Gait-assisted Person Re-identification in Wide Area Surveillance

Apurva Bedagkar-Gala and Shishir K. Shah

Quantitative Imaging Laboratory, University of Houston,
Department of Computer Science, Houston, TX 77204-3010, U.S.A.
avbedagk@central.uh.edu, sshah@central.uh.edu

Abstract. Gait has been shown to be an effective feature for person recognition and could be well suited for the problem of multi-frame person re-identification (Re-ID). However, person Re-ID poses very unique set of challenges, with changes in view angles and environments across cameras. Thus, the feature needs to be highly discriminative as well as robust to drastic variations to be effective for Re-ID. In this paper, we study the applicability of gait to person Re-ID when combined with color features. The combined features based Re-ID is tested for short period Re-ID on dataset we collected using 9 cameras and 40 IDs. Additionally, we also investigate the potential of gait features alone for Re-ID under real world surveillance conditions. This allows us to understand the potential of gait for long period Re-ID as well as under scenarios where color features cannot be leveraged. Both combined and gait-only features based Re-ID is tested on the publicly available, SAIVT SoftBio dataset. We select two popular gait features, namely Gait Energy Images (GEI) and Frame Difference Energy Images (FDEI) for Re-ID and propose a sparsified representation based gait recognition method.

1 Introduction

Person re-identification (Re-ID) is a fundamental task in multi-camera tracking and other surveillance applications. Re-ID is defined as a problem of person recognition across videos or images of the person taken from different cameras. Beyond surveillance, it has applications in robotics, multimedia, and more popular utilities like automated photo/video tagging or browsing [1]. Re-ID has been a topic of intense research in the past six years [2–8]. Most of the current approaches rely on appearance-based similarity between images to establish correspondences. The typical features used to quantify appearance are low-level color and texture extracted from clothing. A review of appearance descriptors for Re-ID is presented in [9]. However, such appearance features are only stable over short time intervals as people dress differently on different days. In case of large temporal separation or long-period Re-ID, more stable person descriptions based on unique features like biometrics are needed. Biometrics like face and gait have been shown to have tremendous utility in person recognition/identification applications [10, 11]. However, leveraging biometrics for Re-ID has its own challenges.

Re-ID data comes from uncontrolled environments with non-cooperative subjects where face of a person will not always be visible. Further, the data is often low quality due to low sensor resolutions and low frame rates. If the face is visible, it varies greatly in pose, facial expressions, and illumination conditions. All these factors make capturing reliable facial data and subsequent face recognition very difficult. Even though the state-of-the-art face recognition techniques yield high recognition rates, it is important to note that these results are obtained on high resolution data captured under controlled lighting and pose settings. Automated facial recognition on low resolution images under variations in pose, age and illumination conditions is still an open problem [12, 10].

(a) (b)

Fig. 1. Images from SAIVT SoftBio dataset and extracted silhouettes of the same person from different cameras, (a) images with visible face regions, (b) images without visible face regions.

Gait is a behavioral biometric that has been effective for human identification [11]. Gait is especially suited for Re-ID as it can be extracted by non-obtrusive methods and does not require co-operative subjects. Figure 1 shows images and corresponding silhouettes of the same person taken from different cameras. In figure 1 (a) the person's face is visible but not in figure 1 (b). This situation occurs frequently in surveillance videos as the camera angles are uncontrolled. Further, we can see that in sequence (a) the face resolution is really low and hence extracting usable facial features is challenging. However, the detected silhouettes can be used to extract gait features for Re-ID, even with low resolution and low frame rate data. The availability of video data makes gait feature extraction feasible as gait is extracted over multiple frames. On the other, gait is sensitive to view angles and walking poses. Surveillance cameras usually have a wide field-of-view (FOV) and people often tend to change their walking pose during the duration of observation. This greatly increases the probability of common walking views across different camera views, which can be leveraged for gait recognition. Nonetheless, silhouette extraction errors due to occlusions, illumination variations as shown in figure 2, can affect gait feature potency.

In this paper, we study the impact of incorporating gait features extracted from real world surveillance videos along with color (clothing-based appearance) features on Re-ID performance. If the number of frames available for a subject are not enough or silhouette extraction is faulty, then the gait features are not used

Fig. 2. Example images from SAIVT SoftBio dataset and erroneous extracted silhouettes.

and the Re-ID is only based on color features. The following are the contributions of our work:

- Investigate if gait extracted from real world video sequences can be successfully leveraged as an additional feature along with appearance features for person Re-ID in the context of surveillance.
- Identify robust gait features that can be extracted from noisy and incomplete silhouettes and yet retain discriminative capability for Re-ID.
- Propose a sparsified representation-based gait recognition method, where probe gait features are represented as a linear combination of gallery gait features and reconstruction residuals are used for recognition.

