GAIT RECOGNITION

Introduction:

Biometric traits such as fingerprint, face, iris and gait can be used to identify or authenticate a person. Gait recognition is non-intrusive and addresses the problem of human identification at a distance based on their unique walking patterns. Gait recognition has been studied through analyzing the gait collection from camera, inertial measurement sensors such as accelerometer and gyroscope, floor sensors and Wi-Fi/radar signalling mechanisms as shown in Fig 1.

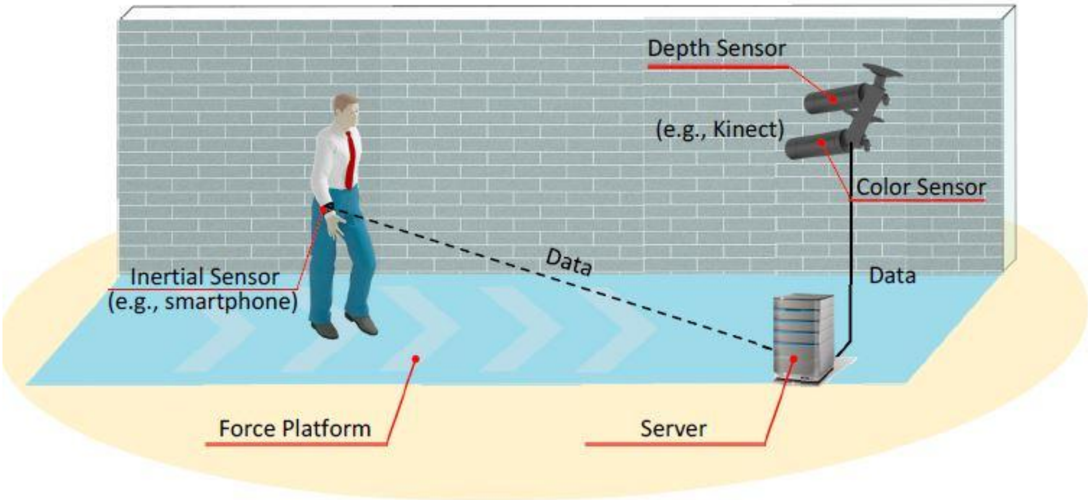

Fig 1: Application scenario for gait recognition

Camera based gait recognition methods can be broadly classified into model based methods and appearance based/ model free methods. The model-based approaches commonly use a priori model to match the data extracted from a video and parameters of the model are then used for gait recognition. An accurate model establishment is necessary to obtain better performance. But the modelling and optimizing processes are both complex and error prone. This necessitates choosing model free methods. Compared to model- based methods, appearance based approaches extract the gait features by analyzing the shapes, texture, color and contours of the silhouettes in successive frames.

Gait recognition proposed by Wang et al [1] is a state of the art algorithm and a brief overview is as follows. The challenges, improvements and recent work are also discussed further to make it a real world product.

Silhouette analysis based gait recognition:

This method proposes a gait recognition algorithm using spatial-temporal silhouette analysis measured during walking. Silhouette is defined as a region of black and white pixels of the walking person. Three major modules of this method are as follows: human detection and tracking, feature extraction, and training/classification as shown in Fig 2. The first module serves to detect and track the walking figure in an image sequence. Least Median of Squares method [2] is used to construct the background by determining background brightness value for each pixel location. Silhouette is the thresholded version of frame difference with the computed background.

In the second module, Image morphological operations followed by connected component labelling were performed to extract the silhouette by removing noisy blobs. The contour in 2D silhouette Image has been unwrapped counter clockwise with respect to the silhouette centroid into a 1D normalized distance signal.

In the third module, Principal Component Analysis (PCA) been applied to these 1D signals to reduce the dimensionality of the input feature space. Classification has been done with the Nearest Neighbours while Dynamic Time Warping (DTW) and Normalised Euclidian distances being similarity measures.

