applied sciences

MDPI

Article

A Study on the Application of Walking Posture for Identifying Persons with Gait Recognition

Yu-Shiuan Tsai * and Si-Jie Chen

Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung City 202, Taiwan
* Correspondence: ystsai@mail.ntou.edu.tw

Abstract: In terms of gait recognition, face recognition is currently the most commonly used technology with high accuracy. However, in an image, there is not necessarily a face. Therefore, face recognition cannot be used if there is no face at all. However, when we cannot obtain facial information, we still want to know the person's identity. Thus, we must use information other than facial features to identify the person. Since each person's behavior will be somewhat different, we hope to learn the difference between one specific human body and others and use this behavior to identify the human body because deep learning technology advances this idea. Therefore, we used OpenPose along with LSTM for personal identification. We found that using people's walking posture is feasible for identifying their identities. Presently, the environment for making judgments is limited, in terms of height, and there will be restrictions on distance. In the future, using various angles and distances will be explored. This method can also solve the problem of half-body identification and is also helpful for finding people.

Keywords: deep learning; human skeleton; identity recognition; walking posture; behavior recognition; LSTM; single camera; OpenPose

check for updates

Citation: Tsai, Y.-S.; Chen, S.-J. A Study on the Application of Walking Posture for Identifying Persons with Gait Recognition. *Appl. Sci.* **2022**, *12*, 7909. https://doi.org/10.3390/app12157909

Academic Editor: Zimi Sawacha

Received: 3 July 2022
Accepted: 4 August 2022
Published: 7 August 2022

1. Introduction

Behavior recognition is generally used to recognize various movements, such as running, jogging, and waving. Behavior recognition is often a series of continuous images, and the general convolutional neural network uses static and single pictures to perform operations. Before the long short-term memory (LSTM) network could calculate continuity pictures [1], neural networks could only perform calculations with a single image. The LSTM network is sufficient to obtain long-term and short-term memory. It has a better solution to a problem that requires timeliness than the general recurrent neural network (RNN), such as continuous action.

Dollar et al. proposed the behavior recognition method in 2005 [2]. However, simple 2D data cannot identify behavior, so it has been proposed to use space and temporal data for behavior recognition. Cohen et al. in 2013, proposed a new behavior recognition method, which used the difference between hand posture (gesture) and the change in foot posture when performing movements. Therefore, a 3D convolutional neural network was proposed to learn time-space data [3]. Meanwhile, a new continuous neural network architecture, long-term recurrent convolutional networks (LRCNs), was proposed in 2015. Combining convolutional neural networks and long- and short-term memory networks enable data to be trained simultaneously [4]. Zhang proposed using 3D convolutional neural networks for behavior recognition [5]. In body identification, RFID is a very convenient technology for body identification. However, the principle of RFID is to use radio frequency signals and wireless transmission to identify the object. The hardware required is an RFID signal receiver and an electronic tag. The method used is that when the tag is close to the signal receiver, they will start transmitting to each other. After receiving the signal, the signal will be sent back to the system for identification of the label, and the label can be processed simply. The advantage of RFID is that it can transmit data without direct contact and can also read multiple signals simultaneously, such as a yoyo card. When the yoyo card is close to the card reader, it will be identified by the label on the yoyo card. After identification, the transaction with this identity is carried out. Zumsteg et al. proposed to read the FRID system in a specific space in 2018 [6], and Greene et al. proposed using RFID to identify patient tags [7].

Face recognition is also a type of body recognition. In recent years, the technology of face recognition has become more and more sophisticated. The accuracy rate is getting higher and higher, and the recognition speed is getting faster and faster. Face recognition has gradually become a kind of body recognition. Representatives and face recognition are also used in many ways. In the research on face recognition, Bronstein et al. proposed three-dimensional face recognition [8], and Naseem et al. used linear regression classification for face recognition [9]. Schroff et al. used FaceNet [10], which can directly project the face image into Euclidean space and use its distance to represent the similarity of the face to those in the database. Liu et al. used face recognition under the open protocol SphereFace [11], using convolutional neural networks, and proposed a new angular-softmax to solve the angle problem. Wang et al. proposed the face recognition method CosFace [12], in which they proposed the loss function angular-softmax to solve the angular problem using the convolutional neural network. The loss function called significant margin cosine loss was proposed.

From this, we can see that the use of facial features for body recognition is now more and more commonly used, and the accuracy is good, but it still has many shortcomings. As the name suggests, face recognition requires a face to perform recognition. Under some particular circumstances, there may be a faceless situation, and there are many factors that can have an effect in addition to this situation. Therefore, using the face for identification will vary in success due to various factors. In addition to the amount of data when face recognition is generally used, most training data are obtained without backlighting and sufficient light, most of which are frontal on the capture lens. However, in the environment, there are various problems. The camera that takes the face photos often has insufficient light, backlighting, and capturing angles, and must deal with obstructions (hair, glasses, accessories, etc.). All of these will affect the accuracy of face recognition in this environment.

Human skeleton detection has always been challenging because humans have dynamic bodies, changing shape during different actions. Hence, it is difficult to identify the human body's skeleton correctly. In recent years, significant progress has been made. For example, probabilistic image segmentation methods were used to outline objects in a picture that were clearly marked [13]. Yang proposed a new method to solve joint detection and preprocess human body features. The method of estimating a person's posture [14] uses the relationships among the human body parts in space to encode movements, and the authors constructed a model architecture accordingly. Ouyang et al. used deformation, appearance mixture type, and visual appearance score to build a depth model [15]. In 2015, Pfister et al. proposed evaluating the human body's posture in multiple frames [16]. The convolution architecture and a heatmap were used to make the effect better. Newell et al. proposed stacking hourglasses for evaluating human posture, to understand the corresponding relationships between the human body and space. There is repeated use of top–down and bottom–up processes in the internal model, so that the performance of the network architecture improves. Cao et al. proposed the concept of OpenPose [17]. OpenPose can grasp the joints of the human body and connect them to form a complete human skeleton, making it useful for obtaining information about the human body. A fast and accurate method will be conducive if it is used in body recognition. After much research, the posture of the human body, which was initially considered difficult, can be grasped more accurately. The accuracy of the posture of the human body is getting higher and higher, which also makes human behavior recognition have better accuracy.

