Person Identification Using Full-Body Motion and Anthropometric Biometrics from Kinect Videos

Brent C. Munsell[1], Andrew Temlyakov[2], Chengzheng Qu[2], and Song Wang[2]

[1] Claflin University, Orangeburg, SC. 29115
[2] University of South Carolina, Columbia, SC. 29208
bmunsell@claflin.edu, {temlyaka,quc,songwang}@cec.sc.edu

Abstract. For person identification, motion and anthropometric biometrics are known to be less sensitive to photometric differences and more robust to obstructions such as glasses, hair, and hats. Existing gait-based methods are based on the accurate identification and acquisition of the gait cycle. This typically requires the subject to repeatedly perform a single action using a costly motion-capture facility, or 2D videos in simple backgrounds where the person can be easily segmented and tracked. For person identification these manufactured requirements limit the use gait-based biometrics in real scenarios that may have a variety of actions with varying levels of complexity. We propose a new person identification method that uses motion and anthropometric biometrics acquired from an inexpensive Kinect RGBD sensor. Different from previous gait-based methods we use all the body joints found by the Kinect SDK to analyze the motion patterns and anthropometric features over the entire track sequence. We show the proposed method can identify people that perform different actions (e.g. walk and run) with varying levels of complexity. When compared to a state-of-the-art gait-based method that uses depth images produced by the Kinect sensor the proposed method demonstrated better person identity performance.

1 Introduction

The development of accurate and efficient person identification methods is a major area of research in the computer vision, biometric, surveillance, and security communities. In general, person identification is typically achieved by measuring and analyzing biologic features, or *biometrics*, where a biometric is some distinguishing characteristic used for recognition. For example, high impact research has been conducted that identify people with similar facial, fingerprint, or iris biometrics [1–6] in 2D images or videos. To date, person identification systems that incorporate one, or more [7, 8], of these biometrics tend to dominate the community. However, the person being identified is usually required to physically touch the sensor, or cooperate with the sensor when acquiring data. Also, in real imagery the accurate identification and location of these biometrics can

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part III, LNCS 7585, pp. 91–100, 2012.
© Springer-Verlag Berlin Heidelberg 2012

be sensitive to photometric differences and obstructions (e.g., glasses, hair, hats) which may severely degrade recognition performance.

To overcome these limitations, person identification methods that use a lower extremity (i.e. below the hips) gait biometric have been proposed that attempt to identify people with similar lower extremity gait kinematics [9–11], stride and cadence [12, 13], and mechanics [14, 15]. Even though these gait-based methods are less restrictive and more robust to obstructions they do require the accurate detection of the motion region, and the coherent segmentation of the object (i.e. person boundary or silhouette) over a specified time sequence to isolate the gait cycle. In [16] the isolated gait cycle is used to construct normalized (i.e. temporally averaged) energy volumes, in [17] the isolated gait cycle is used to align the motion samples, and in [18] gait specific features are normalized using the isolated gait cycle . In real video footage that may have a variety of actions with varying levels of complexity, isolating the gait cycle can be computationally expensive and error prone. If the gait cycle is not properly detected and isolated, gait alignment and normalization errors are likely to be introduced, which may result in very poor recognition performance.

We present a novel person identity method that uses full-body (upper and lower extremity) motion and anthropometric biometrics derived from Kinect videos. The Kinect sensor was chosen because it is an inexpensive, easy-to-use, and accurate 3D motion sensor that is not sensitive to photometric differences. The major contribution is three-fold: 1. We introduce a new motion biometric that examines the coordinated motion of the entire body when a person performs a basic action, 2. The derived motion biometric examines the periodic motion of the tracked joints over the entire track sequence making the proposed motion biometric more robust than those derived by gait-based methods that are sensitive to gait cycle isolation errors, and 3. We introduce an integrated anthropometric biometric to boost biometric authentication if the motion biometric is unable to distinguish two people with similar full-body motion patterns. Unlike existing methods that may use anthropometric data to improve object tracking [18] or pose estimation [17] prior to biometric authentication, the proposed motion and anthropometric biometrics are combined to form a unified person identity classifier.

