3D Tracking for Gait Characterization and Recognition

Raquel Urtasun and Pascal Fua*
Computer Vision Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland

Technical Report No: IC/2004/04

Abstract

We propose an approach to gait analysis that relies on fitting 3–D temporal motion models to synchronized video sequences. These models allow us not only to track but also to recover motion parameters that can be used to recognize people and characterize their style.

Because our method is robust to occlusions and insensitive to changes in direction of motion, our proposed approach has the potential to overcome some of the main limitations of current gait analysis methods. This is an important step towards taking biometrics out of the laboratory and into the real world.

1 Introduction

Most current gait analysis algorithms rely on appearance-based methods that do not explicitly take into account the 3–D nature of the motion. In this work, we propose an approach that relies on robust 3–D tracking and has the potential to overcome the limitations of appearance-based approaches, such as their sensitivity to occlusions and changes in the direction of motion.

In previous work [22], we have shown that using motion models based on Principal Component Analysis (PCA) and inspired by those proposed in [19, 20] lets us formulate the tracking problem as one of minimizing differentiable objective functions whose state variables are the PCA weights. Furthermore, the differential structure of these objective functions is rich enough to take advantage of standard *deterministic* optimization methods, whose computational requirements are much smaller than those of *probabilistic* ones and can nevertheless yield very good results even in difficult situations.

In this paper, we will argue that, in addition to allowing us to track the motion, these methods can also be used to recognize people and characterize their motion. More specifically, we first used an optical motion catpure system and a treadmill to build a database of walking motions for a few subjects. We then captured both the motion of these subjects and of other people by running our PCA-based

*This work was supported in part by the Swiss National Science Foundation.

tracker [22] on low-resolution stereo data. The resulting weights can then be used to recognize the people in the database and to characterize the motion of those who are not.

Because our tracking algorithm is robust to occlusions and insensitive to changes in direction of motion, our proposed approach has the potential to overcome some of the main limitations of current gait analysis methods. This is important if biometrics, defined as a measure taken from a living person and used as a method of identity verification or recognition [5], are to move out of the laboratory and into the real world.

In the remainder of this paper, we first discuss related work and introduce our PCA-based motion tracking algorithm. We then show how its output can be used to recognize and characterize different walking motions and conclude with some perspectives for future work.

2 Related Work

Current approaches to gait identification can be classified into two broad categories: Appearance-based ones that deal directly with image statistics and Model-based ones that first fit a model to the image data and then analyze the variation of its parameters.

Because model-fitting tends to be difficult where images are concerned, the majority of published approaches fall into the first category. Some rely on first processing each frame independently and then using PCA [16, 11] or HMM [10, 15] to model the transitions from one frame to the next. Other methods exploit the spatio-temporal statistics of the image stream. An early example of this approach can be found in the work by Niyogi [17]. More recently, methods that rely on dense optical flow [13] or self similarity plots computed via correlation of pairs of images [2, 6] have been proposed. The main drawback of these appearance-based approaches is that they are usually designed only for a specific viewpoint, usually fronto-parallel. Furthermore guaranteeing robustness against clothing and illumination changes remains difficult even though much effort has been expanded to this end, for example by processing silhouettes or binary masks instead of the image itself [2].

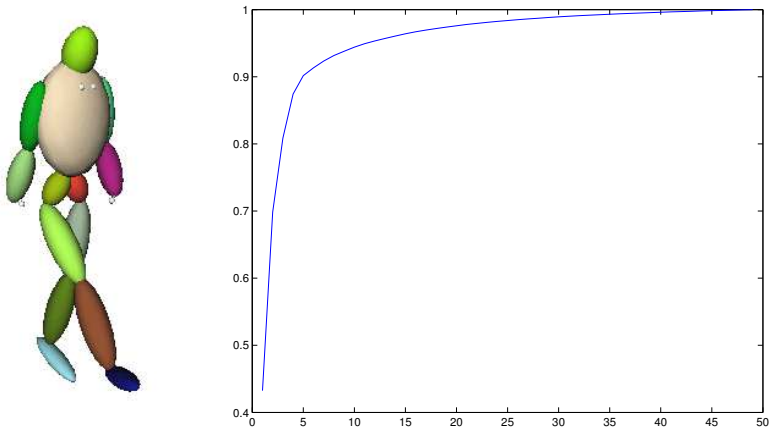
A good recent example of model-based gait recognition can be found in [5]. The gait signature is extracted by using Fourier series to describe the motion of the upper leg and by applying temporal evidence gathering techniques to extract the moving model from a sequence of images. However this technique is still 2–D, which means that a near fronto-parallel view is assumed. The approach we propose can be viewed as an extension of this philosophy to full 3–D modeling by replacing the Fourier analysis by the fitting of our PCA-based motion models.

