Gait Shape Estimation for Identification

David Tolliver and Robert T. Collins

Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213, USA
tolliver@ri.cmu.edu

Abstract. A method is presented for identifying individuals by shape, given a sequence of noisy silhouettes segmented from video. A spectral partitioning framework is used to cluster similar poses and automatically extract gait shapes. The method uses a variance-weighted similarity metric to induce clusters that cover disparate stages in the gait cycle. This technique is applied to the HumanID Gait Challenge dataset to measure the quality of the shape model, and the efficacy of shape statistics in human identification.

1 Introduction

Low-level video segmentation routines such as background subtraction frequently generate mislabeled pixels. Intuitively, this follows from the absence of high level constraints, such as a statistical shape model or a parametric appearance model. In the case of articulated motion, such models require laborious hand training and frequently require hand initialization. In contrast, we approach the shape estimation problem by building a catalogue of prototypical views.

Noting that mislabeled pixels tend not to be strongly correlated over time, we appeal to averaging to recover an estimate of the underlying shape and propose an unsupervised sample selection method for sequences of inaccurately segmented video. Approximate shape images are estimated through a spectral partitioning algorithm that minimizes the variation in each cluster. This results in shape estimates that preserve pose information and mitigate pixel mislabeling. In the case of human motion we must recover shape estimates that preserve the configuration of the body, rather than blurring or distorting pose details. To this end, we employ a spectral embedding using a similarity metric tailored to precondition the clustering step.

The method is tested with a human identification algorithm that utilizes shape statistics. We produce results that are competitive with current algorithms, with less computational cost. Beyond testing our shape estimation procedure, this set of experiments sheds light on the quality of gait shape as a discriminative biometric.

2 Related Work

There is a rich body of work describing vision systems for modeling and tracking human bodies (see [6] for a review). However, the vision research community has only recently begun to investigate gait as a biometric [7, 13, 14, 16].

Recent approaches [3, 20, 21, 10, 2, 1, 5, 11, 8, 17] rely almost exclusively on the information contained in binary silhouettes computed by background subtraction. In [10], each silhouette is divided into seven regions, and the first and second order moments from each region are combined into a feature vector. More direct use of silhouette shape is achieved in [20] by performing a Procrustes shape analysis for points on the silhouette boundary. This generates a set of mean shapes that are used for nearest-neighbor classification. In [8], feature vectors are computed from the widths between left and right edges of the silhouette contour, and these features are clustered to form a set of temporal keyframes. This structural information and the information about the temporal transitions between keyframe states are encoded in an HMM.

The work of [11] projects a sequence of silhouettes along the row and column dimensions to form spatio-temporal projection patterns. Each lattice unit from this 1D periodic pattern forms a gait signature for comparison and classification using nearest-neighbor. In contrast, [2] extracts lattice units from the 2D symmetric pattern formed by all pairwise silhouette correlations. These units are input to PCA, followed by k-nearest neighbor classification.

Some work makes the connection between 2D silhouette and 3D body measurements. In [3], four static parameters (body height, torso length, leg length and stride length) are extracted from silhouettes representing the double support phase of the gait cycle, as seen from the side. The work of [21] also uses side views to determining hip and knee joint angles, and fits trigonometric polynomials to the periodic traces of these angles over time. The linear relationship between between stride length and cadence (frequency) is estimated in [2] using periodicity analysis and knowledge of relationship between the camera and the ground plane, yielding a view-invariant classifier.

The work most related to our own uses direct comparison of silhouette shapes to determine identity. In [5] the periodic gait cycle is analyzed to identify double support and midstance keyframes. The silhouettes in these frames are centered and scaled to form a gallery of templates. Incoming silhouettes are compared using normalized correlation and classified using nearest-neighbor. The baseline algorithm of [18] performs full volumetric correlation on sliding temporal subsequences normalized silhouette frames. These two direct approaches, while being very simplistic, both perform well on the Gait Challenge dataset that is used (and described) later in this paper. We will show that our work compares favorably to these methods, while having a much lower computational cost.

