Gait Recognition using Shadow Analysis

Yumi Iwashita
Graduate School of Information Science and Electrical Engineering
Kyushu University, Fukuoka, Japan

Adrian Stoica
Jet Propulsion Laboratory
California Institute of Technology, Pasadena, USA

Abstract

The exploitation of biometrics information in human shadow silhouettes (shadow biometrics), derived from video imagery after processing by gait analysis methods opens new avenues in remote biometrics. For the first time it becomes possible to obtain "overhead" biometrics, which may lead to recognition of human identity and behavior from high altitude airborne platforms using overhead video sequences. "Shadow biometrics" may use shadow information without body information, or in combination with it, as an additional perspective approximately equivalent to the use of a second camera. We took recordings and created a gait database in which both shadows and bodies are visible and used it to provide a first demonstration of the human gait recognition from shadow analysis. Only the information from shadows was used, which appears appropriate for overhead surveillance. We select a set of horizontal on the silhouette, for which we determine their length; this determines a set of varaibles in time, to which we apply spherical harmonics for each gait cycle. A k-nearest neighbor classification is applied to spherical harmonic coefficients. A subset of only 5 different subjects were used in this work to avoid biasing the results since we did not compensate for changing of sun position wiht time; the correct classification rate (CCR) was 95 %. In additional tests, we reduce spatial and temporal resolution of the images to 50 % each, which reduced the CCR to 75 %.

1. Introduction

The centimeter-level resolution of airborne sensing systems enables people surveillance. Although two individuals seen from above may appear indistinguishable, in particular if wearing similar head covers and robes, their shadow silhouette often offers a larger area and almost certainly more accurate body movement details head and shoulder top view. Shadows offer the connecting link between aerial observation and gait/biometrics/gesture and behavior classification. Shadow biometrics technology [1] (simply defined as biometrics using information from shadows) opens a new research area of applications in "overhead biometrics", which includes the remote observations from airborne/space platforms of biometric characteristics.

To illustrate these observations, Figure 1 shows a (Google-provided) image above a city. A rectangular area from the image is zoomed in, and after rotation and magnification shown in the enlarged window in Figure 2. What appears to be the shape of a human body is in fact the shape of its shadow, a body projection, while the actual body in top view is hard to distinguish and occupies only a minuscule area at the bottom of the shadow.

The paper that introduced the idea of shadow biometrics [1] outlined the generic processing steps for analysis but did no demonstrate it with real data. This paper provides the first demonstration of shadow biometrics, with recognition experiments on 5 subjects. The video/image processing greatly benefits from advances in two main areas: shadow detection/segmentation techniques, that allow extraction of the shadow silhouette, and gait analysis techniques, which extract the information from silhouette movements.

Significant progress has been achieved recently for human detection and identification (see comprehensive overview in [2]). Although there is a large diversity of gait recognition algorithms, a majority have focused on canonical (side) viewing point using silhouettes for human detection or identification, with several public databases available. For individual identification, correct classification rates based on image processing of gait video reaches 60-80 % depending on conditions of observation, with higher values (over 90 % for special conditions) reported for newer gait recognition algorithms [3]. Newer algorithms also effectively compensate for the hard covariates, such as surface, time, carrying condition, and walking speed, by normalizing the gait dynamics based on a population-based generic walking model [4]. Silhouette gait recognition approaches generally fall into two main categories [2]: (1) model-free shape-based analysis, and (2) model-based articulated or structural analysis. Shape-based analysis uses measurements of spatio-temporal features of the silhouette. To characterize shape and its variations different measures may including size (width, height, area), angles between lines (e.g. foot/ankle, upper arm-lower arm), higher order moments around the centroid, measures of symmetry, or other shape representations, and temporal variations such as cyclic oscillations at the stride frequency. Shape-based approaches have been shown very effective for human silhouette detection and have been used with good results on human identification. For studies on review on compact representations on human body shapes see [5]. For a baseline method based on body shape and gait see [6].

