04-04

16th International Conference on Machine Vision Applications (MVA)
National Olympics Memorial Youth Center, Tokyo, Japan, May 27-31, 2019.

Gait Recognition Based on Constrained Mutual Subspace Method with CNN Features

Akinari Sakai, Naoya Sogi and Kazuhiro Fukui
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577
{akinari_sakai, sogi}@cvlab.cs.tsukuba.ac.jp, kfukui@cs.tsukuba.ac.jp

Abstract

Gait recognition has various applications such as surveillance and criminal investigation since it can work even at a distance from a camera without the cooperation of subjects. In the conventional methods, especially when using silhouette images, a typical feature is Gait Energy Image (GEI), which is an average silhouette image of a given sequence. Although the GEI feature can represent the sequence compactly, it cannot capture a more detailed structure of the sequence. To address this issue and increase the robustness to the speed variations of walking, gait recognition based on Mutual Subspace Method (MSM) was proposed, where each image sequence is compactly represented by a subspace. In this paper, we enhance further the MSM based method by introducing two functions: 1) to add a feature extraction by projecting onto a generalized difference subspace, 2) to use Convolutional Neural Network (CNN) features, which are obtained from a fully connected layer of a learned CNN, as an input. The extended MSM with the projection is called Constrained MSM, which has been well known as a useful method for image set based recognition. The proposed method achieved the accuracy of 97.7% on an experiment with 1000 subjects from the OU-ISIR Large population datast.

1 Introduction

In this paper, we propose a new framework of gait recognition, which is based on Constrained mutual subspace method (CMSM) [1, 2], considering CNN features obtained through a learned convolution deep neural networks as its input. Gait recognition has gained much attention in surveillance and criminal investigation, as it is available even at a distance from a monitor camera without the cooperation of subjects unlike other types of biometrics, such as fingerprints, iris, and face. Due to these advantages, gait recognition has started to be used in practical cases for a criminal investigation [3, 4].

Gait recognition often used the silhouette images of a subject, as it can suppress the influence of the difference in the texture of clothes. A typical feature extracted from the silhouette images is Gait Energy Image (GEI) [5, 6, 7, 8, 9, 10], which is defined as the average of sequential silhouette images. GEI feature can represent the overall structure of a gait sequence. However, in the GEI, detailed structural information such as the direction and magnitude of time change in silhouette images may be lost in the compressing process to an average.

To precisely represent the structure information of a gait sequence, the mutual subspace method (MSM)[11] was introduced [12, 13]. In the MSM based method, a set of sequential gait silhouette images is compactly represented by a subspace in a high dimensional vector space, where a class subspace of a subject is generated by applying the principal component analysis (PCA) to a set of learning samples. Such a subspace reflects more rich information on the structure of the sequence than the GEI feature. Another important merit of introducing the subspace representation is to increase the robustness against the change in walking speed, since the subspace is less susceptible to the walking speed[12, 13].

However, the discriminant ability of MSM is still insufficient to perform high performance, since a class subspace of each subject is generated without considering the class subspaces of other subjects. To improve the discriminant ability, we introduce Constrained mutual subspace method (CMSM), which is a powerful extension of MSM that adds the projection of class subspaces onto a generalized difference subspace. Considering high effectiveness of CMSM in image set based recognition such as face, hand and action recognition [1, 2, 14], we expect highly that the performance of the MSM based framework can be further enhanced, while inheriting its robustness against the change in walking speed.

Besides the introduction of CMSM, we utilize CNN features, which are obtained by feeding each silhouette image of the sequence into a learned Convolutional Neural Network. CNN feature can be used validly in various types of recognition methods [15, 16, 17]. Thus, we expect that the CNN feature can work well under our framework based on CMSM to boost its performance.

