

Gait Recognition from Time-normalized Joint-angle Trajectories in the Walking Plane

Rawesak Tanawongsuwan and Aaron Bobick

College of Computing, GVV Center,
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
{tee|afb}@cc.gatech.edu

Abstract

This paper demonstrates gait recognition using only the trajectories of lower body joint angles projected into the walking plane. For this work, we begin with the position of 3D markers as projected into the sagittal or walking plane. We show a simple method for estimating the planar offsets between the markers and the underlying skeleton and joints; given these offsets we compute the joint angle trajectories. To compensate for systematic temporal variations from one instance to the next — predominantly distance and speed of walk — we fix the number of footsteps and time-normalize the trajectories by a variance compensated time warping. We perform recognition on two walking databases of 18 people (over 150 walk instances) using simple nearest neighbor algorithm with Euclidean distance as a measurement criteria. We also use the expected confusion metric as a means to estimate how well joint-angle signals will perform in a larger population.

1. Introduction

There are numerous applications for computer vision that require a system to automatically identify people or at least verify their claimed identity. *Gait* is one of human characteristics that researchers are investigating and hope to use as a signature to recognize people. A walking pattern biometric is appealing from a surveillance standpoint because the data can be measured unobtrusively and from a distance.

Walking is a complex dynamic activity that involves many segments of the body moving and interacting with one another and with the environment. There is evidence from the psychology literature that when humans view degraded displays of gait, they are still able to extract some identity information. For example, experiments in [6, 10] show that people can indicate some identity from the subjects' movement even when viewing only a moving point-light display.

There are many properties of gait that might serve as recognition features. We can categorize them as static fea-

tures and dynamic features that evolve in time. Static features reflect instantaneous, geometry-based measurements such as stride length [e.g. [9]]. Dynamic measurements, in contrast, are sensitive to the temporal structure of the activity. In this work, we analyze only dynamic features, namely lower-body (hip and knee) joint-angle trajectories. Our goal is to determine the extent of identity information present in this type of data.

1.1. Approach

The approach we take here is to look at joint-angle trajectories derived from motion capture data. Our reason for considering motion capture data is that we are primarily interested in determining feasible methods for extracting identity information from the joint angle data. Because we are evaluating the efficacy of such an approach, we consider the ideal case where the joint angle data can be known as well as possible. This will enable the evaluation of particular methods of determining identity independent of the ability to recover such information.

Having said this, we note that in the approach we present we consider only the joint angles projected into the *sagittal* or walking plane. Our reasons for doing so are two fold. First, as will describe, such joint angles are more robust to derive from marker data than complete, full degree of freedom joint angle specifications. Second, the walking plane joint angles are likely to be the most easily determined joint angles for *visual* methods. Previous work on gait recognition that employ frontal parallel views are also motivated by such an assumption.

In the remainder of this paper, we first consider some previous work on the recovery of identity from gait. Next we describe the sagittal joint-angle estimation method we apply to the motion capture data. Given these joint-angle trajectories we devise a temporal normalization that will allow comparisons across subjects. Using a database of 18 people we evaluate the efficacy of attempting recognition from joint-angle data; we report not only percent correct

but also the ability of the data to filter the population for verification.

1.2. Previous work

In vision research, there is a growing number of efforts investigating whether machines can recognize people from gait. Example work includes appearance based approaches where the actual appearance of the motion is characterized [9, 5, 11, 8]. Alternatively, there is work that attempts to model the physical structure of a person, recover that structure, and then recognize the identity. For example, in [13], they detect gait patterns in a video sequence, fit a 2D skeleton of human body to the imagery, estimate lower-body joint angles from the model, and then perform recognition. They employ a small walking database and report reasonable results.

There is also relevant work in the computer animation field, including that of recovering underlying human skeletons from motion capture data [14, 16] and analyzing and adjusting characteristics of joint angle signals [4, 3, 2, 17].

2. Recovering normalized joint angle trajectories

Here we describe the basic motion capture mechanism, the conversion to sagittal-plane projected joint angles, and the temporal normalization that results in comparable joint angle trajectories.

