Víctor Ponce
Miguel Reyes
Xavier Baró
Mario Gorga
Sergio Escalera

Centre de Visió per Computador

Universitat Autònoma de Barcelona

Two-level GMM Clustering of Human Poses for Automatic Human Behavior Analysis

Abstract

Detect human poses is a main step in the study of human behavior analysis. Our achievement is to find a non supervised method to determine key poses of a human gesture for a posterior analysis of a human behavior using machine learning and computer vision techniques. We use Kinect system for body feature extraction from joint positions of the articulated human models and then we apply a global and local feature clustering using a two-level Gaussian Mixture Model approach for the expert evaluation by comparing each clustering levels through their visualizations.

Keywords: Human Pose, Human Behavior Analysis, Clustering.

Methodology

Human Pose Representation

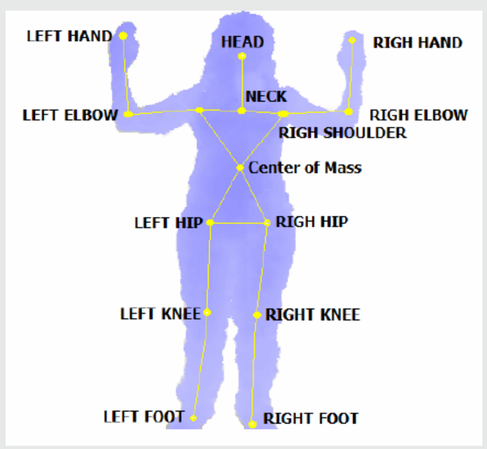

Figure 1: The 3D articulated human model consisting of 15 distinctive points.

Depth maps acquired by public API OpenNI software [1]. These features are able to detect and track people to a maximum distance of six meters from multi-sensor device.
Method of [2] used for the detection of human body and its skeletal model. The approach of [2] uses a huge set of human samples to infer pixel labels through Random Forest estimation, and skeletal model is defined as the centroid of mass of the different dense regions using mean shift algorithm.
The articulated human model is defined by the set of 15 reference points. This model has the advantage of being highly deformable, and thus, able to fit to complex human poses.

$$V_j = \{\{v^1_{j,x}, v^1_{j,y}, v^1_{j,z}\}, \ldots, \{v^{14}_{j,x}, v^{14}_{j,y}, v^{14}_{j,z}\}\}$$

Human Pose Clustering

Grouping the previous pose descriptions in pose clusters with standard Gaussian Mixture Model. Our goal is to group the set of frame pose descriptions in clusters so that posterior learning algorithms can improve generalization in Human Behavior Analysis systems. We use a full covariance GMM of K components parameterized:

$$\theta = \{\pi(k), \mu(k), \Sigma(k), k = 1..K\}$$

Then, a likelihood value based on the probability distributions $p(\cdot)$ of the GMM is obtained:

$$GMM(V, k, \theta) = \sum -\log p(V \mid k_i, \theta) - \log \pi(k_i)$$

1. **First level Clustering**: Use three spatial components of descriptor V for each joint *i, i ∈ [1, … , 14]* and perform GMM of k^1 clusters, namely GMM^i_1
2. **Second level Clustering**:
   1. Define for each pose a new feature vector, $V^* = \{v^1_1, .., v^1_{k^1}, .., v^1_1 4, .., v^1_{k^1} 4\}$ of size $14xk^1$, where v^j_i is the probability of applying GMM model GMM^i_1 at features from V corresponding to spatial coordinates of *j.th* joint.
   2. Use components of descriptor V and perform GMM of k^2 clusters, namely GMM_2.

Given a new frame, then, human skeleton is obtained as described before, and feature vector V is computed and tested using two-level GMM description, obtaining final probability for the most likely cluster from the set of k^2 possible clusters.

Results

Figure 2: Sample examples from clusters defined using one-level GMM (left) and two-level GMM (right), respectively.

Data: Set of gestures using the Kinect device consisting of seven different categories. It has been considered 10 different actors and different environments, having a total of 130 data sequences with 32 frame gestures. Thus, the data set contains the high variability from uncontrolled environments. The resolution of the video depth sequences is 340x280.

Methods and parameters: The people detection system used is provided by the public library OpenNI. This library has a high accuracy in people detection, allowing multiple detection even in cases of partial occlusions.

Figure 3 shows consecutive visual descriptions of some data set gestures. At the bottom of the sequences, we show a first row that represents the cluster number assigned by a one-level GMM, and a second row with the assigned cluster using two-level GMM. One can see that in most cases both grouping techniques assigns consecutive poses to same clusters. However, the clusters assigned by the one-level GMM have more visual variability, being inefficient for human behavior generalization purposes.

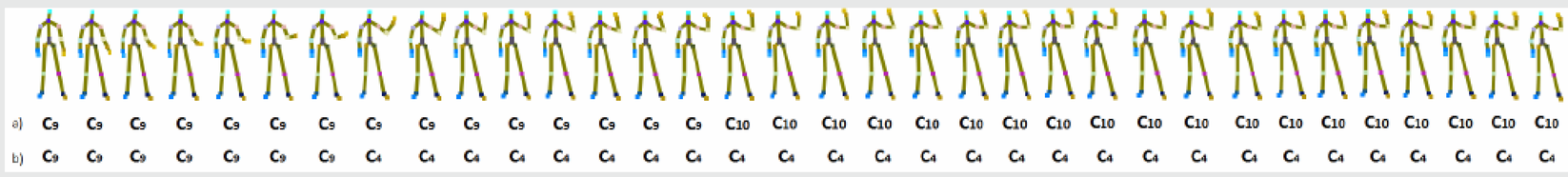

Figure 3: Cluster assignment for some gesture pose sequences.

Conclusion

In this paper, we designed a data set of human actions and described individual frames using pose skeleton models from depth map information. We proposed a two-level GMM clustering algorithm in order to group similar poses so that posterior Human Behavior analysis techniques can improve generalization. We showed some preliminary qualitative results comparing our approach with the classical one-level GMM clustering strategy, showing a more visual coherent grouping of poses.

Refrences

[1] Open natural interface. November 2010. Last viewed 14-07-2011 13:00.
[2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. *CVPR*, 2011.
[3] V. Ponce, M. Gorga, X. Baró, S. Escalera. 'Human Behavior Analysis from Video Data using Bag-of-Gestures. International Joint Conference on Artificial Intelligence' (IJCAI 2011). ISBN 978-1-57735-515-1 Vol 3, ISBN 978-1-57735-516-8 (electronic proceedings), AAAI Press, pp. 2836-2837.

Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Spain
Centre de Visió per Computador, Universitat Autònoma de Barcelona, Spain
E-mail: v88ponce@gmail.com, mreyese@gmail.com, xbaro@cvc.uab.es, mario.gorga@gmail.com, sergio@maia.ub.es