In this way, our framework of gait recognition is built with two steps: 1) extraction of CNN features and 2) classification by CMSM, as shown in Figure 1. We verify the effectiveness of our framework through two types of experiments using OU-ISIR Treadmill dataset [18] and OU-ISIR Large population

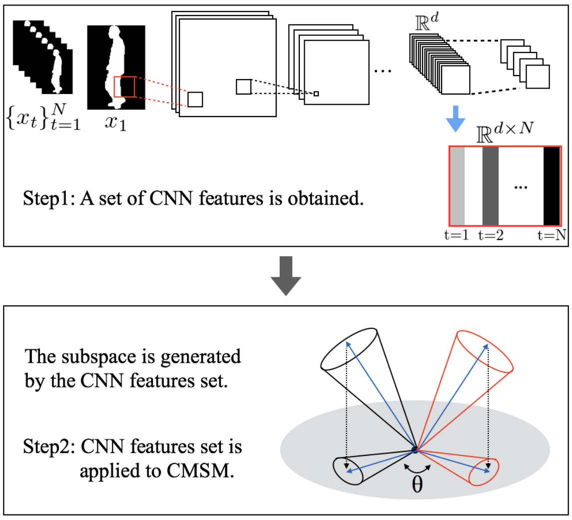

Figure 1. Conceptual diagram of the proposed framework. 1) CNN is learned with training dataset, and then training and test data are transformed to CNN features through the learned CNN, 2) CMSM is performed on CNN features.

dataset [19]. In the former, we evaluate the robustness of our framework against the change in walking speed in details. In the later, we evaluate the performance of our framework on a large scale database.

The rest of this paper is organized as follows. In Section 2, we describe the algorithms of the conventional methods, GEI and MSM. In Section 3, we describe the details of our framework of gait recognition. In Section 4, we demonstrate the validity of the proposed framework through classification experiments on two public datasets. Section 5 concludes the paper.

2 Related Work

In this section, firstly, we describe the representation of a gait sequence by GEI. Then, we overview the concepts of the mutual subspace method and how to apply it to the gait recognition.

2.1 Gait Energy Image

Gait Energy Image (GEI) is a spatio-temporal feature of a gait sequence. Given frame images of the sequence, each frame is converted to a silhouette image as a preprocessing. Let $\mathbf{x}_t \in \mathbb{R}^{h\times w}$ be a gait silhouette image at time t in a sequence. The GEI is obtained by aggregating the silhouette images $\{\mathbf{x}_t\}$ over one gait period as the following:

$$GEI(\{\mathbf{x}_t\}_{t=1}^N) = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}_t, \tag{1}$$

where, N is a length of a gait sequence.

It is shown that this aggregation induces the GEI to robustness against noise included in each frame[5]. Furthermore, as an advantage of the GEI, it is known that the GEI does not depend on the walking speed. For these reasons, there are various gait recognition methods using the GEI as an input [5, 6, 7, 8, 9].

However, the above operation does not consider the direction and magnitude of change in the silhouette images. In other words, GEI decreases the structure of the sequence, although the structure is important for realizing recognition with high accuracy. To overcome this problem while keeping the advantages of the GEI, the mutual subspace method have been introduced for the gait recognition [12].

2.2 Mutual subspace method for the gait recognition

The mutual subspace method (MSM) is a recognition method of a set of images [11]. In MSM, a set of images is represented by a subspace representing structural information of the set. After the subspace representation, a set of input images is classified by comparing input and dictionary subspaces.

In gait recognition, a whole structure of silhouette images $\{\mathbf{x}_t\}$ in a gait sequence is represented by a subspace [12]. Once applying the subspace representation, comparing the structural information between two gait sequences can be done by comparing the two subspaces. To this end, the similarity between the two subspaces is calculated by using canonical angles.

