

Investigating Gait Recognition in the Short-Wave Infrared (SWIR) Spectrum: Dataset and Challenges

Brian DeCann^a, Arun Ross^b, Jeremy Dawson^a

^a West Virginia University, Morgantown, West Virginia, USA

^b Michigan State University, East Lansing, Michigan, USA

ABSTRACT

In the biometrics community, challenge datasets are often released to determine the robustness of state-of-the-art algorithms to conditions that can confound recognition accuracy. In the context of automated human gait recognition, evaluation has predominantly been conducted on video data acquired in the active visible spectral band, although recent literature has explored recognition in the passive thermal band. The advent of sophisticated sensors has piqued interest in performing gait recognition in other spectral bands such as short-wave infrared (SWIR), due to their use in military-based tactical applications and the possibility of operating in nighttime environments. Further, in many operational scenarios, the environmental variables are not controlled, thereby posing several challenges to traditional recognition schemes. In this work, we discuss the possibility of performing gait recognition in the SWIR spectrum by first assembling a dataset, referred to as the WVU Outdoor SWIR Gait (WOSG) Dataset, and then evaluate the performance of three gait recognition algorithms on the dataset. The dataset consists of 155 subjects and represents gait information acquired under multiple walking paths in an uncontrolled, outdoor environment. Detailed experimental analysis suggests the benefits of distributing this new challenging dataset to the broader research community. In particular, the following observations were made: (a) the importance of SWIR imagery in acquiring data covertly for surveillance applications; (b) the difficulty in extracting human silhouettes in low-contrast SWIR imagery; (c) the impact of silhouette quality on overall recognition accuracy; (d) the possibility of matching gait sequences pertaining to different walking trajectories; and (e) the need for developing sophisticated gait recognition algorithms to handle data acquired in unconstrained environments.

Keywords: short wave infrared (SWIR), gait recognition, silhouette extraction, gait curves, gait energy image, frieze pattern, unconstrained biometrics

1. INTRODUCTION

Human recognition using gait has become an active area of study in both the biometrics and computer vision communities. Gait is defined as the pattern of locomotion in animals. Human gait, therefore, is the manner in which people walk. Human gait recognition is perceived as an attractive solution for identification-at-a-distance applications for a number of reasons. First and most importantly, human gait is believed to be unique to the individual. Psychological studies by Cutting and Kozlowski demonstrated that humans are capable of perceiving friends and gender based on gait.^{1,2} Second, gait can be recorded covertly as well as passively. That is, subjects are not required to explicitly interact with a sensor in order for the system to extract their gait patterns. Finally, features for human gait recognition (both human-based and machine-based) can be perceived in low resolutions.³ For automated systems, this enables recognition to be performed with less expensive camera systems at greater distances. Typically, methods for gait recognition fall under two categories: model-based or model-free.³

Model-based approaches use information based on known structure of individuals or through models of the human body. Biped models (models inferring locomotion via two limbs) are common, but vary in terms of computational complexity and information extracted. Examples of extracted features include spectra of thigh

Brian DeCann: E-mail: bdecann@mix.wvu.edu (Contact author)

Arun Ross: E-mail: rossarun@cse.msu.edu

Jeremy Dawson: Email: jeremy.dawson@mail.wvu.edu

inclination,⁴ stride and elevation parameters,^{5,6} and thigh rotation.⁷ The advantage of a model-based algorithm is that by modeling the structural information of the human body, distortions or occlusions induced by apparel and other objects are less likely to adversely affect recognition performance. However, model-based approaches are often computationally complex, which may exclude such algorithms from use in real-time applications.

Model free approaches, on the other hand, generally aim to extract features based on the dynamics of a moving shape. Examples of model-free approaches include spatiotemporal XYT gait signature derivation,⁸ spectra of key poses,⁹ procrustes shape analysis,^{10,11} frieze patterns,¹²⁻¹⁵ and the gait energy image (GEI)¹⁶ and its variants.¹⁷⁻¹⁹ In general, the advantage of a model-free approach is simplicity, as features are derived directly from the dynamics of a moving body.

It can be argued that the advancement of gait recognition algorithms has coincided with the release of challenging datasets characterized by a large number of subjects, unconstrained environment with respect to lighting and objects, variations in clothing and footwear, diverse viewpoints with respect to the camera, etc. For example, early gait datasets, such as the CMU Mobo Database²⁰ and Soton Databases,²¹ had a limited number of subjects and were collected in constrained settings (e.g., indoors, subjects walking on a treadmill). Despite these limitations, these initial datasets aided in demonstrating the feasibility of performing automated gait recognition. The UMD Human Identification at a Distance (HID),³ collected in 2001, was one of the first datasets to consider multiple viewpoints in an outdoor environment. The dataset included an added challenge of matching low resolution silhouettes. This dataset was later superseded by the USF Gait Challenge Dataset,²² which continues to be a benchmark for evaluating and reporting algorithm performance. Published in 2002, the dataset initially consisted of 74 individuals subject to 16 different collection conditions,²³ pertaining to viewpoint, walking surface and time. It has since been expanded to include 122 subjects and 12 specific experiments. The next major datasets were released by the Chinese Academy of Sciences (CASIA). Referred to as the CASIA B²⁴ and C Databases,²⁵ these datasets included a larger number of subjects (124 and 153, respectively) and exhibited diverse variations, such as viewpoint, clothing, cadence and carrying condition (i.e., with or without a backpack). In particular, the CASIA C database was the first large gait database to expand beyond the visible spectrum, using an infrared (thermal) camera to collect video sequences in a nighttime environment. These datasets contributed towards advancing the state-of-the-art in gait recognition allowing researchers to consider issues such as view invariance,^{24,26,27} object detection,¹¹ clothing,²⁸ time²⁹ and framerate.³⁰ A summary of these datasets is provided in Table 1.

In order to continue the advancement of automated gait recognition, it is essential for the next generation datasets to have data acquired from less constrained environments. Toward this end, we introduce the WVU Outdoor SWIR Gait (WOSG) Dataset, a new challenging dataset whose properties are very likely to occur in an operational setting. These properties include:

- Data collection occurs in an *active, outdoor* environment, wherein environmental factors such as cloud cover (that impacts illumination) and scene factors such as motion artifacts due to trees or additional persons (that impacts segmentation and tracking) exist.
- Multiple walking paths, resulting in video sequences representing different viewing angles.
- The spatial resolution of the human subjects is not the same in every video sequence.

Additionally, the WOSG Dataset is assembled using a sensor operating in the short-wave infrared (SWIR) spectrum (900nm-1,700nm), which in an operational setting may be more advantageous than visible light. For example, in low-light conditions, RGB imagery requires an *active* illumination source, which can be detected by the human eye. On the contrary, SWIR illumination is for the most part undetectable to the human eye. Thus, a system operating in the SWIR spectrum can operate covertly. Additionally, light emitted from the sun (and reflection from the moon) and stars can act as natural illumination sources, enabling both daytime and nighttime operation. Further, SWIR is tolerant to obscurants such as dense clouds, fog, and smoke. For these reasons, it is worthwhile to evaluate gait recognition in the SWIR domain. It should be noted that the current WOSG dataset does not have nighttime imagery. We anticipate future datasets to incorporate imagery from both daytime and nighttime environments.

Table 1. Examples of public datasets available for gait recognition research. The column “Covariates” indicates the types of intra-class variations present in the dataset.