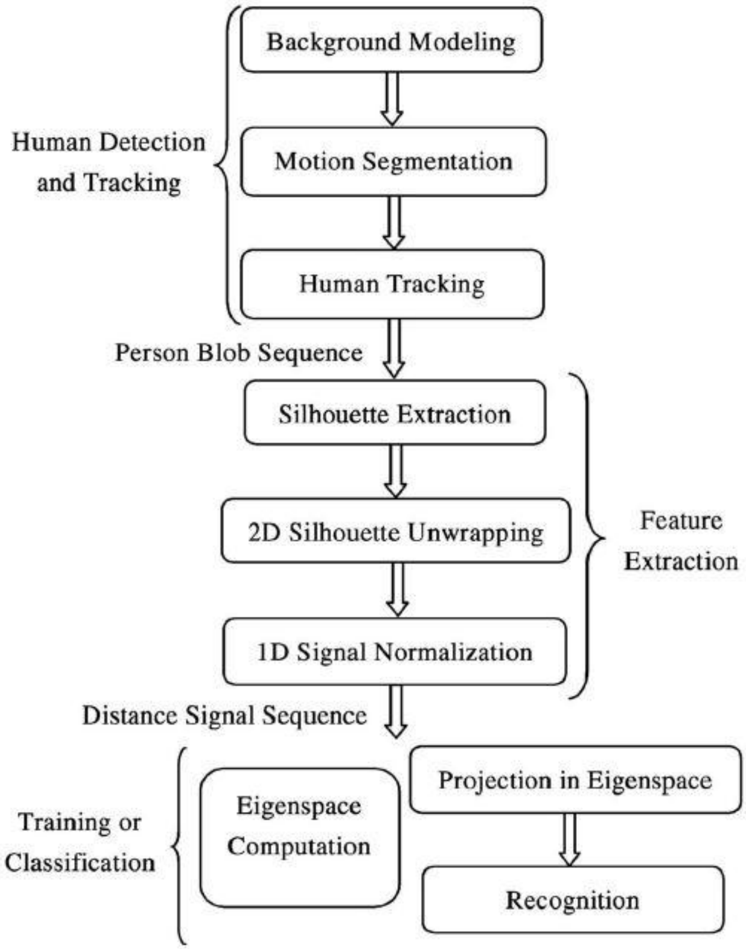

Fig 2: Algorithm flow of gait recognition

Improvements and recent work:

1. The GEI (gait energy image) method [3] represents gait dynamics by aligned and normalized silhouettes over a gait cycle. The GEI provides a compact representation of the spatial occupancy of a person over a gait cycle as shown in Fig 3.

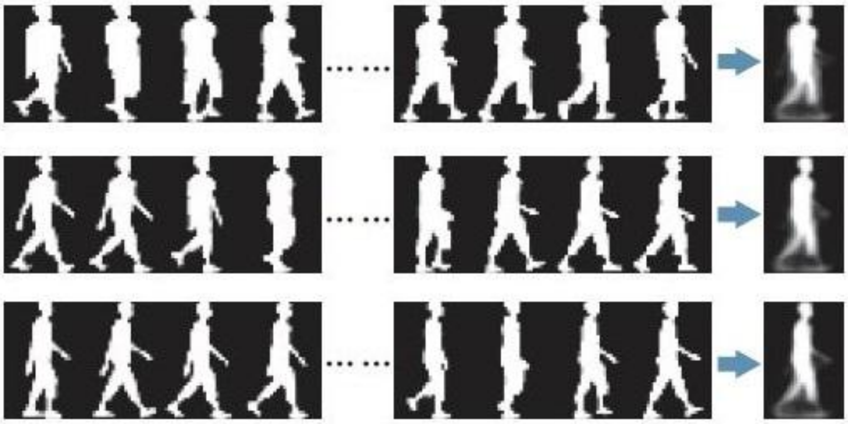

Fig 3: GEI for various silhouette sequences

2. The silhouette mainly depends on the shape information extracted from motion. This varies rapidly with change of clothing of the subject. To improve the Silhouette extraction in different clothing:

   (a) Active Energy Image (AEI) [4] and the gait entropy image (GEnI) [5] templates were able to produce a better recognition accuracy over the clothing and carrying covariates in gait.
   (b) A sway alignment of Silhouette considering the fact that as a person walks, his/her whole body sways forward and backward periodically and this information is lost in the simple alignment technique [6].
   (c) A gait energy response function (GERF) transforms original gait energy into another one, which increases discrimination capability under clothes variation [7].
   (d) Masking the GEI with the image of the respective GEnI [8] and an adaptive outlier detection method to remove the effect of clothing on silhouettes [9].

