

3D POSE ESTIMATION FROM SILHOUETTES IN CYCLIC ACTIVITIES ENCODED BY A DENSE GAUSSIANS MIXTURE MODEL

S. Amin Dadgar, Jean-Christophe Nebel and Dimitrios Makris
Digital Imaging Research Centre, Kingston University, London, U.K.

Keywords: 3D Pose Estimation, Principle Component Analysis, Gaussian Mixture Model, Annealed Particle Filter.

Abstract: This paper presents a system for 3D Pose estimation of cyclic activities (e.g. walking, jogging). Principal Component Analysis is used to compress the high dimensional space of poses. Human activities are encoded by Hidden Markov Models, overlaid on Gaussian Mixture Models. A generative approach based on the Annealed Particle Filter is used to estimate poses from silhouettes derived by a monocular camera. Experimental results indicate the value of the proposed Dense Gaussian Mixture Model when initialised by a gait cycle.

1 INTRODUCTION

Visual analysis of human motion is one of the most active research fields in computer vision, because of its promising applications in many areas such as visual surveillance, perceptual user interface, content-based image storage and retrieval, video conferencing, athletic performance analysis, virtual reality. Human body pose estimation, which is the process of estimating the configuration of the human body parts over time, is a prerequisite to human motion analysis. While specialised Motion Capture (MoCap) systems can produce satisfactory results, their applicability is limited, because of their intrusive character (i.e. requirement to attach sensors/markers on the human body). Therefore, non-intrusive visual-based systems seem to be the attractive alternative that can be easily employed in a variety of sites and applications.

In recent years, there has been a clear shift of research from 2D image-plane (Kou, 2007) to 3D articulated motion analysis (Balan, 2005), (Roth, 2004), (Sidenbladh, 2000) to overcome limitations of specific angle-views and ambiguous depth information in 2D representations. Since 3D pose recovery from a 2D view is an ill-posed problem, researchers attempt to use multiple views (Darby), (Lee 2007), (Bhatia, 2004) and/or learnt models of specific activities (Deutscher, 2005), (Hogg, 1983),

(Howe, 2004), (Howe, 2005), (Lakany, 1999) to constraint pose estimation.

Assuming a specific cyclic activity such as walking and jogging this paper presents a system which is capable of estimating a sequence of 3D poses from a 2D image sequence (2D video) captured by a monocular camera.

Sequences of 3D poses derived by a Motion Capture system are used to train activity models. Principal Component Analysis (PCA) is used to reduce the dimensionality of the pose space and a Hidden Markov Model (HMM) overlaid on a Gaussian Mixture Model (GMM) encodes the dynamics of the activity. In this work, we propose the usage of a Dense GMM initialised by a gait cycle as the base of the HMM.

Inferring the image data is constrained by the learnt activity model. Specifically, an Annealed Particle Filter (APF) (Balan, 2005), is used to generate 3D pose hypotheses. Then, hypotheses are assessed by comparing their 2D projection with the input image silhouettes in a Bayesian Framework.

2 METHODOLOGY

Here we define the estimation of a 3D Human Body pose as the recovery of the configuration of a set of body parts. Given a sequence of image observations (i.e. sequence of 2D contour) $\{y_t\}, t \in \{1 \dots T\}$,

the proposed system aims at estimating the sequence of state vectors $\{\tilde{x}_t\}$, which represent the configuration of the 3D poses of the human body within an specific cyclic activity (e.g. walking).

2.1 Learning

Training data is acquired from a MoCap device for a specific cyclic activity. 3D Poses are represented by the joint angles of a standard human model of 10 body parts. The space of unconstrained poses is inevitably highly dimensional (32D). Since searching in this space is not efficient, we reduce the dimensionality of the solution space to 8D using PCA.

The 8D samples of the training data are then used to estimate a GMM that constrains the solution space. In particular, two different solutions are investigated:

a) A mixture of 4 Gaussians (equal to the number of gait phases in walking/jogging) that is estimated by Expectation Maximisation (EM).

b) A Dense GMM initialised by a gait cycle. This implies that the number of Gaussian clusters is equal to the number of pose samples in the initialising gait cycle.

