Recognizing People in Non-Intersecting Camera Views

I. Bouchrika, J. N. Carter, M. S. Nixon

Department of Electronics and Computer Science. University of Southampton, UK

Abstract

Many studies have now confirmed that it is possible to recognize people by the way they walk. As yet there has been little formal study of identity tracking using gait over different camera views. We present a new approach for people tracking and identification between different non-intersecting uncalibrated cameras based on gait analysis. An identification signature is derived from gait kinematics as well as anthropometric knowledge. Given the nature of surveillance data, we have developed a new feature extraction technique for finding human legs. The novelty of our approach is motivated by the latest research for people identification using gait. The experimental results confirm the robustness of our method to extract gait features in different scenarios. Furthermore, experimental results revealed the potential of our method to work in real surveillance systems to recognize walking people over different views with achieved cross-camera recognition rates of 95% and 90% for two different views.

1 Introduction

Surveillance technology is now ubiquitous in modern society. This is due to the increasing number of crimes as well as the vital need to provide a safer environment. Tracking people over a network of distributed cameras has recently emerged of great interest to surveillance and monitoring applications. This is because of the impossibility of human operators to work simultaneously on multiple video scenes to track people of interest and analyse their activities. Therefore, it is becoming a vital requirement for the computer vision community to research vision-based alternatives that can automate the process for object hand-off over different views including the tracking and identification of subjects as well as analyse their activities in such environment. Intuitively, a number of approaches were recently proposed to accomplish such aim based on deploying basic feature information including colour or shape. However, their practical deployment in real application is limited due the complexity of such problem. Alternatively, we propose a novel method for tracking subjects across different views by the way of their walk defined as Gait.

Recently, the use of gait for people identification in surveillance applications has attracted researchers from computer vision. The suitability of gait recognition for surveillance systems emerge from the fact that gait can be perceived from a distance as well as its non-invasive nature. In fact, early studies by Johansson [7] on human motion perception using Moving Light Displays have revealed that an observer can recognise different types of human motion based on joint motions. Moreover, the observer can make a judgment of the gender of the person, and even further identify the person if they are already familiar with their gait. Early studies by Murray [10] revealed that gait might be a useful biometric for people identification, a total of 20 feature components including ankle rotation, spatial displacement and vertical tipping of the trunk have been identified to render uniquely the gait signature for every individual. This leads to the conclusion that gait might be a potential biometric for surveillance systems.

As recent results have proven that gait can be a potential biometrics for real surveillance and forensic applications, we explore in this research a novel approach based on gait analysis for tracking and identification of walking subjects over a camera network. The approach is based on marker-less feature extraction to recover joint positions of walking subject from uncalibrated single cameras. The gait signature is thereafter derived in the sagittal plane by a rectification process [15] that transforms gait angular data from a particular viewpoint to the normal plane. In an unconstrained environment where camera information is not available, people are tracked from different viewpoints by matching their gait signature against a database of existing signatures. This way we can recognize people in one camera view from data derived from a different view. Thus, we can recognize and track people in non-intersected camera views.

The remainder of this paper is organized as follows. The next section summarises the previous approaches for object handoff between camera and the use of gait for tracking in surveillance applications. The theoretical description of the proposed marker-less approach for deriving gait-based tracking signature is presented in Section 3 and 4. Section 5 introduces the experimental results and analysis applied.

2 Related Work

There is a large collection of literature from various disciplines that proves of the concept of people identification by their gait. The following section sheds light on the recent state of the art studies related to people tracking across multiple Field of Views (FOV) as well as people recognition using gait covering the different methodologies employed for feature extraction.

2.1 Tracking & Handoff Between Multiple Cameras

Most of the early approaches proposed for tracking across multiple views are limited in a way that they require information about the camera calibration as well as overlapping fields to maintain tracks between different views. Camera calibration is an expensive task and the availability of camera parameters in real life surveillance systems is difficult to provide. Therefore, subsequent methods relax the need for calibration parameters but still require the overlapping views to establish correspondences using different types of features including colour information [13] or geomtrical properties. Cai and Aggrawel *et al* [2] proposed an approach for tracking subjects from sequences of synchronized images from calibrated cameras. The correspondence is established using a set of feature points in a Bayesian probability framework. Subjects are tracked using a single camera view until the system predicts that the active camera will no longer have a good view of the candidate object. Features include geometric properties such as the height of the subject.

