Towards a View Invariant Gait Recognition Algorithm

Amit Kale, Amit K. Roy Chowdhury and Rama Chellappa*
Center for Automation Research
University of Maryland at College Park
College Park, MD 20740

Abstract

Human gait is a spatio-temporal phenomenon and typifies the motion characteristics of an individual. The gait of a person is easily recognizable when extracted from a side-view of the person. Accordingly, gait-recognition algorithms work best when presented with images where the person walks parallel to the camera (i.e. the image plane). However, it is not realistic to expect that this assumption will be valid in most real-life scenarios. Hence it is important to develop methods whereby the side-view can be generated from any other arbitrary view in a simple, yet accurate, manner. That is the main theme of this paper. We show that if the person is far enough from the camera, it is possible to synthesize a side view (referred to as canonical view) from any other arbitrary view using a single camera. Two methods are proposed for doing this: i) by using the perspective projection model, and ii) by using the optical flow based structure from motion equations. A simple camera calibration scheme for this method is also proposed. Examples of synthesized views are presented. Preliminary testing with gait recognition algorithms gives encouraging results. A by-product of this method is a simple algorithm for synthesizing novel views of a planar scene.

1 Introduction

Human identification forms an important component of visual surveillance. In many such applications established non-invasive biometrics such as face or iris may not be available at sufficient resolution to be used for recognition. A biometric that can address some of these shortcomings is "gait", which is motivated by the fact that humans exhibit the capability of recognizing people even from impoverished displays of gait [1], indicating the presence of identity information. Gait can be detected and measured even in low resolution video and can also be used with IR imagery [2, 3, 4, 5]. The gait of a person is best reflected when he/she presents a side view (referred to in this paper as a canonical view) to the camera. Hence, most gait recognition algorithms rely on the availability of the side view of the subject. The situation is analogous to face recognition where it is useful to have frontal views of the person's face.

In realistic surveillance scenarios, however, it is unreasonable to assume that a subject would always present a side-view to the camera and hence, gait recognition algorithms need to work in a situation where the person walks at an arbitrary angle to the camera.There are two effects of a change in viewing direction. One is simply the foreshortening or lengthening that occurs as the person walks away or towards the camera. The second is the change in the apparent stride length. The most general solution to this problem would involve estimating a 3D model of the person from which the required canonical view can be generated. This problem requires the solution of the structure from motion (SfM) or stereo reconstruction problems [6, 7], which are known to be hard. To circumvent the problems associated with the estimation of 3D models, several approaches have been proposed for the gait recognition problem. Bobick and Johnson [8] use linear regression to map static parameters across views. In [9], Shakhnarovich et al. compute an image based visual hull from a set of monocular views which is then used to render virtual views for tracking and recognition. In this paper, we propose an alternative approach that can work with only a single camera and can synthesize canonical views of high quality in a way that uses the 3D structure only implicitly. These synthesized views can then be used for gait recognition. The order of computation is $O(mn)$, where m and n are the dimensions of the bounding box around the person.

Consider a person walking along a straight line which subtends an angle θ with the image plane (AC in Figure 1). If the distance, Z_0, of the person from the camera is much larger than the width, ΔZ, of the person, then it is reasonable to replace the scaling factor $\frac{f}{Z_0+\Delta Z}$ for perspective projection by an average scaling factor $\frac{f}{Z_0}$. In other words, for objects far enough from the camera, we can approximate the actual 3D object as being represented by a planar object. Assume that we are given a video of a person walking at a fixed angle θ (Figure 1). We show that by tracking the direction of motion, α, in the video sequence, we can accurately estimate

*Supported by the DARPA/ONR grant N00014-00-1-0908.

