

Performance Analysis for Gait in Camera Networks

Michela Goffredo, Imed Bouchrika, John N. Carter and Mark S. Nixon

ECS, University of Southampton (UK)
{mg2,ib04r,jnc,msn}@ecs.soton.ac.uk

ABSTRACT

This paper deploys gait analysis for subject identification in multi-camera surveillance scenarios. We present a new method for viewpoint independent markerless gait analysis that does not require camera calibration and works with a wide range of directions of walking. These properties make the proposed method particularly suitable for gait identification in real surveillance scenarios where people and their behaviour need to be tracked across a set of cameras. Tests on 300 synthetic and real video sequences, with subjects walking freely along different walking directions, have been performed. Since the choice of the cameras' characteristics is a key-point for the development of a smart surveillance system, the performance of the proposed approach is measured with respect to different video properties: spatial resolution, frame-rate, data compression and image quality. The obtained results show that markerless gait analysis can be achieved without any knowledge of camera's position and subject's pose. The extracted gait parameters allow recognition of people walking from different views with a mean recognition rate of 92.2% and confirm that gait can be effectively used for subjects' identification in a multi-camera surveillance scenario.

Keywords

Gait Analysis, Biometrics, Multi-view, Surveillance, Object Handover.

1. INTRODUCTION

Activity monitoring for security purposes and work flow control in critical infrastructures can be successfully achieved with a distributed network of video sensors, instead of having one high resolution camera with a limited field of view. Multiple-cameras, in fact, provide a solution to wide and complex areas where overlapped and non-overlapped views allow more accurate surveillance of human action. The aim of a smart surveillance system is to automatically track people across a network of cooperating active cameras and de-

tect abnormal behaviours. In this context, "object handover" is defined as the subject's identification and tracking across a distributed network of cameras, that can have both overlapped and non-overlapped views.

Several strategies have been employed for coordinating subject tracking between multiple overlapping camera views [5, 27]. They assumed that 3D ground calibration information was available for each camera or estimate it by analysis of training data. Recently there has been some interest in tracking objects between blind regions [21, 23, 26]. They make use of Kalman filtering assuming the ground plane is known [9] or assume that the transition models are known or hand-labelled [13, 37]. Some approaches recover the relative cameras' position and the subject's movement across them with a statistical approach [28, 40]: two cameras are considered connected if objects seen departing in one camera are seen arriving in the other and the statistical dependence is characterized by the distribution of observation transformations between cameras, such as departure to arrival transition times.

While there is a large number of papers regarding the tracking between blind regions, a few works address the issue of subject's identification across a network of video sensors [7, 14, 25, 31, 39]. Appearance information is indeed one of the most popular solutions for solving the object handover issue since it is relatively robust towards size and orientation changes [7, 14, 25, 31]. Possible appearance features are colour templates, histograms, moments, signatures as dominant colours, and partitive colour layouts. Since illumination, camera distortion and object resolution change in a multi-camera setting, these approaches include an inter-camera distortions and illumination compensation [7, 14, 22]. However, in some environments like the military or industrial scenarios, where everyone dresses in identical clothes, appearance information cannot be applied. Furthermore, the colour-based approaches work in a short time period and are not suitable for a long-term workflow control. To this extent, Stillman *et al.* used face identification in a multimodal approach where acoustic and visual data are combined [39]. However, in typical surveillance applications cameras are usually located far away from the subjects, which result in low-resolution face images. Another concern is the face orientation: most face-based methods work only for frontal images and the accuracy of identification quickly decreases even with the slight orientation differences.

In this paper, we propose gait analysis as a solution for subjects' identification across a network of cameras.

Recognizing individuals by the way they walk is a particularly challenging research area, where the potential for personal identification is supported by a rich literature, including medical and psychological studies [11, 30]. The completely unobtrusiveness without any subject cooperation or contact for data acquisition make gait particularly attractive for identification purposes in camera handover. Gait recognition techniques at the state of the art can be divided into 3D [4, 29] and 2D approaches [20, 33, 45]. In the first group, identification relies on parameters extracted from the 3D limb movement. These methods use a large number of digital cameras and the 3D reconstruction is achieved after a camera calibration process. On the other hand, the 2D gait biometric approaches extract explicit features describing gait by means of human body models [6, 34] or silhouette shape [10, 41, 43]. A rich variety of data has been collected for evaluation of 2D gait biometrics. The widely used and compared databases on gait recognition include: the University of South Florida [35]; Carnegie Mellon University [17]; and the University of Southampton [38] data. The majority of methods and databases found in the literature use a single camera positioned with specific orientations with respect to the subject's walking direction (generally capturing the walk from the lateral view) and a large number of papers for gait recognition have been published [19, 33].

In real surveillance scenarios, however, we need a system that operates in an unconstrained environment where maybe there is no information regarding the relative position between the camera and the walking subject [44]. 2D view independent gait identification is a recent research area and the approaches found in literature extract some gait parameters which are independent of the human pose [3, 42] by analysing the silhouette shape over time or aim at synthesising the sagittal view of the human body from any other arbitrary views [16, 24]. The latter choice is justified by the fact that the lateral view has proven recognition capability in a great number of works [20, 33, 45]. However, they need information on the camera parameters [24] or rely on the use of reflective markers on the body [16].

