articulated body and a

procedure to fit that model to the observed voxel data. In [4], Mikic et al. proposed a hierarchical procedure for acquisition and tracking of full body model from voxel data, starting by locating head and torso using their specific shapes and sizes, then segmenting the remaining voxels to locate the limbs. Although this method is fully automated and may track even for large displacements, it lacks generality (i.e. only apply for full body model) because of using specific information about head and torso. Cheng et al. [2] propose a more general probabilistic method using KC-GMM for articulated body pose inference from voxel data and they did experiment with both full body model and hand model. This method however requires a careful manual initialization of the model, which is obviously an obstacle if we want to use the method for real time application. In [7], Sundaresan et al. proposed using a LE based method for segmenting full body voxel into different body chains then doing body model registration and estimation based on segmented result. In our proposed integrated framework, the LE based segmentation is applied to hand voxel at an initial specific pose, which clearly reveals the hand's structure. This segmentation result is then used to fill the gap of an automated hand model initialization for the KC-GMM method. The outline of this integrated framework is shown in Figure 1.

The remaining sections are organized as follows. Section 2 describes briefly the related research studies in [2, 7] and how they motivate our idea of combining them. Section 3 & 4 are about the implementation steps of the integrated framework shown in Figure 1. The experimental result is presented in section 5 and finally, we have some discussion in section 6.

2. Related Research Studies

2.1 KC-GMM method for articulated body pose inference [2]

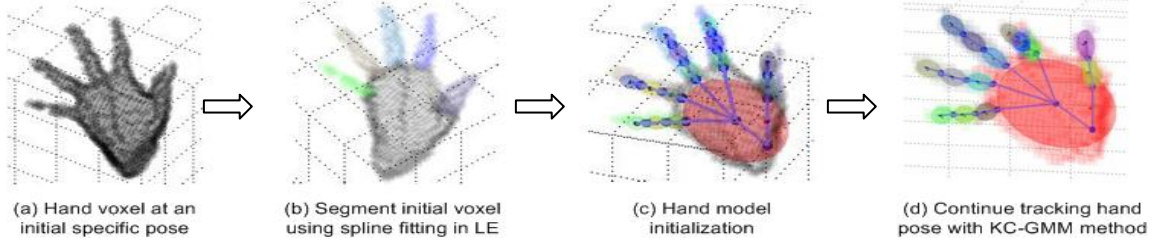


Figure 1. Outline of the proposed integrated framework

In this method, the articulated body pose is estimated using the same paradigm of probabilistic clustering. The hand model used in this method has a total of 16 components and 27 DOF (degree of freedom). Each component is described by a Gaussian and the set of components are kinematically constrained according to a predefined model. The goal is then to estimate optimal value for the Gaussian Mixture Model under those kinematic constraints. In [1] this is done by adding a constraining C-step to EM algorithm. However this C-step may compete with the M-step and cause instability in the optimization. The primary contribution of [2] is to remove this C-step by incorporating kinematic constraints into the probability model in the form of a prior probability to have a KC-GMM and then derive the EM algorithm for this new probability model.

This method is quite general (can be applied the articulated subject’s components can be described by Gaussians) and was applied for both full body and hand [2]. However, this method is not fully automated because it requires a manual initialization step. We may think of using hierarchical model acquisition procedure [4] for the initialization of full body model. In doing so, however, we will lose the generality of KC-GMM method, e.g. cannot apply it to hand model. The voxel segmentation using LE transformation method in [7] has generality so it is a more appropriate choice. Another issue is that due to the nature of EM algorithm, KC-GMM method could stuck in a sub-optimal solution when there is a large displacement.

2.2 Articulated body model segmentation in Laplacian Eigenspace (LE) [7]

This LE-based voxel segmentation is fairly general and can be applied for articulated object which is composed of long chains. It is shown that body chains like limbs, which have their length greater than their thickness, will form a 1-D smooth curve when mapped into high dimensional (e.g. 6D) LE. The procedure for LE mapping is as follows: First, we compute the adjacent matrix W of voxel data, such that $W_{ij} = 1$ only if voxel i is a neighbor of voxel j . Then, we

compute a D matrix, so that $D_{ii} = \sum_{k=1}^m W_{ik}$ and $D_{ij} = 0$ for $i \neq j$. The first d eigen vectors of $L=D-W$ with minimum eigen values give us the d basis of the needed LE.

After mapping into LE, a spline fitting process is used to segment the 1-D curves which results in the segmentation of their respective body chains. In [7], they applied this method to segment voxel data of full body into 6 chains (torso, head and 4 limbs) and they also propose techniques for registration and estimation more detailed body model from the segmented result. Their experiment with HumanEvaII dataset however indicates that the LE-based voxel segmentation is sensitive to voxel noise and will affects all the subsequent steps. This motivates the idea of an integrated framework, in which instead of doing voxel segmentation at every frame, we only use it for initialization purpose. In subsequent frames, using tracking based method like KC-GMM could help to overcome the sensitization to noise to some extent.

