

Gait Appearance for Recognition

L. Lee and W. E. L. Grimson

MIT AI Laboratory, 200 Technology Square, Cambridge, MA 02139.
{lee, welg}@ai.mit.edu

Abstract. This paper describes a set of representations of gait appearance features for the purpose of person identification. Our gait representation has two stages: the first stage computes a set of image features that are based on moments extracted from orthogonal view video silhouettes of human walking motion; the second stage applies three methods of aggregating these image features over time to create the gait sequence features. Despite their simplicity, the resulting gait sequence feature vectors contain enough information to perform well on human identification. We demonstrate the accuracy of recognition using gait video sequences collected over different days and times, under varying lighting environments and explore the differences in the three time-aggregation methods for the purpose of recognition.

1 Introduction

Gait is defined as “a manner of walking” in the Webster’s New Collegiate Dictionary. However, human gait is more than that: it is an idiosyncratic feature of a person that is determined by, among other things, an individual’s weight, limb length, footwear, and posture combined with characteristic motion. Hence, gait can be used as a biometric measure to recognize known persons and classify unknown subjects by gender. Moreover, we extend our definition of gait to include 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. We have designed mathematical representations for the overall appearance of human gait that facilitate the recognition of people by their gait. We described in [LG02] a similar set of representations for person recognition and gender classification. In this paper we explore the different methods of aggregating image features over time and their effects on the accuracy of person recognition.

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. Johansson [Joh75] has shown that joint angles are sufficient for recognition of people by their gait. However, reliably 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, or personal accoutrements. 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 the background scene, changes in clothing of the subject, variations in the camera viewing angle with respect to the walking subjects, and changes in gait as a result of mood or speed change, or as a result of carrying objects. The gait appearance features discussed below will tolerate some imperfection in segmentation and clothing changes, but not drastic style changes such as pants vs. skirts, nor are they impervious to changes in a person's gait. The view-dependent constraint of our gait appearance feature representation can be removed in a separate paper [SLD01] by synthesizing a walking sequence in a canonical view using the visual hull constructed from multiple cameras.

2 Previous Work

Several researchers have concluded that gait is indicative of a person's gender and identity. Johansson[Joh75] 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.* [CPK78] studied human perception of gait using moving light displays (MLD) similar to that used by Johansson and confirmed human person identification results [CK77].

Several computational approaches have been taken to the task of gait recognition from walking video sequences. Cutler and Davis [CD00] used self-correlation of moving foreground objects to distinguish walking humans from other moving objects such as cars. Polana and Nelson [PN93] detected periodicity in optical flow and used these to recognize activities such as frogs jumping and people walking. Bobick [BD01] used a time delayed motion template to classify activities. Little and Boyd [LB98] used moment features and periodicity of foreground silhouettes and optical flow to identify walkers. Nixon, *et al.* [NCN⁺99] used principal component analysis of images of a walking person to identify the walker by gait. Shutler, *et al.* [SNH00] used higher order moments summed over successive images of a walking sequence to identify persons by their gait.

The work described in this paper is closely related to that of Little and Boyd [LB98]. 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 explore three different ways of combining these image-based features: (1) the parametric modeling of the distribution of image features for gait using a normal distribution assumption, (2) the non-parametric method of using histograms to model the distribution of the gait image feature space, and (3) the spectral decomposition of the fundamental frequency of the gait time series. Both distribution models discard the time dimension of the gait time series, while the spectral decomposition retains some information about the temporal relationship between regions.

3 Gait Sequence Representation

We consider the canonical view of a walking person to be one which is perpendicular to the direction of walk. We also assume that the silhouette of the walker is segmented from the background (details to follow). We would like our gait dynamics feature vector to have the following properties: (a) the ability to describe appearance at a level finer than a whole body description without having to segment individual limbs, (b) robustness to noise in video foreground segmentation, and (c) simplicity of representation. A number of features intuitively come to mind that may measure the static aspects of gait and individual traits. One such feature is the height of an individual, which requires calibrating the camera to recover distances. Other features include the amount of bounce of the whole body in a full stride, the side-to-side sway of the torso, the maximum distance between the front and the back legs at the peak of the swing phase of a stride, the amount of arm and leg swing, etc. Our gait sequence appearance features are computed in two stages. In the first stage, a set of image features are computed using moments from parts of the silhouette. In the second stage, we explore three different methods of aggregating the image features across time to for gait sequence features: (1) the averaged appearance, (2) the appearance histograms, and (3) the fundamental spectral decomposition.

