Gait Recognition Using Static, Activity-Specific Parameters

Aaron F. Bobick

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

Amos Y. Johnson

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

Abstract

A gait-recognition technique that recovers static body and stride parameters of subjects as they walk is presented. This approach is an example of an activity-specific biometric*: a method of extracting identifying properties of an individual or of an individual's behavior that is applicable only when a person is performing that specific action. To evaluate our parameters, we derive an expected confusion metric — related to mutual information — as opposed to reporting a percent correct with a limited database. This metric predicts how well a given feature vector will filter identity in a large population. We test the utility of a variety of body and stride parameters recovered in different viewing conditions on a database consisting of 15 to 20 subjects walking at both an angled and frontal-parallel view with respect to the camera, both indoors and out. We also analyze motion-capture data of the subjects to discover whether confusion in the parameters is inherently a physical or a visual measurement error property.*

1. Introduction

Gait recognition is a subfield of biometrics [8] and has the advantage (over other biometrics) of being unobtrusive because body-invasive sensing 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. Most other modalities require proximal sensing, making it difficult to apply unobserved and to many people. Furthermore, humans exhibit the capability of recognizing people from impoverished displays of gait [10, 3, 14] indicating the presence of identity information.

In this paper we develop a gait recognition (or verification) method based upon static body and stride parameters measured during walking. We first discuss some previous work and describe several general deficiencies in those efforts. We then detail our approach to address those concerns.

1.1. Previous work

Previous work in automatic gait recognition from visual measurements can be divided, roughly, into model-free and model-based approaches. By model-free we mean there is no underlying representation of the three-dimensional structure of walking; however, they do have an implicit model of walking built into their methods of extracting features.

Model-free approaches [7, 11, 12, 1] analyze the motion or shape subjects make as they walk, and the features recovered from the motion or shape are used for recognition. Model-based techniques either model the person [13, 9] or explicitly model the walk of the person as it will appear in the imagery [2]. 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, or limb lengths) 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.

Most of the approaches taken to date suffer from three key deficiencies that we seek to address. One is the lack of generality of viewing condition – noting the exceptions of [9, 1]. Second, most researchers do not consider whether the confusions that arise are caused by vision yielding noisy measurements or by having chosen features with low discrimination power.

Perhaps the most significant limitation in most previous work in gait recognition is the manner in which results are reported. Even though the size of the database is typically less than ten people (sometimes as few as six), results are reported as percent correct. 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 four fundamentally different steps than previous researchers.

0-7695-1272-0/01 $10.00 © 2001 IEEE

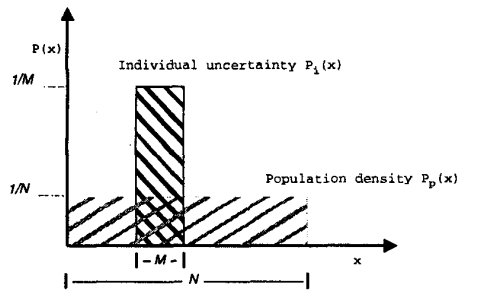


Figure 1: Uniform probability illustration of how the density of the overall population compares to the the individual uncertainty after the measurement is taken. In this case the remaining confusion — the percentage of the population that could have given rise to the measurement — is M/N.

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*. Second, as opposed to reporting percent correct, we will establish the reduction in uncertainty of identity that occurs when a particular measurement is taken. Third, to determine whether resulting confusion is because of poor vision or poor choice of a distinguishing feature, we compare vision-based analysis to a similar measurement derived from motion-capture data. Because our static measurements are based upon physical properties of the person and the behavior, we can compute those properties directly from three-dimensional limb position data.

Lastly, we present an *ad hoc* cross-condition mapping method that allows for the identification of a walking subject viewed under conditions that are different than those at which their initial data were recorded. Here we mean *ad hoc* in the literal sense: designed for the specific situation.

2. Expected Confusion

To evaluate a particular biometric, our goal is to characterize a particular measurement as to how much it reduces the uncertainty of identity after the measurement is taken. Such a measure would be much less sensitive to database size than reporting percent correct.