Dataset	#Subjects	Environment	Spectrum	Covariates
CMU MoBo Dataset ²⁰	25	Static Indoor	RGB	Viewpoint, Pace, Objects
Georgia Tech Dataset ³¹	24	Static Indoor, Static Outdoor	RGB	Pace
UMD HID Database ³	25-55	Static Outdoor	RGB	Viewpoint
Soton Small Dataset ²¹	12	Static Indoor	RGB	Shoe, Clothing, Objects
Soton Large Dataset ²¹	115	Static Indoor Static Outdoor	RGB	Viewpoint, Time
USF HumanID Dataset ²²	122	Static Outdoor	RGB	Viewpoint, Shoe, Surface, Objects, Time
Osaka Treadmill (A) ³²	34	Static Indoor	RGB	Pace
Osaka Treadmill (B) ³²	68	Static Indoor	RGB	Clothing
Osaka Treadmill (C) ³²	200	Static Indoor	RGB	Viewpoint
Osaka Treadmill (D) ³²	185	Static Indoor	RGB	Gait Fluctuation
CASIA B Dataset ²⁴	124	Static Indoor	RGB	Viewpoint, Objects
CASIA C Dataset ²⁵	153	Static Outdoor	Thermal	Pace, Objects
WVU Outdoor SWIR Gait Dataset	155	Active Outdoor	SWIR (1550nm)	Viewpoint, Illumination

In the gait recognition literature, recognition capability in an active outdoor environment has not been adequately tested. In particular, silhouette segmentation is superficially treated and the performance of recognition algorithms on silhouettes extracted from a more operational setting is not well known. On the aspect of multi-directional trajectories, an operational gait recognition system will most certainly encounter individuals walking along arbitrary paths (rather than a simple unidirectional path that is perpendicular to the camera’s optical axis), although this issue has been receiving attention as of late.^{24,27} Finally, the problem of dealing with human silhouettes that vary in resolution across video sequences (or even frames) has not been adequately evaluated in the literature. In an operational setting, particularly with cameras capable of performing a zoom operation, feature extraction may have to be conducted on human objects having variable spatial resolutions. This challenge is evident in the WOSG Dataset as the observed field of view may not be the same for all sequences. In summary, the WOSG Dataset represents a challenging dataset for the biometrics and computer vision communities in that it embodies the following attributes: matching gait sequences across viewpoints, trajectories and distances; and silhouette extraction in low-resolution (e.g., $\approx 50\%$ reduction in silhouette bounding box area compared to the CASIA datasets), SWIR imagery.

To provide a sense of the challenges posed by the dataset, a baseline evaluation is performed using three different gait recognition algorithms from the literature. These algorithms are the gait energy image (GEI)¹⁶ method by Han and Bhanu, gait curve matching by DeCann and Ross¹¹ and frieze pattern matching, as proposed by Liu et al.¹² These algorithms were chosen as each denotes a sufficiently distinct approach for accomplishing the recognition task (i.e., each algorithm is unique in its own regard). The following sections provide details about the WOSG Dataset, as well as a brief description of the matching algorithms used for evaluation. Experimental analysis establishes the performance of current gait recognition algorithms on this new challenging dataset, which will enable researchers to further advance the-state-of-the-art.

2. WVU OUTDOOR SWIR GAIT DATASET

2.1 Hardware Description

The camera used to acquire this dataset was the Sensors Unlimited Goodrich SU640KTSX-1.7RT High Sensitivity InGaAs SWIR Camera (640x512 pixels) with a 50mm f/1.4 SWIR lens. The Goodrich camera was used to capture

video sequences of individuals walking at distances ranging from 20-50m.

2.2 Collection Protocol

Collection occurred between September and November, 2011. Data collection with the Goodrich camera was performed in an outdoor environment during daylight hours, supplying natural illumination via sunlight. Cloud cover (i.e, ambient illumination) varied between clear skies, partly cloudy and overcast. Video sequences were bandpass filtered to 1550nm ($\pm 50\text{nm}$ FWHM). Operational settings such as integration time were adjusted to achieve the best image quality based on daily environmental conditions (e.g., cloudy, sunny, etc.).

During collection, each subject completed one session and during which, was asked to walk in a continuous motion along eight predefined trajectories. The specified walking directions and approximate distances to the camera are denoted in Figure 1. The distances and walking paths were defined such that a minimum of three complete gait cycles could be completed in each direction. The length of each video sequence (in time) varied between 90 and 110 seconds, depending on the walking speed of each subject. Sample video frames from the dataset are shown in Figure 2. In total, data was collected from 155 individuals. In terms of gender, there were 93 males and 62 females. In terms of ethnicity, 46% of subjects identified themselves as Caucasian, 25% as Asian Indian, 13% as Asian, 8% as African, 5% as Middle Eastern, 2% as African American and 1% as Unknown. In terms of age, 71% of subjects were between 20-29 years old, 12% between 18-19, 8% between 30-39, 4% between 40-49, 3% between 60-69 and 2% between 50-59.

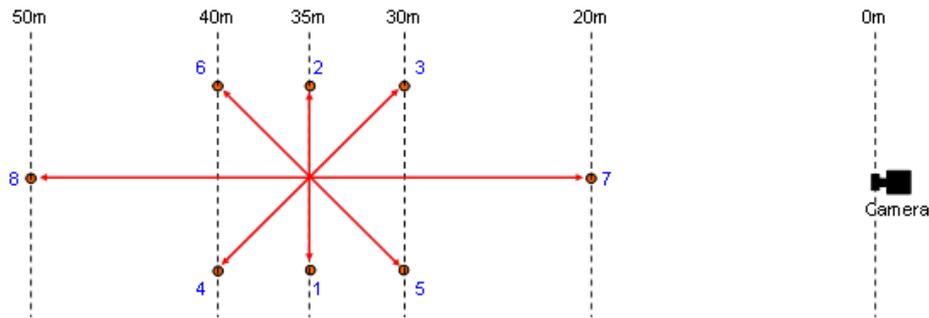


Figure 1. Gait collection layout. Each subject is captured walking in the numbered direction one time.

3. BASELINE ALGORITHMS

3.1 Segmentation and Silhouette Extraction

Algorithms for gait recognition typically begin with a silhouette extraction process where the pixels associated with the human object are detected and isolated from other background pixels. The simplest method for accomplishing this is through background subtraction,^{33,34} although more sophisticated methods can be used as well.³⁵ However, one of the associated challenges with SWIR imagery is that only one channel of information is available in each video. As a result, advanced segmentation methods that make use of multiple image channels may no longer be applicable. An example being the “Codebook Model” for background subtraction by Kim et al.³⁴ In addition, SWIR images are often of lower contrast than RGB, which increases the difficulty of properly identifying foreground pixels (i.e., pixels that denote a human body). To better convey these properties, Figure 3 depicts the intensity histograms from an image in the CASIA B Dataset (RGB) and the WOSG Dataset. Note that in the RGB image, regions depicting the person (intensities between 25-75) and background (intensities between 75-150) can be inferred. In the SWIR image, the range of intensities is smaller and no such distinction between person and background can be made.

For the purposes of this work, silhouette extraction is performed using a background extraction scheme. This was done since this technique has been predominately used in the gait recognition literature. In order to mitigate the challenge of low-contrast imagery (as can be seen in Figure 2) each frame of a video sequence is preprocessed by median filtering followed by contrast enhancement using contrast limited adaptive histogram equalization (CLAHE).³⁶ Following contrast enhancement, a difference image, I_{diff} , is created by subtracting



Figure 2. Sample frames from the WOSG Dataset. Note the variance in contrast and brightness in each frame, occurring as a result of varying environmental conditions.

