

# HUMAN MOTION TRACKING IN 3D VIDEO

P. Huang and A. Hilton

Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

---

## Abstract

This paper presents a method to track human body parts in 3D video sequences for character animation production. A voxel-based Local Shape Histogram descriptor is first extracted as geometric features. A graph-based Multiple Objects Tracking scheme is then used to track these features. Finally, the evaluation against a publicly available real 3D video database demonstrates the performance.

---

**Keywords:** Human Motion Tracking, 3D Video, Local Shape Histogram, and Graph Optimisation.

## 1 Introduction

The reuse and manipulation of captured 3D video sequences for animation is difficult because each frame is reconstructed individually, therefore, no temporal consistent structure is available. Recovering the temporal consistence requires finding dense temporal correspondences. SIFT features are widely used for this purpose since it is local, discriminative and robust [1]. However, these photometric features cannot be guaranteed to be well distributed across the whole human body. For a typical 3D video, body parts like limbs and head have few or less SIFT features available due to less distinguish texture or visual patterns. This work is motivated to compensate the coverage of SIFT features for the human body parts, particularly, focus on the extremities of the limbs and head.

## 2 Local Shape Histograms

Previous work on Global 3D Shape Similarity using a spherical voxel-based Shape Histogram [2] is adapted to a Local Shape Histogram (LSH) with one shell in a local radius. A volumetric representation is first constructed by dividing the space into a voxel grid  $v$  where  $v(i, j, k) = 1$  for an occupied voxel. Since the interest exists in finding surface correspondences, we only compute LSH for surface voxels. Given a surface voxel  $s$ , the LSH is computed as follows,

$$LSH(s) = \sum_{d(s,v) < r} v(i, j, k) \quad (1)$$

where  $v$  denotes a voxel lies in the neighbourhood of  $s$ ,  $(i, j, k)$  the grid indices, and  $r$  the local radius. To reduce the effect of global topology change, e.g., hand touching the body,  $d(s, v)$  calculates geodesic distance from  $s$  to  $v$ . Since each LSH is a scale value, it is easy to sort them from strongest to weakest. Here, we define the strongest feature as the one with smallest scale value, because it is most likely to be a terminal point at limbs and head. An example of the scale value mapping on the whole body is shown in Figure 1, where strongest feature corresponds to dark blue colour.

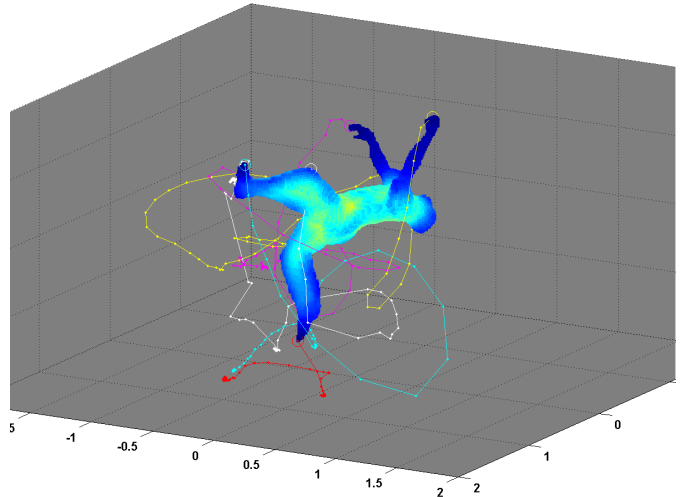


Figure 1: Trajectories for JP-flash-kick 3D video sequence.

## 3 Graph-based Multiple Object Tracking

The Graph-based Multiple Objects Tracking scheme works given two reasonable assumptions: first, a true track tends to be the one with shortest traverse distance; second, two tracks cannot occupy the same position at the same time. To-be-tracked candidates are first selected per frame in this way adding the strongest feature into an empty group and following ones while keeping all added features well distributed by forcing the geodesic distance between each pair of added features greater than a pre-defined threshold. User manually selects start-to-track candidates at the first frame. A standard Kalman filter is applied for each candidate individually to suggest all possible associations in next frame. A trellis-like graph is then constructed where a slice corresponds to a time instant, a node to a candidate, and an edge to a possible association in the next frame with edge cost equal to the Euclidean distance in between. Finally, a graph optimisation is performed to find a group of graph paths, which minimise the total cost of all paths while keeping these paths compatible (not overlapping with each other). The proposed method is evaluated against a publicly available real 3D video database [3]. An example of tracking results is shown in Figure 1.

**Acknowledgements:** This work was funded by EPSRC Grant EP/E001351 and EU IST project i3dpost.

## References

- [1] A. Doshi, A. Hilton and J. Starck. An Empirical Study of Non-rigid surface feature matching. CVMP 2008.
- [2] P. Huang, A. Hilton, and J. Starck. Shape Similarity for 3D Video Sequences of People. IJCV 2010.
- [3] <http://cvssp-data.eps.surrey.ac.uk/cvssp3d>