Information theory argues for computing the mutual information between the uncertainty in identity before the biometric is measured versus the uncertainty after measurement. In this section we derive an expected confusion measure — related to mutual information — that reflects how good a filter a particular biometric measurement is.

We motivate the measure by way of a simple example illustrated in Figure 1. Using a uniform density for illustration we let the density of a feature over the entire population P_p be $1/N$ in the interval $[0, N]$. Further assume that the individual density P_i is much narrower, being uniform in $[x_0 - M/2, x_0 + M/2]$. If we assume uniform priors on identity then 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.

When the densities are not uniform, the situation is a little more complicated. We still need the area under P_p that falls under P_i, but that area needs to be weighted by the value of P_i. Consider a little strip Δx under both P_p and P_i. The incremental area in that strip (with respect to P_p) in the neighborhood of x is just $P_p(x) \cdot \Delta x$. Suppose we have taken a measurement of value η (a slight abuse of notation from above). If we assume that P_i is zero mean noise about the measurement, then the absolute scaled incremental area ΔA is

$$\Delta A^*(\eta, x) = P_p(x)\, P_i(\eta - x)\, \Delta x,$$

and the total absolute scaled area A^* is

$$A^*(\eta) = \int_{-\infty}^{\infty} P_p(x)\, P_i(\eta - x)\, dx.$$

The reason we say "absolute scaled" is that what is needed is not the scaled area, but the *relative* scaled area. This is understood by looking at the one dimensional uniform densities in Figure 1. Each little strip of P_p that falls under P_i should not be weighted by $1/M$. If we did, we would end up with a value for the scaled area of $1/N$. That is, the scale factor would have been used to compute the average value of P_p over P_i. The problem is that the desired relative area A has been scaled and we need to normalize by the average scale factor. The average scale factor $\bar{s}$ is the expected value of P_i taken over P_i:

$$\bar{s} = \int_{-\infty}^{\infty} P_i(x) \cdot P_i(x) dx = \int_{-\infty}^{\infty} (P_i(x))^2 dx.$$

This yields the expression for the relative weighted area when the measurement is η:

$$A(\eta) = \frac{A^*(\eta)}{\bar{s}} = \frac{\int_{-\infty}^{\infty} P_p(x)\, P_i(\eta - x)\, dx}{\int_{-\infty}^{\infty} (P_i(x))^2 dx}. \quad (1)$$

Applying this formula to our uniform density example gives $\frac{1/N}{1/M} = M/N$ as expected. To compute the expected value of the confusion, we need to take the expectation over the population:

$$E[A(x)] = \int_{-\infty}^{\infty} A(x) P_p(x) dx. \quad (2)$$

We can analytically apply this measure of expected confusion to the Gaussian case as well. We derive the expression for the one-dimensional case; the multidimensional case follows naturally. Assume that $P_p(x)$ and $P_i(x)$ are

normal densities $N(\mu_p, \sigma_p^2)$ and $N(\mu_i, \sigma_i^2)$, respectively, where $\sigma_p \gg \sigma_i$, and μ_i is the mean of the individual. Again considering a measurement of η, the value for A^* is

$$A^*(\eta) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_p} e^{\frac{(x-\mu_p)^2}{2\sigma_p^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(x-\eta)^2}{2\sigma_i^2}} dx.$$

Given that $\sigma_p \gg \sigma_i$ we can assume that over the region in which P_i is significantly non-zero, the value of P_p is constant, in this case namely $P_p(\eta)$. If so, then

$$\begin{aligned} A^*(\eta) &= P_p(\eta) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(x-\eta)^2}{2\sigma_i^2}} dx \\ &= P_p(\eta). \end{aligned}$$

The average scale factor $\bar{s}$ is

$$\bar{s} = \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(x-\eta)^2}{2\sigma_i^2}} \right]^2 dx.$$