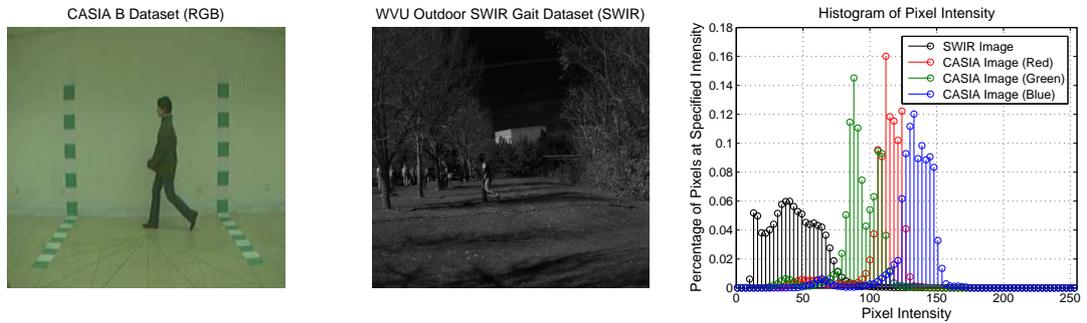


Figure 3. Sample images from the CASIA B (RGB) and WVU Outdoor Datasets and their associated intensity histograms. Note the dynamic range for the SWIR image is less than that of the RGB image.

enhanced image I_t from I_0 , the background image. Here I_0 was generated by averaging pixel regions in a video sequence with no individual present. Using a combination of threshold filters and morphological operations to eliminate noise, a binary image yielding the silhouette is created. The resulting binary image is denoted as S_t .

3.2 Gait Energy Image

The Gait Energy Image (GEI) is a model-free algorithm for human gait recognition proposed by Han and Bhanu.¹⁶ While not the most superior algorithm in terms of performance,¹⁹ the algorithm is conceptually and mathematically simple (although it can be resource intensive). For this reason, GEI is often cited as a benchmark for algorithmic comparison and has spawned many variants.^{17–19} The GEI algorithm attempts to reduce the motion dynamics of an individual represented in multiple frames into a single image. The algorithm computes $G(x, y)$, which is defined as the average of T space-normalized human silhouette images, $S_t(x, y)$, $t = 1, 2, \dots, T$. Commonly, T denotes the number of frames in one gait cycle. Mathematically, this is described in Equation (1).

$$G(x, y) = \frac{1}{T} \sum_{t=1}^T S_t(x, y) \tag{1}$$

By computing the “average” silhouette, the features used for matching correspond to the “moving” pixel intensities as a human silhouette moves in time. Since in the WOSG Dataset, the distance between the subject and the camera changes as a function of time, the height of each resolved silhouette is normalized to a pre-defined number of pixels. This is consistent with similar implementations of the GEI algorithm on other datasets where subjects move towards or away from the camera.²⁴ In addition, silhouette images must also be horizontally aligned to properly describe moving shape dynamics as pixel intensities in the GEI image.

Since raw GEI templates are in the form of images, the resulting feature vector can be very large, which can cause difficulty in matching. In the original work by Han and Bhanu¹⁶ and in many subsequent uses in

the literature, this issue is overcome by a combination of dimensionality reduction methods, notably Principal Component Analysis³⁷ and Linear Discriminant Analysis.³⁸

3.3 Gait Curve Matching

A second algorithm chosen for evaluating baseline performance is the “gait curves” algorithm by DeCann and Ross.¹¹ Inspired by previous work by Wang et al.,¹⁰ the gait curves algorithm is also a conceptually simple gait algorithm of the model-free variety. In comparison to the GEI algorithm, previous studies have shown it to be of comparable performance, while being more computationally efficient.¹¹ A brief description of the Gait Curves algorithm is presented below.

The “gait curve” of a subject is loosely defined as the evolution of the outer contour of a set of T silhouette images, $S_t(x, y)$, $t = 1, 2, \dots, T$, where the outer contour is defined as the maximum and minimum pixel coordinates, for which $S_t(x, y) = 1$, for each unit of y . That is, in a matrix coordinate system, define g_p^{left} and g_p^{right} as the leftmost and rightmost pixel locations for the p^{th} row ($p = 1, 2, \dots, y$). Subtraction of the mean horizontal position (x component) from these point sets results in a space normalized contour, denoted as the gait curve, G_t , for the t^{th} frame in a video sequence. Thus, the evolution of the gait curve across several frames can be regarded as a spatiotemporal feature for matching. To convert each of T gait curves into a single feature vector, the procrustes meanshape representation is adopted.^{39,40} The procrustes meanshape is a mathematically elegant measure of representing and evaluating shape sets. An advantage of this measure is that differences in translation, rotation, or scale do not negatively impact matching between a pair of shapes transformed by procrustes analysis. In applications of gait recognition, this is particularly advantageous as it is likely subjects are observed at varying distances from the camera.

Prior to obtaining the procrustes meanshape, the x- and y- coordinates of each gait curve are stored in a single vector by representing each point as a complex number (Equation (2)). Next, the mean of the gait curves is subtracted, resulting in zero-mean vectors (Equation (4)). Finally, the scatter matrix of the transformed vectors is computed (Equation (5)).

$$\mathbf{z}_i = Re(G_i) + jIm(G_i) \quad (2)$$

$$\bar{\mathbf{z}} = \sum_{i=1}^N \frac{\mathbf{z}_i}{i} \quad (3)$$

$$\mathbf{u}_i = \mathbf{z}_i - \bar{\mathbf{z}}, \quad \mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T] \quad (4)$$

$$S_u = \sum_{k=1}^T \frac{\mathbf{u}_k \mathbf{u}_k^T}{\mathbf{u}_k^T \mathbf{u}_k} \quad (5)$$

The first eigenvector of the scatter matrix, S_u , is used to denote the procrustes meanshape representation, $\bar{\mathbf{G}}$, of T gait curves extracted from a video sequence. In order to evaluate the distance (or similarity) between a pair of procrustes shape representations, the procrustes distance is used. This distance is defined in Equation (6). The range of values produced by Equation (6) is limited to $[0, 1]$, where, the smaller the value (distance), the more similar the shapes.

$$d(\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2) = 1 - \frac{|\bar{\mathbf{G}}_1^T \bar{\mathbf{G}}_2|^2}{\|\bar{\mathbf{G}}_1\|^2 \|\bar{\mathbf{G}}_2\|^2} \quad (6)$$

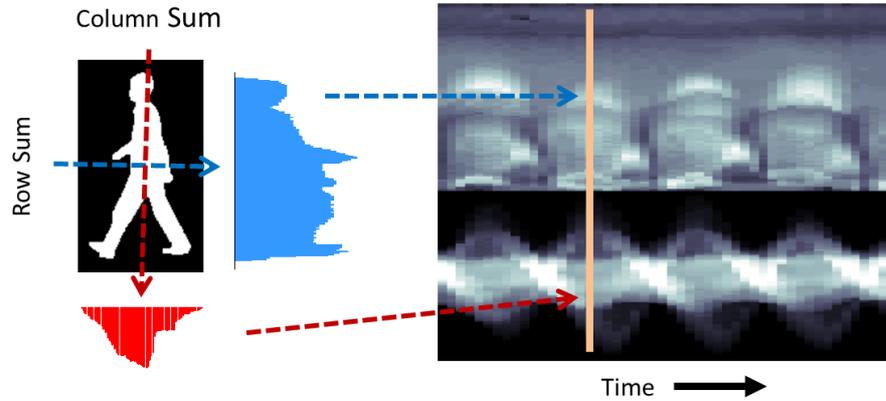


Figure 4. Visual example illustrating how the horizontal and vertical projections of the silhouette can be combined to create a spatiotemporal pattern for human gait recognition.

