Gait as a Biometric for Person Identification in Video Sequences

Chiraz BenAbdelkader

June 5, 2001

Abstract

The term gait recognition *is typically used to signify the identification of individuals in image sequences 'by the way they walk'. There is an increased interest in gait as a biometric, mainly due to its non-intrusive as well as non-concealable nature. Considerable research efforts are being devoted in the computer vision community to characterize and extract gait dynamics automatically from video. The objective is to use gait as a filter (indicator) to effectively enhance the overall recognition performance of a system that uses multiple modalities. In this proposal, we present (describe) two different gait recognition methods; a non-parametric method that uses the self-similarity plot of a walking sequence as the input feature for classification; and a parametric method that estimates the spatiotemporal parameters of gait (the cadence and stride length) and exploits their linear relationship as a cue for identification. Finally, because carried loads are gait-altering, we also present a motion-based method to detect whether a walking person carries an object (load).*

Contents

Chapter 1

Introduction

Recently, gait recognition has received growing interest within the computer vision community, due to its emergent importance as a biometric [10, 4]. The term *gait recognition* is typically used to signify the identification of individuals in image sequences 'by the way they walk'. It can also refer to the recognition of different types of human locomotion, such as running, limping, hopping, etc. The former usage of the term shall be assumed throughout this proposal, however.

A major impetus for gait recognition research lies in psychophysical experiments with Moving Light Displays (MLDs) pioneered by Johansson [27]. Johansson's first experiments demonstrated the ability of human subjects to recognize the type of movement of a person solely from observing the 2D motion pattern generated by light bulbs attached to the person. Similar experiments later showed some indication that even the identity of a familiar person ('a friend') [3], as well as the gender of the person [13], might be recognizable from MLDs.

These experiments not only provided insight about motion perception in the human visual system, they also brought about evidence suggesting that motion patterns generated by the human gait encode information that is characteristic of (and sometimes unique to) the moving person. The goal of gait recognition research is to determine how that information can be extracted from images.

The fact that each person seems to have a distinctive (idiosyncratic) way of walking is hardly surprising from a biomechanics standpoint [52]. Human ambulation consists of synchronized integrated movements of hundreds of muscles and joints in the body. Although these movements follow the same basic pattern for all humans, they seem to vary from one individual to another in certain details such as their relative timing and magnitudes. Much research in biomechanics and clinical gait analysis (among others) is devoted to the study of the inter-person and intra-person variability of gait (albeit not for the purpose of recognition, but rather to determine normal vs. pathological ranges of variation). The major sources of inter-person variability are attributed to physical makeup, such as body mass and lengths of limbs, while the sources for intra-person variability are things like walking surface, footwear, mood and fatigue [26, 52, 40]. Nonetheless, the gait of any one individual is known to be fairly repeatable when walking under the same conditions.

This intra-person consistency and inter-person variability are what makes gait suitable (desirable) for use as a biometric.

Having established that gait as a biometric does have potential, what makes this problem challenging and novel from a computer vision viewpoint, however, is that automatic extraction of gait parameters (i.e. such as joint positions) requires feature tracking, which cannot always be done robustly over long sequences, due to for example occlusions by clothing, changes in lighting, and image noise, all of which are inherent to real imagery of natural scenes. Video-based gait analysis methods rely on markers, wearable instruments or special walking surfaces [40], which are not appropriate for a computer vision approach, since direct participation (or cooperation) of the subject under study cannot be assumed.

In this proposal, we present two gait recognition techniques both of which take a motion-based recognition approach [9], and use correspondence-free image features to extract and characterize gait motion patterns from video. Both techniques are also robust to tracking errors, and changes of lighting and clothing.

In the first method, a sequence of images of a walking person is first mapped to a 2D feature consisting of the matrix of self-similarities between each pair of images in the sequence. We contend that this feature encodes a projection of the planar dynamics of gait, and which can be used for gait recognition much the same way as a face image is used to identify a person. We use principal components analysis (PCA) to reduce the dimensionality of this feature space, and k-nearest neighbor rule for gait classification in the reduced space, termed the *Eigengait*. Because the technique is not view-invariant, we investigate recognition performance as a function of changing camera viewpoint.

The second method takes a parametric approach instead, in that it explicitly estimates actual gait parameters. Specifically, given a calibrated camera and knowledge of the plane of walking, the method exploits the periodicity of human walking to accurately estimate the cadence and stride length (also known as the spatio-temporal parameters of gait). It also exploits the (known) linear functional relationship between cadence and stride length to identify an unknown person in an existing database of people from his/her estimated cadence and stride length.

An important question that arises (or should arise) in any gait recognition algorithm is whether or not it is invariant to gait-altering factors, such as footwear, surface of walking, and load-carrying (i.e. when the person is carrying something). We expect that both techniques are generally not invariant to load-carrying, and we describe a method to detect whether a walking person carries an object (load), so that gait recognition is only attempted under non-load-carrying conditions. This method is view-invariant and uses binary shape and periodicity cues.

The rest of this proposal is organized as follows. In Chapter 2, we present a survey of existing vision methods that are of, or related to, gait recognition, and compare them with our techniques. Chapter 3 describes the proposed non-parametric gait recognition technique, and Chapter 4 the parametric technique. In Chapter 5, we describe our method for determining when a walking person is carrying an object.

Chapter 2

Survey of Related Work

Gait recognition can be generally related to human motion analysis methods (i.e. vision methods that detect, track and/or recognize human movement), and more specifically to methods that deal with whole-body human movement.

2.1 Human Movement Analysis

The extraction and characterization of human movement from video spans several research areas of computer vision, such as gesture recognition, action/activity recognition, lipreading and person identification from gait (or gait recognition). Good comprehensive surveys on this topic are in [9, 1, 17].

Existing methods can be grouped into: (i) structural methods, which recover a structural model of the human body and use this structure for motion recognition [23, 24, 2, 45, 18, 7, 36, 46], and (ii) non-structural methods, which directly model, extract and recognize the motion patterns generated by any particular body movement [53, 39, 37, 30, 25, 22, 12].

2.1.1 Structural Methods

In SFM-based methods, a set of body points are tracked (as a result of body structure recovery), and their motion trajectories are used to characterize, and thereby recognize the motion or action performed by the body. Note that this approach emulates MLD-based motion perception in humans, since the body part trajectories are in fact identical to MLD-type stimuli. Furthermore, this approach is supported by biomedical gait research [38] which found that the dynamics of a certain number of body parts/points totally characterize gait. However, because tracking body parts in 3D over a long period of time remains a challenge in vision, the effectiveness of SFM-based methods remains limited.

A 2D or 3D structural model of the human body is assumed, and body pose is recovered by extracting image features and mapping them to the structural components of the model (i.e. body labelling). Hence a human is detected in the image if there exists a labelling that fits the model well enough (based on some measure of goodness of fit) [23, 24, 45, 18, 46]. Once a person has been detected and tracked in several images, motion recognition is done based on the temporal trajectories of the body parts, typically by mapping them to some low-dimensional feature vector and then applying standard pattern classification techniques [2, 48, 36].

2.1.2 Structure-free Methods

Motion-based recognition methods, on the other hand, characterize the motion pattern of the body, without regard to its underlying structure. Two main approaches exist; one which represents human movement as a sequence (i.e. discrete number) of poses/configurations; and another which characterizes the spatiotemporal distribution generated by the motion in its continuum.

To recognize a moving object (or person), these methods characterize its motion pattern, without regard to its underlying structure. They can be further divided into two main classes. The first class of methods consider the human action or gait to be comprised of a sequence of poses of the moving person, and recognize it by recognizing a sequence of static configurations of the body in each pose [37, 25, 22]. The second class of methods characterizes the spatiotemporal distribution generated by the motion in its continuum, and hence analyze the spatial and temporal dimensions simultaneously [39, 43, 14, 31, 30, 12].

State-space Methods

These methods represent human movement as a sequence of static configurations. Each configuration is recognized by learning the appearance of the body (as a function of its color/texture, shape or motion flow) in the corresponding pose.

Spatiotemporal Methods

Here, the action or motion is characterized via the entire 3D spatiotemporal (XYT) data volume spanned by the moving person in the image. It could for example consist of the sequence of grey-scale images, optical flow images, or binary silhouettes of the person. This volume is hence treated as a 'large' vector, and motion recognition is typically done by mapping this vector to a low-dimensional feature vector, and applying standard pattern classification technique in this space. The following methods describe different ways of doing this.

2.2 Whole-body Movement Analysis

Existing vision methods that analyze whole-body human movement can be classified into (1) gait recognition methods [39, 37, 30, 25, 10, 22], which identify people from their gait (i.e. the 'way they walk'), (2) human detection methods which essentially classify moving objects as human or non-human [53, 12, 46], and (3) human motion classification [8, 36], which recognize different types of human locomotion, such as walking, running, limping, etc.

Niyogi and Adelson [39] extract four silhouette signatures of a moving person, $x_i(y,t), i = 0, 1, 2, 3$, two of which correspond to the outer boundaries of the person, and the other two to the inner edges of each leg. Each signature is normalized via spatial and temporal alignment and scaling (i.e. so that it is stationary in the image and has a fixed height, a fixed period and a fixed phase). These normalized signature defines a spatiotemporal sheet over the entire image sequence. Gait recognition is done by matching these sheets for the model gait and input gait.

Yasutomi and Mori [53] describe a human detection method that computes cadence and stride length based on periodicity of human motion, and classifies the moving object as 'human' based on the likelihood of the computed values in a normal distribution of human walking. Like in our method, they use a calibrated camera to recover the 3D position of the person on a known plane and compute the stride length based on the ground plane distance number of steps travelled. However, while we use cadence and stride length for recognition (i.e. person identification), they use them for human detection.

Murase and Sakai [37] describe a template matching method which uses the parametric eigenspace representation as applied in face recognition [49]. Specifically, they use PCA (Principal Component Analysis) to compute a 16-dimensional manifold for all the possible grey-scale images of a walking person. An input sequence of images (after normalization) is hence mapped to a trajectory in this 16-dimensional feature space, and gait recognition is achieved by computing the distance between the trajectories of the input image sequence and a reference sequence.

Little and Boyd [30] perform a non-parametric method of person identification. They first compute the optical flow of people walking parallel to the image plane, and compute a best-fitting ellipse on the flow. Various statistics of this ellipse are used to identify people. The method is not view invariant, and has not been shown to be robust to lighting or clothing changes.

Huang et al. [25] use a similar technique, as they apply PCA to map the binary silhouette of the moving figure to a low dimensional feature space. The gait of an individual person is represented as a cluster (of silhouettes) in this space, and gait recognition is done by determining if all the input silhouettes belong to this cluster.

He and Debrunner [22] recognize individual gaits via an HMM that uses the quantized vector of Hu moments of a moving person's silhouette as input.

Cutler and Davis [12] describe a method for human detection by recognizing specific periodic patterns in the *similarity plot*, a 2D matrix of all pairwise image matching correlations. They also use the periodicity of these similarity plots to estimate the stride of a walking and running person, assuming a calibrated camera. They contend that stride could be used as a biometric, though they have not conducted any study showing how useful it is as a biometric.

2.3 Carried Object Detection

Haritaoglu's *Backpack* [20] system is the only work we know of that addresses the specific problem of carried object detection for video surveillance applications. Like our method, *Backpack* uses both shape and motion cues. It first locates *outlier regions*, or significantly protruding regions, of the silhouette via static shape analysis that segments silhouette regions that violate the symmetry assumption of the human body. Each outlier region is then classified as being part of the carried object or of the body based on the periodicity of its vertical silhouette profile.

Implicit in this method is the assumption that aperiodic outlier regions correspond to the carried object and periodic regions to the body. However, for this to be true, the following must hold: the person's silhouette is perfectly symmetric (so that not too many non-symmetric regions are detected), and the legs and arms are the only protruding body parts (so that periodic non-symmetric regions truly correspond to body parts). This can often fail for a variety of reasons. For example, the axis of symmetry (which is computed as the blob's major axis) is very sensitive to detection noise, as well as to the size and shape of the carried object itself. Also, using a heuristically-determined threshold to filter out small non-symmetric regions makes this method less robust. One way this method can be more efficient and robust is by constraining the location and number of non-symmetric regions a priori, since the vertical profile of regions other than the legs and arms are generally not periodic.

Like *Backpack*, we use a silhouette signature shape feature to capture the periodicity of the human body. A major difference lies in that we analyze *both* the periodicity and magnitude of these shape features over time to detect the carried object, and only use static shape analysis in the final segmentation phase of the object. Another important difference is that we explicitly constrain the location of the object to be either in the arms region and/or legs region, since as noted above, the silhouette signature of the region above the arms are not periodic.

Chapter 3

Eigengait: a Non-parametric Gait Recognition Method

We know from biomechanics that the dynamics of gait can be fully characterized via the kinematics of a handful of body landmarks such as limbs and joints [26]. The problem with taking this approach in a vision algorithm is that it requires feature tracking (to extract joint positions for example), which typically cannot be done robustly over long image sequences without the use of special markers. Furthermore, we may not necessarily need to extract the complete gait dynamics in order to discriminate different gaits.

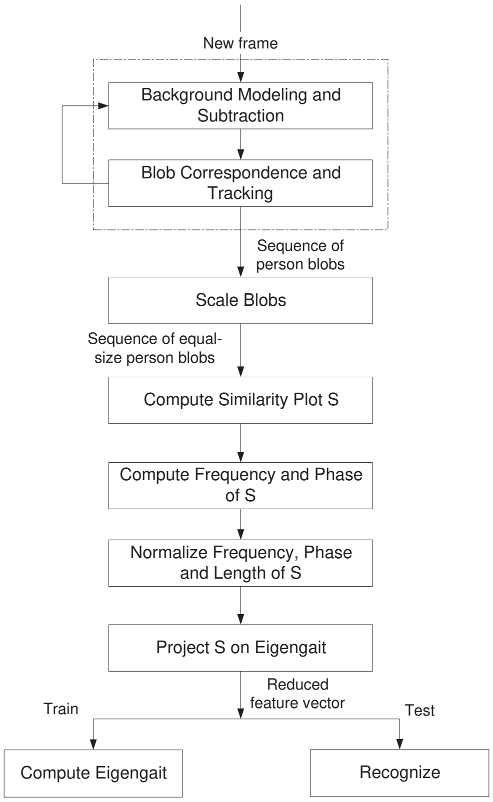


Figure 3.1. Overview of Method.