3 Shape Estimation

Given a set of noisy observations of a shape, such as those obtained from a background subtraction algorithm, we transform the images into a normalized coordinate frame and determine a soft label for each transformed pixel. We seek to label pixels in the normalized coordinate frame as belonging to a shape configuration by appealing to averages taken at each pixel conditioned on each natural shape class.

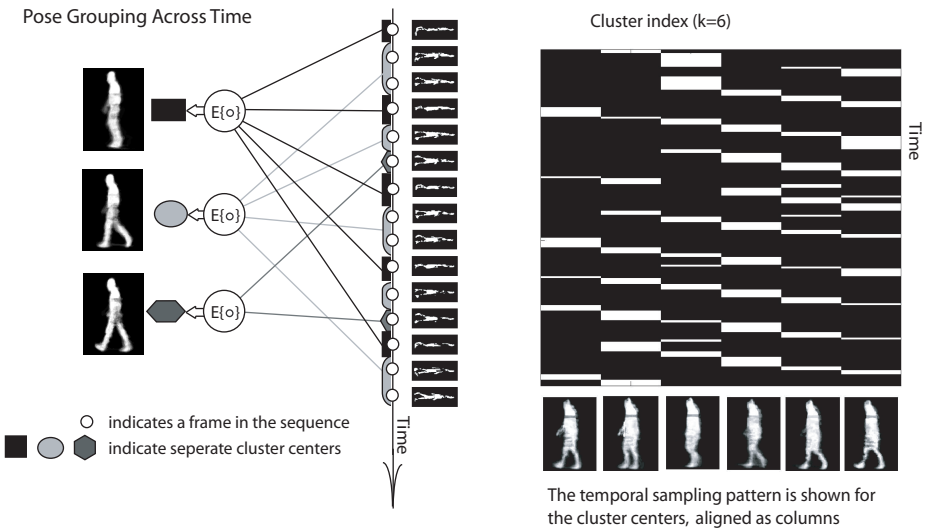


Fig. 1. Selective sample averaging through Spectral Partitioning: (left) illustrates the selection procedure for each natural pose cluster. (right) shows the temporal sampling pattern for each cluster given a periodic motion sequence.

3.1 Shape Normalization

Work on statistical shape theory has provided a set of technologies for making shapes invariant to certain registration parameters. Mardia [12] furnishes a thorough introduction. In our experiments translation and scale were selected as registration parameters.

In preparation for shape clustering, the silhouette images are cropped, and vertically scaled, so that the height is normalized out and the body is centered. The centroid and height of the silhouette in each frame is computed using first and second order moments of the binary silhouette. To be more resilient to noisy silhouette extraction, the time series of moment values over the video sequence is smoothed by averaging moment values within a sliding window that spans one temporal period of the gait cycle, which itself is determined from the time series using a simple periodic frequency analysis.

3.2 Spectral Partitioning

To determine which frames are similar enough to average together we choose a spectral partitioning framework [4] using the Normalized Cuts criteria [19]. We prefer this criteria to the familiar Minimum Cut as we wish to avoid singleton and disproportionately small clusters. Here we construct a graph representation of the inter-image similarity of a sequence, the spectrum of which determines the natural clusters (similar frames) of the sequence. We briefly discuss the graph construction and normalized cuts criteria below.

Define the undirected graph G as $G = (V, E)$, where V is the vertex set corresponding to frames from the sequence and E is the edge set corresponding to the similarity between the incident frames. Given an image sequence the graph is constructed by computing the correlation between frames and storing them in the weight matrix W defined such that $W(i,j) = sim(frame_i, frame_j)$, where *sim* is defined in §3.2.1. As *sim* is symmetric positive semi-definite, the resulting weight W matrix is symmetric non-negative. For efficiency we sample from the sequence of frames such that W remains sparse.

