View Independent Human Identification by Gait Analysis using Skeletal Data and Dynamic Time Warping

Athiwat Cheewakidakarn, Nirattaya Khamsemanan, Cholwich Nattee
School of Information, Computer and Communication Technology,
Sirindhorn International Institute of Technology,
Thammasat University, Thailand
Email: athiwat.cheewakidakarn@studentmail.siit.tu.ac.th, {nirattaya,cholwich}@siit.tu.ac.th

***Abstract*—Microsoft Kinect is a motion sensing device that allows us to conveniently capture human body motions and extract skeletal data. For biometric from human gait, Kinect lets us extract more kind and more complicated gait features which help improve identification accuracy. This paper proposes biometric from human gait technique from view independent gait analysis based on the Kinect's motion tracking ability in 3D coordinate. Dynamic Time Warping is applied to identify users from their walking styles. The selected features include static features e.g. proportion of body parts' length , height as well as kinematic features e.g. head tilt angles, body tilt angles.**

Keywords - View Independent; Human Recognition; Kinect; Gait Analysis; Dynamic Time Warping

I. INTRODUCTION

Using human gait as a biometric is the techniques which identify people from their gait features including walking style and body's dimensions ([1]). Compared to other biometrics such as fingerprint, voice or iris recognition, the advantage of biometric from human gait is gait features are almost impossible to be completely and permanently altered, covered and imitated. Gait features can also be acquired unintrusively in the public space.

For those reasons, biometric from human gait has been increasingly concerned in past few decades. Many techniques of biometric from human gait has been developed such as Background Subtraction[2]–[5], Silhouettes[6], Optical Flow[7], Motion Energy and Motion History Images[8]. Those techniques are on the complicated environmental setup and specific devices such as wearable sensors, multi-video cameras. With the help of Kinect technology from Microsoft, we can obtain more kind of the gait features without the need of those complicated environmental setup or specific devices as suggested in [9]–[13] and [15].

To the best of our knowledge, none of the existing techniques have been designed to identify human with unlimited and uncontrolled movement. However, it is impossible to limit the walking path or to exactly guide peoples how to walk when they are monitored by a gait analysis system. In this work, we explore another possibility for application of Kinect technology to the gait analysis which is to capture gait features without considering of viewpoint to the subject and walking path. Subjects are allowed to walk freely in our experiment.

II. BACKGROUND AND RELATED WORKS

A. Kinect and Kinect SDK

Kinect is a human interface device first introduced by Microsoft in November 2010. The device composed of RGB video camera, depth sensors and array of microphones. Microsoft provided Kinect Software Development Kit (Kinect SDK) which allow developers to develop Kinect applications. In our work, we develop our application on SDK version 1.7.

Kinect is mainly applied for gesture recognition tasks which provide target's joints position in 3D coordinate or so called "Skeletal data". Figure 1 show the 20 joints around the human's body which are traceable through Kinect and Kinect SDK.

B. Dynamic Time Warping

The Dynamic Time Warping (DTW) is nonlinear time normalization technique for measuring dissimilarity (or distance) between time dependent series or sequences with difference in length. Dynamic Time Warping has been applied to various type of applications ([16]). In the work of [14] , they applied Dynamic Time Warping for measuring distance between sequences which represent period of walking cycles with different length. In our work, we apply Dynamic Time Warping to measure distance between the sequences of kinematic gait features captured from each participants' walking sequences.

C. Related Works

Boulgouris et al. [14] proposed the method to extract the period of walking cycles by extraction the sequences of sum of foreground pixels from silhouette images converted from human walking videos. They applied Dynamic Time Warping to measure the distance between each subjects' period of walking cycles. Their work yielded 85% identification accuracy. This shows the promising possibility of applying Dynamic Time Warping to the gait analysis applications. The proposed method of [14] is also view independent that the sum of foreground pixels can be extracted regardless of the

