massachusetts institute of technology — artificial intelligence laboratory

Gait Dynamics for Recognition and Classification

Lily Lee

AI Memo 2001-019 September 2001

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

This paper describes a representation of the dynamics of human walking action for the purpose of person identification and classification by gait appearance. Our gait representation is based on simple features such as moments extracted from video silhouettes of human walking motion. We claim that our gait dynamics representation is rich enough for the task of recognition and classification. The use of our feature representation is demonstrated in the task of person recognition from video sequences of orthogonal views of people walking. We demonstrate the accuracy of recognition on gait video sequences collected over different days and times, and under varying lighting environments. In addition, preliminary results are shown on gender classification using our gait dynamics features. [1]

[1] Research Support: This research is supported in part by DARPA under contract N00014-00-1-0907.

1 Introduction

Human gait is an identifying feature of a person that is determined by his/her weight, limb length, and habitual posture. Hence, gait can be used as a biometric measure to recognize known persons and classify unknown subjects. Gait can be detected and measured at low resolution, and therefore it can be used in situations where face or iris information is not available in high enough resolution for recognition. We have designed a representation for the dynamics of human gait that facilitates the recognition and classification of people by their gait.

Gait is defined as "a manner of walking" in the Webster Collegiate Dictionary. We extend our definition of gait to include both the appearance and the dynamics of human walking motion. In other words, we consider the appearance of the person, the aspect ratio of the torso, the clothing, the amount of arm swing, and the period and phase of a walking cycle, etc., all as part of one's gait. Johansson [7] had shown in the 1970's that observers could recognize walking subjects familiar to them by just watching video sequences of lights affixed to joints of the walker. Hence, in theory, joint angles are sufficient for recognition of people by their gait. However, recovering joint angles from a video of walking person is an unsolved problem. In addition, using only joint angles ignores the appearance traits that are associated with individuals, such as heavy-set vs. slim, long hair vs. bald, and particular objects that one always wears. For these reasons, we have included appearance as part of our gait recognition features.

The challenges involved in gait recognition include imperfect foreground segmentation of the walking subject from background scene, change in clothing of the subject, and variations in the camera viewing angle with respect to the walking subjects. The gait dynamics features discussed below will tolerate some imperfection in segmentation and clothing changes, but not drastic style changes. The view-dependent nature of our gait dynamics feature representation is removed in a separate paper [13].

In the following sections, we summarize some of the previous work on the perception of gait and computational methods to measure gait, introduce our model-based gait dynamics representation, apply our representation to recognition and classification tasks, and suggest future work to improve the representation.

2 Previous Work

Several researchers have concluded that gait is indicative of a person's gender and identity. Johansson[7] used lights affixed to joints of a human body and showed observers motion sequences of the lights as a person walks. The observers were able to identify gender and, in cases where the observer was familiar with the walking subject, the identity of the walker. Cutting, *et. al.*[5] studied human perception of gait using moving light displays similar to that used by Johansson and showed human person identification results [4] and gender classification results[8]. They showed that human observers could identify gender with approximately 70% accuracy using only the visual cues from moving light displays.

Given the ability of humans to identify persons and classify gender by the gait of a walking subject, there have been a few computer vision algorithms developed

for people identification and activity classification. Cutler and Davis [3] used self-correlation of moving foreground objects to distinguish walking humans from other moving objects such as cars. Polana and Nelson[12] detected periodicity in optical flow and used these to recognize activities such as frogs jumping and human walking. Bobick[1] used a time delayed motion template to classify activities. Little and Boyd[9] used moment features and periodicity of foreground silhouettes and optical flow to identify walkers. Nixon, *et. al.*[10] used principal component analysis of images of a walking person to identify the walker by gait. Shutler, *et. al.* [14] used higher order moments summed over successive images of a walking sequence as features in the task of identifying persons by their gait.

The work described in this paper is closely related to that of Little and Boyd [9]. However, instead of using moment descriptions and periodicity of the entire silhouette and optical flow of a walker, we divide the silhouettes into regions and compute statistics on these regions. We also further study the capacity of our features in tasks beyond person identification, such as gender classification.