Heo et al. proposed a method of using foot pressure for identity recognition [18], and used the properties of the rotation matrix to apply various angles. This research used foot width, height, foot pressure, and an embedded system to build a rapid user identification system. Shaik used gait recognition to distinguish the differences in each person's gait for identity recognition [19]. This study output each video frame into 18 feature points and used KNN for classification. Wang et al. proposed using two-channel convolutional neural networks and support vector machines for gait recognition [20]. They proposed a new representation of gait features captured by the interval frame method and a convolutional neural network with two channels.

In the existing identity recognition methods, particular image-based recognition targets require specific contexts. For example, it is challenging to perform identity recognition when face images are unavailable—e.g., when masks and sunglasses cover the face. In addition, the walking posture of each person, such as the frequencies and heights of hand and foot movements, may be habitual. Therefore, we aim to identify a person by his or her walking posture. We can also help determine a person's identity when he or she deliberately covers his or her face in certain behaviors, such as theft or other crimes. In our study, we found that different people do walk differently.

Moreover, although there may be other small movements (grabbing the head, etc.) when walking, the overall frequencies and heights of hand and foot movements are consistent due to habit. Regarding people's emotions that may be involved in walking, we believe if there is variability in walking posture, it will be reflected in the positions of the 18 feature points (the head, arms, and legs). An individual's feature points usually change due to emotions when walking, which can be confirmed in most training sets.

If we want to use information other than the face to identify a person's identity, the first thing that comes to mind is the way he or she moves. Everyone will be somewhat different when doing the same action. If we can find the differences between these movements, it will be of great benefit for automatic judgments, and when it comes to the movements that most people will naturally do, walking is key. Walking depends on many factors, such as height, weight, and specific diseases. Therefore, many factors will cause changes in walking posture. If behavioral recognition is used, it is not necessary to face the camera precisely as in face recognition, making detection quicker and more convenient. Therefore, this study aimed to use human walking behavior for identity recognition with OpenPose. We first recorded the walking postures of multiple people by ourselves and performed image pre-processing to facilitate the subsequent experiments. After that, we obtained skeleton feature points through OpenPose and used these data for training.

2. Material and Methods

The procedure of this experiment can be divided into seven parts (Figure 1). (1) Use the GoPro Hero5 camera to collect the training data. (2) After collecting the data, clean the data and cut out the unnecessary parts (such as turning around and no-person images) to facilitate the subsequent experiments. (3) After removing the unnecessary parts, use OpenPose to detect the skeleton, obtain the human body's feature points (Figure 2), and use these data for training. (4) The default is 30 fps, meaning 30 frames are selected as a data set. (5) Since the starting positions of the selected data are different, we need to normalize the action so that the starting positions of each group are from the same place. (6) The model must be trained after the final data preprocessing. We implemented LSTM on the Tensorflow platform for training and testing. (7) Use the model to perform final identification.

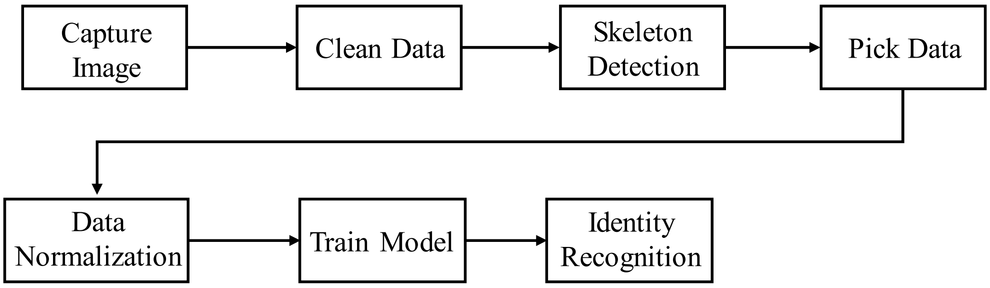


Figure 1. Architecture flowchart.