3 Tracking

In this section we introduce our body and motion models and show how we can use them for tracking.

3.1 Models

In previous work [18] we have developed the body-model depicted by Figure 1(a) that is composed of implicit surfaces attached to an articulated skeleton. Each primitive defines a field function and the skin



(a) (b)

Figure 1: Models. (a) Volumetric primitives attached to an articulated skeleton. (b) Percentage of the database that can be generated with a given number of eigenvectors.

is taken to be a level set of the sum of these fields. Defining surfaces in this manner lets us define a distance function of data points to the model that is easy to evaluate and differentiable.

To build motion models, we have used a Vicon^{tm} optical motion capture system to capture 4 people, 2 men and 2 women walking at 9 different speeds ranging from 3 to 7 km/h by increments of 0.5km/h on a treadmill. The data was segmented into cycles and sampled at regular time intervals using quaternion spherical interpolation so that each example can be treated as $N = 33$ samples of a motion starting at normalized time 0 and ending at normalized time 1. An example is then represented by an *angular motion vector* Θ of dimension $N * NDofs$, where $NDofs = 84$ is the number of angular degrees of freedom in the body model. Θ is of the form

$$\Theta = [\theta_{\mu_1}, \ldots, \theta_{\mu_N}] \ , 0 \leq \mu_i < 1 \ , \tag{1}$$

where the θ_{μ_i} represent the joint angles at normalized time μ_i. The posture at a given time $0 \leq \mu_t \leq 1$ is estimated by interpolating the values of the θ_{μ_i} corresponding to postures immediately before and after μ_t.

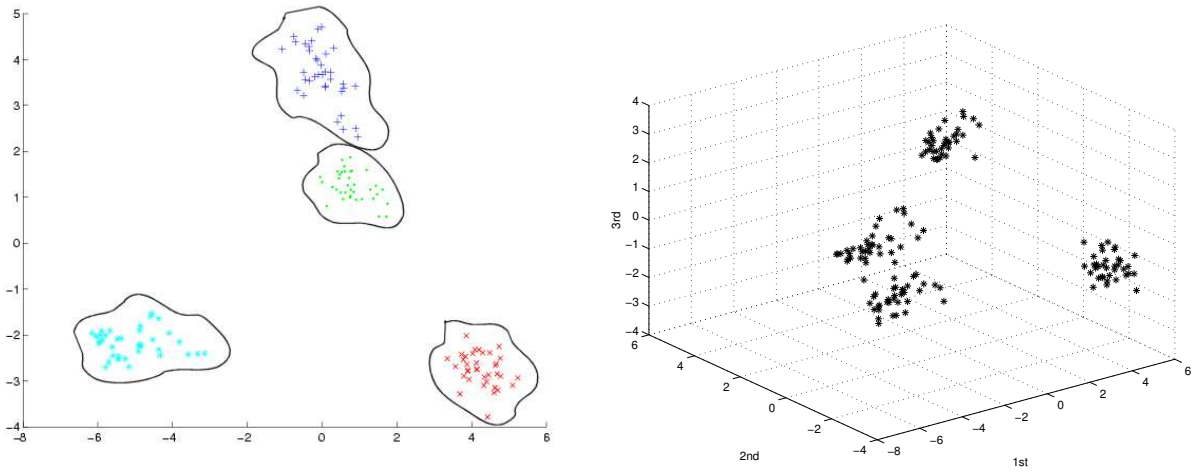
This process produces $M = 144$ angular motion vectors. We form their covariance matrix and compute its eigenvectors $\Theta_{1 \leq i \leq M}$ by Singular Value Decomposition. Assuming our set of examples to be representative, a Θ motion vector can be approximated as a weighted sum of the mean motion Θ_0 and the Θ_i:

$$\Theta \approx \Theta_0 + \sum_{i=1}^{m} \alpha_i \Theta_i \tag{2}$$

where the α_i are scalar coefficients that characterize the motion and $m \leq M$. m controls the percentage of the database that can be represented in this manner. This percentage is defined as

$$\sigma = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{M} \lambda_i} \tag{3}$$

where the λ_i are the eigenvalues corresponding to the Θ_i eigenvectors. It is depicted by Figure 1 (b) as a function of m. The posture at time μ_t is computed by interpolating the components of the Θ vector of Eq. 2 as discussed above.



(a) (b)

Figure 2: Clustering behavior of the first α_i coefficients of Eq. 2 for the motion vectors measured for 4 subjects walking at speeds ranging from 3 to 7km/h. (a) First two coefficients that form the 4 clusters, one for each subject, that we have outlined. (b) First three coefficients that also form relatively compact clusters in 3–D space that can be used for recognition purposes.

Figure 2 depicts the first three α_i coefficients of the original motion vectors when expressed in terms of the Θ_i eigenvectors. Note that the vectors corresponding to specific subjects tend to cluster. Figure 3 depicts the behavior of the fourth component which varies almost monotonically with walking speed. We will take advantage of these facts for recognition and characterization purposes in Section 4.

3.2 Deterministic tracking

Most recent tracking techniques rely on probabilistic methods to increase robustness [12, 8, 7, 4, 21]. While effective, such approaches require large amount of computations. In previous work [22], we developed an approach that relies on the motion models of Section 3.1 to formulate the tracking problem as the one of minimizing differentiable objective functions. The structure of these objective functions is rich enough to take advantage of standard deterministic optimization methods and, thus, reduce the computational costs.

Given a T-frame video sequence in which the motion of a subject remains relatively steady such as those of Figures 4 and 5, the entire motion can be completely described by the angular motion vector of Eq. 2 and, for each frame, a six-dimensional vector G_t that defines the position and orientation of the root body model node with respect to an absolute referential. We therefore take the state vector ϕ to be

$$\phi = [G_1, \ldots, G_T, \mu_1, \ldots, \mu_T, \alpha_1, \ldots, \alpha_m] \ , \tag{4}$$

where the μ_t are the normalized times assigned to each frame and which must also be treated as optimization variables.

In the sequences of Figures 4 and 5, the images we show were acquired using one of three synchronized cameras and used to compute clouds of 3–D points via correlation-based stereo. The ϕ state

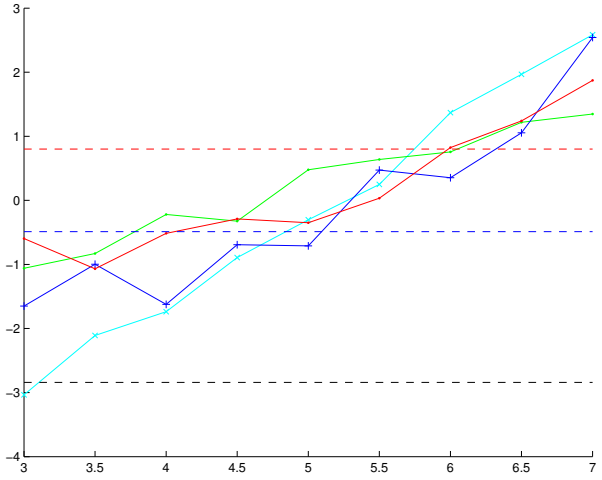


Figure 3: Fourth α_i coefficients of the motion vectors as a function of speed. They evolve almost monotonically with speed for each subject, which will allow us to evaluate the speed of a walking subject. The dashed lines represent the values obtained with the tracking. The upper two lines corresponds to the two women of Figure 4, while the lower one correspondes to the male subject of Figure 5.

vector, and thus the motion, were recovered by minimizing in the least-squares sense the distance of the body model to those clouds in *all* frames simultaneously. The α_i recovered by the system define the style of the motion, and this information will be used in the following section for characterization and recognition purposes.