In this paper a novel 2D markerless view independent gait analysis algorithm is presented and tested for assessing its applicability for object handover in multi-camera scenarios. The method does not need camera calibration or pre-knowledge of subject pose. Since the choice of the cameras' characteristics is a key-point for the development of a smart surveillance system, the performance of the proposed approach is measured with respect to different video properties.

2. THEORY

The proposed method for the view independent gait analysis is based on two consecutive stages: the markerless joints' estimation from the image sequence and the view-point rectification. Gait parameters are then used for identification.

2.1 Markerless Joints' Position Estimation

Let $\mathbf{S}(\mathbf{x}, \mathbf{y}, t_\varphi)$ be the RGB frame (of size $R \times C$ pixels) at time $\mathbf{T} = [t_1, t_2, \dots, t_\varphi, \dots, t_F]$ where (x_1, y_1) is the top-left corner of the image. By applying a background subtraction method based on the threshold of the three components of the color space YUV, the binary image $\bar{\mathbf{S}}(\mathbf{x}, \mathbf{y}, t_\varphi)$ has been extracted. The pixels $(\mathbf{x}_{sil}, \mathbf{y}_{sil})$ containing the hu-

man silhouette have been selected as the object with maximum area [18]. The lower limbs pose estimation algorithm is based on the proportions of the human body segments, following the medical results of anatomical studies [12]:

$$\begin{aligned} y'_{hip} &= \min(\mathbf{y}_{sil}) + 0.5 \cdot H \\ y'_{knee} &= \min(\mathbf{y}_{sil}) + 0.75 \cdot H \\ y'_{ankle} &= \min(\mathbf{y}_{sil}) + 0.90 \cdot H \end{aligned} \quad (1)$$

where H is the silhouette's height.

Subsequently, the shins and thighs have been analysed separately. Let $\hat{\mathbf{X}}$ be the set of horizontal coordinates where $\bar{\mathbf{S}}(\hat{\mathbf{x}}, y'_{knee}, \hat{t}_\delta) = 1$, then the subset of frames

$$\hat{\mathbf{T}} = [\hat{t}_1, \hat{t}_2, \dots, \hat{t}_\delta, \dots, \hat{t}_{D \leq F}] \quad (2)$$

where the legs do not overlap are those where $\hat{\mathbf{X}}$ forms 2 single connected components larger than $0.02H$ pixels.

Consequently, for each frame \hat{t}_δ , the shin extraction algorithm, based on the linear approximation of the skeleton of the portion of image that includes the lower leg, is applied on $\bar{\mathbf{S}}(\mathbf{x}, \mathbf{y}, \hat{t}_\delta)$. The sub-images containing the shins (labeled as \mathbf{S}_{shin1} and \mathbf{S}_{shin2}) are extracted from

$$\mathbf{S}_{lower\ leg}(\mathbf{x}, \mathbf{y}, \hat{t}_\delta) = \begin{cases} \bar{\mathbf{S}}(\mathbf{x}, y_s, \hat{t}_\delta) & \text{if } y_s \in \mathbf{y}_{shin} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where the vertical coordinates belonging to the two shins $\mathbf{y}_{shin} = [y_1, y_2, \dots, y_s, \dots, y_S]$ are defined in the following way:

$$\mathbf{y}_{shin} = [y'_{knee}, y'_{knee} + 1, \dots, \min(y'_{ankle}, y_{overlap} - 1)] \quad (4)$$

which includes the possibility of feet overlapping with $y_{overlap}$: vertical coordinate where the set of correspondent horizontal coordinates $\mathbf{x}_{overlap}$ belong to a single object larger than $0.02H$ pixels.

Therefore, the left and right shins are defined by the pixels with coordinates

$$\begin{cases} \mathbf{x}_{shin\ell}^{skel}(\hat{t}_\delta) = [x_{1\ell}, x_{2\ell}, \dots, x_{s\ell}, \dots, x_{S\ell}] \\ \mathbf{y}_{shin\ell}^{skel}(\hat{t}_\delta) = \mathbf{y}_{shin} \end{cases} \quad \ell = \{1, 2\} \quad (5)$$

where

$$x_{s\ell} = \sum_{j=1}^C x_j \cdot \mathbf{S}_{shin\ell}(x_j, y_s, \hat{t}_\delta) / \sum_{j=1}^C \mathbf{S}_{shin\ell}(x_j, y_s, \hat{t}_\delta) \quad (6)$$

Then, the shins can be linearly approximated by the first order polynomial with coefficients

$$\mathbf{p}_\ell(\mathbf{x}_{shin\ell}^{skel}, \hat{t}_\delta) = p_{\ell 0}(\hat{t}_\delta) + p_{\ell 1}(\hat{t}_\delta) \cdot \mathbf{x}_{shin\ell}^{skel}(\hat{t}_\delta) \quad (7)$$

Therefore, the angle between the shins and the vertical axis (called knee angle) at frame \hat{t}_δ is $\alpha_\ell(\hat{t}_\delta) = \pi - \arctan[p_{\ell 1}(\hat{t}_\delta)]$.

