Abstract—In this paper, we propose a novel direction for gait recognition research by proposing a new capture-modality independent, appearance-based feature which we call the Backfilled Gait Energy Image (BGEI). It can be constructed from both frontal depth images, as well as the more commonly used side-view silhouettes, allowing the feature to be applied across these two differing capturing systems using the same enrolled database. To evaluate this new feature, a frontally captured depth-based gait dataset was created containing 37 unique subjects, a subset of which also contained sequences captured from the side. The results demonstrate that the BGEI can effectively be used to identify subjects through their gait across these two differing input devices, achieving rank-1 match rate of 100%, in our experiments. We also compare the BGEI against the GEI and GEV in their respective domains, using the CASIA dataset and our depth dataset, showing that it compares favourably against them. The experiments conducted were performed using a sparse representation based classifier with a locally discriminating input feature space, which show significant improvement in performance over other classifiers used in gait recognition literature, achieving state of the art results with the GEI on the CASIA dataset.

I. INTRODUCTION

There are number of biometrics being explored for robust automatic human identification. Gait, as a biometric, is attractive due to its ability to operate using low resolution imagery and can be acquired at a distance without alerting the subject. Hence, research in human identification using gait is becoming popular in authentication processes and forensic applications.

The performance of gait based algorithms is continuously improving, with recent algorithms achieving recognition rates of over 90%, albeit under constrained environments and with a limited number of subjects [1]. This recognition rate, however, can still be impacted by various factors, including environmental factors such as changes in view angle and lighting conditions; and subject-based factors, such as changes in clothing, the carrying of goods, and even variations in the subject’s mood.

There are two major approaches to gait recognition: model-based and appearance-based [1]. Model-based approaches represent the dynamic and static parameters of gait by modelling the intrinsic kinematics of human motion. These algorithms are less affected by changes in appearance in the subject, as well as changes in the environment (subject to model accuracy). They, however, can be computationally expensive. Examples of model-based gait recognition systems range from simple pendulum models with Fourier based features [2], to the extraction of fully articulated human skeletons using static (limb lengths) and dynamic (joint angles) gait features [3].

Appearance-based techniques take a holistic approach, attempting recognition based on observations without an explicit representation of underlying gait features. These techniques are usually much simpler as well as generally being computationally less expensive. They are however, more susceptible to environmental changes and appearance/pose changes in the subject. Examples of appearance based methods include temporal matching of silhouettes [4], as well as template methods such as the gait energy image (GEI) [5].

Gait energy based features are popular due to their high recognition rate, with various extensions being proposed [6], [7]. However, like many other appearance based features, recognition is best performed using video recorded from the side, as the gait features, particularly the motion of the legs, are best captured from that view. Limited gait recognition research has been performed from a frontal perspective [8] however, it is difficult due to its inability to comprehensively capture the dynamic details of gait from a 2D image, particularly when only silhouettes are used. The use of stereo cameras [9] and other depth based sensors (e.g. Microsoft Kinect) can overcome this issue, and may even prove to be superior
to a lateral view due to the lack of self occlusion in the legs. Sivapalan et al. [10] demonstrated that good recognition results can be achieved from frontal depth images through the use of the gait energy volume (GEV).

A frontal perspective has various advantages to that of the side, such as for use in narrow corridors, where the limited field-of-view of cameras may prevent the recording of complete gait cycles from the side. They can also be easily integrated into biometric portals such as that used in the Multiple Biometric Grand Challenge (MBGC) [11]. However, there are also situations where side view gait recognition is preferable, such as in surveillance where the distances involved may be unsuitable for many depth sensing devices, or where depth sensing hardware may simply not be present. These two different capture modalities operate in differing image domains, with the gait features used in existing approaches specific to each. This prevents sharing of information without the use of view transformation models.

In this paper, we present a new gait energy based feature that can be constructed from both side view silhouettes and frontal depth images. We call this the backfilled gait energy image (BGEI). This allows the feature to be applied across differing capturing systems using the same enrolled database, such as in a system using both frontal depth cameras mounted on biometric portals and general surveillance cameras as shown in Figure 1.