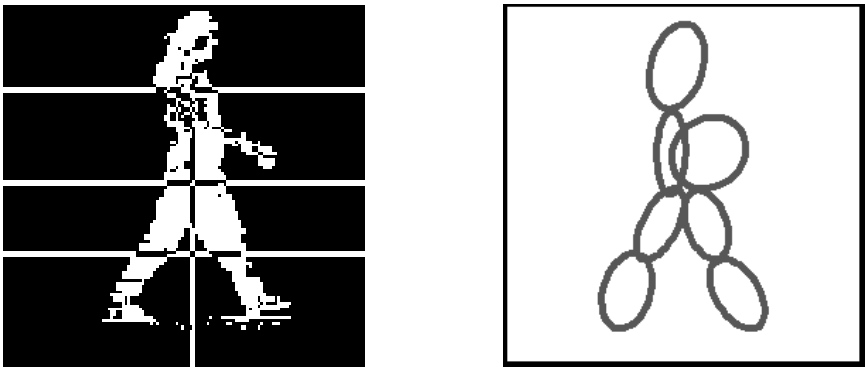
3 Gait dynamics representation

The human gait perception studies by Johansson[7], and Cutting *et. al* [4][8] used moving light display, where the only visual cues that observers received were lights affixed to joints of walking humans. These studies give indications that humans can use the joint angles to recognize action, identify persons, and classify gender. It would be interesting to build a human gait recognition algorithm based entirely on joint angles of a dynamic sequence. However, computing joint angles from video sequence is still a difficult problem, though several attempts have been made on it [2][15][6]. Solely using joint angles also ignores the appearance of a person, which may give strong indications of identity and/or gender of the person. For example, the presence of a pony tail is a good indication of the gender of the subject. Hence, we advocate using the whole appearance of a walker rather than just joint angles. An appearance based gait representation is constrained to be view–dependent. In a separate paper[13] we show that this constraint can be be removed by synthesizing a walking sequence in a canonical view using the visual hull constructed from multiple cameras.

We consider the canonical view of a walking person to be that which is perpendicular to the direction of walk. We also assume that the silhouette of the walker is segmented from the background using a background subtraction algorithm such as [16]. To make our gait representation insensitive to changes of clothing and distance between the camera and the walking subject, we ignore the color of the foreground walker and use only the scale-normalized binary silhouette.

We would like our gait dynamics feature vector to have the following desirable properties: ability to describe appearance at a level finer than whole body description without having to segment individual limbs; robustness to noise in video foreground segmentation; and simplicity of representation.

Our gait dynamics feature vector consists of parameters of smoothed versions of moment features in image regions containing the walking person. For each silhouette of a gait video sequence, we find the centroid and divide the silhouette into 7 parts roughly



(a) (b)

Figure 1: The silhouette of a foreground walking person is divided into 7 regions, and ellipses are fitted to each region.

corresponding to head/shoulder, arms/torso (front and back), thighs (front/back), and calves/feet (front/back) (see Figure 1(a)). These regions are by no means meant to segment the arms and legs precisely. We are interested in a consistent method to divide the silhouette of a walking person into regions that will facilitate the person recognition task.

For each of the 7 regions from a silhouette, we fit an ellipse to the portion of foreground object visible in that region (Figure 1(b)). The features we extract from each of these regions are the centroid, aspect ratio of major and minor axis of the ellipse, and the orientation of major axis of the ellipse. These moment based features are somewhat robust to noise in the silhouettes obtained from background subtraction, as long as the number of noise pixel is small and not consistent over time. The features extracted from each frame of a walking sequences consists of features from each of the 7 regions.

Given the region features across a gait sequence, one could use Fourier transform to recover the periodicity of any of the parameters in a region across time and use that as part of a gait recognition feature. In reality, we found that frequency was not a good indicator of person identity because our video frame capture rate (15 frames/sec) did not allow for high enough resolution measure of the period. In addition, the error in foreground segmentation often makes peak detection difficult after Fourier transforming the time signals. When the peaks can be detected, the inter-person period difference is up to 1 frame, and the intra-person period difference is up to 2 frame for some pairs of people, but within one frame for the majority of pairs of subjects. What is discriminating in person idenfication are factors such as the amount of arm and leg swing, the average orientation of the head (held high vs. looking at the floor) (see Figure 2). These characteristics can be just as easily represented by the mean and standard deviations of region features across time as they are represented by Fourier transforms

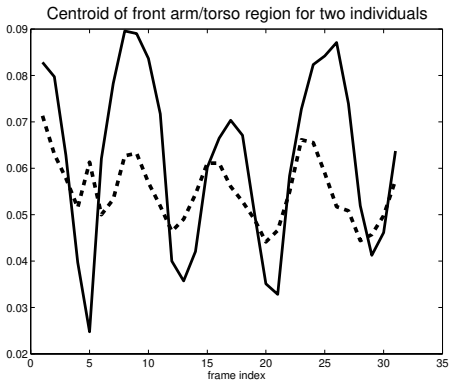


Figure 2: The difference in the amount of arm swing between two individuals.

