

Two-Point Gait: Decoupling Gait from Body Shape

Stephen Lombardi, Ko Nishino
Drexel University, Philadelphia, PA, USA
{sal64, kon}@drexel.edu

Yasushi Makihara, Yasushi Yagi
Osaka University, Suita, Osaka, Japan
{makihara, yagi}@am.sanken.osaka-u.ac.jp

Abstract

Human gait modeling (e.g., for person identification) largely relies on image-based representations that muddle gait with body shape. Silhouettes, for instance, inherently entangle body shape and gait. For gait analysis and recognition, decoupling these two factors is desirable. Most important, once decoupled, they can be combined for the task at hand, but not if left entangled in the first place. In this paper, we introduce Two-Point Gait, a gait representation that encodes the limb motions regardless of the body shape. Two-Point Gait is directly computed on the image sequence based on the two point statistics of optical flow fields. We demonstrate its use for exploring the space of human gait and gait recognition under large clothing variation. The results show that we can achieve state-of-the-art person recognition accuracy on a challenging dataset.

1. Introduction

The study of gait has enjoyed a rich history since its inception with Eadward Muybridge's study [26] of equine locomotion in 1878, and for good reason. A large body of research has shown that human gait patterns contain a great deal of information. For example, it can give information about a person's emotional state [29], health [24], age [11], gender [45], or even distinguish the person themselves [9].

But what is gait, exactly? Gait is defined as the pattern of movement of the limbs of animals during locomotion over a solid substrate [41]. This definition highlights the fact that gait is about motion rather than body shape. Work by Gunnar Johansson has visualized this in his study of pure gait with point light displays [16], which eliminate all body shape information. Stevenage *et al.* later demonstrated the possibility of pure gait person identification with these same displays [34]. On the other hand, many past approaches to modeling gait rely heavily on body shape information for recognition and other computer vision tasks.

Extracting gait from images and video has remained a challenging problem. Gait representations can be roughly categorized into model-based and image-based (model-

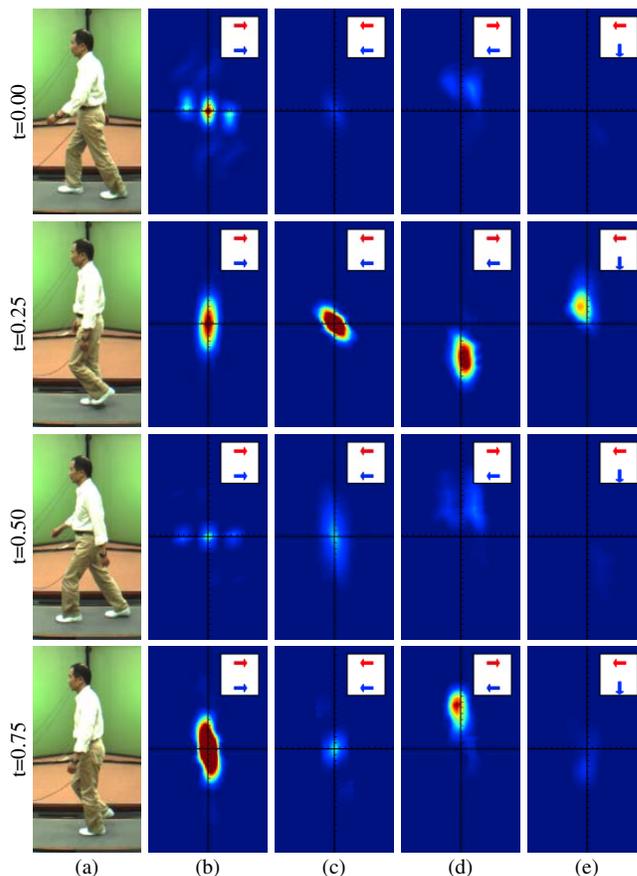


Figure 1: Example Two-Point Gait. Each row corresponds to a different temporal position in the gait cycle. Column (a) shows the original video while columns (b), (c), (d), and (e) show various orientation pairs of the Two-Point Gait. Column (b) is orientation pair (right, right), column (c) is (left, left), column (d) is (right, left), column (e) is (left, down). The Two-Point Gait encodes the body parts that move together in a particular pair of directions as a pseudo-histogram of displacement vectors. It provides an easy-to-compute gait representation that is robust to body shape. Please see text for more details.

free) representations. Model-based gait representations fit limbs or other parts of the body in a gait pattern to a pre-defined 2D or 3D model. Yam *et al.* examine the periodical

and pendulum-like motion of the legs to model gait [43]. Urtasun and Fua fit an articulated 3D skeleton to motion sequences for tracking and recognition [37]. Sigal *et al.* introduce a benchmark for measuring performance on articulated human motion estimation [32] and develop a graphical model for a 3D articulated skeleton [33]. As model-based gait representations typically estimate the 3D motion of the limbs, they are naturally body shape-invariant. The major drawback of these methods is that fitting a latent 3D model to 2D images is inherently ill-posed due to the depth information lost through projection.