3. Hand voxel segmentation

Because fingers also have their length greater than their thickness, they will form 1-D smooth curves in LE but the palm will not. So in case of hand, the spline fitting in LE is only used to segment 5 fingers and the remaining voxels are considered as of palm. The segmentation result of applying spline fitting process in LE for hand voxel is shown in Figure 1.(b).

4. Hand model initialization and tracking

In the initialization step, we have segmented hand voxel and a template of hand model, which contains a predefined hand structure (i.e. number of components, number of joint, number of DOF). The goal is to adjust needed parameters including components dimensions, joints position and joints angles to achieve a hand model that fit the segmented voxel well. Because we require the hand to start with a specific pose (stretch pose), this initialization step can

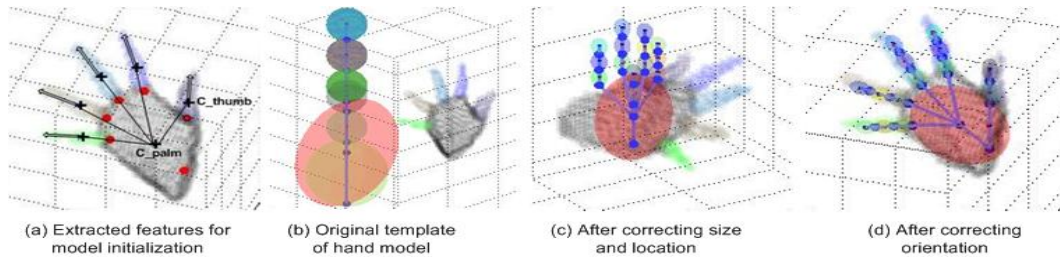


Figure 2. (a)- A simple procedure for hand model initialization
(b)(c)(d)-Hand model initialization result

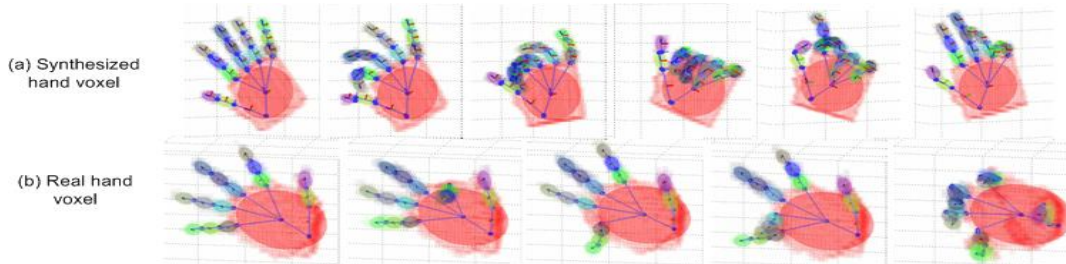


Figure 3. Visual result of hand modeling and tracking using the integrated framework

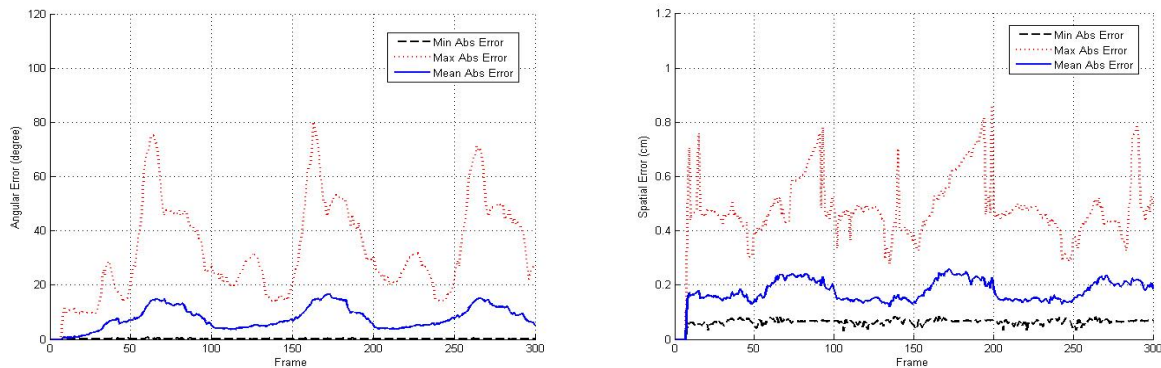


Figure 4. Quantitative result of synthesized hand data: Left - Angular error, Right - Position error

be done with the following simple and fast procedure. As shown in Figure 2.(a):

- For fingers registration, we first compute the center of each segmented voxel region (marked with plus sign). By constructing lines from palm center C_{palm} (which is already known) to all other centers, we see that the angle characteristic of the line from C_{palm} to C_{thumb} is distinguishable (i.e. the minimum angle to all other lines is the largest) and can be used to register voxel region of thumb and then other fingers.
- The local z-axis of each finger is computed as the largest PCA component of corresponding voxel region (marked with arrows). Because all fingers are in the same plane in this stretch pose, we can use the found local z-axis to compute local x-axis

and y-axis for each finger. From these axis, we can compute the orientation and the dimension of each finger (project voxel region on each local axis and find the range). With the assumption that each finger consists of 3 equal segments, all joint positions can also be found (e.g. red circles).