Using a change of variable of $\gamma_i^2 = \sigma_i^2/2$ yields

$$\begin{aligned} \bar{s} &= \frac{1}{\sqrt{2\pi}\sigma_i \cdot \sqrt{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\gamma_i} e^{\frac{(x-\eta)^2}{2\gamma_i^2}} dx \\ &= \frac{1}{2\sqrt{\pi}\sigma_i}. \end{aligned} \quad (3)$$

This yields a confusion at η for the Gaussian case of

$$\begin{aligned} A(\eta) &= \frac{A^*(\eta)}{\bar{s}} = \frac{P_p(\eta)}{\frac{1}{2\sqrt{\pi}\sigma_i}} \\ &= 2\sqrt{\pi}\sigma_i P_p(\eta). \end{aligned}$$

Finally, to compute the expected confusion we integrate over P_p:

$$\begin{aligned} E[A] &= \int_{-\infty}^{\infty} 2\sqrt{\pi}\sigma_i P_p(\eta) \cdot P_p(\eta) d\eta \\ &= 2\sqrt{\pi}\sigma_i \int_{-\infty}^{\infty} (P_p(\eta))^2 d\eta \\ &= 2\sqrt{\pi}\sigma_i \cdot \frac{1}{2\sqrt{\pi}\sigma_p} \\ &= \frac{\sigma_i}{\sigma_p}. \end{aligned} \quad (4)$$

This satisfying result simply states that the percentage of the population remaining is the ratio of standard deviation of the uncertainty after measurement to that before the measurement is taken. The relation to mutual information is also clear. Since the mutual information between the two Gaussian densities, in this case, is simply $\ln \frac{\sigma_p}{\sigma_i}$, our expected confusion measure is just the exponentiation of the negative of the mutual information. However, our measure is a relative one and can more easily be interpreted as a percentage of overlap and thus is perhaps more directly applicable to the field of biometrics.

Also, in concept, our expected confusion measure is related to the area under a receiver operating characteristic (ROC) curve. Medical researchers use the area under a ROC curve to determine a screening test's ability to correctly discriminate between those with and without a particular disease [4].

Extending the expected confusion measure to the multidimensional Gaussian case, the result is

$$\text{Expected Confusion} = \frac{|\Sigma_i|^{1/2}}{|\Sigma_p|^{1/2}}. \quad (5)$$

This quantity is the ratio of the volumes of equal probability hyper-ellipsoids as defined by the Gaussian densities. Essentially, the expected confusion can be defined as the ratio of the individual variation of a feature vector to that of the variation of the same feature vector over an entire population. This analysis also holds for densities that are mixtures of Gaussians.

3. Static Measurements from Gait

In this section we present two sets of activity-specific static body parameters designed for gait. As described earlier, activity specific implies that these parameters are defined only when a person is executing the particular movements associated with the activity. For example, in gait, we may consider the distance between the feet during the double-support phase — the time when both feet are on the ground. This is a *static* physical quantity determined by body geometry and the actual gait of the person, but it can be measured only when someone is performing an activity, in this case walking.

3.1. Gait parameters

The first set of static body parameters our technique measures, as a person walks, are four distances: the vertical distance between the head and foot (d_1), the distance between the head and pelvis (d_2), the distance between the foot and pelvis (d_3), and the distance between the left foot and right foot (d_4). The second set of parameters, a subset of the first and less discriminating, but less sensitive to error introduced by variation in viewing conditions, are d_1 and d_3 (See Figure 2). These distances are measured only at the maximal separation point of the feet during the double-support phase of the gait cycle and are concatenated to form a four-dimensional walk vector $\mathbf{w} =< d_1, d_2, d_3, d_4 >$ and a two-dimensional walk vector $\mathbf{s} =< d_1, d_3 >$ for each subject.

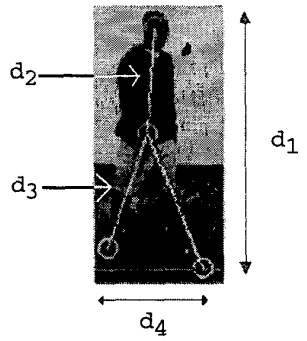


Figure 2: The static body parameters: $\mathbf{w} =< d_1, d_2, d_3, d_4 >$ and $\mathbf{s} =< d_1, d_3 >$.