Image-based representations, on the other hand, draw features directly from 2D images. A major advantage of image-based approaches is that they obviate the estimation of a latent model. Many previous approaches are based on examining the silhouette of a person over time [14, 17, 22, 25, 40]. The primary problem with silhouette-based approaches is that they fundamentally entangle body shape and gait. Another serious drawback is the difficulty of obtaining silhouettes from a real-world video. Although background modeling and subtraction can be used, it leads to noisy silhouettes that usually require manual intervention. Other image-based approaches are based on optical flow or feature trajectory points [5, 18, 20, 21, 46, 47]. Although they provide a better avenue for a body-shape invariant gait descriptor, they have not been able to capture the nuances of human gait as finely as silhouette-based methods.

Separating body shape from gait is related to the classic problem of “style vs. content,” explored by Tenenbaum and Freeman [36]. Elgammal and Lee study the automatic decomposition of style and content specifically for human gait [10]. In our case, body shape is style—it can be modified by clothing or occlusions—but it is the “content” (*i.e.*, the pure gait) that we want to capture. There have been several works that specifically study the entanglement of body shape and gait for person recognition [8, 38, 39]. They all conclude that recognition becomes strong only when both body shape and gait are used together. However, these findings were performed on relatively small databases (USF [28], CMU MoBo [13], UMD [7], U. Southampton [27]) where there is little difference in silhouette shapes across instances of the same person. In such data, body shape is indeed the most discriminative feature.

But can we always assume that the same person will have the same shape? Clothing, for example, can make a person appear to have a different shape despite identical limb movements. Mood can also heavily influence pose and body shape. Weight gain or loss will also have a significant effect on body shape. The advantage of a pure gait representation is that it is optimal for applications where body shape is orthogonal to the things we would like to measure. For example, determining the general emotional state of a person

based on their gait should function the same despite their body shape. In a medical setting, automatically determining rehabilitative progress from a gait sequence would also need to ignore body shape. Silhouettes are fundamentally limited because they entangle shape and gait. On the other hand, accurately recovering 3D posture is difficult unless we rely on a statistical model that heavily regulates potential poses, which would wash out the subtle differences in gait that we are interested in.

Our goal is to devise an image-based gait representation that factors out the body shape. We achieve this with a novel gait representation based on a statistical distribution of the optical flow: the Two-Point Gait. We focus on optical flow rather than silhouette as it primarily encodes the motion of a person rather than shape and can also be directly computed from the image without any manual intervention. Using the optical flow directly, however, would be just as shape-dependent as using the silhouette. Our key idea is to extract the statistical characteristics of these optical flow fields that encode the gait of the person.

We make use of the two-point statistics, which has been used in a wide variety of scientific fields [12, 15, 19, 23, 31, 44]. We introduce the two-point statistics of optical flow. For two optical flow vectors, a and b , the two-point statistics is defined as the spatial distribution of pairs of pixels in the image whose optical flow vectors are a and b , respectively. The central property of Two-Point Gait is that it encodes the changing spatial distribution of limbs moving in directions a and b . Figure 1 shows an example Two-Point Gait sequence. Many of these pairs of directions have an intuitive meaning. For example, if a points left and b right, the two-point statistics will encode information about the arm movement versus the leg movement because they move in opposition during much of the gait cycle. We expect this representation to be very robust to body shape difference because it is principally encoding the changing spatial distribution of limbs rather than their size.

We reveal the properties of the Two-Point Gait representation with a thorough analysis on a synthetic data set. We introduce a synthetic data set that contains gait motion from the Carnegie Mellon University Motion Capture Database [2] realized with a set of synthetic body shapes created with MakeHuman [3]. By examining the distance matrix of these synthetic data sets, we show that the Two-Point Gait is robust to body shape and appearance variations.

We also demonstrate the use of Two-Point Gait in several important computer vision problems. First, we demonstrate how the Two-Point Gait representation naturally encodes gait motion into an intuitive gait space in which the distance between Two-Point Gaits tells us how similar two people walk regardless of their body shapes. Next, we demonstrate the use of Two-Point Gait for person recognition using the

OU-ISIR clothing variation gait database [4] which is the largest and arguably the most challenging of its kind, comprised of 68 subjects each with at most 32 combinations of clothing for a total of 2,746 gait sequences. We show that, when combined with a body shape representation, the Two-Point Gait achieves the state-of-the-art accuracy for gait recognition with clothing variation. These results clearly demonstrate the power and advantage of having a pure gait representation that can be computed from 2D images.

2. Two-Point Gait

Our goal is to develop an image-based gait representation as insensitive as possible to body shape. We want to only encode properties of the motion, such as the amount of arm swing, without encoding body shape or size. To do this we need to decouple the body shape from the representation. By deriving the two-point statistics of optical flow, we achieve a highly discriminative, body-shape robust representation of human gait: the Two-Point Gait.

2.1. Two-point statistics

Two-point statistics, also known as the two-point co-occurrence (correlation) function, is a spatial distribution of two quantities. It has seen use in areas as diverse as statistical mechanics [12, 23] and astrophysics [31, 44]. In computer vision, it has been used to model the spatial distribution of pixels with two given pixel intensities, a and b [15, 19]. Our key idea is to use the two-point statistics to encode the gait observed in optical flow fields (*i.e.*, 2D vector fields).