3.1 Gait Image Features

Our gait image feature vector is comprised of parameters of moment features in image regions containing the walking person. For each silhouette of a gait video sequence, we find the centroid and proportionally divide the silhouette into 7 parts as in Figure 1. The vertical line through the centroid is used to divide the silhouette into front and back sections, except for the top portion. The parts above and below the centroid are each equally divided in the horizontal direction, resulting in 7 regions that roughly correspond to: r_1 , head/shoulder region; r_2 , front of torso; r_3 , back of torso; r_4 , front thigh; r_5 , back thigh; r_6 , front calf/foot; and r_7 , back calf/foot. These regions are by no means meant to segment the body parts precisely. We are only interested in a method to consistently divide the silhouette of a walking person into regions that will facilitate the person recognition task.

For each of the 7 regions in a silhouette, we fit an ellipse to the portion of foreground object visible in that region (Figure 1(b)). The features extracted from each region are the centroid, the aspect ratio (l) between the major and minor axes of the ellipse, and the orientation (α) of the major axis of the ellipse, *i.e.*, the region feature vector $f(r_i)$ is

$$f(r_i) = (\bar{x}_i, \bar{y}_i, l_i, \alpha_i), \text{ where } i = 1, \dots, 7. \quad (1)$$

These moment-based features are robust to noise in the silhouettes obtained from background subtraction, provided the number of noise pixels is small and

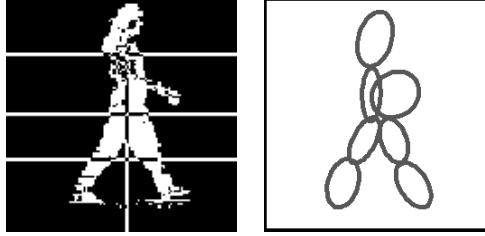


Fig. 1. The silhouette of a foreground walking person is divided into 7 regions, and ellipses are fitted to each region.

not systematically biased. The frame feature vector F_j of the j th frame consists of the 7 region feature vectors, *i.e.*

$$F_j = (f(r_1), \dots, f(r_7)). \quad (2)$$

In addition to these 28 features, we use one additional feature, h , the height (relative to body length) of the centroid of the whole silhouette, to describe the proportions of the torso and legs. This results in a frame feature vector of length 29. The intuition behind this measure is that an individual with a longer torso will have a silhouette centroid that is positioned higher (relative to body length) on the silhouette than someone with a short torso.

3.2 Gait Sequence Features

Given the image-based gait features, we need to concisely represent gait across time. We explore three different methods of aggregating the image features across time to form sequence features: (1) the averaged appearance method, (2) the appearance histogram method, and (3) spectral decomposition. Each of these methods is intended to test assumptions about the nature of the time series that represents gait. The averaged appearance discards the time dimension, decouples the different parts of the components of the body, and assumes that the shape appearance, which we describe from fitted ellipses, is normally distributed. The appearance histogram method differs from the averaged appearance method in that it does not assume any parametric distribution of shape appearance, but it still discards the time dimension and decouples the body components. The spectral decomposition method preserves the coupling between the body components and measures the maximum size of change in appearance, but it assumes that the time series of gait image features are perfectly sinusoidal.

Averaged Appearance Features Specifically, the gait average appearance feature vector of a sequence s is,

$$s = (\text{mean}_j(h_j), \text{mean}_j(F_j), \text{std}_j(F_j)), \quad (3)$$

where $j = 1, \dots, \text{last frame}$, and s is 57-dimensional. This feature set is very simple to compute and robust to noisy foreground silhouettes. Intuitively, the

mean features describe the average-looking ellipses for each of the 7 regions of the body; taken together, the 7 ellipses describe the average shape of the body. The standard deviation features roughly describe the changes in the shape of each region caused by the motion of the body, where the amount of change is affected by factors such as how much one swings one’s arms and legs.

