Tracking 3D Hand Postures from Monocular Video

Iek-Kuong Pun* Hsing-Han Wang† I-Chen Lin‡
National Chiao Tung University

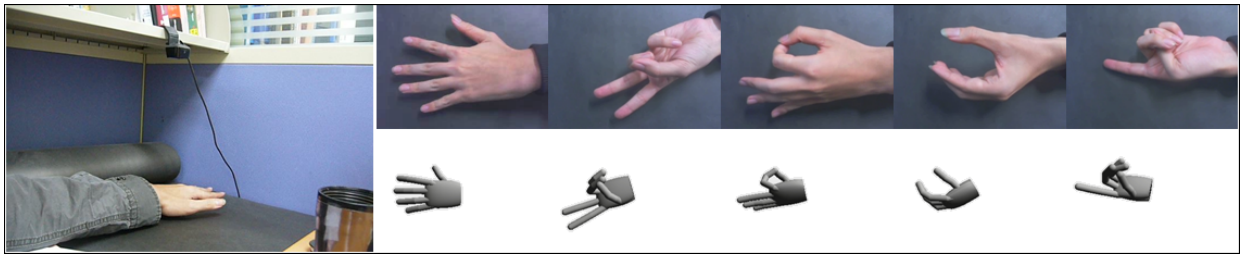

Figure 1: *The proposed interactive-rate method for 3D hand motion estimation. Left: the experiment environment; the upper and lower sequences are the input real images and the estimated hand articulation, respectively.*

Abstract

We estimate 3D hand articulation from monocular video without markers. To efficiently tackle the ill-posed problem, we propose a hierarchical method that searches the approximate nearest neighbors in a small real-hand image set and a large synthetic 3D hand posture database. Temporal consistency is also considered to disambiguate the approximate results. Our prototype system can preliminarily estimate hand gesture, rotation and short-term missing situation in an interactive rate.

1 Introduction

Our human hand is an articulated object with more than 20 degrees of freedom (DOFs). Textures and colors of different articulated parts are similar. Since the captured images results from perspective projection, the shapes in a camera view are of considerable variety and self-occlusions. Without sufficient prior information or constraints, it is difficult to estimate the optimum postures from such high dimensional space in real time.

Martin de La Gorce et al. [2011], built a delicate polygonal mesh and synthesized the hand projection image for hand pose optimization. The model provided state-of-the-art pose estimate on complicated background, high DOFs and occlusion sequence. However, their method took considerable time on synthesizing novel projection images. Wang and Popovi [2009] proposed using a glove with color patches to match hand gestures in database, the results were further refined by inverse kinematics. They successfully tracked many commonly used hand gestures and even random jiggling of the fingers. We aim at tracking an articulated hand without markers. Our motion model includes 20 DOFs for the joint angles and 6 DOFs for orientation and location. We propose a data-driven technique using hierarchical approximate nearest neighbor (ANN) search to efficiently estimate hand motion.

2 Our Approach

The core of our approach is to efficiently search the most likely posture from databases for a given bare-hand image. To balance the user training procedure and data completeness, we utilize two sets of image data sets. During initialization, a user is required to follow a short sequence for gathering training data. These user-dependent and computer-labelded real hand images are then mapped onto a large image database, generated from a 3D hand model. At the present, we match two images by the silhouette area and the Chamfer distance of edges.

Given a query image, our system first find a few ANN groups based on the small real image dataset. These groups are then regarded as searching seeds in a large database. After evaluating the most likely postures according to data similarity and temporal coherence, our system then report the estimated postures by weighted blending. Our experiments perform on a desktop with an Intel i5-760 processor. The interactive posture estimation can reach around 10 frames per second (FPS). Since only one camera is applied , one weakness of this system is its unstability in z-direction movement. We also plan to explore more reliable and efficient features in data matching.

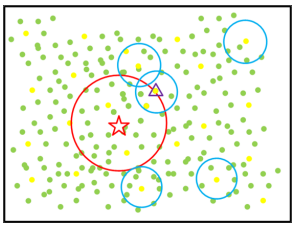

Figure 2: *The relationship of data samples in the searching problem. Yellow points: real image data in the small database; green points: the synthetic data; the red star: the previous best pose; the triangle: the current best pose; the red circle: the neighbors of the previous frame; the blue small circles: the high posterior regions by ANN search on the real image database*

References

DE LA GORCE, M., FLEET, D. J., AND PARAGIOS, N. 2011. Model-based 3d hand pose estimation from monocular video. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 9 (Sept.), 1793–1805.

WANG, R. Y., AND POPOVIĆ, J. 2009. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics 28*, 3 (Aug.), article 63.

*e-mail:louispun.csie94@nctu.edu.tw
†e-mail:amamiya623@gmail.com
‡e-mail:ichenlin@cs.nctu.edu.tw

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008