Estimation of Personal Information Based on Joint Locations Obtained by Pose Estimator

Shiho Hanashiro*, Yui Suzuki†, and Ryusuke Miyamoto†
*Department of Computer Science, Graduate School of Science and Technology, Meiji University, Japan
†Department of Computer Science, School of Science and Technology, Meiji University, Japan

***Abstract*—Estimating personal information using image information is important for analyzing actual social activities. The most popular visual cue for this purpose is facial imaging, however, it is difficult to achieve a high enough accuracy for practical use when the resolution of the facial images is not very high, and the image resolution becomes lower quite easily if the view point is far from the target. Gait may be a powerful cue to obtain accurate classification results about personal identification and gender prediction, using images captured by surveillance cameras; details of human parts are not required for gait analysis. Previous research has shown that the movement of joint locations may be a good feature for gait analysis, but joint locations cannot be estimated directly from input images. In this paper, a novel scheme for gait analysis using the movement of joint locations estimated using only input images is proposed. The experimental results showed that the proposed scheme achieved accurate results when Alpha Pose was used to obtain joint locations with a 93.13 % accuracy rate for personal identification and 93.56 % for gender prediction.**

I. Introduction

It is important to obtain personal information using visual information for several applications, such as analysis of consumer behavior, surveillance, and security. The most popular method to obtain personal information is human facial imaging that can be used to estimate age, sex and emotions amongst other factors. In recent studies, information obtained from facial images has proven to be reliable enough for serious applications, such as identification at passport control, entrances of secured buildings, and criminal databases, where highly accurate classification is indispensable.

Facial images are useful for several purposes, but they are ineffective if the resolution is not of a high enough quality. The resolution becomes lower in generic outdoor scenes, where the face images must be captured from distant viewpoints.

If only images captured from cameras distant from target humans are available, the entire human body may become a reliable visual cue for estimating personal information; to improve the accuracy of estimating personal information using whole-body images of humans, several schemes have been proposed [1]–[8]. A popular cue for this purpose is their gait, where a part-based representation [4], a biologically inspired approach [5], and an HOG-based approach [6] are used for gait analysis.

Gait analysis can also be applied for gender prediction [9]–[11] and personal identification [10], [12], [13]. These schemes that utilize image-based gait analysis are very useful for estimating personal information and a high accuracy can be obtained without any specialized sensing devices. Most schemes tend to actualize gait analysis using several types of image processing techniques, where the classification is performed using image features directly. Meanwhile, some previous schemes [14], [15] have attempted to perform gender prediction and personal identification using feature vectors created from the locations of joints in several frames. The experimental results of the previous work showed that a high accuracy rate can be obtained when the locations of the joints are estimated properly.

In the previous schemes, joint locations were measured using the Microsoft Kinect, which has the capability of obtaining pixel-wise depth information in addition to an RGB image. However, depth information is not available for practical applications, such as surveillance in outdoor scenes. To solve this problem, this study attempts to construct a novel scheme for gait analysis using joint locations obtained from generic 2D images. For this purpose, state-of-the-art pose estimation schemes based on deep learning were adopted, and the performance of gender prediction and personal identification was investigated.

II. Related work

This section details a previous scheme for gait analysis using joint motions [14], [15] and state-of-the-art pose estimation schemes based on deep learning: Open Pose [16] and Alpha Pose [17].

A. Gait analysis using joint motions

The previous scheme [14], [15] used feature vectors created from joint motions in a cycle of walking of target humans. The absolute locations of joints $\boldsymbol{p_{i,t}}$ in the input images changed drastically according to the translational movement of the target humans or camera motions. Here, i = 0 is SPINE BASE, 1 is SPINE MID, 2 is NECK, 3 is HEAD, 4 is SHOULDER LEFT, 5 is ELBOW LEFT, 6 is WRIST LEFT, 7 is HAND LEFT, 8 is SHOULDER RIGHT, 9 is ELBOW RIGHT, 10 is WRIST RIGHT, 11 is HAND RIGHT, 12 is HIP LEFT, 13 is KNEE LEFT, 14 is ANKLE LEFT, 15 is FOOT LEFT, 16 is HIP RIGHT, 17 is KNEE RIGHT, 18 is ANKLE RIGHT, 19 is FOOT RIGHT, 20 is SPINE SHOULDER, 21 is HAND TIP LEFT, 22 is THUMB LEFT, 23 is HAND TIP RIGHT, and 24 is THUMB RIGHT.