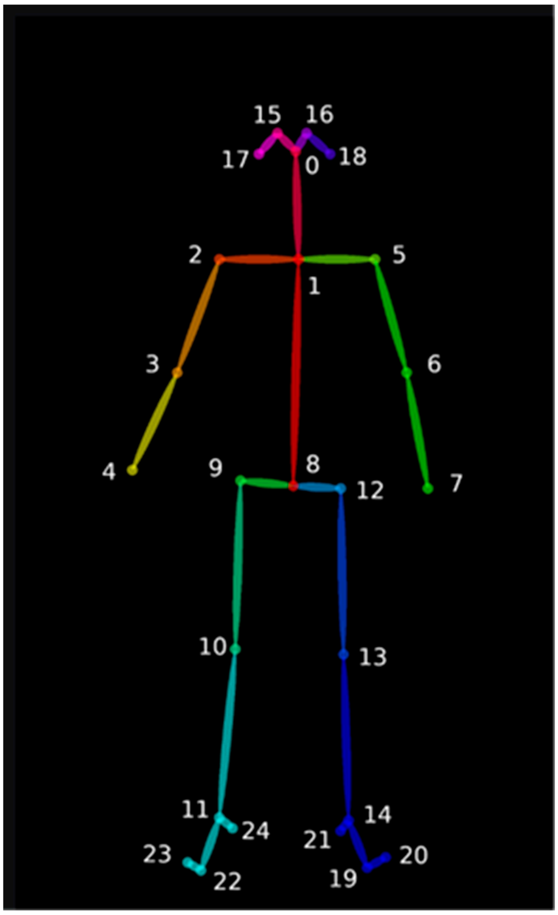


Figure 2. Twenty-five feature points were detected by OpenPose [17]. The points represent as follows: 0 is the nose; 1 is the neck; 2 is the right shoulder; 3 is the right elbow; 4 is the right wrist; 5 is the left shoulder; 6 is the left elbow; 7 is the left wrist; 8 is the middle hip; 9 is the right hip; 10 is the right knee; 11 is the right ankle, 12 is the left hip; 13 is the left knee; 14 is the left ankle; 15 is the right eye; 16 is the left eye; 17 is the right ear; 18 is the left ear; 19 is the left big toe; 20 is the left small toe; 21 is the left heel; 22 is the right big toe; 23 is the right small toe; 24 is the right heel.

- Hardware and equipment

The equipment used in this experiment was a GoPro Hero5, which has 4K resolution and a built-in linear correction function. Therefore, the images obtained will not be distorted. First, we used the GoPro Hero5 camera to create a training video of the subject. As shown in Figure 3, we set up the tripod vertically at a distance of 3.5 m from the subject. After setting up the camera, we asked the subject to walk from the far right side of the camera screen to the left side. After walking to the left side, we continued to walk forward until the subject disappeared from the screen.

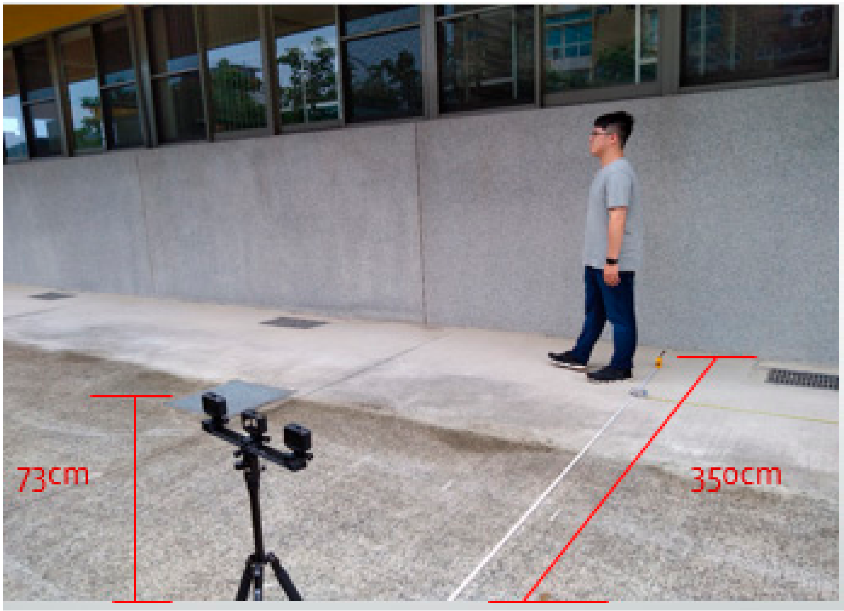


Figure 3. Experimental architecture diagram.

Furthermore, after the character disappeared, we asked the subject to turn 180 degrees and walk back. Next, the subject reappeared in the field of view and then walked to the far right side of the camera until the subject disappears from the screen. Finally, after about 10 min of video capture, the subject was asked to perform the same action again. This step was completed after five subjects were filmed. The resolution of the recorded video was 1920 × 1080 and 60 fps.

- Video cleaning

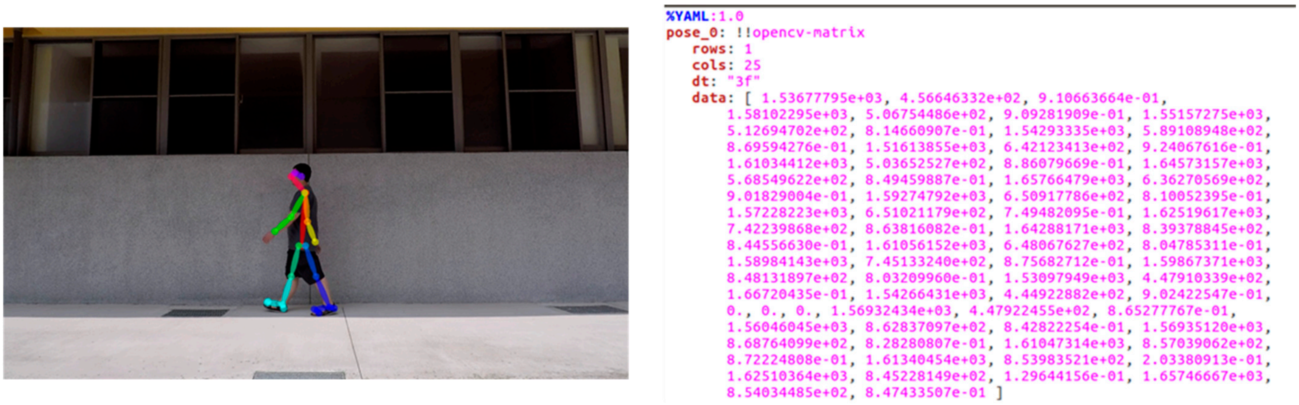
After the image was taken, we had to preprocess (video cleaning) the filmed video to ensure the final data were correct. First, because the film is a repeated video of the same behavior of walking to the left and walking to the right, and because there was a short period when the person disappeared during the two behaviors, the film had to be cut into several segments of varying length. The rules for these cuts were to divide the rightmost walk into the first category and the leftmost walk into the second category, and finally, to cut out the video of the time when the character disappears. After completing the above actions, two categories of different video lengths were obtained. Thus, twenty categories of films were obtained after the experiment with ten people. Since the turning part may be filmed during filming, we also cut out the turning part. In addition, the part that disappeared from the picture had to be cut out. Figure 4 shows that after the video was cleaned, everyone was divided into two categories: one with walking to the left and the other with walking to the right.