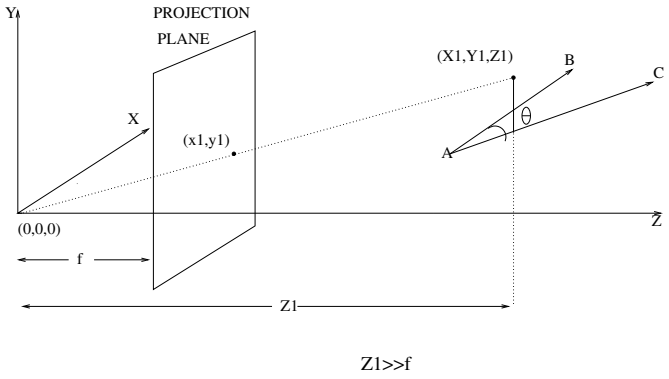


Figure 1: Imaging Geometry

the angle θ in the 3D world. This can be done is two ways: a) by using the perspective projection matrix, or b) by using the optical flow based SfM equations. We also show that a simple, yet precise, camera calibration scheme can be designed for this problem. Under the assumption of planarity, using the angle θ and the calibration parameters, we can synthesize side-views or canonical views of the person, which can then be passed on to the gait recognition algorithms. Since the planar approximation is reasonable for many surveillance scenarios where the distance between the camera and people is large, this is a practical approach for synthesizing canonical views required by many gait recognition algorithms. A by-product of the above method is a simple algorithm to synthesize novel views of a planar scene.

2 Theory

2.1 Imaging Geometry

The imaging setup is shown in Figure 1. The coordinate frame is attached rigidly to a camera with the origin at the center of perspective projection and the z-axis perpendicular to the image plane. Assume that the person walks with a translational velocity $\mathbf{V} = [v_x, 0, v_z]^T$ along the line AC. The line AB is parallel to the image plane XY and this is the direction of the canonical view which needs to be synthesized. The angle between the straight line AB and AC, i.e. θ, represents a rotation about the vertical axis we hence we shall call this the azimuth angle. We will use the notation that $[X, Y, Z]$ denotes the coordinates of a point in 3D and $[x, y]$ its projection on the image plane.

2.2 Estimating the Azimuth Angle from Video Sequence

We present two ways of estimating the angle θ from the video sequence.

Perspective Projection Approach

We assume that the person is walking along the straight line AC in Figure 1. Under exact perspective projection, straight lines map to straight lines. Thus the direction of motion in the 3D world corresponds to a straight line in the image plane, which can be estimated by tracking some points which move approximately rigidly as the person walks. Consider the equation of the 3D line which is at a height k from the ground plane and parallel to it, i.e.

$$Z = Tan(\theta)X + Z_0, Y = k. \tag{1}$$

Under perspective projection this line transforms to (see Appendix)

$$y = \frac{kf}{Z_0} - k\frac{Tan(\theta)}{Z_0}x, \tag{2}$$

where $x = f\frac{X}{Z_0+Tan(\theta)X}$, $y = f\frac{Y}{Z_0+Tan(\theta)X}$ and f denotes the focal length of the camera. Thus if the slope of the line in the image plane, viz. $Tan(\alpha)$, is known, then given $K = -\frac{k}{Z_0}$, the azimuth angle θ can be computed as

$$Tan(\theta) = \frac{1}{K}Tan(\alpha). \tag{3}$$

K can be obtained as a part of the calibration procedure. Note that using the orthographic projection model will result in giving a straight line $y = k$ which does not reflect the azimuth angle variation in the image plane. Thus our method will not work under orthographic projection assumptions.

Optical Flow Based SfM Approach

Assume that the motion between two consecutive frames in the video sequence is small. Using the optical flow based SfM equations, let $p(x, y)$ and $q(x, y)$ represent the horizontal and vertical velocity fields of a point (x, y) in the image plane. Since we consider straight line motion along AC, p and q are related to the 3D object motion and scene depth by [10]

$$p(x, y) = (x - fx_f)h(x, y) \tag{4}$$
$$q(x, y) = yh(x, y), \tag{5}$$

where $h(x, y) = v_z/z(x, y)$ is the scaled inverse scene depth and $x_f = Cot(\theta) = \frac{v_x}{v_z}, y_f = \frac{v_y}{v_z}$ is the focus of expansion (FOE). When $v_z = 0$ but $v_x \neq 0$, we see that $\theta = 0$, i.e. the canonical direction of walk, AB. Also, in this case $q(x, y) = 0$. For the case when the person walks at an azimuth angle $\theta \neq 0$, dividing (4) by (5) we obtain,

$$Cot(\alpha(x, y)) = c(x, y) - m(y, f)Cot(\theta), \tag{6}$$

where $Cot(\alpha(x, y)) = \frac{p(x,y)}{q(x,y)}$. For a fixed point in the image (e.g. centroid of the head) in the (6), we have

$$Cot(\alpha) = C - MCot(\theta). \tag{7}$$

$c(x, y)$ and $m(x, y)$ can be obtained from calibration data. By considering one particular point in a number of images, we can robustly estimate θ from α.