2.2. Two-point statistics of optical flow

The most straightforward application of two-point statistics to optical flow would be to define a probability function,

$$P(\mathbf{d}|\mathbf{a}, \mathbf{b}) \propto \left| \left\{ \mathbf{o}(\mathbf{x}_1) = \mathbf{a}, \mathbf{o}(\mathbf{x}_2) = \mathbf{b} \mid \mathbf{x}_1 + \mathbf{d} = \mathbf{x}_2 \right\} \right|,$$

where \mathbf{a} and \mathbf{b} are the two optical flow vectors, $\mathbf{d} = \begin{pmatrix} d_x \\ d_y \end{pmatrix}$ is the spatial displacement, and $\mathbf{o}(\mathbf{x})$ is the optical flow at pixel \mathbf{x} . The problem with this naive approach is that it requires the two optical flow vectors to be equal not only in orientation but also in magnitude. Even if we quantize the optical flow vectors, we would end up with an extremely large collection of probability distribution functions to describe gait. To simplify things, we can consider only optical flow vectors whose directions are equal,

$$P(\mathbf{d}|\mathbf{a}, \mathbf{b}) \propto \left| \left\{ o_\theta(\mathbf{x}_1) = a_\theta, o_\theta(\mathbf{x}_2) = b_\theta \mid \mathbf{x}_1 + \mathbf{d} = \mathbf{x}_2 \right\} \right|.$$

Ignoring magnitudes completely, however, causes near-zero noisy optical flow vectors to outweigh relevant ones. To

avoid this problem, we can generalize the equation and preserve the optical flow magnitudes by dropping the notion of a probability distribution and writing the two-point statistics as,

$$t_{\mathbf{a},\mathbf{b}}(\mathbf{d}) = \sum \left\{ \begin{array}{l} \|\mathbf{o}(\mathbf{x}_1)\| \|\mathbf{o}(\mathbf{x}_2)\| \\ o_\theta(\mathbf{x}_1) = \mathbf{a}_\theta, o_\theta(\mathbf{x}_2) = \mathbf{b}_\theta, \\ \mathbf{x}_1 + \mathbf{d} = \mathbf{x}_2 \end{array} \right\}. \quad (1)$$

This formulation has the advantage of accounting for optical flow magnitude simply without requiring that we store the two-point statistics for a large amount of optical flow vector pairs (\mathbf{a}, \mathbf{b}) . Discarding the probabilistic interpretation also allows the “intensity” variation of the two-point statistics to vary with time which is an important feature for recognition.

Figure 1 visualizes an example gait sequence alongside its corresponding Two-Point Gait. Each Two-Point Gait image is an unnormalized spatial distribution of displacement vectors. The center of each Two-Point Gait image corresponds to $(0, 0)$ displacement. These spatial distributions tell us what the spatial relationship of limbs moving in two particular directions are. We show several pairs of orientations: (right, right), (left, left), (right, left), and (up, down). Identical pairs like (right, right) and (left, left) show the spatial distribution of movement in their respective directions. For example, the (right, left) orientation pair of the Two-Point Gait tells us the spatial arrangement of body parts moving right versus left. Because the arms and legs have a strong horizontal motion and typically move in opposite directions, the (right, left) orientation pair encodes how the arms and legs are spatially arranged. The representation is robust to body shape variation because it is the *spatial distribution* of limbs that is encoded rather than the *size* of the limbs.

3. Computation of Two-Point Gait

Before we can compute the Two-Point Gait, we must compute the optical flow. In our experiments, we used the OpenCV implementation [6] of the Simple Flow algorithm [35]. After the optical flow is computed, we quantize the orientation vectors into directional bins by projecting them onto each orientation bin direction

$$h_{\mathbf{v}}(\mathbf{x}) = \mathbf{o}(\mathbf{x})^T \mathbf{v}, \quad (2)$$

where $h_{\mathbf{v}}$ is the optical flow “histogram” value for bin orientation \mathbf{v} , \mathbf{x} is the pixel location, and \mathbf{o} is the optical flow vector. This serves two purposes: it provides a more informative discretization and preserves magnitude information. In this sense, the optical flow histogram is not truly a histogram (we leave it unnormalized).

After the optical flow has been quantized, we can compute the two-point statistics of optical flow for two orienta-

tions \mathbf{a} and \mathbf{b} ,

$$t_{\mathbf{a},\mathbf{b}}(\mathbf{d}) = \sum_{\mathbf{x}} h_{\mathbf{a}}(\mathbf{x})h_{\mathbf{b}}(\mathbf{x} + \mathbf{d}). \quad (3)$$

The raw two-point statistics are too high-resolution to use directly. We therefore downsample it by quantizing it into spatial bins. The width and height of the bins are set proportional to the height of the subject so that the representation becomes robust to the height of the subject and the distance from the camera. Because the bounding box of each subject will vary slightly, so too will the size of the Two-Point Gait. To handle this and ensure that each Two-Point Gait can be compared with one another, we discard all Two-Point Gait data outside of a rectangle of displacements. In practice, we make this window large enough that we don't lose any information.

