

A Multi-view Method for Gait Recognition Using Static Body Parameters

Amos Y. Johnson¹ and Aaron F. Bobick²

¹ Electrical and Computer Engineering
Georgia Tech, Atlanta, GA 30332
amos@cc.gatech.edu

² GVU Center/College of Computing
Georgia Tech, Atlanta, GA 30332
afb@cc.gatech.edu

Abstract. A multi-view gait recognition method using recovered static body parameters of subjects is presented; we refer to these parameters as *activity-specific biometrics*. Our data consists of 18 subjects walking at both an angled and frontal-parallel view with respect to the camera. When only considering data from a single view, subjects are easily discriminated; however, discrimination decreases when data across views are considered. To compare between views, we use ground truth motion-capture data of a reference subject to find scale factors that can transform data from different views into a common frame (“walking-space”). Instead of reporting percent correct from a limited database, we report our results using an expected confusion metric that allows us to predict how our static body parameters filter identity in a large population: lower confusion yields higher expected discrimination power. We show that using motion-capture data to adjust vision data of different views to a common reference frame, we can get achieve expected confusions rates on the order of 6%.

1 Introduction

Automatic gait recognition is new emerging research field with only a few researched techniques. It has the advantage of being unobtrusive because body-invading equipment is not needed to capture gait information. From a surveillance perspective, gait recognition is an attractive modality because it may be performed at a distance, surreptitiously.

In this paper we present a gait recognition technique that identifies people based on static body parameters recovered during the walking action across multiple views. The hope is that because these parameters are directly related to the three-dimensional structure of the person they will be less sensitive to error introduced by variation in view angle. Also, instead of reporting percent correct (or recognition rates) in a limited database of subjects, we derive an expected confusion metric that allows us to predict how well a given feature vector will filter identity over a large population.

1.1 Previous Work

Perhaps the first papers in the area of gait recognition comes from the Psychology field. Kozlowski and Cutting [8,4] determined that people could identify other people base solely on gait information. Stevenage, Nixon, and Vince [12] extended the works by exploring the limits of human ability to identify other humans by gait under various viewing conditions.

Automatic gait-recognition techniques can be roughly divided into model-free and model-based approaches. Model-free approaches [7,9,10] only analyze the shape or motion a subject makes as they walk, and the features recovered from the shape and motion are used for recognition. Model-based techniques either model the person [11] or model the walk of the person [3]. In person models, a body model is fit to the person in every frame of the walking sequence, and parameters (i.e. angular velocity, trajectory) are measured on the body model as the model deforms over the walking sequence. In walking models, a model of how the person moves is created, and the parameters of the model are learned for every person.

Because of the recency of the field, most gait recognition approaches only analyze gait from the side view without exploring the variation in gait measurements caused by differing view angles. Also, subject databases used for testing are typically small (often less than ten people); however, even though subject databases are small, results are reported as percent correct. That is, on how many trials could the system correctly recognize the individual by choosing its best match. Such a result gives little insight as to how the technique might scale when the database contains hundreds or thousands or more people.

1.2 Our Approach

Our approach to the study of gait recognition attempts to overcome these deficiencies by taking three fundamentally different steps than previous researchers.

First, we develop a gait-recognition method that recovers static body and stride parameters of subjects as they walk. Our technique does not directly analyze the dynamic gait patterns, but uses the action of walking to extract relative body parameters. This method is an example of what we call *activity-specific biometrics*. That is, we develop a method of extracting some identifying properties of an individual or of an individual's behavior that is only applicable when a person is performing that specific action. Gait is a excellent example of this approach because not only do people walk much of the time making the data accessible, but also many techniques for activity recognition are able to detect when someone is walking. Examples include the motion-history method of Bobick and Davis [5] and even the walker identification method of Nyogi and Adelson [11].

Second, we develop a walking-space adjustment method that allows for the identification of a subject walking at different view angles to the viewing plane of a camera. Our static body parameters are related to the three-dimensional structure of the body so they are less sensitive to variation in view angle. However, because of projection into an image, static body parameters recovered from different views need to be transformed to a common frame.

