المجلة العراقية للهندسة الكهربائية والالكترونية
المجلد 14، العدد 1-2018

The Effect of Using Projective Cameras on View-Independent Gait Recognition Performance

Fatimah S. Abdulsattar
Department of Computer Engineering
University of Al-Mustansiriyah
Iraq, Baghdad
fsa.abdulsattar@uomustansiriyah.edu.iq
ORCID ID/0000-0001-8918-6777

Abstract *Gait as a biometric can be used to identify subjects at a distance and thus it receives great attention from the research community for security and surveillance applications. One of the challenges that affects gait recognition performance is view variation. Much work has been done to tackle this challenge. However, the majority of the work assumes that gait silhouettes are captured by affine cameras where only the height of silhouettes changes and the difference in viewing angle of silhouettes in one gait cycle is relatively small. In this paper, we analyze the variation in gait recognition performance when using silhouettes from projective cameras and from affine cameras with different distance from the center of a walking path. This is done by using 3D models of walking people in the gallery set and 2D gait silhouettes from independent (single) cameras in the probe set. Different factors that affect matching 3D human models with 2D gait silhouettes from single cameras for view-independent gait recognition are analyzed. In all experiments, we use 258 multi-view sequences belong to 46 subjects from Multi-View Soton gait dataset. We evaluate the matching performance for 12 different views using Gait Energy Image (GEI) as gait features. Then, we analyze the effect of using different camera configurations for 3D model reconstruction, the GEI from cameras with different settings, the upper and lower body parts for recognition and different GEI resolutions. The results illustrate that low recognition performance is achieved when using gait silhouettes from affine cameras while lower recognition performance is obtained when using gait silhouettes from projective cameras.*

Keywords: **3D reconstruction, view-independent, gait recognition, Gait Energy Image, K-Nearest Neighbor.**

I. Introduction

Recognizing people by the way they walk, also known as gait recognition, achieve considerable attention from the research community due to its attractive advantage in identifying people at a distance and without their cooperation when other biometric cues, such as face, iris and fingerprint, are inoperative. Gait is hard to conceal and forge. It is also unobtrusive and unique for each person. Furthermore, it is easy to capture where normal cameras can be used to record people’s gait. All these advantages give priority for the gait to be used in many applications such as visual surveillance, security and forensic.

There are two approaches in gait recognition: model-based [1, 2] and appearance-based [3, 4]. The former extract static (e.g. lengths of upper and lower legs) and dynamic (e.g. joint angles) gait features by tracking and modeling the movements of different body parts. These approaches are more robust to appearance changes. However, they require extensive computations and accurate silhouettes.

The latter extract spatio-temporal patterns as gait features. These approaches are more sensitive to appearance changes but they are simpler and do not require any modelling. One of the most widely-used gait representations is Gait Energy Image (GEI) [4] which encodes shape and motion information in a single grayscale image by averaging all silhouettes over one gait cycle. A similar representation to the GEI is Motion Silhouette Image (MSI) [5] where pixel intensity in MSI is computed as a function of motion history of that pixel in one gait cycle. Another representation, called Gait Entropy Image (GEnI) [6], computes Shannon entropy for each pixel over one gait cycle. This representation shows better performance against covariate condition changes. Most of these approaches record high recognition rates when gait sequences are captured from the same viewing angles, while their performance degrades when viewing angles are different.

To deal with the problem of view dependency efficiently, several approaches have been developed but the majority of these techniques assume that view variation in a single gait cycle is small where the viewing angle from the first to last silhouettes in a single gait cycle slightly changes. Iwashita et al. [7] used 3D model sequences with adaptive virtual image synthesis to identify people walking along curved trajectories from single cameras. In this technique, the walking direction for a subject is estimated in each frame and the corresponding virtual image is synthesized from a 3D model. The GEI and the Affine Moments Invariants are used as gait features. The performance is evaluated on the Kyushu University 4D Gait Database which consists of 42 subjects. Muramatsu et al. [8] proposed a framework to identify subjects from an arbitrary view using a view transformation model (VTM) with 3D model sequences. First, 3D models of training subjects, which are different from the test subjects in the target scene, are reconstructed and then projected onto the target views to generate 2D gait sequences. Second, gait features are extracted and used for part-dependent view selection scheme. The technique recorded an improved accuracy for cross-view matching. López-Fernández et. al. [9] introduced appearance-based gait descriptors extracted from 3D human models. A gait signature is built using morphological analysis of a series of 3D occupation human models, which is then classified using a Support Vector Machine with a sliding temporal window for majority voting. This strategy is validated on the AVA Multi-View Gait Dataset and on the Kyushu University 4D Gait Database, and achieved good recognition results.