Equation (3) was derived under the perspective projection model, while equation (7) was derived using perspective projection and optical flow. Hence the difference in the two equations. However, both give numerically close results as explained in Section 3.

2.3 Coordinate Transformation to Canonical View

Having obtained the angle θ, we need to synthesize the canonical view. Let Z denote the distance of the object from the image plane. If the dimensions of the object are small compared to Z, then the variation in θ, $d\theta \approx 0$. This essentially corresponds to assuming a planar approximation to the object. Let $[X_\theta, Y_\theta, Z_\theta]'$ denote the coordinates of any point on the person who is walking at an angle $\theta \geq 0$ to the image plane (as shown in the Figure 1). Then

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = R(\theta) \cdot \begin{bmatrix} X_\theta \\ Y_\theta \\ Z_\theta \end{bmatrix}, \tag{8}$$

where

$$R(\theta) = \begin{bmatrix} Cos(\theta) & 0 & Sin(\theta) \\ 0 & 1 & 0 \\ -Sin(\theta) & 0 & Cos(\theta) \end{bmatrix}. \tag{9}$$

Denoting the corresponding image plane coordinates as $[x_\theta, y_\theta]'$ and $[x_0, y_0]'$ (for $\theta = 0$) and using the perspective transformation, we can obtain the equations for $[x_0, y_0]'$ as (see Appendix)

$$\begin{aligned} x_0 &= f\frac{x_\theta Cos(\theta) - f Sin(\theta)}{-x_\theta Sin(\theta) + f Cos(\theta)} \\ y_0 &= f\frac{y_\theta}{-x_\theta Sin(\theta) + f Cos(\theta)}, \end{aligned} \tag{10}$$

where

$$x = f\frac{X}{z} \text{and } y = f\frac{Y}{z}.$$

Equation (10) is particularly attractive since it does not involve the 3D depth; rather it is a direct transformation of the 2D image plane coordinates in the non-canonical view to get the image plane coordinates in the canonical one. Thus knowing the azimuth angle θ we can obtain a synthetic canonical view using (10) and a suitable texture mapping rule.

Synthesis of Arbitrary Planar Views

The extension of the above method to synthesize arbitrary planar views is straight-forward. Suppose we are given a video sequence of a person walking at an angle θ_1. This can be estimated from the direction of motion of the person in the video sequence (as explained above). Once this is done we can synthesize the view at an angle θ_2 by applying the transformation of (10) with $\theta = \theta_2 - \theta_1$. Thus, for planar scenes, we are able to generate synthetic views purely from the video data. This is important for many applications other than gait recognition, such as multimedia.

2.4 Application to Gait Recognition

Approaches in computer vision to the gait recognition problem can be broadly classified as being either model-based or model-free. Methods which assume *a priori* models [3, 11, 12] match the 2-D image sequences to the model data. In [3], the authors proposed a method where several ellipses are fitted to different parts of the binarized silhouette of the person and the parameters of these ellipses such as location of its centroid, eccentricity etc. are used as a feature to represent the gait of a person. Recognition is achieved by template matching. Model-free methods [4] establish correspondence between successive frames based upon the prediction or estimation of features related to position, velocity, shape, texture and color. In [2, 13], the sum of the white pixels along each row of the boxed silhouette image, referred to as the width vector, has been used as a feature. We now show that it is possible to obtain the transformed width vector in the canonical view directly from images obtained at an arbitrary view. Let $T : [x_\theta, y_\theta] \rightarrow [x_0, y_0]$ as represented in (10) and $J(x, y)$ denote the image intensity at (x, y). Also, let $\tilde{I}_\theta$ represent the synthesized image at angle θ and $\tilde{W}_\theta(y)$ the width vector for a particular row, y, in the image. Given the azimuth angle we can synthesize the width vector in the canonical view as

$$\begin{aligned} \tilde{W}_0(y) &\triangleq \sum_{x_0} \tilde{I}_0(x_0, y_0) \\ &= \sum_{x_\theta : (x_\theta, y_\theta) = T^{-1}(x_0, y_0)} I(x_\theta, y_\theta)), \end{aligned} \tag{11}$$

(12)

where $I(x, y) = 1$ if $J(x, y) \neq 0$ and $I(x, y) = 0$, otherwise.