Finally, as opposed to reporting percent correct, we will establish the uncertainty reduction that occurs when a measurement is taken. For a given measured property, we establish the spread of the density of the overall population. To do so requires only enough subjects such that our estimate of the population density approaches some stable value. It is with respect to that density that we determine the expected variation in the measurement when applied to a given individual.

The remainder of this paper is as follows: we describe the expected confusion metric we used to evaluate our technique, present the gait-recognition method, and describe how to convert the different view-angle spaces to a common walking-space. Last, we will assess the performance of our technique using the expected confusion metric.

2 Expected Confusion

As mentioned our goal is not to report a percent correct of identification. To do so requires us to have an extensive database of thousands of individuals being observed under a variety of conditions. Rather, our goal is to characterize a particular measurement as to how much it reduces the uncertainty of identity after the measurement is taken.

Many approaches are possible. Each entails first estimating the probability density of a given property vector \mathbf{x} for an entire population $P_p(\mathbf{x})$. Next we must estimate the uncertainty of that property for a given individual once the measurement is known $P_I(\mathbf{x}|\eta = \mathbf{x}_0)$ (interpreted as what is the probability density of the true value of the property \mathbf{x} after the measurement η is taken). Finally, we need to express the average reduction in uncertainty or the remaining confusion that results after having taken the measurement.

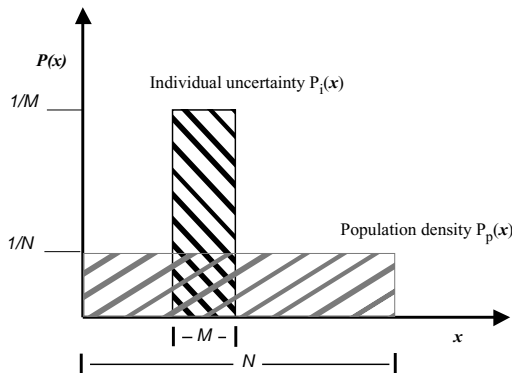


Fig. 1. Uniform probability illustration of how the density of the overall population compares to the the individual uncertainty after the measurement is taken. In this case the remaining confusion — the percentage of the population that could have given rise to the measurement — is M/N .

Information theory argues for a mutual information [2] measure:

$$I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}). \quad (1)$$

where $H(\mathbf{X})$ is the entropy of a random variable \mathbf{X} defined by

$$H(\mathbf{X}) = - \int_x p(x) \ln p(x),$$

and $H(\mathbf{X}|\mathbf{Y})$ is the conditional entropy of a random variable \mathbf{X} given another random variable \mathbf{Y} defined by:

$$H(\mathbf{X}|\mathbf{Y}) = - \int_{x,y} p(x,y) \ln p(x|y).$$

For our case the random variable \mathbf{X} is the underlying property (of identity) of an individual before a measurement is taken and is represented by the population density of the particular metric used for identification. The random variable \mathbf{Y} is an actual measurement retrieved from an individual and is represented by a distribution of the individual variation of an identity measurement. Given these definitions, the uncertainty of the property (of identity) of the individual given a specific measurement, $H(\mathbf{X}|\mathbf{Y})$, is just the uncertainty of the measurement, $H(\mathbf{Y})$. Therefore the mutual information reduces to:

$$I(\mathbf{X}; \mathbf{Y}) \equiv H(\mathbf{X}) - H(\mathbf{Y}). \quad (2)$$

Since the goal of gait recognition is filtering human identity this derivation of mutual information is representative of filtering identity. However, we believe that a better assessment (and comparable to mutual information) of a metric's ability to filter identity is the expected value of the percentage of the population eliminated after the measurement is taken. This is illustrated in Figure 1. Using a uniform density for illustration we let the density of the feature in the population P_p be $1/N$ in the interval $[0, N]$. The individual density P_i is much narrower, being uniform in $[x_0 - M/2, x_0 + M/2]$. The confusion that remains is the area of the density P_p that lies under P_i . In this case, that confusion ratio is M/N .