Given N_1-dimensional subspace $\mathcal{S}_1$ and N_2-dimensional subspace $\mathcal{S}_2$ in d-dimensional vector space, where $N_1 \leq N_2$, the canonical angles $\{0 \leq \theta_1, ..., \theta_{N_1} \leq \frac{\pi}{2}\}$ between the $\mathcal{S}_1$ and $\mathcal{S}_2$ are recursively defined as follows[20, 21]:

$$\cos\theta_i = \max_{\mathbf{u}\in\mathcal{S}_1}\max_{\mathbf{v}\in\mathcal{S}_2}\mathbf{u}^{\mathrm{T}}\mathbf{v} = \mathbf{u}_i^{\mathrm{T}}\mathbf{v}_i, \tag{2}$$

$$s.t. \|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = 1, \mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j = \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = 0, i \neq j,$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the canonical vectors forming the i-th smallest canonical angle θ_i between $\mathcal{S}_1$ and $\mathcal{S}_2$. The j-th canonical angle θ_j is the smallest angle in a direction orthogonal to the canonical angles $\{\theta_k\}_{k=1}^{j-1}$.

The canonical angles can be calculated from the orthogonal projection matrices onto subspaces $\mathcal{S}_1, \mathcal{S}_2$. Let $\{\mathbf{\Phi}_i\}_{i=1}^{N_1}$ be basis vectors of $\mathcal{S}_1$ and $\{\mathbf{\Psi}_i\}_{i=1}^{N_2}$ be basis vectors of $\mathcal{S}_2$. The projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are calculated as $\sum_{i=1}^{N_1}\mathbf{\Phi}_i\mathbf{\Phi}_i{}^{\mathrm{T}}$ and $\sum_{i=1}^{N_2}\mathbf{\Psi}_i\mathbf{\Psi}_i{}^{\mathrm{T}}$, respectively. $\cos^2\theta_i$ is the i-th largest eigenvalue of $\mathbf{P}_1^{\mathrm{T}}\mathbf{P}_2$ or $\mathbf{P}_2^{\mathrm{T}}\mathbf{P}_1$. Alternatively, the canonical angles can be easily obtained by applying the SVD to the orthonormal basis vectors of the subspaces.

The similarity between two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined by using the canonical angles as follows:

$$sim(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{N_1}\sum_{i=1}^{N_1}\cos^2\theta_i. \tag{3}$$

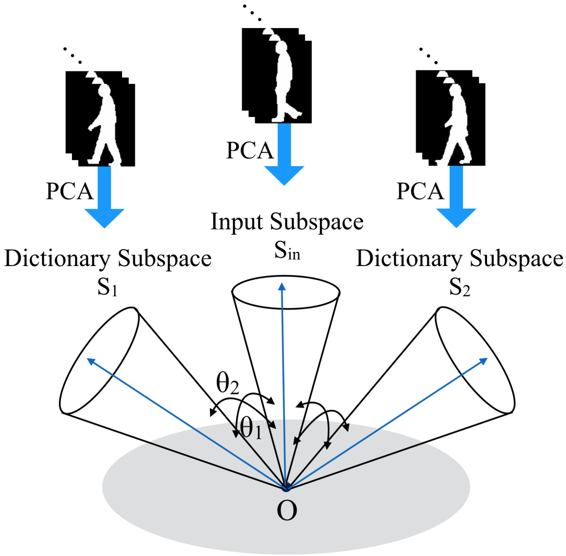

Figure 2. Conceptual diagram of MSM for the gait recognition. Each gait sequence is represented by a subspace, where it is generated by applying PCA to the silhouette images of the sequence. An input subspace is classified based on the similarities between input and dictionary subspaces.

In MSM, an input subspace $\mathcal{S}_{in}$, which is obtained from a gait sequence, is classified by comparing it with dictionary subspaces $\{\mathcal{S}_c\}_{c=1}^{C}$ using this similarity as shown in Fig.2.

3 CMSM with CNN features

In this section, we first describe the concept of CNN features and constrained MSM. Then, we explain the overall method for gait recognition based on the constrained MSM with CNN features.

3.1 CNN features

The underlying architecture of convolutional neural networks (CNN) is composed by connecting three types of layers, convolution, pooling, and fully connected layers. In a CNN, an input image is classified based on the output conviction degrees for each class from the last layer with softmax. The last layer uses a feature, extracted by previous layers stepwise, as evidence for classification. Thus, the feature, extracted from any hidden layer in the CNN, has high discriminative nature. This discriminative feature is called CNN feature.