Articulated model-based approaches incorporate a human body model composed of rigid body parts interacting at joints. Parameters of the model may include kinematics such as link lengths, widths, and (more rarely used) dynamics such as moments of inertia. The model may be normalized to a standard body dimension or adapted with absolute coordinate measures, depending on application (e.g., for individual id, absolute measures are good discriminants; for behavioral id, body normalization may be preferred). This approach may be best suited for shadow biometrics, because the added complexities of viewing angle, sun angle and subject heading direction will likely require model-based estimation and tracking as feedback to reliably extract the desired features.

Structural model-based approaches include parameterization of gait dynamics, such as stride length, cadence, and stride speed. Static body parameters, such as the ratio of sizes of various body parts, can be considered in conjunction with these parameters. Traditionally, these approaches have not reported high performances on common databases, partly due to their need for 3D calibration information. However, this approach may prove more efficient for shadow analysis if multiple shadows are used in training.

Each category can be further segregated by inclusion of dynamics: temporal alignment-based vs. static parameter-based. The temporal alignment-based approach emphasizes both shape and dynamics. It treats the sequence as a time series and alignment of sequences of these features, corresponding to the given two sequences to be matched. The alignment process can be based on simple temporal correlation, dynamic time warping, Hidden Markov models, phase locked-loops, or Fourier analysis. Static parameter-based approaches emphasize the silhouette shape similarity and downplay temporal information. An image sequence can be transformed, for example using an averaged silhouette or silhouette feature, or treated as just a collection of silhouette shapes while disregarding the sequence ordering. The compromise approach in [4] uses stance specific representation, ignoring dynamics between stances, but still preserves the temporal ordering of the individual gait stances.

In [1] we proposed shadow biometrics and outlined the generic processing steps for analysis, however the approach was not demonstrated with real data. In this paper we use a specially created shadow gait database of 5 subjects, to illustrate the video processing necessary for recognition. We demonstrate the approach subjects, walking only in one direction, the extension for other directions being work in progress. Section 2 offers details about the database. Section 3 introduces the processing method. Section 4 presents the results. Section 5 is discussion and future work, while Section 6 offers concluding remarks.

2. Shadow gait database

A shadow database was created based on imagery recorded with a video camera positioned on an upper floor of a building. Twenty subjects were recorded while walking in different directions, in a sequence of walks that changed by 150 degrees when turning (in trigonometric sense, left u-turn), in a specific sequence that maximizes collecting data in various directions for a limited area (see Figure 2). For each person a normal walk and a fast walk were recorded. About 8 to 10 steps were walked in each direction. Turning was not imposed, and it varied from being abrupt at some to being a walk in small radius curve at others. In addition, a number of people were recorded at two times of the day (with different shadow lengths). The ground resolution is approximately 1cm per pixel. There are 30 frames per second.

Figure 2 (a) illustrates the set-up diagram and shows a photo of the real-world set-up, with white markers placed on the ground to indicate the star points (directions of walk and turning points). Figure 2 (b) illustrates different directions of walking.

3. Methodology

This section describes the methodology used in our work.

3.1. Segmentation

To extract shadow silhouettes from images in the database, firstly a background subtraction is used.

In our work we employ a method proposed by Tanaka et al. [7]. This method is based on the probability density function estimated by Parzen density estimation and the

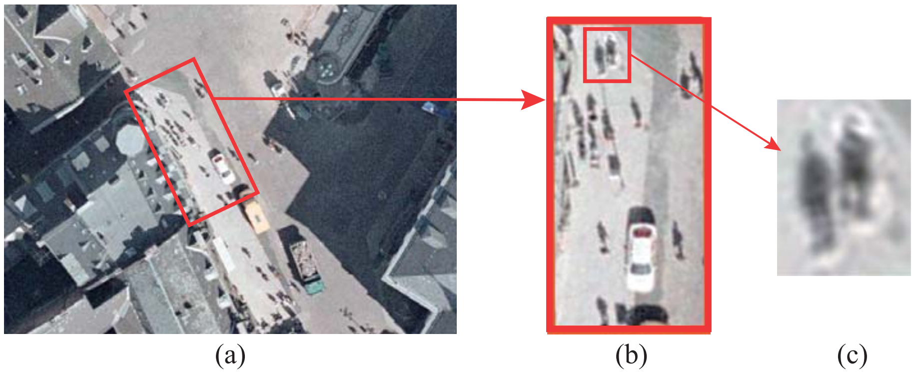

Figure 1. (a) Remote sensing imagery (Google), (b) magnified and rotated 32 degrees clockwise, (c) further magnification of smaller window on top. What looks like humans are, in fact, shadows. Heads and shoulders are small areas at the bottom (legs) of the shadows.