In the experiments, challenging scenarios are performed that study the subtle difference in motion among 10 different people when they perform 2 basic actions (walking and running). In total we collect 100 short[1] Kinect videos (40 videos for training and 60 for testing) resulting in an average ROC Equal Error Rate and Cumulative Match Curve Rank-1 identification rate of 13% and 90% respectively. Using the same Kinect data set, the proposed method is compared to the Gait Energy Volume (GEV) [16] a state-of- the-art lower extremity gait-based method and the experiments show our method out performs GEV. The remainder of this paper is organized as follows: Section-2 the proposed method is described in detail, Section-3 experiments are performed that evaluate person identity performance, and in Section-4 a brief conclusion is given.

[1] Roughly 30 sec of footage at 30 fps

2 Proposed Method

In this section we develop new classification methods that attempt to identify the basic action and identity of unknown persons in Kinect videos. Conceptually, the proposed method is accomplished using the two-stage classification system illustrated in Fig. 1(b). In the first stage, a test Kinect video with an

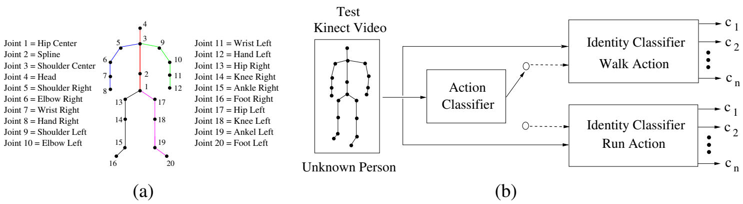

Fig. 1. (a) The 20 skeletal joints found using the Kinect SDK. (b) Two-stage classification system where the first stage recognizes the action and the second stage recognizes the identity of an unknown person in a Kinect video.

unknown action is input into a trained action classifier that is capable of recognizing two basic actions: walking and running. To accomplish this, a training set of Kinect videos are collected that capture subjects, with different gender, executing two basic actions. For each frame, in each training video, the normalized 3D locations of the 20 skeletal joints, illustrated in Fig. 1(a), are projected into a high-dimensional space and hyperplanes that best separate the two actions are found by a Support Vector Machine (SVM). Given a test Kinect video the learned hyperplanes allow us to classify each frame in the video and then the unknown action is recognized using a majority vote algorithm.

In the second stage, the test Kinect video is input into an identity classifier matched to the recognized action. For each basic action, we train n human identity classifiers by considering motion patterns and human anthropometric measures for n different people. In particular, the motion biometric is trained using Kinect videos that describe the radial, azimuth, and elevation motion patterns of 20 skeletal joints performing the same action multiple times. Likewise, the anthropometric biometric is trained using the same Kinect videos, however this biometric is a statistical model that describes the proportions between the 20 skeletal joints. Finally, n identity costs $c_1, c_2, \ldots, c_n$ are calculated and the unknown person in the test Kinect video is recognized by finding the score with the smallest value.

2.1 Action Classification

Let $\mathcal{V}^w = \{\mathcal{V}_i^w;\ i = 1, 2, \ldots, m\}$ be a set of walking and $\mathcal{V}^r = \{\mathcal{V}_i^r;\ i = 1, 2, \ldots, m\}$ be a set of running training Kinect videos, where $\mathcal{V}_i^w = (\mathcal{F}_{i1}^w, \mathcal{F}_{i2}^w, \ldots, \mathcal{F}_{in}^w)$ is an ordered sequence of image frames that capture various people performing normal walk actions. Using each frame in $\mathcal{V}_i^w$ a $60 \times n$