This chapter describes a motion-based recognition technique that uses a correspondence-free image feature for gait classification. Specifically, it maps a sequence of images of a walking person to a *similarity plot* (SP), the matrix of self-similarities between each pair of images of the person in the sequence. We contend that this 2D feature encodes a projection of the planar dynamics of gait, and hence a signature of gait dynamics. We shall use this signature as the gait biometric of choice.

The proposed method essentially treats a similarity plot much the same way that the Eigenfaces technique [49] treats a face image; it uses Principal components analysis to reduce the dimensionality of the feature space, then applies some supervised pattern classification technique (k-nearest neighbor rule in our case) in the reduced feature space for recognition.

An oveview diagram of the method is shown in Figure 3.1. An input image sequence is first processed to segment the moving person from the background and track him in each frame. The obtained sequence of blobs of the person are then properly aligned and scaled to a uniform size (dimensions), to account for tracking errors, as well as any depth changes that occur in non-frontoparallel walking. The similarity of each pair of these blobs is then computed, to obtain a similarity plot of the person. For recognition, the similarity plot is mapped (projected) to a small feature vector in Eigengait space, which is then used for classification. The Eigengait space vectors are computed in a training phase by applying PCA to the similarity plots of a set of known (labeled) people.

In the sequel, we first present the assumptions of the method, then we describe the method in detail, and finally we present a set of experiments in which we test the method on walking sequences of multiple subjects, taken on different days and from different camera viewpoints.

3.1 Assumptions

The method makes the following assumptions:

- People walk on a known plane with constant velocity for about 3-4 seconds.
- The frame rate is greater than twice the frequency of the walking.
- The camera is static.

3.2 Computing Similarity Plots

Since the camera is assumed to be stationnary, we use background modeling and subtraction [15] to segment moving objects in each frame. To track an object, we use a simple correspondence method based on the overlap of blob bounding boxes in any two consecutive frames [21].

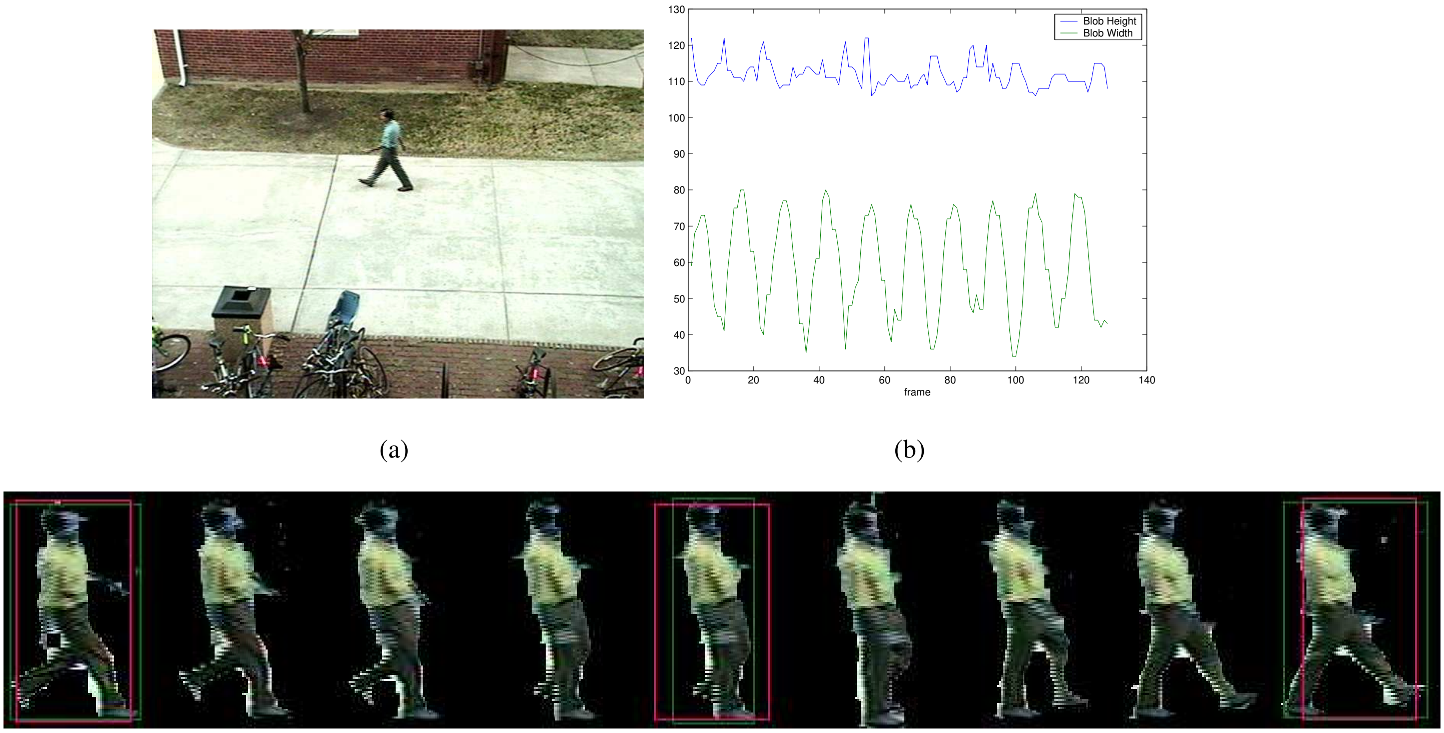
Let $C_{t_1}, C_{t_2}, .., C_{t_N}$ be the N blobs obtained from tracking a person in N consecutive frames. Because of the pendular-like oscillatory motion of the legs and arms, the person's size in the image changes at the frequency of gait, and so these blobs do not have identical sizes (i.e. width and height dimensions).

If we assume fronto-parallel walking, then blob size varies as a stationary process and the average blob size is almost constant, as illustrated by Figure 3.2. Let us denote the average height and width by $\bar{H}$ and $\bar{W}$, respectively. The similarity of any two blobs C_{t_1} and C_{t_2}, for all $1 \leq t_1, t_2 \leq N$, is then computed as follows:

$$S_{t_1,t_2} = \min_{|dx,dy|<r} \sum |C_{t_1}(x+dx, y+dy) - C_{t_2}(x', y')|. \tag{3.1}$$

where B_{t_1} and B_{t_2} are equal rectangular regions of height $\bar{H}$ and width $\bar{W}$, centered at the centroids of C_{t_1} and C_{t_2}, respectively; (x, y) and (x', y') are corresponding pixels in B_{t_1} and B_{t_2} respectively; and r is a small search radius that accounts for small tracking errors. This is illustrated by Figure 3.2(c); green boxes correspond to the actual bounding box of a blob, and red boxes correspond to the average box B_t used to compute the similarity plot.

If the person is not walking fronto-parallel, however, then the blobs need to be scaled to the same dimensions, before computing their similarities. This is because the average blob size changes (increases or decreases) linearly. Figure 3.3 and Figure 3.4 illustrates two examples in which the blob dimensions oscillate with an increasing and decreasing trend, respectively. Hence, the dimensions of the blobs are analyzed, and if a linear trend is detected, then they are all scaled down to the size of the smallest blob using a Mitchell filter [11]. Note that it is also possible to scale each pair of blobs separately, i.e. compute $S(t_1, t_2)$ by scaling blob C_{t_1} to the size of blob C_{t_2}.

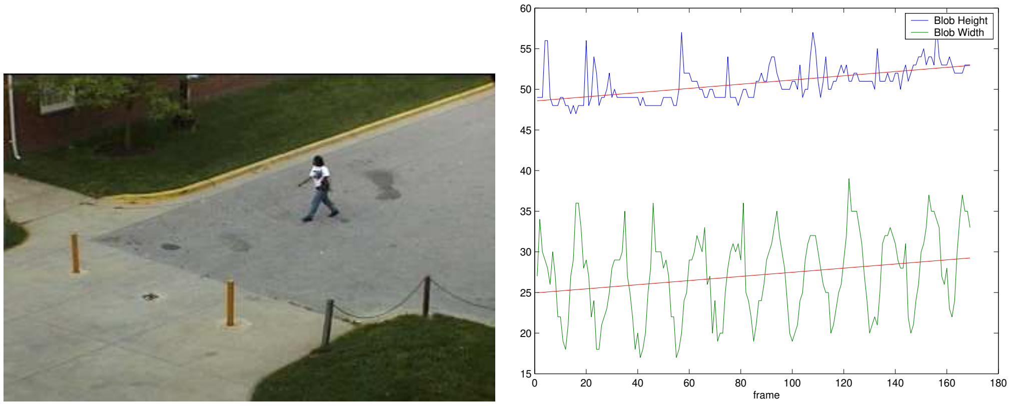


(c)

Figure 3.2. (a) One frame of a fronto-parallel walking sequence. (b) Dimensions (width and height) of the person blob oscillate as a stationary process. (c) First few blobs of person segmented from background; green boxes are actual blob bounding boxes; red boxes bound blob area used to compute similarity plot.

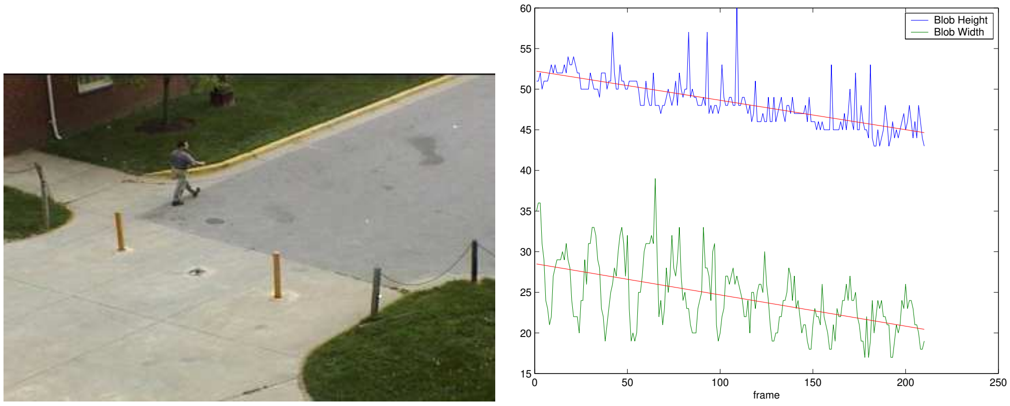
Note that the blob similarity measure in Equation 3.1 can be applied to any one of:

- Binary silhouettes.
- Grey (color) silhouettes, without background.



(a) (b)

Figure 3.3. (a) First frame of a sequence where person walks non-fronto-parallel, *closer* **to the camera. (b) Dimensions of corresponding blob oscillate with an** *increasing linear trend* **(indicated by the red line).**



(a) (b)

Figure 3.4. (a) First frame of a sequence where person walks non-fronto-parallel, *further* **from the camera. (b) Dimensions of corresponding blob oscillate with an** *decreasing linear trend* **(indicated by the red line).**

- Grey (color) silhouettes, with background.

though each has its own merits and drawbacks, and it is not directly obvious which is best to use. For example, 1 and 2 are sensitive to segmentation errors; 2 and 3 are not invariant to change of person's clothing and scene lighting; 3 is not invariant to change in the background scene.

3.2.1 Properties

The similarity plot, S, of a walking person has the following properties:

1. $S(t, t) = 0$, i.e. it has a dark main diagonal.

2. $S(t_1, t_2) = S(t_2, t_1)$, i.e. it is symmetric along the main diagonal.

3. $S(t_1, kp/2 + t_1) \simeq 0$, i.e. it has dark lines parallel to the main diagonal (the off-diagonals).

4. $S(t_1, kp/2 - t_1) \simeq 0$, i.e. it has dark lines perpendicular to the main diagonal (the cross-diagonals).

where $t_1, t_2 \in [1, N]$, p is the period of walking, and k is an integer. The first two properties are generally true for any similarity function (though the second property may not hold if substantial image shifting and scaling are required). The latter two are a direct consequence of the periodicity and the bilateral symmetry, respectively, of the human gait.

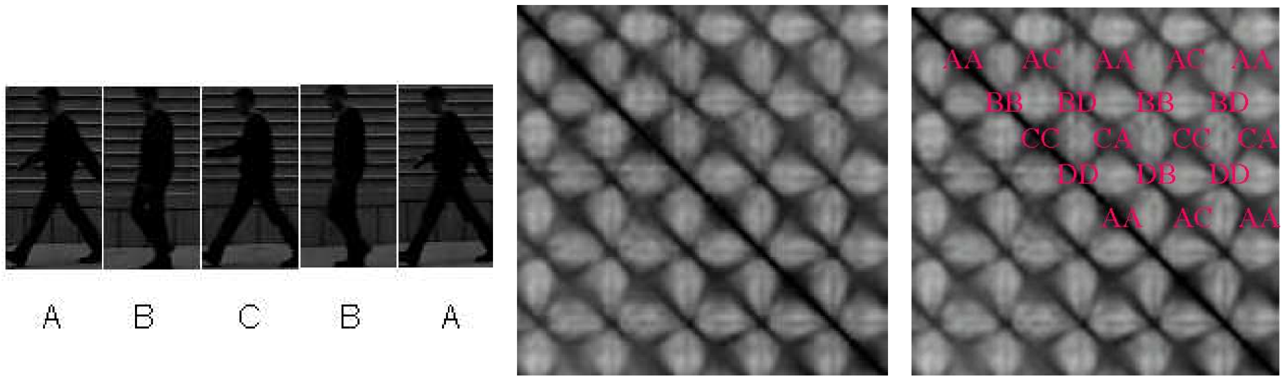
Furthermore, the intersections of the off-diagonals and cross-diagonals encode the frequency and phase of walking [12]. Specifically, each intersection corresponds to combination of two of the four key poses of the gait: (i) when the two legs are furthest apart and the left leg is leading, (ii) when the two legs are joined together and the right leg is leading, (iii) when the two legs are furthest apart and the left leg is leading, and (iv) when the two legs are joined together and the left leg is leading, as illustrated by Figure 3.5. These poses shall be denoted A, B, C, and D, respectively. Due to the bilateral symmetry of human walking, poses A and C, and poses B and D are very similar in appearance when the person is walking fronto-parallel to the camera. However, as the camera viewpoint deviates away from fronto-parallel, the similarity between B and D decreases rapidly to zero, while the similarity between A and C generally only decreases to a small non-zero value.