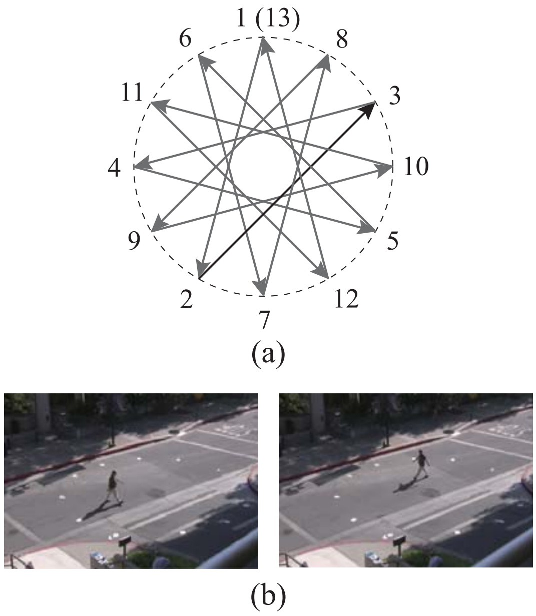

Figure 2. (a) To ensure coverage in database recordings of individual (shadows) walks in different directions and yet minimize area covered by experiment, a 12 point star marking on the ground was designed to guide the direction of individual walk, starting at 1, ending at the same point (13). Note that 2→3 and 8→9 are not identical although directions are parallel, the direction of the walk being opposite in the two cases (e.g. towards the sun, shadow from front lighting, and away from the sun, shadow from back lighting). Compared with walking in a circle the star arrangement offers more steps of walking in the same direction. (b) Image illustrating the real world set-up, and person walking, with shadow visible.

evaluation of the local texture at pixel-level resolution. The method performs robust object detection under varying illuminations. Figure 3 (b) shows an example of extracted silhouettes, from the original image shown in Figure 3 (a).

In this work we extracted the silhouette manually, separating the body and shadow regions (Figure 3 (c)); this should be performed automatically in future work).

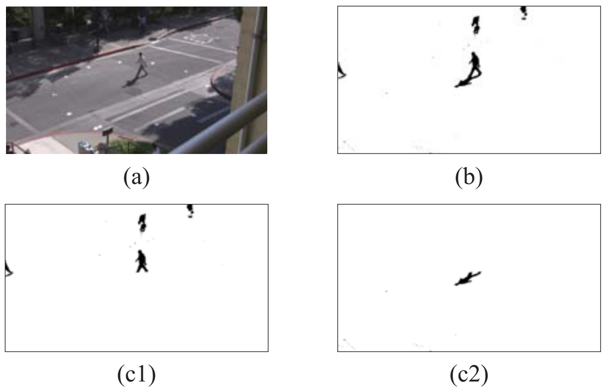

Figure 3. (a) A sample from the database, (b) its extracted target region, (c) body and shadow regions.

3.2. Compensations/scaling to a "standard" silhouette

At this point we would normally apply a transform to compensate for different shadow angles/sizes based on the information of the light source, position of person and observation camera, from know time of the day, inclination of the sun rays from time/position of the sun for given longitude/latitude, position of the platform, etc. (This would also allow to determine the height of the person.) (In these preliminary experiments we did not have yet implemented and thus we skipped the light source compensation step, with the consequence that we maintain a remanent information about the time of the day (in the angle of the shadow) which is in itself a marker of an individual - for individual recordings only at one time of the day. This will be further discussed in the experiment section.)

The image is scaled to a uniform height and aligned with respect to its horizontal centroid. Here, binary image $I(x, y, t)$ is indexed by pixel location (x, y) and time t. In this work we set the uniform height to be 100 pixels.