- Skeleton Detection

After that, we wanted to use the differences in each person's walking style to make a judgment because we needed to perform body recognition, and OpenPose can extract the key points of the human body. Therefore, we used OpenPose to detect the skeleton in the above videos. Figure 5a shows the video of step 1 using OpenPose to capture the nodes. It can be seen that OpenPose can accurately capture the joints of the human body, which is very helpful for us in obtaining information about the human body. Figure 5b shows a keypoint type yml file for each frame. The yml file contains various information about the joints captured by OpenPose, showing the number of people on the screen and the locations of the human joints captured by OpenPose. The first number represents the x-coordinate of the numbered joints, the second number represents the y-coordinate of the numbered joints, and the third number represents the number of points. Since not all frames can catch the character, we deleted the frame if no critical points in the frame were caught.

(**a**) (**b**)

Figure 4. (**a**,**b**) Types of left and right walking.



```
%YAML:1.0
pose_0: !!opencv-matrix
   rows: 1
   cols: 25
   dt: "3f"
   data: [ 1.53677795e+03, 4.56646332e+02, 9.10663664e-01,
       1.58102295e+03, 5.06754486e+02, 9.09281909e-01, 1.55157275e+03,
       5.12694702e+02, 8.14660907e-01, 1.54293335e+03, 5.89108948e+02,
       8.69594276e-01, 1.51613855e+03, 6.42123413e+02, 9.24067616e-01,
       1.61034412e+03, 5.03652527e+02, 8.86079669e-01, 1.64573157e+03,
       5.68549622e+02, 8.49459887e-01, 1.65766479e+03, 6.36270569e+02,
       9.01829004e-01, 1.59274792e+03, 6.50917786e+02, 8.10052395e-01,
       1.57228223e+03, 6.51021179e+02, 7.49482095e-01, 1.62519617e+03,
       7.42239868e+02, 8.63816082e-01, 1.64288171e+03, 8.39378845e+02,
       8.44556630e-01, 1.61056152e+03, 6.48067627e+02, 8.04785311e-01,
       1.58984143e+03, 7.45133240e+02, 8.75682712e-01, 1.59867371e+03,
       8.48131897e+02, 8.03209960e-01, 1.53097949e+03, 4.47910339e+02,
       1.66720435e-01, 1.54266431e+03, 4.44922882e+02, 9.02422547e-01,
       0., 0., 0., 1.56932434e+03, 4.47922455e+02, 8.65277767e-01,
       1.56046045e+03, 8.62837097e+02, 8.42822254e-01, 1.56935120e+03,
       8.68764099e+02, 8.28280807e-01, 1.61047314e+03, 8.57039062e+02,
       8.72224808e-01, 1.61340454e+03, 8.53983521e+02, 2.03380913e-01,
       1.62510364e+03, 8.45228149e+02, 1.29644156e-01, 1.65746667e+03,
       8.54034485e+02, 8.47433507e-01 ]
```

(**a**) (**b**)

Figure 5. (**a**) OpenPose skeleton; (**b**) the yml file obtained from OpenPose.

- Frame selections

The data needed for this experiment were 18 coordinates (x, y) output by OpenPose, and 30 frames as a group of movements. We took these actions as training data and divided the same person into two categories: walking to the left and walking to the right, and then give them labels. The 60 frames were cut into two training data sets by odd and even means. The advantage of this approach is that we can shorten the 60 fps video to 30 fps. In addition, we can use fewer data to represent the same time information.

Figure 6 shows a schematic diagram of a group of data, where each color represents a frame of information, and we visualize 30 frames from 2 s with odd and even numbers. Therefore, each data group will have 30 lines of data, and each line represents a frame to capture the character's joints. As a group of feature points has 25 coordinates, each coordinate has x and y values, so that each line has 50 numbers. In the end, there will be a total of 30 × 50 numbers in a set of data, making 1500 numbers in total.

- Data normalization

Our training data were taken with a GoPro Hero5 camera at a fixed distance. In order to reduce the error caused by different distances, we fixed the distance of people at 3.5 m (Algorithm 1). Additionally, we solved the problem of inconsistent starting positions for each group of training data. The normalization was according to Equation (1). To make the starting position of each group consistent, we first selected a reference point (the reference point selected in this study was the first frame number 1 joint). We set this point as the origin ($x = 0$, $y = 0$) and mapped this point to a positioning point (c_x, c_y). The joints of the first frame, number 1, are (X_0, Y_0); X_{ij} represents the x-coordinate of the jth point of the i-frame; and Y_{ij}

represents the y-coordinate of the jth point of the i-frame. We let $X_{ij}{}'$ be the corrected value of X_{ij}, and $Y_{ij}{}'$ be the corrected value of Y_{ij}. The total data were 60 frames, each with 25 points of X and 25 points of Y. For clarity, we have included the algorithm below.