3.4 Frieze Pattern Matching

The third algorithm chosen for baseline evaluation is referred to as matching “frieze” patterns. In general, a frieze pattern is defined as a two-dimensional pattern that repeats itself infinitely. Frieze pattern matching was first exploited by Liu *et. al.*¹² in the context of gait recognition and is of the model-free class of gait recognition algorithms. A frieze pattern denoting human gait is defined as the concatenation of x- and y- projections of a silhouette moving in time. A mathematical description of such a pattern is presented below.

Consider a set of T silhouette images, denoted as $S_t(x, y)$, $t = 1, 2, \dots, T$. Define a 2-d frieze image, $F(y, t)$ as the horizontal projection (row sum) of each of T silhouette images. Mathematically, this is described in Equation (7).

$$F(y, t) = \sum_x S_t(x, y) \tag{7}$$

$F(y, t)$, denotes the width of the t^{th} silhouette at a “height” of y . Similarly, $F(x, t)$ can be defined as the vertical projection (column sum) of the silhouette, capturing the height of the t^{th} image at a given “width” x . In a set of T silhouette images encompassing at least one gait cycle, the silhouette width varies periodically due to factors such as stride and arm sway. By observing $F(y, t)$ as an image, the row projections combine to form a spatiotemporal pattern. A graphical example of how these patterns are constructed is illustrated in Figure 4.

In the original work by Liu *et al.*, matching of frieze patterns was accomplished by comparing the central moments of a probe and gallery sequence.¹² Alternatively, Dynamic Time Warping (DTW) or Hidden Markov Models (HMM’s) can be used to perform matching, as described by Kale *et al.*^{14,15} In this work, matching is performed using DTW, as described by Kale *et al.*¹⁴ In general, Dynamic Time Warping is a method for evaluating similarity between two sequences that vary in time. For instance, in the context of a frieze pattern for gait recognition, the periodicity varies depending on the speed at which a person is walking. Similarity between patterns is evaluated by generating cost matrix, \mathbf{C} , which stores the difference between each pattern for each unit of time (in this case, each frame). Thus, for a pair of patterns, A and B , with length T_A and T_B , $C(1, 1)$ denotes the difference between them at $t = 1$ and $C(1, T_B)$ denotes the difference between A at $t = 1$ and B at $t = T_B$. The resulting distance score is obtained by summing the smallest valued path from $C(1, 1)$ to $C(T_A, T_B)$ while moving strictly in increments of one in an 8-connected neighborhood.

4. EXPERIMENTAL RESULTS

4.1 Evaluation of Silhouette Quality

Prior to evaluating the matching accuracy of the algorithms described above, a good precursor is to evaluate the quality of the silhouettes produced in the segmentation process. Since gait sequences in the WOSG Dataset

were collected in an active outdoor environment (Figure 2), it is also useful to compare how well the typical segmentation method (i.e., background subtraction) performs on more challenging data. In order to perform this test, a metric for evaluating the quality of a silhouette is needed. Here, the metric defined by Liu et al.⁴¹ is adopted. In short, the metric attempts to measure the noise of the “foreground-sum signal”, which is defined as the summation of all foreground pixels of a silhouette over time. In theory, the foreground-sum signal (denoted by $\phi(t)$), is periodic according to the rate at which an individual performs one half of a gait cycle (Equation (8)).

$$\phi(t) = \sum_{\forall x} \sum_{\forall y} S_t(x, y) \tag{8}$$

Holes in the silhouette, or shadows can greatly impart noise to $\phi(t)$. Noise is measured by performing an eigenvalue decomposition of the autocorrelation matrix on a spatially normalized $\phi(t)$, referred as $f(t)$ (Equation (9)). The normalization parameters a_0 and a_2 in Equation (9) denote the DC component and amplitude of $\phi(t)$, respectively.

$$f(t) = \frac{\phi(t) - a_0(t)}{a_2(t)}, \text{ where } a_0 = E[\phi(t)], \quad a_2 = \frac{\sup\{\phi(t)\} - \inf\{\phi(t)\}}{2} \tag{9}$$

Spatial normalization is necessary to handle data (and datasets) collected at varying resolutions. The resulting quality metric is denoted as ψ and is obtained by summation of the first two eigenvalues (λ_1, λ_2), followed by subtraction of the 5th to M ($\lambda_5, \dots, \lambda_M$) eigenvalues, where M is the dimension of the autocorrelation matrix (Equation (10)). In general, the periodicity of a silhouette is captured by λ_1 and λ_2 , while $\lambda_5, \dots, \lambda_M$ denotes noise. Note that the 3rd and 4th eigenvalues are neglected. This is intentionally done as these components are likely to reflect both periodicity and noise.⁴¹ Thus, a large value of ψ is indicative of higher quality silhouettes. For a comprehensive description, the reader is referred to the original text by Liu et al.⁴¹

$$\psi = \sum_{i=1}^2 \lambda_i - \sum_{i=5}^M \lambda_i \tag{10}$$

Here, ψ is estimated from the silhouettes generated from the WOSG Dataset by utilizing Equations (8)-(10). Specifically, ψ is computed using (a) *all* of the frames in a sequence and (b) the mean ψ computed from a 50-frame moving window. That is, averaging the ψ value obtained from the frame sets: $\{1, 2, \dots, 51\}, \{2, 3, \dots, 52\}, \dots, \{T-50, T-49, \dots, T\}$. For comparison, estimates for ψ are also computed from the CASIA B and C datasets.^{24, 25} To mitigate the effect of outliers (i.e., exceptionally high quality sequences or exceptionally poor quality sequences), the median value of ψ is reported to indicate the global quality of each dataset. These results are tabulated in Table 2, where the ψ computed for each dataset is normalized by the maximum computed value. That is, the dataset with the highest quality corresponds to a value of 1.0. Note that the quality value produced for the WOSG Dataset is *less* than the CASIA B and C datasets, further suggesting its relevance to challenge problems.

Table 2. Median silhouette quality metric for the gait sequences in the WOSG Dataset, CASIA B dataset, and CASIA C dataset.

Dataset	Median ψ (all frames)	Median ψ (50 frame moving window)
WOSG Dataset	0.389	0.443
CASIA B Dataset	1.000	0.971
CASIA C Dataset	0.609	0.654

4.2 Protocol for Identification Analysis

The templates for the GEI, gait curve, and frieze pattern algorithms were constructed from approximately 125 frames, roughly equivalent to four seconds of a video sequence. GEI templates were extracted using a 75-pixel horizontal window, with the silhouette height normalized to 100 pixels. The templates were then downsampled

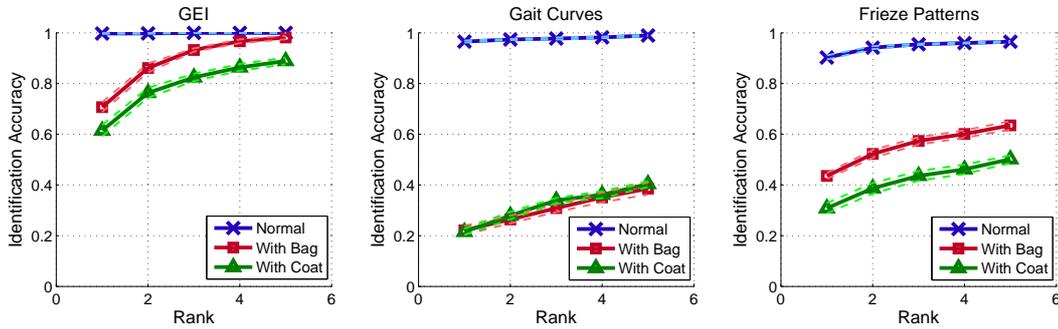


Figure 5. Baseline performance of the GEI, gait curves, and frieze pattern matching algorithms on the CASIA B dataset. Dashed lines indicate one standard deviation above or below the mean for ten trials.