3. This method works well when subjects walk along a straight-line path with the viewing angle of the camera lateral (0 degrees), oblique (45 degrees), and frontal (90 degrees) to the image plane. This necessitates a view invariant model. View Transformation Methods use transforms to match the angle between sequences [10, 11]. View Preserving Models (VPM) incorporates the view information within the feature set for the extraction of relevant view-invariant gait features such as geometric view estimation [12]. A variant of VPM involves extraction of view-independent features through multi-view training and then use a single gallery view for testing. Choudhury et al. [13] designed a VPM named view-invariant multi-scale gait recognition (VIMGR) which applied Shannon's entropy function to the lower limb region of the GEI. The sub-region selection was later modified by Rida et al. [14] automating this segmentation procedure with a process known as group lasso of motion (GLM). Their approach to the problem has shown significant improvement in the covariate recognition accuracy.

4. To establish a cross-view mapping between gallery and probe templates, multi-view feature learning can be done. View-invariant feature selector (ViFS) [15] selects features from multi-view gait templates and reconstructs gallery templates that accurately match the data for a specific view angle. ViFS reconstructs gallery templates from arbitrary view angles, and thus help to transfer the cross-view problem to identical-view gait recognition.

CNN based implementations:

1. Multi-View gait recognition using 3D CNNs [17]: This network as mentioned in Fig 4 was able to learn a representation independent of the view, speed and clothing conditions. The training/test splits was also a reason for achieving it. In case where training and test data have been recorded under different conditions, both were split separately into a 66% and 33% partition. Then the 66% of the training data are combined with the 66% of the test data to form the altered training data. The remaining 33% of the original training and test data were combined to form the altered test data. This network architecture was able to achieve more than 94% accuracy for different view configurations on CASIA Dataset B [16].

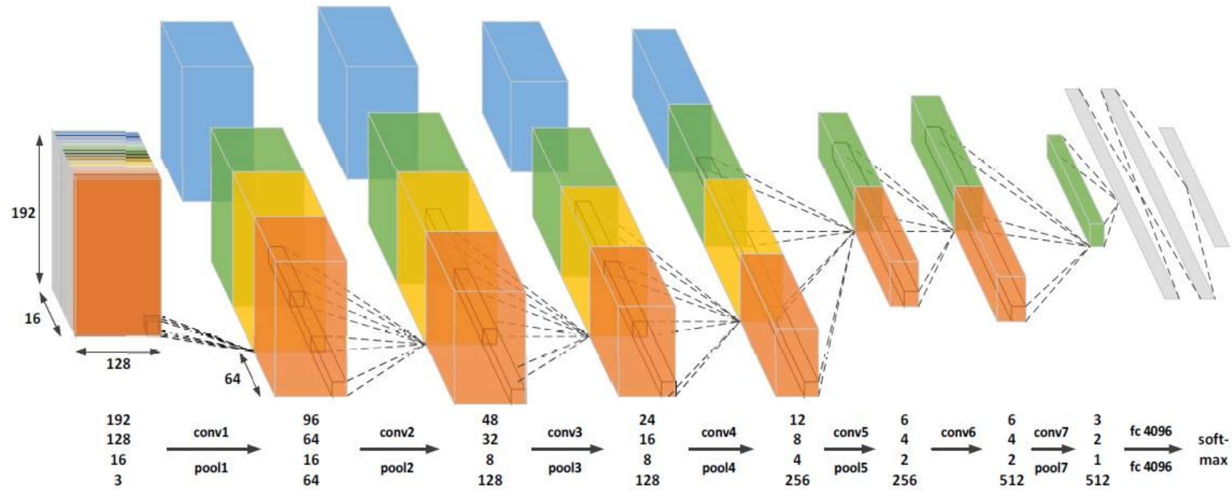

Fig 4: Topology of the network. All pooling layers are max-pooling layers with a size of 2x2x2, except for pool1 and pool3 which are of size 2x2x1. After each pooling layer a ReLU-nonlinearity [16] is implemented. The features aquired in conv7 are used as inputs of two consecutive fully connected layers with 4096 units each applying dropout with a value of 0.5. The final softmax layer produces a probability distribution over all classes.

2. SIAMESE Neural Network for human gait recognition [18]: A Siamese neural network as in Fig 5 contains two parallel CNN architectures sharing the same parameters which learns a function that maps input patterns into a latent space where similarity metric to be small for pairs of the same objects.