Previous works have shown the effectiveness of CNN features for various types of applications [15, 16, 17]. We also use this advantage to the gait recognition by a subspace-based method.

3.2 Constrained Mutual Subspace Method

In the MSM, the subspace of each class is generated without considering a relationship with other class subspaces. Therefore, subspaces that have similar feature may be close together, so that MSM has room for improvement. Constrained MSM (CMSM) is an extension method of the MSM based on this idea [1, 2]. The essence of CMSM is to classify input data by comparing input and dictionary subspaces while limiting the common information between different class subspaces. This is because the common information is unnecessary for the classification. From this reason, in the CMSM, an input subspace is classified after extracting the difference component between classes by projecting the subspaces onto the generalized difference subspace (GDS) [2].

GDS is designed to contain only different components among class subspaces $\{\mathcal{S}_c\}_{c=1}^{C}$ [2]. Thus, the projection of class subspaces onto GDS can increase the class separability, largely improving the classification ability of MSM.

3.3 Gait recognition using CMSM with CNN features

We construct a gait recognition method using the CMSM and CNN features as shown in Fig.3. In the following, we explain training and recognition procedure in the case that C subjects have a silhouette sequence $\{\mathbf{x}_t^c\}_{t=1}^{N}|_{c=1}^{C}$ as training data.

•Training phase

1. CNN is trained to classify each silhouette image by using training images $\{\mathbf{x}_t^c\}$.
2. CNN features $\{\mathbf{f}_t^c\}$ are extracted from the trained CNN.
3. Each class subspace $\{S_c\}$ is generated by applying PCA to a set of CNN features$\{\mathbf{f}_t^c\}$.
4. GDS is generated by using $\{S_c\}$. Reference subspaces $\{\hat{S}_c\}$ are generated by projecting $\{S_c\}$ subspaces onto GDS.

• Recognition phase

1. Given the input silhouette sequence $\{\mathbf{x}_t^{in}\}$.
2. A set of input CNN features $\{\mathbf{f}_t^{in}\}$ is extracted from the trained CNN.
3. A subspace of the input sequence $\{S_{in}\}$ is generated by applying PCA to the input CNN features $\{\mathbf{f}_t^{in}\}$.
4. An input subspace $\{\hat{S}_{in}\}$ is generated by projecting $\{S_{in}\}$ onto the GDS.
5. The input sequence $\{\mathbf{x}_t^{in}\}$ is recognized based on similarities between the input subspace $\{\hat{S}_{in}\}$ and dictionaries $\{\hat{S}_c\}$.

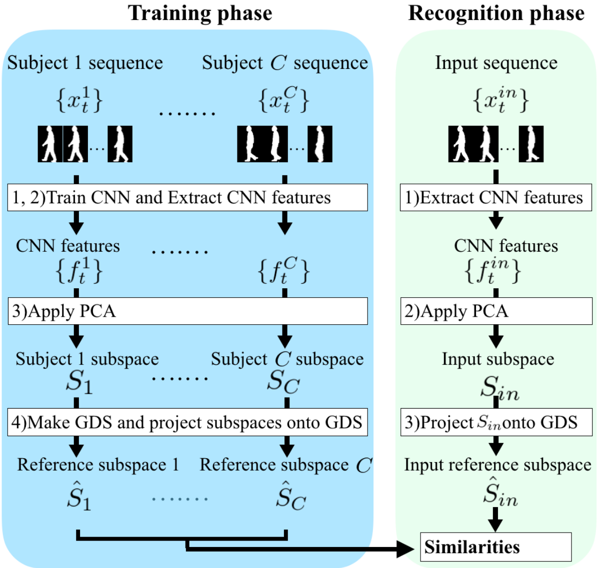

Figure 3. The process flow of the CMSM with CNN features for gait recognition.

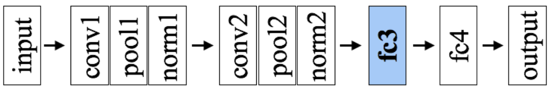

Figure 4. CNN structure for the experiments. This architecture is used for extracting CNN features and classifying frames of a gait sequence.