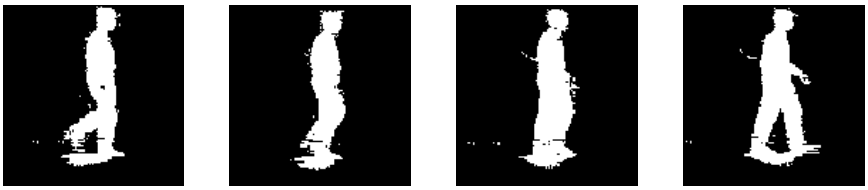
of the same signals. However, the mean and standard deviation representation has the additional advantage that they are not seriously affected by the occasional frame skip that a foreground segmentation algorithm may undergo. Hence, the feature vector we compute for each gait sequence consists of the mean and standard deviations across time of each of the parameters for each region. It may be that for different regions of the silhouette the mean (or the standard deviation) may be better suited for recognition than the standard deviation (or the mean). We are conducting further study to refine the representation for each region.

4 Experimental results

We apply our gait dynamics features to the following tasks: identification of persons and gender classification. We gathered gait data in indoor environments with different backgrounds, on four separate days spanning two months. Twenty-five subjects, 11 women and 14 men, were asked to walk at their normal speed and stride, back and forth, twice in front of a video camera that was placed perpendicular to their walking path. The walking gait sequences going the opposite direction are flipped so that all sequences are of one walking direction. In all, 210 walking sequences were collected, between 4 to 22 sequences for each subject, averaging 8.4 sequences per subject. A minimum of 3 complete walking cycles were captured at 15 frames per second for each walking sequence. We then use an adaptive background subtraction algorithm [16] to segment out the walking person as a moving foreground object and scale-normalize it to fit in a 128x128 pixel binary image. An example of the foreground walking person is shown in Figure 3. The foreground segmentation is not perfect–the shadow on the ground and some background are included. However, our gait representation tolerates this amount of noise in foreground segmentation.

Using the gait dynamics features that we described in the previous section, a feature vector is computed for each walking sequence and used in person identification and

(a) t=1 (b) t=2 (c) t=3 (d) t=4



(e) t=5 (f) t=6 (g) t=7 (h) t=8

Figure 3: A sample sequence of the silhouette of a walking subject after background subtraction.

gender classification.

4.1 Person Identification

While the gait dynamics features that have been described are not completely independent of each other, we use a simplifying assumption that they are independent, which then allows us to compute distances between feature vectors in Euclidean space. Assuming that all components in the feature vector are equally important to gait recognition, using the Mahalanobis distance between feature vectors as a measure of the similarity between two sequences will remove the dominance of parameters with large variance.

We evaluated our gait recognition performance using the cumulative match score described by Phillips [11]. Under the closed-universe assumption, all query sequences contain a walker who has at least one other walking sequence in our collection of known walking sequences with known person ids. We answer the question, "Is the correct person id in the top n matches?" The cumulative match score has the ranked top match on the x-axis and the percentage of correct identification on the y-axis.

Two gait recognition tests were performed. In one, a query walking sequence is compared against all sequences, regardless of the day of recording. The second test compares walking sequences taken on one day against sequences taken on other days only. This is a much more stringent test because sequences taken on different days

have different lighting conditions, different backgrounds, different clothing.

The cumulative match score for recognition test one, where each sequence is compared to all other sequences regardless of date of recording, is very high: $>96\%$ correct identification at the top match, and 100% correct idenfication at the 6th top match (see Figure 4). Closer examination of the gait dynamics feature vector distances shows that the top correct matches are almost always those sequences of the same person taken on the same day.

