Person Re-identification Using 3D Data Analysis Method and Kinect Sensor

Wisrut Kwankhoom, Paisarn Muneesawang
Department of Electrical and Computer Engineering, Faculty of Engineering,
Naresuan University, Phitsanulok 65000, Thailand.
paisarnmu@nu.ac.th

***Abstract*—Automated personal identification systems, such as personal facial recognition systems and automated motor vehicle registration number checking, are examples of public security protection systems. The area of personal identification for security purposes is of growing interest for security assessment in public places, and airports, as examples, now becoming an imperative matter for research in the Internet netscape. We propose a method of immediate recognition of a subject person, based on Incremental Dynamic Time Warping (IDTW) which identifies personal gait patterns recorded via a 3D depth sensing camera such as in Microsoft's Kinect® version 2, by analyzing a dataset of gait gestures derived from a sample of 16 people. The experimental results show that the IDTW algorithm increases the efficiency of recognizing at 81%.**

***Index Terms*—Personal Identification; Gesture Recognition; IDTW; Kinect Camera; Computer Visions.**

I. INTRODUCTION

Multimedia analysis technology for indexing or explaining multimedia files provides significant advantages for information retrieval from huge databases. This technology can be applied for personal identification based on a video database. The science of personal physical behavior pattern searching and indexing is based on the biological features of a subject person that are able to be filtered from multimedia files (image or video) recording imagery of that person. Specific physical aspects of the person are analyzed and are measured and analyzed by a computer based on standard, generic data of biological features for recognition and identification of the person.

Gesture recognition is an important part of communication between humans and computers. Interactive human body movement and gesture tracking is the basis of many applications including gaming, human-computer interaction, telepresence, health-care and security [1]. Generally, the recognition result is acquired after the movement or gesture has been detected and completed.

In this paper, we propose an analysis method for recognition of gait gestures and its application to person identification.

In the person identification process, the skeletal joints of the person can be analyzed to solve the limited problem of face recognition by using the biological features of the gait gestures. There has also been research into the analysis of the human personal cycle of gait gestures [2,3]. Gait gesture analysis requires more than twenty components to identify a person, and uses the kinematics of the person; movement and rhythm of walking, for example. Compared with analyzing the human face, using gait gesture recognition is the more effective method for person identification. Person re-identification has to analyze joint of skeleton

However, the gait gesture recognition methods need video images to track the motion correctly and a segmentation algorithm to separate the human body image area from the overall media image. The human body data area is the only part of the image used to analyze gait gestures [4] and therefore must be isolated from the overall image. This is difficult due to the movement of the image in

the video. Segmenting a moving gesture from the video requires a high degree of computation and can easily be mistaken.

Given the problems of analyzing images and videos, researchers have found that using a 3-dimensional data sensor to record human gait gesture is the preferred option. In 2010, Microsoft Company produced a 3D depth sensing camera called the 'Kinect sensor' which uses X-Box 360 videogame. This player can act as a remote control for gaming by using the gesture of the players. The camera has a sensor and an application programming interface (API) for searching and representing the body as a skeletal image. It identifies the body joints of a player who is standing in front of the sensor. The Kinect camera has the great advantage of being inexpensive but easy to use and gives high accuracy of detection of the motion, in 3D [5].

This paper presents a method for person identification using gait gestures. The computational method uses the anthropometric biological features in the 3D data, captured by the Kinect 3D camera, and analyzes that data.

The method of gait pattern recognition that we propose is based on the Incremental Dynamic Time Warping (IDTW) algorithm. This an extension of the classic Dynamic Time Warping (DTW) algorithm by providing an accurate comparison between the incoming, incomplete data and the complete sample data already available in the image database.

The preliminary results of our work indicate that the proposed system is capable of successful person identification when incomplete data is available.

II. RESEARCH METHODOLOGY

This section describes the process of recording the human gait patterns to enable the creation of the skeletal models of the sample people, to compute the feature vectors, and to measure similarity, as shown in Figure 1.

