Gait Features Extraction Technique for Identity Detection Using CNN for Surveillance Systems

Ali Aouto, Jae-Min Lee and Dong-Seong Kim,
Networked Systems Lab, Dept. of IT Convergence Engineering.
Kumoh National Institute of Technology,
Gumi-si, Gyeongbuk-do, 39177, South Korea.
Email: {ali.aouto,ljmpaul,dskim}@kumoh.ac.kr

***Abstract*—This paper proposes an identification system based onto gait recognition approach that considers distances and orientations between joints into an amount of time. The images processed by transforming the collected 3D skeleton sequences to its geometric attributes. Then, those attributes are used as inputs for pre-trained trained convolutional neural network (CNN) that was modified to suit our purpose. Experimenting the proposed system with UPCV dataset resulted 92.88% of accuracy.**

***Index Terms*—Identity Detection, Skeletal Data, Deep learning**

I. Introduction

Posture-based identification systems for gait recognition are based on processing skeleton data that describes human posture. This kind of technique is preeminent thanks to its noticeable performance. But, a decade ago, there was only a little amount of researches that adapted this approach since there was no commercialized product to be implemented for, until Microsoft announced their Kinect device as a motion sensor for their gaming counsel Xbox 360 Kinect [1].This paper discusses deriving of posture characteristics in a hierarchical geometric approach which apprehends the human posture dynamics in the space and time domains. Then, a trained convolutional neural network is fine-tuned as person classifier via a transfer learning technique, which will grant a complete control over the high level traits derived from an illustrated map of multi-scale trait.

II. Proposed scheme

This section explains the identifier that we are proposing in this paper. A CNN scans the 3D skeleton coordinates that were collected in spatial and temporal domains, doing that for more than one frame by taking temporal domain in account. First, calculating the distance and orientation of joint to joint locomotion. Secondly, calculating the standard and mean deviation at high level to represent the illustrative statistical characteristics. Finally, putting all these together to build a high dimension matrix that will be passed through the trained CNN that is deployed here as an identity classifier. Fig. 1 shows an overview of the system that is proposed in this paper.

A. Extracting Posture Characteristics

Let's say that joint h belongs to a whole set of skeleton σ. Then, that skeleton can be descried as $\sigma = \{\rho_\iota = 1 : h\}$. Also, each joint exists in the 3 dimensional space and presented as (x, y, z), where by assuming that $x = 0$, $y = 0$ and $z = 0$, we can obtain the Euclidean distance in P_{xy}, P_{xz} and P_{yz} planes [2]. Calculating the distance between joint a and joint b can be done with the following equations:

$$\begin{aligned}
\epsilon_x(a,b) &= \sqrt{(y_b - y_a)^2 + (z_b - z_a)^2},\\
\epsilon_y(a,b) &= \sqrt{(x_b - x_a)^2 + (z_b - z_a)^2},\\
\epsilon_z(a,b) &= \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2},
\end{aligned} \tag{1}$$

and to express the distance of two chosen joints we can use $m = [\epsilon_x \ \epsilon_y \ \epsilon_z]$. Our approach also considers joint orientation which is another important geometric characteristic that is used for identification, this geometric illustrates the angle that any joint results with the original axis by limbs locomotion. To do that, we are calculating by considering basic unit vector on each of $\overrightarrow{P_y}, \overrightarrow{P_z}$ and $\overrightarrow{P_x}$ by assuming $x = 0$, $y = 0$ and $z = 0$, respectively. So, the angle that results between the original axis and the vector $\overrightarrow{ab}$ is calculated by

$$\begin{aligned}
\theta_x\left(\overrightarrow{ab}, \overrightarrow{P_y}\right) &= cos^{-1}\left(\frac{\overrightarrow{ab}\cdot\overrightarrow{P_y}}{\left\|\overrightarrow{ab}\right\| \times \left\|\overrightarrow{P_y}\right\|}\right),\\
\theta_y\left(\overrightarrow{ab}, \overrightarrow{P_z}\right) &= cos^{-1}\left(\frac{\overrightarrow{ab}\cdot\overrightarrow{P_z}}{\left\|\overrightarrow{ab}\right\| \times \left\|\overrightarrow{P_z}\right\|}\right),\\
\theta_z\left(\overrightarrow{ab}, \overrightarrow{P_x}\right) &= cos^{-1}\left(\frac{\overrightarrow{ab}\cdot\overrightarrow{P_x}}{\left\|\overrightarrow{ab}\right\| \times \left\|\overrightarrow{P_x}\right\|}\right),
\end{aligned} \tag{2}$$

calculating these parameters will allow us to show the orientation between two chosen joints as $n = [\theta_x \ \theta_y \ \theta_z]$. Since this approach consider the relationship between a joint and all other joints, the amount of distance and orientation parameters for joints of a σ skeleton is $h(h-1)$ at any chosen frame τ. in our approach we consider multi-frame processing which means collecting data over a period of time, that would increase the identifying accuracy. In temporal dimension, a series of S skeletons is expressed as $\gamma = \{\sigma^\tau\}$. Then, we need to consider the statistical descriptive characteristics by