3 Obtaining Camera Calibration Parameters

Using Equations (3), (6) and (10) requires a knowledge of the parameters f, K, C and M, which are essentially the camera calibration parameters for this problem. In order to compute f, we used a calibration grid marked with 20 points

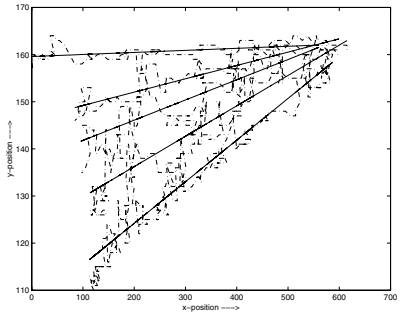
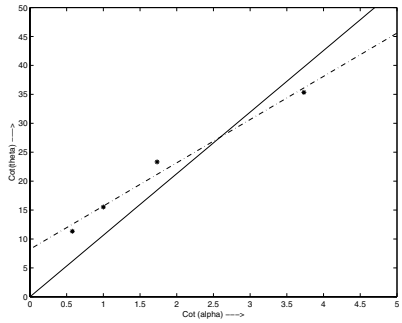


Figure 3: Tracked points in the video sequence and the best fit straight lines.



(a)

Figure 4: Calibration curves for $Cot(\theta)$ vs. $Cot(\alpha)$. The dots represent the true values, the solid line represents the calibration curve using (3) and the dashed and dotted line represents the calibration curve using ((7).

as shown in Figure 2. We placed the grid at 3 different azimuth angles $\theta = 15, 30, 45$ and obtained the point correspondences by hand. The points are related by (10). We represent the three angles by $\theta_j \in \{15, 30, 45\}$ and the coordinates of the i^{th} point by $[x^i_{\theta_j}, y^i_{\theta_j}]'$ and $[x^i_0, y^i_0]'$. Using these points we form a cost function $J(f)$ as shown in Equation (13). We solve this nonlinear regression using the Gauss-Newton method to obtain $f^* = \text{argmin}_f J(f)$, where

$$J = \sum_{i,\theta_j} \left(x^i_0 - f \frac{x^i_{\theta_j} Cos(\theta_j) - f Sin(\theta_j)}{f Cos(\theta_j) + x^i_{\theta_j} Sin(\theta_j)} \right)^2 + \left(y^i_0 - f \frac{y^i_{\theta_j}}{-x^i_{\theta_j} Sin(\theta_j) + f Cos(\theta_j)} \right)^2 . \quad (13)$$

Next, we consider the estimation of K, C and M. In order to do this, we captured videos of a person walking at $\theta = 0, 15, 30, 45, 60$. We tracked the position of a rigid point on the person followed by a median filtering of the trajectory. The resulting tracks are shown for the different θs in the Figure (3). To each of these tracks we fit a line using the least squares criterion. These are the solid lines in Figure (3), with slopes $Tan(\alpha(\theta))$. The top line is the case when $\theta = 0$. The lines for $\theta = 15$ is the one immediately below this line and so on. As may be expected, larger azimuth angles lead to larger image plane angles. The upper right corner where the lines intersect approximately, corresponds to the point from where the subjects start walking. These straight lines are the projections of the straight lines (one for each angle) traced out by the motion of the tracked rigid point in the 3D world. For the calibration procedure, we know the angle θ which traces out the straight line at the angle α. Given the corresponding values of α and θ, we can estimate K from (3).

Similarly, we can obtain $c(x, y)$ and $m(x, y)$ in (6). Let C and M be the corresponding representations for a particular point (x, y). The direction of motion of this line, $Cot(\alpha) = \frac{p}{q}$, is constant, since it moves along a straight line. Hence a robust estimate of α can be obtained by considering the motion of this point over a number of frames of the video sequence, i.e. the slope of the straight lines in Figure (3). Thus C and M can be determined from (7) by considering all the corresponding values of α and θ.