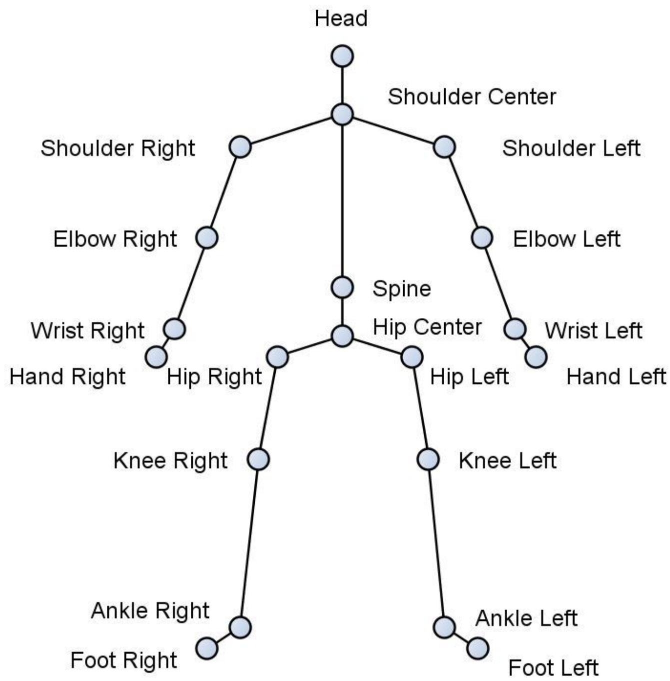

Fig. 1. Twenty joint positions provided by Kinect SDK

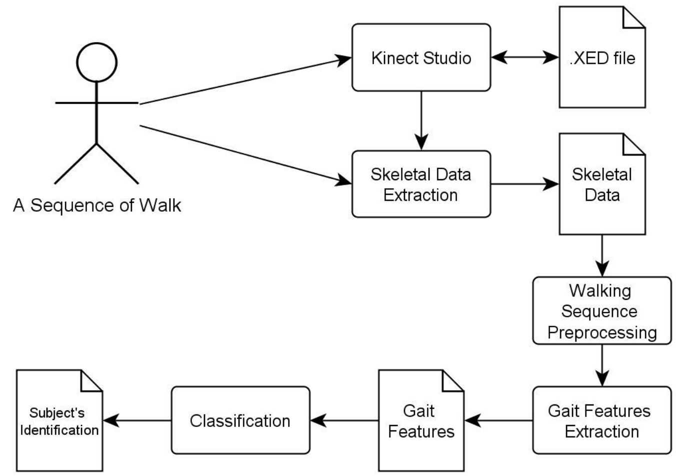

Fig. 2. The overall processes

Fig. 3. Examples of walking sequence in our experiment

view point of the camera. For this reason, we compare our proposed method's classification accuracy to [14]. The result of implementation [14]'s method is shown in Section IV.

Our previous work ([15]) proposed the biometric from gait which based on Euclidean distances from static gait features those are Length of Right Upper Arm, Length of Right Forearm, Length of Right Arm, Length of Right Thigh, Length of Right Calf, Length of Right Leg, Length of Torso and Height. We also measured the distance by using Dynamic Time Warping technique from kinematic gait features those are Right Arm Swing Angles, Body Tilt Angles, Head Tilt Angles and Stride Angles. The gait features in this proposed method are limited to be captured from the right side of human body which is only one view point in [15]'s experiment.

III. METHODOLOGY

Our proposed technique is composed of 5 components, Figure 2 shows how each component related to the system.

We first capture walking sequences from each participants using Kinect. The next is to extract skeletal data which is the sequences of joints coordinates in 3D space from each captured walking scene. Once we have skeletal data of each walking scene, we preprocess the extracted data before calculation of gait features so that the effect from different view point will be eliminated. The identification process will be performed then to measure the identification accuracy of the proposed method.

A. Capturing Walking Sequences

We capture walking sequences from 17 participants, 14 males and 3 females, 5 times each. They are instructed to walk in their natural manner. In our experiment, each participant is allowed to start walking from any direction and changes to other directions as many times as they want. Kinect is fixed at the top end of a 195 centimeters height pole. This is to simulate the real life situation that people is walking freely in the sight of a security camera with carelessness of the camera's presence. The duration of captured walking sequences range from 832 to 4,823 milliseconds. Figure 3 shows the examples of captured walking sequence in our experiment.