Alternatively, Stein *et al* [16] presented a new approach that does not require camera calibration. The camera calibration information is estimated by observing motion trajectories in the scene. Javed et al [8] presented a system for tracking people in multiple uncalibrated cameras. The system is able to discover spatial relationships between the camera fields of views and use this information to correspond between different perspective views of the same person.

2.2 Gait-Based Identification for Tracking

Much of the interest in the field of human gait analysis has originated from physical therapy, orthopaedics and rehabilitation practitioners for the diagnosis and treatment of patients with walking abnormalities. As gait has recently emerged as an attractive biometric, gait analysis has become a challenging computer vision problem. Many research studies have aimed to develop a system capable of overcoming the difficulties imposed by the extraction and tracking of human motion features. Various methods were surveyed in [11].

In 2002 BenAbdelkader *et al.* proposed a pose-free approach, where the moving person is detected and tracked and an image template corresponding to the person's motion blob is extracted in each frame [1]. Subsequently, a self-similarity plot from the obtained sequence of templates has been computed. Experimental tests on outdoor sequences of 44 people with 4 sequences of each taken on two different days achieved a classification rate of 77%. The method has also been tested on indoor sequences of 7 people walking on a treadmill, taken from 8 different viewpoints (from 0° to 120°) and on 7 different days. A classification rate of 78% was obtained for near-fronto-parallel views, and 65% on average over all views.

The pose-based methods, which generate the lateral view from data acquired at different arbitrary views, are the most recent approaches to 2D view-independent gait biometrics. This choice is justified by the fact that the lateral view has proven recognition capability with many approaches [6] and that the pose-free approach works on a small number of camera positions.

In 2003 the group at the University of Maryland developed a gait recognition algorithm showing that if a person is far enough from a single camera, it is possible to synthesize the lateral view from any other arbitrary view by knowing the camera calibration parameters [4]. The method has been tested on 12 people walking along straight lines at different camera views. Considering a gallery of people walking at lateral view, the video sequences where people walk at arbitrary angles were chosen as probes and the Receiver Operating Characteristic (ROC) was computed for each view.

The biometrics research group of the University of Southampton has focused attention on 2D view invariant gait recognition from 1999 [3] where a trajectory-invariant gait signature was presented. The method of Carter and Nixon corrects the variations in gait data by knowing the walking trajectory and modelling the thigh as a simple pendulum. The approach was then reformulated by Spencer and Carter [15] to provide a pose invariant biometric signature which did not require knowledge of the subject's trajectory. Results on synthesized data showed that simple pose correction for geometric targets generalizes well for objects on the optical axis. More recently, these techniques have been refined and applied on subjects wearing reflective markers with success observed from 6 different point views [15].

3 Markerless Extraction of Gait Features

For estimating the lower limb positions, a Haar-like feature template matching approach is proposed for the localization process using motion information. The method does not depend on background subtraction for the derivation of gait features. In contrast to most methods which rely mainly on the quality of the segmented silhouettes or generated background model which are not guaranteed for real surveillance videos. This is due to background clutter, illumination, weather conditions and other outdoor factors in addition to the computing complexity involved when updating the background model within every frame.

For the marker-less extraction of gait feature, we derive motion models based on medical data that describe the angular motion of the knee and hip at different states of the gait cycle. A gait cycle is defined as the time interval between successive instances of initial foot-to-floor contact for the same foot [5]. The hip initially bends or flexes by approximately $20°$ throughout the terminal stance phase, then it extends until it reaches approximately 10 degrees during the stance phase. During the pre-swing and throughout most of swing phase, the hip flexes to nearly $20°$, and then starts to extend just before the next initial contact. The knee angular motion illustrated in Figure (1(b)), shows the knee is almost fully extended then during the first part of the midstance, it gradually begins to flex to its support phase peak which is about 20 degrees. The knee extends again almost fully and then flexes to approximately $40°$ during the pre-swing phase. After toe-off, the knee flexes to reach a peak of 60 to 70 degrees (measured relative to the thigh) at