Note that, since these intersections correspond to the local minima of S, the frequency and phase of gait can hence be automatically estimated by finding the minima of S, as we shall explain later.

The reason that S encodes the frequency and phase of gait may be explained by the fact that it is (approximately) a projection of the planar dynamics of the walking person when viewed sufficiently far from the camera, as argued in [12]. Intuitively, this is because S is obtained via a sequence of transformations (image projection and correlation matching) applied to the set of 3D points on the person's body. It can be shown that these transformations preserve certain properties of the dynamics of these points (and hence of the gait dynamics).

3.3 Gait Classifier

As mentioned in the previous section, the similarity plot is a projection of the dynamics of the walking person that preserves the frequency and phase of the gait. The question then arises as to whether this projection preserves more detailed (higher-dimensional) aspects of gait dynamics, that capture the unique



(a) (b) (c)

Figure 3.5. (a) Key poses in one cycle of a walking gait. (b) Similarity plot S of the corresponding sequence. (c) Intersections of dark lines in S correspond to combinations of key poses.

way a person walks. In other words, does a similarity plot contain sufficient information to distinguish (not necessarily uniquely) the walking gaits of different people?

To evaluate the usefulness of the self-similarity plot in characterizing and recognizing individual gaits, we propose to build a gait pattern classifier that takes an SP (self-similarity plot) as the input feature vector. For this, we take an 'eigenface' approach [49], in which we treat a similarity plot the same way that a face image is used in a face recognizer. The gist of this approach is that it extracts 'relevant information' from input feature vectors (face images or SPs) by finding the principal components of the distribution of the feature space, then applies standard pattern classification of new feature vectors in the lower-dimensional space spanned by the principal components. We use a simple non-parametric pattern classification technique for recognition. In the following, we explain the details of the proposed gait classifier.

3.3.1 Input Variability and Normalization

In any feature classifier, it is important to identify the sources of variation in the input feature of choice. Unwanted sources of variation are those that are not relevant to the classification, and hence should be detected and removed or normalized prior to classification.

Obviously the similarity plot of the same walking person will be different if any of the following are varied:

- Lighting.
- Clothing.
- Number of pixels on target.
- Camera viewpoint.
- Cadence (frequency of gait).

- Phase of walking (body pose at beginning of sequence).
- Length of image sequence.

In the sequel, we discuss how each of these variations in the input image sequence can be normalized or otherwise dealt with when normalization is not possible.

Lighting and Clothing

Variations in clothing and lighting can both be normalized by a using a color-invariant image similarity measure to compute S, such as by applying Equation 3.1 to binary blobs instead of grey (or color) blobs, or by using chamfer matching on edge maps of the person. The former is sensitive to segmentation errors, and the latter may not be a computationally efficient solution. Another method is to lower the resolution of the color blobs by scaling them down, so that most of their color detail is lost. The question that arises here is how much can image resolution be decreased without losing motion information of gait.

Number-of-Pixels-on-Target

The number of pixels-on-target (POT) can be defined as the height of the person in the image grabber (assuming the person appears upright in the image). Assuming fixed image resolution and fixed camera viewpoint, POT can vary due to change in camera depth. To normalize POT for such variations, we can scale *down* the blobs to some fixed size h. Since normalization is not possible when the blob sizes are smaller than h, we should choose h to be small enough so that blobs can generally be scaled down to this size.

Camera Viewpoint

Normalizing for variation in camera viewpoint is not possible except for very small changes. This is because, inherently, a different (planar) projection of gait dynamics is captured in the image plane from any one camera viewpoint.

Hence it is necessary to index our gait recognition method by different ranges of camera viewpoint. The camera viewpoint is defined by the pan and tilt angles of the 3D camera ray passing through the walking person in the sequence (with respect to the camera frame of reference). Hence we can for example define the viewpoint ranges at equal $15 \deg$-intervals of the pan and tilt angles.

Phase, Cadence and Sequence Length

In order to account for different walking paces, starting poses, and length sequences, and assuming these are irrelevant differences, we can normalize the similarity plots so that they start at the same phase, have the same frequency, and contain the same number of cycles. This way, we obtain similarity plots of the same size.

We estimate the fundamental period of S by computing the average 1D power spectrum of its columns, and determining its smallest significant peak [12].

The phase of S can be determined by finding the intersections of its off-diagonals and cross-diagonals, since these correspond to combinations of the key poses A, B, C, and D, as discussed in Section 3.2.1.

Since these intersections occur at local minima of S, then we simply find the first local minimum that occurs on the diagonal, then determine to which pose combination it corresponds to, i.e. AA, BB, CC, or DD (See Figure 3.5). For this, we observe (without proof) that the local minima corresponding to AA and CC tend to be 'flatter' than those that correspond to BB and DD. Hence, we can determine whether the first peak is one of AA, CC or is one of BB, DD. Hence, we can determine phase of S up to (modulo) a half-period. We are unable to resolve the remaining two-way ambiguity between AA and CC, and between BB and DD (which is a result of bilateral symmetry of human gait). However, it might be resolved using other cues, such as shape and direction of motion.

However, gait dynamics are in fact not invariant to large changes in cadence. Hence it is necessary to index for different cadences as well. For now, we shall divide cadence into three ranges; slow, medium and fast. The biomechanics literature shows that any normal individual has a *customary walking speed* (CWS) (also called natural cadence) which is typically in the range $90 - 130$ steps/min [40]. Hence, we can for example classify cadences within this range as *medium*, and cadences outside this range as *slow* or *fast*.

3.3.2 Training the Classifier

Let $S'_1, S'_1, .., S'_M$ be a given training set of M labelled (i.e. corresponding to a known person) normalized similarity plots, of size NxN each, and let s'_i be the vector of length N^2 corresponding to the ith similarity plot S'_i (obtained by concatenating all its rows). We compute the principal components [28] of the space spanned by $s'_1, .., s'_M$ by computing the eigenvalue decomposition (also called Karhunen-Loeve expansion) of their covariance matrix:

$$C_s = \frac{1}{M} \sum_{i=1}^{M} (s'_i - \bar{s'_i})(s'_i - \bar{s'_i})^T$$

where $\bar{s'}$ is the simple mean of all training vectors $s'_1, .., s'_M$. This can be efficiently computed in $O(M)$ time (instead of the brute force $O(N^2)$) [49].

We then consider the space spanned by the n most significant eigenvectors, $u_1, .., u_n$, that account for 90% of the variation in the training SPs [1]. We denote this space the *Eigengait*. Hence each training vector s'_i can be sufficiently approximated by a n-dimensional vector w_i obtained by projecting it onto the Eigengait, i.e. $w_i \equiv \sum_{j=1}^{n} u_j^T . s'_i$. Furthermore, assuming that the training vectors are representative of the variation in the entire feature space, then any new feature vector can be similarly approximated by a point in Eigengait space.

3.3.3 Classification

Gait recognition now reduces to a standard pattern classification in a n-dimensional Eigengait space. The advantage of doing pattern classification in this space is not only that n is typically much smaller than N^2 and M, but also that it contains less unwanted variation (i.e. random noise)[2] and hence provides better separability of the feature vectors, or SPs.

[1] According to the theory of PCA, if $\lambda_1, .., \lambda_n$ are the n largest eigenvalues, then the space spanned by their corresponding eigenvectors account for $\sum_{i=1}^{i=n} \lambda_i / trace(C_s)$ of the total variation in the original feature vectors.

[2] Assuming data variation is much larger than noise variation.

Given a new SP (corresponding to an unknown person), the procedure for recognizing it is to first convert it to a N^2-vector, map it to a point in Eigengait, find the k closest training points to it, then decide its class (or label) via the *k-nearest neighbor rule* [5, 44].

3.4 Experiments

We test our method on two different data sets, and use the leave-one-out cross-validation to obtain a statistically accurate estimate of the recognition rate [51, 44].

3.4.1 Little and Boyd Dataset

The first data set is the same used by Little and Boyd in [30], which consists of 42 image sequences with six different subjects (shown together in Figure 3.6 overlaid on the background image) and 7 samples each. Since the camera is static we used median filtering to recover the background image. Templates of the moving person are extracted from each image by computing the difference of the image and the background and subsequently applying a threshold as well as morphological operations to clean up noise. The self-similarity plots are computed for each of each sequence via absolute correlation, and normalized such that they all contained 4 gait cycles starting on the same phase, and are of size 64x64. Figure 3.7 shows examples of these normalized similarity plots, where each column of three plots corresponds to one person. The recognition rates using the k-nearest neighbor classifier are given in Table 3.1.

Figure 3.6. The six people contained in the test sequences, overlaid on the background image.

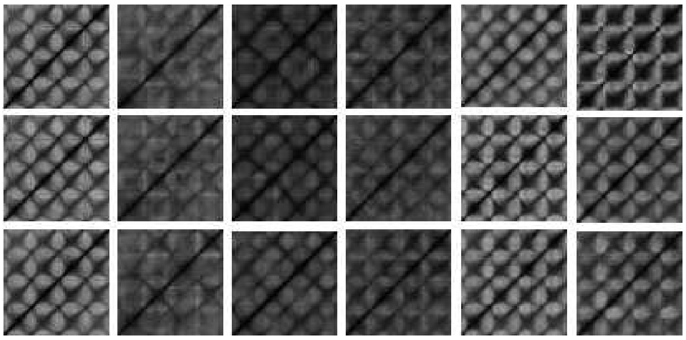


Figure 3.7. Normalized self-similarity plots (columns correspond to a single person).

K_n	Rate
1	0.80
3	0.825
5	0.90

Table 3.1. Classification rates for various K_n values. The classification for a random guess is also given.

3.4.2 Multiview Dataset

A more viable assessment of our gait classifier is by using statistically independent samples of each person. Sequences of one person taken on the same day are not quite independent samples. Sequences taken on different days not only provide useful information about variation of the person's gait. However, they also introduce unwanted variation such as that caused by different color and style of clothing.

We aim to assess the effectiveness of different similarity measures in computing the similarity plots to deal with variations of clothing and lighting. We also aim to evaluate the recognition performance for different camera viewpoints.

To this end, we test the technique on a database consisting of 400 sequences of 7 people (3 females and 4 males) walking on a treadmill, taken on 7 different days and captured simultaneously from 8 different cameras. An average of 56 sequences is provided for each subject. The multiple viewpoints correspond to different pan angles of the camera that are at 15 degree intervals and span a range of about 120 degrees of the camera field of regard. Figure 3.8 illustrates the eight camera viewpoints used in this experiment. The data sequences were captured in the multi-perspective lab [6] at a frame of 60 fps and using greyscale 644x488 images.

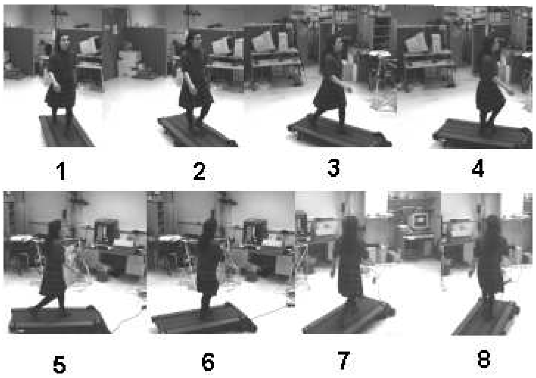


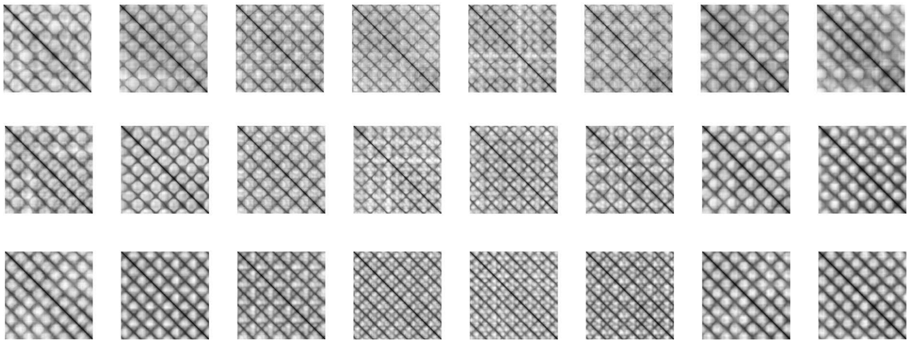
Figure 3.8. Eight camera viewpoints of the sequences in second test data set.

Figure 3.9 shows the similarity plots corresponding to 24 sequences of one subject taken on three different days and from all 8 camera viewpoints.

The subjects have been instructed to walk at their natural medium pace, so that no cadence indexing is needed. However, we used three different similarity measures to compute the similarity plots: (1) correlation of binary blobs, (2) correlation of color blobs without background, and (3) correlation of color blobs with background. Hence, we built and tested the classifier (as described in previous section) separately for each of these similarity measures.

Figure 3.10 shows the classification results for each as a function of three different k-nearest neighbor classifier parameters ($K_n = 1, 3$), the 8 camera viewpoints, and three similarity measures used to compute the SPs.

(a)



(b)

Figure 3.9. (a) Person walking on treadmill as seen from one camera viewpoint on 3 different days (from left to right). (b) Similarity plots of same person in (a) for 3 different days (rows) and 8 camera viewpoints (columns).

3.5 Summary and Discussion

We have used a correspondence-free motion-based method to recognize the gaits of a small population (7) of people, using sequences captured from 8 different viewpoints and taken on different days. The best recognition result (65%) was achieved using correlation on foreground images from a near-frontoparallel viewpoint. This result is 4.6 times better than a random classifier (16.7%).