Therefore, these methods may not be suitable for creating relevant features to classify the properties of target humans.

To solve this problem, this scheme uses joint locations relative to the location of the spine base $\boldsymbol{p_{i,t}}$ represented by $\boldsymbol{q_{i,t}}$ for classification. Fig. 1 shows an overview of the extraction of the relative locations of joints using the location of the spine base. Equation 1 represents the computation to obtain the relative joint locations.

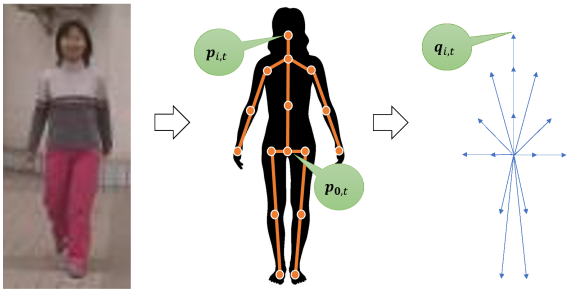

Fig. 1. Overview of extracting relative locations of joints using the location of the spine base.

$$\begin{aligned} \boldsymbol{q_{i,t}} &= \boldsymbol{p_{i,t}} - \boldsymbol{p_{0,t}} \quad (i = 1, 2, \ldots, 24) \\ &= (x_{i,t} - x_{0,t}, y_{i,t} - y_{0,t}) \\ &= \left(x'_{i,t}, y'_{i,t}\right), \end{aligned} \tag{1}$$

Where $x'_{i,t}$ and $y'_{i,t}$ represent the $x-$ and $y-$ components of the joint location $\boldsymbol{q_{i,t}}$ relative to the spine base, respectively. Through this conversion, feature vectors without the influence of the translational motion of a target human and camera motions can be extracted.

Fig. 2. Extracting a walking cycle of a target human.

After the computation of the relative locations of joints $\boldsymbol{q_{i,t}}$, a feature vector is constructed using $\boldsymbol{q_{i,t}}$ for a walking cycle. Here, the number of frames corresponding to a walking cycle differs according to the target human, as shown in Fig. 2. To obtain feature vectors whose length is the same, normalization was applied, as depicted in Fig.3.

Finally, a feature vector corresponding to a walking cycle can be obtained as follows:

$$\boldsymbol{f} = \left(x'_{1,0}, y'_{1,0}, \ldots, x'_{24,0}, y'_{24,0}, \ldots, x'_{24,N}, y'_{24,N}\right), \tag{2}$$

Where N shows the number of frames corresponding to a waking cycle.

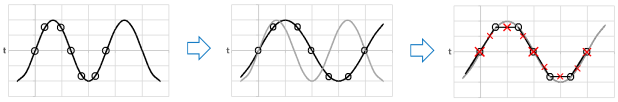

Fig. 3. Linear interpolation for feature generation.

B. Open Pose

The Open Pose was originally proposed by Cao et al. [18], which demonstrated a good accuracy and processing speed for the MPII multi-person dataset and the COCO key points challenge. This scheme adopts a bottom-up approach in which all body parts are detected independently in the first process. However, there are two major problems in the scheme, namely: creating a global connection of local parts and a huge computation effort for several combinations of body parts [19], [20].

To solve the global connection problem, Cao et al. [18] proposed a novel convolutional neural network (CNN) architecture. The architecture adopts detection of body parts using confidence maps and part affinity fields (PAFs) that estimate the connection between two body parts.

The computational problem is solved by transforming a complete graph into many bipartite graphs and applying the greedy algorithm to compute. The greedy algorithm may worsen the accuracy, but the global context from confidence maps and PAFs helps to maintain accuracy. As a result, real-time processing with high accuracy was achieved.