Finally, the dynamics of the activity are modelled by a HMM overlaid on the GMM.

2.2 Testing

Poses are inferred from image data using a generative approach. Specifically an APF is used to generate particles in the 8D PCA space constrained by the HMM. A reverse PCA transformation of these particles results in 3D pose hypotheses.

3D pose hypotheses are projected on the image plane using the camera calibration model, which is assumed to be known. The projections of the 3D hypotheses are then assessed using the following formula:

$$p(y_t|x_t) = \frac{p(s_t)}{C(y_t, x_t^{2D}) \cdot M(x_t, s_t)} \quad (1)$$

where $p(s_t)$ is the posterior probability of a Gaussian cluster according to the HMM formulation, $C(y_t, x_t^{2D})$ is the Chamfer distance between the input image silhouette y_t and the 2D projection of a pose hypothesis x_t^{2D} , and $M(x_t, s_t)$ is the Mahalanobis distance between the pose hypothesis x_t and the Gaussian cluster.

3 EXPERIMENTS

3.1 Datasets

Two sequences from the HumanEva dataset (Sigal, 2007) were used for testing and training our methodology:

1. Sequence of “walking in circle”. The data for the subject 2 was used for training and testing (S2W). This sequence consists of about two circular paths where one circle (i.e. frames from 431 to 870) has been chosen to train the walking activity and the remaining frames (i.e. frames from 1 to 430) were used for testing.

2. Sequence of “jogging in circle”. The data for the subject 2 was used for training and testing (S2J) too. This sequence consists of 5 circular paths where three circles (i.e. frames from 301 to 790) were used for training, while the two other circles (i.e. frames from 1 to 300) were used for testing.

3.2 Results

A PCA transformation is applied to each of the two training datasets (i.e. walking and jogging) to reduce their dimensionality from 32D to 8D and at the same time preserve more than 90% of the data variation (Fig. 1)

Modelling of activities was based on either a mixture of 4 Gaussians initialised according to the 4 phases of a gait cycle (Fig. 2) or a Dense-GMM initialised by the samples of a gait cycle (Fig. 3). Walking model consists of 84 clusters while jogging consists of 48 clusters.

Pose hypotheses were generated by either an exhaustive search constrained by either the GMM model of the activity or APF. Hypotheses were assessed against the image data, i.e. human silhouettes segmented by a foreground/background separation algorithm. The system solution for each frame is selected according to the Maximum Likelihood Estimation, where the likelihood is given by Eq.1.

Table 1 summarises the results of our experiments. They reveal that Dense-GMM based modelling is more accurate than the one based on a small number of Gaussian Clusters. Moreover, Dense-GMM is more computationally efficient since it requires about only 40% of the time that is required by the mixture of 4 Gaussians.

Fig 4 shows in the PCA space the pose estimation results against ground truth when hypotheses are generated by exhaustive search

constrained by the GMM model. Since the centres of Gaussians have high likelihood, estimated poses (blue) are concentrated on the Medial Axis of the search space.

Higher accuracy is achieved by APF (Fig. 5). In our implementation, we used 3 layers and 40 particles per layer. Higher number of particles may improve the accuracy marginally but would increase significantly the computational time. Annealed Particle filtering produced the best results for both walking and jogging activities with mean errors of 101 mm and 162 mm respectively.

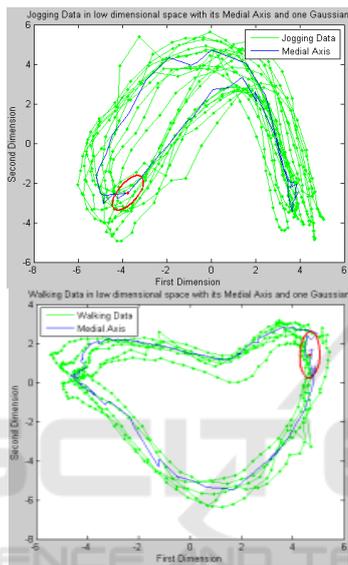


Figure 1: First three dimensions of the PCA space for the walking (left) and jogging (right) datasets.

