Robust gait recognition using adaptive random depth subspace from depth information

YUAN QU † ATSUO YOSHITAKA †

Abstract—Depth information provided by depth sensors, such as Kinect, contains the information of physical distance from the sensor to a human body part at each pixel while walking, and it shows good performance in gait recognition. In this paper we proposed an adaptive random depth subspace (ARDS) framework to explore the applicability of depth information for robust gait recognition of individuals. Without assuming specific clothes type and baggage carrying condition for each input test sequence, proposed method is able to segment human body into very small regions and adaptively select more unaffected random depth subspaces as the reconstructed gait features for classification, which is realistic for real-world applications. Experiments are conducted on a new depth gait database captured with Microsoft Kinect assuming several walking conditions. Experimental results showed the effectiveness compared with other methods using depth information.

Keywords: Robust gait recognition; depth information; Microsoft Kinect; Subspace; Random

1. Introduction

Identifying individual is a challenging task and has been studied extensively. Human gait is advantageous for recognition over other traditional biometric modalities such as face, DNA and fingerprint. Since gait features neither can be easily imitated nor hidden. Furthermore, walking data acquisition doesn't require a closer distance and higher quality sensors. In addition, the complete unobtrusiveness without cooperation or contact of the subject also makes gait attractive in many applications such as access control, biomedical research and criminal investigation.

Over the last twenty years, researchers presented numerous successful gait recognition techniques with RGB images. These techniques were classified into two major categories called model-based and appearance-based approaches [1]. The major limitation of model-based approaches [7] is that it requires high quality gait images and high computational cost for motion model construction on each body part.

In contrast, as a representative appearance-based approach, Gait Energy Image (GEI) is a spatio-temporal template calculated by averaging all images extracted from video in a full gait cycle [2]. This method achieves good experimental results with less computational cost. Nevertheless, appearance-based approaches are view dependent and perform best when a side view point is utilized. Because the directions of legs' movement are invisible in the human body silhouettes while subject walks toward the sensor. In addition, as the limitation of the high sensitivity for the body shape changing, GEI is hard to provide reliable performance in some real-life scenarios, such as variations of clothes and carrying a baggage.

Recent depth sensor provided superiority to solve the problem of viewpoint dependency that the depth information of physical distance from the sensor to a human body part at each pixel can be captured efficiently. Based on this advantage, the accuracy of frontal gait recognition in appearance-based method has great potential to be improved. However, compared with the number of algorithms that are applied for RGB images in gait recognition, studies on depth gait recognition is relatively smaller. Sivapalan et al. [3] improved the concept of GEI to 3D and proposed a new gait representation called Gait Energy Volumes (GEV) for depth gait recognition in frontal view point image. Meanwhile, Hofmann et al. [4] proposed an effective method for depth gait representation. Motivated by the idea of histograms of oriented gradients (HOG) [9], depth gait sequence is represented by a Depth Gradient Histogram Energy Image (DGHEI) which improves the accuracy in depth gait recognition. Both of these two methods demonstrated very low accuracy by different cofactors like as clothes, carrying objects, since they treated each body part equally.

In this work, we concentrate on exploring the applicability of depth information for robust gait recognition of individuals from frontal viewpoint, and propose an appearance-based approach named adaptive random depth subspace (ARDS) for improving accuracy, motivated by Isam et al. [5]. Their approach is based on a strong assumption where the clothing types appear in both probe (test) set and training stage. Based on their assumption, each body part has a fixed weight for every sequence in probe (test) set ignoring the clothing types. To overcome this problem, we apply an automatic feature selection algorithm to extract features in each body part adaptively, which is realistic for assumed applications. Moreover, because of random selection of image chunks, the spatial information is lost. In contrast, we select chunks (as named "cell" in our work) at the same location between probe and gallery. Since their method is applied for gait recognition using RGB images, we replace the gait representation in his work with DGHEI. The experimental results indicated that the proposed method improved the performance in robust depth gait recognition for an individual.

2. Depth Gait Representation

In this section, we will briefly introduce the gait representation technique based on depth oriented gradient histogram [4], which will be applied in our proposed method as a gait template for selecting adaptive random depth subspace.