dimension skeletal matrix $\mathrm{S}_i^w = [\mathbf{s}_{i1} \ \cdots \ \mathbf{s}_{ik} \ \cdots \ \mathbf{s}_{in}]$ is constructed, where $\mathbf{s}_{ik}$ is a column vector that defines the 3D locations of the 20 skeletal joints in the kth frame[2]. This is repeated for each of the m videos in $\mathcal{V}^r$, and the resulting skeletal matrices are concatenated to form one matrix $\mathrm{S} = [\mathrm{S}_1^w \ \mathrm{S}_2^w \ \cdots \ \mathrm{S}_m^w \ \mathrm{S}_1^r \ \mathrm{S}_2^r \ \cdots \ \mathrm{S}_m^r]$ with dimension $60 \times 2nm$. The combined skeletal matrix S is decomposed $\hat{\mathrm{S}} = \mathrm{U}\Sigma\mathrm{D}^T$ into a set of matrices using singular value decomposition, where D is a $2nm \times 2nm$ dimension matrix of right singular vectors, Σ is a $60 \times 2mn$ dimension diagonal matrix of singular values, and U is a 60×60 dimension matrix of left singular vectors. Since most of the singular values are very small or zero, only the 10 largest singular values are considered. Therefore the reduce space is now represented by $\hat{\Sigma}$ that has dimension 10×10, $\hat{\mathrm{D}}$ that has dimension $2nm \times 10$, and $\hat{\mathrm{U}}$ that has dimension 60×10. The reduced dimension row vectors in $\hat{\mathrm{D}}$ are then used to train a multi-class SVM, and because the data is not linearly separable a non-linear Gaussian Radial Basis Function kernel learns the hyperplanes used in classification.

Given a test Kinect video $\mathcal{V}$ the reduced dimension matrices $\hat{\Sigma}$ and $\hat{\mathrm{U}}$ are used to insert each frame $\{\mathcal{F}\}_{k=1}^n$ into the space spanned by the row vectors in $\hat{\mathrm{D}}$ using $\hat{\mathbf{d}}_k = \mathbf{s}_k^T \hat{U} \hat{\Sigma}^{-1}$, where vector $\mathbf{s}_k$ contains the origin translated 3D locations of the 20 skeletal joints in frame $\mathcal{F}_k$. Once each frame in $\mathcal{V}$ is inserted into the high dimensional space they are labeled as a walk or run action by the trained SVM. The action is then recognized using a majority vote algorithm by determining the action that appears the most among all n frames.

2.2 Identity Classification

Given a test Kinect video of an unknown person X performing a recognized action, the identity cost

$$\min_{\forall P}\{\ \Delta_M(X,P) \cdot \Delta_A(X,P)\ \}$$

is calculated for each person P in the training data set, where $\Delta_M(X,P)$ is the motion difference and $\Delta_A(X,P)$ is the anthropometric difference. The identity of person X is recognized when the person with the least cost is found.

The motion biometric is trained as follows: Let $\mathcal{V}^w = (\mathcal{F}_1^w, \mathcal{F}_2^w, \ldots, \mathcal{F}_n^w)$ be a training Kinect video that captures a person performing the walk action. Like Section-2.1, $\mathrm{S}^w = [\mathbf{s}_1 \ \cdots \ \mathbf{s}_k \ \cdots \ \mathbf{s}_n]$ is constructed using each frame in the video, however column vector $\mathbf{s}_k$ now defines the r radius, θ azimuth, and ϕ elevation values of the 20 skeletal joints in the kth frame[3], which is then seperated into three different $20 \times n$ dimension matrices namely: M_r the radius matrix, M_ϕ the elevation matrix, and M_θ the azimuth matrix. Specifically, row vector $\mathbf{r}_1 = (r_{11}, r_{12}, \ldots, r_{1n})$ defines the radial motion, $\theta_1 = (\theta_{11}, \theta_{12}, \ldots, \theta_{1n})$ defines

[2] Translation differences are removed by picking joint-2 as the origin $(0,0,0)$ and the remaining 19 joints are translated relative this joint. E.g., $\mathbf{s}_{ik} = ((x_{1k} - x_{2k}), (y_{1k} - y_{2k}), (z_{1k} - z_{2k}), \ldots, (x_{20k} - x_{2k}), (y_{20k} - y_{2k}), (z_{20k} - z_{2k}))^T$

[3] $\mathbf{s}_k = (r_{1k}, \theta_{1k}, \phi_{1k}, \ldots, r_{20k}, \theta_{20k}, \phi_{20k})^T$

the angular azimuth motion, and $\phi_1 = (\phi_{11}, \phi_{12}, \ldots, \phi_{1n})$ defines the angular elevation motion of joint-1 when the person executes the walk action.