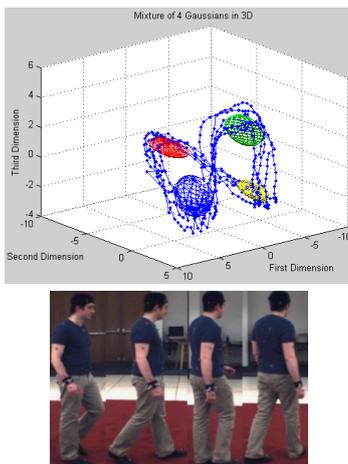


Figure 2: Modelling of walking poses (blue points) based on a mixture of 4 Gaussians (coloured ellipsoids) initialised by the 4 different phases of the walking cycle, i.e. left leg passing right leg, left leg in front, right leg passing left leg and right leg in front.

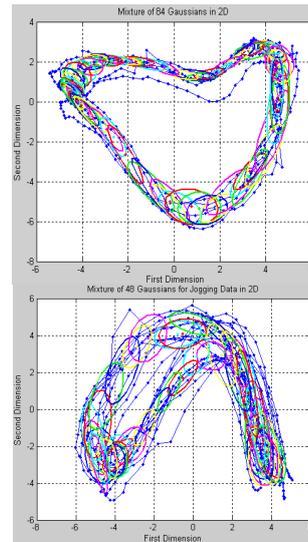


Figure 3: Modelling of walking (left) and jogging (right) poses using Dense GMMs initialised by samples of a gait cycle. Mixtures of 84 and 48 Gaussians are shown for walking and jogging data respectively.

Table 1: Summary of pose estimation results.

Activity	Modelling	Search	Mean Error mm	Std Error mm
Walking	4-GMM	Exhaustive	211	58
Walking	DenseGMM	Exhaustive	129	50
Walking	DenseGMM	APF	101	17
Jogging	DenseGMM	Exhaustive	168	40
Jogging	DenseGMM	APF	163	44

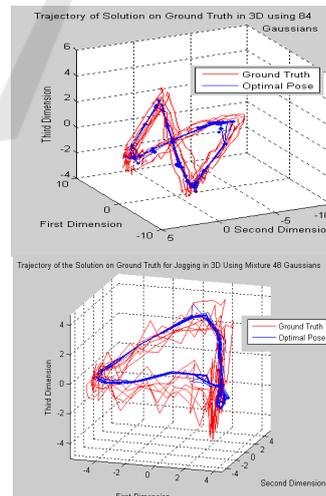


Figure 4: Estimated poses (blue) against ground truth (red) represented in the compressed PCA space for walking (left) and jogging (right) activity. Modelling was based on Dense GMM and hypotheses were generated by exhaustive search.

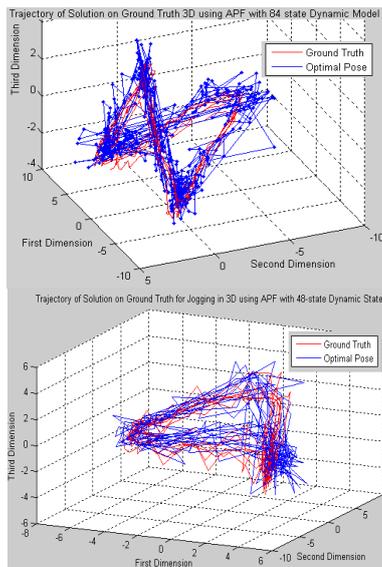


Figure 5: Estimated poses (blue) against ground truth (red), represented in the compressed PCA space for walking (left) and jogging (right) activity. Modelling was based on Dense GMM and hypotheses were generated by APF.