mid-swing, then extends again in preparation for the next initial contact.
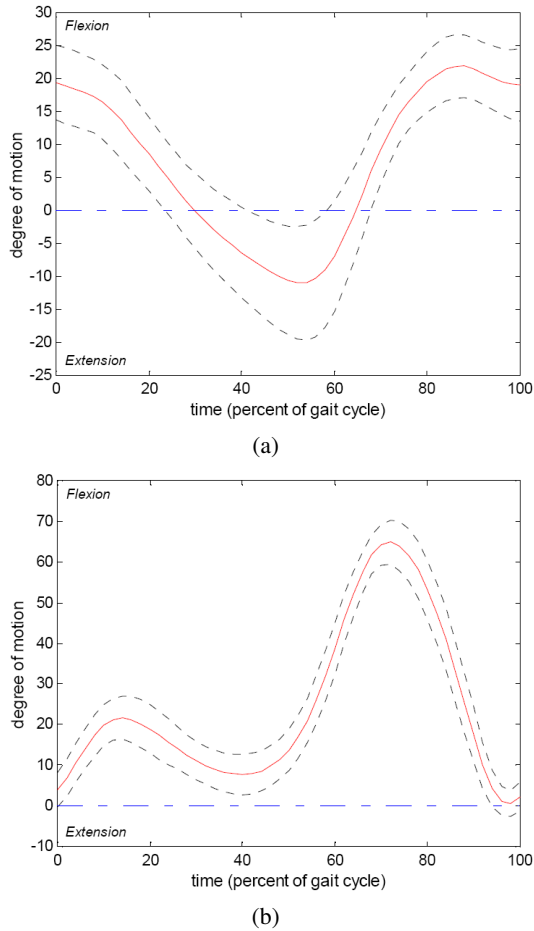

(a)

(b)

Figure 1. Gait Angular Motion: (a) Hip Angular Motion. (b) Knee Angular Motion.

The new algorithm derives the motion map image based on the change detection for the inter-frame difference. Moving pixels for a walking subject across consecutive frames are detected with the emphasis to provide better edge information. The motion map M_t at frame t is computed as the absolute difference of two consecutive frames I_t and I_{t+1} as:

$$M_t = \|I_t - I_{t+1}\| \quad (1)$$

A sample motion image is shown in Figure (2(a)) for a walking subject recorded form a CCTV surveillance camera inside a busy airport.

A *Haar-like* feature [12] is used for the localization of the gait features due their simple and robust performance in real-time application from object recognition to pedestrian detection. The template is shown in Figure (2(b)) which is based on the outlier of the leg. Let p_t^{ankle} is the ankle candidate position at t^{th} frame. To localize the ankle position, different templates are produced accounting for the different possible rotations and translation parameters defined by kinematical knowledge. The templates are superimposed against the motion map image at

the candidate point p computing the match score S as described in Equation (2). The similarity score describes how well is the matching template is superimposed on the motion map. It is estimated as the sum of larger values inside the superimposed region divided by the accumulated lower values inside the area that are less than a certain threshold which is experimentally set to $\tau = 20$.
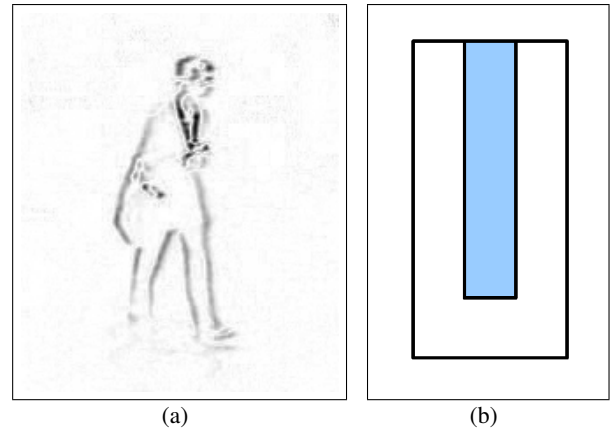

(a)   (b)

Figure 2. Markerless Gait Feature Extration: (a) Motion Image. (b) Haar-based Matching Template.

$$S(x,y,\alpha) = \frac{\sum_{i \in P_{x,y,\alpha}} P_{x,y,\alpha}(i) \times Z(P_{x,y,\alpha}(i))}{\sum_{i \in P_{x,y,\alpha}} P_{x,y,\alpha}(i) \times |1 - Z(P_{x,y,\alpha}(i))|} \quad (2)$$

where and α is the rotation angle and Z is defined as :

$$Z(i) = \begin{cases} 1 & \text{if } i > \tau \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