We also performed the same classification on an existing dataset with 6 people, taken from a single angle on the same day, with no variation on lighting or clothing. The best recognition rate (90%) was achieved using $K_n = 5$, and is 5.3 times better than a random classifier (14.2%).

This method is view dependent, and performs best when frontoparallel images are used. Clothing, lighting, and other variations may degrade the performance of the classifier, though not significantly.

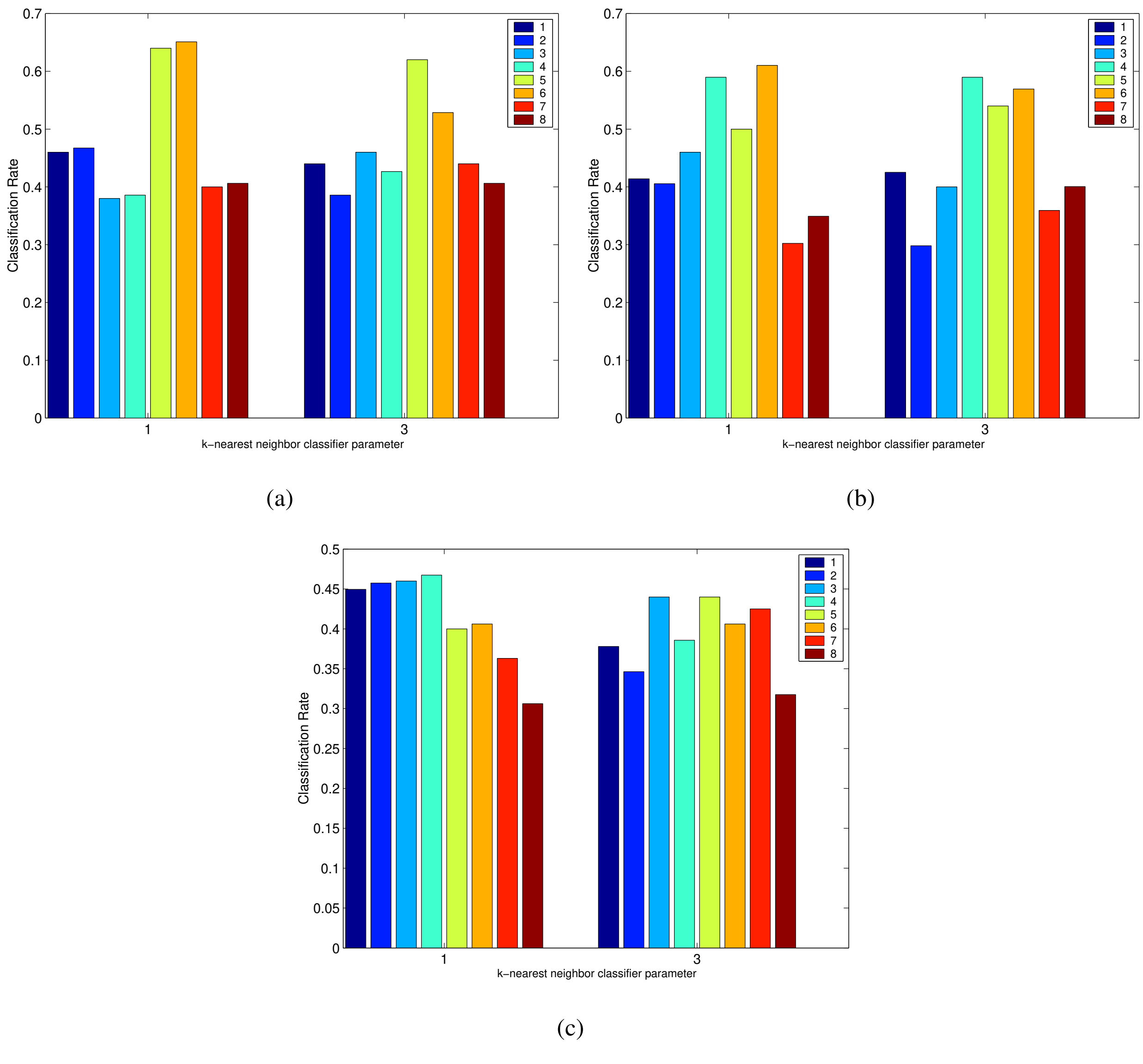


Figure 3.10. Classification rate for the 8 viewpoints when using (a) Correlation of binary images (BC). (b) Correlation of foreground greyscale images (FC). (c) Correlation of greyscale images (GC).

Ang	K_n	BC Rate	FC Rate	GC Rate
15	1	0.365	0.460	0.454
35	1	0.447	0.467	0.467
55	1	0.460	0.380	0.460
75	1	0.589	0.385	0.467
95	1	0.500	0.640	0.400
115	1	0.610	0.651	0.406
135	1	0.357	0.400	0.383
155	1	0.356	0.406	0.328
15	3	0.415	0.440	0.417
35	3	0.319	0.385	0.320
55	3	0.400	0.460	0.440
75	3	0.589	0.426	0.385
95	3	0.540	0.620	0.440
115	3	0.569	0.528	0.406
135	3	0.353	0.440	0.363
155	3	0.355	0.406	0.316

Table 3.2. Classification rates for binary correlation (BC), forground image correlation (FC), and grey image correlation (GC) at various viewing angles (in degrees).

Chapter 4

Parametric Gait Recognition using Cadence and Stride Length

This chapter describes a gait recognition technique which, like the Eigengait method of Chapter 3, takes a motion-based approach in that it extracts correspondence-free motion features to characterize gait from images. However, it is different in that it estimates explicit 3D parameters of gait (and is hence deemed a parametric approach). It exploits the periodicity of human walking to estimate the cadence and stride length of gait, also known as the spatio-temporal parameters of gait [52]. The cadence is estimated using the periodicity of a walking person. Then, using a calibrated camera system, the stride length is estimated by first tracking the person and estimating their distance and number of steps traveled over a period of time.

For a typical outdoor surveillance configuration (with certain assumptions), we are able to estimate the stride length to within 1cm. An error analysis is given in Appendix A.

The stride length and cadence are known to vary linearly for any one person over his/her range of natural walking speeds[1], and because they are also a function of the physical characteristics of the person (most notably leg length and body weight), we develop a parametric gait classifier which takes cadence and stride length as the input features.

This method is view invariant, and robust to changes in lighting, clothing, and tracking errors.

4.1 Assumptions

This technique makes the following assumptions:

- People walk on a known plane with constant velocity for about 3-4 seconds.
- The camera is calibrated.
- The frame rate is greater than twice the frequency of the walking.

[1]Natural cadence corresponds to when a person walks as naturally as possible, and is typically in the range 100-120 steps/minute.

4.2 Overview

The algorithm for gait recognition via cadence and stride length consists of three main modules, as shown in Figure 4.1. The first module tracks the walking person in each frame, extracts a 1D signature of his silhouette shape, and estimates his 3D position on the ground plane. The second module computes the time-distance parameters of gait; it computes the cadence (or frequency of gait) via spectral analysis of the extracted shape feature over time, and deduces the stride length as a ratio of the total distance and the number of cycles traveled. Finally, the third module determines the likelihood (in a probabilistic sense) of the person being any one of the people in an existing database, based on his cadence and stride length thus measured from the video sequence.

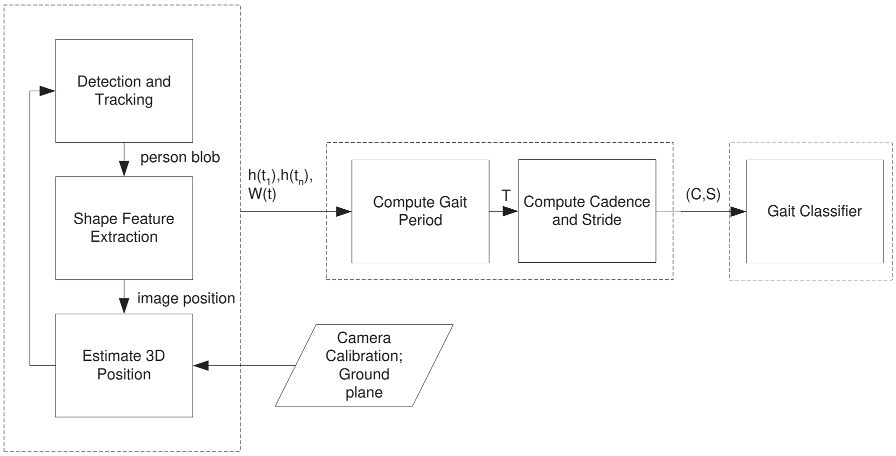


Figure 4.1. Overview of Method.

4.3 Tracking and Feature Extraction

Since the camera is static, we use a non-parametric background modeling technique for foreground detection, which is well suited for outdoor scenes where the background is often not perfectly static (e.g., for occasional movement of tree leaves and grass) [15]. We apply a number of standard morphological cleaning operations to the background-subtracted binary image and use a two-pass algorithm to find connected components (or blobs). Each foreground blob is then tracked in subsequent frames by finding the set of new blobs that overlap it in the current frame. Once a blob has been tracked for a sufficient number of frames, it is classified as *person* or *non-person* based simply on the aspect ratio of its bounding box.

For each *person* blob, we compute two image features that will be used in the next module to estimate the cadence and stride length, namely the width of the blob's bounding box and its lowest leftmost pixel. These correspond to the width of the person and the position of the back heel in the image, respectively.

Assuming the person is moving with constant velocity on a known plane and the camera is calibrated, the 3D position of the back heel point can then be determined as follows. Let K and E be the camera intrinsic and extrinsic matrices respectively, and let $P : AX + BY + CZ + D = 0$ be the parametric equation of the plane of motion, in some defined world frame. Furthermore, we assume perspective projection and hence the camera equations:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = K \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} E \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Therefore, if (X, Y, Z) is any point on plane P, and (x, y) is its projection (pixel coordinates) on the image plane, then we have:

$$\begin{pmatrix} k_{11} & 0 & -x + k_{13} \\ 0 & k_{22} & -y + k_{23} \\ A' & B' & C' \end{pmatrix} E \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -D' \end{pmatrix}$$

where $\begin{pmatrix} A' & B' & C' & D' \end{pmatrix} \equiv \begin{pmatrix} A & B & C & D \end{pmatrix} E^{-1}$ and k_{ij} is the (i, j)th element of K. Hence, given the image position of the back heel point (x, y), its 3D position (X_h, Y_h, Z_h) can be determined via the above linear system of 3 equations and 3 unknowns X_h, Y_h and Z_h.

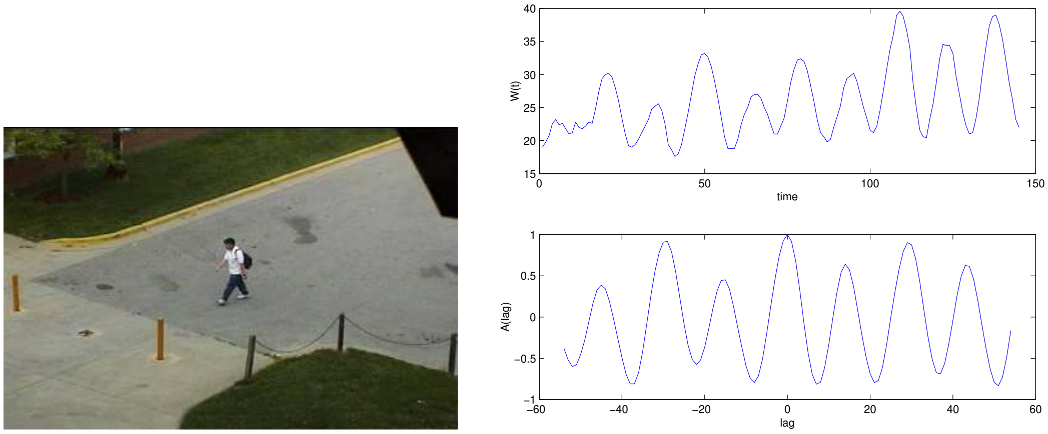
4.4 Estimation of Cadence and Stride Length

Once the person has been tracked for a certain number of frames, n, its spatio-temporal gait parameters, namely the cadence and stride length, are estimated from its 3D trajectory and width time series computed above.

Human gait is a rhythmic (repetitive) phenomenon. Hence the appearance of a moving person viewed from a camera is itself periodic. Existing vision methods that compute the period of gait typically extract a 1D or 2D feature from the image then analyze the spectral properties of the obtained 1D or 2D time series. For example Polana and Nelson [43] use *reference curves*, each of which is a 1D signal of the intensity values of one pixel in the person's image over time; Cutler and Davis [12] compute the *self-similarity plot*, a 2D matrix of the absolute correlation between each pair images of the person in the sequence; and Haritaoglu et al. [20] use the autocorrelation function of the vertical profile of the contour silhouette.

In this method, we simply use the width of the bounding box of the corresponding blob region, henceforth denoted $W(t)$, not only for computational efficiency but also because it has proven to work well with our background subtraction algorithm (which is fairly robust). Niyogi and Adelson [39] also extract this feature from an image sequence and analyze the corresponding spatiotemporal patterns for the detection and recognition of a walking person.

One method for estimating the period of a time series that accounts for colored (non-white) noise, is to use its autocorrelation [12]. To estimate the period of $W(t)$, T_W, we first detrend it (to remove the linear trend if any), smooth it with a symmetric average filter of radius 2, then compute its autocorrelation, $A(i), i \in [-lag, lag]$ where lag is chosen such that it is much larger than the expected period of $W(t)$. The peaks of $A(i)$ correspond to when $W(t)$ is maximally self-similar. However, due to the bilateral symmetry of the human gait (which makes the pose when the left leg is leading similar, though not maximally, to the pose when the right leg is leading), $A(i)$ will sometimes have minor peaks half way between every two major peaks. The strength of these minor peaks diminishes the more the camera viewpoint deviates from fronto-parallel.



(a) (b)

Figure 4.2. Autocorrelation function of the width series of a (nearly) fronto-parallel sequence. It contains major and minor peaks.