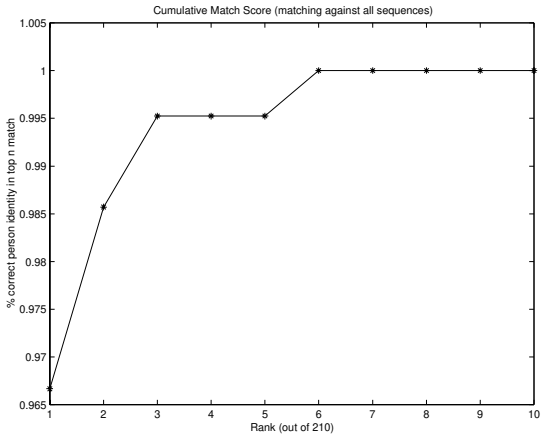
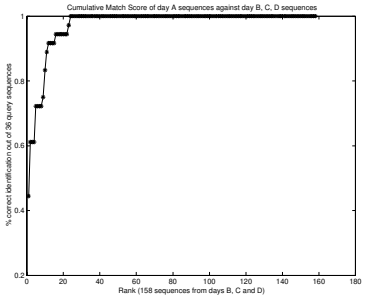


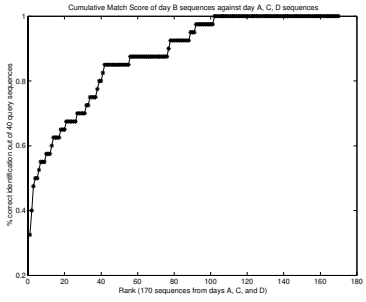
Figure 4: Cumulative match score of top 10 matches for comparison of each gait sequence against all other walking sequences (209 sequences).

The results of the second person identification test, where gait sequences recorded on one day are compared against only gait sequences taken on other days, are shown in Figure 5. Each of these cumulative match score curves represents the comparison between walking sequences taken on one day (days A, B, C, or D) against sequences taken on all other days ((B,C,D), (A,C,D), (A,B,D), or (A,B,C), respectively). Only walking subjects who have gait data collected from different days are used in this recognition test. The gait recognition performance is higher when comparing data collected on day A against all other sequences than matching sequences between those recorded on other days against the rest of the gait data. Closer examinations of the ranked matches show that the better performance of day A sequences against day B, C, and D sequences is the result of lack of similarity in the type of clothing that some of the subjects wore for the day B, C, D gait sequence data. For example, the 7 query sequences with the worst match scores from day B are all from one subject who wore baggy pants and whose only other sequences were collected on day D showing him wearing shorts. For the same reason, recognition results of matching day D gait sequences against all other sequences suffer because of lack of a similar appearance model in day B for the same subject.

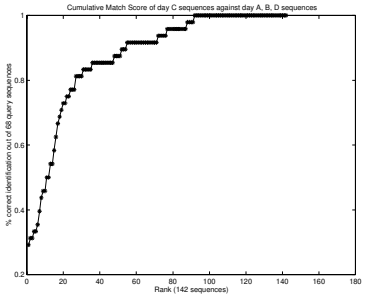
It is clear that the gait dynamics feature representation is sensitive to drastic clothing style changes. For subjects that do not change their clothing style very much,



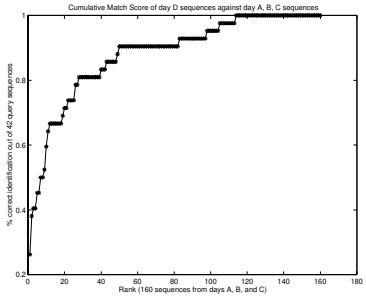
(a) 36 sequences from day A against 158 sequences from other days



(b) 40 sequences from day B against 170 sequences from other days



(c) 48 sequences from day C against 142 sequences from other days



(d) 42 sequences from day C against 160 sequences from other days

Figure 5: Cumulative match score for matching sequences taken on one day against sequences taken on all other days.

their correct identities are retrieved very quickly based on the ranked feature distance. Figure 6 shows one example of a subject with her walking sequences taken on three different days under three different hair styles and two different clothing styles (pants vs. skirt). These drastic style changes need to be represented as multiple gait appearance models for each person. Because the number of clothing styles are very limited after they are binarized (for example, pants vs. skirt vs. shorts), there will only be a small set of gait dynamics/appearance model for each person.

