# Finding people by their shadows: aerial surveillance using body biometrics extracted from ground video

Yumi Iwashita (*1, *2), Adrian Stoica (*1), Ryo Kurazume (*2)
(*1) *Jet Propulsion Laboratory, USA*
(*2) *Kyushu University, Japan*
{Yumi.Iwashita, Adrian.Stoica}@jpl.nasa.gov

## Abstract

*In past work we proved that shadow analysis extends gait biometrics to aerial surveillance, enabling people recognition from aerial imagery. The classifiers used to recognize individuals by their gait were both trained and tested on features derived from people's shadows; this may prove difficult to apply in real-world applications, since video-recordings that include shadows of people of interest may only rarely be available. Videos showing the body of the person of interest are however more often available. In this paper gait/dynamics features from body movement are obtained from ground video, and the search for matching dynamics of shadows (corresponding to the body dynamics) takes place in aerial surveillance video. An example application scenario involves observation of people (body movement) using ground/city surveillance cameras, from which one extracts information used to initiate a wide-area search for corresponding shadows, using aerial platforms. Alternatively one can start with the aerial observation of the shadow of a suspect leaving an incident area; this triggers a city-wide search for matching body/gait biometrics observed with ground surveillance cameras. To illustrate the feasibility of this approach the paper introduces a method that compares contours of bodies in ground image frames and contours of shadows in aerial image frames, for which an alignment is made and a distance is calculated, integrated over a normalized gait cycle. While the results are preliminary, for only 5 people, and using a specific walking arrangement to avoid compensation for changes in the viewing angles, the method obtains a 70% correct classification rate, which is a first step in proving the feasibility of the approach.*

## 1. Introduction

Gait biometrics has shown significant promise as a remote biometrics (see a comprehensive overview in [1]). While the correct classification rate (CCR) of individual recognition depends on a large number of factors, it commonly offers a CCR in the range of 60% to 80%, over a range of factors of variability (ground surface type, shoes, state of etc); higher values of CCR (over 90% for special conditions) are reported for newer gait recognition algorithms [2]. Clearly the smaller the variability in the many dimensions influencing the gait, the higher the possible CCR.

Despite its proven advantage in remote biometrics, gait recognition has been limited to observation from ground surveillance cameras. A huge untapped potential is in overhead remote biometrics obtainable from aerial platforms deployed over areas of interest. This new avenue has recently been opened by the introduction of shadow biometrics [3], [4], which exploits the dynamics of shadows, performing gait recognition not on the body dynamics, but on shadow dynamics, as shadows offer more information on the body than it is seen directly looking at the body from overhead. Shadow biometrics thus allows for the first time the biometrics-based person recognition from aerial images.

This application domain is supported by the fact that aerial surveillance of moving vehicles and people, from aerostats to unmanned planes is becoming common practice in conflict areas or other places of high interest. Fig 1 illustrates a scene showing such a blimp carrying a range of surveillance sensors (notice also the shadows).

Figure 1: Aerial platforms are becoming common for area surveillance. Many regions of interest are in sunny areas, with good visibility of shadows.

978-0-7695-4791-6/12 $26.00 © 2012 IEEE
DOI 10.1109/EST.2012.41

The altitude of observational assets may vary from low to very high; in principle one can use not only high altitude planes but also orbiting satellites. Imaging capabilities of commercial satellites have reached 34 cm ground resolution, as seen from a 637Km altitude orbit (GeoEye2). By going lower, for same imaging system, this results in a ground resolution of 15cm for a lower orbit at ~ 300km. This calculation is only to illustrate the power of present day imagery; the more optimal surveillance platforms would be stationary above the area of interest, most likely up to 15-20km, and the high performance ones should be able to offer 1 cm ground resolution from that altitude.

Past work in shadow-based person recognition [3], [4] focused on classifiers both trained and tested on features of people shadows, extracted by image processing. The systems were Shadow-In, ID-Out (SIIO) classifiers.

Starting with video including shadows of the people to be found/ recognized may be difficult in practice. Most common imagery of people of interest, coming from video recordings with portable camcorders or from ground surveillance cameras, most likely contain lateral/front body images, and only in rare cases contain good shadows of the person.

One option is to synthetically generate shadows from sufficient 3D body information obtained from 2D video. Yet, more importantly, regardless of the means to exploit body information, an Air-Ground Synergistic Surveillance System (AG-S3) can be imagined, in which input information for a desired observation target is provided either in shadow or body form, and the wide-area search, aerial and ground-level, for a matching individual is performed (or, a tracking takes place for an observational target, without a need for knowledge of identity). This corresponds to SIIO, Body-In, ID-Out (BIIO), Shadow-in Track-Out (SITO).