The proposed combined feature method is tested for Re-ID and potential of purely gait features is also tested to analyze its usability for long period Re-ID. We restrict the definition of long period Re-ID as Re-ID using videos captured on different days or where appearance features like clothing color break down. The paper is organized as follows: the next section gives a short overview of Re-ID techniques and gait features. Section 3 introduces the color and gait features and gait recognition methods adopted. Section 4 presents each of the datasets, experimental results and discussion. Section 5 concludes the paper.

2 Related Work

Person Re-ID has received much attention in the past few years [2–7]. In general, recent approaches have focused on two aspects of the problem: 1) design of discriminative, descriptive and robust visual descriptors to characterize a person's appearance [6, 13, 14]; and 2) learning suitable distance metrics that maximize the chance of a correct correspondence [15, 7, 5]. All of these methods are based on clothing-based appearance features, hence applicable to short period Re-ID. Gala *et al.* [16] have investigated the incorporation of low resolution facial features as potential descriptor but they are combined with clothing color, making them suitable only for short period Re-ID. There has been some work [17] that leveraged gait information for Re-ID. However, it combines gait features with motion phase and camera topology, which requires a training phase. Moreover, the datasets on which the method was demonstrated were fairly simplified videos.

Kawai *et al.* [18] propose a view dependent gait and color similarity fusion technique however it required explicit view information regarding gallery and probe poses, which is unrealistic in Re-ID. Color depth cameras are used to extract height temporal information which is combined with color features in [19], however this approach is not applicable with standard surveillance cameras.

2.1 Gait Features

Gait features are divided into two categories, *model-based* and *model-free* [20]. Model-based features like stride and cadence require an explicit model construction, hence, are more sensitive to the accuracy of silhouette extraction techniques and model fitting requires large computational cost. On the other hand, they are invariant to view angle changes and scale. Model-free features capture changes in silhouette shapes or body motion over time. This makes them partially robust to errors in silhouette extraction process but are more sensitive to variations due to pose and scale changes. Since silhouette extraction on surveillance video can be very challenging due to illumination variations, complex backgrounds, occlusions and unconstrained environments [21], model-free gait features are well-suited for Re-ID.

Han *et al.* [11] extract a gait feature called Gait Energy Image (GEI) that is robust to silhouette errors and computationally easy to extract. GEI captures the spatio-temporal description of person's walking pattern into a single image template by averaging silhouette over time, however it fails to retain the dynamic changes in the pattern. GEI represented using Gabor features [22] and wavelet features [23] extracted from different scales, orientations and x-y co-ordinates have been shown to be effective gait features. In order to incorporate the temporal information for gait recognition, Gait History Image (GHI) was proposed in [24]. It used difference between consecutive frames to retain frequency of motion within a GEI. All of these features are robust to silhouettes shapes but not as much to incomplete silhouette detections. The effect of incomplete silhouettes is partially alleviated by the Frame Difference Energy Image (FDEI) [25] which retains only positive portions from consecutive frame differences and combines them with GEIs to retain both static and dynamic portions of the person's walking pattern. In this paper, GEI and FDEI features are selected for gait representation.

2.2 Cross-view Gait Recognition

Gait recognition across different view angles is also an active area of research [26–28]. A view transformation model (VTM) is learnt in [29, 28] that learns a mapping between different view angles to transform data from probe and gallery sequences in the same view angle. A model that projects both the gallery and probe features into a subspace where they are highly correlated is learned to improve gait recognition in [27]. The model is learnt using canonical correlation analysis and combined with a Gaussian process classifier trained for view angle recognition for effective cross-view gait recognition. Gait dynamics from different views are synchronized into fixed length stances using non-linear similarity based HMMs in [26] and multi-linear projection model is learned to help gait recognition. All of these methods require some way to quantify the view angles from either gallery sequence, probe sequence or both, and involves a projection stage that introduces gait feature errors. Due to low quality data available for gait feature extraction such methods prove challenging to leverage for Re-ID. In order, to deal with view angle changes some methods [23, 30, 22] use locality preserving constraints are leveraged during gait recognition. We propose using sparsified representation based gait recognition technique that requires no explicit common view angle synthesis or projection subspace learning and we discuss it in detail in the next section.

3 Approach

Gait is described as the walking characteristic of a person and in order to understand the applicability of gait features to Re-ID they are combined with color features to create a person model for recognition. A person's walking pattern is periodic in nature, thus motion patterns during a walking sequence repeat themselves at regular time intervals. Thus, the features are extracted over a sequence of frames over a gait period. Gait period is defined as the time between two successive strikes of the same foot. Before gait period extraction is performed a silhouette sequence selection step is used to ensure that gait can be reliably extracted. To detect if silhouettes extracted are faulty or do not have enough information, a simple constraint is applied to number of non-zero pixels in the silhouette i.e. they should be more than 40% of the image size. For a sequence, if gait period cannot be detected, then gait features cannot be extracted. This serves as a natural silhouette sequence selection step. We adopt the method proposed in [31] for gait period estimation of arbitrary walking sequences. The aspect ratio of silhouette over a sequence of frames is used to estimate the gait period. Hence the assumption is that silhouettes are already extracted and preprocessed. Pre-processing involves size normalization and horizontally aligning the silhouettes. Figure 3 depicts the step by step results of the gait period estimation method, which is as follows:

- The aspect ratio of the silhouette over a sequence of frames is represented as a 1D temporal signal. The signal is z-normalized (subtracting the signal mean and dividing by signal's standard deviation) and smoothed using a moving average filter.
- Peaks in the aspect ratio signal are magnified by computing its auto-correlation sequence and the first derivative of the auto-correlation signal is used to detect zero-crossings.
- Zero crossing positions of positive and negative peaks are used to compute distance between prominent peaks, average of distances between consecutive peaks result in the gait period in number of frames.