The procedure is applied for each frame in the subset $\hat{\mathbf{T}} = [\hat{t}_1, \hat{t}_2, \dots, \hat{t}_\delta, \dots, \hat{t}_D]$.

In order to also be able to estimate α_1 and α_2 when the legs overlap, the extracted shins are labeled as 'left' and 'right' according to the following rationale without losing generalities: the subject is walking for at least two gait cycles; in t_1 the shin labeled as *shin1* belongs to the right leg; the labels swap after the legs overlap.

Subsequently, a 3rd order polynomial interpolation of α_1 and α_2 has been applied and the knees angles are achieved during the whole video sequence. This choice has been determined experimentally.

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

$$\begin{cases} x'_{hip\ell} = \frac{1}{P} \cdot \sum_{j=1}^P \tilde{x}_j + (2\ell - 3) \cdot H \cdot \mu \cdot 10^{-3} \\ y'_{hip\ell} = y'_{hip} \cdot (2\ell - 3) \cdot \left(\frac{\tilde{x}_P - \tilde{x}_1}{2} \right) \cdot \sin(0.3 \cdot \mu) \end{cases} \quad (8)$$

where $\tilde{\mathbf{X}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_j, \dots, \tilde{x}_P]$ is the subset of P ($P \leq C$) horizontal coordinates such as $\mathbf{S}(\tilde{\mathbf{X}}, y'_{hip}, t_i) = 1$.

Equations 8 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 regressive 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 with coordinates:

$$(\mathbf{x}_{str}, \mathbf{y}_{str}) = (\mathbf{x}_{feet}, \mathbf{y}_{feet}) \mid \sum_{i=1}^P \mathbf{SK}(\mathbf{x}_{feet}, \mathbf{y}_{feet}, t_i) > \tau \quad (9)$$

where \mathbf{SK} is the skeleton [32] of the image, \mathbf{x}_{feet} and \mathbf{y}_{feet} are the coordinates corresponding to the portions of silhouette belonging to the feet and the threshold τ is automatically set so that at least 4 heels strikes are extracted (in accordance with the hypothesis listed above).

Subsequently, a finer hips pose estimation is obtained with a linear approximation of the thighs by the first order polynomial with coefficients

$$\begin{aligned} q_{\ell 0}(t_\varphi) &= \frac{y'_{shin\ell} \cdot x'_{hip\ell} - y'_{hip\ell} \cdot x'_{shin\ell}}{x'_{hip\ell} + x'_{shin\ell}} \\ q_{\ell 1}(t_\varphi) &= \frac{1}{2} \cdot \left(a'_\ell + \frac{y'_{hip\ell} - q_{\ell 0}}{x'_{hip\ell}} \right) \end{aligned} \quad (10)$$

where $y'_{shin\ell} = \min(\mathbf{y}_{shin})$ and $x'_{shin\ell}$ is the mean value of the correspondent horizontal coordinates. a'_1 and a'_2 are the slope of the straight lines that approximate respectively the left and right edges of the positions of silhouette belonging to the thighs.

Therefore, the angle between the thighs and the vertical axis at every frame t_φ is $\gamma_\ell(t_\varphi) = \pi - \arctan[q_{\ell 1}(t_\varphi)]$.

Ultimately, the thighs are labeled as 'left' or 'right' in accordance with the corresponding shin labeling. The angles $\alpha_1, \alpha_2, \gamma_1$ and γ_2 are then corrected by projecting the limbs trajectories on the lateral plane, as explained next.

2.2 View-point Independent Rectification

The method proposed by the authors in [16] 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^ℓ is the set of joints positions for each leg $\ell = \{1, 2\}$ at the i^{th} frame in the image reference system, the relationship between \mathbf{j}_i^ℓ and the corresponding positions in the worldspace is $\mathbf{j}_i^\ell \times \mathbf{P}_i \cdot \mathbf{J}^\ell = 0$, where $\mathbf{P}_i = [\mathbf{R}_e^T, -i\mathbf{e}_0]$ and \mathbf{R}_e^T is the rotation matrix for aligning the epipolar vector \mathbf{e}_0 with the

horizontal axis X. Then,

$$\mathbf{j}_i^\ell = \mathbf{P}_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_V^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_V \end{pmatrix} = \mathbf{H} \cdot \mathbf{J}^\ell \quad (11)$$

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

$$\Delta \mathbf{j}_i^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_i^\ell = \Delta \mathbf{j}_{i+1}^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta \mathbf{j}_{i+1}^\ell \quad (12)$$

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

$$\mathbf{J}_x^\ell(t) = v_x t + \sum_{k=1}^n A_k \cos \left(2\pi k f_0 \left(t + \frac{(\ell-1)T}{2} \right) + \phi_k \right) + \mathbf{J}_{x0}^\ell \quad (13)$$

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

$$\tilde{\mathbf{J}}^\ell(t) = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} g \left(t + \frac{(\ell-1)T}{2} : f_0, \mathbf{D}_\ell, v_x, v_y, F \right) \quad (14)$$

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 [16], estimates the gait parameters projected on the lateral plane.