We also normalize for the length of the gait cycle by quantizing into temporal bins. Rather than putting one frame into one bin, we deposit each frame into the two closest bins with linear weighting. This helps to reduce quantization errors. In our experiments, we used a total of 16 temporal bins. We also divide each Two-Point Gait by its cycle length so that the temporal quantization does not favor longer-length gait cycles.

3.1. Efficient computation

The complexity of the Two-Point Gait computation is $O(N^2)$ where N is the number of pixels. While we can't reduce the complexity, we can reduce the number of pixels in the computation while keeping the resulting Two-Point Gait very similar.

We do this by downsampling the optical flow histogram computed on the full size images and then computing the two-point statistics on the downsampled optical flow. We downsample the optical flow histogram by collecting it into spatial "blocks" of size $B_W \times B_H$. If the original image was $W \times H$ pixels, the running time goes from W^2H^2 without optical flow downsampling to $\frac{W^2H^2}{B_W^2B_H^2}$ with optical flow downsampling. Even with a modest amount of downsampling, say using 4×4 blocks, a $16 \times$ speed-up is achievable.

3.2. Comparing Two-Point Gaits

We need a distance metric to compare the Two-Point Gait of different people. A simple distance metric can be defined by unrolling the Two-Point Gait each into a long vector and using the Euclidean distance. This does work well, but we can take a step further and learn a distance metric for the application at hand. For instance, for recognizing the person based on their Two-Point Gait, we may learn a distance metric from a training data set such that the discriminative power of the overall representation is maximized. For this, we use a supervised learning method that

attempts to find the most discriminative Two-Point Gait orientation pairs.

We define the distance metric to be a linear combination of the distances between the component orientation pairs

$$d_{\mathbf{w}}(t^{(i)}, t^{(j)}) = \sqrt{\sum_{\mathbf{a},\mathbf{b}} w_{\mathbf{a},\mathbf{b}} \|t_{\mathbf{a},\mathbf{b}}^{(i)} - t_{\mathbf{a},\mathbf{b}}^{(j)}\|^2}, \quad (4)$$

where $t^{(i)}$ and $t^{(j)}$ are Two-Point Gaits, \mathbf{w} is the set of weights for all orientation pairs, and $w_{\mathbf{a},\mathbf{b}}$ is the weight of orientation pair (\mathbf{a}, \mathbf{b}) . These weights can then be learned from a training data set for the application at hand. For gait-based person recognition, we formulate this as supervised learning in which the weights are learned to keep the distance of instances of the same person small, while ensuring that large distances between instances of different people are preserved.

We can formulate this distance metric learning as a convex optimization problem [42]. If we call \mathcal{S} the set of all pairs of Two-Point Gait from the same person (*i.e.* $(t^{(i)}, t^{(j)}) \in \mathcal{S}$ if and only if $t^{(i)}$ and $t^{(j)}$ were computed from gait sequences of a single person) and \mathcal{D} the set of all pairs of Two-Point Gait not from the same person (*i.e.* $(t^{(i)}, t^{(j)}) \in \mathcal{D}$ if and only if $t^{(i)}$ and $t^{(j)}$ were computed from gait sequences of different people) then we set up an optimization problem,

$$\begin{aligned} \arg \min_{\mathbf{w}} \quad & \sum_{(t^{(i)}, t^{(j)}) \in \mathcal{S}} d_{\mathbf{w}}(t^{(j)}, t^{(i)}) \\ \text{s.t.} \quad & \sum_{(t^{(i)}, t^{(j)}) \in \mathcal{D}} d_{\mathbf{w}}(t^{(i)}, t^{(j)}) \geq 1, \\ & \mathbf{w} \geq 0. \end{aligned} \quad (5)$$

Details on optimizing this objective function are given by Xing *et al.* [42].

4. What Two-Point Gait encodes

Let us now examine the properties of the Two-Point Gait representation. We will show that many of the orientation pairs have an intuitive meaning and correspond to parts of the gait cycle. Following that, we will demonstrate with a synthetic data set the Two-Point Gait's invariance to body shape and appearance.

4.1. Orientation pairs

Figure 2 shows an example video frame, its optical flow, and the Two-Point Gait for that frame. Specific orientation pairs (*e.g.*, left and right) encode particular gait cycle properties. For example, by examining the Two-Point Gait of optical flow vectors pointing left and right, we can observe the changing spatial relationship of the arms versus the legs. Note also how background noise from the optical flow is included in the Two-Point Gait. This shows how we don't