Most of the dynamic information about walking motion can be deduced from the lower half of the silhouette region [11], hence gait period estimation is performed using aspect ratio of lower body silhouettes.

3.1 Gait Energy Image (GEI)

The first of the two gait features used is called the GEI first proposed in [11]. GEI is a spatio-temporal representation of a person's walking characteristic condensed

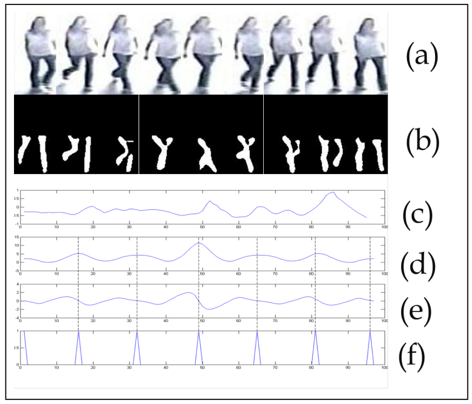


Fig. 3. Gait Period Estimation: (a) Sample image sequence. (b) Sample silhouette sequence. (c) Bounding box aspect ratio after average filtering. (d) Auto-correlation signal of aspect ratio signal. (d) First order derivative of auto-correlation. (f) Detected peak positions.

into a single image and it captures the changes in the shape of the silhouette over a sequence of images. GEI is computed by averaging the silhouettes in the spatial domain over a gait cycle. Given, a sequence of silhouettes, the GEI is given by,

$$GEI(i,j) = \frac{1}{N}\sum_{t=1}^{N} S_t(i,j) \tag{1}$$

where, S_t is the silhouette at frame t, (i,j) are the spatial image co-ordinates and N is the estimated gait cycle period. A given sequence of images can contain multiple gait cycles and hence can be represented by multiple GEIs. It has been shown that GEIs are robust to errors in silhouettes in individual frames [11] and are computationally efficient. Figure 4 shows GEI extracted from a sample sequence from both the SAIVT SoftBio dataset and the dataset we acquired, which we refer to as the multi-camera Re-ID (MCID) dataset. For the MCID dataset, we only used lower body silhouettes to extract gait features due to challenging conditions for silhouette extraction.

3.2 Frame Difference Energy Image (FDEI)

The second gait feature used is called Frame Difference Energy Image (FDEI) and was proposed in [25]. This feature is designed to deal with incomplete silhouettes at the same time retaining the shape and motion changes. This is done by constructing multiple thresholded GEIs from a single gait cycle and then summing these with differences between consecutive silhouettes along the sequence. The difference is taken in such a manner that only the positive values are retained and negative values are discarded so that only parts of the silhouette that are missing or contain motion information is retained. Having estimated the gait cycle, it is further divided into sub-cycles with equal number of frames.

(a) (b)

Fig. 4. GEI features extracted from sample sequences from the 2 datasets, (a) Lower body GEI from MCID dataset and (b) GEI from the SAIVT Soft-Bio dataset.

The silhouettes within each sub-cycle are averaged using Eq. 2 as follows:

$$GEI_c(i,j) = \frac{1}{N_c}\sum_{t=1}^{N_c} S_t(i,j), \tag{2}$$

where N_c denotes the period of the sub-cycle. This creates GEIs per sub-cycle of every gait cycle. These GEIs are thresholded, only retaining the pixels with value greater than or equal to 80% of the maximum value, in order to remove noise due to silhouette errors. These are referred to as dominant GEIs (DGEIs) for each sub-cycle and are computed using Eq. 3 as follows:

$$DGEI_c(i,j) = \begin{cases} GEI_c(i,j), & \text{if} \quad GEI_c(i,j) \geq (0.8xmax(GEI_c)) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The frame difference is computed by subtracting consecutive silhouettes and only the positive portion of the difference is included in the feature as follows:

$$DS_t(i,j) = \begin{cases} 0, & \text{if} \quad S_t(i,j) \geq S_{t-1}(i,j) \\ S_{t-1}(i,j) - S_t(i,j), & \text{otherwise} \end{cases} \tag{4}$$