An analogous measure can be derived for the Gaussian case under the assumption that the population density σ_p is much greater than the individual variation σ_i . In that case the expected confusion is simply the ratio σ_i/σ_p , the ratio of standard deviation of the uncertainty after measurement to that before the measurement is taken. Note that if the negative natural logarithm of this is taken we get:

$$- \ln\left(\frac{\sigma_i}{\sigma_p}\right) = \ln \sigma_p - \ln \sigma_i, \quad (3)$$

we arrive at an expression that is the mutual information (of two 1D Gaussian distributions) from Equation 2. For the multidimensional Gaussian case, the result is

$$\text{Expected Confusion} = \frac{|\Sigma_i|^{1/2}}{|\Sigma_p|^{1/2}}. \quad (4)$$

This quantity is the ratio of the individual variation volume over the population volume. These are volumes of equal probability hyper-ellipsoids as defined by the Gaussian densities. See [1] for complete proof.

3 Gait Recognition Method

Using a single camera with the viewing plane perpendicular to the ground plane, 18 subjects walked in an open indoor-space at two view angles: a 45° path (angle-view) toward the camera, and a frontal-parallel path (side-view) in relation to the viewing plane of the camera. The side-view data was captured at two different depths, 3.9 meters and 8.3 meters from camera. These three viewing conditions are used to evaluate our multi-view technique.

In the following subsections we explain our body part labeling technique and our depth compensation method. The body part labeling technique is used to arrive at the static body parameters of a subject. Depth compensation is used to compensate for depth changes of the subject as the walk. Lastly, before stating the results of the experiments, we present the static body parameters used and how we adjust

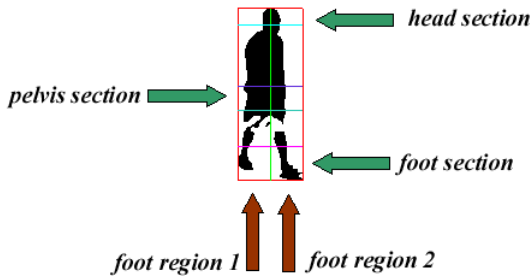


Fig. 2. Automatic segmenting of the body silhouette into regions.

3.1 Body Part Labeling

Body parts are labeled by analyzing the binary silhouette of the subject in each video frame. Silhouettes are created by background subtraction using a static background frame. A series of morphological operations are applied to the resulting images to reduce noise. Once a silhouette is generated, a bounding box is placed around the silhouette and divided into three sections – head section, pelvis section, and foot section (see Figure 2) – of predefined sizes similar to the body part labeling method in [6]. The head is found by finding the centroid of the pixels located in the head section. The pelvis is contained in pelvis section, and is the centroid of this section. The foot section houses the lower legs and feet, and is further sub-divided down the center into foot region 1 and foot region 2. Within foot region 1 and foot region 2, the distance (L2 norm) between each pixel and the previously discovered head location is calculated. The pixel location with the highest distance in each region is labeled foot 1 and foot 2. The labels do not distinguish between left and right foot because it is not necessary in

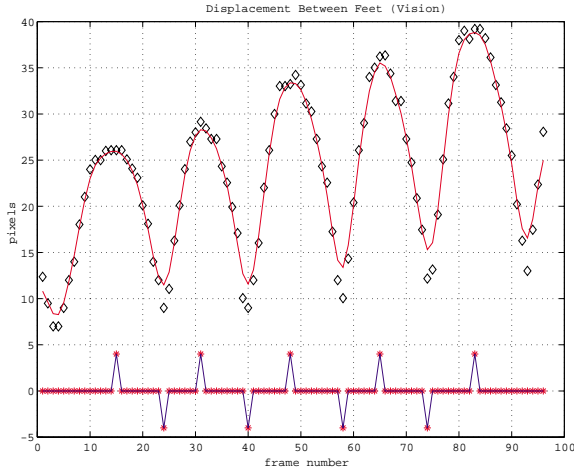


Fig. 3. The distance between the two feet as measured in pixels. The value increases as the subject approaches the camera. The curve is an average that underestimates the value of the peak but well localizes them. The lower trace indicates the maximal and minimal separations.

our technique. This method of body part labeling allows for imperfections in the silhouette due to noisy background subtraction by using local body part searches and placing soft constraints on body part locations.