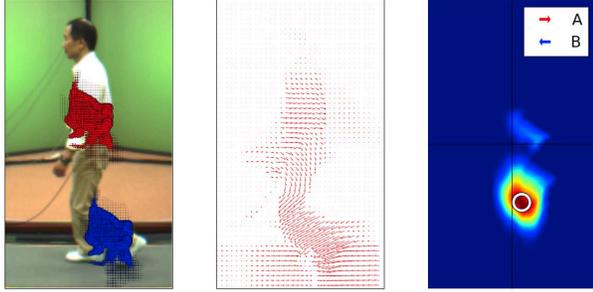


Figure 2: Two-Point Gait and optical flow relationship. The left column shows a frame of video, the middle column shows the optical flow for that frame, and the right column shows the Two-Point Gait for orientation pair (right, left). The red and blue dots in the left column indicate pairs of pixels that contribute to the circled displacement of the Two-Point Gait in the right column. That is, the blue dots have optical flow vectors pointing right, and the red dots have optical flow vectors pointing left. This frame of Two-Point Gait is capturing the spatial distribution between the left arm and left leg because they are moving in opposite directions. Note that it includes some background noise from the optical flow.

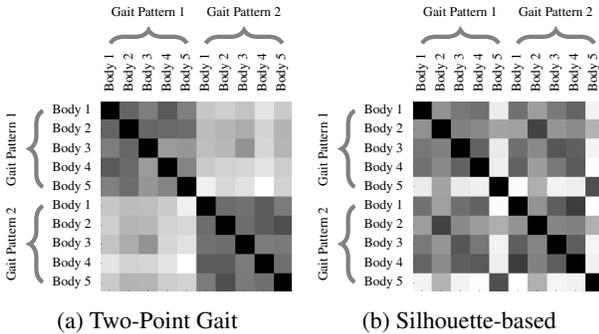


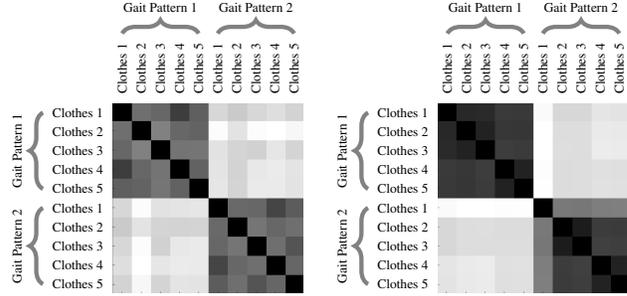
Figure 3: Robustness to Body Shape Variation. Here we show two distance matrices from our synthetic data set to compare the robustness to body shape variation between the Two-Point Gait and a silhouette-based method [4]. The distance matrices have 10 rows and columns, each corresponding to a combination of a motion-captured gait pattern retargeted to a synthetic body shape. The block structure of the matrix in column (a) shows the Two-Point Gait’s robustness to body shape variation.

need an accurate binary mask of the subject, unlike most silhouette-based methods. Despite this inclusion of noise, Two-Point Gait remains discriminative, as we show later.

4.2. Robustness to body shape and appearance

One goal of our representation is to be as insensitive as possible to changes in body shape. We expect that this is the case because the two-point statistics primarily encodes the distribution of moving body parts rather than their sizes.

In this section, we demonstrate the robustness of Two-Point Gait to body shape variation. We created several synthetic human models using MakeHuman [3]. We created human models that vary in body shape and clothing tex-



(a) Two-Point Gait (b) Silhouette-based

Figure 4: Robustness to appearance variation. Here we show two distance matrices from our synthetic data set to compare the robustness to changes in body appearance (texture) between the Two-Point Gait and a silhouette-based method [4]. The distance matrices have 10 rows and columns, each corresponding to a combination of a motion-captured gait pattern retargeted to a synthetic body shape. The block structure of the matrix in column (a) demonstrates the Two-Point Gait’s robustness to appearance variation. The silhouette-based method has this property as well because it ignores all image texture and color by only dealing with the silhouette.

ture. We then selected several motion-captured gait patterns from the Carnegie Mellon University Motion Capture Database [2]. Using Blender [1], we retargeted the motion capture animations to our synthetic human skeletons to create a synthetic gait database.

Figure 3 shows a 10×10 distance matrix of a portion of the synthetic data set. The elements being compared are all possible pairs of 5 synthetic body shapes with two unique gait patterns. The block structure of the distance matrix shows that the representation is extremely robust to body shape variation and discriminates instead against the actual gait. We also compare our distance matrix against a modern silhouette-based method [4]. The silhouette-based method does not have the block structure showing its inability to decouple gait from body shape.

Another important quality for the representation is robustness to changes in body appearance (texture) which also affects optical flow accuracy. Similar to the previous example, we demonstrate this in 4 by showing a 10×10 distance matrix. Here, the elements of the distance matrix consists of all combinations of a synthetic human model with five clothing variations with two unique gait patterns. Again, the strong block structure demonstrates the robustness.

5. Using Two-Point Gait

We now demonstrate the use of Two-Point Gait in important computer vision tasks. We first show how it enables us to map out the gait space, the space of gait patterns of different people. In this space, we can analyze the difference and similarities of how people walk without regard to their

appearance and shapes. We then show that the Two-Point Gait provides us with powerful means to identify people with large clothing variation. We show that Two-Point Gait, without any training, achieve competitive recognition accuracy which is further boosted to achieve the state-of-the-art accuracy when combined with a body shape representation.

