Front View Gait (FVG-B) Database

Yiyang Su and Xiaoming Liu

Michigan State University

August 26, 2022

1 Introduction

To facilitate gait recognition research, we collected the Front-View Gait (FVG) database over the course of two years, 2017 and 2018. To better protect the privacy of the subjects, we created a version of the database called FVG-B where the face area is blurred to the extent that state-of-the-art face recognition algorithms fail to recognize the subject. In this report, we present the technical details for preparing the FVG-B database. For more information on the FVG-B database and to request access, please visit http://cvlab.cse.msu.edu/frontal-view-gaitfvg-database.html.

2 Face Blurring

We use Gaussian filters on the subjects' faces to protect their privacy. Given a gait image I, we first use a face detector $\mathcal{D}$ to predict a bounding box for the face region. Then, we gradually increase the kernel size and evaluate the cosine similarity between the output feature of the blurred gait image and the gallery images from a face recognition model $\mathcal{R}$. We stop when the cosine similarity for each gallery image is below a threshold t [1]. The algorithm is specified in Algorithm 1. For those subjects who have missing gallery images, we use the last frame where the face can be extracted by $\mathcal{D}$ from sequences 1 and 2 as the gallery images, since the face becomes larger and clearer as the subjects walk towards the camera in FVG.

Algorithm 1 Face blurring for gait image

```
procedure BLUR
    I: input gait image
    D: face detector
    R: face recognition model
    F_g^1, F_g^2, ...: face regions in the gallery images
    F_p ← D(I)
    for k = 1, 3, 5, ... do
        F_p^k ← Gaussian_kernel (input = F, kernel_size = k)
        if cosine_similarity (F_p^k, F_g^i) < t for all i then
            return I overlaid with F_p^k
```

In our implementation, we use MTCNN [2] as the face detector and AdaFace [1] as the face recognition model. The PyTorch implementation of MTCNN [2] and the official implementation of AdaFace [3] are used. For both models, we use pretrained models provided by the authors. The threshold t is chosen to be 0.2.

[1] A typical subject in FVG has two gallery images, one frontal view and one side-view.

[2] https://github.com/timesler/facenet-pytorch

[3] https://github.com/mk-minchul/AdaFace

Session	1			2			3		
Collection Year	2017			2018			2018		
Number of Subjects	147			79			12		
Viewing Angle (°)	-45	0	45	-45	0	45	-45	0	45
Normal	1	2	3	1	2	3	1	2	3
Fast/Slow Walking	4/7	5/8	6/9	4	5	6	4	5	6
Carrying Bag/Hat	10	11	12	-	-	-			
Change Clothes	-	-	-	7	8	9	7	8	9
Multiple Person	-	-	-	10	11	12	10	11	12

Table 1: Variations of different sequences in FVG and FVG-B.

3 Post-Processing

In practice, the face detector fails to detect some of the faces, especially when the subject is far away from the camera. Missing frames substantially affect the gait patterns. To remedy this problem, we mainly follow two strategies.

First, since in all gait sequences in FVG the subjects walk towards the camera, it is reasonable to assume that the kernel size increases monotonically in each gait sequence. Moreover, it is also reasonable to assume that the cosine similarity between the output features of the gait images and the gallery images increases monotonically in each gait sequence. Therefore, if a frame is not changed by the Blur procedure in Algorithm 1 and no frame that precedes it is changed by the procedure (i.e., the cosine similarity between the face feature of the frame and the gallery images are already below 0.2), the subject in all frames preceding it cannot be recognized correctly by the face recognition model with respect to the threshold t and thus do not need to be blurred. For example, suppose frame #10 does not require blurring, then we assume neither do frames #1-9. It is straightforward to use a linear search to find such frames and add all frames preceding them to FVG-B as-is.

Also, in a small time window, the head movement of a person walking can roughly be considered as a linear motion. In the presence of undetected faces where blurring is required as opposed to the previous case, we can infer the face location of the subject from the bounding boxes of adjacent frames using linear interpolation. By our assumption, we get a good approximation of the actual location of the face. In this way, we can use Algorithm 1 even if the face detection failed. Due to the strict constraint of this assumption, we limit the application of the strategy to at most three consecutive frames.

After employing the two strategies, we are able to manually annotate the remaining frames.