We explore the effectiveness of this proposed framework by experimentally demonstrating how the BGEI can be used to match subjects across the two modes. This is performed on a new database we have created, which contains 37 subjects under various walking conditions captured from the front using a depth camera. 8 of these subjects also have gait sequences recorded from the side in order for us to perform the cross-modality experiments.

Using this database, we also evaluate the BGEI against the GEV in intra-capture modality experiments, comparing the BGEI to domain-specific features. A similar experiment was performed with the GEI using the CASIA dataset B.

To perform the recognition in our experiments, we utilise sparse representation based classification (SRC) similar to [12]. It is required that the feature space must be sufficiently large for good SRC performance in the generalised feature domain. However, we show that by combining the generalised optimiser from SRC and a locally discriminant optimiser using MDA, the recognition results can be further improved with reduced computational complexity due to the significant reduction in the size of the feature space.

The remainder of this paper is organised as follows. Section II outlines the feature extraction methods used in this paper, including the proposed backfilled gait energy image (BGEI). Section III details the feature modeling and classification while Section IV presents the depth gait database that we have developed and released for public access. Experiments and results are shown in Section V, followed by the conclusion in Section VII.

II. GAIT ENERGY BASED FEATURES

To achieve a cross-capture-modality gait recognition system, we consider the gait features that provide best performance in each specific domain. In this paper, we focus on the popular and high performing algorithms in side-view and frontal-view gait recognition.

Motivated by motion history images and motion energy images used in action recognition, Han and Bhanu [5] proposed a simple and effective gait feature called the gait energy image (GEI). GEI based approaches are simple, fast and perform comparatively well on a side-view. GEI represents the static and dynamic behaviour of human motion within a gait cycle in a single image template by averaging the normalised binary silhouettes over that cycle. The GEI of the $k^{th}$ gait cycle is computed by averaging the silhouettes ($I_t$) corresponding to the frames in the $k^{th}$ gait cycle as follows,

$$GEI(x, y) = \frac{1}{n} \sum_{t=1}^{n} I_t(x, y),$$

where $n$ is the number of frames within a gait cycle. Figure 2(a) shows example silhouettes from a gait cycle and the computed GEI.

Traditional 2D frontal images are poorly suited for gait recognition due to the inability to capture dynamic details of gait [8]. However, depth images, either from stereo cameras or other depth sensing devices, can be used to alleviate this.

The gait energy volume [10] (GEV) was developed to exploit the robustness of gait energy features in the 3D domain using these depth images. It is an volumetric extension of the GEI, where binary voxel volumes are used as an analogue to the binary silhouette. Both full body volumes and frontal surface reconstructions have been used. In the context of frontal-view gait recognition, constructed frontal surface volumes are used to generate frontal GEV. A frontal (or possibly even back) perspective is ideal as it does not suffer from occlusions between the legs, and in theory, it should contain all the relevant dynamic gait information as the hidden surface should only contain relative structure information (i.e. thickness of limbs and torso). Constructed frontal voxels volumes and computed frontal GEV are shown in Figure 3(a).
The GEI and GEV use two different capturing systems and extract features that are heavily dependent on those capture-modalities. As their representations are incompatible, there is a requirement to maintain a separate recognition system for each feature.

In this paper, we propose a new feature, the Backfilled Gait Energy Image (BGEI), that captures the essential but common gait information from the above two models and enables cross-modality comparisons where the user can enroll in side-view and be recognised in frontal-depth or vice-versa.

A. Backfilled Gait Energy Images

The backfilled gait energy image (BGEI) operates on a similar premise to the frontal GEV, where the frontal surface of a model should contain all the relevant gait information. It takes only the frontal contour of the silhouettes and assumes that it contains sufficient information to perform gait recognition. By doing so, there is possibility of losing some of the gait information as the back leg is no longer represented by this feature. This information, however, could potentially be unnecessary for, or at least, contribute minimally to the system’s ability to discriminate between different people.