3.2. Motion-capture data

When evaluating static body parameters in an activity as a biometric assay, there are two potential sources of confusion. Here we consider the question of how discriminating the feature is with respect to the population, *even if the measurement could be made exactly.* For example, if people vary their stride considerably as they walk, and if that variation is significant with respect to the entire variation across the population, then stride length would be a poor biometric feature even if it could be measured error free. In these experiments, we use three-dimensional motion capture data to evaluate the intrinsic discriminability of our measures.

Our motion-capture system uses magnetic sensors to capture the three-dimensional position and orientation of the limbs of the subjects as they walk along a platform. Sixteen sensors in all are used for the head, torso, pelvis, hands, forearms, upper-arms, thighs, calves, and feet. For this experiment, we recorded 20 subjects (11 males and 9 females with heights varying from 149.9 cm to 185.4 cm) walking along a 5 meter long platform. The data obtained from the motion-capture system were used to calculate the walk vectors $\mathbf{w}$ and $\mathbf{s}$ at the maximal separation points of the feet. These points were determined automatically from the displacement of the two foot sensors.

For the motion-capture experiment, the features are computed directly from the distances between the appropriate sensors. Each of the 20 subjects was captured walking across the platform six times; all measurements taken during the double-support phases of a given trial were robustly averaged to generate the two walk vectors $\mathbf{w}$ and $\mathbf{s}$ per trial, yielding a total of 120 instances of both $\mathbf{w}$ and $\mathbf{s}$.

3.3. Population vs. individual variation

Following the analysis of Section 2, we want to measure the ratio between the volume of the individual variation density and that of the overall population. Because of a limited number of subjects, we model the density of the population as a single Gaussian. To determine whether our population density estimation is valid, we plot the value of $|\Sigma_p|^{\frac{1}{2}}$ as a function of the number of subjects (see Figure 3). The data point for k subjects was computed by taking 200 random sets of k subjects and computing the maximum likelihood estimate Gaussian density.[1] The fact that the volume asymptotes as we approach 18 subjects, for the static body parameters associated with walk vector $\mathbf{w}$, implies that we have a reasonable model of the population density, but of course more data is always better. The asymptotic value is 309 cm^4. The volume more obviously asymptotes as we approach 12 subjects for the static body parameters associated with walk vector $\mathbf{s}$. The asymptotic value is 39.7 cm^2. We take this asymptotic behavior to indicate that we have enough subjects to begin to estimate the variance of an entire population.

To compute the individual variation, we subtract the mean walk vector of each subject from each of their six trials and then compute the covariance Σ_i over all the trials. The value of $|\Sigma_i|^{\frac{1}{2}}$ for the first set of static body parameter is 1.3 cm^4. This yields an expected confusion ratio $E[A]$ of 0.0042. This implies that with *relatively perfect* sensor measurements, the variation in these static body parameters during an individual's gait would leave a confusion with an average of less than 1% of the population.

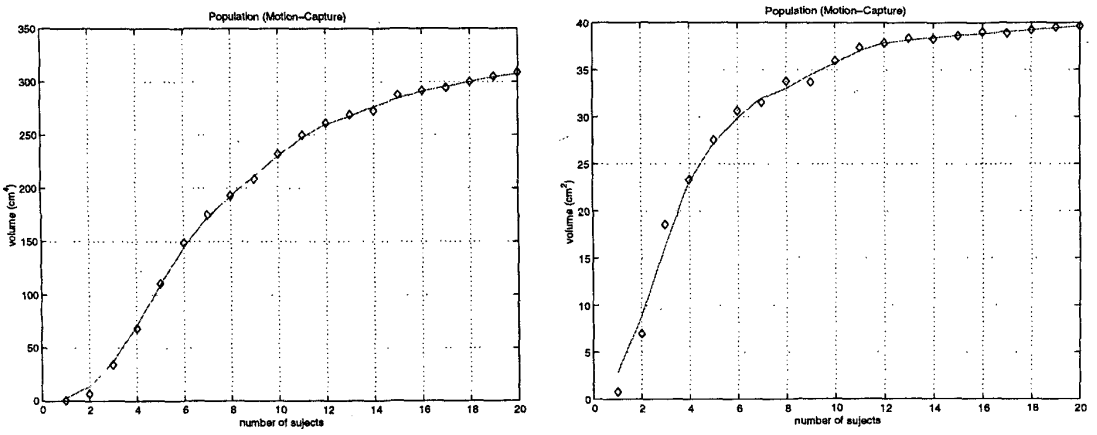
The value of $|\Sigma_i|^{\frac{1}{2}}$ for the second set of static body parameters is 1 cm^2. This yields an expected confusion ratio $E[A]$ of 0.025. Using this second set of parameters, the variation in these parameters of an individual would yield a confusion with 2.5% of the population. This set of parameters yields a higher confusion then set one, so the discrimination power will be less. However, as we will show in the next section, features that might be less discriminating in the ideal case may provide better actual performance because of noise in feature recovery.