For each motion matrix, a motion histogram is constructed using Algorithm-1. For example, given M_r and k_f, in *line-2* $\mathbf{h} = (h_1, h_2, \ldots, h_{N/2})$ a $N/2$ dimension vector is created, where F_s is the sampling frequency and h_1 corresponds to $F_s/N = 0.5$ Hz. Likewise, $h_2 = 1$ Hz, $h_3 = 1.5$ Hz and so forth. In *line-5* a N-point Discrete Fourier Transform (DFT) is performed using the radial values in row i, and then in *line-6* a k_f dimension vector $\mathbf{b}$ of bin values are found for the top k_f frequencies that have the largest magnitude (sorted in descending order). In *lines 7-9*, for each frequency bin the corresponding histogram bin is incremented by one, and then on line *line-11* the radial motion histogram is normalized by n the total number of rows in the motion matrix.

Algorithm 1. Histogram(M, k_f)

1: $N \leftarrow 2048$
2: $F_s \leftarrow N/2$
3: $\mathbf{h} \leftarrow zeros(F_s)$
4: **while** $i \leq n$ **do**
5: $\quad Y \leftarrow |\ DFT(M(i,:), N)\ |$, $i = i + 1$
6: $\quad \mathbf{b} \leftarrow \mathrm{sort}(Y, k_f)$
7: $\quad$ **while** $j \leq k_f$ **do**
8: $\quad\quad \mathbf{h}[\mathbf{b}(j)] \leftarrow \mathbf{h}[\mathbf{b}(j)] + 1$, $j = j + 1$
9: $\quad$ **end while**
10: **end while**
11: $\mathbf{h} = \mathbf{h}/n$

For each person in the training data set 6 motion histograms are computed using Algorithm-1 (i.e. 3 histograms for each action). Given X an unknown person not in the training data set, and P a known person in the training data set the motion difference is calculated using

$$\Delta_M(X, P) = 1 - \frac{R(\mathbf{h}_r^x, \mathbf{h}_r^p) + R(\mathbf{h}_\theta^x, \mathbf{h}_\theta^p) + R(\mathbf{h}_\phi^x, \mathbf{h}_\phi^p)}{3}$$

where $R(\cdot)$ is the correlation coefficient, $(\mathbf{h}_r^x, \mathbf{h}_\theta^x, \mathbf{h}_\phi^a)$ are the motion histograms for X, and $(\mathbf{h}_r^p, \mathbf{h}_\theta^p, \mathbf{h}_\phi^p)$ are the motion histograms for P. The motion difference has a value in [0 1], where a value of 0 indicates the two people have identical motion patterns for the recognized action.

The anthropometric biometric is trained as follows: Let $\mathcal{V}^w = (\mathcal{F}_1^w, \mathcal{F}_2^w, \ldots, \mathcal{F}_n^w)$ be a training Kinect video that captures a person performing the walk action. For each frame in the video a 20×20 dimension joint proportion matrix is constructed using the (x, y, z) locations of the 20 skeletal joints. Specifically, joint proportion is calculated by $p_{ab} = \frac{d(a,b)}{d_{total}}$, where $d(a, b)$ is the

Euclidean distance between joints a and b, and $d_{total} = d(1,4)+d(3,8)+d(3,12)+d(1,16)+d(1,20)$ is the total skeletal distance[4]

The resulting joint proportion matrices are concatenated to form one matrix $\mathrm{J} = [\mathrm{J}_1^w \ \mathrm{J}_2^w \ \cdots \ \mathrm{J}_n^w]$ that has dimension $20 \times 20n$. A statistical model $\mathcal{N}(\bar{\mathbf{p}}, \mathrm{D})$ is constructed using the proportions in J, where $\bar{\mathbf{p}}$ is a 20 dimension vector that describes the mean proportions for all 20 joints, and D is a 20×20 covariance matrix that describes proportion variation for all 20 joints. For each person in the training data set 2 anthropometric statistical models are constructed (i.e. one model for each action). Given X an unknown person not in the training data set, and P a known person in the training data set the anthropometric difference is calculated using the well known KL-distance measure