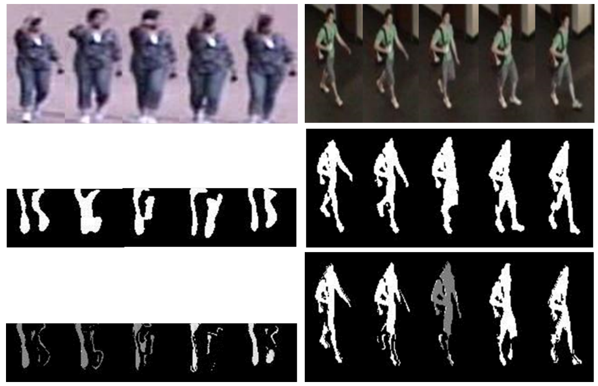
The FDEI is generated by summation of the positive frame difference and corresponding sub-cycles dominant GEI as follows,

$$FDEI(i,j) = DS_t(i,j) + DGEI_c(i,j) \tag{5}$$

If the silhouette in frame t is incomplete, and one at frame $t-1$ is complete, then DS_t contains the incomplete portion of the silhouette and summation of difference image with the dominant GEI compensates for missing portion. Figure 5 shows FDEI extracted from a sample sequence from both the datasets.

3.3 Feature matching for Re-ID

Color features The distribution of colors of the person's clothing is characterized using weighted HSV histogram proposed in [6]. The silhouette is divided into three body parts corresponding to head, torso and legs regions by detection of one vertical axis of symmetry and two horizontal axes of asymmetry. The histogram for each body part is weighted by the distance from the axis of symmetry. The histograms from each body part are concatenated channel-wise to generate a single color feature descriptor.



(a) (b)

Fig. 5. FDEI features extracted from sample sequences from the 2 datasets, (a) Lower body FDEI from MCID dataset and (b) FDEI from the SAIVT Soft-Bio dataset.

Combined similarity measure for Re-ID The similarity measure between a gallery subject G and a probe subject P is a weighted sum of color feature-based similarity and gait features based similarity and is defined as:

$$dist(G, P) = w_{color}.d_{color}(G, P) + w_{gait}.d_{gait}(G, P) \tag{6}$$

where, d_{color} is the color feature-based similarity and d_{gait} is the gait features based similarity. The color similarity is obtained using the Bhattacharyya distance [32]. Since, a sequence of frames are available for every subject, color similarity for a given probe-gallery pair is simply the maximum similarity among all the probe-gallery frame pairs. Gait similarity is computed as the re-construction error between probe gait features and the gallery gait model, and is explained in detail in the next section.

Further, w_{color} and w_{gait} are the weights given to color and gait similarity, respectively, such that: $w_{gait} = 1 - w_{color}$. If good quality silhouette sequences are available for both gallery and probe subject and gait features can reliably be extracted, gait similarity is incorporated in the overall similarity measure by setting w_{gait} to a non-zero value. On the other hand, in the absence of usable gait features, Re-ID is established using only color features, i.e., $w_{gait} = 0$ and $w_{color} = 1$.

In this paper, we define the acceptable quality of the silhouettes simply as to whether or not the number of positive pixels is greater than 40% of the total pixels in the image. Gait period estimation process provides another gait feature selection mechanism. If the silhouettes are too noisy or the sequences are too short, then gait period cannot be reliably estimated.

3.4 Gait Recognition by Sparsified Representation

In order to compute the gait similarity/recognition we utilize a sparsified representation [33]. Sparsified representation is used here as it has been shown to be robust to incomplete or missing data or occlusions in data. The underlying implication is that given a dictionary matrix built using labeled training features of several subjects, a test feature of i-th subject is the linear combination of the training images of only the same subject from the dictionary. Following this implication, we construct a dictionary which is simply a matrix $V = [v_1, ..., v_n]$, where each column is obtained by vectorizing the gait features belonging to all the gallery subjects. Since, each gallery subject can have multiple gait features, either GEIs or FDEIs, multiple columns in the matrix V may belong to the same gallery subject. We build a separate dictionary for each gallery subject, hence the dictionary of a given gallery subjects consists of different gait cycles from the same subject. This ensures that, given a dictionary matrix V, if the probe subject is as close as possible the gallery subject in both identity and view angle, the probe image will lie approximately in a linear span defined by only a subset of the gait features that constitute the matrix V. This implies that given a probe gait feature, for example, GEI_P, it can be expressed as $GEI_P = V.\alpha$ and the intent is to find the sparest α that generated GEI_P in V. Thus, among all possible solutions of α, we want the sparsest. This amounts to solving the following ℓ_1-minimization:

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \quad s.t. \quad GEI_P = V.\alpha \tag{7}$$

This optimization is solved using linear programming that leverages augmented Lagrange multipliers [34]. Thus, d_{gait} is given by Eq. 8 and is an estimate of how well $\hat{\alpha}$ reproduces GEI_P.

$$d_{gait}(G, P) = \|GEI_P - V\hat{\alpha}\|_2 \tag{8}$$

Further, for a given probe subject, there can be multiple GEIs/FDEIs and the gait similarity is simply the minimum among all the probe GEIs/FDEIs.

