Articulated Hand Configuration and Rotation Estimation using Extended Torus Manifold Embedding

Chan-Su Lee, Sung Yong Chun and Shin Won Park
Yeungnam University
{chansu, whiteyongi, psw0085}@ynu.ac.kr

Abstract

Due to the high degree of freedom found in hand motion, it is difficult to model articulated hand configurations. In addition, observed hand shapes vary according to the hand rotation, even when using the same hand configuration. This paper presents a new manifold embedding method for modeling low dimensional hand configurations and hand rotation using a 4D torus manifold, in which the product of three circular manifolds are used. Hand shapes are extracted from 3D depth and skin color models. The experiment results using synthetic motion and real hand motions from the Kinect sensor show simultaneous estimations of hand configuration and hand rotation parameters using particle filters on the extended torus manifold.

1. Introduction

There has been a great deal of research on natural human computer interaction in virtual environments, augmented reality, smart devices, and games. Input devices using human body motion are available for games and in smart devices such as Smart TV. Their functionality, however, is limited and remains far from natural manipulations and signaling gestures. The hand is one of the most effective interaction tools, due to its dexterous functionality in communication and manipulation [2]. One of the key components in understanding the manipulation of objects and signaling by hand is in the estimation of the articulated hand configuration without wearing complicated devices, such as gloves. Therefore, the estimation of hand motion from 2D and 3D images has become an interesting research area in computer vision and pattern recognition [5, 4, 8].

Recently, manifold-based alternative approaches have been applied to articulated human motion tracking that utilize the intrinsic constraints of body configurations using low dimensional manifold embedding [9, 6]. In regards to hand pose estimation, manifold learning based approaches have been applied for hand posture estimation [3], where an Isometric Self-Organizing Map (ISOSOM) was used for effective 3D hand pose estimation using a low dimensional nonlinear manifold. Hand rotation during the hand configuration change between two key poses was modeled using cylindrical manifold embedding [7].

This paper presents a new manifold embedding method to model various key hand configurations and their variations over time (dynamics), and the view change from hand rotation by extending a torus manifold embedding to a higher dimensional manifold. The previous torus manifold embedding method was applied to model cyclic human motions i.e. walking [1]. The motion manifold is homeomorphic to a circle. Cylindrical manifold embedding can be used to model the configuration change between two key poses [7]. However, it is still a difficult problem to extend this model to multiple key poses. 4D torus manifold embedding, which is the product of three one dimensional circle manifolds, is therefore proposed to model key pose variations and simultaneous configuration changes between key poses, along with the circular hand rotation in each configuration.

2. Modeling Arbitrary Hand Pose Variations

All of the possible hand pose variations may need to be captured when modeling a very accurate hand configuration space. However, it is not possible to capture all of the possible hand pose configurations and it may not be necessary to do so.  

2.1 Problem Formulation: Articulated Hand Pose Transitions

We can model arbitrary hand configuration variations from one hand configuration to another by using
the hand transition between two distinctive hand configurations. Therefore, when we collect all of the possible key poses (configurations) and their pairwise transitions, we can model arbitrary hand configuration changes. If we choose a closed hand as a base hand configuration, then all of the extreme key hand poses can be represented by a combination of the open fingers. When we have determined all of the transitions between the closed hand and other extreme key hand poses, we can model the arbitrary in-between pose variations by a combination of the pose transition between the closed hand and each key pose.

2.2 Data Collection

We use graphic tools to collect synthetic hand configuration silhouette data from multiple views simultaneously. Since we model shape variations not only by hand configuration changes but also by hand rotational changes, we need to collect shape variation data along the view circle for each hand configuration. The extreme key hand pose can be modeled by different combinations of bending fingers. We selected 12 sample key poses for our experiments. For each key pose, we generated the configuration changes from a closed hand to the key pose and back again with 18 samples. For each given configuration, we sampled the view variation by rotating the configuration by 20 degree intervals; 18 view sample shapes were collected for each configuration. Therefore, a total 3888 samples (12 × 18 × 18) were collected in our experiments. Figure 1 shows sample sequences of the 12 key poses, the configuration or view variations in a fixed key pose.

3. Extended Torus Manifold Embedding

From the collected dataset, we applied linear and nonlinear dimensionality reduction techniques i.e. PCA, Isomap, and GPLVM in order to find the low dimensional representation of the manifold space. These low dimensional embeddings show partially circular manifold structures from the view and configuration variations. However, it is hard to model or parameterize the embedding space based on the view circle or hand configuration.

When a hand configuration sequence changes from the base hand pose (closed hand) to a specific key hand pose, and returns back to the base hand pose, the sequence can be modeled as a circular one dimensional manifold; the sequence is homeomorphic to a circle. Additionally, the shape variations along the view circle can be modeled into another independent circular manifold. Therefore, the transition from the base pose to a specific key pose with view circles can be modeled by torus manifold [1] when the configuration returns back from the target key pose to the base pose. The question is how to extend the model to cover various key poses to model arbitrary hand configuration transitions.

3.1 4D Torus Manifold Embedding

Key hand configurations can be embedded along circular manifolds based on the pairwise distance among key hand poses. Given a key hand configuration, the pairwise distance can be determined from the pairwise distance of the key hand configuration joint angles. Given a hand configuration distance metric, the optimal key hand pose configuration on the circular manifold can be defined as the shortest-closed loop needed to travel through all of the key hand poses by the distance metric between the key poses.