Form the normalized *Laplacian* $\mathcal{L}$ of the graph G as

$$\mathcal{L} = D^{-1/2}(D - W)D^{-1/2} \tag{1}$$

where the diagonal mass matrix D of W is defined as $D(i,i) = \sum_{j=1}^{|W|} W(i,j)$. We can compute the spectrum of $\mathcal{L}$ in matrix form as

$$\mathcal{L}S = S\Lambda \tag{2}$$

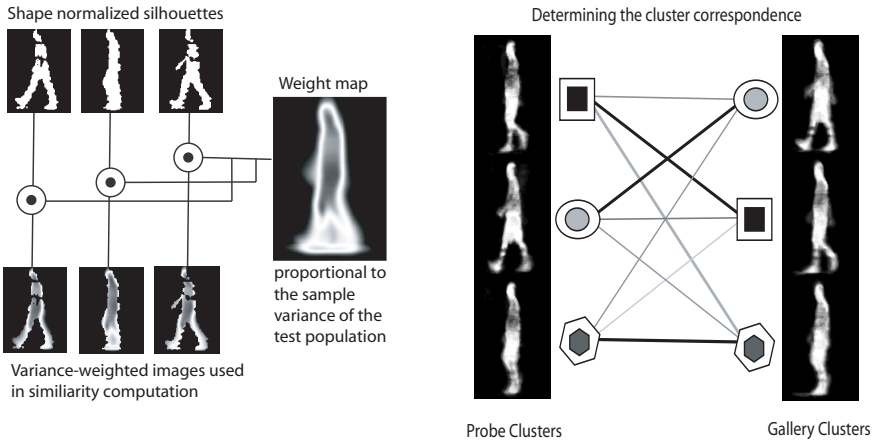
we then take the eigenvectors corresponding to the $\lambda_{2^{nd}}$ through $\lambda_{k+1^{st}}$ smallest eigenvalues of $\mathcal{L}$ and form an embedded space in which to cluster our frames (using k-means). The subspectrum of interest is computed in MATLAB using the `eigs` function. An example of the natural cluster images determined by k-means in the embedded space can be seen in figure 1.

Affinity Metric In constructing W we need to determine the exact form of the affinity function *sim*. Given that we wish to cluster frames into typical poses in the gait sequence we select a weight map proportional to the pixel-wise sample variance of the training silhouettes. This weight map focuses the correlation measure on higher variance areas of the shape, such as arms, legs, and border pixels of the silhouette. Figure 2 depicts the weight map derived from the training silhouettes.

Clustering Clustering into a set of k prototypical shapes is performed using an algorithm similar to Ng *et al.* [15].

1. Compute the frame to frame weighted correlation matrix W from segmented frames in the video sequence.
2. Compute the normalized *Laplacian* $\mathcal{L}$.
3. Compute the spectrum $S\Lambda$ of $\mathcal{L}$.
4. Project the k-dimensional points of $D^{-1/2}S\Lambda$ onto the unit k-sphere.
5. Cluster using k-means on the unit k-sphere.
6. Compute the average frames according to cluster membership, producing k cluster images with pixels values between $[0, 1]$.

In step 5, we use a modification to standard k-means that substitutes geodesic distance on the sphere for the standard L_2-norm.



(a) variance weighted silhouettes (b) Shape matching

Fig. 2. (a) The variance proportional weight map weights leg, arm, and boundary pixels heavily, while down-weighting trunk pixels. (b) The gait shape correspondence is computed as the maximum weighted bipartite graph match.

4 Classification

Given a collection of silhouette shapes, and identities associated with each shape, we wish to identify a probe individual as a particular member of our database. This is accomplished by clustering the probe sequence using the algorithm in §3.2.2, and comparing this collection of shapes to members of the shape database.

Cluster centers are compared using two distinct shape similarity measures described in §5. The measures are positive semi-definite and finite. The match matrix $A \in \Re_+^{k \times k}$ is constructed from all-pairs of correlations between the probe collection clusters and database gallery.

To compare the cluster frames of two sequences we need to match the poses in one frame to those in the other. Given a square match matrix A we seek to find the permutation matrix P_{max} which results in the maximum total correlation. This corresponds to the maximum weight exact bipartite graph match between the clusters from one sequence to those of another. This can be posed as a linear assignment problem for which the Kuhn-Munkres algorithm [9] provides an optimal solution. As the number of clusters centers is small $k \simeq 8$ the $O(n^3)$ computational cost of [9] is acceptable. The final classification is performed by nearest neighbors using the total match score.