4.2 Gender classification

We trained a linear discriminant classifier to distinguish the gender of walking subjects in gait video sequences. Two types of training sets are used. For one type of

(a) hair in French twist, wearing pants

(b) hair down, wearing a dress

(c) hair in pony tail, wearing pants

Figure 6: Silhouette of one walking subject from video sequences taken on three different days, with three different hair styles and two different types of clothing.

training set, we randomly selected gait dynamics feature vectors of approximately half of the sequences, without regard to the identity of the walking subject, and tested on the gait dynamics features of the remaining sequences. The linear discriminant classifiers trained on these data sets have very high classification performances: all have less than 10% error rate, with a mean error rate of approximately 6% over multiple trials at training a classifier. For the second type of training set, we randomly selected approximately half of our walking subjects, trained the linear discriminant classifier on all sequences from these walkers, and tested on all sequences of the remaining walking subjects. The classification results for this type of training set on the linear discriminant classifier are slightly worse. The classification error rate ranges from 12% to 27%, with a mean of 17.5%. We attribute this decline in performance to the sparsity of training data. There are only between 4 to 6 women and 5 to 7 men for each training set, which is not sufficient to capture the full range of gait appearances of each gender.

5 Discussion and Future Work

Our gait dynamics features are view and appearance based. The view dependent nature of this representation can be resolved by synthesizing a view of the walking in the canonical view used in this paper. The dependence of out gait dynamics feature on the appearance of a walking is difficult to remove, unless the joint movements of a walker can be accurately detected. However, as we argued earlier, the appearance of a walking subject actually contains information about the identity of the person. To allow accurate recognition of a person with drastic clothing style change then requires having multiple representations of the person. We propose that a multi-modal recognition algorithm is useful for building such a multiple-appearances gait dynamics model. Recognition modalities such as face recognition can be very useful here because face is

less susceptible to appearance change. Recognition results from face can be combined with gait appearance recognition to establish multiple–appearance gait model.

6 Summary and Conclusions

We have presented a very simple representation of human gait dynamics based on moments computed from the silhouette of the walking person for the purpose of person identification and gender classification. Our representation is rich enough to show promising results in these tasks. This view and appearance dependent model of gait dynamics can be further extended to accommodate a multiple appearance model of a person with the use of other recognition modalities.

References

[1] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3), March 2001.

[2] Chris Bregler and Jitendra Malik. Tracking people with twists and exponential maps. In *Proc. of IEEE CVPR*, 1998.

[3] Ross Cutler and Larry Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781 –796, 2000.

[4] J.E. Cutting and L.T. Kozlowski. Recognizing friends by their walk: gait perception without familiarity cues. *Bull. Psychonometric Soc.*, (9):353–356, 1977.

[5] JE Cutting, DR Proffitt, and LT. Kozlowski. A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3):357–372, 1978.

[6] D. DiFranco, T.J. Cham, and J. Rehg. Recovery of 3d articulated motion from 2d correspondences. Technical report, Compaq Cambridge Research Lab, 1999.

[7] G Johansson. Visual motion perception. *Scientific American*, (232):76–88, 1975.

[8] L.T. Kozlowski and J.E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21(6):575–580, 1977.

[9] James Little and Jeffrey Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2), 1998.

[10] M.S. Nixon, J.N. Carter, J.M. Nash, P.S. Huang, D. Cunado, and S.V. Stevenage. Automatic gait recognition. In *Motion Analysis and Tracking (Ref. No. 1999/103), IEE Colloquium on*, pages 3/1–3/6, 1999.

[11] P.J. Phillips, H. Moon, P. Rauss, and S.A. Rizvi. The feret evaluation methodology for face-recognition algorithms. In *CVPR*, pages 137–143, 1997.

[12] Ramprasad Polana and Randall Nelson. Detecting activities. In *CVPR*, 1993.

[13] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *CVPR*, 2001.

[14] J.D. Shutler, M.S. Nixon, and C.J. Harris. Statistical gait recognition via velocity moments. In *Visual Biometrics, IEEE*, pages 10/1–10/5, 2000.

[15] Hedvig Sidenbladh, Michael Black, and David Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *European Conference on Computer Vision (ECCV)*, volume 2, pages 702–718, 2000.

[16] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, 1999.