Figure 4.2 shows $W(t)$ and the corresponding $A(i)$ for a (almost) frontoparallel sequence. Note that $A(i)$ has both minor and major peaks. Figure 4.3 shows $W(t)$ and the corresponding $A(i)$ for a non-frontoparallel sequence. Note here the linear increasing trend in $W(t)$ due to the changing camera depth. Note also that $A(i)$ does not have any minor peaks.

Hence, we estimate T_W as the average distance between every two consecutive *major* peaks in $A(i)$. The gait period T is then given by T_W, and the cadence and stride length respectively by [40] :

$$
\begin{aligned}
C &= \frac{F_s \cdot 60 \cdot 2}{T} \; (steps/min) \\
S &= \frac{\parallel h(t_n) - h(t_1) \parallel}{n \cdot T} \; (meters)
\end{aligned}
$$

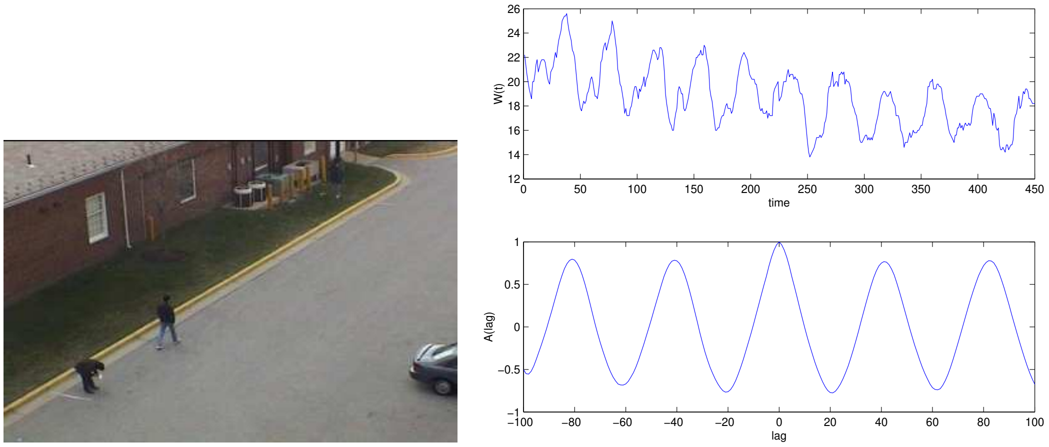
where $h(t_1)$ and $h(t_n)$ are the 3D coordinates of the person's back heel point in the first and nth frames, respectively, and F_s is the frame capture rate.

An analysis, of the estimation error in stride length, based on a typical outdoor setting, is given in Appendix A, and shows an error of about 1cm.

4.5 Person Identification

Here the goal is to build a supervised classifier that uses the cadence and stride length as the input features. While we believe stride length and cadence cannot uniquely identify a person, just as height and hair color do not, building such a classifier helps us to assess the 'separation power' of these features, and hence their usefulness in an actual authentication system that would use several modalities (biometrics) to identify a person.

We use two different parametric classification techniques whereby, given a set of m labeled samples (i.e. cadence and stride length value-pairs of known people), $(C_1, S_1), .., (C_m, S_m)$, the classifier is



(a) (b)

Figure 4.3. Autocorrelation function of the width series of a non-fronto-parallel sequence containing no minor peaks.

trained by fitting the points of each class (person) c to the parameters θ_c of a certain model. Classification of a new sample is then achieved by maximizing the likelihood of that sample given the class model.

4.5.1 Linear Regression Model

Since the stride length and cadence are known to vary linearly for each person within the range of his/her natural walking speeds [19, 52], we use a linear regression model of the form: $S = a \cdot C + b + \omega$, where ω is random noise. We estimate the model parameters $\hat{a}_p$ and $\hat{b}_p$ for each class p in the training set using the linear least squares method, which is known to minimize the square sum of the residuals $r_i = S_i - aC_i - b$ (i.e. vertical distance between fitted line and the sample points) and is optimal when the model noise ω_p is a white noise. The variance of ω_p is estimated using the sample variance, $\hat{\sigma}_{w_p}$, of these residuals.

Given a new sample (C', S') of an unknown person, we compute the probability α_p that it belongs to class p in the database as the likelihood (also called p-value), α_p, of the corresponding residual $r \equiv S' - aC' - b$ with respect to the distribution of ω_p.

4.5.2 Bivariate Gaussian Model

A simpler way to model the relationship between cadence and stride length is as a bivariate Gaussian distribution. Hence the model is trained simply by computing the sample mean, $\hat{\mu}_p$ and sample covariance matrix, $\hat{\Sigma}_p$, of the model from the training samples of each class p. The likelihood (p-value) of a new sample, $u \equiv (C, S)$, with respect to each class p is then computed.

An obvious disadvantage of this model is that, unlike the linear regression model, it cannot extrapolate to values outside the range of its training samples (since they inevitably lie at the tails of the probability distribution). This problem can be avoided if we ensure the training samples indeed span the entire range of natural walking cadences.

4.6 Results

The technique is tested on a database of 131 sequences, consisting of 17 people with an average 10 samples each. The subjects were videotaped with a Sony DCR-VX700 digital camcorder in a typical outdoor setting, while walking at various cadences (paces). Each subject was instructed to walk on a straight line at a fixed speed a distance of about 90 feet (30 meters). Figure 4.4 shows a typical trajectory walked by each person in the experiment. The same camera field of view was used for all subjects.

The video sequences were captured at 30 fps with an image size of 360x240. We used the technique described above to automatically compute the stride length and cadence for each sample sequence. The results are plotted in Figure 4.5, where each subject is shown with a different line label (only 10 of the 17 subjects are shown here for the sake of clarity).

In order to test our gait classifier, we then use these 131 cadence and stride length samples and the leave-one-out cross-validation technique [51, 44] to obtain a statistically accurate estimate of the recognition rate. Specifically, we train the classifier using all but one of the samples, and test it on the sample missed (or left out). This process is hence repeated 131 times, leaving out each of the 131 samples in turn. The classification result of any one test sample consists of a ranking, $(p_{i_1}, .., p_{i_{17}})$, of all 17 persons in the database, where p_{i_1} is the top match and $p_{i_{17}}$ is the least favorable match. The overall classification rate is then obtained as the fraction of test samples (out of the total 131) for which the correct person is found to be the top match.

A more general way of describing the performance of a classifier is via the FERET evaluation methodology (originally developed for evaluating face-recognition algorithms) [41]. Specifically, for each $k = 1, .., 17$, we measure the probability, $\lambda(k)$, that the correct person (or class) is *within* the top k matches for a test sample. Hence the classification rate is equal to $\lambda(1)$. Using the leave-one-out cross-validation procedure, we compute $\lambda(k)$ for each k as the fraction of test samples for which the correct person label is in the subset $(p_{i_1}, ..., p_{i_k})$.

Figure 4.6 shows $\lambda(k)$ for the two different classifier models as well as for the 'chance' classifier. A chance classifier corresponds to when the ranking $(p_{i_1}, .., p_{i_{17}})$ is a random permutation of the the 17 persons, i.e. $\lambda(k) = k/17$.

4.7 Summary and Discussion

We have presented a parametric method for person identification by estimating and classifying their stride and cadence. This approach is view invariant, and robust to changes in lighting, clothing, and tracking errors. It achieves its accuracy by exploiting the nature of human walking, and computing the stride and cadence over many steps.

The classification results are promising, and are over 7 times better than chance for the k=1, linear regression classifier. The linear regression classification can be improved by limiting the extrapolation distance for each person, perhaps using supervised knowledge of the walking speeds of each person.

Further improvements can be achieved by using higher resolution cameras (here we used only 360x240), and by improving the tracking precision.

Figure 4.4. Typical trajectory walked by each subject. Red dots correspond to repeating poses in the gait cycle.

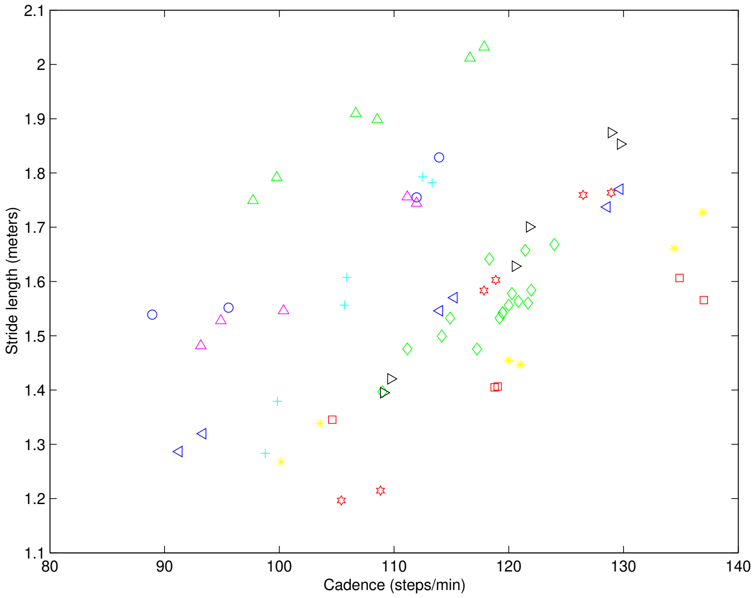


Figure 4.5. Stride length vs. Cadence for 10 of subjects. Note that points of one person are approximately on a line.

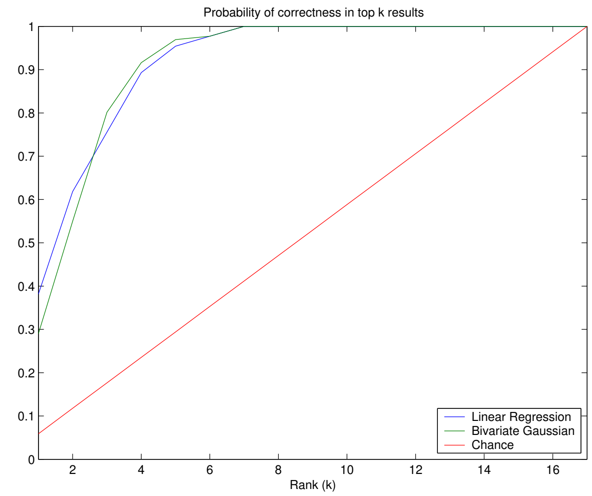


Figure 4.6. Classification results according to the FERET protocol.

Chapter 5

Detection of Load-carrying People for Gait and Activity Recognition

5.1 Motivation

The detection of whether a walking person is carries an object is of interest in gait recognition because carried loads are considered a gait-altering factor (i.e. they alter the the dynamics of walking). Moreover, some gait recognition algorithms are appearance-based, and hence the presence of a large carried object that distorts the silhouette shape of the person is very likely to 'break' such algorithms. Thus, it is essential to determine whether a person is carrying an object before attempting gait recognition.

Carried object detection is also of interest to human activity recognition. In many surveillance applications, an important class of human activities are those involving interactions of people with objects in the scene, which include depositing an object, picking up an object, and the exchange of an object between two people. Given the time intervals during which objects are carried by any one person, we would expect that a temporal logical reasoning system will be able to infer events of object pickup, object deposit and object exchange.

The clinical gait analysis and ergonomics research communities have also been interested in assessing the effect of load-carrying on the human gait, for applications that include the design of 'ergonomically' safe backpacks for army infantrymen and recreational hikers [29, 34]. Their studies typically analyzed the spatio-temporal, kinematic and kinetic parameters of gait as a function of the amount of carried load as well as the manner by which it is carried. According to these studies, people carrying a (heavy) object adjust the way they walk in order to minimize their energy expenditure [1] [26, 32, 33]. Consequently, their cadence tends to be higher and the stride length to be shorter for people carrying an object. Also, the duration of the double-support phase of the gait cycle (i.e. the period of time when both feet are on the ground) tends to be larger for a person carrying an object.

From a computer vision viewpoint, carried objects can be classified into two types:

- Objects that alter the way the person *walks* (i.e. the biomechanics of his gait) because of their sheer weight and/or size.
- Objects that alter the way the person *appears* because they occlude part of the body when carried.

[1] in fact this is a general concept in gait dynamics that applies to any walking conditions

Hence, a vision algorithm can either determine if the person's gait is within the range of normal gait dynamics, in which case the problem becomes one of gait classification, or can try to detect changes in its appearance that might suggest the presence of a carried object.

The detection of deviations from normal gait can most simply be done by measuring certain gait parameters (temporal, kinematic and kinetic) and determining if they correspond to those of a naturally walking person, assuming we have a model of what 'normal gait' is. This is the typical approach taken by clinical gait analysts [40]. It is not a good approach for a computer vision method, however, since it would require accurate tracking of body landmarks, which is still an error prone process. An exception to this are the spatial and temporal periods of gait (i.e. cadence and stride length), which recent work has shown can be computed robustly from video [53, 12]. In the former, pedestrians are detected by estimating the pace (cadence) and stride length of a 'rythmically' moving object, then classifying it as a pedestrian if the pace and stride length lie within some fixed range of 'typical human walking' (they model the latter as two independent Gaussian distributions with fixed parameters).

In [12], Cutler and Davis show that stride length can be automatically estimated from video and an accurately calibrated camera to within 2 inches. Unfortunately, with this much accuracy, it is not possible to confidently distinguish a load-carrying person from a naturally-walking person, since the difference in their stride lengths is typically on the order of only 1-2cm [29].

Thus, a parametric approach for discriminating load-carrying and natural-walking gaits seems impractical. Instead, we shall use periodicity and silhouette shape cues to characterize the differences between these gaits. Specifically, we contend that the spatiotemporal patterns of human silhouette satisfy certain constraints, that are often when the person carries an object. Our method essentially analyzes silhouette shape over many frames, and so it is robust to (spurious) segmentation errors, unlike a static shape analysis method for example that would try to detect a 'bump' in the silhouette contour from one frame.

5.2 Assumptions

We limit the scope of the problem by making the following assumptions:

- The camera is stationary. This simplifies the foreground detection procedure, and helps decouple detection problem from the problem at hand.
- The person is walking in upright pose. This is a reasonable assumption for a person to carry an object.
- The person walks with a constant velocity for a few seconds (i.e throughout the analysis for carried object detection).