- For the palm, the local palm z-axis is determined by the line from C_{palm} to the “lowest” joint of the middle finger. Other palm parameters are then computed similarly as described above.

The result of hand model initialization is shown in Figure 2.(b)-(d). After initialization, KC-GMM method [2] is used for hand pose inference in subsequent frames.

5. Experimental results

We did experiment with both synthesized hand voxel and real hand voxel. With the same hand model in [2], synthesized data is constructed from cylinders of voxel and it simulates a periodic wave pattern moving. For real hand voxel, we set up 4 thermal cameras to capture hand images from multiple views. The background subtraction with thermal images is simple with an upper and a lower threshold respective to the temperature range of skin. Real hand voxel is then reconstructed using shape-from-silhouette technique. The results of automated hand model initialization and tracking in both cases were good as shown in Figure 3. With synthesized data, we also have quantitative result of angular and position error of hand components (Figure 4). We see that the plots are periodic with peaks at times when the hand is in nearly closed fist pose. However the error reduces when the hand opens and we do not lose track. For comparison, figure 5.(a) shows the failure of LE based hand voxel segmentation in the nearly closed fist pose mentioned above. Figure 5.(b) shows an incorrect pose estimation of KC-GMM when the hand model is manually initialized with correct scale but incorrect orientation. This incorrectness is conceivable because the nature of EM algorithm makes it easily stuck at a sub-optimal solution. We meet the same issue when there is a large displacement between frames. This result implies that our automated hand model initialization works well and it is definitely a better replacement of the previous manual initialization step.

6. Discussion

We have proposed an integrated framework that combines the KC-GMM method for articulated body pose inference and the spline fitting method in LE for articulated body voxel segmentation to have an automated hand model initialization and tracking system. Our experiment shows that this combination provides better result than using each previous method separately. This integrated framework also preserves the generality of both methods. When applying to other articulated body model, the initialization procedure needs to be changed. However, because we require the subject to start at a specific pose, it should not be difficult to develop a simple and good procedure for model initialization, e.g. we have applied the same integrated framework to full body model. Regarding the issue of sub-optimal solution, we may reinitialize the model when there is a suspect of sub-optimal solution, e.g. the likelihood is below a threshold. Nevertheless, a more general model initialization procedure is needed for doing this.

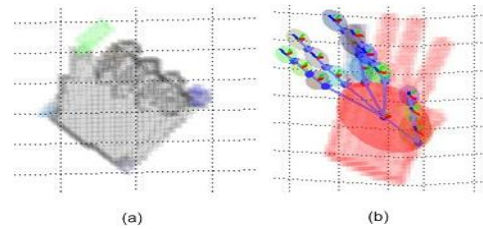


Figure 5. Results for comparison with successful tracking results in Fig. 3: (a) LE based voxel segmentation [7] failed in the nearly closed fist pose (fingers are not well separated) and (b) KC-GMM method [2] failed without careful manual initialization (initialized with correct scale but incorrect orientation).

Acknowledgement

We thank our colleagues at CVRR lab, especially Dr. Shinko Cheng for useful discussions and assistances. The first author also thanks Vietnam Education Foundation (VEF) for its sponsorship.

References

- [1]. S. Cheng, M. Trivedi. *Multimodal Voxelization and Kinematically Constrained Gaussian Mixture Model for Full Hand Pose Estimation: An Integrated Systems Approach*. IEEE ICVS, 2006.
- [2]. S. Cheng, M. Trivedi. *Articulated Human Body Pose Inference from Voxel Data using a Kinematically Constrained Gaussian Mixture Model*. CVPR EHM2, 2007.
- [3]. A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. *A Review on Vision-based Full DOF Hand Motion Estimation*. IEEE CVPR, 2005.
- [4]. I. Mikic, M. Trivedi, E. Hunter, P. Cosman. *Human Body Model Acquisition and Tracking using Voxel Data*. IJCV, 2003.
- [5]. T. Moeslund, A. Hilton, and V. Kruger. *A Survey on Advances in Vision-based Human Motion Capture and Analysis*. CVIU, 2006.
- [6]. G. Slabaugh, B. Culbertson, and T. Malzbender. *A Survey of Methods for Volumetric Scene Reconstruction for Photographs*. International Workshop on Volume Graphics, 2001.
- [7]. A. Sundaresan, R. Chellappa. *Model Driven Segmentation of Articulating Humans in Laplacian Eigenspace*. TPAMI, 2007.
- [8]. C. Tran, M. Trivedi. *Human Body Modeling and Tracking using Volumetric Representation: Selected Recent Studies and Possibilities for Extensions*. AMMCSS Workshop, 2008.
- [9]. E. Ueda, Y. Matsumoto, M. Imai, and T. Ogasawara. *A hand-pose estimation for vision-based human interfaces*. IEEE Transactions on Industrial Electronics, 2003.