We found that the joint positions of each participant's walking sequences provided by the Kinect SDK were abnormally arranged or have unnatural trajectories during a split of second before they disappeared from the Kinect's sight. This would cause an error to the experiment outcome. We decided to exclude last 10 frames from each walking sequence. This significantly improve identification accuracy of our experiment.

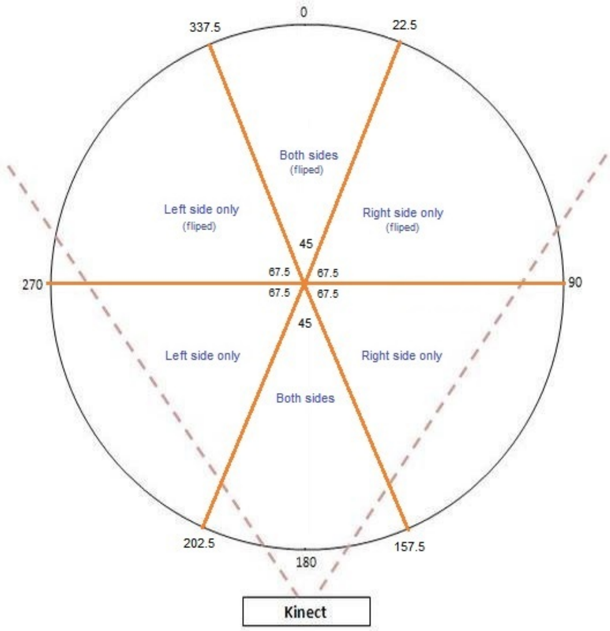

Fig. 4. Direction diagram

B. Skeletal Data Extraction

The captured walking sequences contain both the sequences of color images from Kinect's video camera and the sequences of depth streams from Kinect's depth sensors. We utilize Kinect SDK to extract the skeletal data which are 20 joints' position sequences around the participants' body while they are walking from the captured walking sequences.

C. Skeletal Data Preprocessing and Gait Features Extraction

In order to calculate gait features from each walking sequence correctly, we must know the walking direction of the subject at each moment. How does the system calculate gait features depends on the walking direction. Figure 4 shows the direction diagram of the system.

The direction in the diagram ranged from 0 to 359 degrees depends on the walking direction of the subjects. For example, the subjects will have direction between 157.5 to 202.5 degrees if they presence their front side of the body to the Kinect. The system will calculate gait features from both side of the body. At anywhere between 90 to 157.5 degree, the subjects presence their right side to the Kinect and the left side will be hidden so the system calculate gait features from only right side of the body and vice versa.

Note that for the 0 to 90 and 270 to 359 degrees, the Kinect detect the joints' position in horizontally inverted manner because the Kinect SDK would only interpret the front side of the body even though actually the back side of the body is presented.

To obtain the walking direction of a participant at walking sequence frame t, we calculate the direction vector from the position of participant's Hip Center in x and z-axis , where:

$$\vec{S}_{w,t} = (h_{X_t}, h_{Z_t})$$

is the starting vector of the direction vector for frame t in walking sequence w and

$$\vec{E}_{w,t} = (\frac{h_{X_{t+1,t+2,t+3}}}{3}, \frac{h_{Z_{t+1,t+2,t+3}}}{3})$$

is the end vector of the direction vector for frame t in walking sequence w which is the average value from frame $t+1$, $t+2$ and $t+3$. Note that h_X is the position of Hip Center in x-axis and h_Z is the position of Hip Center in z-axis. The direction vector, therefore, is defined as:

$$\vec{D}_{w,t} = (\vec{E}_{w,t} - \vec{S}_{w,t}) \tag{1}$$

We calculate 20 gait features from each captured walking sequence. These features can be categorized into 2 different groups; static and kinematic. Static features are the ratios or the lengths which depend on the proportion of the body only. Kinematic features are the patterns of the movement which depend on the walking style. Table I and Table II show the list of the static and kinematic gait features in our experiment.

