2009 11th IEEE International Symposium on Multimedia

Introducing "XMOB": Extremity Movement Observation Framework for Upper Body Pose Tracking in 3D [1]

Cuong Tran Mohan M. Trivedi
Computer Vision and Robotics Research Laboratory (CVRR)
University of California, San Diego, CA 92092
{cutran,mtrivedi}@ucsd.edu

***Abstract*— We introduce a vision based, markerless upper body pose tracking approach that first tracks the 3D movements of extremities, including head and hands. Then based on the knowledge of upper body model, these extremity movements are used to predict the whole upper body motion as an inverse kinematics problem. The experimental validation showed the promise of applying this approach in several smart environments and HCI situations, e.g. user activity observation in driving scene, meeting room, teleconference scene.**

Keywords- ***Upper Body Motion Tracking; Upper Body Pose Estimation; Inverse Kinematics; Head and Hands Tracking***

I. INTRODUCTION AND MOTIVATION

Vision based human pose tracking is challenging due to the exponentially large search space of possible human poses, the self occlusion, the variations in human appearance and lighting condition. In several realistic applications including driver assistance system, smart teleconference, smart meeting room, we observe that a user typically sits in a fixed position and the arms carry the most influential information of body motion. Compared to the general case, these observations imply more restrictions on the space of possible poses which can be exploited to develop an efficient upper body pose tracking system for those situations.

We can roughly categorize human pose tracking approaches into monocular approaches (e.g. [2]) and multiview approaches (e.g. [1, 3, 7]). With the goal of tracking real 3D pose, we chose to use multiview inputs which help to reduce the self occlusion issue and provide more helpful information Motivated from studies in neurophysiology which found that the desired position of the hand roughly determines the arm posture [4], we introduce a multiview approach for upper body pose tracking using 3D movement of extremities (head and hands) which is called XMOB (eXtremity Movement OBservation) upper body pose tracker. In this approach, the high dimensional search problem of 3D upper body pose tracking is broken into two sub-problems (reduce the complexity): First, the 3D movements of head and hands are tracked from multiview inputs. Then based on the knowledge of upper body model, these 3D extremity movements are used to predict the whole upper body motion as an inverse kinematics problem. Compared to related studies in human pose tracking [1, 2, 3, 5, 7], XMOB only requires image evidences to track extremities (head and hands) which are the easiest parts to track with little occlusion. This means XMOB could work even when some inner body parts are occluded or when their colors (e.g. user clothes) are mixed with the background.

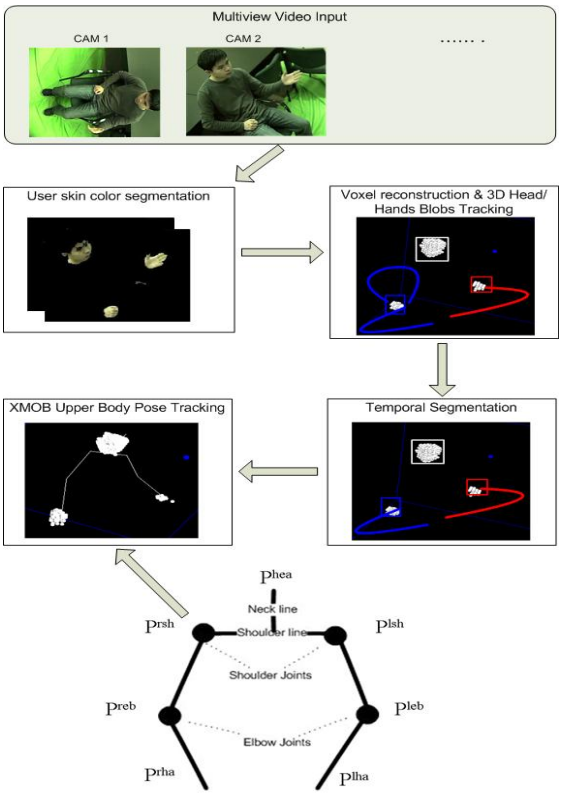

Figure 1. Outline of the XMOB framework for upper body tracking The used upper body model is shown at the bottom.

II. XMOB FRAMEWORK FOR 3D UPPER BODY POSE TRACKING

Fig. 1 shows steps in the proposed XMOB framework for upper body pose tracking. From multiview video input, the skin regions of the face and hands are segmented to produce silhouettes which are used to reconstruct and track 3D voxel blobs of the head and hands. Since we need to observe the dynamics of the head and hands (not just their positions in a single frame) to predict the motion of the whole upper body, we first temporally split the head and hands movements into separate subsequences. Then using the knowledge of the upper body model, XMOB predicts the most likely upper body motion corresponding to each subsequence of head and hand movements.

[1] Please review the video demonstration (4 minutes) of the XMOB in real-time operation in a Laboratory and Vehicle settings at: http://cvrr.ucsd.edu/~ctran/papers/XMOB_Result.avi

978-0-7695-3890-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ISM.2009.89

Two main contributions in this framework are: (i) A semi-supervised procedure, in which user starts by moving only their extremities, to robustly segment skin color of a particular user from background colors of a particular scene; and (ii) an inverse kinematics method to predict the whole upper body motion from 3D extremity movements. Here we took a numerical approach: Using geometrical constraints of upper body joints and extremities, at each frame XMOB determines a set of hypotheses for possible inner joint locations. By observing extremity movements over a period of time, XMOB selects the hypotheses sequence that minimizes the joint displacement

III. Experimental Validation

XMOB uses 2 color cameras configured with a wide baseline to observe 3D movement of head and hands. Several data sequences were captured. In each sequences, there is a single user sitting in a fixed position and doing some random activities with their arms like clapping, simulating some gesticulations in their normal speaking, and in a driving situation. In order to have a quantitative evaluation for a few sequences, we also use marker based motion capture system to obtain the ground truth of upper body motion simultaneously with the video data.