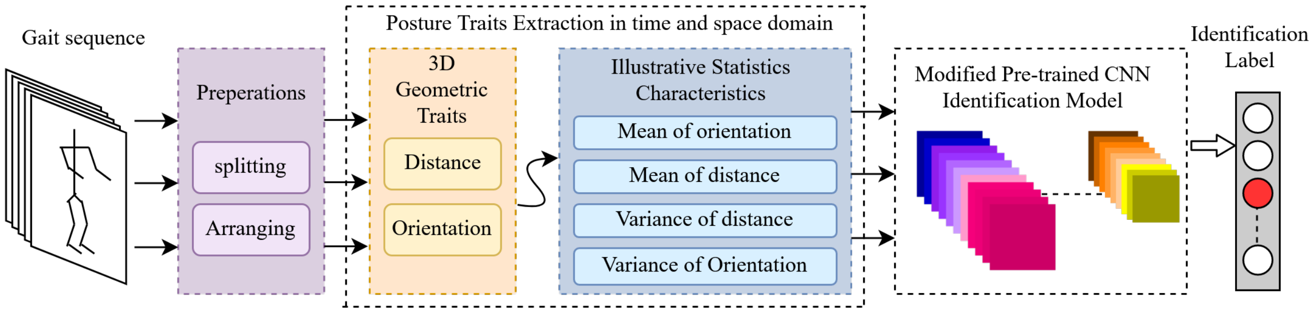

Fig. 1. Overview of the proposed CNN identification system.

calculating the standard and mean deviation, to take in count data that is collected over time as geometric characteristics matrices using the next equations

$$
\begin{aligned}
\eta_{\epsilon_{\{x,y,z\}}} &= \frac{1}{S}\sum_{\tau=1}^{S} \epsilon^{\tau}_{\{x,y,z\}},\\
\eta_{\theta_{\{x,y,z\}}} &= \frac{1}{S}\sum_{\tau=1}^{S} \theta^{\tau}_{\{x,y,z\}},\\
\xi_{\epsilon_{\{x,y,z\}}} &= \sqrt{\frac{1}{S}\sum_{\tau=1}^{S}\left(\epsilon^{\tau}_{\{x,y,z\}} - \eta_{\epsilon_{\{x,y,z\}}}\right)^2},\\
\xi_{\theta_{\{x,y,z\}}} &= \sqrt{\frac{1}{S}\sum_{\tau=1}^{S}\left(\theta^{\tau}_{\{x,y,z\}} - \eta_{\theta_{\{x,y,z\}}}\right)^2},
\end{aligned}
\tag{3}
$$

where $\eta_{\epsilon_{\{x,y,z\}}}$ is the mean distance deviation and $\eta_{\theta_{\{x,y,z\}}}$ is the mean orientation deviation, while the standard deviation is described as $\xi_{\epsilon_{\{x,y,z\}}}$ and $\xi_{\theta_{\{x,y,z\}}}$ for distance and orientation respectively. By building the higher matrix, we obtain the final higher level characteristics matrix that would be applied to the convolutional neural network identifier that will be described in the next subsection.

B. CNN Identity Detection Model

In our approach, we compare four CNN Models which are ResNet101, GoogleNet, Inception-v3 and VGG-19. We tweaked these models by modifying the input layer to suit our input and the output layer to comprehend the labels that we are aiming for.

III. Experimental Results

Here, we are using the UPCV Gait dataset [3] to test our model. It has 150 sequences of 30 people walking on a straight line in normal speed. We start the experiment by splitting up the skeleton sequences to a 1-second parts (30fps) which means we are setting $S = 30$ in the gait characteristics derivation part. Furthermore, skeleton samples are split into multiple sequences with 75% overlapping to increase the accuracy. Using stochastic gradient descent with momentum (SDGM) optimizer, the identity classifier models are modified to process the new type of characteristics set in 40 epochs. Starting with 0.01 learning rate that every 20 epochs will drop by 90%, and the mini batch size is set to 16. Analyses are performed on a low-cost benchmark system equipped with NVIDIA GeForce RTX 2070 GPU.

Table I shows the outcome of testing our proposed scheme using the UPCV dataset for identity classification while considering different already trained CNNs. Our approach resulted an accuracy of 92.88% wtih Inception-v3.

TABLE I
Comparison on Different Models

Models	Accuracy (%)
$\alpha_{\uparrow}$ + GoogleNet	87.76
$\alpha_{\uparrow}$ + VGG-19	89.35
$\alpha_{\uparrow}$ + ResNet101	91.10
$\alpha_{\uparrow}$ + Inception-v3	92.88

IV. Conclusion

This paper proposed surveillance system that can detect the identity of a person without the cooperation from person of interest. The system is based on gait posture to identify someone by extracting the spatiotemporal gait features then presenting them as geometric characteristics to be applied to the identity classifier. A modified pre-trained convolutional neural network is used as classifier in the proposed system. Evaluating the model using UPCV Gait dataset, Inception-v3 showed the highest results of 92.88% accuracy.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2020-2020-0-01612) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

References

[1] N. Saitong-in, K. Assantachai, N. Khamsemanan, and C. Nattee, “Human identification from gait analysis using microsoft kinect,” in *Proc. 2rd ICT Int. Student Project Conf.(ICT-ISPC)*, 2013, pp. 101–104.

[2] T. Huynh-The, C.-H. Hua, N. A. Tu, T. Hur, J. Bang, D. Kim, M. B. Amin, B. H. Kang, H. Seung, S.-Y. Shin, E.-S. Kim, and S. Lee, “Hierarchical topic modeling with pose-transition feature for action recognition using 3d skeleton data,” *Information Sciences*, vol. 444, pp. 20 – 35, 2018.

[3] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, “A framework for gait-based recognition using kinect,” *Pattern Recognition Letters*, vol. 68, pp. 327–335, 2015.