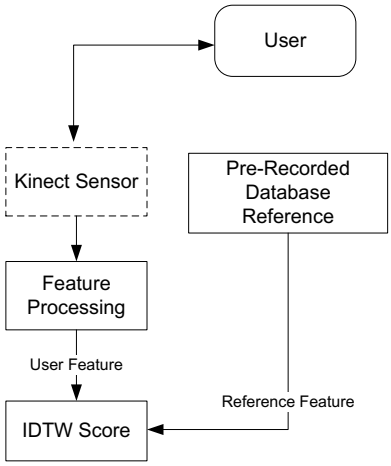

Figure 1: Block diagram of the proposed method for person re-identification

A. *Kinect® Camera version 2*

The Kinect camera is a human motion tracking peripheral for the Xbox 360 console from Microsoft. It was the first applied in gaming systems but it has been found to have many other applications and uses, such as human motion and features recognition, 3D model reconstruction, robot navigation, medical applications and dance training [6-8].

Kinect version 2 can extract many aspects of human body poses from the 3D depth images, representing the body as a skeletal image with identifiable body joints. The positions being tracked and recorded are more anatomically correct and stable the longer the activity proceeds, given that more video data is being recorded. Overall, twenty-five body joints can be represented, as shown in Figure 2.

From the Skeleton representation shown in Figure 2, the Spine base joint is used as the origin of the skeleton and the relative positions of the other joints are calculated from this joint.

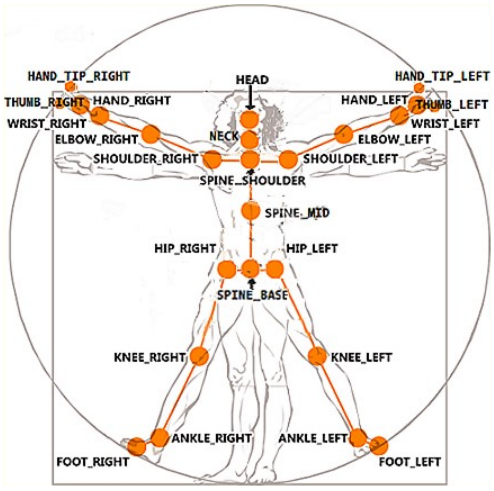

Figure 2: Microsoft SDK Skeleton representation [9]

B. *Incremental Dynamic Time Warping (IDTW)*

IDTW is an algorithm for comparing video image sequences taken at varying speeds and over different time periods by compressing or expanding them in the time domain. This is an extension of the DTW as described in [10]. We used the stored complete data of skeletal features as the reference sequence, as the benchmark, and did a real-time comparison between the new sequence with the reference sequence data. The algorithm calculates a distance score between sequences, providing a comparison between the incomplete sequence and the reference complete sequence, thereby indicating the degree of sameness between the images.

IDTW algorithm works by comparing each sequence with the best starting segment of reference. It terminates with the result after applying the full (reference) sequence with all possible frames, computing the DTW distances for every possible comparison and reporting the minimum distance as follows [7]:

I. Initialize parameters: $\mathbf{U}$ is the user sequence up to current time (length N), $\mathbf{E}$ is the reference full sequence (length M), $\mathbf{G}$ is the matrix ($M \times N$) cumulative cost matrix up to current time, and V is next frame in user sequence.

II. Inserting a new frame: $Q = N + 1$ and $U_Q = V$, where U_Q is the Q-th frame of $\mathbf{U}$.

III. The DTW distance is calculated for all $\mathbf{E}^j, j = 1, \ldots, M$ where each $\mathbf{E}^j$ is the reference sequence $\mathbf{E}$ truncated at j^{th} frame, i.e., $D(\mathbf{U}, \mathbf{E}^j), j = 1, \ldots, M$ is the DTW distance obtained by the optimal warping path.

IV. The minimum is found using the following formula:

$$D_{IDTW}(\mathbf{U}, \mathbf{E}) = \min_{j=1,\ldots,M} D(\mathbf{U}, \mathbf{E}^j) \quad (1)$$

V. Step II and IV are iterated for each the new sequence and updated IDTW distance (the cost matrix $\mathbf{G}$) by using the following formula:

$$\mathbf{G} = \min\left(\mathbf{G}(1 \ldots M, Q)\right)/Q) \quad (2)$$

Note that the components of DTW from previous states are reused as the new sequence movement advances, to optimize computation time.