4 CONCLUSIONS

We presented a pose estimation method for monocular image sequences, where the human object is assumed to perform a known cyclic activity (e.g. walking, jogging). We demonstrated the value of using a Dense GMM initialised by a gait cycle of the activity as the base of a HMM-like dynamic model. Such modelling improves the accuracy and decreases the computational time of pose estimation compared to a mixture of few Gaussians.

Hypotheses were generated by APF and poses were estimated according to the Maximum Likelihood Estimation. APF improves accuracy when compared to an exhaustive search of the PCA space constrained by the GMM model.

Future work will focus on pose estimation in complex scenarios consisting of more than one activity.

REFERENCES

Balan, A. O., Sigal, L. and Black, M. J., 2005. A quantitative evaluation of video-based 3D person tracking, *IEEE workshop on VPSETS*, pp. 349-356.
 Bhatia, S., Sigal, L., Isard, M. and Black, M. J., 2004. 3D Human Limb Detection using Space Carving and Multi-view Eigen Models, *ANM workshop*, pp. 17.

Darby, J., Li, B. and Costen, N. 2008, Human Activity Tracking from Moving Camera Stereo Data, *BMVC*.
 Deutscher, J., Reid, I., 2005. Articulated body motion capture by stochastic search, *International Journal of Computer Vision*, 61, 2, 185-205.
 Hogg, D., 1983. Model-based vision: a program to see a walking person, *Image and Vision Computing*, 1, 1, 5-20.
 Howe, N. R., 2005. Silhouette lookup for automatic pose tracking, *IEEE workshop on Articulated and Nonrigid Motion*, pp. 3
 Howe, NR. Flow Lookup and Biological Motion Perception, *ICIP*, 3, 1168-1171.
 Kuo, P., Nebel, J-C., Makris, D., 2007. Camera Auto-Calibration from Articulated Motion, *AVSS*.
 Lakany, H. M., G.M. Haycs, M. Hazlewood, S.J. Hillman, 1999. Human walking: tracking and analysis, *Proceedings of the IEE Colloquium on Motion Analysis and Tracking*, pp. 5/1-5/14.
 Lee, C. S., Elgammal, A., 2007, Modelling view and posture manifolds for tracking, *ICCV*, 1-8.
 Roth, S., Sigal, L. and Black, M. Gibbs, 2004. Likelihoods for Bayesian tracking, *CVPR*, 1, 886-893.
 Sidenbladh, H., Black, M. and Fleet, D., 2000. Stochastic tracking of 3D human figures using 2D image motion, *Lecture Notes in Computer Vision*, 1843, 702-718.
 Sigal, L., Black, M. J., 2006. HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion, Technical Report CS-06-08.

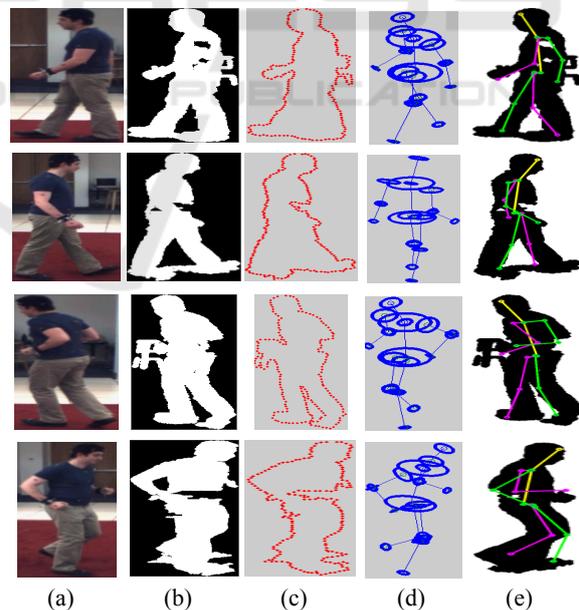


Figure 6: Indicative results for walking (first and second row) and jogging (third and fourth row) sequences using Dense GMM and APF. (a) input images, (b) silhouette extracted by foreground/background separation, (c) boundary of silhouette, (d) estimated pose, (e) estimated pose overlaid on image silhouette.