3.2 Depth Compensation

The static body parameters used for identification will be a set of distances between the body parts locations, and the distances will be measured in pixels; however, a conversion factor from pixels to centimeters is needed for the possible depth locations of the subjects in the video footage. We have created a depth compensation method to handle this situation by having a subject of known height walk at an angle towards the camera. At the points of minimal separation of the subject’s feet (see Figure 3), the system measures the height (this is taken to be the height of the bounding box around the subject) of the subject in pixels at that location on the ground plane. The minimal point represents the time instances where the subject is at his or her maximal height during the walking action. A conversion factor from pixels to centimeters at each known location on the ground (taken to be the lower y -value of the bounding box) is calculated by:

$$\text{Conversion Factor} = \frac{\text{known height (centimeters)}}{\text{measured height (pixels)}}. \tag{5}$$

To extrapolate the conversion factors for the other unknown locations on the ground plane a hyperbola is fit to the known conversion factors. Assuming a world coordinate system located at the camera focal point and an image plane perpendicular to ground plane, using perspective projection we derive a

conversion factor hyperbola,

$$\text{Conversion Factor}(y_b) = \frac{A}{B - y_b}, \tag{6}$$

where A is the vertical distance between the ground and focal point times the focal length, B is the optical center (y component) of the image plus a residual (if the image plane is not exactly perpendicular to the ground), and y_b is the current y -location of the subject’s feet. We implicitly estimate the parameters A and B by fitting the conversion factor hyperbola (Equation 6) to the known locations of the subject and the required conversion factors needed to covert the measured height in pixels to its known height in centimeters (see Figure 4).

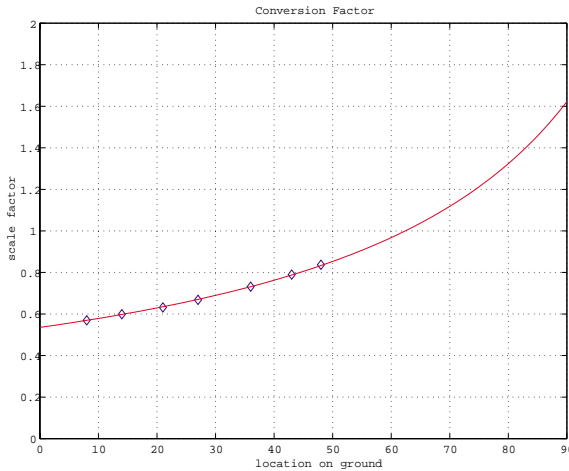


Fig. 4. Hyperbola fit to the data relating the lower y position of the bounding box to the required conversion factor. The data points are generated by observing a subject of known height walking in the space.

3.3 Static Body Parameters

After body labeling and depth compensation, a 4D-walk vector (which are the static body parameters) is computed as (see Figure 5):

- d_1 : The height of the bounding box around the silhouette.
- d_2 : The distance (L2 norm) between the head and pelvis locations.
- d_3 : The maximum value of the distance between the pelvis and left foot location, and the distance between the pelvis and right foot location.
- d_4 : The distance between the left and right foot.

These distances are concatenated to form a 4D-walk vector $\mathbf{w} = \langle d_1, d_2, d_3, d_4 \rangle$, and they are only measured when the subjects’ feet are maximally spread during the walking action. As subjects walk they have multiple maximally spread