We note that when performing recognition on our database of 20 people, our recognition rate was 97.5% for the first set of static body parameters and 86.7% for the second set. But clearly, this number is a function of how many people are in our database. The confusion measure, however, is valid as soon as the number of subjects allows the estimate of the population density to converge.

4. Vision-based Recovery

With the motion capture data as context, we now consider the same features but now as recovered from video imagery. Using a single camera with the viewing plane perpendicular to the ground plane, 18 subjects walked in an open indoor

[1] In cases where $\binom{n}{k}$ is less than 200, all possible sets of k were used.



(a) (b)

Figure 3: Volumetric measure of magnitude of probability density $|\Sigma_p|^{\frac{1}{2}}$ of the population using the motion-capture data. (a) For walk vector **w** the curve reaches a stable asymptote after approximately 18 subjects indicating a reasonable coverage of the population. (b) For walk vector **s** the asymptote occurs around 12 subjects.

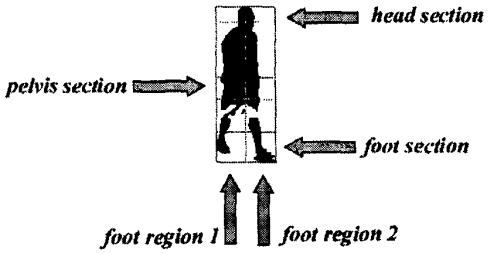


Figure 4: Automatic segmenting of the body silhouette into regions.

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 were captured at two different depths, 3.9 meters and 8.3 meters from the camera. Also, 15 of the 18 subjects were recorded in an open outdoor space in the presence of significant shadows at roughly a 55° path (angle-view-outdoors) toward the camera.

4.1. Body part labeling

Body parts are labeled by analyzing the binary silhouette of the subject in each video frame. The silhouettes are created by background subtraction using a static background frame followed by a series of morphological operations 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 4) — of predefined sizes similar to the body

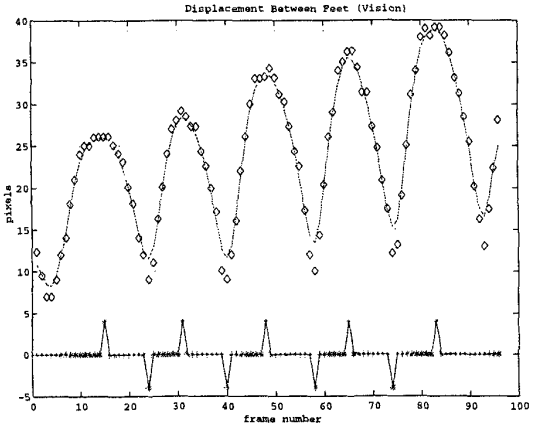


Figure 5: The distance between the two feet as measured in pixels. The value increases as the subject approaches the camera.

part labeling method in [5]. The head and pelvis are defined by the centroids of their respective regions. The foot section houses the lower legs and feet, and is further subdivided into foot region 1 and foot region 2. The pixel within each foot region that is furthest from the head is labeled as the position of the respective foot. For our features, we do not need to distinguish between left and right foot.