Although the previous techniques used 3D models, also called 3D gait recognition techniques, to solve the problem of view-dependency, they have the following limitations: (1) the technique in [7] did not analyze the effect of using independent views in the probe set where all the target (test) cameras are also used to build 3D models of walking people in the gallery (training) set, (2) the work in [8] did not analyze the effect of using cameras with different settings (i.e. affine and projective cameras, far-away and near from a walking subject, varying camera lenses,…etc.) and (3) the authors in [9] used 3D models of people in both gallery and probe sets which make gait analysis techniques impractical to be applied in practice since they assume that multiple synchronous cameras are available always.

Therefore, the main objective of this work is to (1) analyze the variation in recognition performance when using gait silhouettes from affine cameras and gait silhouettes from projective cameras, (2) analyze the performance of using gait silhouettes from independent cameras with different settings in the probe set while using 3D models from multiple synchronous cameras in the gallery set and (3) study the effect of various factors that affect the performance of matching 3D gait models with 2D gait silhouettes using GEI as gait features. More specifically, the following factors will be analyzed.

1- Which camera configuration is effective for matching?

2- Can traditional period-based gait feature extraction (e.g. GEI) deal efficiently with silhouettes captured with different camera settings?

3- Can efficiently be identified subjects at low resolutions using their gait?

4- Which part in the body is effective for person identification?

II. GAIT DATASET

Gait recognition is a subset of action recognition since gait is a behaviour biometric. A few multi-view datasets are designed for action recognition and some of them are synchronous and designed specifically for gait recognition. The Kyushu University 4D gait dataset [7] which consists of 3D models and 2D silhouettes captured by 16 synchronized cameras. The 3D models and 2D silhouettes are dependent because all the cameras that captured the silhouettes are used to build 3D models. The cameras are also arranged around a circular studio so that they have the same distance to walking subjects and therefore they have the same settings. Another one is the AVA Multi-View Dataset for gait recognition [9]. This dataset includes 20 subjects recorded by 6 synchronized cameras. The number of cameras in this dataset is not enough to build 3D models and 2D gait silhouettes from independent cameras.

Therefore, all the experiments are implemented using Soton Multi-View Gait Dataset [10]. This dataset was captured using the University of Southampton multi-biometric tunnel. The tunnel consists of a narrow walkway in the middle surrounding by two walls, to represent a walkway in airports. There are 12 synchronized cameras to record people's gait as they walk from multiple viewpoints simultaneously. Four of them in the middle of the tunnel (close to a walking subject) have wide-angle lenses and capture images from approximately side-views. These cameras are modelled as projective cameras where silhouettes captured by these cameras are distorted according to their positions. Fig. 4 (b) shows several silhouettes in one gait cycle captured by one of the middle cameras in the tunnel.

The remaining cameras at the far ends of the tunnel have narrow-angle lenses and provide nearly front/rear views. These cameras are placed at two different heights and can be modeled as affine cameras. An example of silhouettes captured by one of the far-ends cameras is shown in Fig. 4(a). As can be seen only the height of the silhouettes varies in one gait cycle.

Fig. 1 shows the positions of the middle and far-ends cameras in the tunnel with their IDs. A subset of 43 subjects and a total of 258 sequences are used from this dataset in the experiments. Each subject has 6 walking sequences and each sequence was recorded using all the twelve cameras.