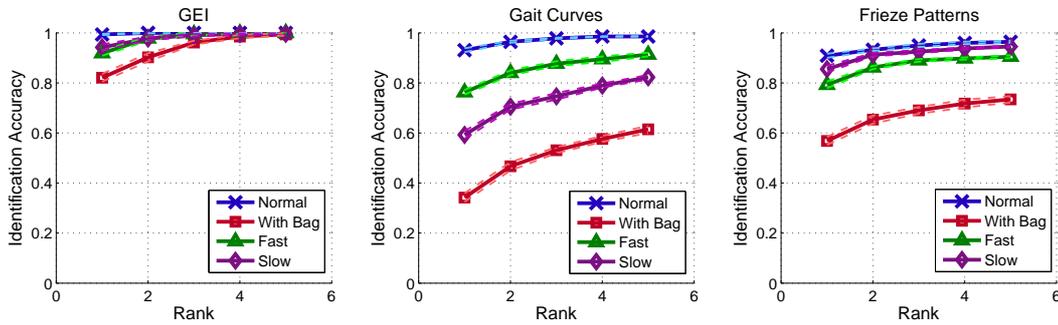


Figure 6. Baseline performance of the GEI, gait curves, and frieze pattern matching algorithms on the CASIA C dataset. Dashed lines indicate one standard deviation above or below the mean for ten trials.

to a size of 75x50. Subspace optimization was performed using principal component analysis, wherein a principal component was retained if the associated eigenvalue was greater than 0.001. Euclidean distance metric was then used to compare GEI templates. Templates for the gait curves algorithm were normalized to contain 300 elements and the procrustes distance metric (Section 3.3) was used for matching. Similar to GEI, templates for frieze patterns were extracted from a 75-pixel horizontal window and a normalized height of 100 pixels. Matching was performed using Dynamic Time Warping, as described by Kale et al.¹⁴ Since the GEI algorithm requires a training subset for computing the subspace projection matrices, templates for 15% of the total number of subjects were randomly selected for this purpose. Identification performance of the remaining subjects was then summarized using Cumulative Match Characteristic (CMC) curves. To remove selection bias, this process was repeated 10 times for each experiment with the reported CMC curves denoting the average of these 10 trials.

4.3 Baseline Evaluation

Here, a baseline evaluation is performed, in order to demonstrate performance on established gait datasets. Again, the CASIA B and C datasets are chosen for comparison. For both datasets, CMC curves are computed using the protocol established in Section 4.2. Rank-1 identification rates are computed for each covariate, using “normal” walking sequences as the gallery and the covariate (e.g., with bag, with coat, fast walk, slow walk) as the probe. When comparing “normal” to “normal” sequences, half the templates are selected as probe and gallery, respectively. In the interest of being concise, viewpoint (CASIA B), is not considered in the baseline evaluation. These results are tabulated in Figures 5-6 for the CASIA B and C datasets, respectively.

4.4 Identification Performance on the WVU Outdoor SWIR Gait Dataset

In order to perform a comprehensive evaluation of the GEI, gait curve and frieze pattern algorithms, a number of experiments were developed. Due to difficulties in completing segmentation, only 114 of the 155 subjects were used in the experimental analysis. In addition, segmentation was not possible in instances when subjects walked directly towards or away from the camera (labels 7 and 8 from Figure 1). To be consistent with related gait literature, unless otherwise stated, one feature vector was then extracted from each of the remaining six directions for each subject. This results in a total number of 684 gait sequences.

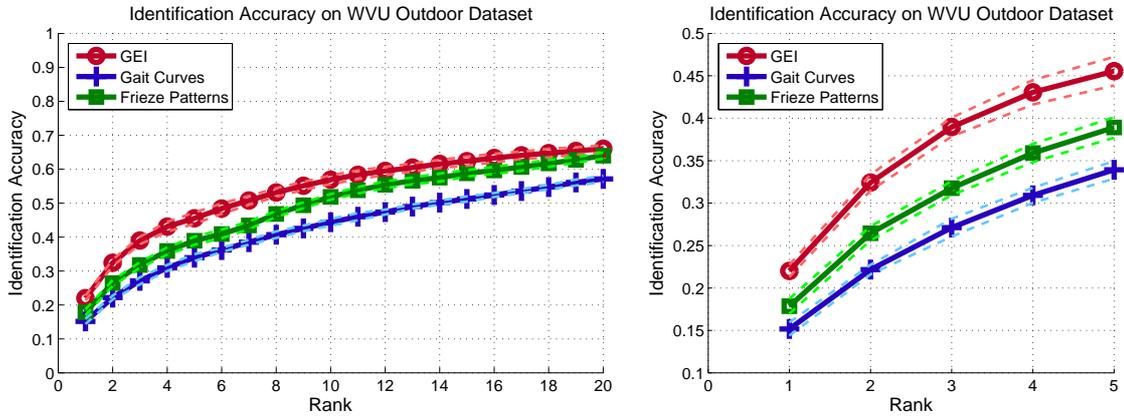


Figure 7. Identification performance of the GEI and Gait Curve algorithms on the WOSG Dataset, illustrating the combined matching performance across all walking directions using leave-one-out cross validation (LOOCV). Dashed lines indicate one standard deviation above or below the mean for ten trials. Left) Detail for ranks 1 to 25. Right) Detail for ranks 1 to 5.

In the first identification experiment, matching is performed using a leave-one-out cross validation scheme. That is, each of N_T gait sequences are compared against the remaining $N_T - 1$ sequences, regardless of walking direction. Establishing the gallery in this way illustrates the matching performance when the constraint of viewing angle is reduced, but not eliminated. These results are illustrated in Figure 7.

In the second experiment, the ability to match gait sequences corresponding to different viewing angles is evaluated. To avoid redundancies in experimental data, sequences of “leftward” walk are compared against sequences of “rightward” walk. Here, “leftward” walk is defined as those sequences wherein the subject is walking from the left to the right (from the camera’s perspective). This includes sequences with direction labels “1”, “4”, and “5” from Figure 1. Similarly, “rightward” walk is defined as those sequences wherein the subject is walking from the right to the left. This includes sequences with direction labels “2”, “3”, and “6” in Figure 1. In particular, the experiment is designed such that each of the three leftward sequences are matched against the three rightward sequences. This results in a total of nine probe and gallery combinations, which are defined in Table 3. These results are illustrated in Figure 8.

Table 3. Probe and gallery combinations for matching gait sequences corresponding to different viewing angles. The arrows denote the direction of walk (in the image plane).