5 Experimental Results

5.1 The HumanID Gait Challenge dataset

The HumanID Gait Challenge dataset contains 452 video sequences of 74 walking individuals, making it the largest gait dataset currently available [18, 17]. The videos are collected outdoors under natural illumination, and consist of each person walking around a prespecified test path. Different videos of the same person are collected to explore variations in gait recognition performance with respect to variation in three test factors: small changes in viewpoint, difference in shoe types (hard vs soft heel), and differences in ground surface (concrete vs. grass). The dataset is distributed along with a suggested testing protocol that specifies which sequences to use as training data, and which to use for testing. The test sequences are grouped into seven sets that span the space of covariation between the three binary test factors.

5.2 Registered Matching with Weighted Correlation

We present results for two distinct shape silimarity measures that both use the shape clusters derived in §3. The first measure compares probe and gallery shapes using weighted correlation. We employ the same weight map used in the previous clustering step. The recognition rates are computed for each probe in the Gait Challenge dataset and are presented as SPS_{corr} in tables 1 and 2, along with results for the CMU [5] and USF [17] gait recognition algorithms for comparison.

5.3 Matching with Median Weighted Distance

We define a new similarity measure called median weighted distance (MWD). Probe and gallery cluster images A and B are compared using MWD as

$$D(f(A), f(B)) = \underset{x \in f(A)}{median}(w(x) \min_{y \in f(B)} \|x - y\|_2) \tag{3}$$

where $f(A)$ extracts a level-set contour of the cluster image A, thresholded at a certain frequency value, *e.g.* $A(x) = .5$. The weight map $w(x)$ is inversely proportional to the sample variance of B taken at each pixel. Subsequently, distances in high variance areas of B, such as the boundary region of the shape, are down-weighted. For efficiency, the weight map w and a distance transform is precomputed for each gallery center image in the shape database. Thus computing the match score with a probe image A is a largely a look-up operation. Results for this approach are shown as SPS_{MWD} in tables 1 and 2.

6 Discussion

We have presented a method for the automatic extraction of frequently observed poses in noisy silhouette data. The extracted shape clusters are applied to the

	A	B	C	D	E	F	G	CPU/Subject
SPS_{corr}	85%	81 %	60 %	23 %	17 %	25 %	21 %	$t < 26^{\dagger}$ sec
SPS_{mwd}	82%	66 %	54 %	20 %	18 %	21 %	21 %	$t < 27^{\dagger}$ sec
CMU	87%	81 %	66 %	21 %	19 %	27 %	23 %	$t > 3^{\ddagger}$ min
USF	73%	66 %	56 %	30 %	29 %	18 %	10 %	$t > 3^{\ddagger}$ min

Table 1. Comparison 1, the probe's correct identity is the best match The last column denotes the average CPU time, on an Intel Pentium *III* 1.6 GHz, per subject during classification. † implemented in MATLAB. ‡ implemented in C.

	A	B	C	D	E	F	G	CPU/Subject
SPS_{corr}	90%	87 %	80 %	52 %	43 %	48 %	44 %	$t < 26^{\dagger}$ sec
SPS_{mwd}	98%	90 %	81 %	46 %	43 %	46 %	43 %	$t < 27^{\dagger}$ sec
CMU	100%	90 %	83 %	59 %	50 %	53 %	43 %	$t > 3^{\ddagger}$ min
USF	82%	76 %	54 %	48 %	48 %	41 %	34 %	$t > 3^{\ddagger}$ min

Table 2. Comparison 2, the probe's correct identity occurs in the top 5

HumanID Gait Challenge dataset, with promising results. Our algorithm produces competitive classification results while reducing computational cost.

The results show *rank 5* classification numbers competitive with algorithms [5] and [17]. The *rank 1* numbers are notably lower for tests D and F. Unfortunately segmentation errors and the test covariates are conflated in this dataset, making it difficult to determine the causal factor in classification error.