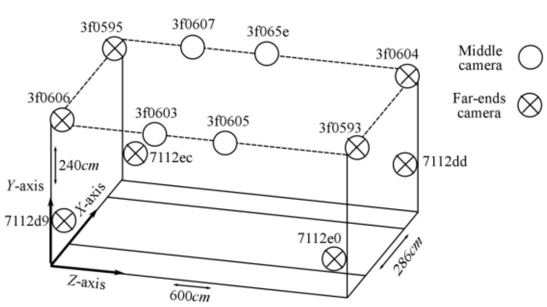


Fig. 1 The placement of the middle and far-ends synchronized cameras in the tunnel.

III. MATCHING PROCEDURE

In order to study the effect of 3D human model accuracy on the performance of 3D gait recognition techniques, the following assumptions are considered. **First,** multiple synchronized cameras are installed for access control scenarios where a subject walks along a straight line of a pre-determined path. Therefore, the cameras are distributed at the start, middle and end of the walking path [see Fig. 1]. **Second,** the cameras have two different settings as explained in section II. **Third,** all the cameras are calibrated (i.e. their intrinsic and extrinsic parameters are known).

The following procedure is implemented to study the performance of matching 3D models with 2D gait silhouettes using Soton multi-biometrics tunnel.

Step 1: One camera from the tunnel is selected each time as a target camera for testing and its silhouette sequences are used to build the testing dataset (probe) for recognition (identification).

Step 2: For view independence, 3D human models are reconstructed from all the cameras in the tunnel except the target camera using a visual hull strategy [11] by computing the intersection of back-projected silhouettes into 3D space.

Step 3: To build the training dataset (gallery), synthetic silhouettes are generated from the 3D models by projecting them onto the target camera's viewpoint using its intrinsic and extrinsic calibration information.

Step 4: Gait features are calculated from the synthetic silhouettes in the gallery and from the real silhouettes in the probe using Gait Energy Image (GEI) [4]. The GEI is a widely used gait representation in the literature due to its robustness and insensitivity to noise in a single silhouette. To compute the GEI, one gait cycle is extracted from each sequence which defines the pattern of walking for an individual. Each silhouette in one gait cycle is then cropped and resized to a same height by keeping the width-to-height ratio constant. Finally, the resulting silhouettes are averaged to obtain the GEI as follows

$$GEI = \frac{1}{C} \sum_{t=1}^{C} S(t) ....(1)$$

Where C is the number of silhouettes in one gait cycle and $S(t)$ is the silhouette image at t index. After that all GEIs are resized to 64×64 pixels for dimensionality reduction.

Step 5: For recognition, a distance matrix is computed by calculating the Euclidean distance between each pair of the GEIs in the gallery and probe as follows

$$D(i,j) = \left\| GEI_p^i - GEI_g^j \right\| = \sqrt{\sum_{x=1}^{X}\sum_{y=1}^{Y}\left(GEI_p^i(x,y) - GEI_g^j(x,y)\right)^2} \quad \text{..........(2)}$$

Where D is the distance matrix, GEI_p^i is the i^{th} GEI in the probe, GEI_g^j is the j^{th} GEI in the gallery and X and Y are the number of the GEIs in the probe and gallery respectively. K-Nearest Neighbor classification is then used to find the closest GEI in the gallery to each GEI in the probe and compute the final recognition results.

Step 6: To evaluate the recognition performance, the Correct Classification Rate (CCR) is used because it is a commonly used metric to evaluate the performance of a recognition problem. The CCR is defined as the percentage of the correctly classified samples in the probe.

IV. EXPERIMENTAL RESULTS

This section displays the recognition results of matching 3D human models with 2D gait silhouettes. The gallery consists of synthetic silhouettes that are generated from projecting 3D models onto the same viewing angles as the probe silhouettes while the probe consists of real silhouettes from the test (target) camera. Fig. 2 shows the CCR for the matching where the results are splited into two groups. High recognition performance (the average CCR is about 97%) is obtained for identifying the silhouettes from the far-ends cameras (3f0595, 3f0606, 3f0604, 3f0593, 7112ec, 7112d9, 7112d9, 7112dd, 7112e0) while low recognition performance (the average CCR is about 42%) is recorded for the silhouettes from the middle central cameras (3f0607, 3f0603, 3f065e, 3f0605). In order to better understand the variation in the CCR between the two groups of cameras, different factors affecting the performance of matching are analyzed in the following experiments.