2.3 View-point Invariant Gait Identification

The processing and derivation of good gait features from this trajectory-based data, is still an unsolved and challenging problem due to the complexity of the human visual perception system as well as the compound nature of gait motion inherent in the numerous variables associated with it including kinematics, kinetics and anthropometrics [8]. An important issue in gait recognition is the derivation of appropriate features that can capture the discriminative individuality from the subject's gait. Such features should respond to crucial criteria such as robustness and invariance to weather conditions, clothing and operating conditions.

In order to identify a subject by their gait, we derive the angular measurements, anthropometric measurements as well as the trunk spatial displacement which best describe the gait kinematics. The use of angular motion is very common in gait analysis and recognition. The angles of the joints including the hip and the knee; are considered the most important kinematics of the lower limbs. The anthropometric measurements include the subject height and lengths of the lower limbs. Feature selection is employed to derive as many discriminative cues as possible whilst removing the redundant and irrelevant gait features which may degrade the recognition rate. It is practically infeasible to run an exhaustive search for all the possible combinations of features in order to obtain the optimal subset for recognition due to

the high dimensionality of the feature space. For this reason, we employed the Adaptive Sequential Forward Floating Selection (ASFFS) search algorithm. The algorithm uses a validation-based evaluation criterion which is proposed to find the subset of features that minimises the classification errors as well as ensure good separability between the different classes. In contrast to the voting scheme used in the KNN, the evaluation function uses different weights w to signify the importance of the most nearest neighbours. The probability score for a sample s_c to belong to class c is expressed in the following equation (15):

$$f(s_c) = \frac{\sum_{i=1}^{N_c-1} z_i w_i}{\sum_{i=1}^{N_c-1} w_i} \quad (15)$$

where N_c is the number of instances in class c , and the weight w_i for the i^{th} nearest instance is related to proximity as:

$$w_i = (N_c - i)^2 \quad (16)$$

The value of z_i is defined as:

$$z_i = \begin{cases} 1 & \text{if } nearest(s_c, i) \in c \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

such that the $nearest(s_c, i)$ function returns the i^{th} nearest instance to the sample s_c . The Euclidean distance metric is employed to find the nearest neighbours.

3. EXPERIMENTAL TESTS

The proposed approach for subject identification from different view-points has been tested both on synthetic and on real data. Different video properties have been considered in order to compare the performance of the method with respect to the cameras' properties.

3.1 Performance Analysis of Gait Feature Extraction

A quantitative validation of gait analysis has been obtained in a virtual environment: synthetic video sequences (500x490 pixels, 30fps) of a humanoid walking for one gait cycle have been created with Poser 7[®] (by e-frontier) in order to provide the ground truth for validation purposes. Different spatial resolutions, image qualities and compressions have been considered. A range of spatial resolutions (250x245, 163x163 pixels) have been considered by sub-sampling the original images with factors 2 and 3. Table 1 shows the comparative results in terms of mean (m_e) and standard deviation (σ_e) of joint's angles compared to the ground truth over the gait cycle.

The results, with a mean value of 2.6 ± 32.61 , are particularly encouraging and present the same magnitude to the ones obtained with 3D markerless systems or 2D complex model based ones [9][12]. The image resolution and silhouette's height reduce the algorithm performance, as predictable. On the other hand, a 50-pixels-height silhouette still maintains an acceptable result. Furthermore, Table 1 shows the mean time for processing one frame (Matlab7[®] code processed with 2GHz Intel CPU): the model-free proposed approach allows to extract the kinematic information with computational cost per silhouette's pixel lower than $2 \cdot 10^{-4}$ s.

In addition, to simulate background subtraction imprecision,

Image resolution/ mean sil's height	m_e (deg)	σ_e (deg)	Proc. time (s/frame)
500x490/200	2.04	2.12	1.44
250x245/100	2.80	2.15	0.97
163x163/50	3.06	3.56	0.37

Table 1: Angle's errors (in terms of mean and standard deviation) on synthetic gait with different spatial resolution.

σ_n / PSNR(dB)	m_e (deg)	σ_e (deg)
0/ ∞	2.04	2.12
5/28.12	2.30	2.13
25/24.25	3.01	3.38
50/21.89	3.30	3.72
100/15.23	3.25	3.68
150/7.01	4.25	3.49

Table 2: Angle's errors (in terms of mean and standard deviation) on synthetic gait with different zero-mean Gaussian noises.

zero-mean Gaussian noise has been added to the synthetic images. The standard deviation (σ_n) of the added noise varies from 0 (original data) to 150 corresponding to PSNR from ∞ to 7.01dB.