Each subject's gait sequence is defined as a series of static images which are extracted frame-by-frame from the depth gait

† School of Information Science-Japan Advanced Institute of Science and Technology

video in one full gait cycle. In order to extract the depth gait feature from the gait sequence, subject's silhouette segmentation is carried out by adopting a skeleton-tracking technique provided by Microsoft [10]. Then, each image is normalized into the same size by scale variation and alignment. In traditional gait recognition with RGB images, each pre-processed image is converted into binary image as shown in Figure 1 (a). In contrast, the depth information within person's silhouette is represented by grayscale in our work. The higher intensity value in each grayscale means that longer physical distance occurs at this position, as illustrated in Figure 1 (b). The grayscale contain luxuriant edges and depth gradients information, which provide additional discrimination clues in the following recognition.

Computation of the gradient values at each pre-processed image in a gait sequence is the first step of calculation. The magnitude γ and orientation θ of the gradient at a pixel (x, y) is calculated as the following equations:

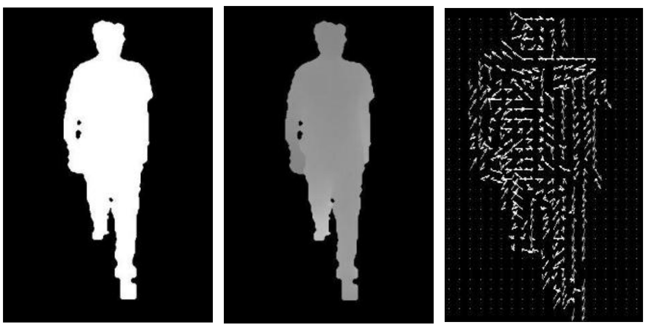
$$u(x,y) = I(x-1,y) - I(x+1,y) \quad (1)$$

$$v(x,y) = I(x,y-1) - I(x,y+1) \quad (2)$$

$$\gamma(x,y) = \sqrt{u(x,y)^2 + v(x,y)^2} \quad (3)$$

$$\theta(x,y) = arctan\big(u(x,y), v(x,y)\big) + \pi \quad (4)$$

Here, $I(x,y)$ is the gray intensity value at a pixel(x,y). Then the gradient orientations at pixel(x,y) are discretized into 9 orientations. In the next step, each pixel within an appropriate square image chunk (typically with fixed size of 8×8 pixels) is weighted by an orientation-based histogram channel with regard to the orientation θ. Thus, the chunk, which is called 'cell', is represented by 9-bin histogram of oriented gradients and reconstructed to a matrix based on the location in the original image. Thus, the pre-processed image is replaced with cell-based matrix. The cell-based matrix is visualized in Figure 1 (c).



(a) Binary image (b) Grayscale image (c) Visual cell-based matrix

Figure 1. Various representations of a single gait image

Finally, the cell-based matrices calculated from a single image are averaged over a gait sequence, which is composed of T images, and is called DGHEI.

$$H(i,j,f) = \frac{1}{T}\sum_{t=1}^{T} h_t(i,j,f) \quad (5)$$

Here, i and j denote the position of a cell and f denotes the numbers of the histogram of this cell. The final DGHEI representation is visualized in Figure 2. In this figure, the origin of each gradient vector denotes the position of a cell, and each cell is considered as a minimum feature unit for random selection.

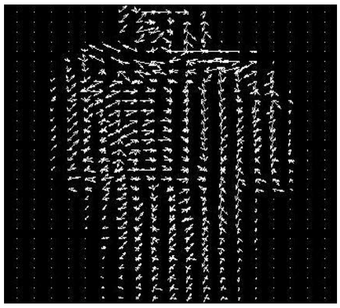


Figure 2. Depth Gradient Histogram Energy Image

3. Adaptive Random Depth Subspace Framework

3.1 Body-region Segmentation

In a large intra-class variation database, each subject walks with various conditions such as changing clothes, carrying baggage. In this case, some parts of depth gait image may be affected by partial occlusion or body shape changing. Thus, the body shape and depth gradient may not reliable anymore in these body parts. In contrast, other unaffected parts maintain nearly the same. These parts remain the useful discrimination clues for individual recognition. In this case, each body part should be treated independently. To tackle robust gait recognition for individual under various walking conditions, we proposed an unsupervised adaptive subspace selection framework. First, we segmented the body part into 4 parts based on a prior knowledge by anatomical body properties discussed by Dempster et al. [8]. When a body's height is H, the upper most section is neck (0.87H). Then the following one is pelvis (0.48H). The final section is the knee (0.285H). On the basis of these premise, the DGHEI is divided into 4 parts marked 1, 2, 3 and 4, as shown in Figure 3. Body part 1, 2, 3 and 4 represent separated regions which is called head, trunk, thigh and crus in anatomy respectively. Therefore, this study takes the various contributions of different part for robust individual recognition into account. The gait features will be selected and extracted independently from each body part in the following work.