3.3. Extraction of spatial features in each image frame

After extracting image sequences of shadow silhouettes of one gait cycle, which is used to partition the sequences, dynamic features of gait sequence are obtained as follows: in each row of the image $I(x, y, t)$ the largest distance $H(y, t)$, which we refer here as "gait stripe", is determined as the absolute value of the distance between extremities of this shape, illustrated in Figure 4 (a) as a horizontal stripe measured from left to right.

3.4. Frequency analysis and classification

Since the gait motion repeats in time, the gait stripe $H(y, t)$ is also repetitive as shown in Figure 4 (b).

Here, one gait cycle is determined manually, and is a fundamental unit to describe the gait during ambulation, which occurs from the time when the heel of one foot strikes the ground to the time at which the same foot contacts the ground again.

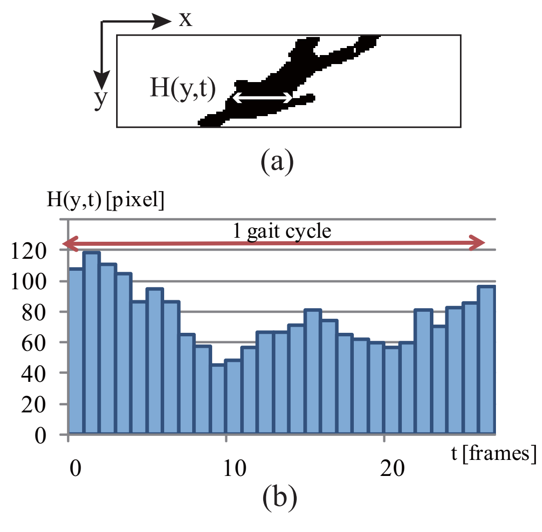

Figure 4. (a) Determination of $H(y, t)$, (b) An example of gait stripes ($H(y, t)$).

We decided to apply spherical harmonics transforms [8] because of its robustness to noise. Spherical harmonic transforms are essentially Fourier transforms on the sphere. Spherical harmonic functions $\{Y_l^m(\theta, \phi) : |m| \leq l \in \boldsymbol{N}\}$ are special functions defined on the unit sphere $\boldsymbol{S^2}$ [8] as:

$$Y_l^m(\theta, \phi) = \begin{cases} \sqrt{2}K_l^m cos(m\phi)P_l^m(cos\theta) & m > 0 \\ \sqrt{2}K_l^0 P_l^0(cos\theta) & m = 0 \\ \sqrt{2}K_l^m sin(-m\phi)P_l^{-m}(cos\theta) & m < 0 \end{cases} \tag{1}$$

where $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, K_l^m is a scaling factor defined $K_l^m = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}$, and P_l^m is the associated Legendre polynomial. The spherical harmonic functions are projected into spherical harmonic coefficients c_l^m as follows:

$$c_l^m = \int_S f(\theta, \phi)Y_l^m(\theta, \phi)ds. \tag{2}$$

Here, f is an original function, and its n-th order approximated function $\tilde{f}$ is reconstructed as

$$\tilde{f} = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta, \phi). \tag{3}$$

The n-th order approximation needs n^2 coefficients, and we define the coefficient c_l^m

$$c_l^m = c_i \quad \text{where} \quad i = l(l+1) + m. \tag{4}$$

For characterizing gait stripes, we obtain spherical harmonic coefficients of gait stripes $H(y, t)$ as features. To obtain spherical harmonic coefficients c_i of $H(y, t)$, we define the original function $f(\theta, \phi)$ as $f(\theta, \phi) = H(y, t)$ where $\theta = \frac{t}{T} \times \pi, \phi = \frac{y}{Y} \times 2\pi$. Here, T is the number of frames in one gait cycle and Y represents the height of images.

Although the extracted target region of the silhouette image may include noise, by using the spherical harmonics, features of each person can be obtained robustly against noise. In the proposed method, we extract spherical harmonic coefficients from gait stripes at first, and the classifier is trained by using training data sets. Then in the identification phase, the same spherical harmonic coefficients are extracted from gait stripes of test data sets, and the subject is identified by the classifier.