Table 1. Layer configurations for CNN feature extraction.

Layer	Kernels	Size/ Stride	Activation	Pooling
conv1	18	$7 \times 7 \times 1/1$	ReLU	-
pool1	-	$2 \times 2/2$	-	Max pooling
conv2	45	$5 \times 5 \times 18/1$	ReLU	-
pool2	-	$3 \times 3/2$	-	Max pooling

Figure 5. Example of the gait sequence.

4 Experiments

In this section, we confirm the effectiveness of using CMSM with CNN features for gait recognition in two experiments. The first experiment focuses on the robustness against the change in walking speed by using OU-ISIR Treadmill dataset A [18]. The second experiment thoroughly investigates the performance using enormous dataset, OU-ISIR large population dataset [19].

4.1 Verification of speed robustness

Details of the OU-ISIR Treadmill dataset A: The treadmill database A includes 34 subjects. For each subject, gait sequences are taken in 1 km/h units from 2 km/h to 10 km/h. Each sequence includes multiple gait cycles. The above procedure was repeated twice for training and testing sequences. Thus, training and testing data have 306 (= 34 *subjects* $\times$ 9 *speeds*) sequences, respectively.

After the shooting, each RGB frame is converted to the silhouette image. The size of the silhouette image is 128×88.

Experimental protocol: In this experiment, 3 gait cycles were extracted from each subject's sequence at random. The number of images included in a gait cycle was decided by subjective evaluation. Each cycle extracted from testing data is classified.

Training and testing data were selected from the whole dataset, according to the following procedure. Cycles taken at one speed were used as training. On the other hand, the same speed cycles from the predefined training set and within ±2km/h cycles of that were used for the testing. For instance, when the 3 cycles included in 4km/h are used as training data, 2km/h, 3km/h, 4km/h, 5km/h, and 6km/h are used for testing.

We used the architecture of the CNN shown in Fig.4 to extract the CNN features. The parameters of each layer of CNN are shown in Table 1. The CNN is trained beforehand, using the training dataset, to classify each frame. The CNN features are extracted from the fc3 layer, which is a fully connected layer of 1024 units, included in the trained CNN.

Besides, for the evaluation, we used the overall of the trained CNN. After inputting silhouette images of an input cycle into the CNN, the input cycle is classified based on the average value of the output conviction degrees of each class from the fc4 layer with softmax. We refer to this method as CNN(softmax).

As a baseline, we set a classification method using CNN with the GEI. The baseline method classifies a cycle by inputting the GEI of the cycle into CNN. This CNN is the same architecture as the Fig.4, but training is conducted to classify GEIs, not each frame. We refer to this method as GEI-CNN.

Result: Table 2 shows the rank-1 identification rates for each method. CMSM is superior to MSM. This indicates that GDS projection works effectively as a valid feature extraction in a task of gait recognition as well as in face and hand recognition. Furthermore, we can see that CNN-CMSM is superior to conventional methods, GEI-CNN, CMSM, and CNN (softmax) in most cases. This result suggests that it is important to extract and compare the whole structure of each gait sequence.

Table 2. The result of the experiment using OU-ISIR treadmill dataset(%).