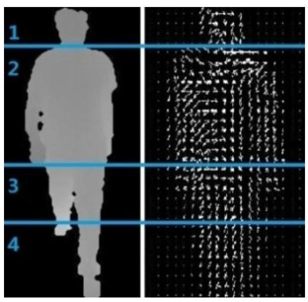


Figure 3. Body-part segmentation

3.2 Random Depth Subspace Selection

At the beginning of this section, we will introduce the definition of subspace. Each subspace is defined as a reconstructed matrix by selecting N cells from p^{th} part of DGHEI randomly.

The coordinates of each cell are represented by i and j. Thus, for recognizing an individual from an input test sequence, the procedure of random depth subspace selection and construction in p^{th} part is given in Algorithm 1. In this algorithm, the gallery (reference) set includes all subjects and each subject has several gait sequences with normal walking condition (described in Section 5.1). The probe (test) set contains the same subjects in gallery set and each subject has several gait sequences with various walking conditions.

Algorithm 1

Input:

T: The DGHEI in p^{th} part of one input gait sequence in probe set.

G^k: The DGHEI in p^{th} part of the k^{th} gait sequence in gallery set.

$Tcell_{ij}$: The cell of T which is located at(x, y).

$Gcell_{ij}^k$: The cell of G^k which is located at (x, y).

N: Total number of cells in each subspace.

K: Total number of gait sequences in the gallery set G^k.

m, n: $m \times n$ cells constitute each DGHEI in p^{th} part.

Output:

ST: Subspace of one input gait sequence in probe set.

SG^k: Subspace of the k^{th} gait sequence in the gallery set G^k.

Procedure:

```
ST = ∅;
SG^k = ∅;
for n ← 1 to N do
    select a pair values of i, j in the range randomly
    [i = 1,2,3 ..., m; j = 1,2,3 ..., n];
    ST ← [ST; Tcell_ij];  ([A; B] denotes that matrices A
    and B integrate vertically into a new matrix)
    for k ← 1 to K do
        SG_k ← [SG_k; Gcell^k_ij];
    end
end
```

In particular, all of cells which contain the same value of zero are discarded from the candidate cells, because most of these cells correspond to the background and do not contribute discrimination for individual recognition. Differently from the method by Isam et al. [5], the location of each selected cell for each sequence in probe set is consistent with all sequences in gallery set for retaining the spatial information efficiently.

Assume that we select L_p subspaces from the p^{th} part in a probe sequence, and then the same number of subspaces L will be selected from corresponding part for each sequence belonging to gallery set. Thus the total number of subspace in each sequence is:

$$\sum_{p=1}^{P} L_p = L \tag{6}$$

In this equation, P is the total number of body parts and $P = 4$ in our experiment (as described in Section 3.1). These subspaces are considered as the new gait features for classification instead of the DGHEI.

3.3 Adaptive Proportion Assignment

In this section, we will apply an adaptive feature selection technique to decide the number of subspaces (L_p) for p^{th} body part. The proportion of L is variable for each input sequence in a probe set.

Without assuming specific clothes type and baggage carrying condition for each input test sequence, we cannot decide what part of DGHEI is affected by walking condition. However, Li et al. [6] tested on a RGB gait database with the variation of walking condition and obtained the results that notable changes, especially the pixel number, appeared in subject's silhouette image under the variance of walking conditions.

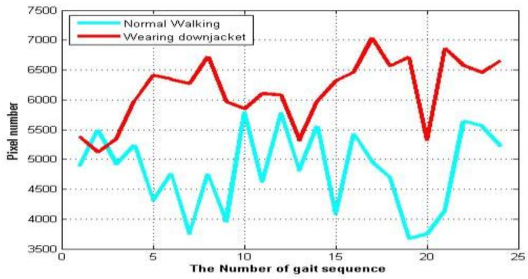


Figure 4. Pixel number distributions in "trunk" part

Figure 4 shows a sample about the distribution of pixel number in "trunk" part under two different walking conditions. It is observed that the number of pixels mainly falls into two scopes. As the described in Section 3.1, the subspaces extracted from this part when subjects walk with down jacket are interfered features for individual recognition.