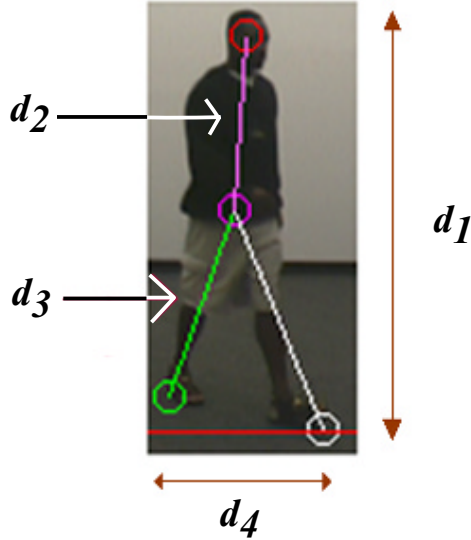


Fig. 5. The 4 static body parameters: $\mathbf{w} = \langle d_1, d_2, d_3, d_4 \rangle$.

points (see Figure 3), and the mean value of \mathbf{w} at these points is found to generate one walk vector per walking sequence. Measurements are taken only at these points because the body parts are not self-occluding at these points, and this is a repeatable point in the walk action to record similar measurements.

3.4 Walking-Space Adjustment

The static body parameters recovered from subjects, from a single view angle, produce high discrimination power. When comparing across views, however, discrimination power decreases. The most obvious reason is that forshortening changes the value of many of the features. Furthermore, variations in how the part labeling techniques work in the different views can lead to a systematic variation between the views. And finally, other random error can occur when doing vision processing on actual imagery; this error will tend to be larger across different views.

In this paper we did not attempt to adjust for random error, but instead compensate for a variety of systematic error including forshortening. We assume that the same systematic error is being made for all subjects for each view angle. Therefore, we can use one subject as a reference subject and use his vision data, from different view angles, to find a scale factor to convert his vision data to a common frame using his motion-capture data as the reference.

Motion-capture data of a reference subject, is considered to be the ground truth information from the subject with minimal error. Our motion-capture system uses magnetic sensors to capture the three-dimensional position and orientation of the limbs of the subject as he (or she) walks along a platform. Sixteen

sensors in all are used: (1) head, (2) torso, (1) pelvis, (2) hands, (2) forearms, (2) upper-arms, (2) thighs, (2) calves, (2) feet. If the error is truly systematic, then the scale factor found, using the motion-capture system, can be applied to the other subjects' vision data.

To achieve this, we model the error as a simple scaling in each dimension of the 4D-walk vector, which can be removed by a simple constant scale factor for each dimension. A mean 4D-walk vector

$$\bar{\mathbf{x}} = \langle d_{x1}, d_{x2}, d_{x3}, d_{x4} \rangle$$

from motion-capture walking sequences of a reference subject is recovered. Next, several (vision recovered) 4D-walk vectors,

$$\mathbf{w}_{ij} = \langle d_{w1}, d_{w2}, d_{w3}, d_{w4} \rangle$$

where i is the view angle and j is the walk vector number, are found of the reference subject from the angle-view, the near-side-view, and the far-side-view. The walk-vector, $\bar{\mathbf{x}}$, from the motion-capture system is used to find the constant scale factors needed to convert the vision data of the reference subject for each dimension and view angle separately by:

$$\mathbf{S}_{ij} = \langle \frac{d_{x1}}{d_{w1}}, \frac{d_{x2}}{d_{w2}}, \frac{d_{x3}}{d_{w3}}, \frac{d_{x4}}{d_{w4}} \rangle$$

where \mathbf{S}_{ij} is scale factor vector for view angle i and walk vector j , and the scale factor vector for a given view angle is

$$\mathbf{SF}_i = \langle sf_1, sf_2, sf_3, sf_4 \rangle = \frac{1}{N} \sum_{j=1}^N S_{ij}. \tag{7}$$

The 4D-walk vectors of each subject are converted to walking-space by

$$\mathbf{w}_{ij} \cdot \mathbf{SF}_i = \langle d_1 \cdot sf_1, d_2 \cdot sf_2, d_3 \cdot sf_3, d_4 \cdot sf_4 \rangle .$$

3.5 Results

We recorded 18 subjects, walking at the angle-view, far-side-view, and near-side-view. There are six data points (walk vectors) per subject for the angle-view, three data points per subject for the side-view far away, and three data per subject for the side-view close up yielding 108 walk vectors for the angle-view and 108 walk vectors for the side-view (54 far way, and 54 close up). The results are listed in Table 1.