TABLE I
STATIC GAIT FEATURES

Order	Feature Name
1	Average Ratio of Upper Arm Length to Forearm Length Right
2	Average Ratio of Upper Arm Length to Forearm Length Left
3	Average Ratio of Arm Length to Leg Length Right
4	Average Ratio of Arm Length to Leg Length Left
5	Average Ratio of Head Length to Torso Length
6	Average Ratio of Head Length to Height
7	Average Upper Arm Length Right
8	Average Upper Arm Length Left
9	Average Forearm Length Right
10	Average Forearm Length Left
11	Average Shoulder Length
12	Average Torso Length
13	Average Thigh Length Right
14	Average Thigh Length Left
15	Average Calf Length Right
16	Average Calf Length Left
17	Average Height

TABLE II
KINEMATIC GAIT FEATURES

Order	Feature Name
1	Body Tilt Angles (to y-axis)
2	Head Tilt Angles (to y-axis)
3	Height Variation During Walking

We calculate each gait feature value for each frame captured in the walking sequence. So the number of gait feature values depend on the duration of the corresponding captured walking sequences.

We calculate the average value of static gait features for each walking sequence. This is used for calculation of static distance (see further explanation in Section III-E) between any pair of walking sequences.

D. Preparation Gallery Database and Test Database

In our experiment, each participant took walking sequence capturing for 5 times. Four out of 5 walking sequences from each participant will be used to construct the gallery database and the other one to construct the test database.

For each participant's gait features instance in gallery database, we calculate average value for each static gait feature from 4 walking sequences. The kinematic gait features from 4 walking sequences are included into the corresponding participant's gait features instance in form of time-dependent sequences.

For each participant's gait features instance in test database, the average values of each static gait feature from every frame in that walking sequence are included. The kinematic gait features from that walking sequence are included into the corresponding participant's gait features instance in form of time-dependent sequences.

We have 17 participants in our experiment, so there will be 17 gait features instances in the gallery database, each of them included averaged static gait features and sequences of kinematic gait features from 4 walking sequences, and 17 gait features instances in the test database.

Note that, for the gait features that involve the body part below knees, we exclude the frames from the walking sequence which the participant belong under 200 centimeters from the Kinect. The reason is the lower parts of the body are disappeared from the Kinect's sight when they are at that close.

E. Calculation of distance between walking sequences

We separated the calculation procedure into 2 parts: static distance and kinematic distance.

For static distance $D_s(G_s, T_s)$ between static gait features instance G_s from gallery database and T_s from test database, we find Euclidean distance between Euclidean vector G_s and T_s where:

$$G_s = (g_{s_1}, g_{s_2}, g_{s_3}, \ldots, g_{s_{17}})$$

and

$$T_s = (t_{s_1}, t_{s_2}, t_{s_3}, \ldots, t_{s_{17}})$$

Note that the order of static gait features is as listed in Table I. So, g_{s_i} and t_{s_i} represent the ith static gait features value for walking sequence G_s and T_s respectively. Static distance $D_s(G_s, T_s)$ is defined as:

$$D_s(G_s, T_s) = \sqrt{\sum_{i=1}^{17} (g_{s_i} - t_{s_i})^2} \tag{2}$$

For kinematic distance $D_k(G_k, T_k)$ between kinematic gait features instance G_k from gallery database and T_k from test database, first we defined G_k as:

$$G_k = (g_{k_1}, g_{k_2}, g_{k_3})$$

where g_{k_i} represents the set of sequences which is kinematic gait features order i obtained from 4 walking sequences. So:

$$g_{k_i} = (g_{k_{i,1}}, g_{k_{i,2}}, g_{k_{i,3}}, g_{k_{i,4}})$$

where $g_{k_{i,j}}$ represents the sequence of kinematic gait features order i obtained from walking sequence number j. In particular;

$$g_{k_{i,j}} = (g_{k_{i,j,1}}, g_{k_{i,j,2}}, g_{k_{i,j,3}}, \ldots, g_{k_{i,j,m}})$$

where $m = 1, 2, \ldots, M$ and M is the length of the sequence. For example $g_{k_{2,3,9}}$ is the Head Tilt Angle of frame 9 obtained from the walking sequence number 3 in the gallery database. Kinematic gait features instance T_k is defined as:

$$T_k = (t_{k_1}, t_{k_2}, t_{k_3})$$

where t_{k_i} represents the sequence which is kinematic gait features order i and

$$t_{k_i} = (t_{k_{i,1}}, t_{k_{i,2}}, t_{k_{i,3}}, \ldots, t_{k_{i,n}})$$

where $n = 1, 2, \ldots, N$ and N is the length of the sequence which is kinematic gait features order i. For example $t_{k_{1,15}}$ is the Body Tilt Angle of frame 15 from the test database. The order of kinematic gait features is listed in Table II.

We apply Dynamic Time Warping to find distance between $g_{k_{i,j}}$ and t_{k_i}. We defined cost matrix C which is an M-by-N matrix composed of Euclidean distances between all possible pairs of an element $g_{k_{i,j,m}}$ in $g_{k_{i,j}}$ and $t_{k_{i,n}}$ in t_{k_i}:

$$C_{m,n} = \sqrt{(g_{k_{i,j,m}} - t_{k_{i,n}})^2}$$

We then defined cumulative distance matrix γ which is an $(M+1)$-by-$(N+1)$ matrix as:

$$\gamma_{i,j} = C_{i-1,j-1} + min\{\gamma_{i-1,j}, \gamma_{i,j-1}, \gamma_{i-1,j-1}\}$$

and defined initial condition of matrix γ as: $\gamma_{1,1} = 0$, $\gamma_{i,1} = \infty$ for $i \in [2, M+1]$ and $\gamma_{1,j} = \infty$ for $j \in [2, N+1]$ where $min\{\gamma_{i-1,j}, \gamma_{i,j-1}, \gamma_{i-1,j-1}\}$ is the minimum cumulative distance from adjacent elements $\gamma_{i-1,j}, \gamma_{i,j-1}$ and $\gamma_{i-1,j-1}$. The distance between $g_{k_{i,j}}$ and t_{k_i} is cumulative distance from matching entire of $g_{k_{i,j}}$ to the entire of t_{k_i}:

$$dtw(g_{k_{i,j}}, t_{k_i}) = \gamma_{M+1,N+1}$$

where $dtw(g_{k_{i,j}}, t_{k_i})$ is the distance between $g_{k_{i,j}}$ and t_{k_i} calculated using Dynamic Time Warping technique. We defined the calculation of distance between g_{k_i} and t_{k_i} as:

$$dist(g_{k_i}, t_{k_i}) = \sum_{j=1}^{4} dtw(g_{k_{i,j}}, t_{k_i})$$

where $dist(g_{k_i}, t_{k_i})$ is the sum of distance between all possible pairs of $g_{k_{i,j}}$ and t_{k_i}. The complete kinematic distance $D_k(G_k, T_k)$ is defined as:

$$D_k(G_k, T_k) = \sum_{i=1}^{3} dist(g_{k_i}, t_{k_i}) \tag{3}$$

The distance $D(G, T)$ between a pair of gait features instance $G = (G_s, G_k)$ and $T = (T_s, T_k)$ is defined as:

$$D(G, T) = \alpha\Big(\frac{D_s(G_s, T_s)}{Ds_{max}}\Big) + (1-\alpha)\Big(\frac{D_k(G_k, T_k)}{Dk_{max}}\Big) \tag{4}$$

where α is the weighting factor for combining static and kinematic distance which is a virtual parameter and can be obtained from the experiment. The Ds_{max} is maximum static distance from all static distance in our experiment and Dk_{max} is maximum kinematic distance from all kinematic distance in our experiment. The Ds_{max}, and Dk_{max} are needed to normalize kinetic and static distances since they are calculated based on different types of units. Next section shows the experimental results on the dataset.

IV. RESULT AND DISCUSSION

In order to match a gait features instance T from test database to the most similar G in gallery database, we apply 1-Nearest Neighbor algorithm on the set of distances between T and every gait features instance from gallery database.