Based on the distribution of pixel number, we can calculate a probability by the fuzzy membership for an input test sequence as follows:

$$Mem_p = \begin{cases} 0, & {\theta_p}^2 \le X_p; \\ 1 - \frac{X_p - {\theta_p}^1}{{\theta_p}^2 - {\theta_p}^1}, & {\theta_p}^1 \le X_p \le {\theta_p}^2; \\ 0, & X_p \le {\theta_p}^1; \end{cases} \tag{7}$$

In equation (7), Mem_p denotes the probability towards p^{th} part. Higher value of the probability means that human silhouette has a greater chance of being unaffected by walking conditions in p^{th} part. X_p is the average pixel number of all images in p^{th} part. Meanwhile, ${\theta_p}^2$ and ${\theta_p}^1$ is the upper and lower bound of pixel number in gallery set, respectively. The thresholds are decided only from the normal walking sequences in the gallery set. From this equation we can see that we will not select the subspaces from the p^{th} body part if the number of its pixels is deviated from the range of $[{\theta_p}^1, {\theta_p}^2]$. Then the proportion (Pro_p) of L to be calculated for p^{th} body part within a probe gait sequence is obtained as follows:

$$Pro_p = \frac{Mem_p}{\sum_{p=1}^{P} Mem_p} \tag{8}$$

Finally, the number L_p of subspaces selected from the p^{th} part in a test sequence is calculated as follows:

$$L_p = Pro_p \times L \tag{9}$$

Note that the proportion obtained from equation (8) satisfies the constraint of depth subspace selection framework as shown in Section 3.2 such that $0 \le Pro_p \le 1$ and $\sum_{p=1}^{P} Pro_p = 1$.

Figure 5 shows the results of proportion assignment when a subject walks under four different walking conditions. The numbers of selected subspaces are similar in four kinds of body parts under the normal walking condition because each part is unobstructed and contains useful discrimination clues for individual recognition. Nevertheless, when a person carries a bag while walking, the pixel number increases and goes beyond the thresholds for two lower body parts due to the bag's volume. In this case, we will select subspaces only from "head" part and "trunk" part for discarding the interfered information from the region of bag.

Figure 5. Proportion of subspace number L in each body part under four different walking conditions

4. Individual Classification

Let ST and SG^k be two corresponding subspaces selected in Algorithm 1 for a sequence in probe set and k^{th} sequence in gallery set, respectively. We apply Euclidean distance to measure the similarity of two subspaces:

$$d(p, g^k) = \|ST - SG^{\boldsymbol{k}}\| \tag{10}$$

Where $d(p, g^k)$ is the distance between this test sequence (p) and k^{th} sequence g^k in gallery set. Then p is classified as the nearest neighbor strategy. Since L random depth subspaces are extracted from each test sequence, we will acquire L classification results for each sequence in the probe set. The final decision is taken by majority voting from L results. A flow chart of recognizing an individual from an input test sequence is shown in Figure 6. The gallery set comprising K normal walking sequences while the test sequence is a subject's gait sequence of walking with unknown clothes types and postures. The preprocessed images are applied for both DGHEI calculation and adaptive proportion assignment. Each subspace from the test sequence is matched with K subspaces of the gallery set with a distance described in equation (10).