5.3 Overview of Method

The method consists of a sequence of three processing steps or modules. First we detect and track the person for some N frames in the video sequence and obtain N binary blobs of the person. Then, we classify the person as *naturally-walking* or *object-carrying*, based on spatiotemporal analysis of certain features of binary silhouette of the blobs. Finally, we segment the object via static shape analysis of a select frame.

5.4 Foreground Detection and Tracking

Since the camera is assumed static, foreground detection can be achieved via background modelling and subtraction. We use the non-parametric background modelling technique that is essentially a generalization of the mixed-Gaussian background modelling approach [15, 47], and is well suited for outdoor scenes in which the background is often not perfectly static (for e.g. occasional movement of tree leaves and grass). A number of standard morphological cleaning operations are applied to the detected blobs to correct for random noise. Frame-to-frame tracking of a moving object is done via simple overlap of its blob bounding boxes in the current and previous frames.

5.5 Carried Object Detection

Given N blobs of a moving person in N frames, we determine whether it corresponds to a naturally-walking person, or a person *possibly* carrying an object. For this, we formulate two constraints that characterize natural walking as a function of the temporal characteristics (periodicity and amplitude) of certain binary shape features of the person's silhouette. A moving person is deemed as possibly carrying an object if either one of these constraints is violated, and as naturally-walking otherwise.

5.5.1 The Algorithm

The human gait is highly structured both in space and time, due to the bilateral symmetry of the human body and the cyclic coordinated movement patterns of the various body parts, which repeat at the fundamental frequency of walking. A good model for the oscillatory motion of the legs is a pair of planar pendula oscillating 180^o out of phase [35, 42, 32]. The same can be said about the swinging of the arms [50].

The differences between natural walking gait and load-carrying gait may be attributed to any of the following (this list does not claim to be exhaustive):

- The manner by which the person carries the object; e.g. when holding a box with both hands, the arms no longer swing.
- Occlusion of part of the silhouette, such as when a handbag or suitcase held on the side occludes the legs.
- Protrusion of the object outside silhouette, hence distorting its contour shape.
- The sheer weight of an object; a heavy object will most likely cause a person not to swing his arms as much.

We capture these differences between natural gait and load-carrying gait via temporal behavior of correspondence-free binary shape features, consisting of the bounding box widths of horizontal segments of the silhouette. We formulate constraints on the periodicity and amplitude of these features, and claim that these constraints are typically violated when the person is carrying an object.

Consider the subdivision of the silhouette into 4 segments, shown in Figure 5.1; three equal contiguous horizontal segments over the lower body region, denoted $L1$, $L2$, $L3$ (bottom segments first), and one

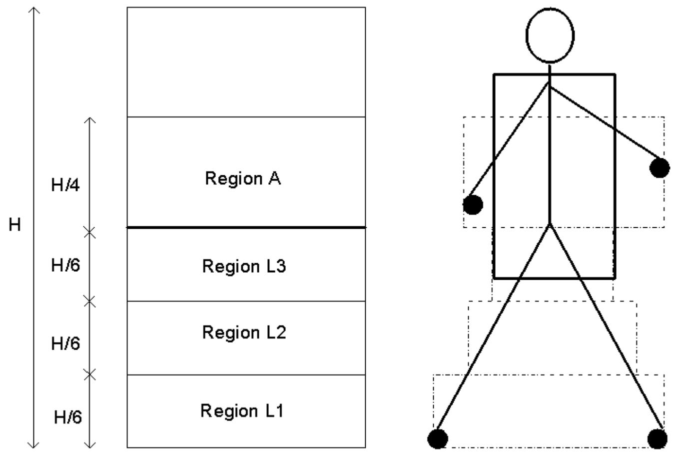


Figure 5.1. Subdivision of body silhouette into 5 segments for shape feature computation.

segment for the upper body region, denoted U. We also define $L = L1 \cup L2 \cup L3$ (i.e. the lower half of the body).

We compute the boundary box width over each of the defined segments of the silhouette, for each blob in the sequence. The time series thus obtained are denoted by $\beta_L(t)$, $\beta_{L1}(t)$, $\beta_{L2}(t)$, $\beta_{L3}(t)$ and $\beta_U(t)$, corresponding to segments L, $L1$, $L2$, $L3$ and U, respectively. Since natural walking gait is characterized by the oscillation of the legs and swinging of the arms at the period of gait [32], we contend that:

$$Period(\beta_L(t)) = Period(\beta_U(t)) = T_0 \tag{5.1}$$

where $Period(.)$ denotes the fundamental period of a time series. If Equation 5.1 is not satisfied, then the person is deemed to be possibly carrying an object. Furthermore, we claim that their mean amplitudes, denoted by $\bar{\beta_L}$, $\bar{\beta_U}$, etc., satisfy:

$$\bar{\beta_L} > \bar{\beta_U} \wedge \bar{\beta_{L1}} > \bar{\beta_{L2}} > \bar{\beta_{L3}} \tag{5.2}$$

Though we do not formally prove this claim here, intuitively it is an artifact of the pendular-like motion of the arms and legs, as mentioned above. The person is also deemed to be possibly carrying an object if Inequation 5.2 is not satisfied.

The binary shape features used by our method to capture the motion patterns of human gait, namely the width of the silhouette bounding box, are indeed akin to those used in [39] and those in [16]; Niyogi and Adelson [39] extract these same shape features and use them to detect humans in video by finding the characteristic 'spatiotemporal braids' created by a walking person in an image sequence; Fujiyoshi and Lipton [16] extract the five (most salient) curvature maxima of the silhouette contour to track the 'gross extremities' of the body, and then analyze the 1D trajectory of each point to detect cyclic motion.

We estimate the period of $\beta_L(t)$ and $\beta_U(t)$ via the autocorrelation method, which is robust to colored noise and non-linear amplitude modulations, unlike Fourier analysis [12]. For this, we smooth each signal, detrend it to account for any linear increase caused by change in viewpoint, and compute its autocorrelation $A(i)$ for i in some interval $[-lag, lag]$, where lag is chosen such that it is sufficiently larger than the expected period of gait. The period is then estimated as the average distance between each two consecutive *major* peaks in $A(i)$.

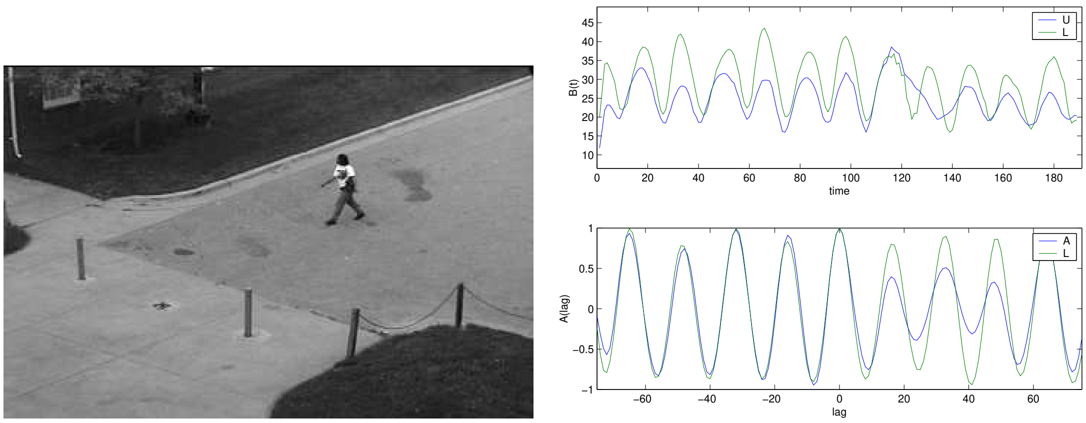
$A(i)$ may also contain *minor* (small) peaks between every two major peaks. A minor peak corresponds to the similarity of the two symmetric poses; when the left leg is leading and when the right leg is leading [12]. This is due to the bilateral symmetry of the human gait. Hence, we first detect all peaks in $A(i)$, then determine whether there are minor peaks by computing the average strength of all odd peaks (i.e. first, third, etc.) and all even peaks (i.e. second, fourth, etc.). If they are comparable, then there are no minor peaks. Otherwise, the odd peaks are minor peaks.

In summary, our algorithm determines that a person is possibly carrying an object if any of the properties given in 5.1 and 5.2 which we claim characterize natural walking gait, is not satisfied (since load-carrying is considered un-natural gait). In the next section, we give examples to illustrate these claims.

5.5.2 Examples

Figure 5.2 shows one frame of a person walking frontoparallel to the camera, along with the corresponding $\beta_L(t)$ and $\beta_U(t)$ (topmost plot) and their respective autocorrelation functions (in bottom plot), $A_L(t)$ and $A_U(t)$. Note that both $A_L(i)$ and $A_U(i)$ have minor and major peaks.

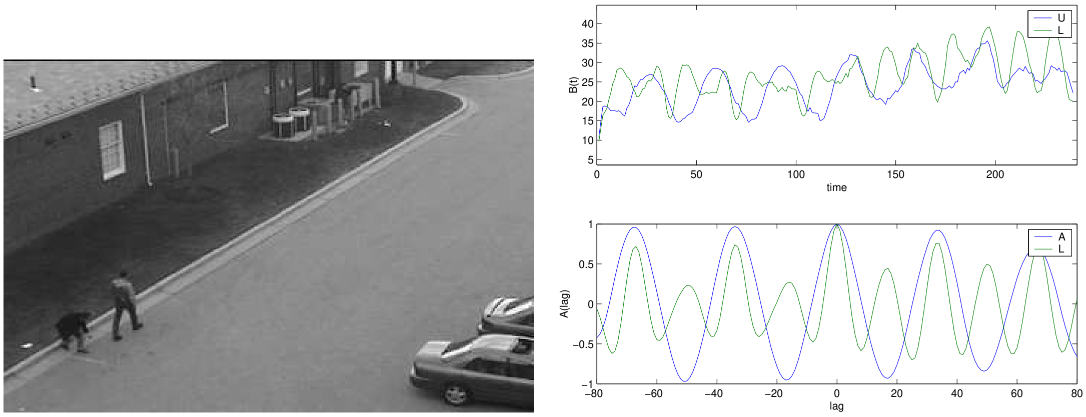
Figure 5.3 shows the same thing for a person walking non-frontoparallel to the camera. Note that, while $A_L(i)$ has major and minor peaks, $A_U(i)$ does not. The bilateral symmetry of the upper region is lost when the person is walking non-frontoparallel to the camera. However, in both cases, $\beta_L(t)$ and $\beta_U(t)$ have the same period, since only major peaks are relevant in computing the period.



(a) (b)

Figure 5.2. (a) Person walking frontoparallel to camera, and (b) corresponding width series and autocorrelation functions.

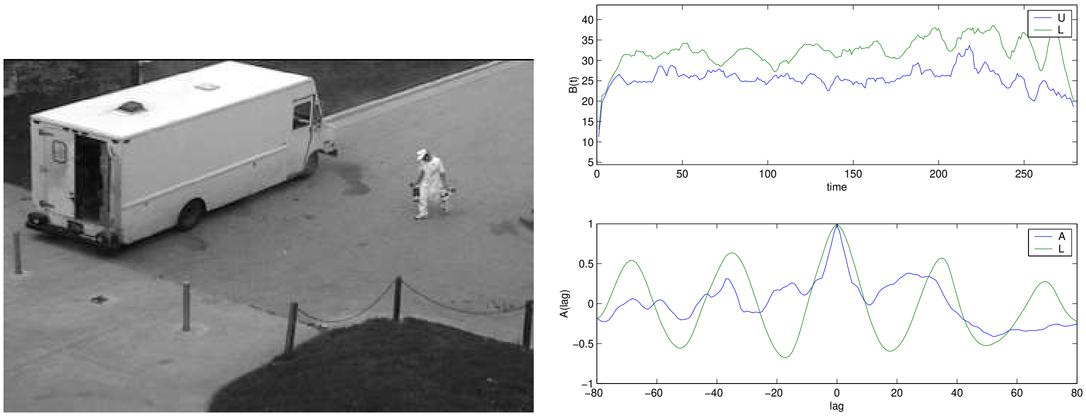
The following two examples illustrate cases in which either of $A_L(i)$ or $A_U(i)$ is not periodic due to the presence of a carried object. Figure 5.4 shows a person carrying a bucket in each arm. He is hardly swinging his arms, perhaps because the buckets must be heavy. This explains why $A_U(i)$ is not periodic, while $A_L(i)$ is. Figure 5.5 illustrates the same case with a different person. Note here that $A_U(i)$ seems to oscillate at a higher frequency than the legs, which maybe due to independent oscillation of the carried handbag (particularly if it's lightweight).



(a) (b)

Figure 5.3. (a) Person walking non-frontoparallel to camera, and (b) corresponding width series and autocorrelation functions.

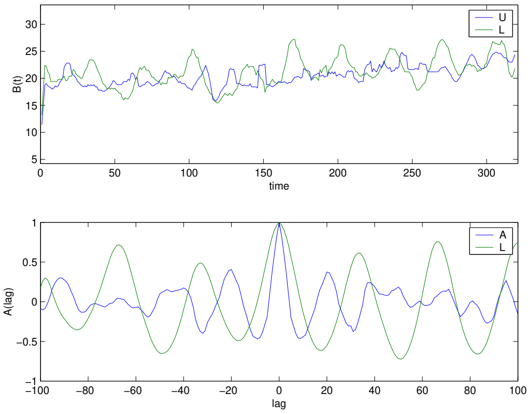
Another case when $A_L(i)$ is periodic while $A_U(i)$ is not, happens when a person is carrying an object with both hands, as illustrated by both examples in Figure 5.6 and Figure 5.7. This can be easily explained by the fact that both arms are not swinging (since they are holding the object). Note that for the second example, Equation 5.1 is not satisfied because the carried box wider than legs-width. This provides further evidence that the person might be carrying an object.