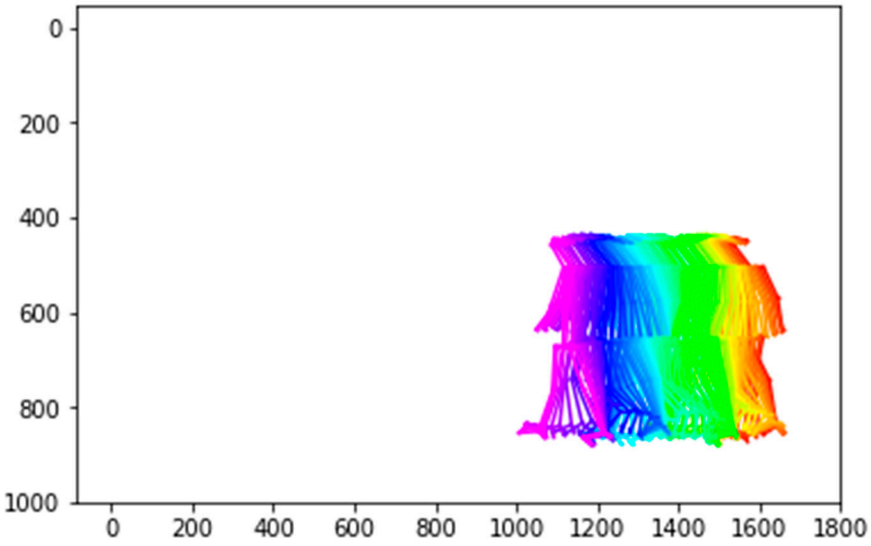


Figure 6. Schematic diagram of walking data.

Algorithm 1. Data normalization algorithm

```
Step 1. Find the first frame number 1 point's x coordinate and y coordinate as a datum point.
Step 2. Set a positioning point.
Step 3. From the first to the last frame, for every point in each frame, we subtract by datum point and add by positioning point.
Let datum point = $(X_0,\ Y_0)$, positioning point = $(c_x, c_y)$
If class is left then an = −1, else $k$ = 1
for $i$ = 1 to 60
        for $j$ = 1 to 25
```

$$\begin{aligned} X'_{ij} &= \left(X_{ij} - X_0 \times k + c_x\right) \\ Y'_{ij} &= \left(Y_{ij} - Y_0 + c_y\right) \end{aligned} \tag{1}$$

Figure 7 shows the key points of the character before and after normalization. Again, it can be seen that no matter where the character starts and ends, the starting position will be the same after normalization.

- Model training

After obtaining the training data, we applied deep learning by taking 80% of the data and the remaining 20% as test data. Next, we used Tensorflow to implement LSTM-RNN. We used two layers of LSTM-RNN; each LSTM had a hidden layer; the model architecture was shown in Figure 8.

- Identity recognition

Last, the test data were put into the training model for the final judgment. Figure 9 shows the result of putting the data into the training model and taking 60 frames to make a judgment.

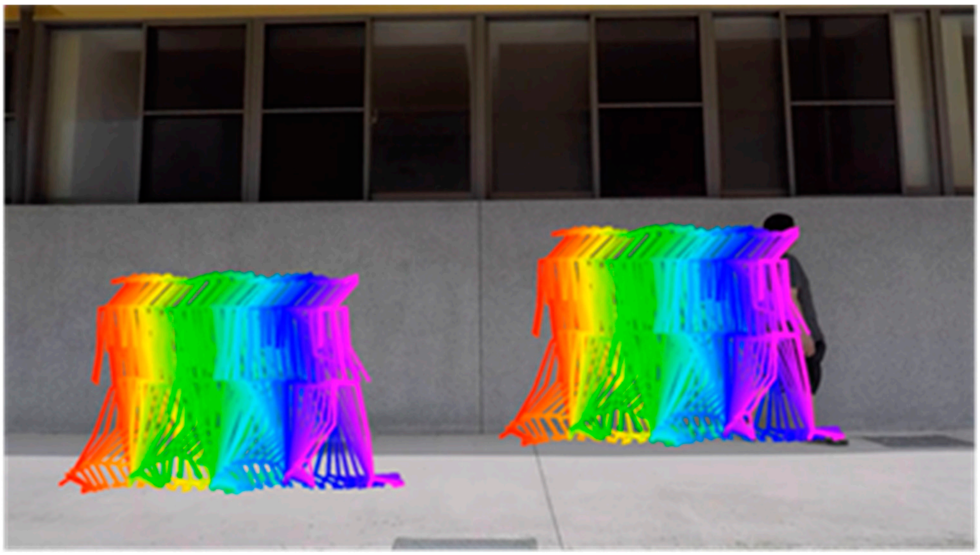


Figure 7. Schematic diagram before and after normalization.

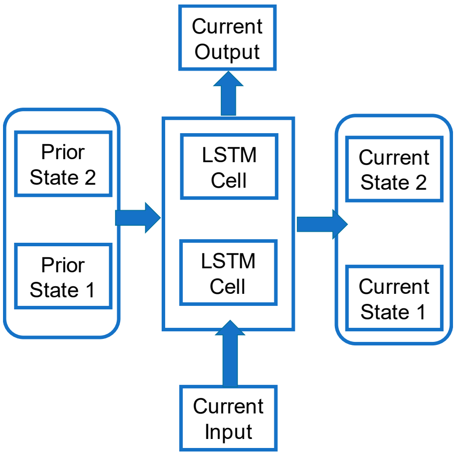


Figure 8. LSTM-RNN architecture.