As opposed to using a per-frame approach for pose recovery [9], a frame-to-frame approach is used for the extraction process where the results from the previous frame are used to guide the matching process in the next frame. To limit the search space for a candidate point and therefore refine the extraction accuracy, Anthropometric and kinematical constraints including spatial as well as angular derived from the gait motion model described in the previous section are imposed. For instance, during the striking phase, one of the legs will be almost stationary at the same position and therefore the ankle spatial movement is limited to a small area whilst the rotation parameter α will be limited within some specific range depending on the stage of the gait cycle. Initially, the bounding box for the walking subject is needed for the first frame which can be derived using a Histogram of oriented Gradients (HoG) approach proposed by Roth *at al.* [14] for people detection.

During the double-support stage of the gait cycle where the legs overlap, it becomes impossible to extract the lower limbs because of the self-occlusion due to the overlap. Therefore, we extract only the striking leg as kinematical gait constraints can be applied aide the localization, meanwhile the swinging leg is skipped during the overlap. The overlapping stage is started whenever the Euclidian distance between the two ankle joints

of both legs is less then a certain threshold related to subject height. The extraction of the skipped leg is resumed after a certain number of frames defined from the average gait cycle model. (Experimentally is set to 6 for a video stream recorded at 25 frames/seconds). In order to estimate the joint positions as well the angular values α when the legs overlap, a 3rd order polynomial interpolation has been applied. The choice of this polynomial has been determined experimentally.

The upper legs orientation is extracted at every frame $\mathbf{T} = [t_1, t_2, \ldots, t_\varphi, \ldots, t_F]$ with a coarse to fine hips estimation procedure where at first, the hips position is achieved with

$$
\begin{cases}
x'_{hip\ell} = \frac{1}{P} \cdot \sum_{j=1}^{P} \widetilde{x}_j + (2\ell - 3) \cdot H \cdot \mu \cdot 10^{-3} \\
y'_{hip\ell} = y'_{hip} \cdot (2\ell - 3) \cdot \left(\frac{\widetilde{x}_P - \widetilde{x}_1}{2} \right) \cdot \sin(0.3 \cdot \mu)
\end{cases}
\tag{4}
$$

where $\widetilde{\mathbf{X}} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_j, \ldots, \widetilde{x}_P]$ is the subset of P ($P \leq C$) horizontal coordinates such as $\overline{\mathbf{S}}(\widetilde{\mathbf{X}}, y'_{hip}, t_i) = 1$.
Equations 4 puts in relationship the horizontal hips position and the walking direction μ, calculated with respect to the horizontal axes of the image reference system. These relationships have been obtained with a regression analysis of the 3D Georgia Tech motion capture data [1] by considering different camera positions. The angle μ is estimated as the angle of inclination of the straight line which approximates the heel strikes points.

4 View-point Rectification

The method proposed by the authors in [15] is based on four main assumptions: the nature of human gait is cyclic; subjects walk along a straight line; the distances between the bone joints are constant; and the articulated leg motion is approximately planar. Therefore, multiple periods of linear gait motion appear analogous to a single period viewed from many cameras related by linear translation and the positions of the legs joints lie in an auto-epipolar configuration.
If $\mathbf{j}_i^\ell$ is the set of joints positions for each leg $\ell = \{1, 2\}$ at the ith frame in the image reference system, the relationship between $\mathbf{j}_i^\ell$ and the corresponding positions in the worldspace is $\mathbf{j}_i^\ell \times \mathbf{P}_i \cdot \mathbf{J}^\ell = 0$, where $\mathbf{P}_i = \left[\mathbf{R}_\mathbf{e}^T, -i\mathbf{e}_0 \right]$ and $\mathbf{R}_\mathbf{e}^T$ is the rotation matrix for aligning the epipolar vector $\mathbf{e}_0$ with the horizontal axis X. Then,

$$
\mathbf{j}_i^\ell = \mathbf{P}_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H_V^{-1}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H_V} \end{pmatrix} = \mathbf{H} \cdot \mathbf{J}^\ell
\tag{5}
$$

having expressed the limb plane transformation matrix with $\mathbf{H_V}$ so that the two cross section plane lines are centred and normalised respect to Y and Z axes and parallel with Y. By assuming the lengths of the articulated limbs $\mathbf{D}_\ell^2 = \Delta \mathbf{j}_i^{\ell T} \Delta \mathbf{j}_i^\ell$ are constant over all the frames, the pose difference vectors for the limb segments at two consecutive frames, $\Delta \mathbf{j}_i^\ell$ and $\Delta \mathbf{j}_{i+1}^\ell$, are related by

$$
\Delta \mathbf{j}_i^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_i^\ell = \Delta \mathbf{j}_{i+1}^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_{i+1}^\ell
\tag{6}
$$