4. Experiments

In this section, firstly we show experimental results of gait recognition with gait silhouettes after applying the described methodology. Next, we explore the degradation in classification/recognition with the reduction in spatial and temporal resolution of the images, which allows us to tailor an observational platform to the needed/sufficient classification rate.

Finally, we used body silhouettes for gait recognition to show that the shadow could have gait information as much as body. In our experiments, we used 5 subjects walking in the direction from point 2 to point 3 (Fig 2 (a)), and the database contains 20 raw image sequences, which contain 5 different subjects with 4 sequences for every subject. As the classifier, the k-nearest neighbor classification was applied to the spherical harmonic coefficients. We used the leave-one-out cross validation to estimate the classification error rate.

4.1. Shadow only, full spatio-temporal resolution

In the first experiment, we used coefficients c_i of gait stripes H for classification. The number of features of each person depends on the order l of spherical harmonic functions, and Figure 5 shows obtained coefficients c_i by varying $l = [1, \cdots, 10]$ (the number of coefficients i varies $i = [1, \cdots, 100]$). Figure 6 represents the correct classification rate with respect to order of coefficients. The correct classification rate exceeded 95.0 % for more than 49 coefficients (l=7).

4.2. Shadow only, reduced spatio-temporal resolution

In the next experiment, we changed the resolution both spatially and temporally for aerial based images. The classifier is trained by using the training data sets of full resolution. For the test data sets, we decreased the spatial and temporal resolution to quarter and half, respectively. Table 1 shows the result of classification for coefficients. The CCR at half spatial and half temporal resolution shows 75.0 %.

4.3. Recognition from body silhouette only, and comparison in classifications from body only and shadow only

In the final experiment, we used body silhouettes and changed the resolution both spatially and temporally by the same procedure with the second experiment. Table 2 shows the result of classification for coefficients. Almost all results of body silhouettes are the same with shadow silhouettes, and we could say that the shadow has as much information as the body.

5. Discussion and Future Work

The results clearly indicate the feasibility of gait recognition based on shadow analysis. While more work is needed

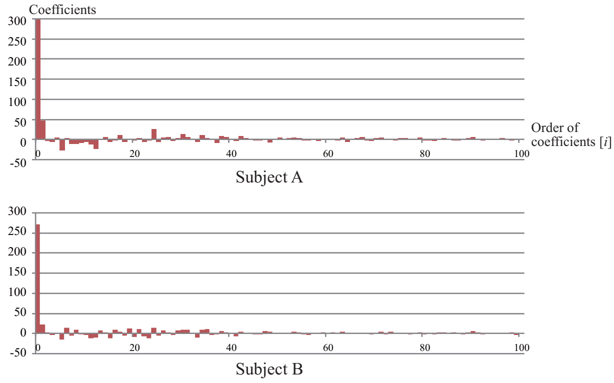

Figure 5. Examples of spherical harmonic coefficients.

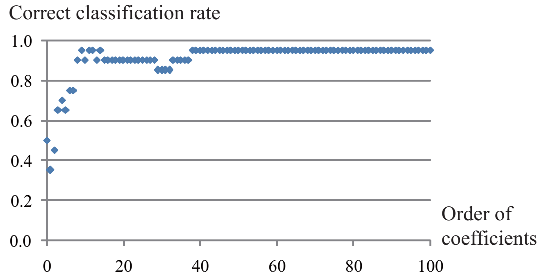

Figure 6. Correct classification rate with respect to the order of the spherical harmonic coefficients.

for determining CCR in a more general context, preliminary results offer a reasonable idea about needed resolution and how it degrades with decreased image resolution (spatial, temporal). This is of immediate interest since it leads to specifications for an airborne platform that would do the surveillance and shadow-based gait recognition. The lower the resolution the higher the platform can be placed, or the cheaper the equipment used for a given altitude. Two near term directions of work are:

1. Implement the equations that provide a corrective scaling with different positions of the sun. These equations are known and involve simple trigonometry. Forestry is the main field that studied the problem in the context of and determination of trees height from shadows (see for example the equations in [9]). In addition corrections may be needed to account for the position of recording camera (e.g. position of aerial platform, such as unmanned aerial vehicle, etc).