Considering the steps, we notice that the next frame of the user sequence will be input to the process. Then just one column is added to the cumulative cost matrix G. This column is filled in line with a chosen updated rule in a similar way to classic DTW. Finally, the normalized minimum value from the newly added column of the cost matrix G comes out as the current IDTW distance.

III. Experimental Results

We obtained a dataset of gait gestures from 16 people by each person walking at a different distance from the Kinect sensor, at the Near Recording (R1) distance of two meters and the Far Recording (R2) distance of three meters, and another record distance between Near Recording and Far Recording for testing data (R3) as shown in Figure 3.

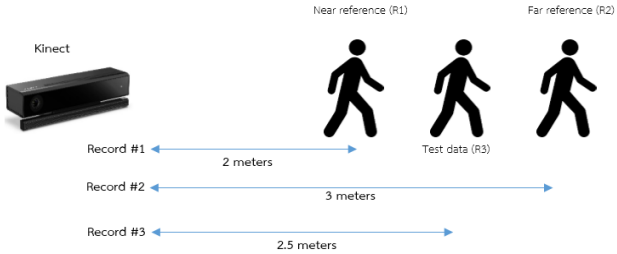

Figure 3: The experimental setting

In this experiment, we selected the twenty joints shown in Figure 2, as a way to evaluate the accuracy of the program. They are SpineMid, Neck, Head, LeftShoulder, LeftElbow, LeftWrist, LeftHand, RightShoulder, RightElbow, RightWrist, RightH, HipLeft, KneeLeft, AnkleLeft, FootLeft, HipRight, KneeRight, AnkleRight, FootRight, and SpineShoulder. Since the movement of each gesture is strongly related to these joints, the application of this image data can classify person gait gestures effectively.

The experimental results as shown in Table 1 which shows the results of the IDTW algorithm in recognizing the between Near record and Far record. For each query instance, the IDTW scores were calculated and nearest neighbor search was applied to select the best match.

Table 1
Recognition results of Near record and Far record

No.	Ref.	Result (Far)		Result (Near)	
1	a	a2	✓	a1	✓
2	ad	ad2	✓	ad1	✓
3	ant	ant2	✓	ant1	✓
4	ap	pae2	×	so1	×
5	b3	b2	✓	b1	✓
6	bo	va2	×	kani1	×
7	bu	ta2	×	bu1	✓
8	ch	ch2	✓	sakul	×
9	cu	cu2	✓	cu1	✓
10	ja	ja2	✓	ja1	✓
11	ka	ka2	✓	ka1	✓
12	kani	si2	×	ta1	×
13	ko	va2	×	ko1	✓
14	mo	mo2	✓	mo1	✓
15	mum	tum2	×	mum1	✓
16	n	n12	✓	n11	✓
Accuracy		10/16	63%	12/16	75%

From the results in Table 1, it can be seen that the accuracy of recognition can be improved by using the average IDTW scores from the Near record and the Far record, as shown in Table 2. These were calculated by the following formula:

$$Best\_Match = \underset{k=1,\ldots,K}{\operatorname{argmin}} \left(\frac{IDTW\,\mathrm{near}_k + IDTW\,\mathrm{far}_k}{2} \right) \quad (3)$$

where $IDTW$ near is the IDTW score of Near record, $IDTW$far is the IDTW score of Far record, and K is the total number of files.

Usually, when the skeleton files from the Far record are compared against those of the Near record, each containing the same reference gestures of the same person, the IDTW score should be close to zero. However, as demonstrated in our experiments, it is very difficult to achieve an IDTW score of zero, or close to zero. Thus, it is important to obtain the best possible recordings and images for inclusion in the reference gesture database.

Some query instances (as shown in Table 1) provided positive and negative results; for example, persons numbered 4, 6 and 12. The application of equation 2 reduced the confusion when making decision with nearest neighbor search on Far and Near records. The new results are represented in Table 2 and the average result was 81%.

Figure 4 shows a graphical comparison of distance scores of both the classic DTW and IDTW algorithms for each frame during the comparison tests between two data sequences. It shows that the distance scores approach zero at the last data frame for both algorithms, which means that both algorithms can ultimately correctly identify the subject with enough data. However, the IDTW provides the answer faster than the DTW.