[1] http://www.cc.gatech.edu/cpl/projects/hid

After recovering the fronto-parallel structure of subject gait, the representation of the leg joints function $\left[\mathbf{J}_x^\ell(t), \mathbf{J}_y^\ell(t) \right]$ is found by fitting a modified Fourier series to the data with fixed fundamental frequency f_0 and period T:

$$
\mathbf{J}_x^\ell(t) = v_x t + \sum_{k=1}^{n} A_k \cos \left(2\pi k f_0 \left(t + \frac{(\ell - 1) T}{2} \right) + \phi_k \right) + \mathbf{J}_{x0}^\ell
\tag{7}
$$

analogously for $\mathbf{J}_y^\ell(t)$. Thus, the projection of the leg joints on the lateral plane is obtained with an optimized procedure in the following way

$$
\breve{\mathbf{J}}^\ell(t) = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} g \left(t + \frac{(\ell - 1) T}{2} : f_0, \mathbf{D}_\ell, v_x, v_y, F \right)
\tag{8}
$$

where $g(t)$ is the bilateral Fourier series function with coefficients F and h are the values of the inverse normalization transform matrix.
Therefore, starting from a video sequence from a single camera and without any calibration, the proposed markerless system, in junction with [15], estimates the gait parameters projected on the lateral plane.

5 Experimental Results

To demonstrate the efficacy of our method for tracking walking subjects in multi-camera surveillance network, the system has been extensively evaluated on a variety of scenarios and conditions. The Casia-B database [17] has been considered as the real test-bed of the proposed view-invariant method since every single subject's walk is simultaneously captured from a large number of different views, and this establishes confidence in the viewpoint invariant approach. The markerless feature extraction method has been applied to the CASIA-B gait dataset consisting of 2270 video sequences for 65 different walking subjects with an average of 6 sessions for every viewpoint per subject. There are 6 different camera orientations (36°, 54°, 72°, 90°, 108°, 126°). The 90° corresponds to the side view walking direction. The limbs pose has been estimated frame by frame and the hip and knee angles have been extracted for each camera position and for each subject. Figure 3 shows an example of the limbs pose estimation in the 6 different directions.
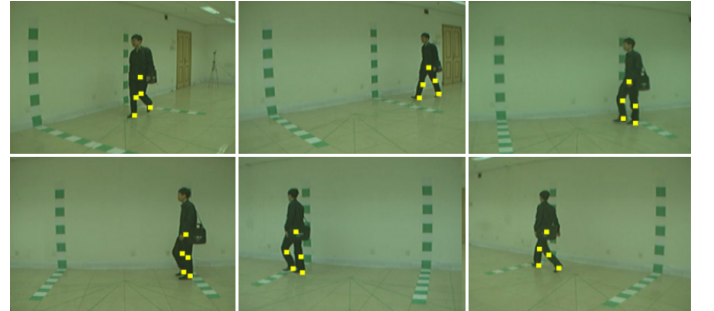

Figure 3. Markerless Extraction of Gait Features.

To investigate the effects of the viewpoint as a covariate factor for the performance of gait recognition, an initial experiment was carried out to measure the recognition rate using the non-rectified gait data. Successively, the viewpoint rectification method described in section 2.4 is applied in order to reconstruct the gait angular features into the lateral plane (i.e. lateral view). Figure 4(a) shows an example of the variations of hip angular motion during two gait cycle for the six different camera positions in the real experimental tests (CASIA B). Predictably, the angles trends, extracted in the image reference system, are influenced by the subject pose with respect to the camera and they cannot be used directly for identification. For this reason, the view point correction algorithm is applied and the angle trends after the correction are shown in figure 4(b).