5.1. The space of gait patterns

We can use the Two-Point Gait as a representation for an individual’s walk and use the distance between different people’s Two-Point Gait to explore the space of gait patterns—that is, the entire space of how people walk. This space is high-dimensional as each person will be represented by a Two-Point Gait. We may, however, use manifold learning to extract a lower-dimensional manifold on which these gait patterns of different people lie. We use locally linear embedding [30] to visualize this low-dimensional gait space. Figure 5 shows the gait space together with several individuals who lie close to each other in this space.

The low-dimensional embedding shows that people lying close to each other in this space have a characteristic similarity in the way that they walk. Please see the supplemental video for details. The Two-Point Gait faithfully encodes the subtle similarities and dissimilarities of how these people walk, regardless of body shape. The extraction of such a space cannot be done with silhouette-based representations as the body shape is not decoupled from the gait. This gait space can facilitate the study of human locomotion as a function of other attributes such as gender and age, which is only possible with a representation that decouples gait and body shape.

5.2. Gait recognition

The Two-Point Gait, as a pure gait representation, can be used for gait recognition. Conventional “gait recognition,” however, does not necessarily mean recognition of people purely by their gait. Current methods heavily rely on silhouette-based representations, which strongly indicate that the recognition is done based on a mixture of gait and body shape. In fact, when there is no apparent body shape variation (*i.e.*, the person is walking with almost the same clothes and viewing conditions in both the gallery and probe), it is obvious that representations that primarily encode the body shape would suffice.

Obtaining a pure gait representation, however, is still crucial to tackle gait-based person identification. Without disentangling gait and body shape, we cannot understand which contributes more to the recognition and hope to combine them in an optimal manner. This becomes even more important when there is, in fact, change in the body shape between the gallery and probe. Clothing variation is a typical example that causes such shape changes.

To this end, we evaluate the use of Two-Point Gait for person identification on the OU-ISIR clothing variation data set which is arguably the largest and most challenging gait recognition data [4]. The data set consists of 68 subjects each with at most 32 combinations of very different clothing for a total of 2,746 gait sequences divided into three subsets: training, gallery, and probe. To demonstrate how the performance can vary, we evaluate the representation in several ways. First, we evaluate our results on the gallery and probe subsets using a simple Euclidean distance to compare two Two-Point Gaits. Next, we evaluate our results using the learned distance metric described above using the training set. Finally, we combine our representation, with a simple silhouette-based representation, Gait Energy Image (GEI) [14], that encodes the average of the silhouette sequence (*i.e.*, it essentially encodes the body shape as the mean silhouette of the gait cycle). The GEI is computed for each person and the recognition is done using Linear Discriminant Analysis using the training set. We combine Two-Point gait with GEI by first computing the distance matrices for all gallery and probe pairs separately for each representation, and then by taking a linear combination to form a new combined distance matrix (in our experiments we used 45% GEI + 55% TPG) after proper scaling.

Figure 6 shows the receiver operator characteristic curves of the Two-Point Gait with Euclidean distance, Two-Point Gait with learned weighting, GEI, and combination of Two-Point Gait with GEI without and with LDA on the OU-ISIR clothing variation data set. We also compare with the results of Hossain *et al.* [4] with their part-based model, whole-based model, and LDA model. This shows that we achieve state-of-the-art results using a combination of Two-Point Gait and GEI.

These results are noteworthy for multiple reasons. What we notice from the ROC curves is that the Two-Point Gait, without any training, already does fairly well on the challenging clothing variation data set (EER=0.232). This is expected because the Two-Point Gait recognizes gait and ignores body shape. These results also show the limitation of “gait recognition.” This is the accuracy one can hope to achieve for person identification solely based on pure gait observation. The small increase in accuracy from learning the set of orientation pair weights (EER=0.216) also supports that the Two-Point Gait is largely shape robust already. The GEI is the average silhouette over the gait period and therefore is essentially encoding the body shape. When trained on the clothing variation training data, the GEI learns to discriminate body shapes in the space of intra-class variation of specific clothing types. It performs well on the testing data, but what it’s truly doing is learning to classify specific types of shapes as being the same. This can be seen from the fact that GEI gains a significant performance boost with LDA (from EER=0.165 to 0.091),