The averaged appearance feature makes the assumption that the underlying distribution of the image features across time is a normal distribution. We use a Mahalanobis distance on the features as the distance between a pair of gait video sequences.

Appearance Histogram Given the 29 gait image features extracted from each frame of a gait video sequence, we tally the features each into a 20-bin histogram, resulting in a 20×29 matrix of gait features. Each histogram is normalized to sum to 1, making them probability distribution functions. The similarity between two gait sequences is measured by a normalized correlation of the histograms and summed over the 29 features. We use a sum instead of a product of the 29 features to reduce the effect from any extremely noisy component.

Fundamental Spectral Decomposition Our gait fundamental spectral decomposition feature vector for a sequence is

$$t = (\Omega_d, |X_i(\Omega_d)|, \text{phase}(X_i(\Omega_d))), \quad (4)$$

where

$$X_i = \text{FourierTransform}(F_{j=1\dots last}(f(r_i))), \quad (5)$$

where Ω_d is the dominant walking frequency of a given sequence. Intuitively, the magnitude feature components measure the amount of change in each of the 7 regions due to motion of the walking body, and the phase components measure the time delay between the different regions of the silhouette. Because of noise in the silhouettes, the time series of region features is also noisy. The power spectra of many region features do not show an obvious dominant peak indicating the fundamental walking frequency. Even when the peak frequencies are found, they often do not agree between different region features. Therefore, we use a normalized averaging of power spectra of all region features resulting in a much more dominant peak frequency that is also consistent across all signals. The magnitude of each region feature at the dominant frequency Ω_d can be directly used, but the phase cannot be directly used because each gait sequence is not predetermined to start at a particular point of a walking cycle. Instead, we compute the phases of all region features relative to the one particular region feature that is “most stable.” We define stability as how closely does a gait time series resemble a pure harmonic, because the phase can be more accurately estimate in a pure harmonic signal. The corresponding quality in the frequency domain is the sharpness of the power spectra around the fundamental frequency, which we measure using the 2nd moment of the power spectra about the fundamental frequency. In our case, the standard phase is that of the x position

of the front calf/foot region. The relative phase features preserves the coupling between components of the walking silhouette. The gait fundamental spectral decomposition feature vector has $57-1$ (the standard phase)=56 dimensions. We use the Mahalanobis distance to compare two gait sequences.

4 Experimental methods and results

We applied our gait dynamics features to person identification. We gathered gait data in indoor environments with different backgrounds on four separate days spanning two months. Twenty-four subjects, 10 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 were flipped so that all sequences are of one walking direction. In all, 194 walking sequences were collected, between 4 to 22 sequences for each subject, averaging 8 sequences per subject. A minimum of 3 complete walking cycles were captured at 15 frames per second for each walking sequence. We then used an adaptive background subtraction algorithm [SG99] to segment out the walking person as a moving foreground object and scale-normalize the silhouettes. An example of the foreground walking person is shown in Figure 2. The foreground segmentation is not perfect—the shadow on the ground and in some cases portions of the background are included. However, our gait representation tolerates this amount of noise in foreground segmentation.



Fig. 2. A sample sequence of the silhouettes of a subject after background subtraction.

Using the gait averaged appearance, gait appearance histogram, and fundamental spectral decomposition features described in the previous section, three feature vectors were computed for each walking sequence in our gait database and used in person identification.

Under the assumption that all components in the feature vectors 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 bias of parameters with large variance. We then use a nearest-neighbor approach to rank a set of gait sequences (and their corresponding features) by their distances to a query sequence. In addition, we applied ANOVA to rank the parameters of our gait features in the multi-person classification task and selected all features with $p < 10^{-9}$, which leaves us with the best 41 feature parameters in the case of the gait averaged appearance feature set, and 32 parameters in the case of the gait fundamental spectral decomposition feature set.

We evaluated our gait recognition performance using the cumulative match score described by Phillips [PMRR97]. Under the closed-universe assumption, all query sequences contain a walker who has at least one other walking sequence in our collection of walking sequences with known person ID’s. We answer the question, “Is the correct person ID in the top n matches?” The cumulative match score has the rank of top matches on the x -axis and the percentage of correct identification on the y -axis.