4.2. Depth compensation

The static body parameters used for identification are a set of distances between body part locations. Of course, distances measured in the image are in units of pixels. To convert from pixels to world units (e.g. centimeters) we construct a depth conversion factor as a function of the depth of the subject from the camera. To estimate this function,

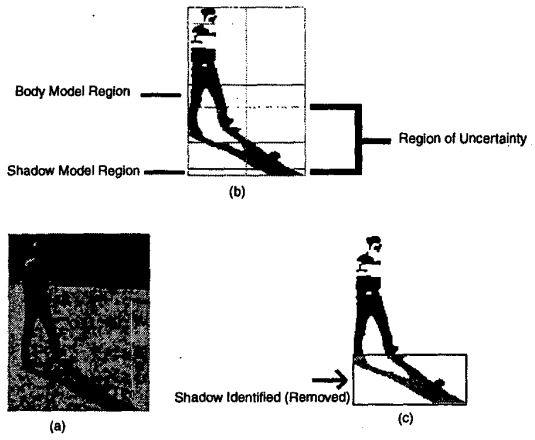


Figure 6: Outdoor image of a person walking where shadows are present. (a) Original image. (b) Background subtraction image with shadow. (c) Shadow separated from person.

we record a subject of known height walking at an angle to the camera. At the points of minimal separation of the subject's feet (see Figure 5), the system measures the height of the subject in pixels at that location on the ground plane. These minimal separation points represents the time instances when the subject is at his maximal height during the walking cycle. For each of these points a conversion factor from pixels to centimeters is recorded as a function of location on the ground plane (taken to be the lower y-value of the bounding box).

Assuming a world coordinate system located at the camera focal point and an image plane perpendicular to ground plane, and using perspective projection we can 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, 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. Using the recorded points and image heights we can explicitly estimate A and B.

4.3. Shadow removal

Shadows are a visual problem in outdoor scenes as they complicate background subtraction. We have implemented a shadow-removal method based upon two assumptions. Our first assumption is that we know the location of the camera with respect the sun's position, allowing our algorithm to know which general quadrant with respect to the subject contains the shadow.

Figure 6(a) gives an example of the camera being in front of the sun allowing the subject's shadow to be projected in front and to the side of the subject. Figure 6(b) is the binary silhouette of the subject after background subtraction; as expected, the shadow appears as part of the silhouette.

The goal now is to separate the shadow from the subject. This is accomplished by sampling the intensity and color characteristics of the shadow and the subject, and building two Gaussian densities, one for shadow and one for non-shadow. Our second assumption is that it is reasonable to represent the color density of each region by a single Gaussian.

The characteristic whose density is estimated is the brightness distortion [6] α which is the scalar brightness value of a pixel relative to the background color. It is found by finding the α that minimizes

$$\phi(\alpha_i) = ||\mathbf{I}_i - \alpha_i \mathbf{B}_i||, \tag{7}$$

where $\mathbf{I_i}$ is the current RGB color intensity of the image at location i, and $\mathbf{B}_i$ is the RGB color intensity of the background image (used for background subtraction). Since we know the general location of the shadow, the system can find the Shadow Model Region and the Body Model Region (see Figure 6(b)); these regions are defined as predetermined proportions of the bounding box around the subject and his shadow. Once the densities of each region are estimated, each pixel in the Region of Uncertainty (see Figure 6(b)), containing the lowers legs and the shadow, is classified as either belonging to the shadow density or the body density. Figure 6(c) shows the results of classification.

4.4. Static body parameters

After body labeling, depth compensation, and shadow removal (for outdoor data), walk vectors $\mathbf{w}$ and $\mathbf{s}$ are computed as (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; and (d_4) the distance between the left and right foot. As before, these distances are measured only when the subjects' feet are maximally spread during the walking action and are averaged over each instance.