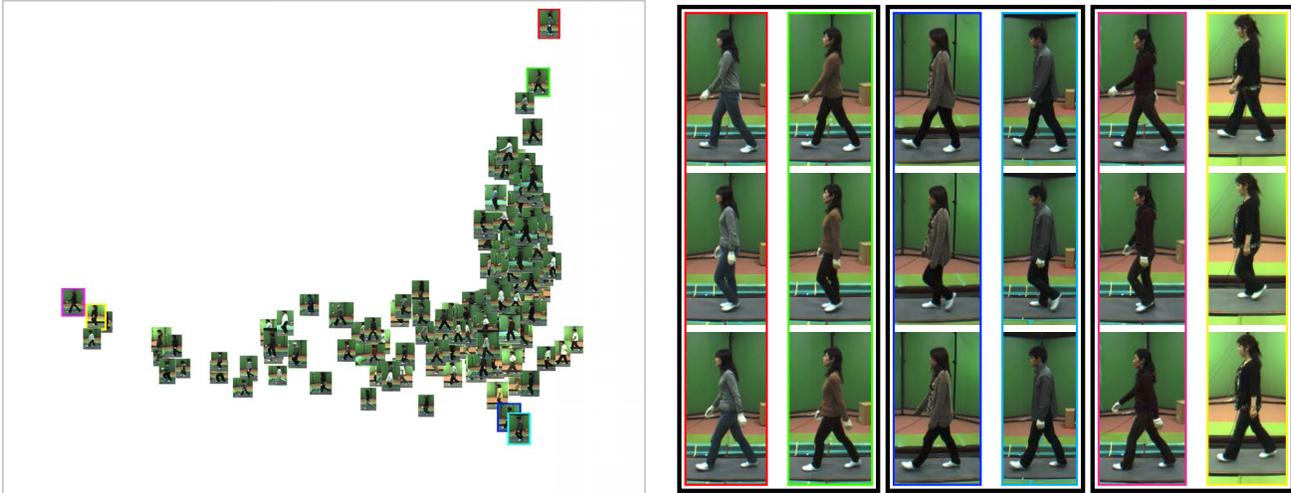


Figure 5: The gait space computed from Two-Point Gait. The left subfigure shows a non-linear embedding of the Two-Point Gait of 370 individuals from the OU-ISIR treadmill data set (without any cloth variation) [4]. Highlighted are three pairs of individuals located in three different parts of the space. On the right, we show three frames of video each from the corresponding people. The frames are at time 0, 0.25, and 0.5, respectively, in the gait cycle. We can see that the pairs of people have similar styles of walking. In particular, notice the similarity of the arm and leg positions for a given gait cycle position. Please see supplementary video for details.

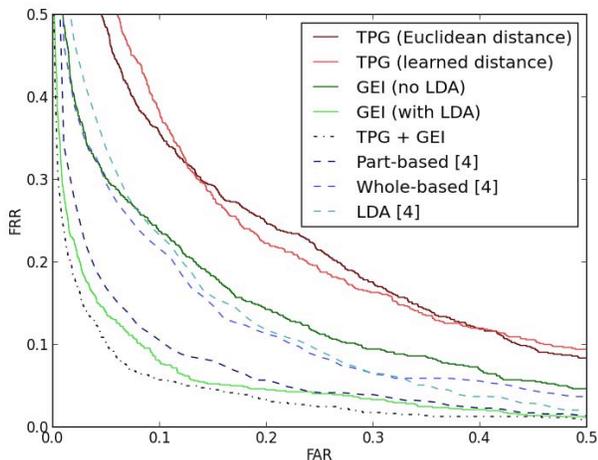


Figure 6: Receiver operator characteristic curves for person recognition with clothing variation using the OU-ISIR database [4]. The ROC curves show that we achieve state-of-the-art results with Two-Point Gait combined with GEI and clearly show the advantage of disentangling gait and body shape. See text for details.

which is in large contrast to the small increase of accuracy of TPG with the metric learning. In other words, GEI with LDA is not learning gait motion but rather learning a better and more subtle shape classifier. When combined with GEI, the Two-Point Gait gives state-of-the-art performance (EER=0.071), which shows the strength of modeling gait separately from body shape. Although the improvement seems small, it is important to note that it becomes significantly more difficult to improve the ROC curve each time (*i.e.*, the improvements are not linear).

We can also see from the figure that the combination of



Two-Point Gait and GEI greatly outperform the approach from Hossain *et al.* Outperforming the LDA and whole-based model from Hossain *et al.* shows that disentangling shape from gait is extremely important: the only way to get such greatly improved performance is by starting with a pure gait and combining it with a pure shape representation. These results clearly show that simply trying to learn invariance to shape from an already shape-entangled representation is limited. We even outperform the part-based method from the same paper. This part-based method relies on adhoc partitioning and would be difficult to apply in general. These results demonstrate the effectiveness of Two-Point Gait as a discriminative gait representation for person identification.

6. Conclusion

In this paper, we introduced a novel gait representation: Two-Point Gait. Two-Point Gait is unique in that it can be directly computed from 2D images yet it encodes purely the gait without regard to the body shape. The experimental results using synthetic gait patterns demonstrate the invariance of the representation to body shape and appearance. We demonstrated the use of Two-Point Gait for exploring the space of people based on their gait and also showed that it allows us to achieve state-of-the-art recognition performance on a challenging gait recognition dataset with clothing variation. We believe this representation has strong implications in a wide range of areas, ranging from human behavior understanding including mood and emotion recognition to medical diagnosis such as assessing the effectiveness of certain rehabilitation regimen. Our code for com-

puting and visualizing Two-Point Gait can be downloaded from <http://www.cs.drexel.edu/~kon/tpg>