Two gait recognition tests were performed using each of the three gait feature vectors. In one, the any-day test, a query walking sequence is compared against all sequences except itself, regardless of the day of recording. The second test, which we call the cross-day 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, and subjects may be wearing different styles of clothing (*e.g.* pants vs. skirt).

4.1 Recognition using Averaged Appearance Features

The cumulative match score using the best 41 features in the any-day test is 100% correct identification for the first match. In the case where all 57 features are used, it gives 97% correct identification for the first match and 100% by the 3rd match (Figure 3). The lower performance using the full set of 57 features suggest that the 16 features eliminated by a threshold on the p -value of ANOVA are unstable measures of gait appearance. Closer examination of the gait feature vector distances shows that the top matches are always sequences of the same person taken on the same day, effectively making the any-day test into a same-day test, hence necessitating the second test—comparing sequences taken on different days.

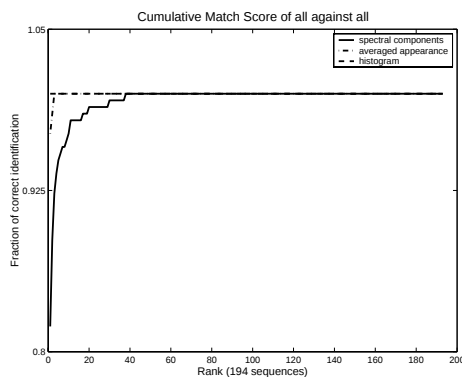


Fig. 3. Cumulative match scores for the all sequences against all sequences test using the 20×29 appearance histogram, all 56 fundamental spectral decomposition features and all 57 averaged appearance features.

The results of the second person identification test, the cross-day test, are shown in Figure 4. 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). Table 1 shows the percentage of correct matches using the 41 features selected using ANOVA vs. using all 57 features. The overall recognition performance is better when using the best 41 features than when using all 57 features.

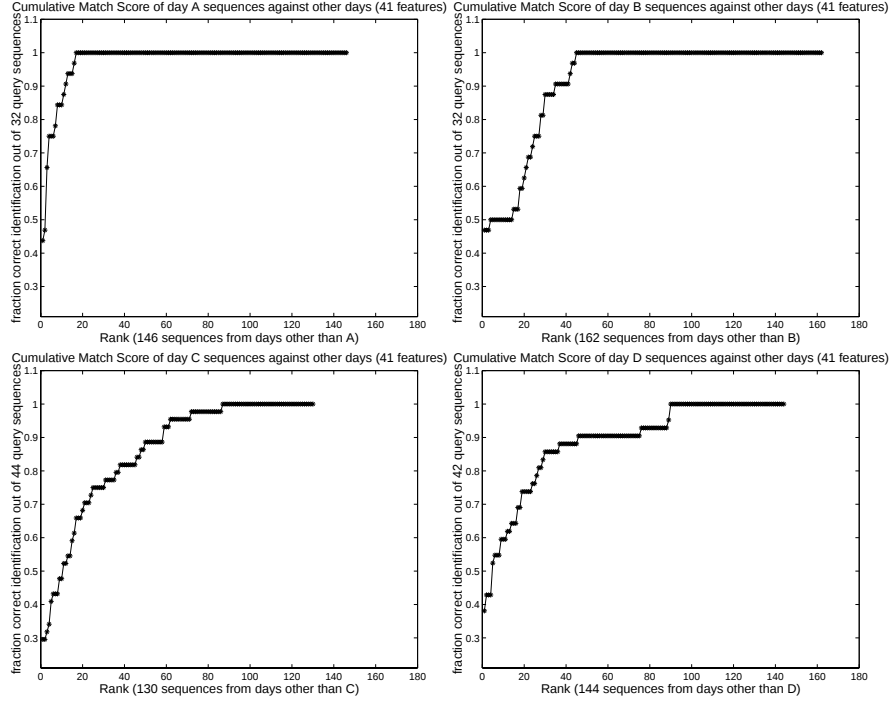


Fig. 4. Cumulative match score for comparing sequences taken on one day against sequences taken on all other days using the 41 best averaged appearance features.