2. Study the robustness to changed direction of walk. This includes studies on how CCR degradations with different angles for different cases of training in one or multiple directions. We would look at both robust methods that provide highest CCR on many directions

Table 1. Correct classification rate [%] of shadow silhouettes with respect to spatial and temporal resolution. T is the number of frames in one gait cycle.

	Image size (width, height)		
	(400, 100) (100 %)	(200, 50) (50 %)	(100, 25) (25 %)
T [frames]	95.0	75.0	50.0
$T/2$ [frames]	85.0	75.0	35.0
$T/4$ [frames]	55.0	55.0	30.0

Table 2. Correct classification rate [%] of body silhouettes with respect to spatial and temporal resolution. T is the number of frames in one gait cycle.

	Image size (width, height)		
	(400, 100) (100 %)	(200, 50) (50 %)	(100, 25) (25 %)
T [frames]	95.0	70.0	45.0
$T/2$ [frames]	80.0	75.0	35.0
$T/4$ [frames]	35.0	35.0	20.0

while being trained with only one, as well as providing high CCR when multi-directional training data is available.

It is perceived that shadow analysis from top view may offer similar capability of recognition as body motion analysis in level view. Furthermore, the addition of shadow analysis to body motion analysis in level view is expected to increase recognition beyond body-only motion analysis, and offers good promise for future work in ground-based surveillance.

6. Conclusion

This paper demonstrated for the first time gait recognition from shadow analysis. Shadows were extracted by segmentation from image sequences from a gait database with both shadows and bodies visible. Shadow silhouettes were scaled to the same height and a set of silhouette stripes were determined, for which the length was determined in each image. Spherical harmonics were applied to the stripes for each gait sequence, followed by a k-nearest neighbor classification to spherical harmonic coefficients; the leave-one out cross validation was used to estimate the classification rate. On a set of 5 different subjects walking in the same direction, at approximately the same time of the day (at a few minutes difference), 4 sequences per subject resulted in a reduced set database of 20 raw image sequences. A correct classification rate (CCR) of 95.0 % from 49 coefficients was obtained. A reduction of resolution to 50 % (from 1 to 2cm per pixel) reduced the CCR from 95.0 % to 75.0 %.

7 Acknowledgments

Adrian Stoica acknowledges the support from Michael Kuhlman, a student from Ranseelar Polytechnic Institute, who during his internship at JPL in 2008 helped in recordings and creation of the shadow database, and also provided useful comments related to shadow gait analysis.

References

[1] A. Stoica, *Towards Recognition of Humans and their behaviors from Space and Airborne Platforms: Extracting the Information in the Dynamics of Human Shadows*, Proc. the 2008 Bio-inspired, Learning and Intelligent Systems for Security, pp.125-128, 2008.

[2] M. S. Nixon, T. N. Tan, and R. Chellappar, *Human Identification Based on Gait*, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.

[3] Z. Liu, and S. Sarkar, *Improved Gait Recognition by Gait Dynamics Normalization*, IEEE Trans Pat Anal and Machine Intelligence, Vol 28, No 6, pp.863-876, June 2006.

[4] S. Lee, Y. Liu, and R. Collins, *Shape Variation-based Frieze Pattern for Robust Gait Recognition*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.1-8, 2007.

[5] A. Godil, *Advanced Human Body and Head Shape Representation and Analysis*, Springer, Lecture Notes in Computer Science, 2007.

[6] R. Collins, R. Gross, and J. Shi, *Silhouette-based Human Identification from Body Shape and Gait*, 5th Intl. Conf. on Automatic Face and Gesture Recognition, 2002.

[7] T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi, *Object Detection under Varying Illumination based on Adaptive Background Modeling Considering Spatial Locality*, Proc. the Third Pacific-Rim Symposium on Image and Video Technology, pp.645-656, 2009.

[8] R. Green, *Spherical Harmonic Lighting: The Gritty Details*, SCEA Research and Development, 2003.

[9] Tree Shadows, *http://www.aie.org.uk/resources/tree_shadows/aie_sunlight1.html*.