In practice, recorded data sequences may be incomplete, or the query sequence is incomplete and may not be effective for comparison with the completed reference data. In such cases, the IDTW algorithm was superior in recognition with incomplete recorded sequences. Moreover, the IDTW algorithm can support the real-time classification because it can issue the comparing result within a second.

Table 2
Recognition results when calculated average IDTW scores between the Near record and the Far record

No.	Ref.	Result (Average)	
1	a	a	✓
2	ad	ad	✓
3	ant	ant	✓
4	ap	pae	×
5	b3	b	✓
6	bo	va	×
7	bu	bu	✓
8	ch	ch	✓
9	cu	cu	✓
10	ja	Ja	✓
11	ka	ka	✓
12	kani	si	×
13	ko	ko	✓
14	mo	mo	✓
15	mum	mum	✓
16	n	n	✓
Accuracy		13/16	81%

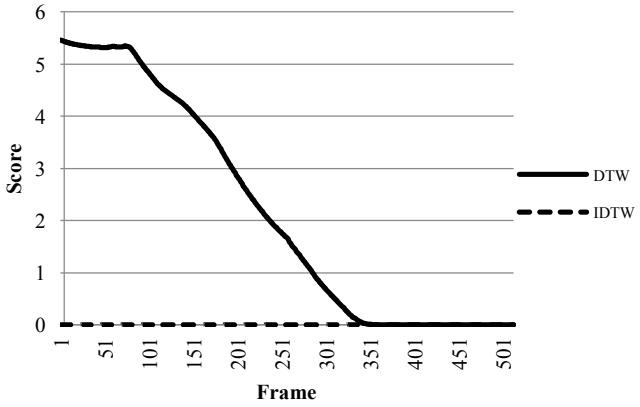

Figure 4: Comparison of distance scores between DTW and IDTW

IV. CONCLUSIONS

We have proposed and tested a method for early recognition of human gait gestures, for person identification, using the Kinect camera. The Incremental Dynamic Time Warping (IDTW) algorithm was applied to calculate the similarity between human gait gestures. We showed that the IDTW algorithm can calculate the distance score between two sequences faster and with greater accuracy, than the classic DTW. The IDTW algorithm accurately identified the test subject 81% of the time. A larger volume of data will enhance this outcome, and we will develop the image database to achieve this end. We will also move the database to be remotely stored in a 'cloud' system.

REFERENCES

[1] Fothergill, S., Mentis, H., Kohli, P. and Nowozin, S. Instructing people for training gestural interactive systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, (2012).

[2] Maquet, Paul GJ. Biomechanics of the knee: with application to the pathogenesis and the surgical treatment of osteoarthritis. Springer Science & Business Media, (2012).

[3] Dikovski, Bojan, Gjorgji Madjarov, and Dejan Gjorgjevikj. Evaluation of different feature sets for gait recognition using skeletal data from Kinect. Information and Communication Technology, Electronics and Microelectronics (MIPRO), 37th International Convention on. IEEE, (2014).

[4] Iosifidis, Alexandros, Anastasios Tefas, Nikolaos Nikolaidis, and Ioannis Pitas. Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis. Computer Vision and Image Understanding, 116.3 (2012): 347-360.

[5] Munsell, B. C., Temlyakov, A., Qu, C., and Wang, S. Person identification using full-body motion and anthropometric biometrics from kinect videos. In Computer Vision–ECCV 2012. Workshops and Demonstrations. Springer Berlin Heidelberg, (2012) 91-100.

[6] Naimul Mefraz Khan, Stephen Lin, Ling Guan, and Baining Guo. A Visual Evaluation Framework for In-Home Physical Rehabilitation. IEEE International Symposium on Multimedia, (2014).

[7] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. Special Issue on Visual Understanding with RGB-D Sensors, ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 6(2), Article ID 23 (2015).

[8] W. Kwankhoom and P. Muneesawang. Recognition of Standard Thai Traditional Dance Through 3D Data Analysis. Naresuan University Engineering Journal, Vol.11, No.2 (2016) 75-84.

[9] Microsoft Developer Network. JointType Enumeration. Available at https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx. Accessed Nov 11, (2016).

[10] Miguel Reyes, Gabriel Dominguez, and Sergio Escalera. Feature Weighting in Dynamic Time Warping for Gesture Recognition in Depth Data. IEEE International Conference on Computer Vision, (2011).