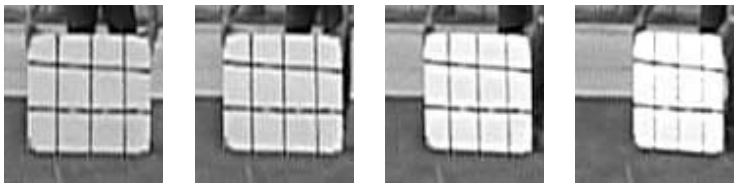
Figure 4 plots the true values of α and θ, as well as the two regression lines (3) and (7) obtained from the calibration procedure. Even though (3) and (7) were derived under different physical models, the straight lines in both the cases are good approximations of the true data. The main source of error is due to the assumption of straight line motion of a point, which is never precisely true in practice. Also, the effect of changing the image coordinate system can be taken into account easily by making the appropriate modifications in the perspective projection equations [6], at the cost of increasing the number of intrinsic camera parameters that need to be estimated during the calibration procedure.

Given a test video, we can estimate the value of α in a way similar to that shown in Figure 3. Thereafter, the value of θ can be read off directly from the calibration lines in Figure 4. The choice of which of the two lines should be used is left to the discretion of the user, who can determine that depending upon the validity of the assumptions in the particular case.

4 Experimental Results

In this section, we present results of our method for synthesizing canonical views of people from videos of them walking along arbitrary directions. We use the canonical views for gait recognition. The experiments are on a small number of people and conducted with the motivation of presenting a proof of concept for our algorithm. Detailed gait recognition experiments is the focus of future work.

Our database consists of 12 people, who walk along



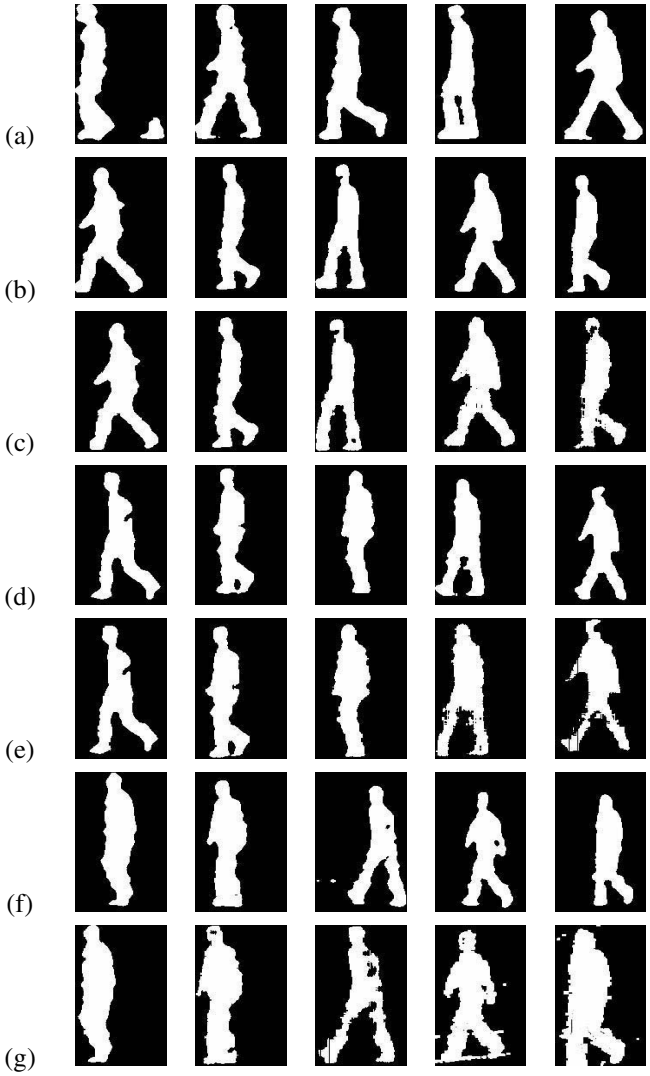
(a) (b) (c) (d)

Figure 2: Calibration grid placed at (a) 0 (b) 15 (c) 30 and (d) 45 degrees.

straight lines at different values of azimuth angle $\theta = 0, 15, 30, 45$ and 60 degrees. For the initial calibration, we consider one person walking at all the above angles. Background subtraction as discussed in [14] is first applied to all the image sequences, one for each angle. To remove spurious noise, a standard 3×3 low-pass filter is applied to the resultant motion image. A bounding box is then placed around the part of the motion image that maximally contains the moving person. The size of the box is chosen to accommodate the extreme cases of individuals in the database as regards height and girth. Further operations are carried out on this 'box'. The upper left corner of the box was tracked in the video. This is approximately the same as tracking a rigid point on the persons body. The angle α is obtained from the median filtered tracks in the image plane. Since θ is known for the calibration procedure, K, C and M were computed using the method described in Section 3.