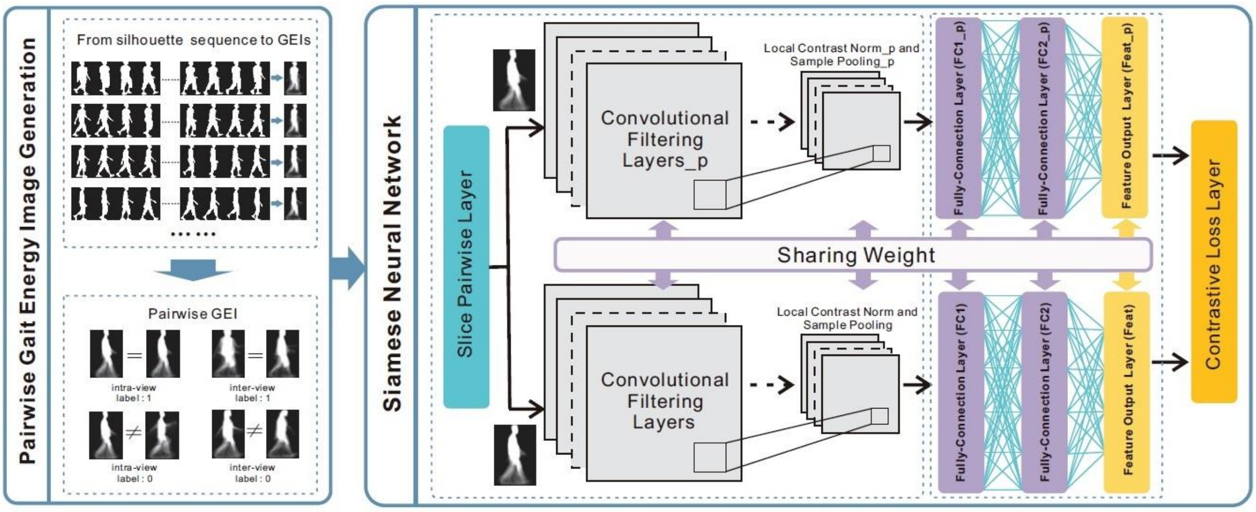

Fig 5: Siamese neural network based gait recognition for human identification

They sent both intra-view and inter-view pair wise GEIs into the Siamese neural network. In this way, they trained the similarity metric to be small for pairs from same subjects, and large for pairs from different subjects. It enhanced the robustness of the gait-based human identification method under the view change condition.

3. Cross-View Gait for Human Identification with Deep CNNs [19]: With the average recognition rate more than 94 %, much better than the previous best result (less than 65%) when the cross view angle is less than 36 degrees, Cross-view and cross-walking were handled with suitable pre-processing approaches and network architectures. This method also performs the best on the USF gait dataset [27], where the gait sequences were imaged in a real outdoor scene. It was also trained on OU-ISIR [26] for generalization with 98% accuracy on similar view appearance and 91% on cross view appearance. These results show great potential of this method for practical applications.

4. EV-Gait [20]: The Dynamic Vision Sensors (Event Cameras) were used instead of traditional RGB sensors. Event-based Gait Recognition (EV-Gait) approach exploits the motion consistency to effectively remove noise, and uses a deep neural network as shown in Fig 6 to recognise gait from the event streams. The event stream has 128 x 128 resolution and the view angle ranges from 0 to 180 degrees. The accuracy of the model is close to 96% with less computational time compared to 3D-CNN.