In [22], we also showed that, if the motion is not steady, using a single set of α_i parameters for the whole sequence over constrains the fitting and that allowing these coefficients to vary produces more accurate results. However, for recognition purposes, we want to recover global motion characteristics and, therefore, we only use a single set of coefficients.

4 Characterization and Recognition

The motion style is encoded by the α_i coefficients of Eq. 2, since they measure the deviation from the average motion along orthogonal directions. The tracker can thus be used for motion characterization and recognition since these coefficients are its output.

Our experiments show that, for characterization purposes, using the first six coefficients that represent 90% of the database, as shown in Figure 1(b), appears to be optimal. Using more coefficients results in overfitting while using fewer yields a system that is not flexible enough to produce good tracking results and tends to change the value of the first components to compensate for missing ones, which results in a poor classification.

We use the three examples of Figures 4 and 5 to highlight our system's behavior. The sequences were captured using a 640x480 Digiclops^{tm} triplet of cameras operating at 14Hz. Because we had to ensure that the subjects would be visible in the whole sequence without moving the Digiclops^{tm}, they only occupy relatively small parts of the images, which results in very low resolution stereo data. Figure 4 depicts the tracking results for the two women whose motion has been recorded in the database.

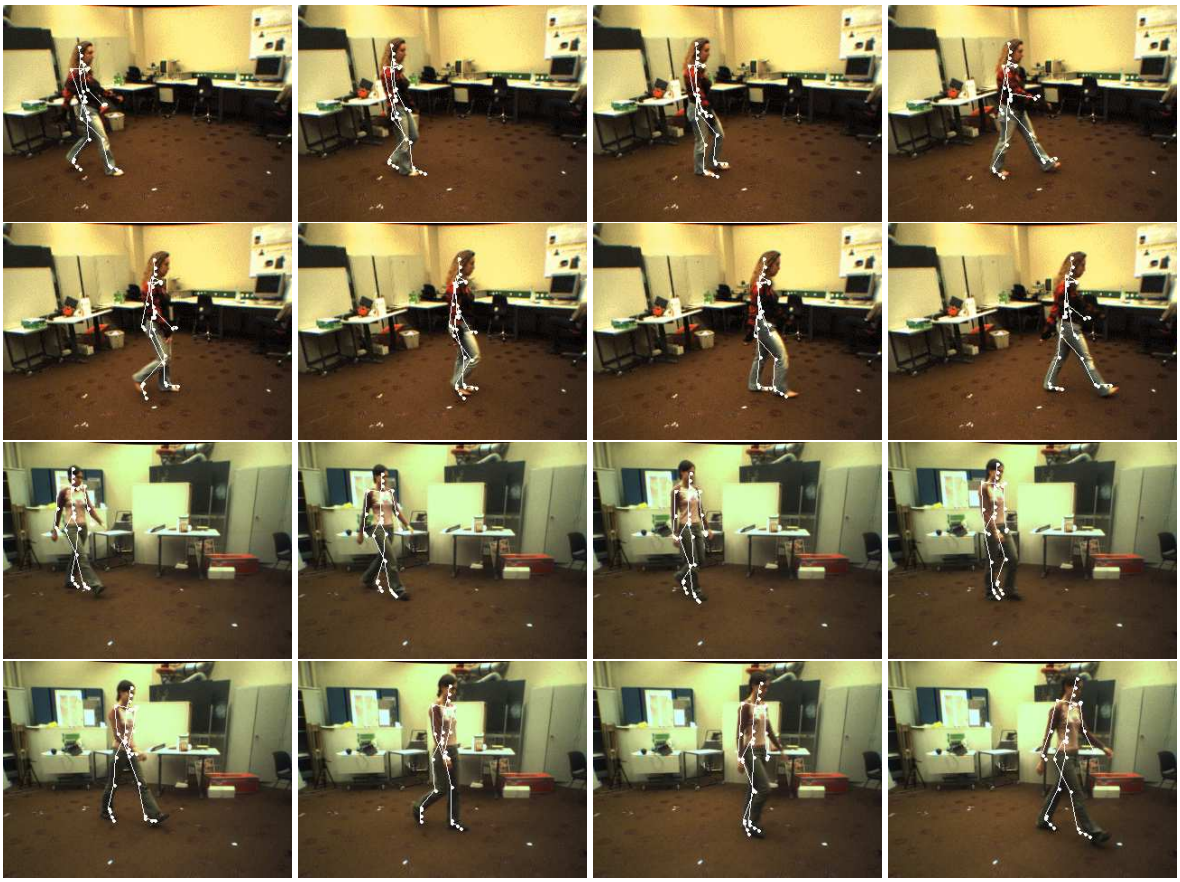


Figure 4: Using low resolution stereo data to track the two women whose motion was recorded in the database. The recovered skeleton poses are overlaid in white.

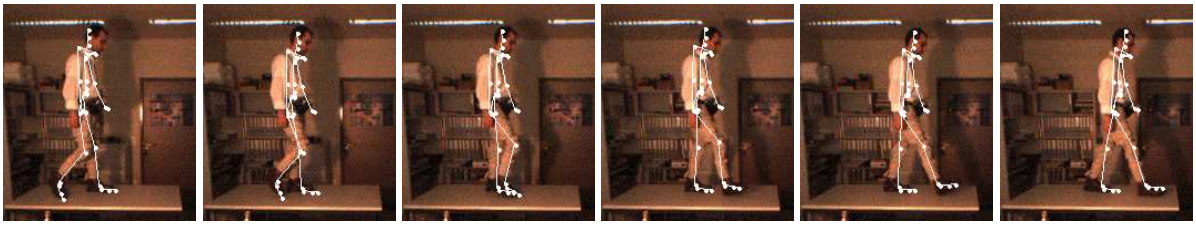


Figure 5: Tracking the walking motion of a man whose motion was *not* recorded in the database.

The motion of their legs is accurately captured. The arms, however, have not been tracked as precisely for two main reasons: Arm motion on a treadmill is quite different from what can be observed when walking naturally and the errors in the noisy cloud of 3D points we use for fitting are sometimes bigger than the distance from the arms to the torso. In [22] we have shown that these results can be improved by dropping the steady motion assumption and allowing the α_i to vary. These errors, however, do not appear to affect recognition as will be shown below.