2.1. Motion capture

Our data are collected using an Ascension electro-magnetic motion capture system. The user wears 16 sensors, each measuring the six degrees of freedom (DOFs) of position and orientation. We place one sensor on each of the ankles, lower legs, upper legs, hands, lower arms, upper arms, one on the head, one of the pelvis and 2 on the torso. Figure 1 is an example of a typical setup in our experiment and figure 2 shows the sensors configuration of the lower part of the body projected into the *sagittal* or walking plane. The data are captured at 30 Hz and directly recorded.

For the work reported here, we focus on the joint angles most likely to be recoverable from vision: the sagittal plane joint angles of the lower body. Continued efforts are ongoing to recover such joint-angle data from video (e.g. [1]) and much work in vision based gait recognition presumes frontal-parallel imagery.

2.2. Skeleton fitting

Because each sensor reports orientation directly, it is possible to estimate the joint angle between two connected limbs with internal angle between the two sensors spanning the



Figure 1: Experimental apparatus of our motion capture system.

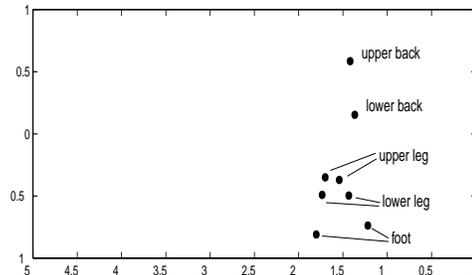


Figure 2: Markers considered in our experiment as viewed in the walking plane .

articulation. Such an estimate assumes that the sensors are parallel to the underlying skeleton.

In an initial study of recognition from joint-angle trajectories, we found that such estimated joint angles gave remarkably high recognition rates within a single database. In this database, each subject donned the motion capture equipment once and was recorded performing a number of walks across the platform. To further test the recognition method we captured an additional data set from a subset of the same subjects. The recognition rate plummeted. The difficulty was that there were biases in the data related to how the motion capture suit was worn. The high recognition rate arose from only using one recording session per subject. The recognition method exploited a bias in the signal caused by the actual placement of the markers on the subject.

This experience necessitated the estimation of joint angle trajectories that was as insensitive as possible to the details of the motion capture session. To accomplish this, we developed the following skeleton estimation method.

In [14], they present an algorithm for recovering a skeleton from magnetic motion capture data using both position and orientation information. The algorithm works well if data has only small noise and subjects exercise enough DOFs for each joint. Our data, however, did not fully exercise all the DOFs and therefore would induce numerical instabilities in such a skeleton recovery.

The above technique, as well as in [16], proceed by estimating one joint location at a time (two limbs and one joint) and then assuming the body joints are spherical. The basic

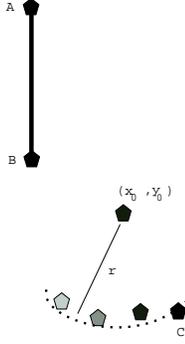


Figure 3: Stabilize two markers as a fixed axis and let the other marker move respect to that axis.

idea for calculating the joint location is to stabilize or fix one limb and let the other limb rotate. To stabilize the two body systems over time, we need to know at least the orientation of one body.

By operating in the sagittal plane, we eliminate both the issue of excessive DOFs, and the question of which articulated element to fix. First, by considering just position, it is straightforward to see that if there is little or no out-of-plane motion of a limb, then the relative movement of a limb is a circle centered about the joint to which the limb is attached. Second, people normally do not bend their backs while they walk. The back axis calculated from two sensors on the back then can be used in the stabilization process to find the hip joints, the angle between the femur and then back.

Figure 3 shows the case where we fix two sensors (A and B) on the back and let a sensor (C) on the upper leg move respect to those sensors. Using the trivial, two-dimensional planar version of the spherical equation in [16], we can solve for joint locations in the two-dimensional coordinate system defined by the back:

$$\arg \min_{r, x_0, y_0} \sum_{i=1}^N (\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} - r)^2 \quad (1)$$

Once we locate the hip joints, we can then propagate the results to find knee joints by stabilizing the estimated hip joint and the upper leg sensor and considering the trajectory of the lower leg.