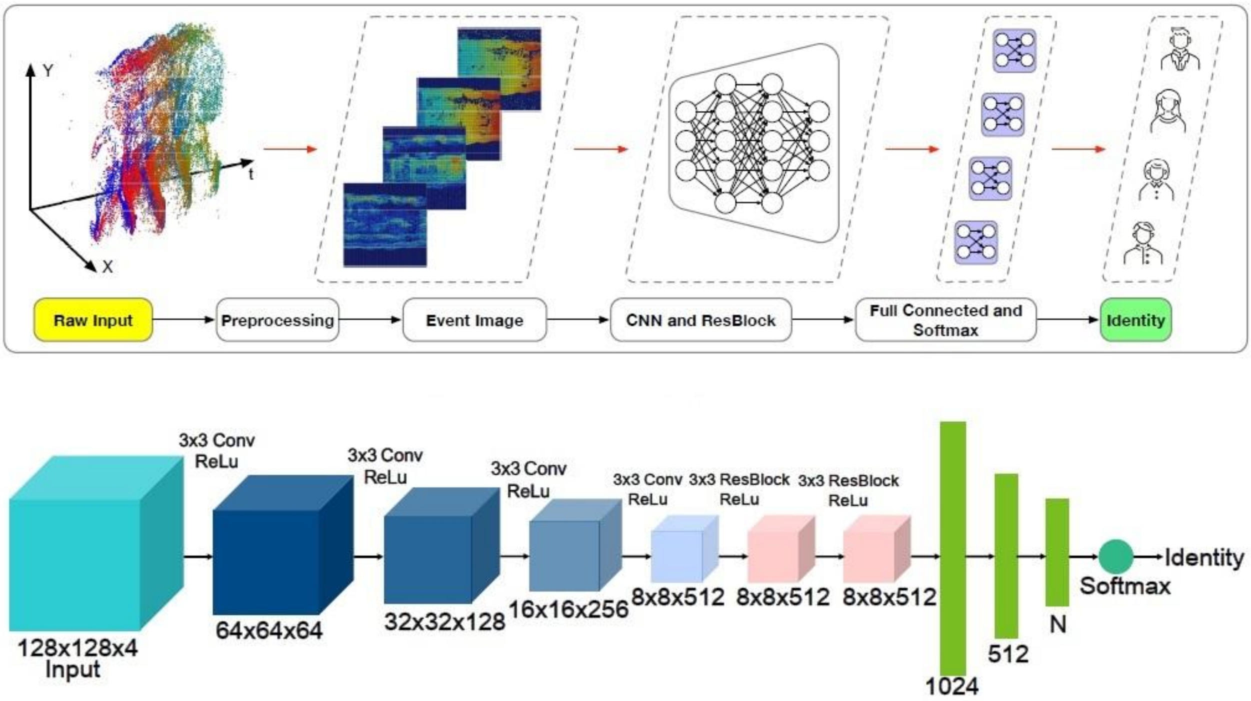

Fig 6: EV-Gait network architecture for gait recognition

5. A Tensor representation framework for cross-view gait recognition [21]: Learning cross-view gait with different view angles is difficult by a set of multi-linear projection matrices. To address this problem, three tensorial coupled mappings were proposed. (1) Coupled multi-linear locality-preserved criterion (CMLP) to detect the essential tensorial manifold structure via preserving local information. (2) Coupled multi-linear marginal fisher criterion (CMMF) to encode the intra-class compactness and inter-class separability with local relationships. (3) Coupled multi-linear discriminant analysis criterion (CMDA) to minimize the intra-class scattering and maximize the inter-class scattering.

Pose kinematics and pose based gait recognition methods consider moving parts around human joints [22-24]. Temporal time resolutions may be interpolated to increase the frame rate of gait sequence [25].

Datasets:

The following datasets contain limited data samples to develop and evaluate gait recognition models.

1. **CASIA-B gait dataset [16]:** 124 subjects in total, and 110 sequences per subject. Namely, there are eleven views (0^0 , 18^0 , · · · , 180^0) and ten sequences per subject for each view.
2. **OU-ISIR gait dataset [26]:** There are 4,007 subjects (2,135 males and 1,872 females) with ages ranging from one to 94 years old. For each subject, there are two sequences available, one in the gallery and the other as a probe sample. Four view angles are considered (55^0, 65^0, 75^0, 85^0).
3. **USF gait dataset [27]**: Widely-used benchmark in the gait recognition community. There are 122 subjects in total, for each of whom there are five covariates, leading to 32 possible conditions under which gait sequences can be imaged. These include two different shoe types (A and B), two carrying conditions (with or without a briefcase), two surface types (grass and concrete), two viewpoints (left and right) and two time instants. The sequences in this dataset are recorded in an outdoor scene, with more complex backgrounds, which is more close to a practical scenario.

Camera setup:

If the detection range is 5-50 meters, surveillance cameras with a spatial resolution more than 1280*720, minimum frame rate of 30 frames per sec are recommended. Multi-camera setup also makes it easy to estimate view transformations. If the detection range is less than 4 meters, Kinect sensor with SDK would be preferred as it provides both color and depth information. Experimentation with event Cameras is also a good direction to explore.

Project plan:

The primary goal could be to achieve excellent recognition accuracies with the models on the image sequences taken from lateral (0 degrees), oblique (45 degrees), and frontal (90 degrees) to the image plane. The secondary goal could be to make the model more generalised in the sense of subject clothing, shoe, backpack and occlusion. Next, experiment and come up with features or efficient feature mappings, Transfer learning, pose or depth based feature augmentation to adopt multi view gait recognition. Further, models could be tried on low resolution camera setups, crowd atmosphere.

References:

[1] Wang, Liang, et al. "Silhouette analysis-based gait recognition for human identification." *IEEE transactions on pattern analysis and machine intelligence* 25.12 (2003): 1505-1518.