Since the frontal surface is available in both the side-view and frontal-depth, by applying the above concept, the BGEI becomes as a common feature for both.

For the side-view silhouettes, the BGEI is constructed by first back filling the binary silhouettes. For this, the frontmost pixel on each row is found and from it, filled to the back of the image. This backfilled binary silhouettes are aligned based on the centroid of the frontal surface. The BGEI is then constructed from these silhouettes in the same manner as a GEI, by averaging within a gait cycle. Figure 2(b) shows example images of backfilled silhouettes and and computed BGEI constructed from side-view images.

To create a BGEI from frontal depth images, first frontal binary voxel volumes are constructed as outlined in Section II. These frontal volumes are projected into the sagittal plane to produce the backfilled binary silhouettes. These backfilled silhouettes are used to generate the BGEI as we do in the side-view. Example backfilled silhouettes and the computed BGEI using depth images are shown in Figure 3(b).

III. FEATURE MODELING AND CLASSIFICATION

As with many other appearance based recognition tasks in computer vision (e.g. face recognition), some form of Euclidean distance based nearest neighbour algorithm is commonly used for classification, typically after applying principal component analysis (PCA) [13], or some other dimensionality reduction technique. Recently, sparse representation based classification [14] has become popular and has been used as an effective classification method for face [12] and gait [15].

Since initial investigation on sparse representation results shows the improvement over traditional classification methods, we investigate the applicability of SRC based classification similar that used in face recognition [12], and propose using a better discriminated input feature space to improve performance.

Initially, the feature vector is formulated by wrapping the gait energy feature into a single column vector, $v$. To discard the unwanted parts in the image and to effectively use the major contributing gait energy features, PCA is applied to the wrapped feature, such that we minimise the error function,

$$J = \sum_{i=1}^{n} \left\| (U + \sum_{j=1}^{dp} a_{ij} e_j) - v_i \right\|^2,$$

where $dp$ is the reduced dimensionality; $U$ is the mean of the template features; and $e_1, e_2, ..., e_{dp}$ is a set of orthogonal unit vectors that minimise $J$ when it represents the eigenvectors of the largest eigen values. In our experiments, $dp$ has been chosen to achieve a re-projection error of less than 1%. This generates the transformation matrix $T_{pca}$, that is applied to the input vector to reduce dimensionality as follows,

$$V = T_{pca} \times v.$$  

To initialise the classification problem, we construct a dictionary that represents all the subjects in the gallery with a total of $N$ extracted feature vectors ($V_1, V_2, ..., V_N$) packaged into the columns of a matrix, $A \in \mathbb{R}^{dp \times N}$. The objective is to identify the test subject, $\gamma$, in terms of the dictionary matrix that satisfies the following linear equation,

$$\gamma = A\alpha,$$

where by determining the coefficient vector $\alpha = [0, ..., 0, \delta_{i1}, ..., \delta_{ik}, 0, ..., 0]$ (where $\delta$ are non zero elements and $k$ is a number of feature vectors in the $i^{th}$ subject), we can determine that $\gamma$ corresponds to the $i^{th}$ class. The above objective can be achieved by computing the most sparse solution to the following optimisation problem,

$$\arg \min \| \alpha \|_0, \text{ s.t. } \gamma = A\alpha.$$  

The above $l_0$ norm optimisation problem can be solved in polynomial time when the system is overdetermined ($dp > N$). However, the dimension of the feature space is limited or maintained as low as possible for computational efficiency and the number of subjects enrolled in the system is comparatively high in a typical classification problem, particularly in
human recognition. Therefore, Equation 4 is generally under-determined, and solving the optimisation problem using the $l_0$ norm is NP-hard.