Figure 5. CMS curve of the gait identification for the rectified and unrectified data.

using the iLids dataset provided the UK Home Office. The dataset is taken from CCTV cameras at Gatwick International Airport with overlapping and non-overlapping views. We have gathered a set of 20 different walking subjects who are seen on different camera views (Camera 2 and 3). The marker-less feature extraction method is applied on the dataset of walking people for pose recovery in order to derive gait signatures from the gait and anthropometric features. Sample results of the extraction are shown in Figure 6.
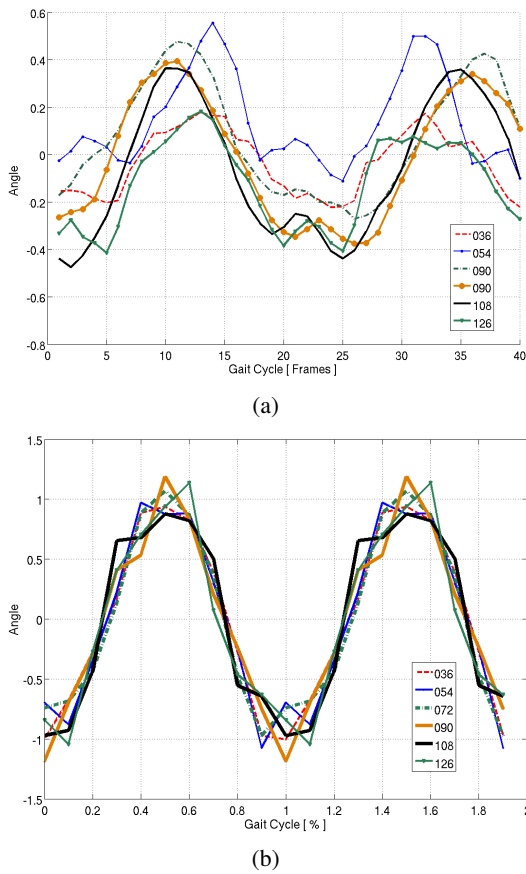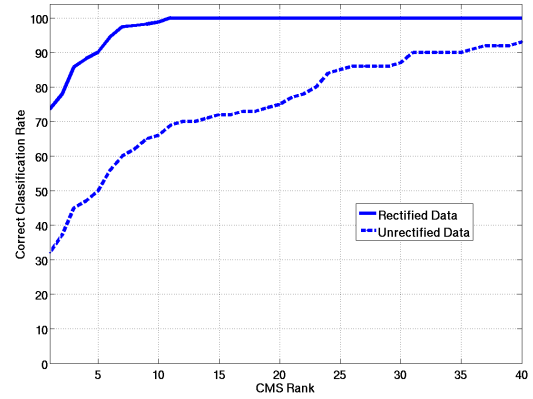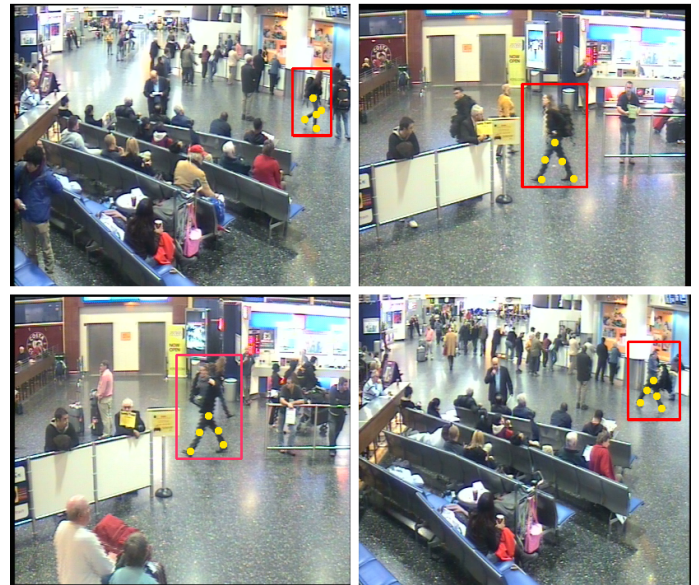

(a)

(b)

Figure 4. Hip Angular Motion from different View-points: (a) Unrectified Data. (b) Rectified Angular Data.

Another useful evaluation measure is the Cumulative Match Score (CMS). The measure assesses the ranking capabilities of the recognition system by producing a list of scores that indicates the probabilities that the correct classification for a given test sample is within the top n matched class labels. A classification score of 73.6% and 100% are achieved at the 1st and 11th rank respectively. The CMS score at 1st rank is the correct classification rate. Figure 5 shows the CMS curve of the gait identification for the rectified and unrectified data.