$$\varDelta_A(X,P) = \frac{1}{2}\ \Big(\ \log \frac{|\mathrm{D}_p|}{|\mathrm{D}_x|}\ + Tr(\mathrm{D}_p^{-1}\mathrm{D}_x)\ + \\ (\bar{\mathbf{p}}_x - \bar{\mathbf{p}}_p)^T\ \mathrm{D}_p^{-1}\ (\bar{\mathbf{p}}_x - \bar{\mathbf{p}}_p)\ - d\Big)$$

where $\mathcal{N}(\bar{\mathbf{p}}_x, \mathrm{D}_x)$ is the anthropometric statistical model for unknown person X, and $\mathcal{N}(\bar{\mathbf{p}}_p, \mathrm{D}_p)$ is the learned anthropometric statistical model for person P, and $d = 20$ the dimension of the covariance matrix. In general, a small anthropometric difference value indicates the two people have very similar joint proportions for the recognized action.

3 Experiments

In this section experiments are performed that evaluate the proposed system's ability to correctly classify unknown people and their action in Kinect videos. The performance of the proposed method is compared to the Gait Energy Volume (GEV) method [16]. In general, GEV is the 3D extension of the 2D Gait Energy Image (GEI) [19]. Using the depth images, the tracked human silhouettes are segmented and the segmentation results are used to isolate each gait cycle in the video sequence. For each isolated gait cycle the results are aligned and averaged to form the GEV. Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) are used to find a reduced dimension feature vector that well describes the GEV. This unknown feature vector is compared to known feature vectors using a distance based measurement to recognize the identity of the tracked person in the Kinect video. In these experiments we manually identified the gait cycle and then used the recommended settings to perform PCA and MDA dimensionality reduction.

In Section 3.1 we describe the Kinect data sets used to train the action and identity classifiers, and in Section 3.2 we describe the Kinect data sets used to test the accuracy of the system and for performance comparison. Both data sets were collected using a Kinect sensor mounted on a movable cart that faced the person performing the action. During data collection the distance between the apparatus and the subject was roughly 1.5 to 3 meters. Lastly, we evaluate the

[4] For example, $p_{18} = (\ d(1,2) + d(2,3) + d(3,5) + d(5,6) + d(6,7) + d(7,8)\ )/d_{total}$.

performance of the action and person identity classifiers in Section 3.3 using the well known Receiver Operating Characteristic (ROC) curve and the Cumulative Match Curve (CMC) [20]. The ROC is also used to evaluate the sensitivity of the method when: 1. only one biometric is used for identity classification, and 2. the number of frequencies k_f (see Section 2.2) used to construct the (r, θ, ϕ) motion histograms are changed over a range of values (*Note: This is the only free parameter in the proposed person identification system*).

3.1 Training Data

The training data set included 10 people, 6 males and 4 females, where each person executed each of the 2 basic actions 2 times. For instance, each person has 4 Kinect videos: 2 walking and 2 running. In total, the training data set has 40 videos. Example 3D skeletons found by the Kinect sensor that illustrate the 2 basic actions are shown in Fig. 2. The activity classifier was trained using all 40 videos. For each walk identity classifier the motion and anthropometric biometrics were trained using the two collections for that person. Likewise, for each run identity classifer the motion and anthropometric biometrics were trained using the two collections for that person. For the 6 male subjects the age range was between 25 and 40 years old, and the height range was roughly between 1.73 and 1.8 meters. For the 4 female subjects the age range was between 25 and 37 years old, and the height range was roughly between 1.55 to 1.6 meters.

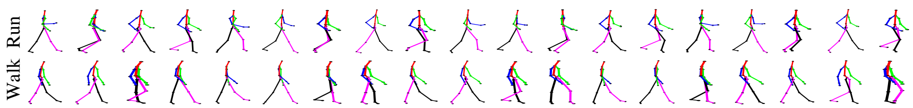

Fig. 2. Example training data. *Top row*: example skeletons of a person performing a normal running action. *Bottom row*: example skeletons of a person performing a normal walking action.

3.2 Test Data

Using the same 10 people in the training data set, each person in the test data set executed each of the 2 basic actions 3 times, i.e. each person has 6 Kinect videos: 3 walking and 3 running. In total, the test data set has 60 videos. To make the test data set challenging, each person was asked to perform the additional actions: *First Collection* (least challenging) wear a backpack that contained 20 lbs of books, *Second Collection* (moderately challenging) wear the same 20 lb backpack, and carry an object in their right hand, and *Third Collection* (most challenging) perform the slow moving “S” motion shown in Fig. 3. In general, these collections simulate real scenarios that may be found in public gathering areas such as airports, train stations, or shopping malls.