The encompassing scenario is illustrated in Figure 2. Assume in first instance that gait/dynamics features of body movement are obtained from a ground video, and the search for matching dynamics of shadows (corresponding to the body dynamics) takes place in aerial surveillance video. A common scenario would be to observe/record people using ground/city surveillance cameras and obtain information to initiate a wide-area search for shadows using aerial platforms. In Fig 2 this is illustrated as obtaining information from imagery as shown in Fig. 2-Left image, showing body info, extracting characteristics and passing them to the aerial platform, that does a search for shadows in the overhead imagery of a wide area, with shadows as shown in Fig. 2-Right.

Vice-versa, having a record of the shadow of a suspect leaving an incident area, detected by aerial surveillance, for example as in Fig. 2-Right, can trigger a city-wide search for matching body/gait biometrics, in imagery observed with city/ground surveillance cameras, such as shown in Fig. 2-Left. These examples focus on obtaining target information in one modality (e.g. body) and performing the search in the other modality (e.g. shadow). Of course one can use both modalities in the search, a suspect's shadow in aerial imagery can be sought as a shadow in aerial imagery at the same time its body dynamics are matched in ground surveillance systems.

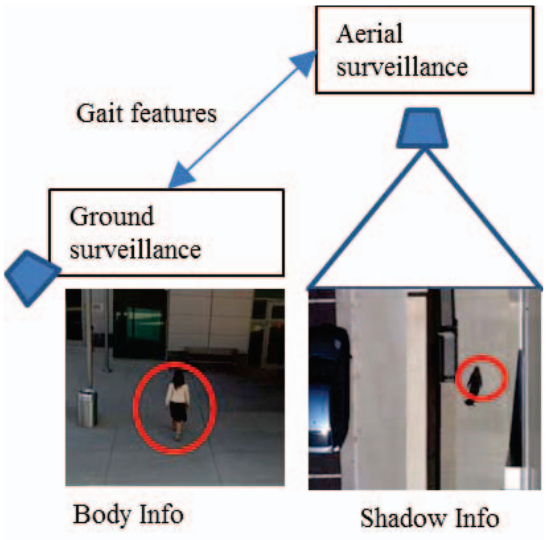


Figure 2: Schematic of the concept: body information obtained by ground surveillance leads to gait features communicate to aerial surveillance that initiates the search, or vice-versa

Figure 3: Photos suggesting similarities sought, although here in frames and not capturing dynamics of movement. The top shows two body postures, the bottom shows two, somehow similar, shadows.

## 2. Shadow database

In this section we introduce a shadow database. Two kinds of images were collected: (A) top view images (Fig. 4(a)), and (B) oblique view images (Fig. 4(b)). To collect top view images and oblique view images, we placed a video camera (Canon, iVIS HF S10) on a rooftop of a building so that it takes images of a subject from straight above and obliquely, respectively, as shown in Fig. 5. Subjects walked straight under the camera. Five people with 2 sequences for each top and oblique image were recorded. The image resolution was 1,920×1,080 and the frame rate was 30 Hz.

As shown in Fig. 4, a captured image contains both body and shadow areas, which can be considered to be captured from two different viewpoints - camera and Sun direction. In our method, we use shadow areas of top view images as training datasets and body areas of oblique view images as test datasets. In general, in case that the viewpoint of the test dataset is different from that of the training dataset, the correct classification rate decreases due to appearance change. Thus to capture body areas from the same viewpoint of the shadows, we calculated the position where subjects should walk based on the sun angle of top view image. The calculated distance is 7.5 m as shown in Fig. 5 (b).

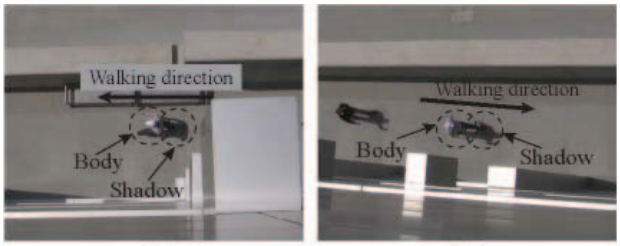


Figure 4: (a) Top view image, (b) side view image

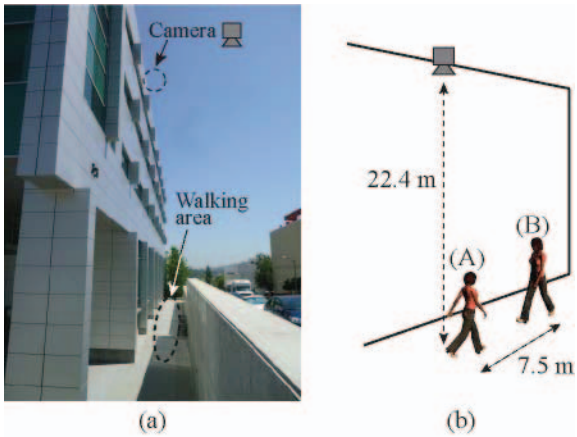


Figure 5: (a) A building for experiments,
(b) drawing illustrating the experimental setting