C. Alpha Pose

Alpha Pose is a scheme using a top-down approach, where a target human is detected before applying pose estimation to humans. This approach can reduce the number of combinations of body parts needed to be considered by the pose estimator because body parts corresponding to only a person are used to create a whole-body model. The accuracy of pose estimation can be easily improved because different CNNs can be adopted for detection and pose estimation. However, the computation time increases according to the number of target humans in an image. Alpha Pose primarily consists of an RMPE for pose estimation and pose flow for pose tracking. Estimation performance for the COCO and MPII multi-person datasets of this scheme is excellent.

III. Estimation of Personal Information Based on Joint Locations Obtained by Pose Estimator

The proposed scheme applies motion-based gait analysis to key points extracted from an input image using a CNN-based pose estimation. Fig. 4 shows an overview of the proposed scheme to estimate human information, where gender prediction and personal identification are selected as target applications.

In the first process of the proposed scheme, the joint locations are extracted by a pose estimator, unlike in previous studies [14], [15] which used three-dimensional coordinates

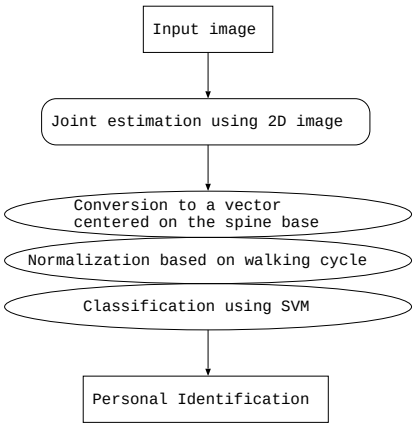

Fig. 4. Overview of estimating human information

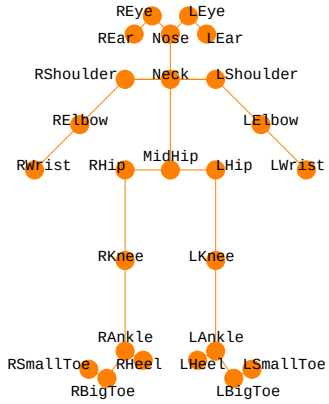

Fig. 5. Twenty five joints used in pose estimation by Open Pose.

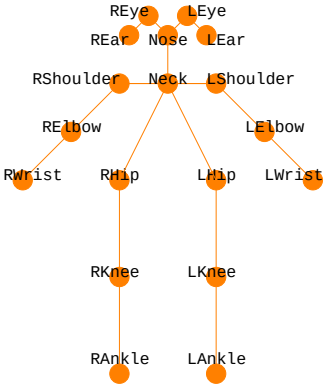

Fig. 6. Eighteen joints used in pose estimation by Alpha Pose.

obtained by Kinect v2. In the proposed scheme, two state-of-the-art schemes are tested to estimate the joint locations of a target human: Open Pose and Alpha Pose. In our implementation, 25 joints shown in Fig. 5 and 18 joints depicted in Fig. 6 are used for pose estimation by Open Pose and Alpha Pose, respectively.

Once joint locations are estimated by a pose estimator, relative locations are computed the same way as in the previous scheme, as explained in Section II. When Open Pose is used as the pose estimator, MidHip is treated as the origin for computing the relative locations. However, the output from Alpha Pose does not have MidHip, which corresponds to the location of the spine base in the previous scheme. Therefore, the center point between RHip and LHip is treated as the origin of the relative locations in this case.

To construct feature vectors corresponding to a walking cycle, the start and end points of a walking cycle are defined manually. The extracted feature vectors are normalized as the number of samples in a walking cycle becomes 40, using interpolation in the same way as in the previous scheme. After normalization, a feature vector is constructed by concatenating the time-series locations of the joints, and an SVM-based classifier is trained using the feature vector.

IV. Evaluation

This section describes the dataset used in the evaluation, and the classifier training method, and evaluates the performance of personal identification and gender prediction.

