GAIT3: An Event-based, Visible and Thermal Database for Gait Recognition

Mohamed Jamel Eddine
Department of Digital Security
EURECOM
Sophia Antipolis, France
jameledd@eurecom.fr

Jean-Luc Dugelay
Department of Digital Security
EURECOM
Sophia Antipolis, France
jld@eurecom.fr

***Abstract*—Identifying people by their gait has gained popularity in the last twenty years. Recent gait recognition methods use acquisitions extracted from advanced sensors such as cameras, depth sensors, microphones, etc. Recently, event-based cameras, a new family of cameras, are gaining popularity. They are vision sensors that differ completely from conventional cameras: instead of capturing images at a fixed rate, they asynchronously measure per-pixel brightness changes generated by moving objects. This motivated us to use it for individual recognition by gait.**

In this paper, we provide means for multimodal gait recognition, by introducing the "Event-based, RGB, and Thermal Gait" database. This database is the first that contains event-camera acquisition, simultaneously with conventional RGB and thermal videos. It contains recordings of people in three variations: normal walking, quick walking, and walking with a backpack. We also present experiments using a baseline algorithm based on gait energy images adapted to event-based camera output. Then we present a comparative experiment against RGB and thermal videos, using the same algorithm, that shows an advantage for event-based data.

***Index Terms*—Event-based camera, Gait recognition, Gait Data Base, Gait Energy Image**

I. Introduction

Gait is a biometric trait aimed at recognizing people by the way they walk. It received lots of interest during the last twenty years, and it has seen considerable progresses. A unique advantage of gait is that it is non-invasive and offers potential for recognition at a distance or at low resolution or when other biometrics might not be perceivable. It is less likely to be obscured since it appears to be difficult to camouflage. [1].

There are many gait datasets published, each containing many variations and using one or more recording sensors. We can mention the "CASIA Gait Database" [2], [3]: it is an indoor multiview dataset. It is constructed to evaluate the ability of the algorithms/methods to manage the carrying conditions, clothing, and view angle distortions. 124 subjects were recorded from 11 different view angles. Each subject is recorded six times under normal conditions (NL), twice under carrying bag conditions, and twice under clothing variation conditions (CL). "OU-ISIR (dataset B)" [4], [5] was built to study the effect of clothing. 48 subjects were recorded on a treadmill under 32 types of clothes. "The TUM Gait from Audio, Image, and Depth (GAID) database" [6]. This database simultaneously contains RGB video, depth, and audio. With 305 people in three variations. and "USF HumanID" [7]: it is an outdoor dataset. It contains 122 subjects recorded under several covariates: viewpoints, shoes, surfaces, carrying conditions, time, and clothing.

Event cameras are a new family of vision sensors that differ completely from conventional cameras: instead of capturing images at a fixed rate, they asynchronously measure per-pixel brightness changes and output the event stream as a sequence of quadruplets $[t, x, y, p]$ that encode the time t, pixel coordinates (x, y) and p, a binary value 0 or 1, that encodes the sign of the brightness changes.

Events are generated when an object is moving in front of the camera. When a person is walking in the receptive field of an event-based camera only his silhouette is recorded, other information such as the background is not. This way only the necessary information to recognize the person's gait is recorded. Also, event-based cameras have very low latency, leading to records without any motion blur. In addition to that, they have a high dynamic range (140 dB instead of 60 dB for standard cameras) which makes them more appropriate for challenging conditions (illumination saturation at underground stations' exits). Another advantage of event-based cameras is that they are better than other visual sensors for personal privacy preservation. They do not provide rich information (details such as the face, clothes, etc.), making the recorded people not recognizable by humans (see Fig. 1).
It cannot be doubted that biometry research and event-based vision research will benefit from new event-based databases since it seems more natural and compatible with gait recognition.

Thermal imaging also is a very powerful remote sensing technique. It can be superior to visible imaging in hard conditions because thermal radiations can penetrate smoke, mist, dust, etc. It is also a technique capable of imaging under both daytime and night-time conditions. Recognizing individuals by their gait can be easier than RGB recordings in such conditions.