Train-Test	GEI CNN	MSM	CMSM	CNN (softmax)	**CNN CMSM**
2km-2km	74.2	81.7	97.1	**100**	99.7
2km-3km	64.7	72.2	92.5	**99.3**	97.4
2km-4km	30.7	46.7	80.7	92.8	**94.1**
3km-2km	53.9	62.4	97.1	98.4	**100**
3km-3km	82.4	87.3	99.7	**100**	**100**
3km-4km	54.9	67.0	96.1	98.7	**99.7**
3km-5km	19.9	41.2	86.6	94.8	**97.7**
4km-2km	20.3	29.4	86.9	92.5	**95.8**
4km-3km	51.6	63.4	97.7	**99.3**	**99.3**
4km-4km	84.3	87.6	94.8	**98.7**	98.0
4km-5km	50.7	68.6	93.8	94.8	**96.1**
4km-6km	21.9	40.5	81.4	91.2	**93.5**
5km-3km	15.4	26.1	87.9	95.4	**97.4**
5km-4km	45.8	50.7	94.1	94.8	**97.1**
5km-5km	85.3	90.8	94.8	**100**	**100**
5km-6km	58.8	73.2	88.9	93.1	**99.3**
5km-7km	35.6	51.0	75.8	86.9	**94.8**
6km-4km	7.8	30.7	83.3	93.5	**95.8**
6km-5km	46.7	64.7	95.4	96.1	**99.7**
6km-6km	84.0	87.9	96.4	**100**	99.7
6km-7km	60.1	82.4	93.5	95.1	**97.4**
6km-8km	8.8	10.8	16.7	29.1	**36.9**
7km-5km	18.0	25.8	78.8	87.6	**88.6**
7km-6km	56.2	59.2	94.8	97.4	**98.0**
7km-7km	75.5	88.6	**100**	**100**	99.7
7km-8km	9.2	16.0	35.6	45.1	**55.9**
7km-9km	3.9	9.8	21.9	36.9	**44.1**
8km-6km	9.5	5.6	7.8	13.7	**14.1**
8km-7km	15.7	6.9	11.1	18.0	**19.3**
8km-8km	65.4	80.7	90.8	93.8	**94.4**
8km-9km	69.3	78.8	94.4	94.1	**98.0**
8km-10km	52.3	72.2	85.9	90.5	**95.4**
9km-7km	5.2	8.5	12.7	13.1	**20.3**
9km-8km	56.2	71.2	80.7	84.0	**89.5**
9km-9km	81.4	92.8	98.7	97.1	**100**
9km-10km	70.6	87.9	92.5	97.1	**98.7**
10km-8km	46.7	51.6	69.3	83.0	**86.9**
10km-9km	73.9	83.7	97.4	**99.3**	99.0
10km-10km	75.2	92.1	98.7	**100**	**100**
Average	47.2	57.6	79.5	84.5	**87.0**

Table 3. The result of the experiment using OU-ISIR large population dataset(%).

Number of Subjects	GEI CNN	MSM	CMSM	CNN (softmax)	**CNN CMSM**
200	77.0	91.0	95.5	98.5	**99.5**
500	80.0	83.8	95.8	96.6	**98.4**
1000	76.5	81.7	94.1	94.5	**97.7**
1500	53.4	78.9	92.4	93.3	**94.1**

4.2 Verification of performance in large population

Details of the OU-ISIR large population dataset: The OU-ISIR large population database includes approximately 4000 subjects. For each subject, gait sequences are observed from four view angles, 55 degrees, 65 degrees, 75 degrees, and 85 degrees. Each sequence approximately includes a gait cycle. The above procedure was repeated twice for training and testing sequences. Thus, training and testing data have $C\ subjects \times 4\ angles$ sequences, respectively.

After the shooting, each RGB frame is converted to the silhouette image. The size of the silhouette image is 128×88.

Experimental protocol: In this experiment, we used only sequences taken at 75 degrees. This experiment was conducted with 200, 500, 1000 and 1500 subjects.

CNN features were extracted from the trained CNN under this experimental setting, according to the same procedure described in the previous experiment. For comparison, the softmax values of the trained CNN, GEI-CNN, and other methods were used according to the previous experiment.

Result: Table 3 shows the rank-1 identification rates for each method. We can see that CNN-CMSM achieved the highest results in all of the cases. For GEI+CNN, we could not train CNN completely due to the lack of training data, which caused the poor performance. When the number of subjects is 1500, CNN-CMSM achieved 94.1% accuracy, which is the highest in this experiment. Besides, it is practical that when the number of subjects increases, there is no sharp decline in the identification rate.

5 Conclusion

In this paper, we proposed a new method for gait recognition, based on CMSM and CNN features. We verified the effectiveness of our method using two different databases. In the experiment with the OU-ISIR large population dataset, CMSM+CNN showed the 97.7% accuracy in 1000 subjects classification. Also, even when the number of subjects increased, no sudden decline in the recognition rate was observed. This indicates that the gait recognition system can be operated with high accuracy regardless of the number of the subjects.