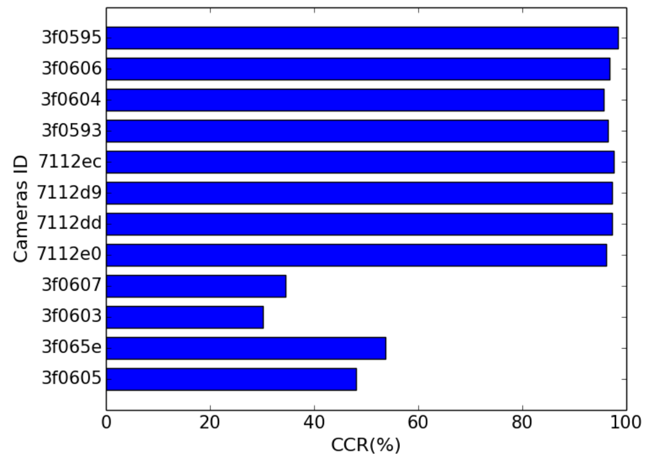


Fig. 2 Performance of matching synthetic silhouettes with real silhouettes.

Experiment A: Effect of Using Real Silhouettes in the Gallery

This experiment aims to illustrate the contribution of the GEI in the final gait recognition performance. The discriminatory power of the GEI as gait features is studied for different cameras in the tunnel by isolating the effect of other factors such as the accuracy of synthetic silhouettes. This is done by using the real silhouettes from the test camera in both gallery and probe. The GEI is computed from each sequence as described in section III and the recognition results are summarized in Fig. 3. The results using real silhouettes only are better than using synthetic silhouettes in the gallery for all cameras.

The improvement in the CCR is about 3% for the far-ends cameras and is about 38% for the middle cameras. It is believed that the reason for this improvement is because of using the same type of silhouettes (i.e. real silhouettes) in both gallery and probe. However, the CCR of the middle cameras is still lower than that of the far-ends cameras. Fig. 4 shows an example for several silhouettes in one gait cycle from two different cameras in the tunnel with their GEIs where there is an obvious transition in the orientation, local viewing angle and appearance of silhouettes from the middle camera as compared to the far-end camera. As a result, the relevant body parts cannot efficiently be matched when computing the average image in the GEI. This means that the GEI cannot capture rapidly changing shapes. The results reveal that the kind of gait features is an effective factor on matching performance.

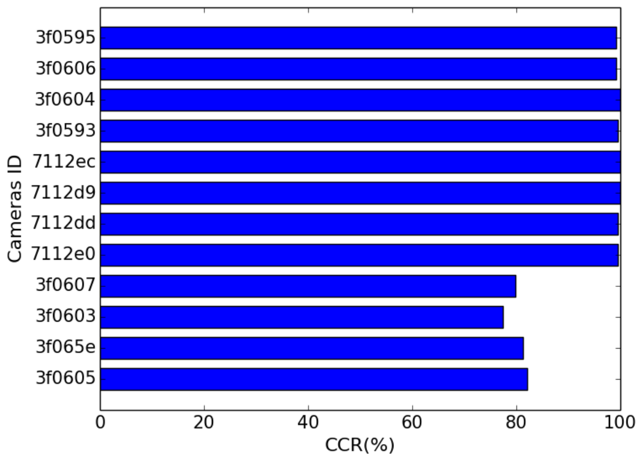


Fig. 3 Performance of using real silhouettes in the gallery and probe.

(a) Far-end camera (3f0593)

(b) Middle camera (3f0605)

Fig. 4 Several silhouettes (left) with their GEIs (right) from two different cameras.