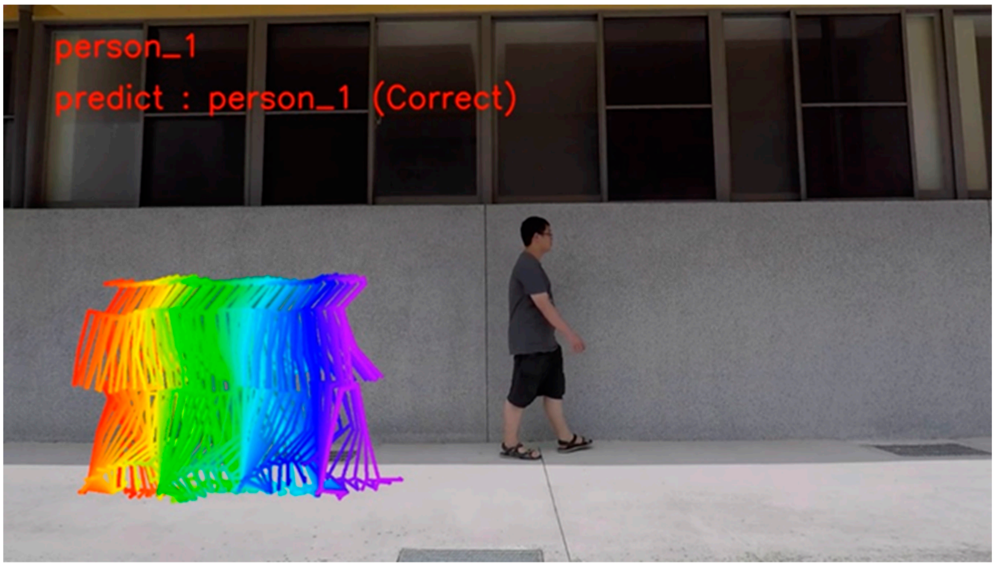
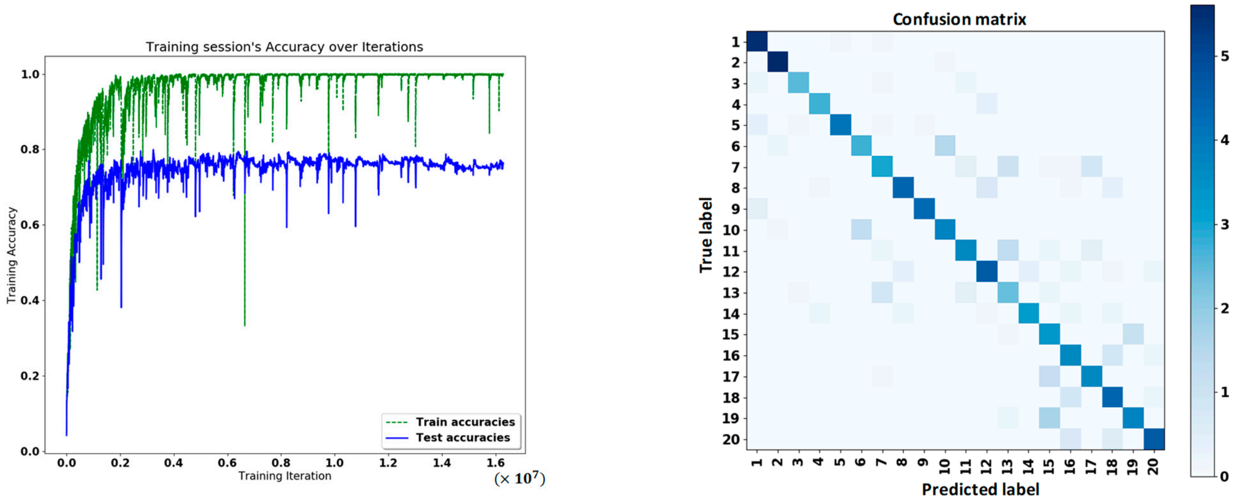


Figure 9. Identity recognition.

3. Results and Discussion

The total number of experimental participants in this study was 10—the training data for the ten individuals summed up to 3364 sets, each with 60 frames. The test data for the ten individuals added up to 862 sets, each with 60 frames. We took 60 frames of 30 fps video for training, and the data we took were all videos taken at 3.5 m, so the only processing performed on the data was the location normalization. After deep learning, the training accuracy and classification results are shown in Figure 10a,b, respectively.



(**a**) (**b**)

Figure 10. Twenty categories of 10 people. (**a**) Training accuracy results; (**b**) classification results.

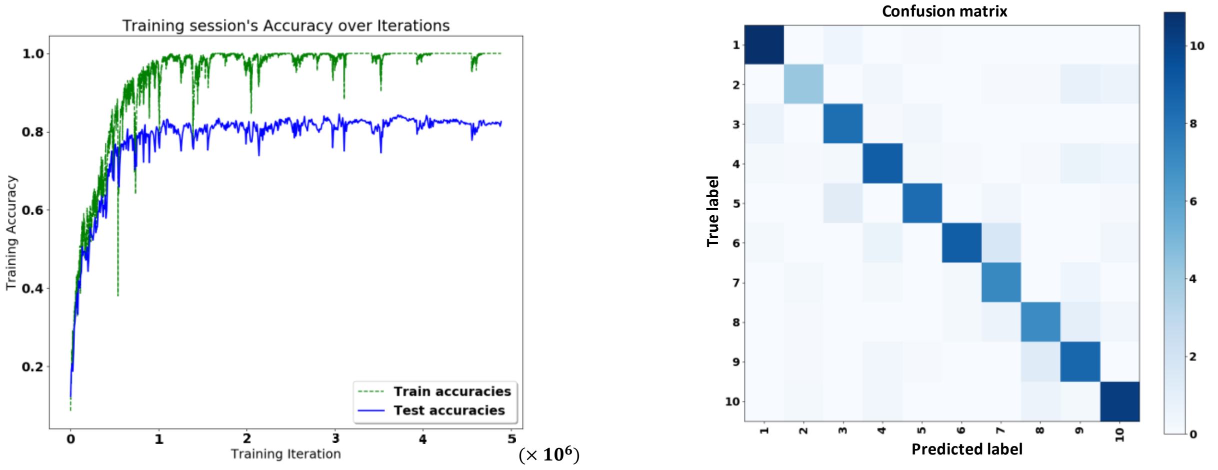
In training and testing ten categories of 10 people, we set the standard point of regularization at the left end of the screen (xx, yy). Since this experiment was to convert two categories of the same individual into one category, there are two different scenarios for the two categories: the leftmost and the rightmost. Therefore, we changed the two categories to one with Equation (2). In the category with people walking to the right, we multiplied -1 in the calculation of x and treated it as the same category as the one walking to the left.