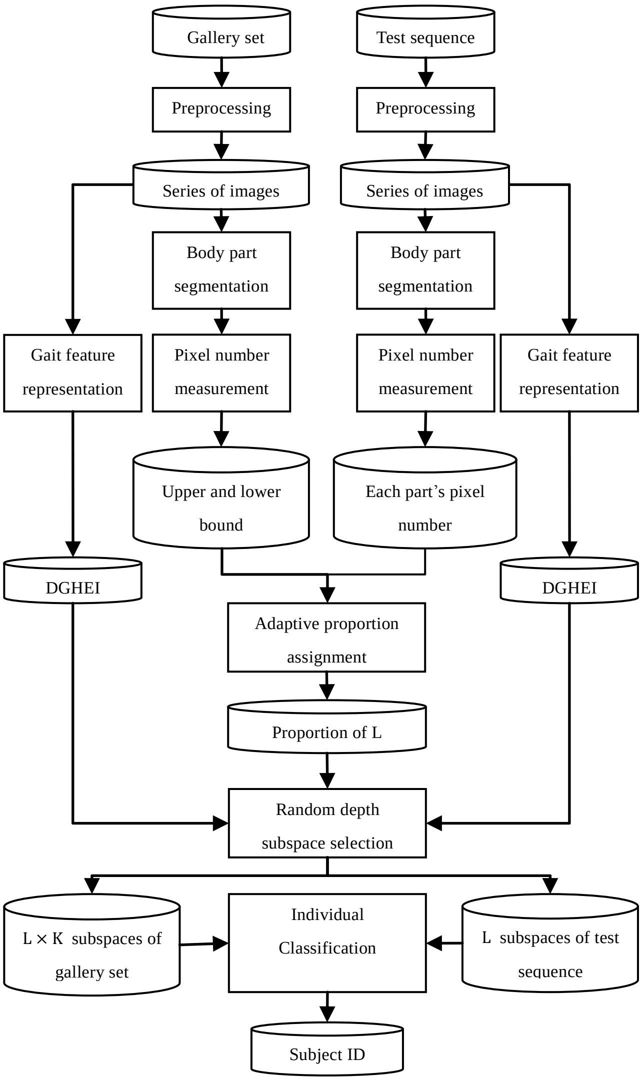


Figure 6. Flow chart of the ARDS framework

5. Experiments

5.1 Robust Depth Gait Database

Since no depth gait database containing multiple walking conditions from frontal viewpoint is available in public domain, we prepared our own depth gait database of frontal depth video frames by Microsoft Kinect XBOX 360. This database is captured in an indoor environment and consists of 12 subjects with several walking conditions. Each subject was instructed to start at the back maker and walk towards the sensor up to the front marker 28 times.

The gait data of each subject is recorded as both color (not used) and depth video frames, including 4 normal walking sequences (wearing T-shirt, regular pants and natural swinging arms), 8 to 12 clothes-varying sequences and 12 to 16 sequences of other walking conditions, named as Set A, Set B and Set C respectively. As shown in Figure 7, we chose some typical dress, such as skirt, down jacket and long coat, for simulating real application. Meanwhile, some common walking posture is selected, including walking with a side-carrying bag and burying his/her hands in the pockets.

Figure 7. Sample images of walking condition

5.2 Experimental Setup

To evaluate the robustness of our method, we segment Set A into two different subsets: Set A1 consists of the 2 sequences of normal walking, taken as the gallery set; Set A2 includes the rest 2 of sequences. Note that only normal walking sequences constitute the gallery set, which is more realistic in real-world applications. For each subject, the probe set is composed of three subsets: Set A2, Set B and Set C. The purpose of testing on three subsets is to evaluate the effectiveness of each method toward different walking conditions. In addition, the gallery set is also referred for calculating the thresholds as described in section 3.3. Since our proposed method is a random system, only one experiment is not enough for evaluation and comparison. We measured the mean accuracy by running ten times.

5.3 Results and Discussion

Table 1 shows the maximum and minimum of the accuracy obtained by ten times of run for our depth gait database. We implemented the GEI, GEV and DGHEI for our depth gait database, and compared our experimental results with these three methods for demonstrating the effectiveness of our method. Considering real-world gait recognition of individual, it is not realistic to train a model for all possible walking conditions for each subject. Thus, the methods breached the above-mentioned requirements are eliminated in comparison prior to the experiment.

As shown in Table 2, it is obvious that the GEI using binary silhouettes that is extracted from the depth images achieved worse performance than these which is applied for the depth information, because the physical distance provides additional discrimination clues from the depth images. GEV and DGHEI, as the methods focusing on depth gait images, extended the concept of the GEI to 3D and showed good performance in Set A2. However, the performance becomes worse when the intra-class variation exceeds individual variation for different walking conditions such as Set B and Set C. This is because the GEV features, as well as DGHEI, are extracted from the whole body silhouettes, which is affected by the variation of walking conditions. Performance is remarkably improved in DGHEI because some body parts with significant discrimination clues, such as legs, are not obstructed. In this case, using DGHEI to extract the gradient information from these body parts becomes smooth, and it is possible to obtain better accuracy than other above mentioned methods.

As seen in the Table 2, the accuracy (Mean accuracy with running ten times) of our proposed method is superior to other methods in these data sets. Specifically, proposed method improves the accuracy higher in Set B than other subsets. Based on the body local gradient information of body and adaptive proportion assignment, proposed method is capable of avoiding the interfered information as much as possible under the variation of clothing effectively.