## 3. Methodology

Although captured images (Fig. 4) contain whole body and shadow areas, in our scenario we consider that shadow areas of top view images and body areas of oblique view images to be used for person identification while the rest of the areas are noise. In this section we propose a method for person identification which is robust to noise.

To summarize, the main steps of processing are as follows.

(Step 1) Extract subject's area by background subtraction, and extract the contour of the subject's area. This process is done for all images in the training datasets and test datasets.

(Step 2) A contour of a frame $F^P$ from a gait sequence $S^P$ in the test datasets is aligned to that of a corresponding frame $F^G$ from a gait sequence $S^G$ in the training datasets. The alignment is done by minimizing the error defined as sum of distances of two contours.

(Step 3) Calculate the similarity between two contours.

(Step 4) Step 2 and Step 3 are repeated for all frames of a sequence $S^P$, and the sum of similarities between a sequence $S^P$ and $S^G$ is calculated for person identification.

The following sections further explain the procedure and illustrate it with examples.

### 3.1. Extraction of contour and alignment

Figure 6 shows an example of extraction of a subject's area by background subtraction and its extracted contour. For fast alignment as introduced in [5], at first we build a distance map of a contour of a frame $F^G$. One of the popular techniques for alignment of two contours is the point-based method such as the ICP algorithm [7]. However, the calculation cost for determining point correspondences is expensive. To reduce the calculation cost, we build a distance map of a contour based on the fast marching method. For further details, please refer to [5].

After obtaining the distance map, a contour of a frame $F^P$ is superimposed in the distance map of a frame $F^G$. For alignment we rotate the test datasets with a certain degree, based on the subject's walking direction and the sun angle in advance (in this case 180 degree, i.e. Fig. 4(b) $\rightarrow$ Fig. 7(c)).

The force $f_k$ ($0 \le k < K$, K is the number of points on the contour of a frame $F^P$) is applied to the center of a contour of a frame $F^P$. The force $f_k$ is defined as follows:

$$f_k = DM_{i,j} \frac{\nabla DM_{i,j}}{\left|\nabla DM_{i,j}\right|} \quad (1)$$

where $DM_{i,j}$ is the value of the distance map at *(i, j)* ((*i, j*) is the position of a point *k*).

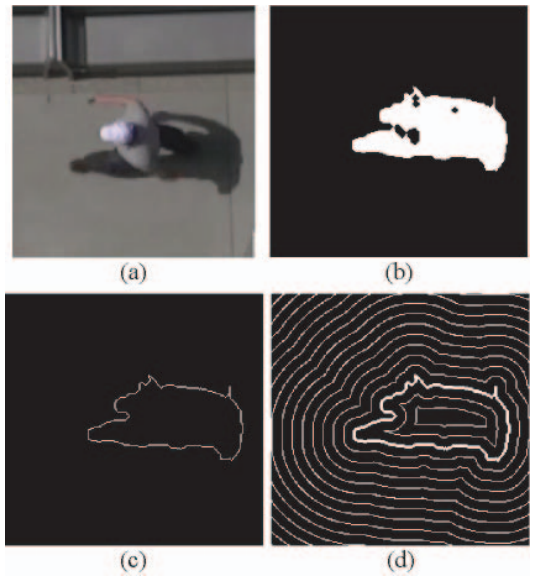


Figure 6: (a) An example captured top view images, (b) silhouette area, (c) contour, (d) distance map

The total force and moment around the COG of a contour of $F^P$ are calculated using the following equations as shown in Fig. 7 (e).

$$F = \sum_k \rho(f_k) \quad (2)$$

$$M = \sum_k \rho(r_k \times f_k) \quad (3)$$

where *r* is the vector from the COG of a contour of $F^P$ to a point on the contour and $\rho$ is a particular estimate function. The current position T and R are updated as follows;

$$T \leftarrow T + k_T F \quad (4)$$

$$R \leftarrow E_{k_R M} \cdot R \quad (5)$$

where $k_T$ and $k_R$ are constant gains and $E_\omega$ is the coordinate transformation matrix around the axis of ω. As we mentioned before, extracted subject area includes noise. Thus a part of the subject's contour including noise does not coincide the contour in the training dataset, and the correct distance value cannot be obtained. Therefore, we introduce the robust M-estimator to ignore contour points with a large amount of errors. The robust M-estimator is generally utilized to reduce the effect of outliers by the weight estimation.