(a) (b)

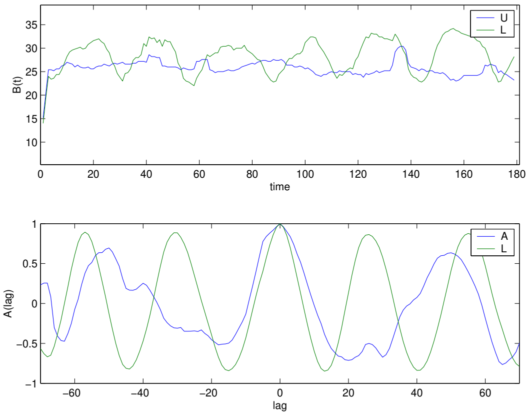
Figure 5.4. Person carrying two objects on the side. Width series of lower body region is periodic, while that of upper body is not.

The examples in Figure 5.8, Figure 5.9, and Figure 5.10 illustrate cases of a person carrying an object for which Equation 5.1 and the first part of Inequation 5.2, i.e. $\bar{\beta_L} > \bar{\beta_U}$ are satisfied. However, in all three cases, detection of the object is in fact made possible by the fact that the second part of this



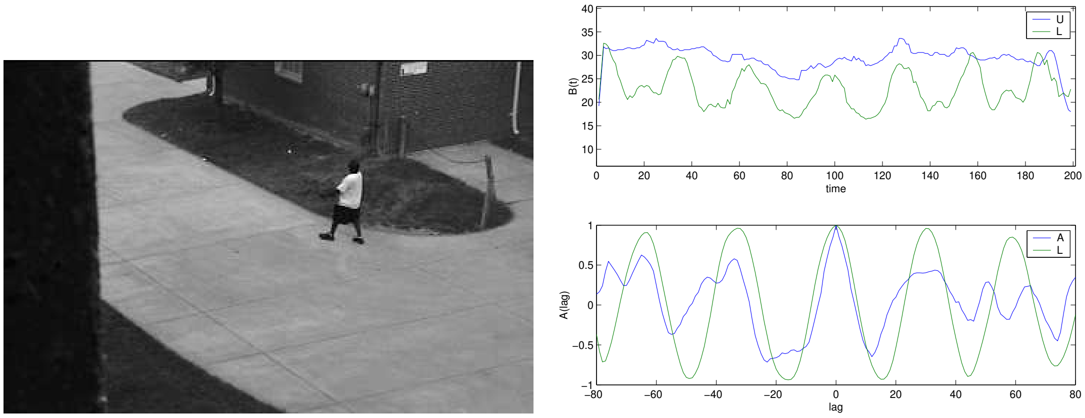
(a) (b)

Figure 5.5. Person carrying a handbag on the side. Width series of lower body region is periodic and of upper body region is aperiodic.



(a) (b)

Figure 5.6. Person carrying a box in front with two hands. Width series of lower body region is periodic and of upper body region is aperiodic.



(a) (b)

Figure 5.7. Person carrying a box in front with two hands. Width series of lower body region is periodic and of upper body region is aperiodic.Furthermore, average width of upper body region is larger than that of lower body region.

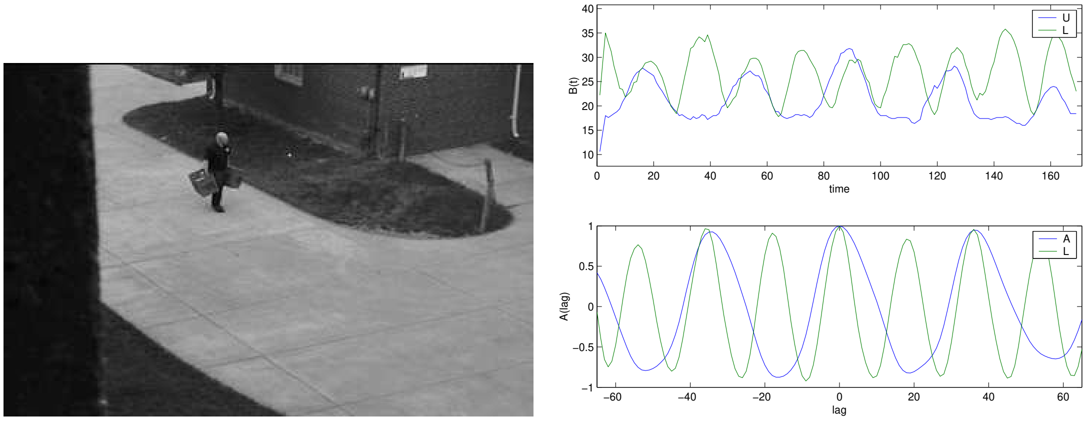
Inequation, i.e. $\bar{\beta}_{L1} > \bar{\beta}_{L2} > \bar{\beta}_{L3}$, evaluates to false. The underlying reason for this is because the carried object occludes part of the legs, and hence distorts the pendular-like motion pattern as seen by the camera. Figure 5.11 shows β_{L1}, β_{L2} and β_{L3} for the examples in Figure 5.2, Figure 5.8, Figure 5.9, and Figure 5.10, respectively.

5.5.3 Carried Object Segmentation

To segment the object, we first observe that the silhouette contour of a walking person is smooth, except when the arms swing outside the body. Hence, if we choose the frame that corresponds to the gait pose when the two legs are joined together (and hence arms are next to the body), then we may assume that any remaining 'significant' curvature maxima in the contour occur because of a protruding carried object. The region enclosed within the convex hull provides a crude initial estimate of the object. We extend this region by adding any connected blob pixels that lie to the left or right of it (i.e. are more extreme than it).

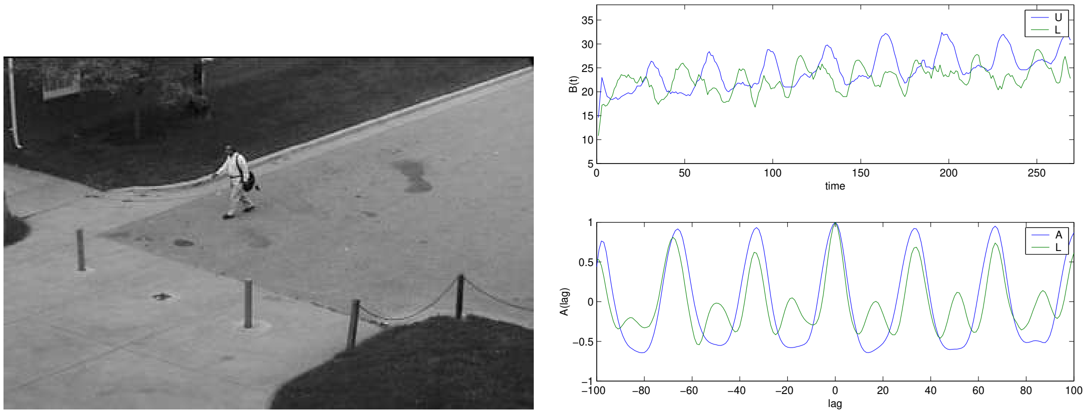
5.6 Experiments and Results

We tested the method on 41 outdoor sequences taken from various camera viewpoints, and captured at 30 fps and an image size of 360x240. The examples given above were samples of these sequences. All of these sequences were depictions of spontaneous, un-orchestrated human activity within the parking lot of a university building. Table 5.1 summarizes the type of sequences used and the detection results for each category. The detection rate is hence 83.33% (20 out of 24) and the false alarm rate is at 11.76% (2 out of 17). We notice that the mis-detections occurred when the object was too small to distort the spatio-temporal shape patterns of the silhouette, while false alarms occurred due to segmentation errors.



(a) (b)

Figure 5.8. Person carrying two objects on the side. Width series of upper and lower body regions are periodic with same period.

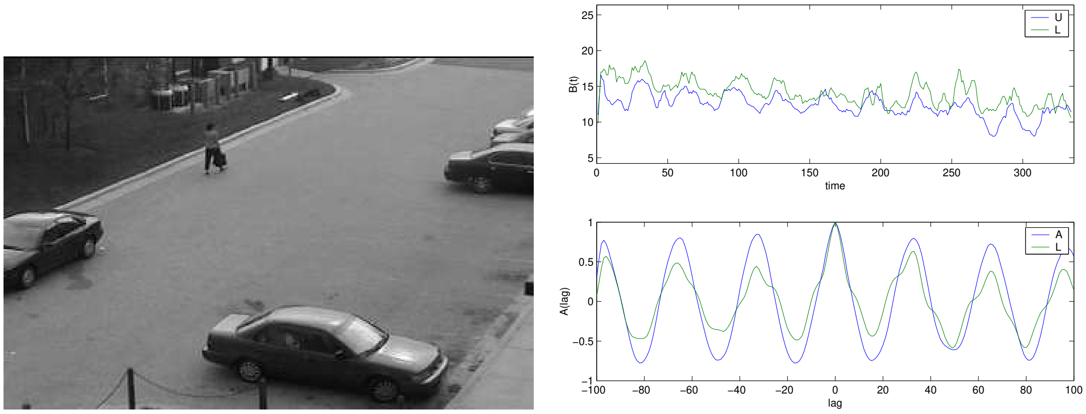


(a) (b)

Figure 5.9. Person carrying two objects on the side. Width series of upper and lower body regions are periodic with same period.

	Total	Natural-walking	Load-carrying
Walking	17	15	2
Backpack	7	2	5
On the side	13	2	11
With both hands	4	0	4

Table 5.1. Carried Object Detection Results.



(a) (b)

Figure 5.10. Person carrying two objects on the side. Width series of upper and lower body regions are periodic with same period.

5.7 Summary and Discussion

We have described a novel method for determining whether a person in carrying an object in monocular sequences seen from a stationary camera. This is achieved via temporal analysis of simple correspondence-free binary shape features, that exploit the periodic and pendular-like motion of legs and arms. The final phase of our method is a crude segmentation of the object via static shape analysis of the blob silhouette. We plan to improve this technique in future work in at least two ways; first by using color/edge information to refine the segmented region and obtain a more accurate estimate of the object, and second by applying the segmentation process on blobs in multiple frames.

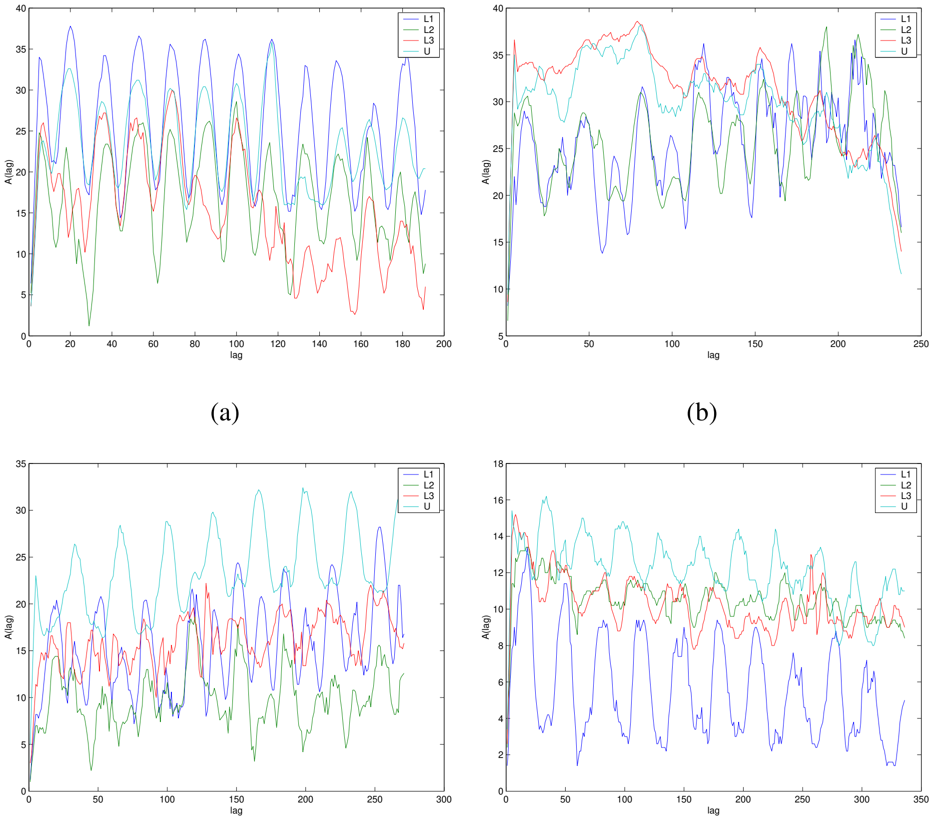


Figure 5.11. β_{L1}**,** β_{L2} **and** β_{L3} **for (a) naturally-walking person, and (b),(c),(d) a person carrying an object. Inequation 5.2 is only violated for the latter three.**

Chapter 6

Conclusions and Future Work

We presented two methods for gait recognition and a method for the detection of carried objects. We take a motion-based recognition approach in all three methods, as we extract correspondence-free motion features from images and use them to directly characterize body movement. We plan to refine and extend this work in the following ways:

6.1 Eigengait Method

- Test on a larger population of *outdoor* sequences taken from arbitrary viewpoints. These sequences present challenges and difficulties that did not arise with frontoparallel and treadmill walking. For example, the scaling of images required for non-frontoparallel sequences may prove to be a significant source of error in the similarity plot. Note also that we need to automatically estimate the camera viewpoint of a walking person, which requires knowledge of both the camera calibration parameters and the walking plane of the person.

- Assess the number of camera viewpoints needed to index recognition (i.e. the granularity of pan and tilt angles ranges).

- Assess the effect of image resolution on recognition performance. We currently scale down the person blobs to a fixed number-of- pixels-on-target. We aim to determine the effect of varying this parameter.

6.2 Cadence/Stride-based Method

- Test on a larger, more diverse population of people.

- Assess the effect of using higher resolution cameras (we currently use only 360x240), and improving the tracking precision on performance.

- Assess the effect of camera viewpoint on the accuracy of the method, via more detailed analysis of the estimation error of cadence and stride length. Although the method is view-invariant, in that the same algorithm can be applied regardless of viewpoint. However, we believe its performance is *not* view-invariant. The presented error analysis assumes constant camera depth throughout the sequence. It also assumes a constant pixel error, whereas it typically varies as a function of camera viewpoint and position in the image. We expect that estimation error will degrade very quickly for near front-on viewpoints (i.e. where the person is walking towards or away from the camera). Hence, it is important to evaluate just how much performance degrades as a function of camera viewpoint.