Each time we propagate the results, errors accumulate. In our case we can estimate reasonably well the hip and knee joints (and therefore the angle at the hip between femur and back). The location of the ankle joint is difficult to determine because the motion of the foot about the ankle joint is small. Instead, we approximate the location of the ankle joint by the location of the foot marker. Because the distance between the foot marker and the true ankle joint is small with respect to the distance between the knee joint and the ankle joint, this approximation is acceptable. Fig-

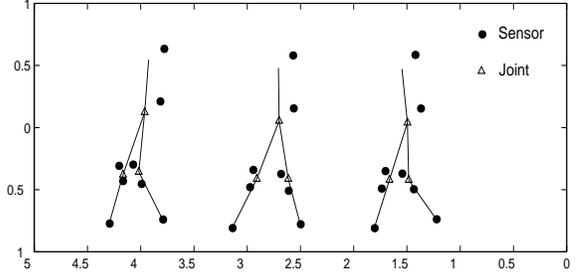


Figure 4: Sensor positions and recovered joint locations at 3 different time instants.

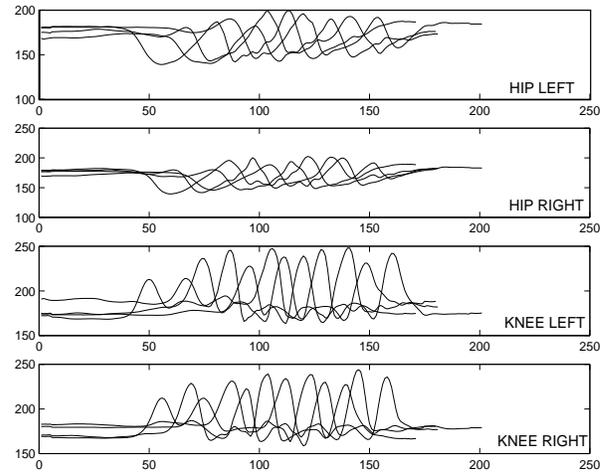


Figure 5: Joint-angle trajectories from several subjects.

ure 4 shows an estimated skeleton in a walk sequence at three different time instants.

2.3. Joint angle trajectories

Estimating an underlying skeleton enables us to measure the joint angle trajectories of four joints: left and right hips', left and right knees' angles. Figure 5 shows 4 joint signals from several walk instances. It is the variation in these joint signals that we wish to consider as information for identity.

Differences in body morphology and dynamics (e.g. height and strength) cause joint-angle trajectories to differ in both magnitude and time and also the number of foot steps taken to cover a fixed distance. To analyze these signals for recognition, we need to normalize them with respect to duration and foot step count (or walk cycles).

To standardize the number of foot steps, we use the difference in foot sensors displacement to segment different parts of walk cycles consistently across the subjects (see Figure 6). Typically peaks of the difference in the foot sensor displacement curve are easy to detect. We, therefore, choose to segment each walk using those peaks as boundaries. Since the magnetic motion capture system works well over a short range, we can select only one and a half walk-

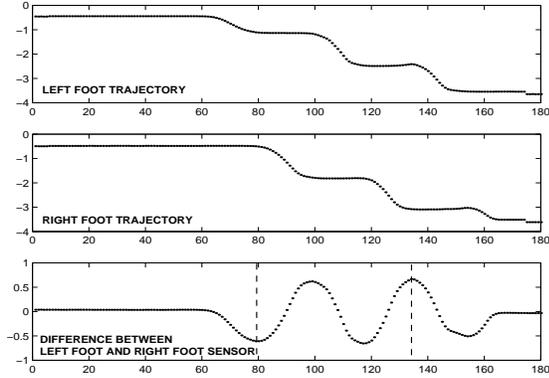


Figure 6: Left-foot and right-foot sensor x-coordinates and its difference over time.