Results in Table 2 show m_e and σ_e with respect to the different PSNR. Obviously the error increases with the level of noise added to the frames but it is lower than 5 degrees even at high noise levels. The noise robustness allows to understand how the proposed method depends on an accurate silhouette extraction and thus to extend the approach in a real context.

Since images from a network of cameras can be compressed for data storage problems, different JPEG compressions have been compared. Seven quality factors between 100 to 5 have been used for compressing the synthetic images, allowing a compression ratio $CR = S_u/S_c$ between 40 to 146, where S_u and S_c are the sizes of the uncompressed and compressed images respectively.

Figure 1 shows m_e with respect to the image compression. The value obtained with uncompressed images is also reported. The mean error increases rapidly with the compression, then remaining constant at about 3.3 pixels for JPEG qualities between 75 and 25.

The proposed method for gait analysis has been tested on real data from the CASIA-B database [2] where 12 young healthy subjects walk along straight lines with 6 different camera orientations ($36^\circ, 54^\circ, 72^\circ, 90^\circ, 108^\circ, 126^\circ$). The 90° corresponds to the side view walking direction as shown in Figure 2. The video sequences have a spatial resolution and frame rate of 320x240 pixels and 25fps respectively. The method has been applied to the video sequences, 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 2 shows an example of the limbs pose estimation in the 6 directions. The frames have been compressed with different JPEG qualities and the joints' position has been compared with the one extracted from