Figure 5 depicts our tracking results for a man whose motion was *not* recorded in the database. This example shows the ability of our system to handle motions that are not originally part of the database.

4.1 Recognition

Recall that, as shown in Figure 2(b), the first three α_i coefficients of the motion vectors in the database tend to form separate clusters for each subject. Our approach handles the motion sequences as a whole as opposed to a set of individual poses, thus making the clusters compact enough for direct classification. A given pose can be common to more than one subject, but not the whole movement.

As the clusters are sufficiently far appart, a simple k-means algorithm can be used to compute them and yield the result depicted by the outlines of Figure 2(b). We take the distance measure between two sets of coefficients α^1 and α^2 to be

$$d(\alpha^1, \alpha^2) = \sum_{i=1}^{m} \lambda_i |\alpha_i^1 - \alpha_i^2| \ , \tag{5}$$

where the λ_i are the eigenvalues corresponding to the eigenvectors of Eq. 1. This distance logically gives more weight to the first coefficients, which are the most descriptive ones as shown in Figure 1(b). In fact, it can be easily shown to be equivalent to a Mahalanobis distance

$$d(\alpha^1, \alpha^2) = [(\alpha^1 - \alpha^2)^T \Sigma^{-1} (\alpha^1 - \alpha^2)]^{1/2} \tag{6}$$

where Σ is the covariance matrix of the database motion vectors.

Figure 6 depicts the first two α_i coefficients computed for each one of the three examples of Figures 4 and 5 as circles drawn on the plot of Figure 2(b). For both women, the first two recovered coefficients fall in the center of the cluster formed by their recorded motion vectors. Higher order coefficients exhibit small variations that can be ascribed to the fact that walking on a treadmill changes the style. For the man whose motion was not recorded in the database, the recovered coefficients fall within the cluster that corresponds to the men whose motion was recorded and whose weight and height are closest. The third coefficient, however, can be used to differentiate the two men.