Probe (Test data)	Gallery (Training data)
Direction 1 (←)	Direction 6 (↗)
Direction 1 (←)	Direction 2 (→)
Direction 1 (←)	Direction 3 (↘)
Direction 4 (↖)	Direction 6 (↗)
Direction 4 (↖)	Direction 2 (→)
Direction 4 (↖)	Direction 3 (↘)
Direction 5 (↙)	Direction 6 (↗)
Direction 5 (↙)	Direction 2 (→)
Direction 5 (↙)	Direction 3 (↘)

In the third experiment, the gallery and probe sequences correspond to the same direction. To accommodate this, two feature vectors are extracted per video sequence, where frames belonging to the first $\frac{T}{2}$ images are used to generate the first feature vector, while the remaining $\frac{T}{2}$ images are used to generate the second feature vector. Extraction of two feature vectors from a single video sequence is necessary since each subject in the database provided data in only one session. Since two templates for recognition are extracted sequentially, this experiment also acts as a measure of local feature variance. If the variance is low, the matching performance should be very

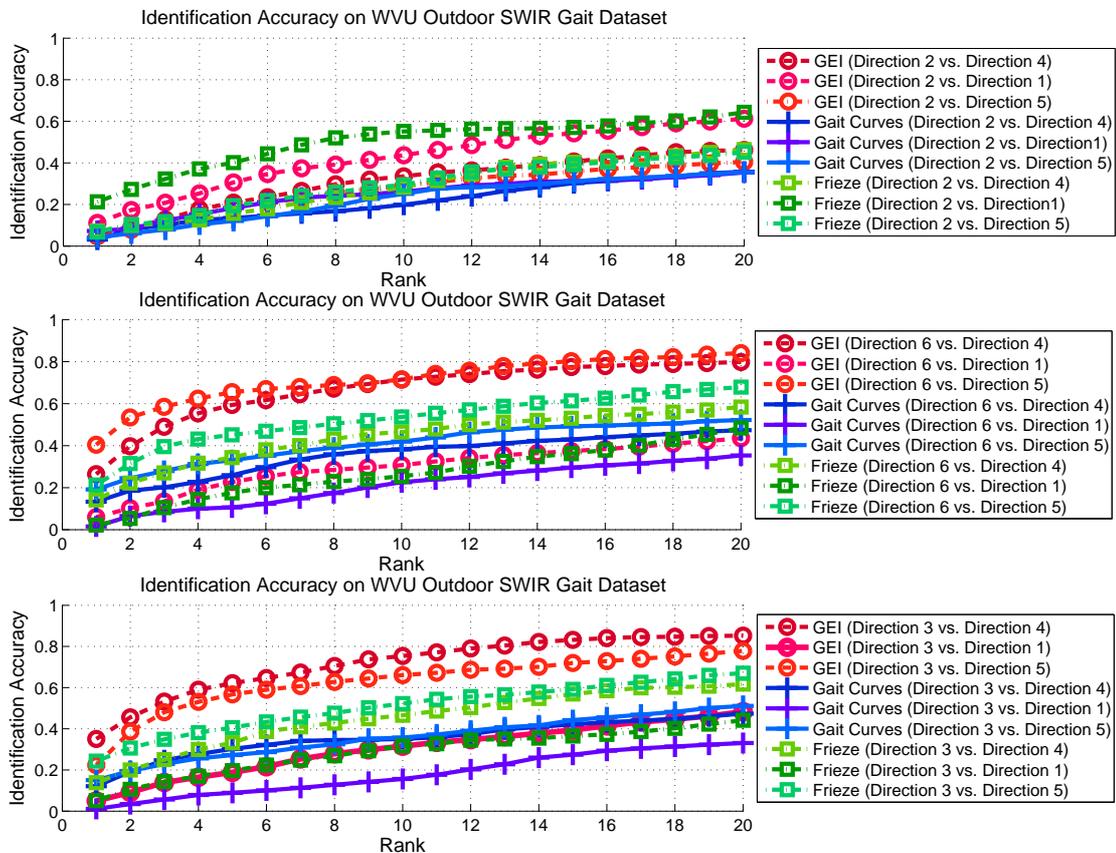


Figure 8. Identification performance when comparing different viewing angles.

high, as both feature vectors for each subject would be approximately equal to one another. If the variance is high, matching performance will be degraded. These results are illustrated in Figure 9.

5. DISCUSSION

The initial experiment regarding silhouette quality qualitatively demonstrated that the silhouettes extracted from the WOSG Dataset contained a larger degree of noise than the CASIA B and C datasets. Numerically, the measure of silhouette quality for the WVU Outdoor data was found to be $\approx 39\%$ and $\approx 64\%$ of the values for the CASIA B and C datasets, respectively. Thus, as potentially confounding environmental factors are introduced (e.g, illumination variance) in conjunction with a reduction in the number of channels in the image data (i.e., grayscale imagery), simple methods such as background subtraction become less adept at performing silhouette extraction. Figure 10 highlights a video sequence with extreme short-term illumination variance, which increases the difficulty of identifying foreground pixels. Additionally, Figure 11 compares silhouette images extracted from the WOSG Dataset, CASIA C Dataset and CASIA B Dataset. Note the gradual degradation of silhouette quality, as constraints such as a fixed background (CASIA B and C) are removed and the number of channels in the image data is reduced from 3 (CASIA B) to 1 (CASIA C and WOSG).

Regarding the basic identification performance of the GEI, gait curves and frieze pattern algorithms, the CMC curves in Figures 7-9 illustrate that the performance numbers of all three algorithms are significantly less than the numbers published for existing datasets^{11,14,16} and for the baseline evaluation (Figures 5-6). In the general matching scenario (Figure 7), the GEI algorithm outperforms both the gait curves and frieze pattern algorithms by about 5-10% between rank-1 and rank-25.

Expanding the analysis to compare matching sequences of differing walking directions yields similar, but interesting results. As with Figure 7, the initial observation is that the GEI algorithm generally outperforms