Accordingly, we obtain three circle manifolds (S1), all of which are independent from each other. These three circle manifolds product(S1 × S1 × S1) can be represented by a 4 dimensional Euclidian space with three independent parameters as follows;

\[
\begin{align*}
  w &= (r_3 + (r_2 + r_1 \cos(\theta))\cos(\phi))\cos(\psi), \\
  x &= (r_3 + (r_2 + r_1 \cos(\theta))\cos(\phi))\sin(\psi), \\
  y &= (r_2 + r_1 \cos(\theta))\sin(\psi), \\
  z &= r_1 \sin(\theta),
\end{align*}
\]

where \( r_1, r_2, r_3 \) are the radius of each circular manifold, \( \theta, \phi, \psi \) are the parameters for the key hand pose, hand temporal configuration between the base and key hand pose, and the hand rotation(view angle) respectively. Therefore, the collected data for each key pose with hand configuration variations between the base hand pose and key hand pose with the circular view change in each configuration can be mapped on a one-to-one basis to this extended torus manifold.

3.2 The Nonlinear Generative Model with 4D Manifold Embedding and Tracking

Nonlinear mapping can be learned between the embedding space and the original shape images of the hand with hand configuration and view variations. The hand shapes are represented by a signed distance function, where the boundary silhouette is represented by a 0 value. Nonlinear mapping from embedding points \( x_i = (w_i, x_i, y_i, z_i) \) to the corresponding hand shapes \( y_i \in \mathbb{R}^n \), where \( n \) is the number of observation shape, can be learned using a generalized radial basis function (GRBF) similar to in [1]. Figure 2 shows the interpolation of new sequences from the trajectories on 4D torus embedding manifolds.
Using the proposed 4D torus manifold embedding and its generative model, the hand shape with the hand configuration change and hand rotation can be estimated based on the 4D torus embedding points. The estimation is accomplished through the key hand pose parameter \( \theta \), temporal configuration parameter \( \phi \), and hand rotation parameter \( \psi \). Particles on the manifold can be determined using particle filtering on the manifold. The hand state \( x_t = (\theta_t, \phi_t, \psi_t) \) is estimated from a given observation \( y_t \) using particle filter as follows:

\[
P(x_t | y_t) \propto P(y_t | x_t) \int_{x_{t-1}} P(x_t | x_{t-1}) P(x_{t-1} | Y^{t-1}) dx_{t-1}
\]

As a result of this particle filtering, the mode or the MAP (maximum a posterior) of each parameter can be estimated.

The hand rotation or view parameter can be computed directly from the estimated hand rotation parameter, \( \psi \), by simple scaling as necessary. In the case of body configuration, we have to estimate one body configuration from the estimated key hand pose \( \theta \) and the phase of the configuration \( \phi \). We learn a interpolation function based on the Radial Basis function for the given key hand pose parameter and phase of the configuration parameter to the joint angles of the hand configuration as \( f(\theta_t, \phi_t) = \Omega(q_1, \cdots, q_{N_h}) \). This allows the synthesis of the hand configuration from any estimated key hand pose parameter and phase of the configuration.

4. Experiment Results

We performed experiments on two data sets: a synthetic data set, and a real data set from a Kinect sensor.

4.1 Synthetic Data

For the synthetic data, we generated 37 frames having simultaneous view and hand configuration variations. The generated sequence of data were rearranged to have the same center of gravity for each foreground shape, and resized into 66 \( \times \) 73. 15 particles for the key pose \( \theta \), 15 particles for the phase of hand configuration \( \phi \), and 15 particles for the hand rotation \( \psi \), total of 3375 \((15 \times 15 \times 15)\) particles were used for the estimation of the 4D torus embedding from the observed shapes. Figure 3 shows the sample input sequences, estimated parameters, and reconstructed sequence from the estimated parameters. Poser® was used to synthesize the new hand shape from the estimated finger joint angles.

4.2 Real Data from Kinect

We used both the depth information and skin color information for the robust hand shape extraction. To extract hand shape from real data, skin color or foreground image after background subtraction can be used. Through a combination of the depth and skin color, we can extract an accurate hand shape from just the image sequence. At first, we used a hand detector, which was developed based on a cascade Adaboost [10] for several static hand shapes. This detector was used to initialize the ROI (Region of Interest), used to define the boundary for the shape depth segmentation and skin color estimation. In the case of the depth segmentation, we specified a depth (z axis) boundary to extract only the hand area; skin color was also applied to find hand area. By using a logical operation and morphological operation from the extracted two type of binary hand shape, we can extract an accurate hand shape. The extracted hand shape images are normalized in a similar
5. Conclusion

In this paper, we presented a new framework to estimate hand configuration and hand rotation for various hand shapes using higher order torus manifold embedding. The proposed method can be applied to obtain accurate estimations of hand configuration and rotation during the manipulation of virtual objects in virtual environments. It can also be used for sign language recognition, especially for hand spelling based on estimated hand configuration in various view environment.

Acknowledgements: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology(2012003830) and by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation.

References