The system runs at about 6 fps on a Pentium(R) D CPU 2.8 GHz. Fig 2 shows some results for visual evaluation of 3D upper body pose tracking compared to the ground truth as well as the superimposed 3D result on input images. Fig 3 shows the estimated 3D position of elbows compared to the ground truth which indicates the ability of XMOB in capturing pattern of joints movement. This means that further analysis such as gesture recognition can use these estimates instead of the marker-based motion capture data. Additional experimental results can be found here [1].

IV. Concluding Remarks

We have proposed XMOB system for upper body pose tracking. XMOB only requires image evidence of extremities which typically have less occlusion than other body parts. By breaking the problem of high dimensional search for upper body pose tracking into two sub problems, the complexity is also reduced to achieve real time performance. The experimental validation showed the promise of using this approach for several realistic applications, in which we are particularly interested in applications for driver assistance systems, e.g. [6].

References

[1] S. Cheng and M. M. Trivedi, "Articulated Human Body Pose Inference from Voxel Data Using a Kinematically Constrained Gaussian Mixture Model", *IEEE CVPR EHuM2*, 2007

[2] V. Ferrari, M. Jiménez, A. Zisserman, "Progressive Search Space Reduction for Human Pose Estimation", *IEEE CVPR*, 2008.

[3] I. Mikic, M. M. Trivedi, E. Hunter, and P. Cosman, "Human Body Model Acquisition and Tracking using Voxel Data", *IJCV, Vol. 51, Issue 3*, July-August, 2003.

[4] J. Soechting and M. Flanders, "Errors in pointing are due to approximations in sensorimotor transformations", *Journal of Neurophysiology, 62(2):595–608*, 1989.

[5] C. Tran and M. M. Trivedi, "Hand Modeling and Tracking from Voxel Data: An Integrated Framework with Automatic Initialization", *IEEE ICPR*, 2008.

[6] C. Tran and M. M. Trivedi, "Driver Assistance for Keeping Hands on the Wheel and Eyes on the Road", *ICVES*, 2009.

[7] J. Ziegler, K. Nickel, and R. Stiefelhagen, "Tracking of the Articulated Upper Body on Multi-View Stereo Image Sequences", *IEEE CVPR*, 2006.

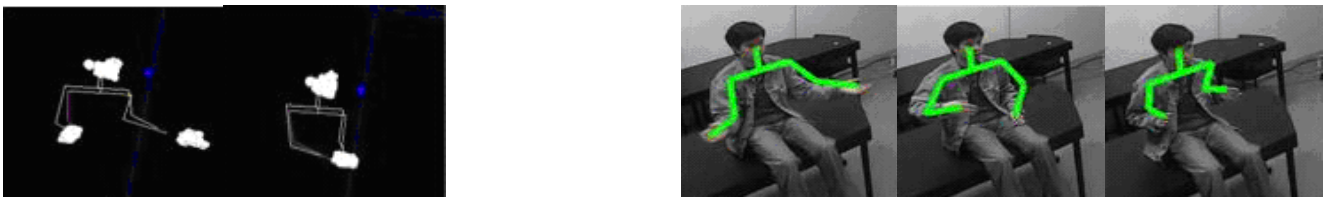

Figure 2. Left - Upper body pose tracking results in 3D (color lines) compared to the ground truth (white lines). Right – Overlayed 3D pose tracking results on image.

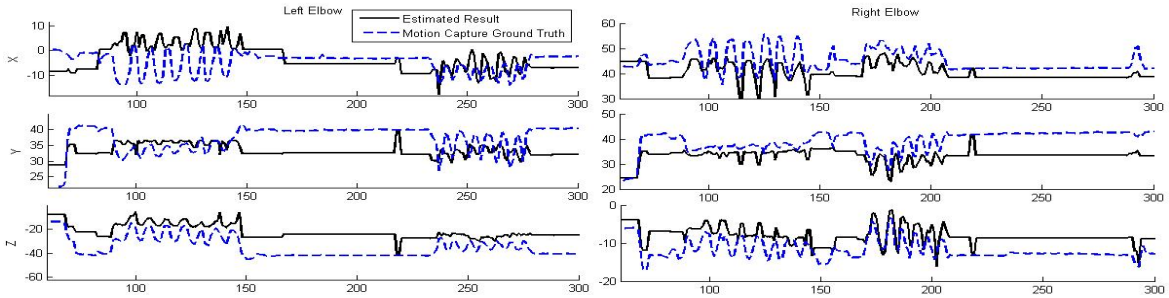

Figure 3. Quantitative plot of the estimated elbow positions (solid black lines) compared to the ground truth (dotted blue lines). We see that the estimates can capture movement patterns of the elbows. A base error in those plots is understandable since the ground truths (marker positions) in this experiment are also not absolutely exact elbow positions.

[1] Please review the video demonstration (4 minutes) of the XMOB in real-time operation in a Laboratory and Vehicle settings at: http://cvrr.ucsd.edu/~ctran/papers/XMOB_Result.avi