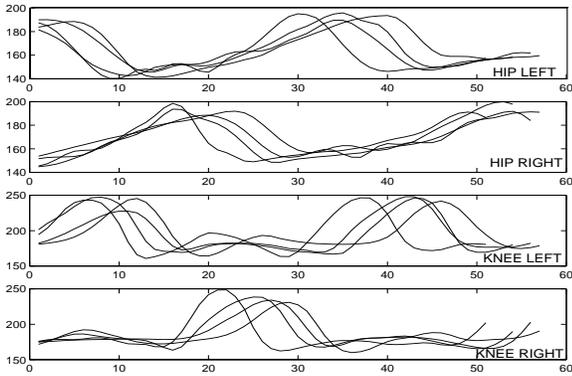


Figure 7: Joint-angles trajectories with same structure (same number of steps)

ing cycles from each of our subjects' walks. Figure 7 shows several segmented joint angle trajectories.

We now have the signals with the same structure (one and a half cycles). Time-normalization is needed for adjusting the signals to have the same length. Dynamic Time Warping (DTW) is a well-known technique to perform non-linear time alignment. [3] applied this technique to align joint angle trajectory in an animation application.

DTW works well for signals shifting, stretching, or shrinking in time, but not for shifting or changing in magnitude. Sometimes, the signals are noisy due to some systematic errors. Therefore, instead of using the joint angle trajectories in the warping process, we perform a variance normalization to reduce the effect of noise by subtracting the mean of each signal and then by dividing by the estimated standard deviation. This normalized signal has unit variance and can be more effective in performing matching in the dynamic time warping algorithm because all dimensions are weighted equally. Once we complete the DTW process on the variance-normalized data, we use the recov-

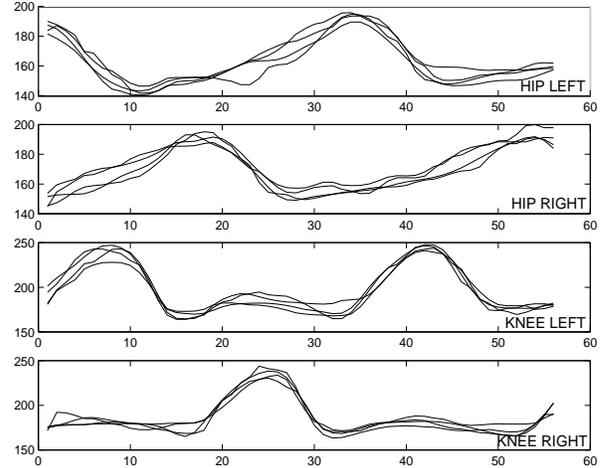


Figure 8: Time-normalized joint-angle trajectories

ered distortion map to align temporally the original (not variance compensated) signals to some arbitrary reference. Figure 8 shows the results after we time-normalize the original signals. It is the differences between these traces that represents the remaining identity information.

3. Gait recognition experiments

3.1. Data collection

We first captured 18 people walking with the magnetic motion capture suit. We attempted to place the sensors at similar body locations for all subjects. Each individual performs 6 walks of approximately same distance (4.0m). Before we use the 2D positions of the markers to estimate the joint locations, we filter the data using a technique similar to [17]. Taking all the data for each subject, we recover hip and knee joint locations. The underlying skeleton is connected through these joints and foot sensors and joint angle trajectories are recovered. We normalize the number of steps for all data to have one and a half walking cycles as shown in Figure 7. Then we perform time-normalization on unit-variance signals using DTW. We select randomly a walk sequence from the database to be a walking template and then we time-warp all the data to that template. After the signal normalization process, all the signals have the same footstep structure and same temporal length.

We also created another database by capturing walking data from 8 of the initial 18 subjects. This capture session took place months later than the first one. Again, each subject performs 6 walks. Skeleton recovery and temporal normalization were performed as before.