A. Dataset and training

In the evaluation, the CASIA Gait Database [21] was used, which includes 240 sequences of walking 20 persons: 12 sequences for a person whose camera angles are $0°$,$45°$, and $90°$. The number of persons was 20, of which 15 were male and five were females. The total number of still images included in the dataset was 19 319. Fig. 7 shows example images included in the database.

For the evaluation of personal identification and gender prediction, the training samples consisted of successive images corresponding to a walking cycle. Table I lists the number of available samples for each camera angle. The number of samples differs according to the camera angles because the number of cycles included in the test images was not constant, owing to the distance between the target humans and the camera. In some images, it was difficult to extract successive images for only a cycle, and the number of samples for $0°$ was significantly smaller than for other angles. A total of 233 samples were used for testing, and 208 samples were used for training.

TABLE I
THE NUMBER OF SAMPLES.

	$0°, 45°, 90°$	$0°$	$45°$	$90°$
test	233	40	107	86
train	208	35	97	76

Liblinear [22] was used to construct a final classifier after extracting the joint locations. For gender prediction, a two-class classification was simply adopted. For personal identification, multiclass classification based on one-vs.-all approaches implemented in liblinear.

B. Personal Identification

Table II shows the classification accuracy when the joint locations were estimated using Open Pose. These results showed that the best accuracy was approximately 89.53 % when the camera angle was $90°$. The worst accuracy was 50.00 % when the camera angle was $00°$. If samples from all angles were used for training and evaluation, the accuracy was 76.39 %.

Table III shows the classification accuracy when the joint locations were estimated using Alpha Pose. The best accuracy was 94.19 % at $90°$ and at worst was 87.50 % at $0°$. If samples from all angles were used for training and evaluation, the accuracy was 93.13 %.

(a)0° (b)45° (c)90°

Fig. 7. Example images of the CASIA dataset.

(b)0° (b)45° (c)90°

Fig. 8. Estimation results by Open Pose.

(a)0° (b)45° (c)90°

Fig. 9. Estimation results by Alpha Pose.

These results showed that joint locations extracted by Alpha Pose were more suitable than Open Pose, and the accuracy was quite good for personal identification in outdoor scenes.

Fig. 10 shows the confusion matrix when the Alpha Pose was used. As shown in the results, the accuracy was quite good.

TABLE II
CLASSIFICATION ACCURACY WHEN JOINT LOCATIONS WERE ESTIMATED BY OPEN POSE.

Camera angleθ	$0°, 45°, 90°$	$00°$	$45°$	$90°$
Accuracy(%)	76.39	50.00	79.44	89.53

C. Gender prediction

Table IV shows the accuracy of the gender prediction when joint locations were obtained using Open Pose. The best and worst accuracies were 92.52 % at the $45°$ camera angle and 75.00 % at the $00°$ camera angle, respectively. If samples from all angles were used for training and evaluation, the accuracy was 89.70 %. Table V shows the accuracy of gender prediction when joint locations are obtained using Alpha Pose. The best and worst accuracies were 100.00 % at the $45°$ camera angle and 67.50 % at the $00°$ camera angle, respectively. If samples from all angles were used for training and evaluation, the accuracy was 93.56 %.

TABLE III
CLASSIFICATION ACCURACY WHEN JOINT LOCATIONS WERE ESTIMATED BY ALPHA POSE.

Camera angleθ	$0°, 45°, 90°$	$0°$	$45°$	$90°$
Accuracy(%)	93.13	87.50	88.79	94.19

Alpha Pose showed better accuracy for the $45°$ and $90°$ camera angles, but it was worse when the camera angle was $0°$.