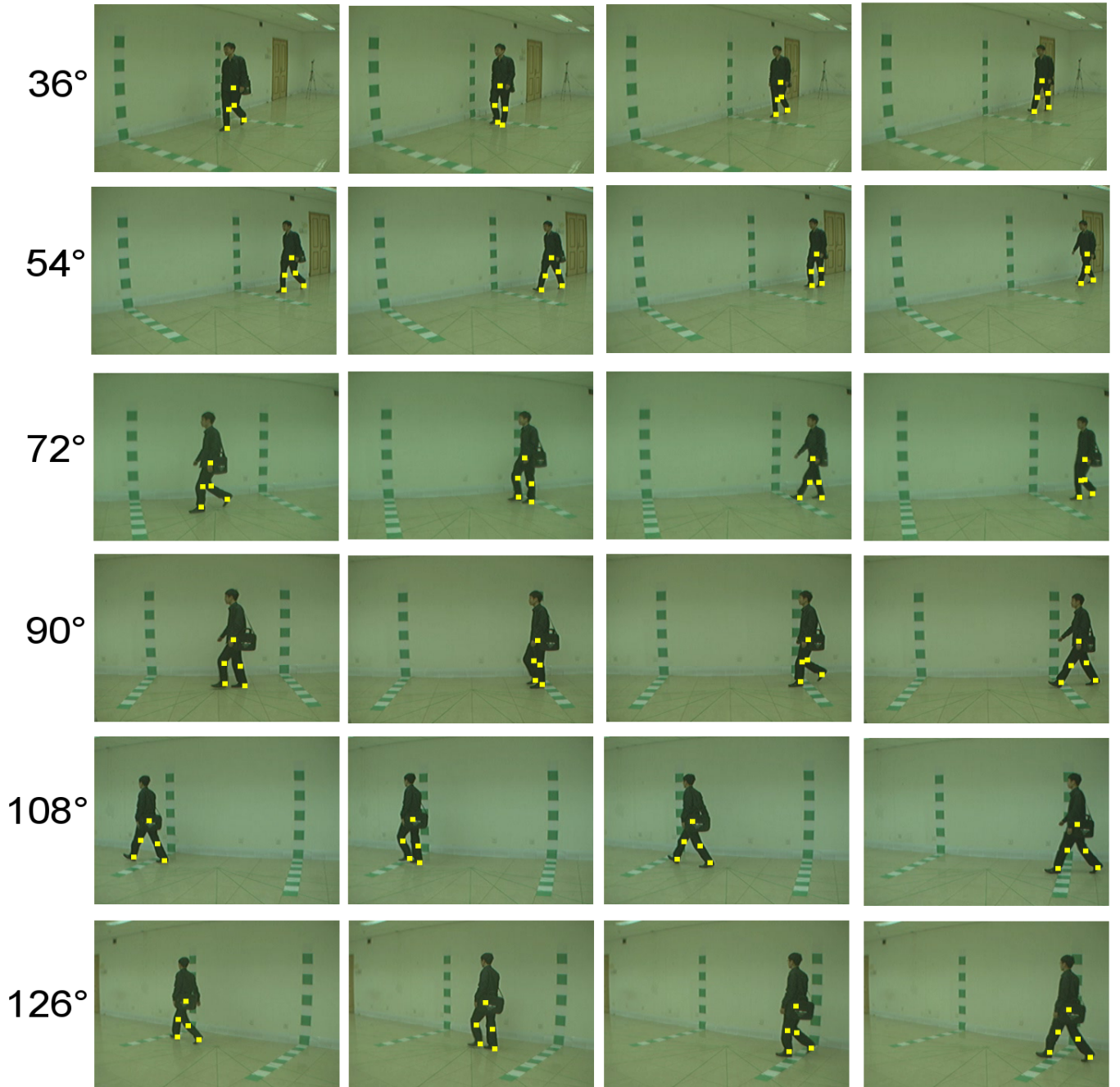


Figure 2: Markerless joints extraction in different view points

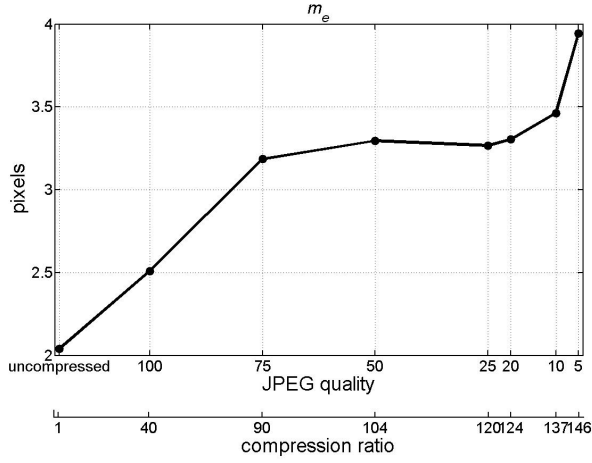


Figure 1: Mean angle's errors on synthetic gait with different image compressions.

the uncompressed images. The mean joints' distance is then measured:

$$m_d = \frac{\sum_{i=1}^I \sum_{v=1}^V \sum_{j=1}^J \sqrt{(x_{ivj}^c - x_{ivj}^u)^2 + (y_{ivj}^c - y_{ivj}^u)^2}}{I \cdot V \cdot J} \quad (18)$$

where the superscripts '*u*' and '*c*' indicate the joints' position from the uncompressed and compressed images respectively. *I* is the number of individuals, *V* the number of video sequence per person and *J* the number of tracked joints. Figure 3 depicts *m_d* with respect to the image compressions. The results are in accordance with the one obtained on synthetic images and show an increase of vertex's de-location especially for qualities higher than 25. The results obtained with different compressions confirm the applicability of gait analysis in real surveillance scenarios where cameras usually presents JPEG compression qualities higher than 50.

3.2 Viewpoint Analysis of Gait Recognition

In order to assess the performance of gait recognition from different viewpoints using a single uncalibrated camera, a set of 275 video sequences with 12 different subjects recorded at 6 viewpoints are taken from the Casia gait database. The Correct Classification Rate (CCR) is computed using the *K*-nearest neighbour (KNN) classifier with the Leave-one-out cross-validation rule. The *KNN* rule is applied at the classification phase due to its low complexity and hence fast computation besides the ease of comparison to other methods. In the leave-one-out validation, every instance from the original sample is used for testing and is validated against the remaining observations. This is repeated for all the observations in the dataset. The recognition rate is computed as the average of all validations.

To investigate the effects of the viewpoint, an initial experimental is carried out to measure the recognition rate using the non-rectified gait data. The CCR is computed separately for every viewpoint. Figure 4 shows the variation of the CCRs with respect to the different viewpoints.

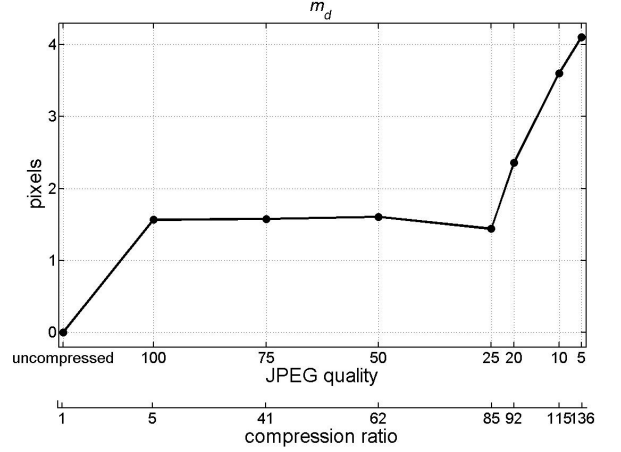


Figure 3: Mean joints' distance on real gait with different image compressions.

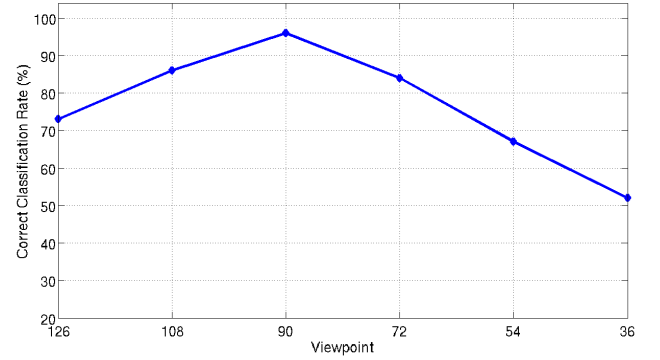


Figure 4: Gait Recognition from different Viewpoints using Non-Rectified Data.