References

- [1] Blender. <http://www.blender.org/>.
- [2] CMU graphics lab motion capture database. <http://mocap.cs.cmu.edu/>.
- [3] MakeHuman. <http://www.makehuman.org/>.
- [4] M. Altab Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, June 2010.
- [5] K. Bashir, T. Xiang, and S. Gong. Gait representation using flow fields. In *BMVC*, London, UK, Sep. 2009.
- [6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [7] T. Chalidabhongse, V. Kruger, and R. Chellappa. The UMD database for human identification at a distance. Technical report, University of Maryland, 2001.
- [8] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *IEEE AFGR*, pages 366–371, 2002.
- [9] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5):353–356, 1977.
- [10] A. Elgammal and C. S. Lee. Separating style and content on a non-linear manifold. In *CVPR*, pages 478–485, 2004.
- [11] A. Gabell and U. S. Nayak. The effect of age on variability in gait. *Journal of Gerontology*, 39(6):662–666, 1984.
- [12] C. J. Gommers, Y. Jiao, and S. Torquato. Microstructural Degeneracy associated with a Two-Point Correlation Function and its Information Content. *ArXiv e-prints*, May 2012.
- [13] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, June 2001.
- [14] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE TPAMI*, 28(2):316–322, 2006.
- [15] J. Huang and D. Mumford. Statistics of natural images and models. In *CVPR*, volume 1, pages 547–554, 1999.
- [16] G. Johansson. Visual motion perception. *Scientific American*, 232:75–88, Jun. 1976.
- [17] T. Kobayashi and N. Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *ICPR*, volume 3, pages 741–744, Aug. 2004.
- [18] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44:973–987, April 2011.
- [19] A. B. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *IJCV*, 41(1-2):35–59, Jan. 2001.
- [20] J. Little and J. Boyd. Recognizing people by their gait: The shape of motion. *Videre: JCVR*, 1(2):1–13, 1996.
- [21] Y. Makihara, B. S. Rossa, and Y. Yagi. Gait recognition using images of oriented smooth pseudo motion. In *IEEE SMC*, pages 1309–1314, 2012.
- [22] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *ECCV*, pages 151–163, 2006.
- [23] C. Meneveau and A. B. Chhabra. Two-point statistics of multifractal measures. *Physica A: Statistical Mechanics and its Applications*, 164(3):564–574, 1990.
- [24] M. Montero-Odasso, M. Schapira, E. R. Soriano, M. Varela, R. Kaplan, L. A. Camera, and L. M. Mayorga. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *The journals of gerontology Series A Biological sciences and medical sciences*, 60(10):1304–1309, 2005.
- [25] S. Mowbray and M. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. In *IEEE AVSBS*, pages 566–573, 2003.
- [26] E. Muybridge. Sallie gardner at a gallop. <http://www.sfmuseum.org/hist3/sallie.html>.
- [27] M. Nixon, J. Carter, J. Shutler, and M. Grant. Experimental plan for automatic gait recognition. Technical report, University of Southampton, 2001.
- [28] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. In *ICPR*, volume 1, pages 385–388 vol.1, 2002.
- [29] C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese. Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6):1–32, 2009.
- [30] S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(2323–2326), 2000.
- [31] P. Schneider, L. van Waerbeke, M. Kilbinger, and Y. Mellier. Analysis of two-point statistics of cosmic shear: I. estimators and covariances. 2002.
- [32] L. Sigal, A. O. Balan, and M. J. Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *IJCV*, 87(1-2):4–27, Mar. 2010.
- [33] L. Sigal, M. Isard, H. Haussecker, and M. J. Black. Loose-limbed people: Estimating 3d human pose and motion using non-parametric belief propagation. *IJCV*, 98(1):15–48, May 2012.
- [34] S. Stevenage, M. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13(6):513–526, Dec. 1999.
- [35] M. W. Tao, J. Bai, P. Kohli, and S. Paris. Simpleflow: A non-iterative, sublinear optical flow algorithm. *EG*, 31(2), May 2012.
- [36] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Comp.*, 12(6):1247–1283, June 2000.
- [37] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In *IEEE AFGR*, pages 17–22, 2004.
- [38] A. Veeraraghavan, A. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *IEEE CVPR*, volume 1, pages I-730–I-737 Vol.1, 2004.
- [39] A. Veeraraghavan, A. Roy-Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1896–1909, 2005.
- [40] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan. Human identification using temporal information preserving gait template. *IEEE TPAMI*, 34(11):2164–2176, nov. 2012.
- [41] Wikipedia. Gait — Wikipedia, the free encyclopedia, 2013.
- [42] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *NIPS*, pages 505–512. MIT Press, 2002.
- [43] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.
- [44] J. Yoo. Complete treatment of galaxy two-point statistics: Gravitational lensing effects and redshift-space distortions. *prd*, 79(2):023517, Jan. 2009.
- [45] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE TIP*, 18(8):1905–1910, 2009.
- [46] G. Zhao, R. Chen, G. Chen, and H. Li. Amplitude spectrum-based gait recognition. In *IEEE AFGR*, pages 23–30, 2004.
- [47] G. Zhao, R. Chen, G. Chen, and H. Li. Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach. *IVC*, 24:795–809, 2006.