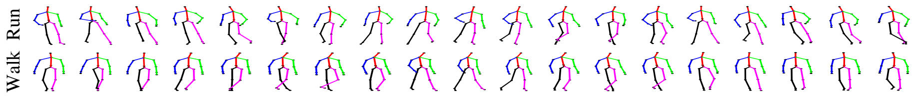

Fig. 3. Example Third Collection testing data. *Top row*: example skeletons of a person performing the "S" the running action. *Bottom row*: example skeletons of a person performing the "S" walking action.

3.3 Results

Action classification performance using majority vote was 100% for both actions, where action classification performance per video ranges from 50.30% to 100% with the average being 93.47%. The ROC curves in Fig. 4(a) show the Verification Rate (VR) and Equal Error Rate (EER) performance for our method and GEV. This figure also shows the CMC Rank-1 through 6 performance for our method and GEV. For both actions the EER and CMC Rank-1 performance of our method is better than GEV. For the walk and run actions our method shows a 11% and 4% EER increase in performance respectively. For the walk action our method is 90% accurate by Rank-3, whereas GEV is still hovering around 88% by Rank-6, and for the run action the Rank-1 performance of our method is 90% while GEV does not achieve 90% until Rank-3. The ROC curves in this figure also show the motion biometric is not overly sensitive to k_f the number of frequencies used to construct the motion histograms. In fact, the ROC curves show roughly the same performance when k_f is three or five times greater than 10. This suggests the joint motion patterns may be adequately described by the top 10 frequencies with the largest magnitude.

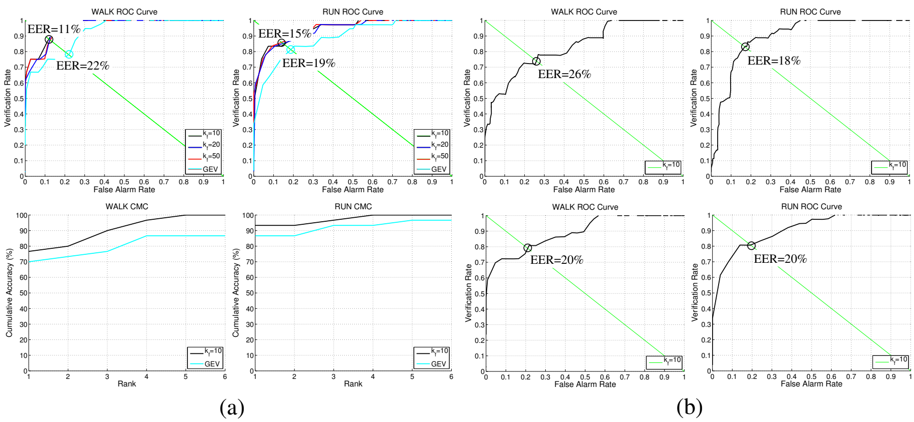

Fig. 4. For both actions **(a)** *Top Row*: VR and EER performance comparison between our method ($k_f = 10, 30, 50$) and GEV [16]. *Bottom Row*: CMC Rank-1 through 6 performance comparison between our method ($k_f = 10$) and GEV. **(b)** *Top Row*: VR and EER performance using only the motion biometric ($k_f = 10$). *Bottom Row*: VR and EER performance using only the anthropometric biometric ($k_f = 10$).

Figure 4(b) shows the VR and EER performance when only the motion and anthropometric biometric is used by the identity classifier. As seen in these ROC curves, person identification is more accurate when both biometrics are used by the identity classifier. For the walk action the anthropometric EER performance is slightly better than the motion biometric, which suggests the anthropometric biometric guides the motion biometric. However, for the run action the discriminative power of the motion biometric is high, requiring less help from the anthropometric biometric.