The highest recognition rate is 96% for the side view whilst the recognition rate decreases when the walking direction changes from the lateral view. This is because of the variations in the extracted angular features from different views. Therefore, non-rectified gait angular data cannot be used directly for biometric applications. This was confirmed by the recent research study by Goffredo [15].

In the second experiment, the viewpoint rectification method described in section (2.2) is applied to reconstruct the gait angular features into the lateral plane (i.e. sideview). Based on a set of 275 video sequences of walking subjects, a high recognition rate of 92.2% is achieved for all the viewpoints combined together. It is worth to note that the database has other covariate factors including clothing and load carriage. This is to signify the importance of the gait reconstruction into translating gait as a biometric method into real world surveillance and monitoring applications.

Another useful evaluation measure is the Cumulative Match Score (CMS) which was introduced by Phillips *et al* in the FERET protocol [36] for the evaluation of face recognition

algorithm. 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 92.2% and 100% are achieved at the 1st and 7th 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 data.

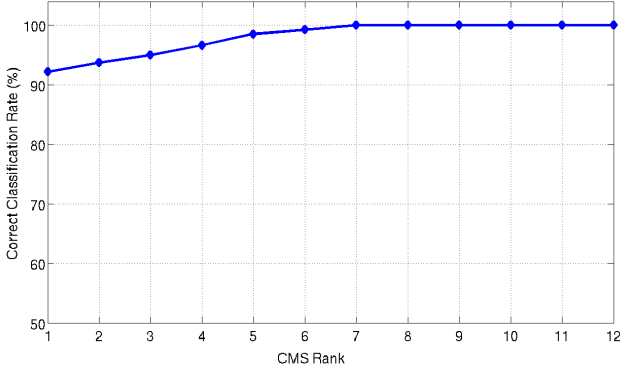


Figure 5: Mean angle's errors on synthetic gait with different image compressions.

Furthermore, a number of experiments are carried out using the same video set to investigate the algorithm potentials for data recorded at low frame rate. The performance error is simulated by dropping a number of frames from every 25 frames (25 is the original frame rate) of the video sequences which is equivalent to changing the frame rates. Table 3 shows the recognition rates for data recorded at different frame rates. It is observed that the algorithm performance is not much affected even when dropping 40% of the frames as the algorithm predicts the joint positions for the missing frames through the use of data fitting.

Video Frame Rate	Recognition Rate
25 fps	92.2%
15 fps	81%

Table 3: Video Frame Rate Effects on the Performance of Gait Recognition .

4. CONCLUSIONS

A smart surveillance system aims at automatically tracking people across a network of cameras and detect abnormal behaviours. In this context, the subject's identification and tracking between the cameras is critical point. Possible solutions found in literature are based on colour or appearance information. However, most approaches to date limit their application to a short time period and environments where subjects does not wear uniform.

In this paper, we have proposed gait analysis as a solution for subjects' identification across a network of cameras. The completely unobtrusiveness without any subject cooperation or contact for data acquisition make gait particularly attractive for identification purposes in camera handover.

A novel 2D markerless view-independent gait analysis algorithm has been presented: the method does not need camera calibration or pre-knowledge of subject pose. Since the choose of the cameras' characteristics is a key-point for the development of a smart surveillance system, the performance of the proposed approach has been measured with respect to different video properties.

Tests on both synthetic and real video sequences allowed to evaluate the performance of the proposed approach with respect to different spatial resolution, frame-rate, data compression and image quality. The obtained results show that gait analysis can be efficiently used for view-independent subjects' identification with commercially available video cameras.

5. ACKNOWLEDGMENTS

The study was supported by SCOVIS: FP 7 ICT project no. 216465 and uses the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

6. REFERENCES

- [1] www.cc.gatech.edu/cpl/projects/hid.
- [2] Casia gait database. online, 2006.
- [3] C. BenAbdelkader, L. S. Davis, and R. Cutler. Motion-based recognition of people in eigengait space. In *Proc IEEE Conf AFG*, pages 267–274, 2002.
- [4] B. Bhanu and J. Han. Human recognition on combining kinematic and stationary features. In *Proc Int Conf AVBPA*, pages 600–608, 2003.
- [5] J. Black, T. Ellis, and P. Rosin. Multi view image surveillance and tracking. *Motion and Video Computing. Proc. Workshop on*, pages 169–174, Dec. 2002.
- [6] I. Bouchrika and M. S. Nixon. *Model-Based Feature Extraction for Gait Analysis and Recognition*, volume 4418 of *LNCS*. Springer, 2007.
- [7] Y. Cai, W. Chen, K. Huang, and T. Tan. Continuously tracking objects across multiple widely separated cameras. In *ACCV (1)*, pages 843–852, 2007.
- [8] T. Chau. A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait Posture*, 13(1):49–66, 2001.
- [9] A. Chilgunde, P. Kumar, S. Ranganath, and H. WeiMin. Multi-camera target tracking in blind regions of cameras with non-overlapping fields of view. In *British Machine Vision Conference*, pages –, 2004.
- [10] R. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, pages 351–356, 2002.
- [11] J. Cutting and L. Kozlowski. Recognizing friends by their walk: gait perception without familiarity cues. *Bull. Psychonom. Soc.*, 9:353–356, 1977.
- [12] W. T. Dempster and G. R. L. Gaughran. Properties of body segments based on size and weight. *Am J Anat*, 120:33–54, 1965.
- [13] R. Fisher. Self-organization of randomly placed sensors. page IV: 146 ff., 2002.
- [14] A. Gilbert and R. Bowden. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. pages 125–136, 2006.