Experiment B: Effect of Camera Configurations

Another factor that affects the matching performance is the accuracy of the 3D human models. As the accuracy of the 3D models is higher, the accuracy of the resulting synthetic silhouettes will be higher and the difference between the synthetic silhouettes and the real silhouettes from the camera will be minimized. In this experiment, three different camera configurations are used to reconstruct 3D models for the gallery. The first configuration (A) includes the 8 top cameras, the second configuration (B) includes the 8 far-ends cameras and the last configuration (C) consists of the 4 middle cameras and the 4 bottom cameras.

The 3D models are reconstructed according to each one of these configurations, and then projected onto the viewing angle of the probe camera to produce a synthetic silhouette.

To evaluate the matching performance, two different probe cameras are selected: the middle camera (3f0605) and the far-end camera (3f0593). The GEIs of the synthetic silhouettes are calculated and compared against the GEIs of the real silhouettes for comparison and the results are summarized in Table 1. As can be seen from the results, when the probe camera is used in the 3D reconstruction, the performance of matching improves especially for the middle camera (3f0605) in the configuration (A) and (C) as compared to the results in Fig. 2. However, in real world scenarios the target (probe) camera may not always be available in the main (gallery) dataset.

On the other hand, the matching performance deteriorates significantly when the 4 middle cameras are excluded from the 3D reconstruction in configuration (B) where the CCR for the middle camera (3f0605) decreases to only 9.3%.

A highly distorted 3D model is obtained when discarding the 4 middle (side-view) cameras in configuration (B) while no noticeable distortion can be achieved when excluding the far-ends cameras in configuration (A) and (C) as shown in Fig. 5, which highlights the importance of including the side-view cameras for a better quality of 3D reconstruction process. As a result, there is a high discrepancy between the GEI of the synthetic silhouettes produced from configuration (B) and the GEI of the real silhouettes from the middle camera (3f0605) while no big discrepancy when the synthetic silhouettes produced from configuration (C) in Fig. 6.

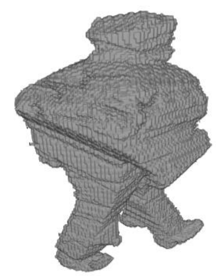
These results reveal that the camera configuration used in 3D reconstruction process has a great impact on matching performance and that the effective camera configuration should involve side-view cameras.

المجلة العراقية للهندسة الكهربائية والالكترونية
المجلد 14، العدد 1-2018

Table 1 CCR(%) using different camera configurations (A) 8 top cameras, (B) without middle cameras and (C) 4 middle and 4 far-bottom cameras.

Camera ID	Configuration (A)	Configuration (A)	Configuration (A)
3f0593	99.6	97.3	97.7
3f0605	81.4	9.3	76.4

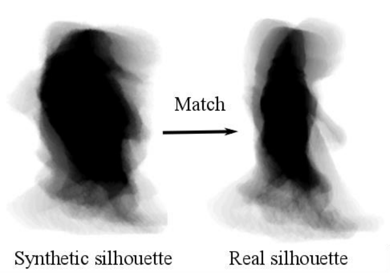
(a) Configuration (A)



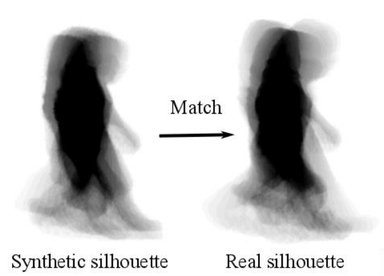
(b) Configuration (B)

(c) Configuration (C)

Fig. 5 3D human model reconstruction using different camera configurations.



(a) Configuration (B)



(b) Configuration (C)

Fig. 6 Example of the GEI for synthetic silhouettes on the left generated from (a) configuration B and (b) configuration C, and the GEI computed from the real silhouettes on the right for the middle camera (3f0605).