Only 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 the analogous test for any other day. Closer examinations of the ranked matches show that the worse performances on the cross-day B, C, and D tests are the result of lack of similarity between the types of clothing that some of the subjects wore for the day B, C, or D gait sequence data and what they each wore on the other days. In other words, the query sequences show subjects wearing clothing that is substantially

	41 features (% correct)				57 features (% correct)			
	1st	10%	20%	50%	1st	10%	20%	50%
cross-day A	44	94	100	100	47	91	100	100
cross-day B	47	53	88	100	25	56	69	91
cross-day C	30	55	75	95	25	55	80	91
cross-day D	38	64	83	90	26	69	81	90

Table 1. Percentage of correct identification using 41 and 57 averaged appearance features at the best match, the top 10%, 20%, and 50% recalls.

different from that in the library sequences. 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 when he wore 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. Day C contains sequences of one subject wearing a short dress while the only other sequences in the database show her wearing pants.

4.2 Recognition using Appearance Histograms

The appearance histogram gait feature has a 100% recognition rate at the first recall in the any-day test (see Figure 3). Its performance in the cross-day tests are shown in Figure 5.

The correct identification results based on the percentage of recall in Table 2 show that the appearance histogram performs consistently better than the 57-feature averaged appearance features and is better than the 41-feature averaged appearance features beyond the 5% recall level. We conclude that the normal distribution assumed in the averaged appearance features does not adequately represent the underlying distribution. A histogram is more suited to represent the distribution.

test	1st	5%	10%	20%	30%
any-day	100	100	100	100	100
cross-day A	56	100	100	100	100
cross-day B	28	78	97	100	100
cross-day C	52	89	98	100	100
cross-day D	36	79	90	98	100

Table 2. Percentage of correct identification using gait appearance histogram features at the best match, top 5%, 10%, 20%, and 30% recalls

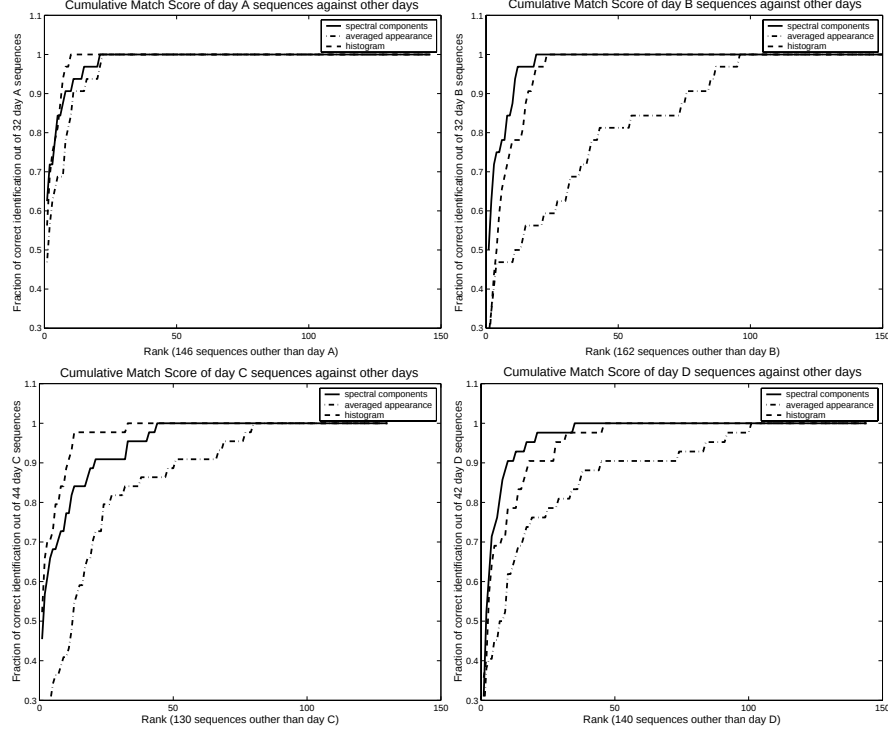


Fig. 5. Cumulative match score for the sequences taken on one day against sequences taken on all other days tests using the complete feature set of each of the fundamental spectral decomposition, the appearance histograms, and the averaged appearance.