For the video of an unknown person in the database the above image processing operations are repeated to compute the image plane angle α. Using the calibration line shown in Figure 4, the azimuth angle θ was obtained. Using this value of θ and the value of f obtained as a part of the calibration procedure, the view of the person was synthesized using the (10). Some of the synthesis results are shown in Figure 5, along with the images from the original video sequences. Note that the height of the synthesized silhouette is almost constant similar to the true zero azimuth case shown in Figure 5(a). It is also instructive to look at the width profile (defined as the number of pixels in each row between the extremities of the binarized silhouette) of one person plotted as a function of time as shown in Figure 6. The lower halves of these width plots correspond to the leg regions. In Figure 6 both the foreshortening and the effect of viewing direction on the leg-swings can be observed. In particular it can be seen that the leg swing as observed from a non-canonical view is smaller than what it would be from the exact side-view. Usage of such an unnormalized gait sequence for recognition will give poor results. Our method provides a systematic way of handling both these effects as can be seen from the width vector of the synthesized images in Figures 6(c) and (d).

For the case when $\theta = 45$ we find that in the torso region, the reconstructed silhouette is broader than the orig-



(a) (b) (c) (d) (e) (f) (g)

Figure 5: (a) represents different stances of a person walking parallel to the camera; (b) (d) and (f) represent different stances of a person walking at angles 15, 30 and 45 degrees to the camera; (c) (e) and (g) represent side-views synthesized from original videos where the person walks at angles of 15, 30 and 45 degrees to the camera.