However, compressive sensing states that we can find the sparsest solution for the under-determined system by using the $l_1$ norm minimisation, defined as

$$\hat{\alpha} = \arg \min \| \alpha \|_1, \text{ s.t. } \gamma = A\alpha. \quad (6)$$

A. Improved SRC by locally discriminating the input feature space

The objective of SRC is to represent the test subject with a sparse combination of the learned dictionary subjects. However, it is impossible to achieve the exact representation of a test subject by a sparse superposition with only non-zero coefficients of the relevant subject because of appearance changes and other external varying factors (e.g., the carrying of goods). The assumption that SRC can work effectively regardless of the feature space [14] needs to be revised in this scenario. The $l_1$ norm optimisation as explained in Equation 6 tries to find the solution globally, and hence fails to identify the similarities and differentiating attributes within and between subjects by local analysis, even though the local subject labels are available.

To solve this issue, we apply multiple discriminant analysis (MDA) to analyse the class-labelled data for intra-class (within the subjects) similarities and inter-class (between the subjects) dissimilarities.

To extract the most discriminant features from the projected feature vectors, MDA is applied to learn the transformation matrix, $T_{mda}$, that maximises the ratio of the between-subject scatter matrix to the within-subject scatter matrix by computing the generalised eigenvectors that correspond to the largest eigen-values of the within-subject and between-subject scatter matrices. The dimension of the projected feature space has been chosen as one less than the number of subjects as explained in [5].

Each column vector in the dictionary and the feature vector of the test subject has been transformed using $T_{MDA}$ as follows,

$$\hat{V} = T_{mda}^T V, \quad (7)$$

where $\hat{V}$ is the final feature vector after PCA and MDA transforms have been applied. The above locally discriminated features form the new dictionary, $\hat{A}$, and the transformed feature vector of a test subject, $\hat{\gamma}$, and $l_1$ minimisation problem becomes,

$$\hat{\alpha} = \arg \min \| \alpha \|_1, \text{ s.t. } \hat{\gamma} = \hat{A}\alpha. \quad (8)$$

The dictionary, $\hat{A}$, now becomes more skewed since the within-subject variation is minimised and between-subject variation is maximised. The more the dictionary is skewed, the sparser the solution becomes. This can avoid the mis-representation of sparse signals when there is similar global effect on the raw feature vector, and it results in a more robust sparsifying solution for the $l_1$ norm minimisation.

B. Classification

Figure 4 shows the comparison of sparse solutions resulting from an example test subject (subject 25) on the labeled dictionary. Each bar shows the coefficients of the test subject on the dictionary. Using MDA, coefficients for the associated subject are significantly higher compared to the others and the coefficients are more sparse compared to PCA only approach.

IV. DEPTH GAIT DATABASE

There is only one depth gait database in the literature [10], and it only consists of frontal depth sequences for 15 subjects under a single walking condition. To facilitate future research in frontal gait recognition using depth, and to evaluate the proposed algorithms in this paper, the depth gait database (DGD) is proposed. The DGD consists of 37 subjects walking towards the camera under various walking scenarios. The main dataset consists of 35 subjects under 6 different walking conditions: normal walk (nw), fast walk (fw), back carrying (bc), side carrying (sc), front carrying (fc) and no shoes (ns). Multiple sequences are captured for each subject and walking scenarios.
V. EXPERIMENTS

Two sets of experiments are carried out in this paper to evaluate the BGEI on inter-capture modality and intra-capture modality platforms. Both the DGD as well as the CASIA dataset B [16] are used in these experiments.

To obtain our gait energy features, voxel volumes, as well as silhouettes need to be extracted from the databases. For the DGD, voxel volumes are constructed first by projecting the depth images into world coordinates. Segmentation planes are used to remove the background as shown in Figure 5(a) and a surface mesh of the subject is created in 3D. Holes in the data are interpolated, and the mesh is filled backwards to create the frontal binary volume. Gait cycles are identified based on detecting the oscillating pattern of the width profile in the volumes’ lateral view [10]. The GEVs and BGEIs for each gait cycle are then computed from these volumes as explained in Section II.