These experimental observations lead us to formulate the following hypothesis that will have to be validated using a much larger database: The first two coefficients encode general characteristics such as weight, height, gender, or age and the third one can be used to distinguish among people who share these characteristics.

4.2 Characterization

For the man of Figure 5, the recorded motion that is closest in terms of the distance of Eq. 5 to the one our system recovers corresponds to a slow 3km/h walk, which is correct in this case. This makes sense since, as shown in Figure 3, the fourth α_i coefficient encodes speed. On this figure, we also represent the respective speeds of the two women of Figure 4 as horizontal dashed lines. The system yields a higher speed for the woman shown in the lower half of the figure than for the other one, which can be simply confirmed by counting the number of frames it takes each one of them to cross the room.

In previous work [22] depicted by Figure 7, we showed that our tracking approach can also handle running motions and transitions from walking to running. This was done by replacing the walking database of Figure 4 by one that encompasses both activities and is depicted by Figure 8. In this more complete database, the first coefficients of vectors corresponding to walking motions are more compressed but still form separate clusters. Similarly, the vectors corresponding to running still cluster by

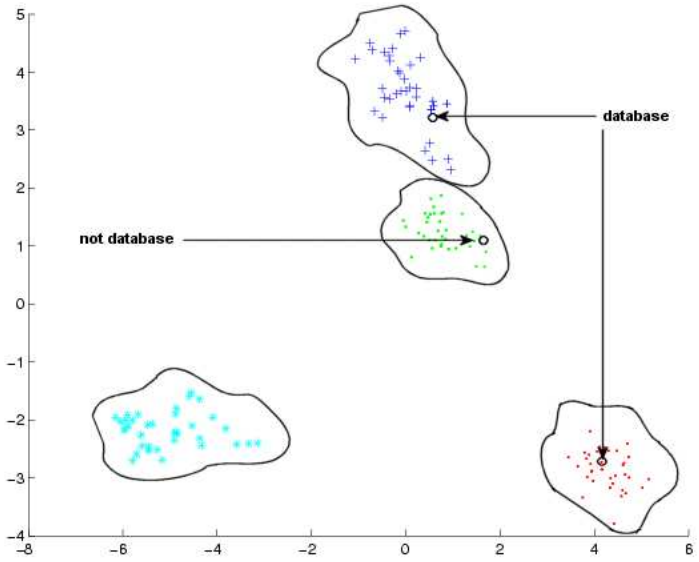


Figure 6: Recognition from stereo data. For both women of Figure 4, the first two α_i coefficients recovered by the tracking are represented by a black circle that falls in the center of the clusters formed by their recorded motion vectors, thus allowing recognition. For the man of Figure 5, the first two coefficients depicted by a black square fall into the cluster corresponding to one of the men whose motion was recorded and whose stature and weight are closest to his.

Figure 7: Tracking a running motion.

identity even though these clusters are much more elongated, reflecting the fact that there is more variation in running style and that the database should be expanded. However, this gives us a hope that the methodology we propose should extend naturally to identity *and* activity recognition.

5 Conclusion

We have presented an approach to motion characterization and recognition that is robust to view-point, clothing and illumination changes as well as to occlusions because it relies on extracting style parameters of the 3–D motion over a whole sequence and using them to perform the classification. Using low-quality stereo data, we have demonstrated that these parameters can indeed be recovered and used. Of course, using higher quality data could only improve the results.

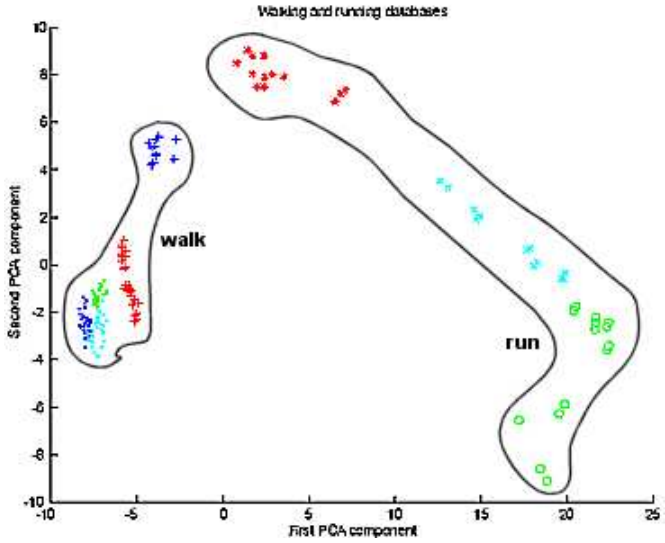


Figure 8: First two α_i components for a mixed database composed of walking and running motion. The coefficients tend to cluster by subject and activity.