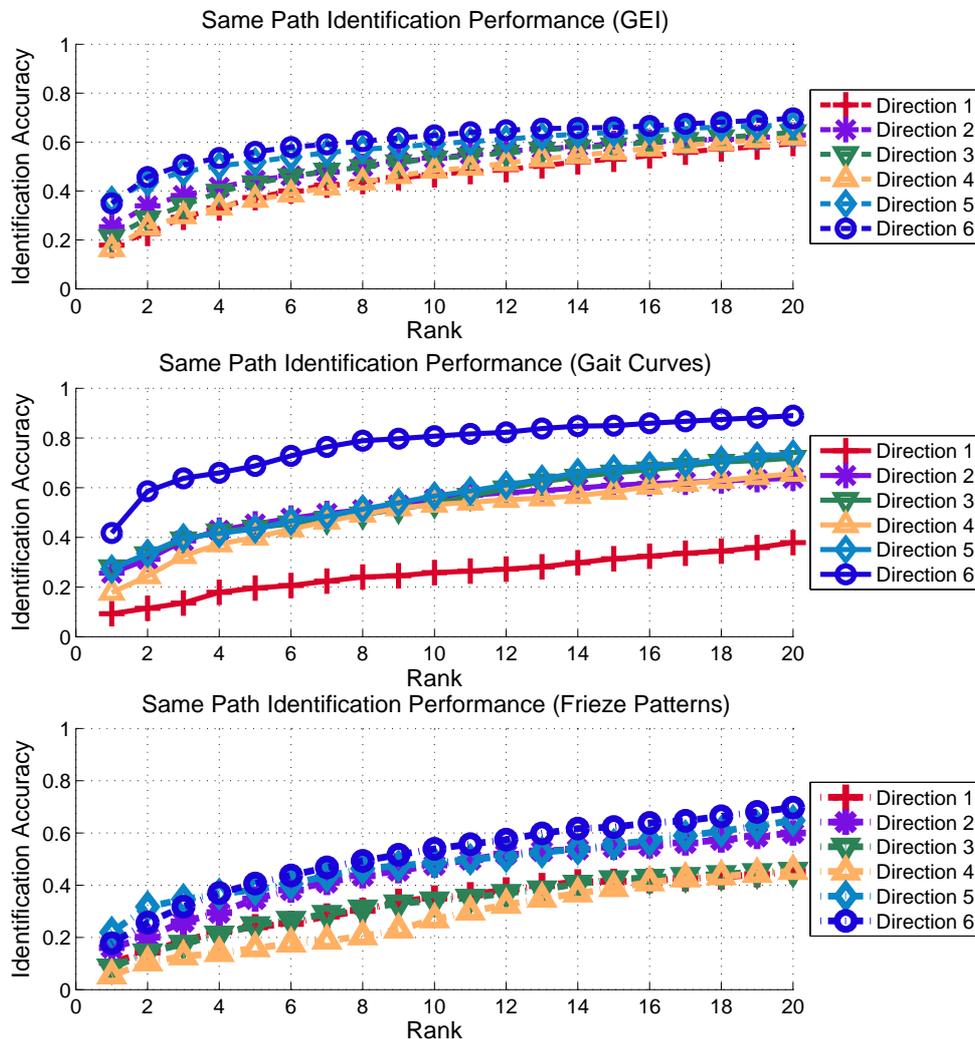


Figure 9. Identification performance when comparing templates extracted from the same walking direction Top) Gait Energy Image. Middle) Gait Curve Matching. Bottom) Frieze Pattern Matching.

both the gait curves and frieze pattern algorithms and in some cases, the difference is fairly significant. A closer look into the results in terms of direction indicates that each algorithm tended to yield reduced performance numbers when matching against sequences of horizontal walk (i.e., directions 1 and 2 from Table 3). This result is particularly interesting as it is commonly believed that gait recognition is optimally performed when the subject is viewed moving perpendicular to the field of view (i.e., moving horizontally in the image plane). In some instances, this reduction of performance is noticeably significant, as in the instances in Figure 8 where the gallery consists of directions 3 and 6.

In the third experiment, the ability of all three algorithms to match the same direction of walk is investigated. This was accomplished by extracting two feature vectors from one gait sequence, rather than using one feature vector per sequence. In addition to evaluating performance, this experiment indirectly serves as an additional measure of silhouette quality. In theory, a pair of feature vectors extracted along the same walking direction, seconds apart, should match very well to one another as the walking dynamics are (presumably) consistent in the short-term. Reductions in matching accuracy, therefore, reflect the lack of constancy in the shape of the silhouette in the short-term. Noting this, although each algorithm in Figure 9 shows an increase in rank-1 (and beyond) identification accuracy, the increase is not substantial and is direction dependent. Direction dependence



Figure 10. Example of a challenging video sequence. Here, changing cloud cover results in a significant change in background pixel intensity over a short period of time, resulting in difficulty in identifying foreground (silhouette) pixels.

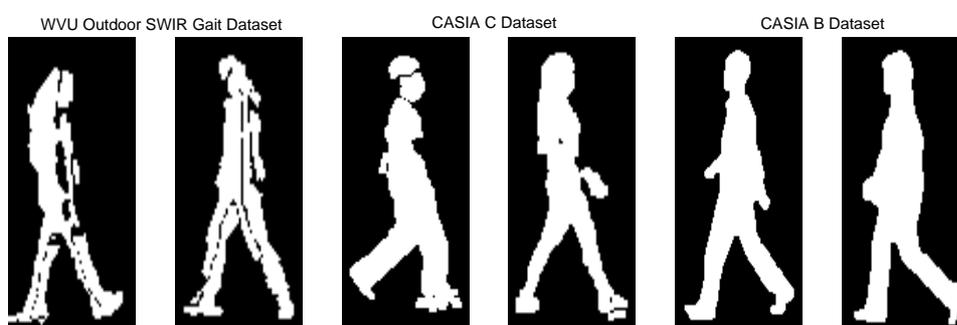


Figure 11. Comparison of silhouette quality using the procedure in Section 3.1. Note that the silhouettes produced in the WOSG Dataset are of lower quality.

is noteworthy, as a more uniform performance increase would suggest issues in the matching algorithms, rather than the data. The results across all algorithms indicate that, in general, higher quality silhouettes were extracted from direction 6 (bottom left to upper right), while the lowest quality silhouettes were extracted from direction 1 (right to left). Additional evidence of this can be found in Figure 8, where matching with direction 1 consistently yields low rank-accuracies.

In summary, of the three evaluated algorithms, *none* exhibited a high identification performance on the WOSG Dataset. However, based on the experimental results, an argument can be made that the likely source of error was the lower quality silhouettes generated from the segmentation process. Additionally, it should be pointed out that the segmentation process used for this work is consistent with related gait literature. This leads to one of two directions for future work. Firstly, there should be an emphasis on improving silhouette extraction in dynamic environments for low-contrast, single channel (grayscale) imagery. This should be in addition to developing methods that are robust to covariate changes as suggested by Liu and Sarkar.⁴² It is naive to presume that robust silhouette extraction is not a critical component towards the realization of an operational gait recognition system. As with any biometric modality, the matching process is certain to be less than ideal if the data passed to a feature extraction algorithm is noisy or otherwise corrupted. Further, due to the operational benefits of the SWIR spectrum compared to visible spectrum (passive illumination, nighttime operation, etc.), it is essential to develop segmentation methods explicitly for SWIR imagery, or methods capable of functioning in single channel video and images. Secondly, future work should also concentrate on developing algorithmic feature extraction and matching solutions for low-quality silhouette images. Certainly, in an operational scenario, segmentation may never be perfect and it is therefore essential to identify and develop solutions to such scenarios. Further, future work and datasets should consider silhouette resolution, a factor which persists in the WOSG Dataset.

6. CONCLUSION

In this study, an evaluation of three gait recognition algorithms is conducted on a novel, challenging dataset. The dataset, referred to as the WVU Outdoor SWIR Gait (WOSG) Dataset is a unique challenge dataset for human gait recognition with many properties that differ from existing datasets. Notably, the dataset was collected in an *unrestricted, outdoor* environment, in the short-wave infrared (SWIR) spectrum. Challenges in the dataset include segmentation in low-contrast, dynamic environments and matching gait sequences wherein subjects walk in differing directions, at various distances from the camera. By performing an evaluation using recent and well-known algorithms for gait recognition on this dataset, an understanding of the challenges that must be mitigated in order to advance the state of the art in gait recognition can be made. In particular, there should be increased emphasis on the design of algorithms for segmentation and recognition of individuals in the SWIR domain, as it represents an imaging modality with significant desirable operational benefits in contrast to RGB images. The results of the study indicate that matching gait sequences in less constrained environments is a difficult problem and should be of interest to researchers working in the biometrics and computer vision communities.

For information on procuring a copy of the WOSG dataset, please send an email to rossarun@cse.msu.edu.

7. ACKNOWLEDGMENTS

This work is sponsored by grant number N00014-08-1-0895 from the Office of Naval Research. Thanks to the WVU Night Biometrics group for useful discussions.