4 Experiments

To compare FVG-B with the original FVG without face blurring and to provide a benchmark on FVG-B, we train a GaitNet [3] model on the FVG-B dataset.

Evaluation Protocols. We used the 5 evaluation protocols defined in [3], namely WS, BGHT, CL, MP, and ALL. For references, we reproduce their definitions in Table 3 and the variations of different videos for each subject in Table 1.

Data Split. In [3], the subjects were randomly divided into the training set and the test set with a fixed seed. As a result, some subjects in Session 3 have been placed in the training set and the test set. However, Session 3 was collected in 2018 while both Sessions 1 and 2 were collected in 2017, which makes it suitable for testing the ability of gait recognition models to generalize to data collected with large time gaps (e.g., 1 year). Therefore, we modify the random train/test split such that all subjects in Session 3 are in the test set. The resulting train/test split is presented in detail in Table 2.

Train						Test			
3	5	6	9	10	11	**1**	**2**	**4**	**7**
14	15	16	18	19	20	**8**	**12**	**13**	**17**
21	22	23	24	29	30	25	26	27	28
32	33	34	35	36	37	**31**	**40**	42	45
38	39	41	43	44	46	**48**	49	54	56
47	50	51	52	53	55	57	59	60	63
58	61	62	64	65	67	66	69	73	75
68	70	71	72	74	76	**77**	78	82	86
79	80	81	83	84	85	89	94	96	98
87	88	90	91	92	93	100	101	104	107
95	97	99	102	103	105	109	111	113	115
106	108	110	112	114	117	116	118	121	123
119	120	122	124	126	132	125	127	128	129
133	134	135	137	138	140	130	131	136	139
144	145	147	148	149	150	141	142	143	146
153	154	155	157	159	160	151	152	156	158
161	163	164	166	168	170	162	165	167	169
171	172	173	175	178	182	174	176	177	179
183	186	187	188	189	190	180	181	184	185
191	194	198	199	200	201	192	193	195	196
203	204	205	207	209	210	197	202	206	208
211	212	215	216	220	221	213	214	217	218
223	224	225	226			219	222		

Table 2: The new train/test split. Subjects 1-147 are in Session 1. Subjects 148-226 are in Session 2. Subjects in Session 3 are highlighted in **bold**.

Protocols		WS		BGHT		CL		MP		ALL	
		Gal.	Prb.	Gal.	Prb.	Gal.	Prb.	Gal.	Prb.	Gal.	Prb.
Session1		2	4-9	2	10-12	-	-	-	-	2	1, 3-12
Session2		2	4-6	-	-	2	7-9	2	10-12	2	1, 3-12
Session3		-	-	-	-	-	-	-	-	-	1-12
Database	Split	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
FVG	old	96.2	97.5	92.3	96.4	70.4	**87.5**	92.5	96.0	91.9	96.3
FVG	new	**97.6**	**98.7**	**93.3**	**96.7**	72.2	86.7	**95.6**	96.7	95.2	97.4
FVG-B	new	96.9	98.4	**93.3**	95.6	**80.0**	85.6	94.4	**97.8**	**95.4**	**97.5**

Table 3: Definition of FVG/FVG-B protocols and experimental results for GaitNet [3] on FVG and FVG-B. In the protocols, "Gal." and "Prb." are the abbreviations for "Gallery" and "Probe", respectively.

Figure 1: Three examples of gait sequences in FVG and FVG-B. FVG images are shown on the left and FVG-B images are shown on the right.

4.1 Quantitative Results

The numerical results are presented in Table 3. From the results, we can see that on FVG-B, GaitNet [3] can achieve comparable performance to that on FVG. Our results demonstrate that the model does not rely on facial details when identifying the subjects.

4.2 Qualitative Analysis

To ensure the face blurring process is properly performed, we further compile the blurred faces into a video and visually checked their quality. The difference before and after the blurring is illustrated in Figure 1.

References

[1] Minchul Kim, Anil K Jain, and Xiaoming Liu. "AdaFace: Quality Adaptive Margin for Face Recognition". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2022, pp. 18750–18759.

[2] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. "Joint face detection and alignment using multitask cascaded convolutional networks". In: *IEEE signal processing letters* 23.10 (2016), pp. 1499–1503.

[3] Ziyuan Zhang, Luan Tran, Feng Liu, and Xiaoming Liu. "On learning disentangled representations for gait recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).