3.2. Recognition results

Our initial database has 106 time-normalized signals from 18 people. Using the nearest-neighbor technique, we can

Table 1: The recognition results using the nearest neighbor from both databases.

| Database | Recognition Results |
|---|---------------------|
| Database 1 (18 people, 106 walks) | 78/106 = 73% |
| Database 2 (8 people, 48 walks) match against database 1 | 20/48 = 42% |

Table 2: The expected confusion numbers on 4-subspace features.

| Database | Expected confusion number |
|-----------------------------------|---------------------------|
| Database 1 (18 people, 106 walks) | 0.097 |
| Database 2 (8 people, 48 walks) | 0.15 |
| Database 3 (8 people, 96 walks) | 0.27 |

count how many times we can recognize a particular walk in the database. For each walk in the database, we find the closest walk using direct Euclidean distance as a measurement. If the closest match comes from the same individual, we count as correctly classified. For those 106 walks, we correctly classify 78 walk instances, or 73% (table 1).

We also perform recognition of database 2 against database 1. In the parlance of the face recognition community, we use database 1 as the *gallery* and each trial of database 2 as the *probes* [12]. For each walk in the second database, we find the nearest neighbor match from the first database, and consider it correct if it comes from the same subject. As show in Table 1, we classify correctly 20 out of 48 probes, or 42%.

Comparing to a random guess result — 1 out of 18 or 6% — the recognition results indicate that there is indeed identity information in the joint-angle trajectories that can be exploited in recognition tasks. The lower recognition rate in our second experiment is the combined effect of an artificially high result caused by session bias when using only one database, and the noise in the joint recovery algorithm in general. Such difficulties suggest that merely reporting recognition results is inadequate, and that better measures of discrimination are required.

One such measure of performance proposed in the face recognition community is known as a "Cumulative Match Characteristic" (CMC) curve (e.g. [12]). This curve indicates the probability that the correct match is included in the top n matches. If we use the term *gallery* to mean the group of walking data in database 1 (of size m), and the term *probe* to mean an unknown walking data in database 2 to be compared to the entire gallery. In figure 9, the horizontal axis shows the top rank, n , or the n closest matches returned from a gallery of size m . In this case, we show only up to

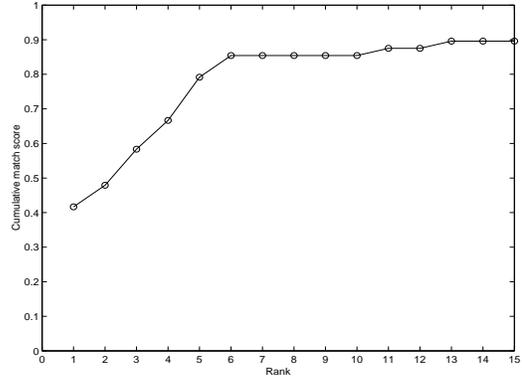


Figure 9: Cumulative Match Characteristic (CMC)

rank 15th. The vertical axis shows the probability that the correct match happens to be in the top n matches.

3.3. Expected confusion

Even though we have performed this experiment with 18 people and over 150 walks, the database size is still considered small. The recognition numbers usually depend on how many people and how many instances of them are in the database. Sometimes the numbers do not tell us much about the scalability or the expected performance of the features used in a larger database.

In [9], they introduce a new metric that predicts how well a given feature vector will filter identity in a larger database. The basic idea of this metric is to measure, on average, the ratio of the average individual variation of the feature vectors to the variation of that feature vector over the population. This ratio is shown to be equivalent to the average percentage of the population with which any given individual will be confused. The simple formula for computing this ratio is:

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

where Σ_i is the average covariance mean for all individuals in the database, and Σ_p is the covariance for the entire population in the database.

This metric is appropriate only in a small feature space where the probability densities can be reasonably estimated. In our work, the features are joint-angle signals of four joints (hips and knees). Usually the normalized signals that we used in the recognition lie within 2 seconds or 60 time samples. To calculate the nearest neighbor, we concatenate the four joint-angle signals into one long signal of dimensionality 240 for each walk instance. This yields a feature space too large to perform any probability density estimates.

To reduce the dimension of the data, we use principal component analysis (PCA). Analysis of variance indicates that the first four eigenvectors capture approximately 75%

of the variance, and adding more vectors does not improve the result significantly. Thus we projected the 240 dimensional vectors into that four-dimensional subspace and used the coefficients as the new feature space.