In our proposed method, there are two significant parameters which need to be investigated for effectiveness: (1) N; the number of cells in each subspace; (2) L; the total number of subspaces in each gait sequence. We obtained the mean accuracy by running ten times under each combination of N and L as shown in Figure 8. In our experiment, each value of N has been tested by setting the range from 2 to 8. The best mean accuracy is obtained for N=3. Though it is not proved in Isam et al. [5], the mean accuracy will be decreased as well, when the value of N decrease from the most suitable value (N=3 in our experiment). In case where the size of each subspace is too small, the corresponding subspaces from same subject have more probability to be mismatched by the slight discrepancies of the same body part between probe and gallery. Meanwhile, if the N is too large, the mean accuracy will be decreased due to overfitting problem.

As Shown in Figure 8, in accordance with the growth of L, the mean accuracy is increased and is stable in a range for each decided value of N. Thus we decided the parameter combination as N=3 and L=300, which is also lead to the best performance in our experiment.

Table 1. Maximum and minimum accuracy of proposed method with running ten times

Condition: L=300, N=3	Set A2	Set B	Set C	ALL
Maximum	95.83% (23/24)	77.42% (96/124)	76.83% (126/164)	78.53% (245/312)
Minimum	87.50% (21/24)	69.35% (86/124)	73.78% (121/164)	73.08% (228/312)
Mean	93.33%	71.85%	75.43%	75.38%

Table 2. Performance comparison by algorithm

Methods	Set A2	Set B	Set C	ALL
GEI [2]	83.33% (20/24)	31.45% (39/124)	36.59% (60/164)	38.14% (119/132)
GEV [3]	91.86% (22/24)	55.65% (69/124)	62.20% (102/164)	61.86% (193/312)
DGHEI [4]	91.86% (22/24)	42.74% (53/124)	73.78% (121/164)	62.82% (196/312)
Proposed	93.33%	71.85%	75.43%	75.38%

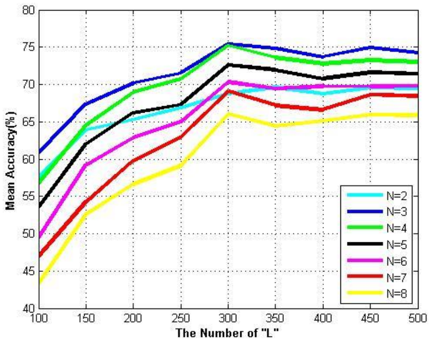


Figure 8. Performance comparison with respect to N

6. Conclusion

This paper introduced an effective method, which is called Adaptive Random Depth Subspace (ARDS) framework in robust depth gait recognition for individual. Different form the related method of depth gait recognition, we extend the whole-based feature extraction to part-based selection for adjusting the walking conditions variation. For real-world application, we combine the adaptive feature selection technique with the random selection procedure to assign the number of subspaces for each body part. Experimental results showed superior performance to the existing methods. In this work, body-part segmentation is accomplished by the premise of anatomical body properties, which may cause the mismatch due to the body proportion variation of the subjects. Open issues include the investigation of other techniques for dynamic body-part detection and segmentation.

Reference

1) S. A. Niyogi and E. H. Adelson, “Analyzing and recognizing walking figures in XYT,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 469-474, 1994.

2) J. Han and B. Bhanu, “Individual recognition using gait energy image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 316-322, 2006.

3) S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “Gait energy volumes and frontal gait recognition using depth images,” *Proceedings of International Joint Conference on Biometrics*, pp. 1-6, 2011.

4) M. Hofmann, S. Bachmann, and G. Rigoll, “2.5D Gait biometrics using the Depth Gradient Histogram Energy Image,” *Proceedings of IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, pp. 399-403, 2012.

5) M. S. Islam, M. R. Islam, M. S. Akter, M. A. Hossain, and M. K. I. Molla, “Window based clothing invariant gait recognition,” *Proceedings of International Conference on Advances in Electrical Engineering*, pp. 411-414, 2013.

6) N. Li, Y. Xu, and X. Yang, “Part-based human gait identification under clothing and carrying condition variations,” *Proceedings of International Conference on Machine Learning and Cybernetics*, pp. 268-273, 2010.

7) C. Yam, M. S. Nixon, and J. N. Carter, “Automated person recognition by walking and running via model-based approaches,” *Pattern Recognition*, pp. 1057-1072, 2004.

8) W. Dempster and G. Gaughran, “Properties of body segments based on size and weight,” *American Journal of Anatomy*, pp.33-54. 1967.

9) N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 886-893, 2005.

10) Tracking Users with Kinect Skeletal Tracking: <http://msdn.microsoft.com/en-us/library/jj131025.aspx>