To address this, we introduce a new gait recognition database collected using an event-based camera along with a conventional camera and a thermal camera. It should allow designing new algorithms compatible with the new asynchronous

event-based data and allow for a fair comparison with frame-based and thermal computer vision.

In this paper, we introduce, in section II, our new gait database. Section III describes the experiments conducted to provide results of gait recognition performance on event-based data using a baseline algorithm based on the gait energy image. We also provide a comparative study between performances of the algorithm on event-based, RGB, and thermal data. The final section concludes the paper and points out directions for future work.

The dataset is publicly available at https://gait3.netlify.app/.

II. A MULTIMODAL DATABASE FOR GAIT RECOGNITION

In this section, we first introduce the recording setup of the database and the database design.

- **Database Recording Setup** The database is recorded using 2 cameras and a laptop. The DAVIS346 [8], is used to record events. Its resolution is 346×260. For each event $[t, x, y, p]$, $x \in [0, 345]$ and $y \in [0, 259]$ and the data is saved with the DV software. The RGB and thermal recordings were acquired with the dual sensor, visible and thermal, camera FLIR Duo R developed by FLIR Systems. The visible sensor is a CCD sensor with a pixel resolution of 1920×1080. The thermal sensor of this camera is an uncooled VOx microbolometer and has a pixel resolution of 640×512.
- **Recording Procedure** The data collection took place in a controlled environment. Videos were recorded in an empty room equipped with two lightings in addition to the room lights. The cameras were put about 1 meter from the ground and facing perpendicularly to the background (the wall of the room) at about 6 meters from it. Fig. 1 shows an example of an individual from our database, recorded simultaneously with an event-based camera, RGB camera, and thermal camera.

  The database was collected from 56 subjects. Each subject participated in one session and does three walking variations: two walking patterns: Normal walking and quick walking and one carrying condition carrying a backpack. For each variation, the subject is recorded twice, walking from left to right and walking from right to left.

III. PRELIMINARY EVALUATION OF THE DATABASE

We present a preliminary evaluation of event-based data, using a baseline algorithm based on Gait Energy Image (GEI). A comparison between event-based, RGB, and thermal introduced in our database is also performed.

A. Gait energy image for event-based data

Gait Energy Image (GEI) is a Spatio-temporal template of a sequence of images of a walking person used for gait recognition. It was first proposed by Han and Bhanu [9]. Gait energy image of a sequence of frames of a walking human is built by extracting the silhouettes from individual frames (results in a binary image of the silhouette), cropping the silhouettes, and resizing them to a fixed height and width. Then, the gait energy image is the average of the silhouette sequence.

In other words, given a size-normalized human walking binary silhouette sequence $B(x, y, t)$, the gray level GEI $G(x, y)$ is defined as follows:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B(x, y, t) \tag{1}$$

GEI has several advantages, it reflects the major shapes of the human silhouettes and their changes over the sequence (stretching of arms and legs). Another advantage, is its lightweight, instead of processing a long sequence of silhouettes to recognize the person, which can be computationally heavy, we use the compact GEI template.

In the following paragraphs, we describe the approach to adapting the GEI generation to event-based data, which is a stream of events rather than a sequence of frames. This data requires a specific encoding analogous to conventional frame-based data. Then, we show the steps to create the gait energy image.

1) Data encoding: For visible and thermal videos, the data is encoded as a sequence of frames taken at a fixed rate. But the event camera outputs a sequence of asynchronous quadruplets encoding the time, location of the pixel that triggered the event, and the polarity of the illumination change. A difficulty arises with this kind of data because we do not have a straightforward way to process it.

To overcome this limitation, an appealing approach is to represent the stream of events (which is a flat stream of asynchronous quadruplets $[t, x, y, p]$: timestamp, x and y coordinates of the pixel and polarity (1 or 0)) into an image-like representation (a 2D matrix with a fixed number of channels). This representation will fit into any standard computer vision tool or algorithm.