Using this reduced feature space we estimated the expected confusion components. For database 1 (18 people, 106 walks), the expected confusion number is 0.097. For database 2 (8 people, 48 walks), the number is 0.15. We also want to calculate the expected confusion number if we combine the two databases together. Since the number of subjects in both databases are not the same, we combine the data from the same 8 people and call it database 3. For database 3 (8 people, 96 walks), the expected confusion number is 0.27. The growth in expected confusion in this combined database is due to variation in data collection: the additional collection enlarged the variation of data for any given subject.

4. Summary and conclusions

This paper presents work of human gait recognition that analyzes joint-angle trajectories measured from normal walks. We choose to use a motion capture system as a tool to measure human movement because our goal is to assess the potential identity information contained in gait data. As many vision systems become more robust and can reliably track human limbs and recover actual joint angles, the recognition from visual data should reflect the results presented here.

We present a method for recovering human joint locations in the walking plane. Using those joints, we can recover the joint angle trajectories. By normalizing these signals so that they have the same structure (same number of steps) and the same duration, we show that the remaining variation contains significant identity information that can be exploited in recognition tasks.

Acknowledgments

Funding for the research was supported in part by the DARPA HID Program, contract #F49620-00-1-0376. We thank Amos Johnson for assistance with the motion capture apparatus and data.

References

- [1] C. Bregler and J. Malik, "Tracking people with twists and exponential maps", In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbera, 1998.
- [2] Armin Bruderlin Kenji Amaya and Tom Calvert, "Emotion from motion", In *Proceedings of Computer Animation (CA'96)*, 1996.
- [3] Armin Bruderlin and Lance Williams, "Motion signal processing", In *Computer Graphics Proceedings*, pages 97–104, August 1995.
- [4] M. Brand and A. Hertzmann, "Style machines", In *The Proc. of ACM SIGGRAPH 2000*, pages 183–192, 2000.
- [5] S. Carlsson, "Recognizing walking people", In *ECCV00*, 2000.
- [6] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues", In *Bulletin of the Psychonomic Society*, 9:353-356, 1977.
- [7] K. Halvorsen, A. Lundberg, M. Lesser, "A new method for estimating the axis of rotation and the center of rotation" In *J Biomech* 32 (11), 1221-1227, 1999.
- [8] P.S. Huang, C.J. Harris, and M.S. Nixon, "Human gait recognition in canonical space using temporal templates", In *IEEE Proceedings on Visual Image Signal Processing*, 146:93–100, 1999.
- [9] A.Y. Johnson and A.F. Bobick, "A Multi-View Method for Gait Recognition Using Static Body Parameters", *The 3rd International Conference on Audio- and Video- Based Biometric Person Authentication (2001)*.
- [10] L. Kozlowski and J. Cutting, "Recognizing the sex of a walker from a dynamic point-light display", In *Perception and Psychophysics*, 21:575-580, 1977.
- [11] J.J. Little and J.E. Boyd, "Recognizing people by their gait: the shape of motion", In *Videre*, 1(2), 1998.
- [12] H. Moon and P.J. Phillips, "Computational and performance aspects of PCA-based face-recognition algorithms", In *Perception 2001*, volume 30, number 3, pages 303-321.
- [13] S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in XYT", In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.
- [14] J.F. O'Brien, E.E. Bodenheimer, G.J. Brostow, J.K. Hodgins, "Automatic Joint Parameter Estimation from Magnetic Motion Capture Data", In *Proceedings of Graphics Interface 2000*, Montreal, Quebec, Canada, May 15-17, pp. 53-60.
- [15] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", In *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, Feb. 1978.
- [16] M. Silaghi and R. Plankers and R. Boulic and P. Fua and D. Thalmann, "Local and global skeleton fitting techniques for optical motion capture", In *Workshop on Modelling and Motion Capture Techniques for Virtual Environments*, Geneva, Switzerland, November 1998.
- [17] S. Sudarsky and D. House, "An Integrated Approach towards the Representation, Manipulation and Reuse of Pre-recorded Motion", In *Proceedings of Computer Animation'2000*, IEEE Computer Society Press.