4.5. Cross-condition mapping method

In previous work [9] we noted that static body parameters recovered from only a single view condition produce high discrimination power because the recovered measurements exist in the same parameter space as the training data. For example, when considering only one view angle, the effects of foreshortening do not need to be considered. With multiple viewpoints, variations in everything from foreshortening to how the body part labeling technique works in the

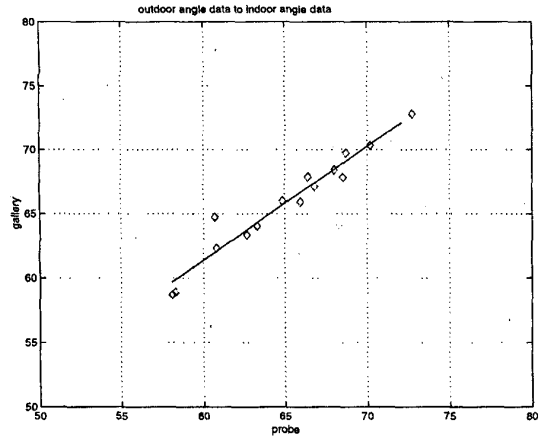


Figure 7: This figure shows the relationship between the d_1 static body parameter between indoor angle data as the gallery and outdoor angle data as the probe.

different views can lead to a systematic variation between the views. In our outdoor data we have errors introduced by shadows that are not present in the indoor footage.

To match between views we need to compensate for systematic differences between them. We assume that for a given set of viewing conditions, the same systematic error is being made for all subjects. Therefore, we can use a few subjects that span various body types as reference subjects and use their data (from different view angles and environments) to define a mapping function between viewing conditions. In the following experiments, we use indoor angle data as the gallery – meaning the viewing condition to which all other viewing conditions are warped. The other viewing conditions are referred to as probe.

We map between features from different views using linear regression. Figure 7 gives an example of the relationship of the d_1 parameter between indoor-angle data and outdoor-angle data.

4.6. Results

We recorded 18 subjects indoors, walking at the angle-view, far-side-view, and near-side-view. There are 6 data points 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). Also, we recorded 15 of the 18 subjects six months later outdoors in the presence of shadows for the angle-view-outdoors. There are six data points per subject for a total of 90 walk vectors.

Table 1 shows the within condition Expected Confusion for the two different walk vectors, **w** and **s**. The results for the angle view, side-view far, and side-view near indoors are similar in magnitude to those of the motion capture system (.42% and 2.5%). This reflects very little measurement noise in the indoor vision data. However the noisy environment of the angle-view outdoors data raises the expected confusion to 9.9% and 14.2% for **w** and **s**, respectively.

Table 1: Expected confusion for single viewing condition results.

Viewing Condition	*Expected Confusion* (**w**)	*Expected Confusion* (**s**)
Angle View indoors	1.53%	2.71%
Side View Far indoors	.71%	2.57%
Side View Near indoors	.43%	1.94%
Angle View outdoors	9.9%	14.2%

Table 2: Expected confusion for between viewing conditions using cross-view mapping.

Viewing Condition	*Expected Confusion* (**w**)	*Expected Confusion* (**s**)
Side View Far indoors	13%	9%
Side View Near indoors	37%	17%
Angle View outdoors	53%	31%

The results also shows that the expected confusion of the smaller subset of static body parameters **s** is higher than the larger set **w**. This is to be expected because the parameters of **w** incorporate more of the physical characteristics of the walking subject.

Table 2 shows results when using the mapping functions to map the probe measurements (side-view far indoors, side-view near indoors, and angle-view outdoors) into the indoor-angle parameter space. The expected confusion is computed by using the population density of the gallery to calculate the population covariance Σ_p. The individual variation is computed by subtracting the mean walk vector of each subject (from the gallery) from each of their probe walk vectors mapped into the gallery space and then the covariance Σ_i over all trials is computed.