Experiment C: Effect of GEI Resolutions

In this experiment, the impact of reducing the resolution of the GEI is evaluated. As the resolution of the GEI decreases, the number of features available for recognition decreases. The GEIs of the synthetic and real silhouettes used in the experiment are resized from 64×64 to 32×32, 16×16 and 8×8 respectively. The results are illustrated in Fig. 7. In general, there is a slight degradation in performance when the GEI resolution decreases to half (32×32) and then to three-quarter (16×16). The degradation becomes sharper when the resolution decreases to the lowest level (8×8) since the number of features decreases. The degradation in performance is higher for the far-ends cameras than that for the middle cameras. This could be because the far-ends cameras are farther to the walking subjects than the middle cameras. However, the CCR of the former cameras is still higher than the latter. Reducing the resolution of the GEI image is equivalent to increasing the distance between the camera and the walking subject. These results illustrate that people's gait can still be used as an efficient biometric cue at low resolutions. Moreover, the recognition of people by their gait can be done using a fewer number of features.

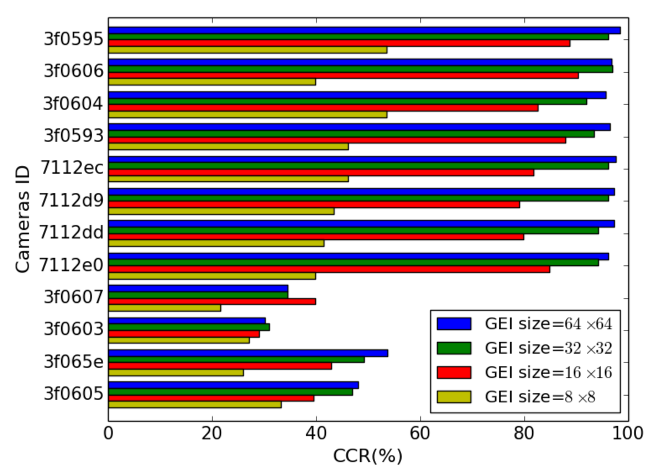


Fig. 7 CCR(%) versus GEI resolutions.

Experiment D: Effect of Upper and Lower Parts

This experiment aims to show which part in the GEI representation contributes more in gait recognition performance and is it possible to identify a subject using only part of his/her body when there is an obstacle preventing the camera from recording the whole shape. To do this analysis, the synthetic silhouettes are used in the gallery and the real silhouettes are used in the probe. Then, each GEI in both gallery and probe is divided horizontally into two equal parts: upper part (covers head and arm movement) and lower part (covers hip and leg movement). Each part is used alone for recognition and the results are illustrated in Fig. 8. As can be seen for the far-ends cameras, the lower body part which relates to the movement of the legs (human gait) has higher discriminatory power as compared to the upper part. The difference in the CCR is more than 10% between the two parts.

However, the effective body part for the middle cameras depends on the camera's viewing angle. The results indicate that it is possible to identify a subject efficiently using only his/her lower body part when the GEI is used as gait features.

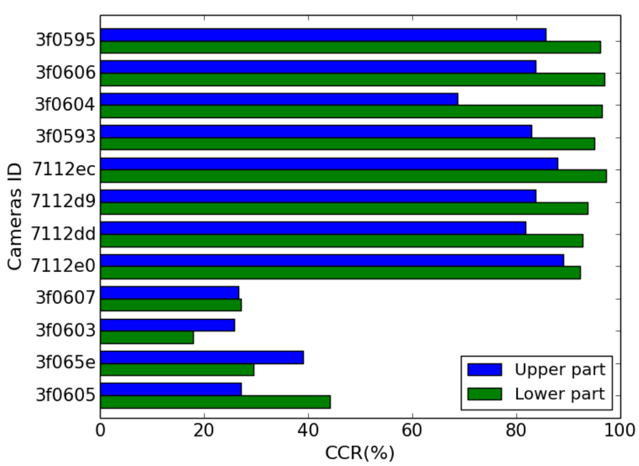


Fig. 8 CCR(%) of the upper and lower body parts.

Experiment G: Effect of Other Gait Representations

In this experiment, we evaluate the performance of other gait representations such as Gait Entropy Image (GEnI) [10] and Motion Silhouette Image (MSI) [11]. For comparison purpose, The GEI, GEnI and MSI are computed from the synthetic silhouettes in the gallery and real silhouettes in the probe for each camera in the tunnel and resized to 64×64 pixels.