4 Results and Discussion

The combined gait and color features based Re-ID is tested on two datasets, our multi-camera Re-ID (MCID) dataset and SAIVT SoftBio dataset [35]. The SAIVT SoftBio dataset was selected as it provides real word multi-camera surveillance videos with multiple frames per person from different camera views.

4.1 SAIVT SoftBio dataset

Combined Color and Gait features model for Re-ID The dataset contains 150 subjects in up to 8 camera views. Only 122 subjects are seen in at least 2 camera views. In our experiments, the gallery is formed by different cameras and so are the probe subjects, so it is as close to the real world Re-ID scenario as possible. The only constraint is that the same subjects gallery and probe camera views are different. The dataset provides background images for each subject per camera view, hence simple background subtraction followed by low level image processing are sufficient to extract silhouette sequences for gait feature extraction. For each subject, in a given camera view, multiple frames are available. Only 5 randomly selected frames from a sequence are used to extract the color features, while all frames are used to extract gait features. Table 1 shows the Re-ID performance of only color, color & GEI (color+GEI) and color & FDEI (color+FDEI) in terms of rank 1, 5, and 20 Re-ID accuracy and normalized AUC extracted from cumulative matching characteristic (CMC) curves using all the 122 subjects. When gait features cannot be extracted, Re-ID relies only on color features. If gait is available, then a weighted sum of color and gait similarity is used to establish a match. The table shows the performance variations with different color and gait weight combinations. We can see that

	Rank 1(%)	Rank 5(%)	Rank 20(%)	nAUC(%)
Color Model	0.82	3.28	18.85	56.55
Color(0.8)+GEI(Sparse,0.2)	0.82	3.28	21.31	59.09
Color(0.8)+GEI(NN,0.2)	0.82	3.28	20.49	56.45
Color(0.5)+GEI(Sparse,0.5)	0.82	4.92	27.05	62.34
Color(0.5)+GEI(NN,0.5)	0.82	3.27	22.13	56.61
Color(0.1)+GEI(Sparse,0.9)	**4.10**	**9.84**	**36.06**	**64.75**
Color(0.1)+GEI(NN,0.9)	0.82	4.09	23.77	57.06
Color(0.8)+FDEI(Sparse,0.2)	0.82	3.28	21.31	58.98
Color(0.8)+FDEI(NN,0.2)	0.82	3.28	18.85	55.04
Color(0.5)+FDEI(Sparse,0.5)	0.82	5.74	27.05	62.24
Color(0.5)+FDEI(NN,0.5)	0.82	3.28	19.67	53.86
Color(0.1)+FDEI(Sparse,0.9)	**4.10**	**11.48**	**36.06**	**64.82**
Color(0.1)+FDEI(NN,0.9)	0.82	4.09	19.67	53.63

Table 1. Rank 1, 5 and 20 matching accuracy and nAUC measures for color, color+GEI and color+FDEI with varying weights.

overall using combination of color and gait performs better than only color. As we increase the weight assigned to gait features, we notice a significant boost in performance. The improvement in rank 1 matching accuracy is most notable with changes in weights suggesting that the gait features can be more effective than color features, when available.

For comparison, the gait similarity is also computed using nearest neighbor technique [36] which uses the euclidean distance as the underlying distance metric. The nearest neighbor gait similarity is combined with color similarity as described in the approach section. From the above table we can see that with any weight combination of color and sparsified representation gait similarity based Re-ID outperforms the color and gait nearest neighbor similarity. Figure 6 shows the CMC curves obtained by using only color, combined color and GEI, and combined color and FDEI. The gallery size = 122 subjects. Figure 6 (a), (b) and (c) show the CMC curves obtained by varying weights given to color and GEI/FDEI. We see that the difference in performance using GEI and FDEI is negligible. Thus, gait features even when extracted from imperfect silhouettes and varying viewpoints provide discriminative value. They add value to color features even in short period Re-ID where clothing color is a reasonable feature.

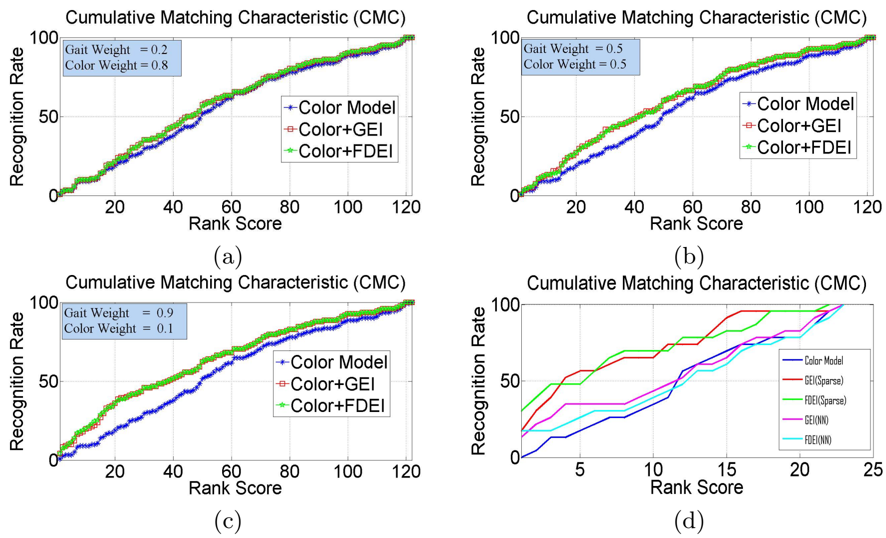


Fig. 6. CMC Curves on the SAIVT SoftBio dataset, (a), (b) and (c) show performance combining color and gait features with varying weights for Re-ID and (d) shows performance using only color, only GEI and only FDEI features for Re-ID (Gallery size = 23).