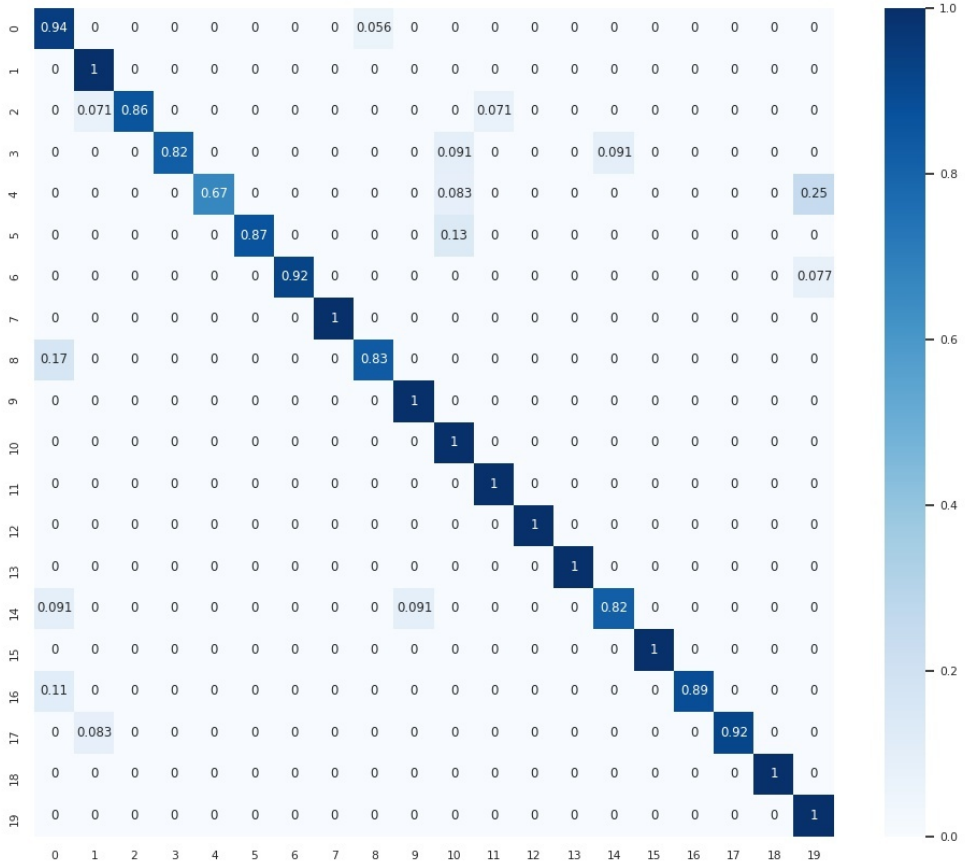

Fig. 10. Confusion matrix when Alpha Pose was used for key point extraction.

TABLE IV
THE ACCURACY OF GENDER PREDICTION WHEN JOINT LOCATIONS WERE ESTIMATED BY OPEN POSE.

Camera angleθ	$0°, 45°, 90°$	$00°$	$45°$	$90°$
Accuracy(%)	89.70	75.00	92.52	91.86

TABLE V
THE ACCURACY OF GENDER PREDICTION WHEN JOINT LOCATIONS WERE ESTIMATED BY ALPHA POSE.

Camera angleθ	$0°, 45°, 90°$	$0°$	$45°$	$90°$
Accuracy(%)	93.56	67.50	96.26	100.00

D. Discussion

The classification accuracy at a camera angle of $0°$ was the worst for both pose estimation schemes. The primary reason seems to be the lack of samples for the $0°$ camera angle: the number of images available for our evaluation was not very large. To evaluate the estimation accuracy of personal information at the $0°$ camera angle, more training and testing samples for this angle must be used.

The classification accuracy improved with an increase in the camera angle for both tasks. This result was attributed to the visibility area of the human body, where nearly half of the human body was invisible when the camera angle is $0°$ and $45°$. Fig. 11 shows failure examples for pose estimation: the left forearm could not be estimated in Fig. 11 (a), and the right arm was predicted at the opposite side of the actual location in Fig. 11 (b). These failures may degrade the estimation performance of the personal information using joint motions. To enhance the classification accuracy, the robustness of pose estimation against partial occlusion by the human body itself

(a) An image of F recorded at the 45° angle. (b) An image of C recorded at the 0° angle.

Fig. 11. Failure examples of pose estimation.

should be improved through future research.