For the side-view sequences in the multi-modal subset, silhouettes are extracted from the depth images as opposed to the colour images. This is chosen as the extracted silhouettes are of higher accuracy, and the poor lighting and sensor quality in the colour camera makes clean segmentation from the RGB images difficult. Note that the depth information of these side-view sequences is only used to obtain the silhouettes and not used in the experiments themselves. Examples of this silhouette extraction are shown in Figure 5(b). Again, the gait cycles are detected based on the width profile, and the GEI and BGEI of the side-view sequences are computed according to Section II.

Only one gait cycle is extracted from each sequence in the DGD. For frontal depth sequences, the closest complete cycle is used in order to maximise the depth resolution (depth resolution in the Kinect sensor decreases with distance). For the side-view sequences, the central cycle in the image is used. This is to minimise any changes to the apparent size due to changes in distance as the subject moves across the camera’s field of view.

In the CASIA database, the dataset B, 90° (side-view) sequences are used. Background subtraction is used to extract the silhouettes from the video sequences. Graph cuts, similar to that used in [17] are used to improve segmentation quality. Once the silhouettes are obtained, gait cycle detection and gait energy feature construction is identical to the process used for the side-view DGD sequences. Two to three complete gait cycles exist in each side-view sequence, however, once again, only the centre-most cycle is used in the experiments.

All gait energy features are scaled to 96 pixel height. Only the lower half of the body is used to remove unwanted motion and appearance changes from the upper body. It also significantly reduces computational cost by decreasing the feature dimension. This results in a height of 48 and width of 84 in the feature image. The GEVs use the same dimension in sagittal plane, but with an additional depth of 60 voxels.

In all our experiments, the improved sparse representation based classifier, as mentioned in Section III-A, is used. The gallery cycles in each experiment form the training set for our classifier. These are used to learn the PCA-MDA basis ($T_{PCA}$ and $T_{MDA}$) and to formulate the dictionary matrix, $\hat{A}$ as explained in Section III-A. Each probe cycle is treated independently in our experiments. The distance scores for the various probe cycles are not combined with other cycles belonging to the same subject ID to perform the classification.

A. Experiments on inter-capture modality platform

For the first set of experiments, we will evaluate the key novelty in this paper; the use of the BGEI in a cross-capture modality platform. This is performed on the multi-modal segment of the DGD. The frontal depth sequences are used...
for each subject exists for the outlined in [18]. The intra-class experiment is performed on classes. The experiments on this dataset follow the evaluation probe. In inter-class tests, again 4 cycles from all the experiments is detailed in Table I.

As the gallery, while the side-view sequences are used as the probe, the BGEI can also be used in domain specific applications. Therefore, we also compare the BGEI to other gait features in their respective imaging domains to see how well this feature performs. First, we compare the BGEI to the GEV using the main set of the DGD, containing 35 subjects. An intra-class test is performed using the nw sequences. 3 of the 5 nw cycles for each subject is assigned as the gallery while the remaining is used as the probe. Interclass tests are also performed, with all 5 cycles in the nw forming the gallery, and all available cycles in each of the other classes (fw, sc, fc, bc, ns) forming the probe in their respective experiments.