$$\begin{aligned} X'_{ij} &= (X_{ij} - X_0 \times (-1) + xx) \\ Y'_{ij} &= (Y_{ij} - Y_0 + yy) \end{aligned} \quad (2)$$

After deep learning, the training accuracy and classification results are shown in Figure 11a,b, respectively.

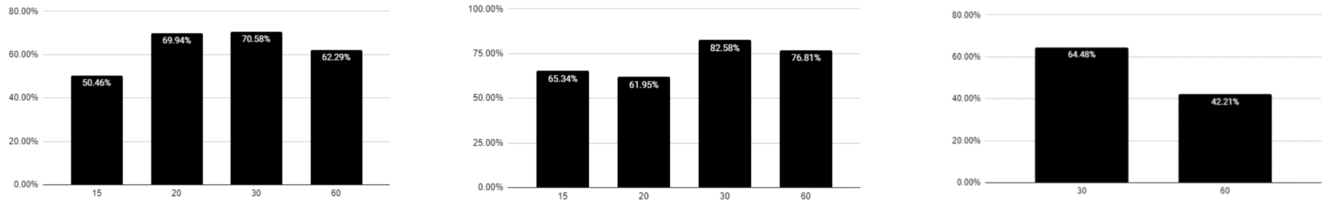
Next, we explore the image's frame rate difference in the recognition rate. Since the frame rate of GoPro was set to 60 fps, the jump points 1, 2, 3, 4, and 5 frames are equivalent to 30, 15, 20, 15, and 12 fps. Figure 12a–c show the accuracy of 30 frames, 60 frames, and 90 frames corresponding to each fps, respectively. It can be seen that the best accuracy in the case of 60 frames is 30 fps.

Next, we masked the image below the knees. This experiment aimed to determine if the part of the lower leg below the knee could still be practical if the image were occluded. We removed the OpenPose information below the knee, and the points we removed were 11, 14, 19, 20, 21, 22, 23, and 24. After removing these eight points, we used the remaining 17 points for the experiment. We took 60 frames and 30 fps videos for training and testing; the results are shown in Figure 13. We can see that the accuracy rate was about 75%, even though the calf information was gone, so we think this method is feasible for calf masking.



(**a**) (**b**)

Figure 11. Ten categories of 10 people. (**a**) Training accuracy results; (**b**) classification results.



(**a**) (**b**) (**c**)

Figure 12. The accuracy of fps for each video. (**a**) Taking 30 frames as a group, the accuracy (*y*-axis) of each fps (*x*-axis). (**b**) Taking 60 frames as a group, the accuracy (*y*-axis) of each fps (*x*-axis). (**c**) Taking 90 frames as a group, the accuracy (*y*-axis) of each fps (*x*-axis).

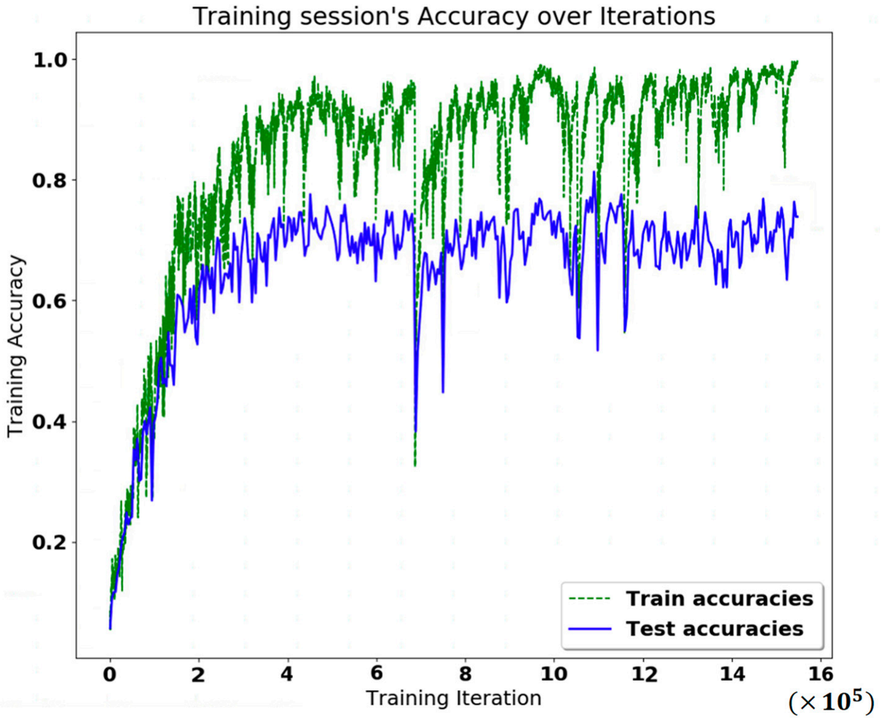


Figure 13. Training results for masking the image below the knees.