Using only static gait features (as listed in Table I) in classification (in other words, α = 1.00) yielded 27.06% $\pm$ 5.46% average identification accuracy while using only kinematic gait features (as listed in Table II) in classification (in other words, α = 0.00) yielded 28.24% $\pm$ 3.43% average identification accuracy. The better result yielded from using both static and kinematic gait features in classification.

TABLE III
IDENTIFICATION ACCURACY USING BOTH STATIC AND KINEMATIC GAIT FEATURES WITH DIFFERENT WEIGHTING FACTORS

(α)	Fold1	Fold2	Fold3	Fold4	Fold5	Average $\pm$ Std. Error
0.00	41.18%	23.53%	29.42%	23.53%	23.53%	28.24% $\pm$ 3.43%
0.05	47.06%	29.42%	29.42%	23.53%	29.42%	31.77% $\pm$ 3.99%
0.10	47.06%	35.3%	41.18%	23.53%	41.18%	37.65% $\pm$ 3.99%
0.15	41.18%	35.3%	41.18%	29.42%	41.18%	37.65% $\pm$ 2.36%
0.20	47.06%	41.18%	41.18%	29.42%	41.18%	40.00% $\pm$ 2.89%
0.25	52.95%	41.18%	41.18%	35.3%	41.18%	42.36% $\pm$ 2.89%
0.30	58.83%	41.18%	47.06%	41.18%	41.18%	45.89% $\pm$ 3.43%
0.35	64.71%	47.06%	47.06%	41.18%	41.18%	48.24% $\pm$ 4.33%
0.40	64.71%	47.06%	52.95%	52.95%	41.18%	**51.77%** $\pm$ **3.91%**
0.41	64.71%	47.06%	52.95%	52.95%	41.18%	**51.77%** $\pm$ **3.91%**
0.45	52.95%	47.06%	47.06%	52.95%	41.18%	48.24% $\pm$ 2.21%
0.50	52.95%	47.06%	47.06%	58.83%	41.18%	49.42% $\pm$ 3.00%
0.55	52.95%	35.3%	47.06%	47.06%	41.18%	44.71% $\pm$ 3.00%
0.60	47.06%	35.3%	47.06%	35.3%	35.3%	40.00% $\pm$ 2.89%
0.65	41.18%	41.18%	41.18%	29.42%	35.3%	37.65% $\pm$ 2.36%
0.70	41.18%	41.18%	41.18%	29.42%	35.3%	37.65% $\pm$ 2.36%
0.75	41.18%	35.3%	35.3%	23.53%	35.3%	34.12% $\pm$ 2.89%
0.80	47.06%	35.3%	23.53%	23.53%	29.42%	31.77% $\pm$ 4.41%
0.85	47.06%	29.42%	23.53%	35.3%	29.42%	32.95% $\pm$ 3.99%
0.90	41.18%	29.42%	17.65%	41.18%	29.42%	31.77% $\pm$ 4.41%
0.95	35.3%	23.53%	11.77%	29.42%	35.3%	27.06% $\pm$ 4.41%
1.00	35.3%	17.65%	11.77%	29.42%	41.18%	27.06% $\pm$ 5.46%

We tried 101 weighting factors from 0.00 to 1.00 increasing by 0.01. For each weighting factor we apply 5-fold cross validation on our gait features samples. We calculate average identification accuracy for each weighting factor yielded from 5 folds.

Table III shows our experiment result, the best identification accuracy is 51.77% $\pm$ 3.91% by using weighting factors of 0.40 and 0.41.

We compared our proposed method to the method of [14]. We use the sequence of Body Tilt Angles to represent period of walking which is only gait feature in [14]. We applied Dynamic Time Warping to match a gait feature from test database to the most corresponding one in gallery database. Table IV shows the identification accuracy of implementing the technique from [14] in our work.

TABLE IV
IDENTIFICATION ACCURACY OF IMPLEMENTATION THE TECHNIQUE IN [14] ON OUR WALKING SEQUENCES

Method	Fold1	Fold2	Fold3	Fold4	Fold5	Average $\pm$ Std. Error
DTW on Walking Period	11.76%	5.88%	5.88%	5.88%	17.65%	**9.41%** $\pm$ **2.35%**

V. CONCLUSION AND FUTURE WORK

With the ability to track human movement in 3D coordinate provided by Kinect, we can extract more kind and more complicated gait features from human walking sequences compared to conventional devices such as video cameras or wearable sensors. Kinect also brings more robustness, less effort, less cost, to the gait analysis tasks.