Only Gait features model for Re-ID In order to better understand the potential of gait features, we performed another experiment with the same dataset. Re-ID was performed using only gait features without the help of color and then compared with pure color based performance. Since only gait features are used the number of subjects for which usable gait features available across views were limited to 23 subjects. Figure 6 (d) shows the CMC curves obtained using color, GEI and FDEI features. From the figure, the power of gait features is even more prominent. Either of the gait features outperform the color features significantly. FDEI features yeild a much better Rank 1 matching accuracy when compared to GEI features, yet overall their performance is still comparable. Once again, the effectiveness of sparsified representation for gait recognition is evident. Re-ID performance using sparsified representation is clearly better than nearest neighbor gait recognition. Table 2 summarizes the rank 1, 5, and 20 matching accuracy and nAUC measures for the 3 features using both gait recognition techniques. From the table we can see that in terms of rank 1 and nAUC, FDEI features perform better than GEI. These results can be viewed as performance of gait features for long period Re-ID as they do not utilize any clothing or appearance features. This speaks to the ability of gait features for long period Re-ID and also demonstrates the robustness of the proposed gait recognition method for cross view angle gait matching. Figure 7 shows the first 5 ranked images from the gallery set given two different probes using the 3 models, we can see that in both cases the color only Re-ID cannot find an ID match to the probe in the top

	Rank 1(%)	Rank 5(%)	Rank 20(%)	nAUC(%)
Color Model	0	17.40	78.26	49.72
GEI (Sparse)	17.40	**56.52**	95.65	72.78
FDEI (Sparse)	**30.43**	47.83	**95.65**	**73.16**
GEI (NN)	13.04	34.78	82.6	55.95
FDEI (NN)	17.4	26.08	78.26	51.42

Table 2. Rank 1, 5 and 20 matching accuracy and nAUC measures for color, GEI and FDEI features.

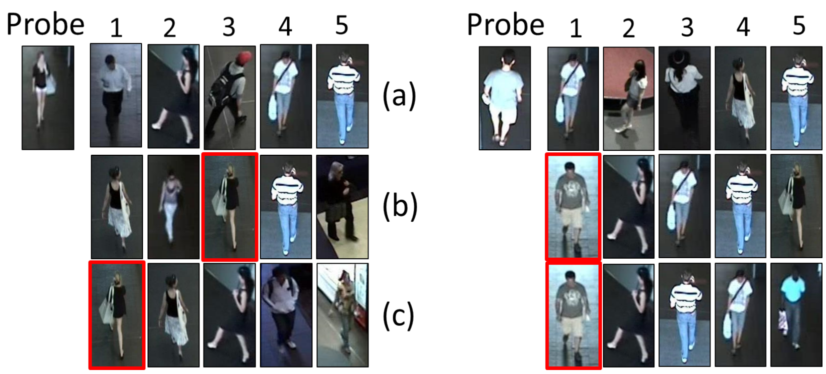


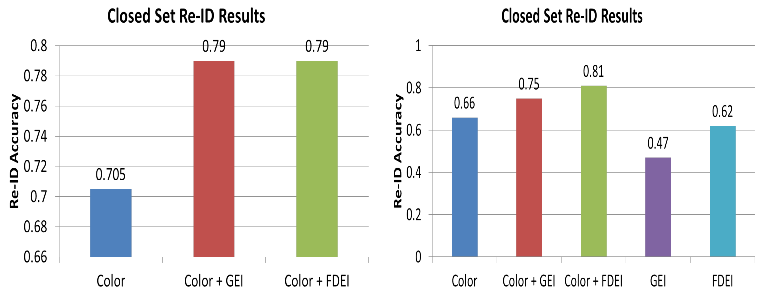
Fig. 7. Examples of Re-ID gallery ranking, showing the first 5 ranked on SAIVT Soft-Bio dataset using: (a) only color, (b) color+GEI and (c) color+FDEI, for two different IDs. The images highlighted by the red border denote the true match.

5 ranked IDs. In the probe ID shown in the left column, the color+GEI Re-ID located the true ID in the top 5 ranked IDs. On the other hand, the color+FDEI Re-ID rank 1 ID is the correct ID as the probe.