4.3 Recognition using Fundamental Spectral Decomposition Features

The cumulative match scores using the gait spectral decomposition features are shown in Figure 3 for the any-day test, and in Figure 5 for the cross-day test.

As is evident in the recognition results, while the fundamental spectral decomposition features do not perform as well as the average appearance features or the appearance histogram features in the any-day test, its recognition performance is much better than that of the averaged appearance features, and is comparable to the performance of the appearance histogram features in the cross-day tests. The fundamental spectral decomposition features performs only marginally better in the day A against day B, C, D test, but significantly better in the other three cross-day tests. Those three cross-day tests are exactly the ones that contain query sequences showing a subject wearing a style of clothing that is not present in the library sequences. This behavior results from the characteristics of the features themselves. The average appearance gait features contain the mean shape of the 7 regions of the silhouettes. The shapes of these regions are very much affected by drastic changes in clothing style, such as pants

test	1st	5%	10%	20%	30%
any-day	82	97	98	100	100
cross-day A	63	88	97	100	100
cross-day B	50	84	97	100	100
cross-day C	45	70	84	91	95
cross-day D	31	81	93	98	100

Table 3. Percentage of correct identification using the gait fundamental spectral decomposition features at the best match, top 5%, 10%, 20%, and 30% recalls

vs. dress, shorts vs. pants. The gait spectral component features only reveal the changes in the shape and the time delay between the different regions, hence it is less affected by the change of clothing style.

We also performed the test using the best 32 fundamental spectral components features in the above two recognition tests. Our results show that the reduced feature set has a similar (although slightly lower) recognition rate.

5 Discussion

The person recognition results using the three aggregations of gait image features—(1) the averaged appearance, (2) the appearance histogram, and (3) the fundamental spectral decomposition—show that the appearance histogram and fundamental spectral decomposition have similar performances, and both are better than the averaged appearance. The appearance histogram outperforms the averaged appearance aggregation method because it captures the underlying image feature distribution more accurately. The fundamental spectral decomposition retains information about the coupling between parts of a walking silhouette, but it assumes that the time series of the gait image features is sinusoidal, hence also assumes a fixed shaped (but not fixed width) distribution of the gait image features. Therefore, we believe that a better set of gait features for person recognition needs to describe not only the coupling between the silhouette components, but also the non-sinusoidal nature of the periodic gait signal. We are actively testing such a set of features.

6 Conclusions

We have presented a set of three gait features for person recognition based on three different time-aggregations of rough silhouette descriptions. We explored the differences between the three aggregation methods and showed their performance on realistic data. Through the recognition results of the three aggregation methods, we inferred the qualities necessary for a better set of gait features. Further work is needed to test our hypotheses.

Acknowledgments

We would like to thank Trevor Darrel for his illuminating discussions. This research is supported by DARPA contract N00014-00-1-0907.

References

- [BD01] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *PAMI*, 23(3), March 2001.
- [CD00] Ross Cutler and Larry Davis. Robust real-time periodic motion detection, analysis, and applications. *PAMI*, 22(8):781–796, 2000.
- [CK77] J.E. Cutting and L.T. Kozlowski. Recognizing friends by their walk: gait perception without familiarity cues. *Bull. Psychonomic Soc.*, (9):353–356, 1977.
- [CPK78] 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.
- [Joh75] G Johansson. Visual motion perception. *Scientific American*, (232):76–88, 1975.
- [LB98] James Little and Jeffrey Boyd. Recognizing people by their gait: The shape of motion. *Videre*, 1(2), 1998.
- [LG02] L. Lee and W.E.L. Grimson. Gait analysis for recognition and classification. In *Proceedings of The 5th International Conference on Automatic Face and Gesture Recognition*, 2002.
- [NCN⁺99] 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.
- [PMRR97] 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.
- [PN93] Ramprasad Polana and Randall Nelson. Detecting activities. In *CVPR*, 1993.
- [SG99] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, 1999.
- [SLD01] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *CVPR*, 2001.
- [SNH00] 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.