In comparison with the studies [18–20], we determined that the image of the figure is sideways and that the two directions can be combined into one direction. Therefore, our method is feasible for both directions. In addition, our method uses an outdoor context and does not require a particular sensor—i.e., only a lens is needed. In addition, we applied LSTM instead of SVM for feature classification. To further explain, in comparison with [19] (Table 1), both scenarios are side images, and our method can combine two directions into one. Therefore, our method works for both directions, not for walking to the right or the left. Although our method had less data, our accuracy rate was 82%, which is higher.

Table 1. Comparison of Shaik [19] to our method.

	Angle of View	People Number	Lower-Leg Image Occluded	Accuracy
Shaik [19]	Side (one direction)	124	No	65%
Our method	Side (two directions)	10	Yes	82%

Compared to the method [19] (Table 2), which uses the scenario indoors, our method uses the outdoor scenario. In addition, our method is passive. That is, no additional sensors are needed, and the subject does not need to step on the hardware; only a lens is needed, and the people just walk by.

Table 2. Comparison of Heo [19] to our method.

	Scenatio	Method	Hardware	Accuracy
Heo [18]	Indoor	Active	$48 * 48$ foot pressure sensors	94%
Our method	Outdoor	Passive	1 camera sensor	82%

Compared with [20] (Table 3), the method uses TCNN for feature crawling, and we use OpenPose for feature classification. Therefore, the accuracy of our method is higher.

Table 3. Comparison of Wang [20] to our method.

	Image Capture	Feature Capture	Feature Classifies	Accuracy
Wang [20]	MF-GEIs	TCNN	SVM	68%
Our method	Jumping frames	OpenPose	LSTM	82%

Regarding research limitations, since this study was conducted on campus (for training and test images), we have not yet tested many people. In the future, we will study a more significant number of people.

4. Conclusions

In this study, we proposed a new method of body identification. The accuracy of this method can reach 80%, which means that this method is effective and that the walking posture differs between people. Gender, height, weight, and even the presence or absence of disease will significantly affect the human body's walking posture, and these differences can be used to distinguish between people. If we want to perform gait recognition, in addition to the commonly used methods of RFID and face recognition, we can also use people's walking postures for recognition. This method can also solve the body recognition issue in which only half of the body can be seen. It is also helpful for finding missing persons. It will be suitable if the camera captures the person's side and the face is also covered. At present, the judgment environment is limited. It can be challenging to face various angles in terms of ground, and the distance is limited to about 3.5 m.

In the future, we will use the dual-lens images to extract the joint points and calculate the differences between two images. After the calculation is completed, the data will be projected to a distance of 3.5 m in proportion. To improve accuracy, we expect to change the parameters of the training alliance, which is also our future research aim. Moreover, there are several possibilities for expansion in the future. For example, step recognition could be used to preliminarily determine whether a person is the owner of a certain car or the owner of a certain house. Alternatively, we could add step recognition from different perspectives.

Author Contributions: Conceptualization, Y.-S.T. and S.-J.C.; methodology, Y.-S.T.; software, S.-J.C.; validation, S.-J.C.; formal analysis, S.-J.C.; writing—original draft preparation, S.-J.C.; writing—review and editing, Y.-S.T.; visualization, S.-J.C.; supervision, Y.-S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Ministry of Science and Technology (contracts MOST-110-2634-F-A49-006 and 110-2634-F-009-018).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
2. Dollár, P.; Rabaud, V.; Cottrell, G.; Belongie, S. Behavior recognition via sparse spatio-temporal features. In Proceedings of the 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China, 15–16 October 2005.
3. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
4. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Darrell, T.; Saenko, K. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
5. Zhang, R.; Ni, B. Learning Behavior Recognition and Analysis by Using 3D Convolutional Neural Networks. In Proceedings of the 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), Luang Prabang, Laos, 2–5 July 2019.
6. Zumsteg, P.; Qu, H. Reading RFID Tags in Defined Spatial Locations. U.S. Patent US9892289B2, 13 February 2018.
7. Greene, J.E.; Rulkov, N.F. RFID Markers and Systems and Methods for Identifying and Locating Them. WO Patent WO/2018/222777, 30 May 2020.
8. Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. Three-dimensional face recognition. *Int. J. Comput. Vis.* **2005**, *64*, 5–30. [CrossRef]
9. Naseem, I.; Togneri, R.; Bennamoun, M. Linear regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2106–2112. [CrossRef] [PubMed]
10. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
11. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
12. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
13. Andres, B.; Kappes, J.H.; Beier, T.; Kothe, U.; Hamprecht, F.A. Probabilistic image segmentation with closedness constraints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
14. Yang, Y.; Ramanan, D. Articulated human detection with flexible mixtures of parts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 2878–2890. [CrossRef] [PubMed]
15. Ouyang, W.; Chu, X.; Wang, X. Multi-source deep learning for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
16. Pfister, T.; Charles, J.; Zisserman, A. Flowing convnets for human pose estimation in videos. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
17. Cao, Z.; Simon, T.; Wei, S.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
18. Heo, K.H.; Jeong, S.Y.; Kang, S.J. Real-time user identification and behavior prediction based on foot-pad recognition. *Sensors* **2019**, *19*, 2899. [CrossRef] [PubMed]

19. Shaik, S. *OpenPose Based Gait Recognition Using Triplet Loss Architecture*; National College of Ireland: Dublin, Ireland, 2020.
20. Wang, X.; Zhang, J. Gait feature extraction and gait classification using two-branch CNN. *Multimed. Tools Appl.* **2020**, *79*, 2917–2930. [CrossRef]