REFERENCES

- [1] L. Kozlowski and J. Cutting, "Recognizing the Sex of a Walker from a Dynamic Point Light Display," *Perception and Psychophysics* **21**, pp. 575–580, 1977.
- [2] J. Cutting and L. Kozlowski, "Recognizing Friends by Their Walk," *Bulletin of the Psychonomic Society* **9(5)**, pp. 353–356, 1977.
- [3] M. Nixon, T. Tan, and R. Chellappa, *Human Identification Based on Gait*, Springer, 2006.
- [4] D. Cunado, M. Nixon, and J. Carter, "Using Gait as a Biometric via Phase-Weighted Magnitude Spectra," *Proceedings of the First International Conference on Audio Visual Biometric Person Authentication* **1206**, pp. 95–102, 1997.
- [5] D. Wagg and M. Nixon, "On Automated Model-Based Extraction and Analysis of Gait," *Proceedings of the IEEE International Conference on Face and Gesture Recognition*, pp. 11–16, 2004.
- [6] R. Zhang, C. Vogler, and D. Metaxas, "Human Gait Recognition," *Proceedings of the IEEE International Workshop on Computer Vision and Pattern Recognition*, 2004.
- [7] C. Yam, M. Nixon, and J. Carter, "Automated Person Recognition by Walking and Running via Model-Based Approaches," *Pattern Recognition* **37(5)**, pp. 1057–1072, 2004.
- [8] S. Niyogi and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 469–474, 1994.
- [9] G. Zhao, R. Chen, G. Liu, and L. Hua, "Amplitude Spectrum-based Gait Recognition," *Proceedings of the IEEE International Conference on Face and Gesture Recognition*, pp. 23–28, 2004.
- [10] L. Wang, H. Ning, W. Hu, and T. Tan, "Gait Recognition Based on Procrustes Shape Analysis," *Gait Recognition Based on Procrustes Shape Analysis*, pp. 433–436, Proceedings of the IEEE International Conference on Image Processing.
- [11] B. DeCann and A. Ross, "Gait Curves for Human Identification, Backpack Detection, and Silhouette Correction in a Nighttime Environment," *SPIE Conference on Biometric Technology for Human Identification VII*, April 2010.
- [12] Y. Liu, R. Collins, and Y. Tsin, "Gait Sequence Analysis Using Frieze Patterns," in *Computer Vision - ECCV 2002*, **2351**, pp. 733–736, Springer Berlin / Heidelberg, 2002.
- [13] S. Lee, Y. Liu, and R. Collins, "Shape Variation-Based Frieze Pattern for Robust Gait Recognition," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

- [14] A. Kale, N. Cuntoor, B. Yegnanarayana, A. Rajagopalan, and R. Chellappa, "Gait Analysis for Human Identification," in *Audio- and Video-Based Biometric Person Authentication*, **2688**, pp. 706–714, Springer Berlin Heidelberg, 2003.
- [15] A. Kale, A. Sunderesan, A. Rajagopalan, N. Cuntoor, A. Roy-Chowdhury, V. Kruger, and R. Chellappa, "Identification of Humans Using Gait," *IEEE Transactions on Image Processing* **13**, pp. 1163–1173, 2004.
- [16] J. Han and B. Bhanu, "Human Activity Recognition in Thermal Infrared Imagery," *Proceedings of the IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 2005.
- [17] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient Night Gait Recognition Based on Template Matching," *Proceedings of the International Conference on Pattern Recognition*, 2006.
- [18] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang, "Gait Recognition Based on Dynamic Region Analysis," *Signal Processing* **88**(2), pp. 316–322, 2008.
- [19] E. Zhang, Y. Zhao, and W. Xiong, "Active Energy Image plus 2DLPP for Gait Recognition," *Signal Processing* **90**, pp. 2295–2302, July 2010.
- [20] R. Gross and J. Shi, "The CMU Motion of Body (MoBo) Database," Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, 2001.
- [21] J. Shutler, M. Grant, M. Nixon, and J. Carter, "On a Large Sequence-Based Human Gait Database," *4th International Conference on Recent Advances in Soft Computing*, pp. 66–71, 2002.
- [22] S. Sarkar, P. Phillips, Z. Liu, and I. Vega, "The HumanID Gait Challenge Problem: Datasets, Performance, and Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2**, pp. 162–177, 2005.
- [23] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "Baseline Results for the Challenge Problem of Human ID Using Gait Analysis," *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02)*, 2002.
- [24] S. Yu, D. Tan, and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," *Proc. 18th International Conference on Pattern Recognition (ICPR06)*, pp. 441–444, August 2006.
- [25] C. A. of Sciences, "CASIA Night Gait Dataset (Dataset C)." <http://www.cbsr.ia.ac.cn/english/Gait>
- [26] A. Kale, A. Roy-Chowdhury, and R. Chellappa, "Towards a View Invariant Gait Recognition Algorithm," *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 143–150, July 2003.
- [27] M. Goffredo, I. Bouchrika, J. Carter, and M. Nixon, "Self-Calibrating View-Invariant Gait Biometrics," *IEEE Transactions on Systems, Man and Cybernetics* **40**, pp. 997–1008, August 2010.
- [28] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi, "Clothing-invariant Gait Identification Using Part-based Clothing Categorization and Adaptive Weight Control," *Pattern Recognition* **43**, pp. 2281–2291, June 2010.
- [29] D. Matovski, M. Nixon, S. Mahmoodi, and J. Carter, "The Effect of Time on the Performance of Gait Biometrics," *IEEE Conference on Biometrics: Theory Applications and Systems (BTAS)*, pp. 1–6, September 2010.
- [30] N. Akae, A. Mansur, Y. Makihara, and Y. Yagi, "Video from Nearly Still: An Application to Low Frame-rate Gait Recognition," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1537–1543, 2012.
- [31] R. Tanawongsuwan and A. Bobick, "Modeling the Effects of Walking Speed on Appearance-Based Gait Recognition," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 783–790, 2004.
- [32] Y. Makihara, A. Tisuji, M. Hossain, K. Sugiura, A. Mori, and Y. Yagi, "The OU-ISIR Gait Database Comprising the Treadmill Dataset," *IPSJ Transactions on Computer Vision and Applications* **4**, pp. 53–62, 2012.
- [33] Y. Yang and M. Levine, "The Background Primal Sketch: An Approach for Tracking Moving Objects," *Machine Vision and Applications* **5**, pp. 17–34, 2002.
- [34] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Background Modeling and Subtraction by Codebook Construction," *IEEE International Conference on Image Processing*, 2004.
- [35] X. Chen, Z. He, J. Keller, D. Anderson, and M. Skubic, "Adaptive Silhouette Extraction in Dynamic Environments Using Fuzzy Logic," *IEEE International Conference on Fuzzy Systems*, pp. 236–243, 2006.

- [36] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. Romney, J. B. Zimmerman, and K. Zuiderveld, "Adaptive Histogram Equalization and its Variations," *Computer Vision, Graphics, and Image Processing* **39**(3), pp. 355–368, 1987.
- [37] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, pp. 1505–1518, 2003.
- [38] P. Huang, C. Harris, and M. Nixon, "Recognizing Humans by Gait via parametric Canonical Space," *Artificial Intelligence in Engineering* **13**(4), pp. 359–366, 1999.
- [39] J. Kent, *New Directions in Shape Analysis*, Wiley, 1992.
- [40] M. Stegmann and D. Gomez, "A Brief Introduction to Statistical Shape Analysis." March 2002.
- [41] J. Liu, N. Zheng, and L. Xiong, "Silhouette Quality Quantification for Gait Sequence Analysis and Recognition," *Signal Processing* **89**, pp. 1417–1427, July 2009.
- [42] Z. Liu and S. Sarkar, "Effect of Silhouette Quality on Hard Problems in Gait Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **35**, pp. 170–183, 2005.