For the future works, we are finding other kinds of classifier and distance measure techniques to improve the accuracy of the proposed method.

VI. ACKNOWLEDGEMENT

This research is financially supported by Thailand Advanced Institute of Science and Technology (TAIST), National Science and Technology Development Agency (NSTDA) Tokyo Institute of Technology and Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU) and the National Research University Project, Thailand Office of Higher Education Commission.

REFERENCES

[1] Jeffrey E. Boyd, James J. Little, *Biometric Gait Recognition*, Advanced Studies in Biometrics Lecture Notes in Computer Science Volume 3161. (2005) 19-42

[2] Sudderth, E., Hunter, E., Kreutz-Delgado, K., Kelly, P.H., Jain, R., *Adaptive video segmentation: theory and real-time implementation*, 1998 Image Understanding Workshop. Volume 1. (1998) 177-181

[3] Horprasert, T., Harwood, D., Davis, L., *A statistical approach for real time robust background subtraction*, IEEE Frame Rate Workshop. (1999)

[4] Stauffer, C., Grimson, W.E.L., *Adaptive background mixture models for real-time tracking*, Computer Vision and Pattern Recognition 99. Volume II. (1999) 246-252

[5] Javed, O., Shafique, K., Shah, M., *A hierarchical approach to robust background subtraction using color and gradient information*, IEEE Workshop on Motion and Video Computing, Orlando, FL (2002) 22-27

[6] Baumberg, A.M., Hogg, D.C., *Learning spatiotemporal models from training examples*, British Machine Vision Conference, Birmingham (1995)

[7] Barron, J.L., Fleet, D.J., Beauchemin, S.S., *Performance of optical flow techniques*, International Journal of Computer Vision 12 (1994) 43-77

[8] Davis, J.W., Bobick, A.F., *The representation and recognition of human movement using temporal templates*, IEEE Computer Vision and Pattern Recognition. (1997) 928-934

[9] J. Preis, M. Kessel, M. Werner, C. Linnhoff-Popien, *Gait recognition with kinect*, 1st International Workshop on Kinect in Pervasive Computing. (2012)

[10] Gabel, Moshe, Ran Gilad-Bachrach, Erin Renshaw, Assaf Schuster, *Full body gait analysis with Kinect*, Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. (2012) 1964-1967

[11] Natdanai Saitong-in, Kanin Assantachai, Nirattaya Khamsemanan, Cholwich Nattee, *Human Identification from Gait Analysis using Microsoft Kinect*, Proceedings of ICT-ISPC. (2013)

[12] Milovanovic M., Miroslav M., Dusan S., *Walking in Colors: Human Gait Recognition Using Kinect and CBIR*, (2013)

[13] Chatzichristofis, Savvas A., Yiannis S. Boutalis, and Mathias Lux, *Img (rummager): An interactive content based image retrieval system*, Similarity Search and Applications, SISAP'09, Second International Workshop, IEEE. (2009)
[14] Boulgouris Nikolaos V., Konstantinos N. Plataniotis, Dimitrios Hatzinakos, *Gait recognition using dynamic time warping*, Multimedia Signal Processing, 2004 IEEE 6th Workshop. (2004) 263-266
[15] Athiwat Cheewakidakarn, Nirattaya Khamsemanan, Cholwich Nattee, *Human Identification by Gait Analysis using Skeletal Data and Dynamic Time Warping*, Proceeding of the 14th International Symposium on Advanced Intelligent Systems (ISIS). (2013)
[16] Ratanamahatana, Chotirat Ann, Eamonn Keogh, *Everything you know about dynamic time warping is wrong*, Third Workshop on Mining Temporal and Sequential Data. (2004)
[17] *Microsoft Kinect SDK*
http://www.microsoft.com/en-us/kinectforwindows/