A similar set of experiments is performed on the CASIA dataset B, in which the BGEI is compared to the GEI. The dataset contains 124 subjects with 3 different walking classes: normal walk (nw), bag (bg) and clothing (cl). 6 sequences for each subject exists for nw, while 2 for each of the other classes. The experiments on this dataset follow the evaluation outlined in [18]. The intra-class experiment is performed on the nw sequences, with 4 allocated to the gallery and 2 to the probe. In inter-class tests, again 4 cycles from nw are used as gallery, while the 2 sequences in each of the other classes make up the probe in their individual experiments. A summery of all the experiments is detailed in Table I.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Feature</th>
<th>Gallery</th>
<th>Cycles</th>
<th>Probe</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>BGEI</td>
<td>DGD side</td>
<td>8×5</td>
<td>DGD front</td>
<td>8×5</td>
</tr>
<tr>
<td>1b</td>
<td>BGEI</td>
<td>DGD front</td>
<td>2×5</td>
<td>DGD nw</td>
<td>35×5</td>
</tr>
<tr>
<td>2a (nw-nw)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×3</td>
<td>DGD nw</td>
<td>35×2</td>
</tr>
<tr>
<td>2a (nw-fw)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×5</td>
<td>DGD fw</td>
<td>35×5</td>
</tr>
<tr>
<td>2a (nw-sc)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×5</td>
<td>DGD sc</td>
<td>35×2</td>
</tr>
<tr>
<td>2a (nw-fc)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×5</td>
<td>DGD fc</td>
<td>35×2</td>
</tr>
<tr>
<td>2a (nw-bc)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×5</td>
<td>DGD bc</td>
<td>35×2</td>
</tr>
<tr>
<td>2a (nw-ns)</td>
<td>GEV / BGEI</td>
<td>DGD nw</td>
<td>35×5</td>
<td>DGD ns</td>
<td>35×4</td>
</tr>
<tr>
<td>2b (nw-bg)</td>
<td>GEI / BGEI</td>
<td>CASIA nw</td>
<td>124×4</td>
<td>CASIA nw</td>
<td>124×2</td>
</tr>
<tr>
<td>2b (nw-cl)</td>
<td>GEI / BGEI</td>
<td>CASIA nw</td>
<td>124×4</td>
<td>CASIA cl</td>
<td>124×2</td>
</tr>
<tr>
<td>2b (nw-bc)</td>
<td>GEI / BGEI</td>
<td>CASIA nw</td>
<td>124×4</td>
<td>CASIA bg</td>
<td>124×2</td>
</tr>
</tbody>
</table>

Table I: Experiments

A rank-1 accuracy of 100% is achieved in the first cross-capture modality experiment. The ROC in Figure 7, however, shows some mis-verification at low false positive rates. This initial experiment shows great potential in the proposed system, though it is limited to a small dataset of 8 subjects. The accuracy is not likely to hold in the expanded experiment, as the larger gallery size makes identification and verification more difficult. The rank-1 accuracy drops to 88.5% in Exp 1b, though 100% accuracy is still able to be achieved at rank 2. Overall, these results are promising, and demonstrate that performing appearance-based gait recognition using the BGEI across frontal-depth and side-view images is possible.

In the intra-capture modality experiments, we see an overall drop in performance of the BGEI compared to the GEI and GEV (Figures 8 and 9). While good results can still be obtained, a fairly notable drop can be seen in the experiments against the GEV. This can be attributed to the significant loss of information in the transition from a 3D representation to a 2D one; the separate motions of the left and right legs are no longer retained, and the entirety of the back legs in the gait cycle is lost in order to construct the BGEI. Still, the lowering of performance does not exceed 5% at a false alarm rate.
Fig. 8. ROC curves for tests in Experiment 2 (DGD). (—) BGEI features. (—-–) GEV features.

Fig. 9. ROC curves for tests in Experiment 3 (CASIA). (—) BGEI features. (—-–) GEI features.
Finally, the benefits of applying SRC to a MDA-based discriminated input space has been shown through a comparison with the traditional PCA-kNN and SRC-PCA based classifiers. The significant improvement of the proposed method over the recent approaches shows that discriminating the input space based on the local class-labels and applying an SRC based classifier is the optimum solution for future classification tasks.