4.2 MCID dataset

The MCID dataset is acquired using a camera network consisting of 9 cameras in and around a building. The re-identification data consists of 40 subjects, 19 of them are seen in multiple cameras. As the re-identification is in the context of multi-camera tracking, Re-ID is established for each camera pair. Consistent with the previous dataset experiments, only 5 randomly selected frames from a sequence are used to extract the color features, while all frames are used to extract gait features. Due to complex and uncontrolled conditions of acquisition, traditional background modeling techniques did not perform well for silhouette extraction. Gray level thresholding combined with morphological operations were used to extract lower body silhouettes. Thus, both gait features were extracted only from lower body dynamics. For this dataset, we empirically set $w_{gait} = 0.2$ and $w_{color} = 0.8$ as this gives the best possible Re-ID accuracy.

Combined Color and Gait features model for Re-ID Since Re-ID is performed between camera pairs, all the subjects in the probe set might not be



(a) (b)

Fig. 8. Closed set Re-ID performance: (a)Bar graph shows the results obtained by our color, color+GEI and color+FDEI on the entire MCID dataset and (b) Bar graph shows the results obtained by our color, color+GEI, color+FDEI, only GEI and only FDEI on subset of MCID dataset.

included in the gallery set. Thus, for each camera pair, only a subset of the probe set or closed probe set, i.e., an intersection between the probe and gallery set IDs is used to establish re-identification. This is a closed set experiment, in the sense that it is only possible to have correct matches or mismatches and results are evaluated using matching accuracy, i.e., number of probe subjects matched correctly. Again, closed set experiment is consistent with our experiments on the SAIVT SoftBio dataset. Figure 8(a) shows the Re-ID performance using all 3 models: one based on only color, color+GEI and color+FDEI. The value of adding gait features to color is very evident, as we observe a significant improvement in the re-identification accuracy. One important observation here is that for the gait features of the MCID dataset are extracted from only lower body silhouettes or truncated GEIs. Even so the truncated gait features combined with color provide a significant boost.

Only Gait features model for Re-ID To better understand the role of gait features in Re-ID performance, we perform another experiment, using only gait features without the help of color and then compared with pure color based performance. Since only gait features are used, only subjects from each camera pair for which usable gait features are available for Re-ID. The results of this experiment is shown in figure 8(b). This figure provides balanced view of performance of all the features, alone and combined. Both purely GEI and FDEI perform reasonably well as compared to the color even though they do not outperform color. It should be noted that only lower body silhouettes are used to generate gait features. The Re-ID accuracy using FDEI comes within 4% of the color accuracy. Combining either GEI or FDEI boosts the Re-ID accuracy significantly. Incorporating FDEI with color helps enhance Re-ID performance to 81% from 66%, a considerable boost. Purely GEI or FDEI features performs better on SAIVT SoftBio dataset compared to MCID, as the MCID features are truncated gait features, which slightly affects the performance using pure gait features. However, we argue that combining gait features with color is an effective strategy to improve Re-ID performance. Figure 9 shows the first 5 ranked

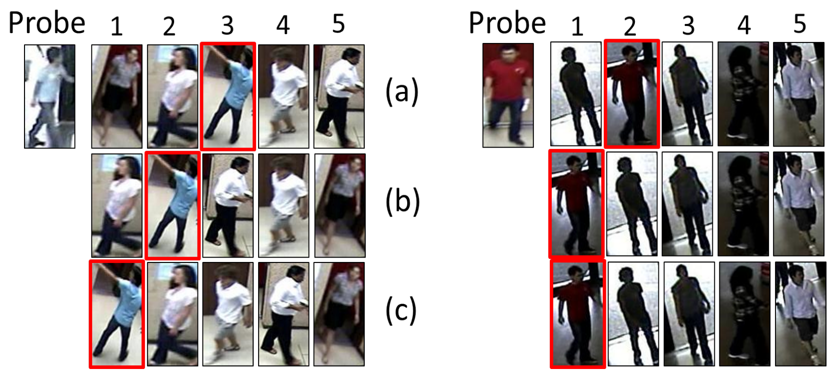


Fig. 9. Examples of Re-ID gallery ranking, showing the first 5 ranked on MCID dataset using: (a) only color, (b) color+GEI and (c) color+FDEI , for two different IDs. The images highlighted by the red border denote the true match.

images from the gallery set given two different probes using the 3 models, we can see that in both cases the color only Re-ID finds the true ID match to the probe in the top 5 ranked IDs at ranks 3 and 2 for the 2 IDs, respectively. In the probe ID shown in the left column, the color+GEI Re-ID located the true ID at rank 2. On the other hand, the color+FDEI Re-ID rank 1 ID is the correct ID as the probe in both probe IDs.