[2] Y. Yang and M. Levine, “The Background Primal Sketch: An Approach for Tracking Moving Objects,” Machine Vision and Applications, vol. 5, pp. 17-34, 1992.

[3] J. Han and B. Bhanu. 2006. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 28, 2, 316–322.

[4] Erhu Zhang, Yongwei Zhao, and Wei Xiong. 2010. Active energy image plus 2DLPP for gait recognition. *Signal Processing* 90, 7, 2295–2302.

[5] K. Bashir, Tao Xiang, and Shaogang Gong. 2010. Gait recognition using gait entropy image. In *International Conference on Crime Detection and Prevention*. 1–6.

[6] Singh, Shamsher, and K. K. Biswas. "Biometric gait recognition with carrying and clothing variants." *International Conference on Pattern Recognition and Machine Intelligence*. Springer, Berlin, Heidelberg, 2009.

[7] Li, Xiang,et al. "Gait energy response function for clothing-invariant gait recognition." In *Asian Conference on Computer Vision*, pp. 257-272. Springer, Cham, 2016.

[8] K. Bashir, T. Xiang, and S. Gong, “Gait recognition without subject cooperation,” Pattern Recognition Letters, vol. 31, no. 13, pp. 2052– 2060, 2010.

[9] Ghebleh, A., & Ebrahimi Moghaddam, M. (2017). *Clothing-invariant human gait recognition using an adaptive outlier detection method. Multimedia Tools and Applications, 77(7), 8237–8257.* doi:10.1007/s11042-017-4712-z

[10] Muramatsu, Daigo, et al. "Arbitrary view transformation model for gait person authentication." 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, 2012.

[11]Makihara, Yasushi, et al. "Gait recognition using a view transformation model in the frequency domain." european conference on computer vision. Springer, Berlin, Heidelberg, 2006.

[12] J. Tang, J. Luo, T. Tjahjadi, and F. Guo, “Robust arbitrary-view gait recognition based on 3d partial similarity matching,” IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 7–22, 2017.

[13] S. D. Choudhury and T. Tjahjadi, “Robust view-invariant multiscale gait recognition,” Pattern Recognition, vol. 48, no. 3, pp. 798–811, 2015

[14] I. Rida, X. Jiang, and G. L. Marcialis, “Human body part selection by group lasso of motion for model-free gait recognition,” IEEE Signal Processing Letters, vol. 23, no. 1, pp. 154–158, 2016.

[15]Jia, Ning, et al. "Fast and robust framework for view-invariant gait recognition." 2017 5th International Workshop on Biometrics and Forensics (IWBF). IEEE, 2017.

[16] S. Yu, D. Tan, and T. 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. IEEE, 2006, pp. 441–444.

[17] Wolf, Thomas, Mohammadreza Babaee, and Gerhard Rigoll. "Multi-view gait recognition using 3D convolutional neural networks." *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.

[18]Zhang, Cheng, et al. "Siamese neural network based gait recognition for human identification." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[19] Wu, Zifeng, et al. "A comprehensive study on cross-view gait based human identification with deep cnns." *IEEE transactions on pattern analysis and machine intelligence* 39.2 (2016): 209-226.

[20] Wang, Yanxiang, et al. "EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

[21] Ben, Xianye, et al. "A general tensor representation framework for cross-view gait recognition." *Pattern Recognition* 90 (2019): 87-98.
[22] Martín-Félez, Raúl, Ramón A. Mollineda, and Javier Salvador Sánchez. "Human recognition based on gait poses." *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, Berlin, Heidelberg, 2011.
[23] Sokolova, Anna, and Anton Konushin. "Pose-based deep gait recognition." IET Biometrics 8.2 (2018): 134-143.
[24] Roy, Aditi, Shamik Sural, and Jayanta Mukherjee. "Gait recognition using pose kinematics and pose energy image." *Signal Processing* 92.3 (2012): 780-792.
[25] Makihara, Yasushi, Atsushi Mori, and Yasushi Yagi. "Temporal super resolution from a single quasi-periodic image sequence based on phase registration." Asian Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010.
[26]] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The OU-ISIR gait database: Comprising the large population dataset and performance evaluation of gait recognition,” IEEE Trans. Information Forensics and Security, vol. 7(5), pp. 1511–1521, 2012.
[27] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, “The humanID gait challenge problem: Data sets, performance, and analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27(2), pp. 162–177, Feb. 2005.