The CCRs are illustrated in Fig. 9. As can be seen, the performance of the GEI and GEnI are better than that of the MSI especially for the far-ends cameras. However, there is still difference in performance between the two types of cameras in the tunnel where the middle cameras still show lower performance than far-ends cameras

In summary, the gait representations that encode motion information over one gait cycle in a single image cannot efficiently deal with rapidly changing shapes. More sophisticated matching procedures that depend on a frame-by-frame comparison may better deal with this type of gait data.

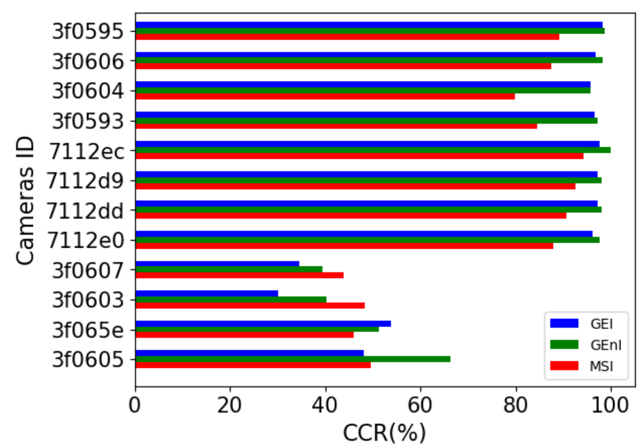


Fig. 9 Performance of different gait representations.

V. CONCLUSIONS

This paper analyzes the effect of using cameras with different settings on view-independent gait recognition. This is done by matching 3D models of walking people with 2D gait silhouettes from single cameras using GEI as gait features. The results showed that matching 3D human models can be done at high recognition performance with silhouettes from a narrow-angle and front/rear-view (affine) cameras and at lower recognition performance with silhouettes from a wide-angle and side-view (projective) cameras. Moreover, there are two important factors that affect the performance of matching. The first factor is the discriminatory power of the gait features. The second factor is the camera configuration used to build 3D human models. Further analyses show that the matching can be done at low resolutions and using only the lower body part when the GEI is used as gait features. Different matching algorithms and other types of features are required to confirm the results obtained in this work.

Acknowledgment

I would like to thank Dr. John Carter and Prof. Mark Nixon from the University of Southampton in the UK for their help and providing facilities to accomplish this work.

References

[1] A. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. I-I, 2001.

[2] C. Yam, M. S. Nixon and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057-1072, 2004.

[3] R. Sagawa, Y. Makihara, T. Echigo, and Y. Yagi, "Matching gait image sequences in the frequency domain for tracking people at a distance," *Computer Vision–ACCV 2006*, pp. 141-150, 2006.

[4] J. Man and B. Bhanu, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316-322, 2006.

[5] T. H. W. Lam and R. S. T. Lee. "A new representation for human gait recognition: Motion silhouettes image (MSI)." *International Conference on Biometrics*. Springer, Berlin, Heidelberg, 2006.

[6] K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image." 3rd *International Conference on Imaging for Crime Detection and Prevention (ICDP)*, UK, 2009.

[7] Y. Iwashita, K. Ogawara and R. Kurazume, "Identification of people walking along curved trajectories," *Pattern Recognition Letters*, vol. 48, pp. 60-69, 2014.

[8] D. Muramatsu, A. Shiraishi, Y. Makihara, M. Z. Uddin and Y. Yagi, "Gait-based person recognition using arbitrary view transformation model," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 140-154, 2015.

[9] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, M. J. Marín-Jiménez, R. Muñoz-Salinas and R. Medina-Carnicer, "Independent gait recognition through morphological descriptions of 3D human reconstructions," *Image and Vision Computing*, vol. 48, pp. 1-13, 2016.

[10] R. D. Seely, "On a three-dimensional gait recognition system," *PhD thesis*, University of Southampton, UK, 2010.

[11] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on pattern analysis and machine intelligence*," vol. 16, no. 2, pp. 150-162, 1994.