Since the computation complexity of the SVD and SVM algorithms are $\mathcal{O}(pq^2 + p^2q + q^3)$ [21] and $\mathcal{O}(q^2)$ [22] respectively, the computational complexity of the action classifier $\approx \mathcal{O}(q^3)$ where $q = 2mn$. The space complexity of the action classifier is $\mathcal{O}(q^2)$, i.e. the size of the right singular value matrix D. An analysis of Algorithm-1 shows the computational complexity of the identity classifier is $\mathcal{O}(nNlogN)$, and the space complexity of the identity classifier is $\mathcal{O}(n)$, i.e. the column dimension of the radius, azimuth, and elevation matrices. On a 2.4GHz Intel Core 2 Quad CPU, the total time needed to train the action classifier was 32 min, and the time needed to train one identity classifer was 30 ms.

4 Conclusion

In conclusion, a novel person identity method that uses full-body motion and anthropometric biometrics derived from Kinect videos was presented. Different form traditional gait-based methods that attempt to isolate and examine the gait cycle in the video sequence, our method considers the entire track sequence and examines the periodic motion of upper and lower extremity joints found by the Kinect SDK that have the largest contribution to the action being performed. Challenging test data sets where constructed that have a variety of basic actions with varying levels of complexity. Experiments showed that the proposed method has an average ROC EER of 13% and an average CMC Rank-1 identification rate of 90%. Performance comparisons where conducted using a gait-based method that uses depth images produced by the Kinect sensor. The results showed our method to have better performance. Experiments were also conducted to assess the individual sensitivities of the two biometrics, and the results suggest both biometrics are needed for person identification. We also show the motion biometric is not overly sensitive to the number of frequencies used to build the motion histograms.

References

1. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
2. Jain, A., Hong, L., Bolle, R.: Online fingerprint verification. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 302–314 (1997)

3. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1519–1533 (2003)
4. Ross, A., Dass, S., Jain, A.: Fingerprint warping using ridge curve correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 19–30 (2006)
5. Lu, X., Jain, A.: Deformation modeling for robust 3D face matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 1346–1357 (2008)
6. Pillai, J., Patel, V., Chellappa, R., Ratha, N.: Secure and robust iris recognition using random projections and sparse representations. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 1877–1893 (2011)
7. Hong, L., Jain, A.: Integrating faces and fingerprints for personal identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1295–1307 (1998)
8. Chang, K., Bowyer, K., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1160–1165 (2003)
9. Murase, H., Sakai, R.: Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern Recognition Letters 17, 155–162 (1996)
10. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 781–796 (2000)
11. Boyd, J.E., Little, J.J.: Biometric Gait Recognition. In: Tistarelli, M., Bigun, J., Grosso, E. (eds.) Advanced Studies in Biometrics. LNCS, vol. 3161, pp. 19–42. Springer, Heidelberg (2005)
12. Gafurov, D., Helkala, K., Søndrol, T.: Biometric gait authentication using accelerometer sensor. JCP 1, 51–59 (2006)
13. Abdelkader, C.B., Davis, L., Cutler, R.: Stride and cadence as a biometric in automatic person identification and verification. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 372–377 (2002)
14. Campbell, L., Bobick, A.: Recognition of human body motion using phase space constraints. In: International Conference on Computer Vision, pp. 624–630 (1995)
15. Little, J., Boyd, J.E.: Recognizing people by their gait: The shape of motion. Videre 1, 1–32 (1996)
16. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.B.: Gait energy volumes and frontal gait recognition using depth images. In: International Joint Conference on Biometrics (2011)
17. Gu, J., Ding, X., Wang, S., Wu, Y.: Action and gait recognition from recovered 3-D human joints. Trans. Sys. Man Cyber. Part B 40, 1021–1033 (2010)
18. Green, R.D., Guan, L.: Quantifying and recognizing human movement patterns from monocular video images - part ii: Applications to biometrics. IEEE Transactions on Circuits and Systems for Video Technology 14, 179–190 (2003)
19. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 316–322 (2006)
20. Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: The relation between the roc curve and the cmc. In: Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp. 15–20 (2005)
21. Brand, M.: Incremental Singular Value Decomposition of Uncertain Data with Missing Values. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part I. LNCS, vol. 2350, pp. 707–720. Springer, Heidelberg (2002)
22. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using second order information for training support vector machines. Journal of Machine Learning Research 6, 1889–1918 (2005)