To explore the performance of the proposed method in real life surveillance applications, we have evaluated the approach

Figure 6. Feature Extration applied on the iLids dataset.

To evaluate the usefulness of the proposed approach in real cases, we have used the leave-one-out cross validation with the KNN classifier to initially estimate the performance across all the 20 subjects across two different cameras. The achieved correct classification rate (CCR) is 97% for the value of $k = 1$. Further, we have matched the data from camera 3 against camera 2 and vice versa in a probe to gallery fashion. We have

	Camera 2 vs Camera 3	Camera 3 vs Camera 2
CCR	90%	95%

Table 1. Correct Classification Rate Analysis

increased the size of people in camera 3 to include 10 more random subjects that do not exist in the 2nd camera dataset when probing camera 2 against the camera 3. In the same way, we have increased the size of dataset camera 2. The achieved classification results are over 90% for the cross-camera matching as shown in Table 1. This shows that gait features can be used in surveillance applications for identity tracking over different camera views.

6 Conclusions

We have taken an important step in deploying gait recognition for the analysis of surveillance video. We present a new approach for people tracking and identification between different non-intersecting uncalibrated cameras based on gait analysis. Identification signature is derived from gait kinematical as well as anthropometric knowledge that are obtained using a markerless feature extraction algorithm. Experimental results revealed the potential of our method to work in real surveillance systems to recognize walking people over different views using the marker-less pose recovery with an achieved recognition rate of 97%. This concludes that identity tracking i.e. object handoff can be achieved by gait analysis and we have encouraging results using the marker-less feature extraction and rectification approach. This is an important step in translating gait biometrics into real scenarios where calibration information cannot be recovered such as in surveillance applications.

Acknowledgements

References

[1] C. BenAbdelkader, R. Cutler, and L. Davis. View-invariant Estimation of Height and Stride for Gait Recognition. *In Proceedings of Workshop on Biometric Authentication ECCV*, 2002.

[2] Q. Cai, JK Aggarwal, R. Inc, and WA Seattle. Tracking human motion in structured environments using adistributed-camera system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1241–1247, 1999.

[3] J. N. Carter and M. S. Nixon. Measuring gait signatures which are invariant to their trajectory. *Measurement+Control*, 32(9):265–9, 1999.

[4] A.K.R. Chowdhury, A. Kale, and A. Kale. Towards a view invariant gait recognition algorithm. In *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 143–150. Citeseer, 2003.

[5] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic Extraction and Description of Human Gait Models for Recognition Purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003.

[6] P. S. Huang, C. J. Harris, and M. S. Nixon. Human Gait Recognition in Canonical Space using Temporal Templates. *In IEE Proceedings of Vision, Image and Signal Processing*, 146(2):93–100, 1999.

[7] G. Johansson. Visual Perception of Biological Motion and a Model for its Analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[8] S. Khan, O. Javed, Z. Rasheed, and M. Shah. Human tracking in multiple cameras. In *Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. Proceedings*, volume 1, 2001.

[9] G. Mori and J. Malik. Recovering 3d human body configurations using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006.

[10] M. P. Murray. Gait as a Total Pattern of Movement. *American Journal of Physical Medicine*, 46(1):290–333, 1967.

[11] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait (The Kluwer International Series on Biometrics)*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2005.

[12] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. Proceedings.*, pages 193–199, 1997.

[13] J. Orwell, P. Remagnino, and GA Jones. Multi-camera colour tracking. In *Second IEEE Workshop on Visual Surveillance*, pages 14–21, 1999.

[14] P.M. Roth, H. Grabner, D. Skocaj, H. Bischof, and A. Leonardis. On-line conservative learning for person detection. In *Proc. IEEE Workshop on VS-PETS*, pages 223–230, 2005.

[15] N. Spencer and J. N. Carter. Towards Pose Invariant Gait Reconstruction. *In Proceedings of IEEE International Conference on Image Processing*, 3, 2005.

[16] G.P. Stein. Tracking from multiple view points: Self-calibration of space and time. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 521–527, 1999.

[17] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proc. of International Conference on Pattern Recognition*, volume 4, pages 441–444, 2006.