Many encoding approaches can be used to convert the asynchronous event stream to frames, such as Frequency [10], Surface of Active Events (SAE) [11], Leaky Integrate-and-Fire (LIF) [12], Histogram of events [13]. In our work, the histogram of events is the encoding approach used as it is easy and fast to construct and has shown robustness compared to other approaches in tasks like pedestrian detection [14]. The approach consists of splitting the stream of events into non-overlapping chunks using a fixed time window (each chunk contains the events that occurred during the interval $[t_{start}, t_{start} + window]$ where t_{start} is the timestamp of the first event in that chunk) $window$ is set to 20 ms in our experiments. Then each chunk is represented in a one-channel image form called a histogram of events or event-frame for simplicity. A Histogram of events is simply given by the count of triggered events per pixel, independently from the polarity, divided by a constant.

$$H(x_i, y_i) = \min \left(1,\ \frac{1}{m} \sum_{e \in chunck} \chi_{x_i, y_i} \left(e(x), e(y) \right) \right) \tag{2}$$

(a) RGB (b) Thermal (c) Event-based

Fig. 1: Example of an individual, RGB, thermal and event-based recordings

Fig. 2: Histogram of events.

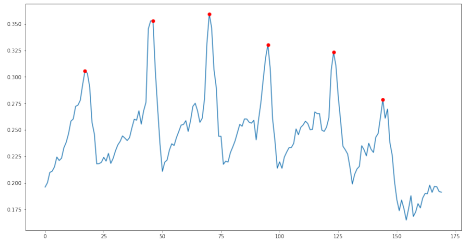

Fig. 3: Normalized correlation of a sequence of event-frames silhouettes with a template silhouette. One step corresponds to two consecutive maximums (red points). We extract two steps to estimate the energy image (GEI)

Where $e(x)$ and $e(y)$ are respectively the x and y coordinates of the pixel that triggered e, m is a normalization factor (10 in our experiment since in our database the number of events triggered by one pixel rarely exceeds 10) and $\chi_{a,b}$ is the indicator function of (a, b) defined as follows:

$$\chi_{a,b} : [0, 345] \times [0, 259] \longrightarrow \{0, 1\}$$
$$(x, y) \longmapsto \begin{cases} 1, & \text{if } x = a \text{ and } y = b \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Then the histogram is clipped, and values that are bigger than 1 become 1. Fig. 2 shows an example of a histogram of events.

2) Steps for GEI: For visible video, generating GEI from the sequence of frames takes three steps: **Human body silhouette segmentation**, then **Gait cycle detection**, and then **GEI calculation**.

a) Human body silhouette segmentation: It means detecting regions in the image that belong to the silhouette of the walking person. For visible data, a lot of algorithms can be used to extract the silhouette. For a static scene where the only moving object is the person, a background segmentation algorithm can be used. For event-based data, non-moving objects do not generate any events. Generating the histogram of events for a chunk of events gives the silhouette of the walking person (see Fig. 2). Background activity noise (events that are triggered but not real) is present but can be eliminated by a threshold on the histogram of events. In the end, the silhouette is cropped with 10 pixels of margin.

b) Gait cycle detection: Gait is a periodic motion at a stable frequency. We aim to detect a subsequence corresponding to a gait cycle (two steps). From the sequence of event frames, we choose the one corresponding to the widest silhouette as a template since it corresponds to stretching legs and arms. Then we perform normalized cross-correlation with all other event frames. Fig. 3 shows the result of normalized cross-correlation of sequence frames with such template. The local maximums in this plot correspond to the starting of gait cycles. The starting and ending of a single cycle can be easily retrieved.

c) GEI calculation: Once we have the silhouettes of the person corresponding to a gait cycle. We align them to have the heads and the trunks superposed (the horizontal translation that maximizes the normalized correlation), then average them to get the GEI (see equation 1). Fig. 4 shows an example of silhouette event-frames in a gait cycle and the rightmost image is the corresponding GEI.