The static body parameters **w** of the probe views yield higher confusion when compared with the gallery because these parameters are sensitive to foreshortening and other visual artifacts. The cross-condition mapping method is able to adjust for systemic variation between conditions; however, it is unable to compensate for random visual error in estimating the static body parameters. For example, the outdoor view-angle data yields a 53% confusion because of the shadows presence in the video footage of the walking subjects, which obscure the actual foot placements of the

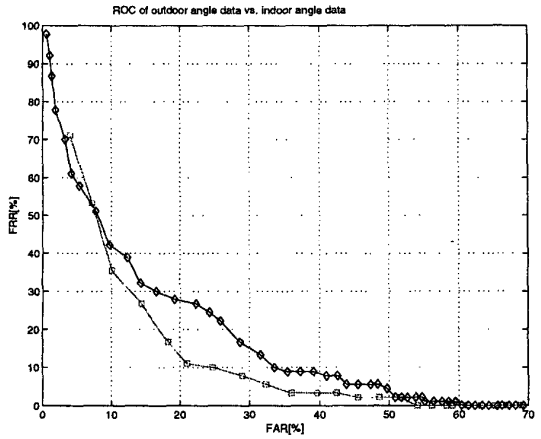


Figure 8: The ROC curves between outdoor angle-view and indoor angle-view. The indoor angle-view is the gallery, and the outdoor angle-view is the probe. ◇ represents the ROC of **w**, and the other is the ROC of **s**.

subjects.

However, if we use the simpler subset of static body parameters **s**, we achieve lower confusion (e.g. 31% confusion for outdoor angle-view compared the previous 53% confusion) because these parameters are less sensitive to foreshortening and visual artifacts. This result highlights the fact that the optimal set of parameters that gives higher discrimination in the ideal case may not be the best when considering measurement quality.

For completeness, we use a traditional method of reporting results, ROC curves. We use the ROC curve (see Figure 8) to evaluate the discrimination between the indoor angle-view data (gallery) and outdoor angle-view data (probe). These two conditions are most interesting to compare because the data were obtain in extremely different environments and with a large displacement in time. We create our ROC curve by using a Gaussian density to modeling each individual and accepting a data point if it falls within a certain standard deviation. By varying the acceptance threshold (the standard deviation), we can spawn various combination pairs of false acceptance rates (FAR) and false rejection rates (FRR).

5. Conclusion

A gait-recognition method using recovered static body and stride parameters of subjects has been presented in this paper. Two set of parameters were presented, and the within and between condition discrimination power of each set were analyzed. Also, we have demonstrated a paradigm to developing and testing gait-recognition methods.

As with any new work, there are several next steps to be undertaken. We must expand our database to test the expected confusion's ability to predict performance over larger databases using our static parameters. More subjects, will also permit us to consider whether representing the population as a single Gaussian density is wise. Finally, experiments must be ran under more viewing conditions, so the error over other possible views can be characterized.

Acknowledgments

The authors acknowledge Rawesak Tanawongsuwan for assistance in processing motion-capture data, and Vivek Kwatra for various discussions. Funding for the research was supported in part by the DARPA HID Program, contract #F49620-00-1-0376.

References

[1] S. Carlsson. Recognizing walking people. In *ECCV00*, 2000.

[2] D. Cunado, M. S. Nixion, and J. N. Carter. Automatic gait recognition via model-based evidence gathering. *IEEE AutoID99*, Summint NJ, 1999.

[3] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9:353–356, 1977.

[4] J. A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 82(143):29–36, 1982.

[5] I. Haritaoglu, 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*, pages 222–227, April 1998.

[6] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. *Proc. IEEC ICCV'99 FRAME-RATE Workshop*, Kerkyra, Greece, September 1999.

[7] P.S. Huang, 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):93–100, 1999.

[8] A. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, Boston, 1999.

[9] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recogntion using static body parameters. *3rd International Conference on AUDIO- and VIDEO-BASED BIOMETRIC PERSON AUTHENTICATION*, pages 301–311, Halmstad, Sweden, June 6-8, 2001.

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

[11] J.J. Little and J.E. Boyd. Recognizing people by their gait: the shape of motion. *Videre*, 1(2), 1998. http://mitpress.mit.edu/e-journals/Videre/.

[12] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[13] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. *Proc. Computer Vision and Pattern Recognition*, pages 469–474, Seattle, 1994.

[14] S. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13, 1999.