V. Conclusion

This paper proposed a novel scheme to estimate the joint locations of personal information obtained only from input images. To construct feature vectors using joint locations, normalization and interpolation techniques were applied in the same way as in previous schemes [14], [15]. For the extraction of joint locations, Open Pose and Alpha Pose were adopted, which are capable of estimating the locations of body parts using only an input image without any depth sensors.

To evaluate the classification accuracy of the proposed scheme, a quantitative evaluation using the CASIA dataset including successive images of 20 humans walking outside: 15 males and five females. At the evaluation, the locations of 25 and 18 joints were estimated using Open Pose and Alpha Pose, respectively. After obtaining the joint locations, a feature vector corresponding to a walking cycle was created using the proposed scheme to construct a classifier using liblinear.

Experimental results show that the proposed scheme achieved accurate results using only two-dimensional images as input: 93.13 % for personal identification and 93.56 % for gender prediction. The best accuracy was obtained when Alpha Pose was used to estimate joint locations. The accuracy is expected to improve if the pose estimation is improved to reduce estimation failures at pose estimation explained in Section IV-D.

References

[1] Y. Ohtaki, K. Sagawa, and H. Inooka, "A method for the long-term gait assessment utilizing accelerometers and gyroscopes," *Transactions of The Japan Society of Mechanical Engineers C*, vol. 67, pp. 782–788, 2001.

[2] K. Sudo, J. Yamato, A. Tomono, and K. Ishii, "Fusing multiple sensor information for a gender determining system," *IEICE Trans. on Information and Systems*, vol. J83-D1, pp. 882–890, 2000.

[3] R. K. Begg, "Support vector machines for automated gait classification," *IEEE Trans. Biomedical Engineering*, vol. 52, pp. 828–838, 2005.

[4] L. Cao, M. Dikmen, Y. Fu, and T. Huang, "Gender recognition from body," in *Proc. ACM Int. Conf. Multimedia*, 2008, pp. 725–728.

[5] G. Guo, G. Mu, and Y. Fu, "Gender form body: A biologically-inspired approach with manifold learning," in *Proc. Asian Conf. Comput. Vis.*, 2009, pp. 236–245.

[6] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," in *Proc. Int. Conf. Comput. Vis. Workshop*, 2009, pp. 1235–1242.

[7] D. Adjroh, D. Cao, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *Proc. IEEE Int. Workshop on Information Forensics and Security*, 2010.

[8] D. Cao, C. Chen, D. Adjeroh, and A. Ross, "Predicting gender and weight from human metrology using a copula model," in *Proc. IEEE Int. Conf. Biometrics: Theory, Applications and Systems*, 2012, pp. 162–169.

[9] J. Yoo, D. Hwang, and M. Nixon, "Gender classification in human gait using support vector machine," in *Proc. Int. Conf. Advanced Concepts for Intelligent Vision Systems*, 2005, pp. 138–145.

[10] L. Lee and W. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2002, pp. 148–155.

[11] H. Mannami, Y. Makihara, and Y. Yagi, "Gait-based categorization and feature analysis of gender and age," *IEICE Trans. on Information and Systems*, vol. J92-D, no. 8, pp. 1373–1382, 2009 (in Japanese).

[12] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, 2003.

[13] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "Eigengait: Motion-based recognition of people using image self-similarity," in *Proc. Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 284–294.

[14] R. Miyamoto and R. Aoki, "Gender prediction by gait analysis based on time series variation on joint position," *J. Systemics, Cybernetics and Informatics*, vol. 13, no. 3, pp. 75–82, 2015.

[15] R. Aoki and R. Miyamoto, "Personal identification based on feature extraction using motions of a reduced set of joints," in *Proc. IEEE Control, Decision and Information Technologies*, 2016.

[16] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2021.

[17] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2353–2362.

[18] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields." 2017, pp. 1302–1310.

[19] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele, "Deepcut: Joint subset partition and labeling for multi person pose estimation." 2016, pp. 4929–4937.

[20] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model." 2016, pp. 34–50.

[21] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhoutte analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1505–1518, 2003.

[22] "https://www.csie.ntu.edu.tw/ cjlin/liblinear/."