B. Gait energy image for RGB and thermal videos

For RGB and thermal videos, the only step that differs from the event-based approach is silhouette extraction.
To extract silhouettes from RGB images, we performed background segmentation (the K nearest neighbor algorithm) since the scenes are static and the only moving object is the person. For thermal images, extracting the silhouette is performed by thresholding since human body heat radiation is high compared to other objects.
The rest of the pipeline is the same as described in III-A2.

C. Individual recognition

Let $\{S_1, \ldots, S_n\}$ be the set of training gait silhouettes sequences (event-frames, RGB frames or thermal frames) for

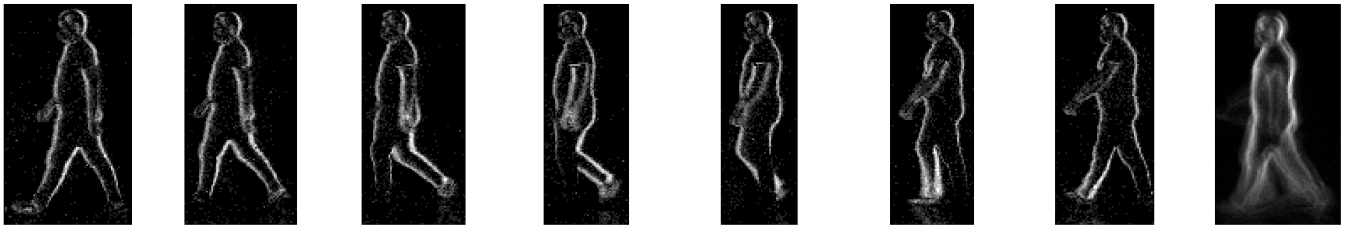

Fig. 4: Examples of silhouette event-frames in a human walking sequence. The rightmost image is the corresponding gait energy image (GEI).

the set of individuals $\{1, \ldots, n\}$. We generate the set of GEI training templates $\{\hat{G}_1, \ldots, \hat{G}_n\}$, according to the procedure described in III-A2 (if we have more than one sequence for the same person, their corresponding GEIs are generated then aligned and averaged into one GEI).

For the classifier based on GEI we define the similarity score of two GEIs $\hat{G}_a$ and $\hat{G}_b$:

$$SScore(\hat{G}_a, \hat{G}_b) = \max\left(normXcorr(\hat{G}_a, \hat{G}_b)\right) \quad (4)$$

Where $normXcorr$ is the normalized cross-correlation function. Its maximum is the similarity score of $\hat{G}_a$ and $\hat{G}_b$.

Given a probe gait silhouette sequence S_p, we follow again the same procedure to generate its GEI $\hat{G}_p$. The similarity score of $\hat{G}_p$ to all classes are computed and ranked in an descending order, and the class with the highest score is the best match of the sequence S_p.

$$Prediction(S_p) = \arg\max_{i=1}^{n}(SScore(\hat{G}_i, \hat{G}_p)) \quad (5)$$

D. Experiments and results

The database contains three walking variations: normal walking, quick walking, and carrying a backpack. For the three data types, the normal walking sequences are used as training or gallery set and the two other sets are used as test sets, analogously to [2].

The procedure described in III-A2 is used to generate GEIs for every sequence. Since every variation has two sequences for every person, the average GEI is calculated to be used for the experiment. Then the recognition is performed as described in III-C to get the performance for every data type (event-based, RGB, and thermal) and every test variation (quick walking and carrying a backpack). For each experiment two performance metrics are calculated, Rank 1 and Rank 3 performance, shown respectively in Tab. I and Tab. II. Rank 1 performance means the percentage of correctly recognized persons. Rank 3 performance means the percentage of correct individuals appearing in 3 closest individuals from the training set.

The results show an advantage of the performance of recognition by GEI using Event-based data over RGB and thermal data.