Let's consider the force $f_k$ and the moment $r_k \times f_k$ as an error $\varepsilon_k$. We modify Eqs.(2) and (3) as

$$E(P) = \begin{pmatrix} F \\ M \end{pmatrix} = \sum_k \rho(\varepsilon_k) \quad (6)$$

where *P* is the pose of a contour of a frame $F^P$. The pose *P* which minimizes the *E(P)* is obtained by the steepest descent method. In our implementation, we adopt the Lorentzian function for the estimation function $\rho(z)$, and gradually minimize the error *E* in Eq.(6) by the steepest descent method.

$$\rho(z) = \frac{\sigma^2}{2} \log\left(1 + (z/\sigma)^2\right) \quad (7)$$

Figure 8 shows an example of alignment.

Yu *et al.* [5] proposed a method which reduced the effect of noise on a contour. This method is based on dynamic time warping, and for matching between two contours, a point from a contour which corresponds to a point from the other contour should be chosen. If these two points do not correspond each other, the effect of noise cannot be reduced. On the other hand, our method does not require to choose the corresponding points in advance, since our method aligns the two contours automatically.

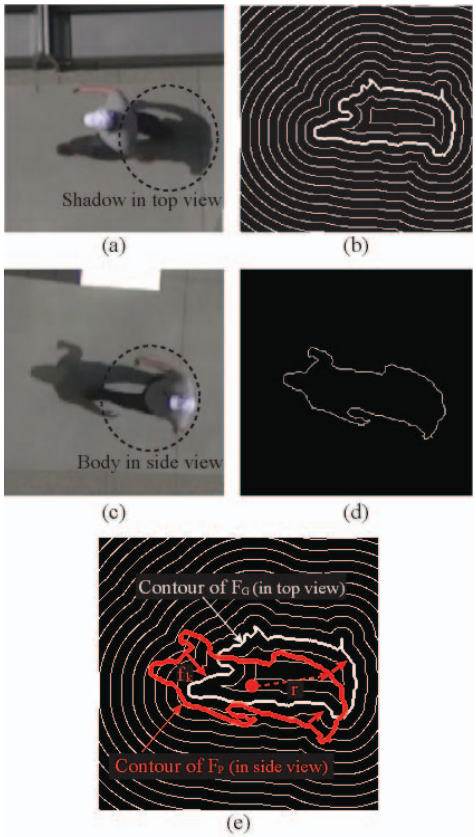


Figure 7: (a) a top view image ($F_G$), (b) distance map of (a), (c) a side view images ($F_P$), (d) a contour of (c), (e) Apply the force *f* to all points on contour of ($F_P$)

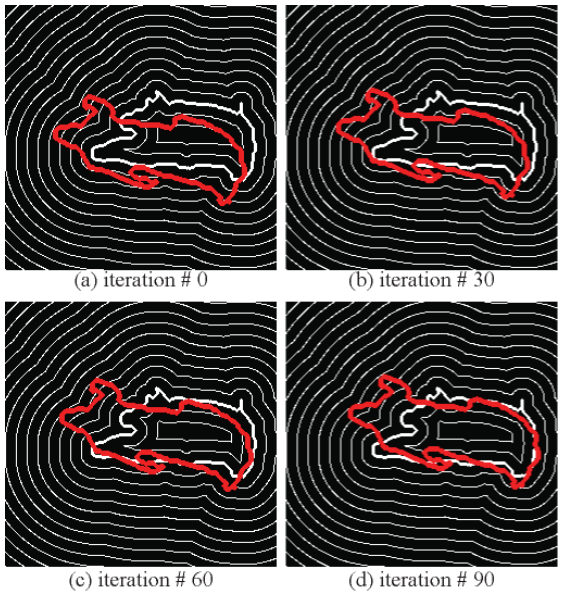


Figure 8: Alignment of contour of ($F_G$) and contour of ($F_P$)

### 3.2. Person identification

After the alignment, we calculate the similarity $s$ between a contour of a frame $F^P$ and that of $F^G$ as follows;

$$s(F^P, F^G) = \sum_k \frac{1}{1 + DM_{i,j}} \cos\theta_{i,j} \tag{8}$$

$$\cos\theta_{i,j} = \frac{\nabla DM_{i,j} \cdot \nabla DM'_{i,j}}{\left|\nabla DM_{i,j}\right| \left|\nabla DM'_{i,j}\right|} \tag{9}$$

where $DM'$ is a distance map constructed on a narrow area along a contour of a frame $F^P$. The reason why we introduce $\cos\theta$ is that after the alignment, in the area which does not contain noise (area A in Fig. 9) gradients of a point on a contour of $F^P$ and that on a contour of $F^G$ coincide, but in the area which contain noise (are B in Fig. 9) gradients of two points are different. To reduce the effect of noise area to the similarity, we introduce $\cos\theta$.