Several gait databases[22, 23, 24] have been released to solve various problems in gait recognition. As a future work, we aim to establish a gait recognition method that can be used in any situations.

Also, in this paper, the neural network model construction to make feature extractor is not considered. In the current gait recognition, the gait image is first converted to silhouette as a pre-processing. This silhouette image is simple so that high-performance recognition could be achieved with even a relatively shallow architecture. However, in the case of inputting raw images, we need to consider how to design a network structure to achieve a high recognition rate, while reducing the computational cost.

Acknowledgement

We would like to express our special thanks of gratitude to the Yagi Laboratory in the Osaka University that provided us the gait databases.

References

[1] Kazuhiro Fukui and Osamu Yamaguchi. Face Recognition Using Multi-viewpoint Patterns for Robot Vision. In *The Eleventh International Symposium of Robotics Research*, pages 192–201, 2005.

[2] Kazuhiro Fukui and Atsuto Maki. Difference Subspace and Its Generalization for Subspace-Based Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(11):2164–2177, 2015.

[3] Haruyuki Iwama, Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi. Gait verification system for criminal investigation. *Information and Media Technologies*, 8(4):1187–1199, 2013.

[4] Imed Bouchrika, Michaela Goffredo, John Carter, and Mark Nixon. On using gait in forensic biometrics. *Journal of forensic sciences*, 56(4):882–889, 2011.

[5] Jinguang Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2):316–322, 2006.

[6] Worapan Kusakunniran, Qiang Wu, Hongdong Li, and Jian Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *IEEE 12th International Conference on Computer Vision Workshops*, pages 1058–1064. IEEE, 2009.

[7] Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2016.

[8] Khalid Bashir, Tao Xiang, and Shaogang Gong. Feature selection on gait energy image for human identification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 985–988. IEEE, 2008.

[9] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11):977–984, 2009.

[10] Maryam Babaee, Linwei Li, and Gerhard Rigoll. Person identification from partial gait cycle using fully convolutional neural network. *Neurocomputing*, 2019.

[11] Osamu Yamaguchi, Kazuhiro Fukui, and Ken-ichi Maeda. Face recognition using temporal image sequence. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998.

[12] Yumi Iwashita, Hitoshi Sakano, and Ryo Kurazume. Gait recognition robust to speed transition using mutual subspace method. In *International Conference on Image Analysis and Processing*, pages 141–149. Springer, 2015.

[13] Yumi Iwashita, Mafune Kakeshita, Hitoshi Sakano, and Ryo Kurazume. Making gait recognition robust to speed changes using mutual subspace method. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2273–2278. IEEE, 2017.

[14] Kazuhiro Fukui, Björn Stenger, and Osamu Yamaguchi. A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method. In *Asian Conference on Computer Vision*, pages 315–324. 2006.

[15] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition workshops*, pages 806–813, 2014.

[16] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. Factors of Transferability for a Generic ConvNet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1790–1802, 2016.

[17] Naoya Sogi, Taku Nakayama, and Kazuhiro Fukui. A method based on convex cone model for image-set classification with cnn features. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[18] Yasushi Makihara, 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*, 4:53–62, 2012.

[19] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, 2012.

[20] Harold Hotelling. RELATIONS BETWEEN TWO SETS OF VARIATES. *Biometrika*, 28(3-4):321–377, 1936.

[21] Sydney N Afriat. Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 53(04):800–816, 1957.

[22] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, 2012.

[23] Yasushi Makihara, Atsuyuki Suzuki, Daigo Muramatsu, Xiang Li, and Yasushi Yagi. Joint intensity and spatial metric learning for robust gait recognition. In *Proc. 30th IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2017)*, pages 5705–5715, 2017.

[24] Al Mansur, Yasushi Makihara, Rasyid Aqmar, and Yasushi Yagi. Gait recognition under speed transition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2528, 2014.