The classification time for a subject can be factored into two sources, building the shape representation and testing against the database of known subjects. The shape modeling requires approximately 20 seconds, and is performed once per subject. Each comparison against a member of the gallery database requires approximately 8 milliseconds. Consequently our method scales well with additional members, as the computation cost of adding new tests is low.

The results suggest that shape is an appropriate biometric, but that it is a tool best employed when the probe subject is viewed under conditions similar to the gallery subjects. This suggests that local models for each camera would be most successful in a typical surveillance environment. The rank performance of our method indicates that gait shape is an effective winnowing feature, reducing the number of candidates that more computationally intensive methods must analyze.

We plan to apply our technique to human activity analysis, by defining collections of key-shapes that are associated with target activities rather than an identity. We are currently exploring bootstrap estimates of cluster statistics as an analytical tool for determining cluster validity, a problem that is frequently left as an engineering detail in the grouping literature.

Acknowledgment This work is supported by DARPA/IAO HumanID under ONR contract N00014-00-1-0915 and by NSF/RHA grant IIS-0208965.

References

1. C. Ben-Abdelkader, R. Cutler, and L. Davis. Motion-based recognition of people in eigengait space. In *IEEE Conf Automatic Face and Gesture Recognition*, pages 254–259, 2002.
2. C. Ben-Abdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *IEEE Conf Automatic Face and Gesture Recognition*, pages 357–362, 2002.
3. A. Bobick and A. Johnson. Gait recognition using static, activity-specific parameters. In *IEEE Computer Vision and Pattern Recognition*, pages I:423–430, 2001.
4. F. R. K. Chung. *Spectral Graph Theory*. AMS, 2nd edition, 1997.
5. R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *IEEE Conf Automatic Face and Gesture Recognition*, pages 351–356, 2002.
6. D. Gavrila. The visual analysis of human movement: A survey. *CVIU*, 73(1):82–98, January 1999.
7. J.J.Little and J.E.Boyd. Recognizing people by their gait: The shape of motion. In *Videre (online journal)*, volume 1(2), Winter 1998.
8. A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger. Gait-based recognition of humans using continuous HMMs. In *IEEE Conf Automatic Face and Gesture Recognition*, pages 321–326, 2002.
9. H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955.
10. L. Lee and W. Grimson. Gait analysis for recognition and classification. In *IEEE Conf Automatic Face and Gesture Recognition*, pages 148–155, 2002.
11. Y. Liu, R. Collins, and Y. Tsin. Gait sequence analysis using frieze patterns. In *European Conference on Computer Vision*, pages II: 657–671., 2002.
12. K. V. Mardia and I. L. Dryden. *Statistical Shape Analysis*. John Wiley and Son, 1st edition, 1999.
13. M.Nixon, J.Carter, D.Cunado, P.Huang, and S.Stevenage. Automatic gait recognition. In A.Jain, R.Bolle, and S.Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 231–249. Kluwer Academic Publishers, 1999.
14. H. Murase and R. Sakai. Moving object recognition in eigenspace representation: Gait analysis and lip reading. *Pattern Recognition Letters*, 17(2):155–162, Feb 1996.
15. A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Neural Information Processing Systems*, 2002.
16. S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *IEEE Proceedings Computer Vision and Pattern Recognition*, pages 469–474, 1994.
17. P. J. Philips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. Baseline results for the challenge problem of human ID. In *IEEE Conf Automatic Face and Gesture Recognition*, 2002.
18. P. J. Philips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. Gait identification challenge problem: Data sets and baseline. 1:385–388, 2002.
19. J. Shi and J. Malik. Normalized cuts and image segmentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
20. L. Wang, H. Ning, W. Hu, and T. Tan. Gait recognition based on procrustes shape analysis. In *International Conference on Image Processing (ICIP)*, pages III: 433–436, 2002.
21. J. Yoo, M. Nixon, and C. Harris. Model-driven statistical analysis of human gait motion. In *International Conference on Image Processing*, pages I: 285–288, 2002.