Currently, the major limitation comes from the small size of the database we use, which we will endeavor to complete. This may result in clusters corresponding to different people or styles in PCA space becoming more difficult to separate. If such is the case, we plan to investigate the use of appropriate classification techniques as EM, SVM or novel techniques [1, 3, 9, 14] to segment those clusters and allow us to handle potentially many classes that can exhibit huge variability.

References

[1] Y. Amit, D. Geman, and K. Wilder. Joint induction of shape features and tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:1300–1306, 1997.

[2] C. BenAbdelkader, R. Cutler, and L. Davis. Motion-Based Recognition of People in EigenGait Space. In *Automated Face and Gesture Recognition*, pages 267–272, may 2002.

[3] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[4] K. Choo and D.J. Fleet. People tracking using hybrid monte carlo filtering. In *International Conference on Computer Vision*, Vancouver, Canada, July 2001.

[5] D. Cunado, M.S. Nixon, and J.N. Carter. Automatic Extraction and Description of Human Gait Models for Recognition Purposes. *Computer Vision and Image Understanding*, 90(1):1–41, April 2003.

[6] R. Cutler and L. Davis. Robust real-time periodic motion detection, analysis and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(13), 2000.

[7] A. J. Davison, J. Deutscher, and I. D. Reid. Markerless motion capture of complex full-body movement for character animation. In *Eurographics Workshop on Computer Animation and Simulation*. Springer-Verlag LNCS, 2001.

[8] J. Deutscher, A. Blake, and I. Reid. Articulated Body Motion Capture by Annealed Particle Filtering. In *CVPR*, Hilton Head Island, SC, 2000.

[9] Y. Freund and R.E. Schapire. Experiments with a New Boosting Algorithm. In *International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann, 1996.

[10] Q. He and C. Debrunner. Individual recognition from periodic activity using Hidden Markov Models. In *IEEE Workshop on Human Motion*, Austin, Texas, December 2000.

[11] P.S. Huang, C.J. Harris, and M.S. Nixon. Comparing Different Template Features for Recognizing People by their Gait. In *British Machine Vision Conference*, volume 2, pages 639–648, Southampton, UK, September 1998.

[12] M. Isard. and A. Blake. CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, August 1998.

[13] J.J. Little and J.E. Boyd. Recognizing People by Their Gait: The Shape of Motion. *Videre*, 1(2):1–32, 1986.

[14] L. Mason, J. Baxter, P. Bartlett, and M. Frean. Boosting algorithms as gradient descent. In *Neural Information Processing Systems*, pages 512–518. MIT Press, 2000.

[15] D. Meyer, J. Pösl, and H. Niemann. Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture densities. In *British Machine Vision Conference*, pages 459–468, Southampton, UK, 1998.

[16] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[17] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in XYZ. In *Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[18] R. Plänkers and P. Fua. Articulated Soft Objects for Multi-View Shape and Motion Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003.

[19] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3D human figures using 2D image motion. In *European Conference on Computer Vision*, June 2000.

[20] H. Sidenbladh, M. J. Black, and L. Sigal. Implicit Probabilistic Models of Human Motion for Synthesis and Tracking. In *European Conference on Computer Vision*, Copenhagen, Denmark, May 2002.

[21] C. Sminchisescu and B. Triggs. Kinematic Jump Processes for Monocular 3D Human Tracking. In *Conference on Computer Vision and Pattern Recognition*, Madison, WI, June 2003.

[22] R. Urtasun and P. Fua. 3d human body tracking using deterministic temporal motion models. Technical Report IC/2004/03, EPFL, January 2004.