5 Conclusions

In this paper, we studied the applicability of gait features for Re-ID. The potential of gait features was explored in combination with color features as well as gait as the only feature for Re-ID. A sparsified representation based cross view gait recognition method is proposed that does not require view angle estimation or gait feature reconstruction for gait matching. We demonstrated its effectiveness for person Re-ID for short and long period Re-ID. We would like to point out that this was the first study of its kind offering a strategy for incorporating gait features for Re-ID and extensive analysis. The most important observation is that in spite of silhouette imperfections and view angle variations, gait can be effectively leveraged for Re-ID and is very well equipped for long period Re-ID. One of the future directions will be to explore different strategies for gait and color similarity measure fusion. Another area that needs more study is to estimate definitively the conditions under which gait features become unusable for Re-ID.

References

1. Sivic, J., Zitnick, C.L., Szeliski, R.: Finding people in repeated shots of the same scene. In: Proceedings of the British Machine Vision Conference. (2006) 909–918

2. Gheissari, N., Sebastian, T., Hartley, R.: Person reidentification using spatiotemporal appearance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2006)
3. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and appearance context modeling. In: International Conference on Computer Vision. (2007) 1–8
4. Bedagkar-Gala, A., Shah, S.: Multiple person re-identification using part based spatio-temporal color appearance model. In: The Eleventh International Workshop on Visual Surveillance (in conjunction with ICCV 2011). (2011)
5. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: Proceedings of the British Machine Vision Conference. (2010) 21.1–21.11
6. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. Computer Vision and Image Understanding **117** (2013) 130–144
7. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 653–668
8. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. Image and Vision Computing **32** (2014) 270–286
9. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. CoRR (2013)
10. Zhao, W., Chellappa, R., Phillips, J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys **35** (2003) 399–458
11. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 316–322
12. Ao, M., Yi, D., Lei, Z., Li, S.: Face recognition at a distance: System issues. In Tistarelli, M., Li, S., Chellappa, R., eds.: Handbook of Remote Biometrics. Springer London (2009) 155–167
13. Cheng, D., Cristani, M., Stoppa, M., Bazzani, L., Murinog, V.: Custom pictorial structures for re-identification. In: Proceedings of the British Machine Vision Conference. (2011) 68.1–68.11
14. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: IEEE International Conference on Advanced Video and Signal Based Surveillance. (2010) 435–440
15. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Asian Conference in Computer Vision. (2010) 501–512
16. Bedagkar-Gala, A., Shah, S.K.: Part-based spatio-temporal model for Multi-Person re-identification. Pattern Recognition Letters (2011)
17. Roy, A., Sural, S., Mukherjee, J.: A hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification. Pattern Recognition Letters **33** (2012) 1891–1901
18. Kawai, R., Makihara, Y., Chunsheng, H., Iwama, H., Yagi, Y.: Person re-identification using view-dependent score-level fusion of gait and color features. In: International Conference on Pattern Recognition (ICPR). (2012) 2694–2697
19. John, V., Englebienne, G., Krose, B.: Person re-identification using height-based gait in colour depth camera. In: IEEE International Conference on Image Processing (ICIP). (2013) 3345–3349
20. Wang, J., She, M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification. In: International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2010. (2010) 320–327

21. Bouwmans, T.: Recent advanced statistical background modeling for foreground detection: A systematic survey. Recent Patents on Computer Science **4** (2011)
22. Xu, D., Huang, Y., Zeng, Z., Xu, X.: Human gait recognition using patch distribution feature and locality-constrained group sparse representation. IEEE Transactions on Image Processing **21** (2012) 316–326
23. Hu, H.: Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis. Circuits and Systems for Video Technology, IEEE Transactions on **24** (2014) 617–630
24. Liu, J., Zheng, N.: Gait history image: A novel temporal template for gait recognition. In: IEEE International Conference on Multimedia and Expo, 2007. (2007) 663–666
25. Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J.: Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters **30** (2009) 977–984
26. Hu, M., Wang, Y., Zhang, Z., Zhang, Z.: Multi-view multi-stance gait identification. In: IEEE International Conference on Image Processing (ICIP), 2011. (2011) 541–544
27. Bashir, K., Xiang, T., Gong, S.: Cross view gait recognition using correlation strength. In: Procedings of the British Machine Vision Conference 2010. (2010)
28. Kusakunniran, W., Wu, Q., Li, H., Zhang, J.: Multiple views gait recognition using view transformation model based on optimized gait energy image. In: IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009. (2009)
29. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: European Conference om Computer Vision. (2006) 151–163
30. Tee, C., Goh, M., Teoh, A.: Gait recognition using sparse grassmannian locality preserving discriminant analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (2013) 2989–2993
31. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1505 – 1518
32. Bhattacharyya, A.: On a measure of divergence between two multinomial populations. Sankhy: The Indian Journal of Statistics (1933-1960) **7** (1946) 401–406
33. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2009)
34. Yang, J., Zhang, Y.: Alternating direction algorithms for l1-problems in compressive sensing. Technical report (2009)
35. Bialkowski, A., Denman, S., Sridharan, S., Fookes, C., Lucey, P.: A database for person re-identification in multi-camera surveillance networks. In: Digital Image Computing: Techniques and Applications. (2012) 1–8
36. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer-Verlag New York, Inc. (2006)