**ACKNOWLEDGMENTS**

Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences. Computational recourses and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

**REFERENCES**


<table>
<thead>
<tr>
<th></th>
<th>nw-nw</th>
<th>nw-cl</th>
<th>nw-bg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank-1</td>
<td>TP @  FAR 3%</td>
<td>Rank-1</td>
<td>TP @  FAR 3%</td>
</tr>
<tr>
<td>SRC-MDA on GEI</td>
<td>100.0</td>
<td>99.57</td>
<td>80.3</td>
</tr>
<tr>
<td>SRC-MDA on BGEI</td>
<td>95.0</td>
<td>94.42</td>
<td>79.5</td>
</tr>
<tr>
<td>SRC on GEI</td>
<td>97.5</td>
<td>99.14</td>
<td>72.1</td>
</tr>
<tr>
<td>LDA on SEIS [20]</td>
<td>99.0</td>
<td>–</td>
<td>64.0</td>
</tr>
<tr>
<td>NN-MDA on GEI [5]</td>
<td>99.0</td>
<td>–</td>
<td>60.0</td>
</tr>
<tr>
<td>CAS [18]</td>
<td>97.6</td>
<td>–</td>
<td>32.7</td>
</tr>
<tr>
<td>KPCA on GEI [19]</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**TABLE II**

**Recognition Performances (CMS at Rank-1 and ROC at False Alarm Rate of 3%) of the Proposed Algorithms and Other Reported Results in Literature on CASIA Database.**

rate of 10%, even though the GEV uses these domain specific 3D appearance information, indicating that the BGEI retains significant discriminating power.

The BGEI fares better against the GEI feature (Figure 9), likely due to the fact that less information is lost going from the GEI to the BGEI than from the GEV. Overall, these results on the CASIA dataset are in fact quite similar (the BGEI achieves a slightly lower accuracy in nw-nw and nw-cl tests, but higher in the nw-bg test), showing that the loss of the back contour and back leg does not severely impede its ability to discriminate between different people under those conditions. Some of these results may also be attributed to the use of the proposed classification method. Table II lists the rank-1 accuracy and the true positive rate at FAR of 3% of the experiments we obtained on the CASIA dataset. Listed also, are the results that have been reported in other papers on this dataset. We can see that the proposed classifier significantly improves upon the systems that share the common GEI feature, such as nearest neighbour with MDA [5] or KPCA [19] classifiers. We also see an improvement over modified gait energy features such as the SESI [20], though it just loses out to it in the nw-bg case.

**VII. CONCLUSION**

In this paper, we propose a novel approach for gait recognition that enables the use of multiple independent capture sources within a single gait recognition system through a feature, the BGEI, that can be synthesised from multiple input sources such as frontal-depth or side-view data. We show that the proposed BGEI has the potential to work in a cross-capture platform with the our initial results of a CMS of 100% at rank-1, albeit on a small database. Performance of the BGEI can be further improved by incorporating advanced spatio-temporal alignment and scaling between cross-capturing platform.

The construction of the BGEI discards information, most notably, the entirety of the back leg, that are retained in domain specific features such as the GEI and GEV. However, through an evaluation on the CASIA database, we have demonstrated that there is sufficient information in the frontal plane of motion to recognise subjects at comparable, and even sometimes greater accuracy compared to the traditional GEI.

<table>
<thead>
<tr>
<th></th>
<th>nw-nw</th>
<th>nw-cl</th>
<th>nw-bg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank-1</td>
<td>TP @  FAR 3%</td>
<td>Rank-1</td>
<td>TP @  FAR 3%</td>
</tr>
<tr>
<td>SRC-MDA on GEI</td>
<td>100.0</td>
<td>99.57</td>
<td>80.3</td>
</tr>
<tr>
<td>SRC-MDA on BGEI</td>
<td>95.0</td>
<td>94.42</td>
<td>79.5</td>
</tr>
<tr>
<td>SRC on GEI</td>
<td>97.5</td>
<td>99.14</td>
<td>72.1</td>
</tr>
<tr>
<td>LDA on SEIS [20]</td>
<td>99.0</td>
<td>–</td>
<td>64.0</td>
</tr>
<tr>
<td>NN-MDA on GEI [5]</td>
<td>99.0</td>
<td>–</td>
<td>60.0</td>
</tr>
<tr>
<td>CAS [18]</td>
<td>97.6</td>
<td>–</td>
<td>32.7</td>
</tr>
<tr>
<td>KPCA on GEI [19]</td>
<td>90.0</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>