Table 1 is divided into two sets of results: Expected Confusion and Recognition Rates. The Expected Confusion is the metric discussed in Section 2. The Recognition Rates are obtain using Maxim Likelihood. Where, recognition is computed by modeling each individual as a single Gaussian and selecting the class with the greater likelihood.

Results are reported from the angle-view, near-side-view and far-side-view. Finally results are posted after the vision data was scaled to walking-space using

Table 1. The results of the multi-view gait-recognition method using static body parameters.

<i>Viewing Condition</i>	<i>Expected Confusion</i>	<i>Recognition Rates</i>
Angle View	1.53%	100%
Side View Far	.71%	91%
Side View Near	.43%	96%
Side View Adjusted (far and near)	4.57%	100%
Combine Angle and Side Views Adjusted	6.37%	94%

the appropriate scale factor based on the viewing condition. The results in the last row, titled *Combine Angle and Side Views Adjusted*, are the numbers of interests because this data set contains all data adjusted using the walking-space adjustment technique.

Once the data points are adjusted by the appropriate scale factors the expected confusion of the Side View (combining near and far) is only 4.57%. Also, the Combined Angle and Side views yield an expected confusion of 6.37%. This tells us that an individual's static body parameters will yield on average 6% confusion with another individual's parameters under these different views.

4 Conclusion

This paper has demonstrated that gait recognition can be achieved by static body parameters. In addition, a method to reduce the variance between static body parameters recovered from different views was present by using the actual ground truth information (using motion-capture data) of the static body parameters found for a reference subject. As with any new work, there are several next steps to be undertaken. We must expand our database to test how well the expected confusion metric predicts performance over larger databases. Experiments must be ran under more view angles, so the error over other possible views can be characterize. Also, the relationship between the motion-capture data and vision data needs to be explored further to find the best possible scaling parameters to reduce the expected confusion even lower than presented here. Lastly, in this paper we compensated for systematic error, and not random error. In future work, we will analyze how to determine random error, and attempt to compensate for (or minimize the effects of) the random error.

References

1. Bobick, A. F. and A. Y. Johnson, "Expected Confusion as a Method of Evaluating Recognition Techniques," Technical Report GIT.GVU-01-01, Georgia Institute of Technology, 2001. <http://www.gvu.gatech.edu/reports/2001/>.
2. Cover, T. M. and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.
3. Cunado, D., M. S. Nixon, and J. N. Carter, "Automatic Gait Recognition via Model-Based Evidence Gathering," accepted for *IEEE AutoID99*, Summit NJ, 1999.
4. Cutting, J. and L. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society* **9** pp. 353–356, 1977.
5. Davis, J.W. and A.F. Bobick, "The representation and recognition of action using temporal templates," *Proc. IEEE Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp 928–934, 1997.
6. Haritaoglu, I., D. Harwood, and L. Davis, "W4: Who, When, Where, What: A real time system for detecting and tracking people," *Proc. of Third Face and Gesture Recognition Conference*, pp. 222–227, April 1998.
7. Huang, P.S., C. J. Harris, and M. S. Nixon, "Human Gait Recognition in Canonical Space using Temporal Templates," *IEEE Procs. Vision Image and Signal Processing*, **146**(2), pp. 93–100, 1999.
8. Kozlowski, L. and J. Cutting, "Recognizing the sex of a walker from a dynamic point-light display," *Perception and Psychophysics*, **21** pp. 575–580, 1977.
9. Little, J.J. and J.E. Boyd, "Recognizing people by their gait: the shape of motion," *Videre*, **1**, 1996.
10. Murase, H. and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognition Letters*, **17**, pp. 155–162, 1996.
11. Niyogi, S. and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," *Proc. Computer Vision and Pattern Recognition*, pp. 469–474, Seattle, 1994.
12. Stevenage, S., M. S. Nixon, and K. Vince, "Visual Analysis of Gait as a Cue to Identity," *Applied Cognitive Psychology*, **13**, pp. 00-00, 1999.