The sum of similarities $SeqS$ of a sequence $S^P$ and $S^G$ is calculated by

$$SeqS(S^P, S^G) = \sum_n S_n\left(F_n^P, F_m^G\right) \tag{10}$$

where $0 \le n < N$ (N is the total number of frames of a sequence $F^P$), and $m = \frac{M}{N} n$ (M is the total number of all frames of a sequence $F^G$). Here, a sequence consist of frames of one gait cycle, and we assume that the phase of the first frame of a sequence $S^P$ is the same with that of a sequence $S^G$.

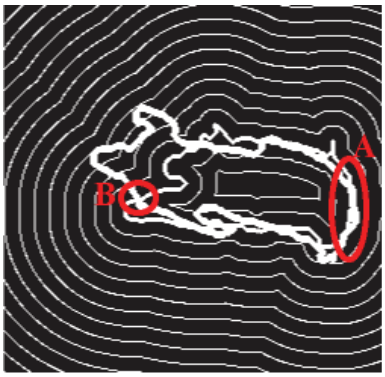


Figure 9: (A) Gradients of both contours coincide, (B) gradients are different

## 4. Experiments

Person identification experiments were performed using the shadow database. We use the top view images as the training datasets, and the oblique view images as the test datasets. As we mentioned in Section 3, we consider that shadow areas of top view images and body areas of oblique view images to be used for person identification. Both images include 5 subjects with 2 sequences for every subject, respectively.

The correct classification rate (CCR) was 70 %. We did experiments using Eq.(8) without $\cos\theta$, and CCR was 50 %. From these results, we can see the effectiveness of the $\cos\theta$. However, the CCR by the proposed method (Eq.(8)) is still low. One of reasons is that the shapes of two people at a certain frame are similar, and the proposed method uses only shape information but not dynamic information.

Future work includes the use of dynamic information. It will also deal with the compensation for the sun angle-performing a distortion of the shadow. Knowing the location on Earth and the time of the day one can know the exact position of the Sun and one can estimate the shadow from a model of the body inferred from 2D frames that allow a 3D body reconstruction.

## 5. Conclusions

In this paper we introduced an Air-Ground Synergistic Surveillance System (AG-S3) which allows a search for an individual to be performed starting either with its body dynamics information, or his shadow dynamics, i.e. either input from ground/city surveillance cameras or images

with shadows of the individual obtained from aerial imagery. Regardless the input, a search can be performed both by the aerial, wide-areas overhead surveillance system or by ground/city cameras.

We illustrated the feasibility of the approach by an implementation that compares the body shape of an input case with potential shadows, of which one identifies the one corresponding to the suspect. While the tests were performed only on 5 people and we used a position of the cameras that facilitates the match, the results are are promising and illustrate the feasibility of the approach. Future work will extend the current set-up to use a model-based approach that takes in consideration changes with the angle of the Sun and that of the direction of walking.

## Acknowledgment

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors thank Dr. Robert C. Stirbl, Program Manager for Navy, Marines & Other DoD Agencies at the Jet Propulsion Laboratory, for his support and encouragement, and Dr. Michael Ryoo for his feedback and suggestions related to this work.

## References

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

[2] 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.

[3] 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.

[4] Y. Iwashita and A. Stoica, Gait Recognition using Shadow Analysis 2009 Symposium on Bio-inspired Learning and Intelligent Systems for Security

[5] S. Yu, D. Tan, K. Huang, and T. Tan, "Reducing the Effect of Noise on Human Contour in Gait Recognition", *International Conference Advances in Biometrics 2007*, pp. 338-346.

[6] Y. Iwashita, R. Kurazume, K. Hara, and T. Hasegawa, "Fast Alignment of 3D Geometrical Models and 2D Color Images using 2D Distance Maps", 5th International Conference on 3-D Digital Imaging and Modeling (3DIM'05), 2005.

[7] P. Besl, and N. MacKay, "A method for registration of 3D shapes", Trans. PAMI, vol. 14, no. 2, 1992.