- [15] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon. Markerless view independent gait analysis with self-camera calibration. *Proc. IEEE Conf. on Automatic Face and Gesture Recognition*, 2008.
- [16] M. Goffredo, N. Spencer, D. Pearce, J. N. Carter, and M. S. Nixon. *Human Perambulation as a Self Calibrating Biometric.*, volume 4778 of *LNCS*. Springer, 2007.
- [17] R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2001.
- [18] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.
- [19] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Trans Systems Man Cybernetics*, 34(3):334–352, 2004.
- [20] P. Huang, C. Harris, and M. Nixon. Recognising humans by gait via parametric canonical space. *J Artif Intell Eng*, 13(4):359–366, 1999.
- [21] T. Huang and S. Russell. Object identification in a bayesian context. In *IJCAI97*, pages 1276–1282, 1997.
- [22] A. Ilie and G. Welch. Ensuring color consistency across multiple cameras. *Tenth IEEE Conf. on Computer Vision.*, 2:1268–1275 Vol. 2, Oct. 2005.
- [23] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking across multiple cameras with disjoint views. *Proc. Ninth IEEE Conf. on Computer Vision*, pages 952–957 vol.2, Oct. 2003.
- [24] A. Kale, A. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. *Proc IEEE Conf AVSS*, pages 143–150, 2003.
- [25] S. Kang, J. Min, and J.-K. Paik. Multiple-camera tracking system with seamless object handover. volume 4303, pages 27–34. SPIE, 2001.
- [26] V. Kettner and R. Zabih. Bayesian multi-camera surveillance. *IEEE Conf. on Computer Vision and Pattern Recognition*, 2:–259 Vol. 2, 1999.
- [27] S. Khan and M. Shah. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1355–1360, Oct. 2003.
- [28] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. *Proc. IEEE Computer Vision and Pattern Recognition*, 2:II–205–II–210 Vol.2, June-2 July 2004.
- [29] L. Middleton, D. K. Wagg, A. I. Bazin, J. N. Carter, and M. S. Nixon. Developing a non-intrusive biometric environment. *Proc IEEE Int Conf IROS*, pages 723–728, 2006.
- [30] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *J Bone Joint Surgery*, 46:335, 1964.
- [31] C. Niu and E. Grimson. Recovering non-overlapping network topology using far-field vehicle tracking data. In *ICPR (4)*, pages 944–949, 2006.
- [32] M. Nixon and A. Aguado. *Feature Extraction & Image Processing*. Academic Press, 2nd edition, Dec. 2007.
- [33] M. S. Nixon and J. N. Carter. Automatic recognition by gait. *Proc. of the IEEE*, 94(11):2013–2024, 2006.
- [34] C. Orrite-Urunuela, J. del Rincon, J. Herrero-Jaraba, and G. Rogez. 2d silhouette and 3d skeletal models for human detection and tracking. *Proc IEEE Conf ICPR*, 4:244–247, 2004.
- [35] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. *Proc Int Conf Pattern Recognition*, 1:385–388 vol.1, 2002.
- [36] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [37] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. *Proc. Computer Vision and Pattern Recognition*, 1:1–187–I–194 Vol.1, June-2 July 2004.
- [38] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *Proc Int Conf Recent Advances in Soft Computing*, pages 66–72, 2002.
- [39] S. Stillman and I. Essa. Towards reliable multimodal sensing in aware environments. In *PUI '01: Proc. workshop on Perceptive user interfaces*, pages 1–6, New York, NY, USA, 2001. ACM.
- [40] K. Tieu, G. Dalley, and W. Grimson. Inference of non-overlapping camera network topology by measuring statistical dependence. *Tenth IEEE Conf. on Computer Vision*, 2:1842–1849 Vol. 2, Oct. 2005.
- [41] G. Veres, L. Gordon, J. Carter, and M. Nixon. What image information is important in silhouette-based gait recognition? *Proc IEEE Conf CVPR*, 2:II–776–II–782 Vol.2, 2004.
- [42] L. Wang, T. Tan, W. Hu, and H. Ning. Automatic gait recognition based on statistical shape analysis. *IEEE Trans IP*, 12(9):1120–1131, 2003.
- [43] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans PAMI*, 25(12):1505–1518, 2003.
- [44] Y. Yamada. Advanced method for improvement of obscure video image. *Proc. IEEE 33rd Annual Carnahan Conference on Security Technology*, pages 440–445, 1999.
- [45] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition at sagittal plane. *J Image Vision Comput*, 25(3):321–330, 2007.