TABLE I: Comparison of recognition performance between Event-based, RGB and thermal on Rank 1 performance

	Rank 1 performance	
	Quick walking	Carrying backpack
Event-based	84	87.5
RGB	80	85.7
Thermal	64	78.5

TABLE II: Comparison of recognition performance between Event-based, RGB and thermal on Rank 3 performance

	Rank 3 performance	
	Quick walking	Carrying backpack
Event-based	94.6	96
RGB	93	94.6
Thermal	80	87.5

IV. Conclusion

We presented an Event-based, RGB, and thermal database for gait recognition. This database contains recordings of 56 persons, walking in three variations: normally, quickly, and carrying a backpack, using an event-based camera (DAVIS346) and a visible/thermal camera (FLIR Duo R). To the best of our knowledge, this is the first database to provide gait recordings with a real event-based camera (not synthetic or converted from RGB) that will help unlock the potential of this new family of cameras for performing tasks such as gait recognition. Also, it provides simultaneous RGB and thermal recordings allowing comparison or fusion of different data types.

We also presented preliminary experimental of gait recognition using a baseline algorithm based on gait energy image adapted to event-based camera output and comparative results against RGB data. Results show decent performance even with a simple approach, and a slight advantage of event-based data over RGB and thermal.

Based on these promising results for event-based data, future work will focus on adapting more powerful computer vision algorithms to it and fusing different types of data.

References

[1] Wan, Changsheng, Li Wang, and Vir V. Phoha, eds. "A survey on gait recognition." ACM Computing Surveys (CSUR) 51, no. 5 (2018): 1-35.

[2] Yu, Shiqi, Daoliang Tan, and Tieniu Tan. "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition." In 18th International Conference on Pattern Recognition (ICPR'06), vol. 4, pp. 441-444. IEEE, 2006.

[3] Zheng, Shuai, Junge Zhang, Kaiqi Huang, Ran He, and Tieniu Tan. "Robust view transformation model for gait recognition." In 2011 18th

IEEE international conference on image processing, pp. 2073-2076. IEEE, 2011.

[4] Hossain, Md Altab, Yasushi Makihara, Junqiu Wang, and Yasushi Yagi. "Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control." Pattern Recognition 43, no. 6 (2010): 2281-2291.

[5] Makihara, Yasushi, Hidetoshi Mannami, Akira Tsuji, Md Altab Hossain, Kazushige Sugiura, Atsushi Mori, and Yasushi Yagi. "The OU-ISIR gait database comprising the treadmill dataset." IPSJ Transactions on Computer Vision and Applications 4 (2012): 53-62.

[6] Hofmann, Martin, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. "The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits." Journal of Visual Communication and Image Representation 25, no. 1 (2014): 195-206.

[7] Sarkar, Sudeep, P. Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W. Bowyer. "The humanid gait challenge problem: Data sets, performance, and analysis." IEEE transactions on pattern analysis and machine intelligence 27, no. 2 (2005): 162-177.

[8] Gallego, Guillermo, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger et al. "Event-based vision: A survey." IEEE transactions on pattern analysis and machine intelligence 44, no. 1 (2020): 154-180.

[9] Han, Jinguang, and Bir Bhanu. "Individual recognition using gait energy image." IEEE transactions on pattern analysis and machine intelligence 28, no. 2 (2005): 316-322.

[10] Chen, Nicholas FY. "Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018.

[11] Mueggler, Elias, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM." The International Journal of Robotics Research 36, no. 2 (2017): 142-149.

[12] Burkitt, Anthony N. "A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input." Biological cybernetics 95.1 (2006): 1-19.

[13] Maqueda, Ana I., Antonio Loquercio, Guillermo Gallego, Narciso García, and Davide Scaramuzza. "Event-based vision meets deep learning on steering prediction for self-driving cars." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5419-5427. 2018.

[14] Perot, Etienne, Pierre de Tournemire, Davide Nitti, Jonathan Masci, and Amos Sironi. "Learning to detect objects with a 1 megapixel event camera." Advances in Neural Information Processing Systems 33 (2020): 16639-16652.