6.3 Towards an Integrated Multi-camera Surveillance System

- Integration of multiple features for gait recognition:

  Perhaps the best approach for achieving better person identification results is to combine the Eigengait features with the stride length/cadence features. However, because of the different nature of these features, a careful analysis of variance is needed.

- Best-view selection:

  The performance (in terms of accuracy and robustness) of any image-based human movement (gait) analysis method is inherently view-dependent, since inevitably images capture only a planar projection of gait dynamics, and not its entire phase space. Hence, a different subset of this phase space is extracted from different camera viewpoints. Since gait dynamics are dominated by leg and arm movements, this subset is largest when the sagittal plane[1] is parallel to the image plane. This happens in fronto-parallel walking.

  Thus, for best performance of gait recognition and human movement analysis methods, it is ideal to use a camera that is nearly fronto-parallel to the walking person. This 'view-selection' capability can be provided by a distributed multi-camera system; the camera that has the 'best view' of the moving person is selected to analyze its movement.

  We plan to develop the camera control and planning algorithms that implement this capability, using the existing network of outdoor Philips cameras. This problem also involves a great deal of low-level vision issues, associated with ensuring smooth successful camera handoff, i.e. data transfer from one camera to another.

[1]The Sagittal plane corresponds to the side view of a walking person, and is hence the plane of movement of the legs and arms.

Appendix A

Stride Length Error Analysis

Consider the outdoor camera configuration shown in Figure A.1. The camera has a height H, and looks down on the ground plane with tilt angle T and vertical field of view F. Suppose that a person walks roughly parallel to the image plane (to simplify the analysis), and with roughly constant velocity. The stride S is estimated using the number of steps walked, N, and the distance walked, W:

$$S = W/N \tag{A.1}$$

The uncertainty in S is σ_S:

$$\sigma_S = S\sqrt{(\sigma_W^2/W^2 + \sigma_N^2/N^2)} \tag{A.2}$$

where σ_N is the uncertainty in N, and σ_W is the uncertainty in W. We assume the uncertainty in W and N are uncorrelated (to simplify the analysis). We have empirically estimated $\sigma_N = 0.07$ steps with our configuration and method of estimating periodic motion.

The horizontal ground sampling distance per pixel is $g = F * R/(V \cos T)$, where V is the vertical resolution, $R = \sqrt{D^2 + H^2}$ is the distance from the camera to the person, and $D = H * \tan T$ is the distance from the camera base to the person. We approximate $\sigma_W = 2 * g$, which assumes we can track

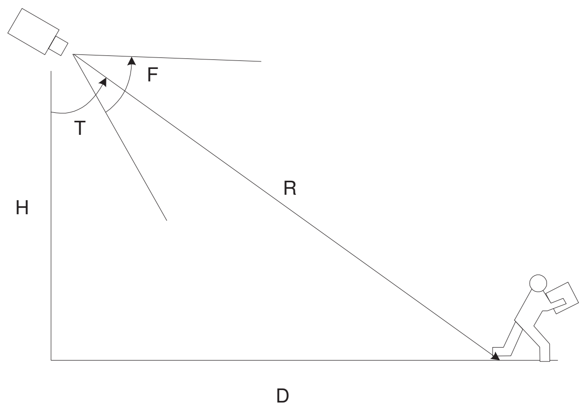


Figure A.1. Outdoor surveillance camera configuration.

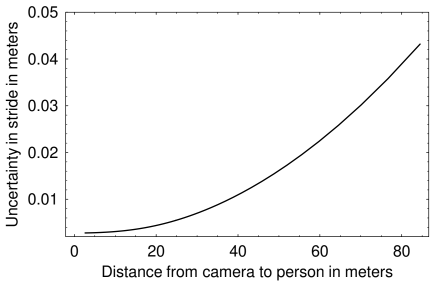


Figure A.2. Stride uncertainty as a function of distance from camera to person.

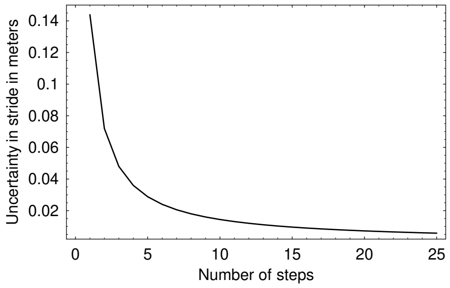


Figure A.3. Stride uncertainty as a function of number of steps.

people with two pixel accuracy in the video image. We neglect camera calibration errors in this analysis, as they they were small.

For our outdoor surveillance configuration, $F = 12\,\text{deg}$, $T = 64\,\text{deg}$, $H = 15\text{m}$, $V = 240$ pixels, $N = 16$ steps, $D = 30.5\text{m}$; the resultant $\sigma_S = 7\text{mm}$. Figure A.2 gives a plot of σ_S as a function of D. Note that the stride error is relatively low compared to the horizontal ground sampling distance since we are estimating the stride over many steps. We are exploiting the fact that the total distance traveled is the sum of the individual steps; implicitly, we use the fact that people move with piecewise contiguous steps. Therefore, as W and N increases and D remains relatively constant, then σ_S significantly decreases. For example, at the distance of $D = 30.5\text{m}$, the horizontal ground sampling distance $g = 68\text{mm}$, while $\sigma_S = 7\text{mm}$. Figure A.3 shows the uncertainty in stride as a function of number of steps taken[1].

Note that we cannot exploit the same error-reduction method in estimating a person's height. However, while we cannot accurately estimate height with a low resolution video camera in this configuration, we can accurately esimate stride and cadence.

[1]To further explain the reduction of error, consider the following problem: suppose you were asked to measure the length of a poker card, and you were given a tape ruler that is accurate to 1cm. To achieve greater accuracy, you take 20 cards from the same deck, and align them to be piecewise contiguous. You measure the length of all 20 cards, and divide this measure by number number of cards. The cards can now be measured with 20 times the accuracy as before with using just a single card.

Bibliography

[1] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: a Review," in *Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, 1997.

[2] K. Akita, "Image Sequence Analysis of Real World Human Motion," *Pattern Recognition*, Vol. 17, No. 1, pp. 73–8, 1984.

[3] C. Barclay, J. Cutting, and L. Kozlowski, "Temporal and Spatial Factors in Gait Perception that Influence Gender Recognition," *Perception and Psychophysics*, Vol. 23, No. 2, pp. 145–152, 1978.

[4] J. Bigun, G. Chollet, and G. Borgefors, *Audio- and Video-based Biometric Person Authentication*, Springer, 1997.

[5] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford: Clarendon Press, 1995.

[6] E. Borovikov, R. Cutler, T. Horprasert, and L. Davis, "Multi-perspective Analysis of Human Actions," in *Third International Workshop on Cooperative Distributed Vision*, 1999.

[7] C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences," in *Proceedings of the Computer Vision and Pattern Recognition*, 1997.

[8] L. W. Campbell and A. Bobick, "Recognition of Human Body Motion Using Phase Space Constraints," in *International Conference on Computer Vision*, 1995.

[9] C. Cedras and M. Shah, "A survey of motion analysis from moving light displays," in *Proceedings of the Computer Vision and Pattern Recognition*, pp. 214–221, 1994.

[10] D. Cunado, M. Nixon, and J. Carter, "Gait Extraction and Description by Evidence Gathering," in *Proceedings of 2nd Int. Conf. on Audio- and Video-Based Biometric Person Authentication*, 1999.

[11] R. Cutler, *On the detection, analysis, and applications of oscillatory motions in video sequences*, Ph.D. dissertation, Phd Dissertation, University of Maryland, College Park, 2000.

[12] R. Cutler and L. Davis, "Robust Real-time Periodic Motion Detection, Analysis and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 2, pp. 129–155, 2000.

[13] J. Cutting and L. Kozlowski, "Recognizing Friends by Their Walk: Gait Perception Without Familiarity Cues," *Bulletin Psychonomic Soc.*, Vol. 9, No. 5, pp. 353–356, 1977.

[14] J. W. Davis and A. F. Bobick, "The Representation and Recognition of Action Using Temporal Templates," in *Proceedings of the Computer Vision and Pattern Recognition*, pp. 928–934, 1997.

[15] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proceedings of International Conference on Computer Vision*, 2000.

[16] H. Fujiyoshi and A. Lipton, "Real-time Human Motion Analysis by Image Skeletonization," in *IEEE Workshop on Applications of Computer Vision*, October 1998.

[17] D. Gavrila, "The visual analysis of human movement: a survey," *Computer Vision and Image Understanding*, Vol. 73, pp. 82–98, January 1999.

[18] D. M. Gavrila and L. Davis, "Towards 3-D Model-based Tracking and Recognition of Human Movement: a Multi-View Approach," in *IEEE International Conference on Automatic Face and Gesture Recognition*, (Zurich, Switzerland), 1995.

[19] D. W. Grieve and R. Gear, "The Relationship Between Length of Stride, Step Frequency, Time of Swing and Speed of Walking for Children and Adults," *Journal of Ergonomics*, Vol. 5, No. 9, 1966.

[20] I. Haritaoglu, R. Cutler, D. Harwood, and L. Davis, "Backpack: Detection of people carrying objects using silhouettes," *Computer Vision and Image Understanding*, Vol. 6, No. 2, 2000.

[21] I. Haritaoglu, D. Harwood, and L. Davis, "W4S: A Real-Time System for Detecting and Tracking People in 21/2 D," in *European Conference on Computer Vision*, June 1998.

[22] Q. He and C. Debrunner, "Individual Recognition from Periodic Activity Using Hidden Markov Models," in *IEEE Workshop on Human Motion*, 2000.

[23] D. Hoffman and B. Flinchbaugh, "The interpretation of biological motion," *Biological Cybernetics*, 1982.

[24] D. Hogg, "Model-based vision: a program to see a walking person," *Image and Vision Computing*, Vol. 1, No. 1, 1983.

[25] P. S. Huang, C. J. Harris, and M. S. Nixon, "Comparing Different Template Features for Recognizing People by their Gait," in *BMVC*, 1998.

[26] V. Inman, H. J. Ralston, and F. Todd, *Human Walking*, Williams and Wilkins, 1981.

[27] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis," *Perception and Psychophysics*, Vol. 14, No. 2, pp. 201–211, 1973.

[28] I. T. Joliffe, *Principal Component Analysis*, Springer-Verlag, 1986.

[29] H. Kinoshita, "Effects of Different Loads and Carrying Systems on Selected Biomechanical Parameters Describing Walking Gait," *Ergonomics*, Vol. 28, No. 9, pp. 1347–1362, 1985.

[30] J. Little and J. Boyd, "Recognizing People by their Gait: the Shape of Motion," *Videre*, Vol. 1, No. 2, 1998.

[31] F. Liu and R. Picard, "Finding periodicity in space and time," *International Conference on Computer Vision*, pp. 376–383, January 1998.

[32] K. Luttgens and K. Wells, *Kinesiology: Scientific Basis of Human Motion*, Saunders College Publishing, 7th ed., 1982.

[33] R. A. Mann and J. Hagy, "Biomechanics of Walking, Running and Sprinting," *American Journal of Sports Medicine*, Vol. 8, No. 8, pp. 345–350, 1980.

[34] P. Martin and R. Nelson, "The effect of Carried Loads on the Walking Patterns of Men and Women," *Ergonomics*, Vol. 29, No. 10, pp. 1191–1202, 1986.

[35] T. A. McMahon, *Muscles, Reflexes, and Locomotion*, Princeton University Press, 1984.

[36] D. Meyer, J. Psl, and H. Niemann, "Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture Densities," in *BMVC*, pp. 459–468, 1998.

[37] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognition Letters*, Vol. 17, pp. 155–162, 1996.

[38] M. Murray, "Gait as a total pattern of movement," *American Journal of Physical Medicine*, Vol. 46, No. 1, pp. 290–332, 1967.

[39] S. Niyogi and E. Adelson, "Analyzing and Recognizing Walking Figures in XYT," in *Proceedings of the Computer Vision and Pattern Recognition*, pp. 469–474, 1994.

[40] J. Perry, *Gait Analysis: Normal and Pathological Function*, SLACK Inc., 1992.

[41] J. Philips, Hyeonjoon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *PAMI*, Vol. 22, No. 10, pp. 1090–1103, 2000.

[42] J. Piscopo and J. A. Baley, *Kinesiology: the Science of Movement*, John Wiley and Sons, 1st ed., 1981.

[43] R. Polana and R. Nelson, "Detection and Recognition of Periodic, Non-rigid Motion," *International Journal of Computer Vision*, Vol. 23, pp. 261–282, June/July 1997.

[44] B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press, 1996.

[45] K. Rohr, "Towards Model-based Recognition of Human Movements in Image Sequences," in *CVGIP*, vol. 59, 1994.

[46] Y. Song, X. Feng, and P. Perona, "Towards Detection of Human Motion," in *Proceedings of the Computer Vision and Pattern Recognition*, 2000.

[47] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins*, 1999.

[48] P. Tsai, M. Shah, K. Keiter, and T. Kasparis, “Cyclic Motion Detection for Motion–based Recognition,” *Pattern Recognition*, Vol. 27, No. 12, pp. 1591–1603, 1994.

[49] M. Turk and A. Pentland, “Face Recognition using Eigenfaces,” in *Proceedings of the Computer Vision and Pattern Recognition*, 1991.

[50] D. Webb, R. H. Tuttle, and M. Baksh, “Pendular Activity of Human Upper Limbs During Slow and Normal Walking,” *American Journal of Physical Anthropology*, Vol. 93, pp. 477–89, 1994.

[51] S. Weiss and C. Kulikowski, *Computer Systems that Learn*, Morgan Kaufman, 1991.

[52] D. Winter, *The Biomechanics and Motor Control of Human Gait*, Univesity of Waterloo Press, 1987.

[53] S. Yasutomi and H. Mori, “A Method for Discriminating Pedestrians based on Rythm,” in *IEEE/RSG Intl Conf. on Intelligent Robots and Systems*, 1994.