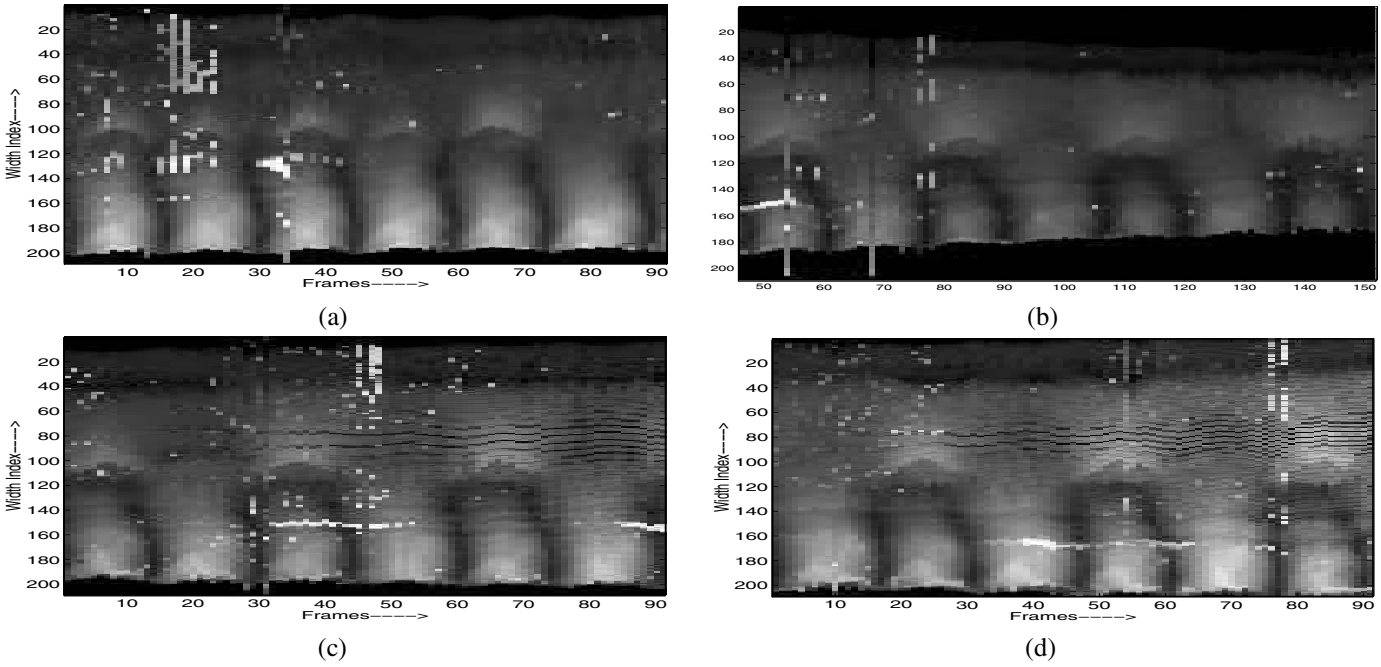


Figure 6: Width profile as a function of time for (a) Canonical View ($\theta = 0$);(b) Unnormalized sequence for $\theta = 45$;(c) and (d) Synthesized non-canonical Views for $\theta = 30$ and $\theta = 45$ respectively.

inal. The reason for this is the limitation of the planarity assumption for the torso region. For non-canonical views, parts of the torso unseen in the canonical view, appear. To appreciate this better, consider that we approximate the torso as a rectangular block. In the canonical view, just one face of this block is visible. For non-canonical views parts of the other faces of this block are visible too. The synthesis algorithm, which interprets this as a plane, renders a broader reproduction of the torso part. Notice however that this effect is somewhat lesser in the leg portions of the silhouette. As can be seen from the Figure 6, the lower parts are clearly distinguishable and similar for the different angles, though the upper halves of the plots become more and more noisy as the value of θ increases. In order to study the performance of gait recognition on the synthesized images we used a rather simple variant of the baseline gait recognition algorithm [15]. Our gallery consists of people walking at $\theta = 0$, i.e. the canonical view. The probes are video sequences where people walk at arbitrary angles θ. We take N contiguous boxed images of a person i in the gallery when he/she is walking at an azimuth $\theta = 0$. For every image of the probe j transformed to the zero azimuth from the azimuth θ_x, we compute the similarity matrix $S^{\theta_x} = s^{\theta_x}(i,j)$, where

$$s^{\theta_x}(i,j) = \sum_{k=1}^{M_j} \text{maxcorr}(B_k^{\theta_x,j}, \mathcal{A}_N^i), \qquad (14)$$

and $B_k^{\theta_x,j}$ refers to the kth image in the sequence synthesized from θ_x for the probe j. $\mathcal{A}_N^i = \{A_1^i, \cdots, A_N^i\}$ is the set of N contiguous images for the zero azimuth for gallery person i, and

$$\text{maxcorr}(B_k^{\theta_x,j}, \mathcal{A}^i{}_N) = \max_l \frac{Num(A_l^i \bigcap B_k^{\theta_x,j})}{Num(A_l^i \bigcup B_k^{\theta_x,j})}.$$

Besides taking the usual binary correlation in (14), we also computed the similarity matrices for just the lower half of the bounding box. This is approximately equivalent to considering just the leg portion of the body. This is motivated by the fact that the planarity assumption is more strictly adhered to in the leg portion than in the torso. Gait recognition performance can be improved further by fusing other static cues about the person, such as height. We fuse height information with the leg dynamics by scaling each entry $s(i,j)$ of the similarity matrix by the corresponding height ratio, $\max(\frac{h(i)}{h(j)}, 2 - \frac{h(i)}{h(j)})$.

The similarity matrices, yield as a by-product a quantitative assessment of the quality of the synthesized images as

$$Q^\theta = \frac{1}{P} \sum_{i=1}^{P} S^\theta(i,i), \qquad (15)$$

for each $\theta = \theta_x$ and P persons. This is plotted as a function of θ in Figure (7). The cumulative match characteristics [15] are shown in Figure (9) for the full body, leg only and leg and height fusion cases. The rise of the solid curves (representing the leg dynamics, with or without height fusion) is faster

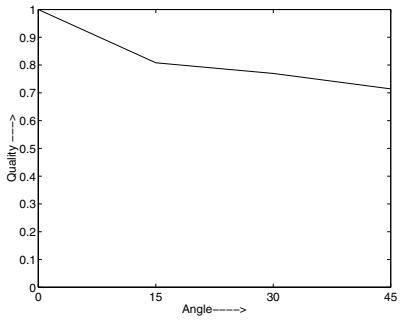


Figure 7: Quality degradation of the synthesized images as a function of angle.

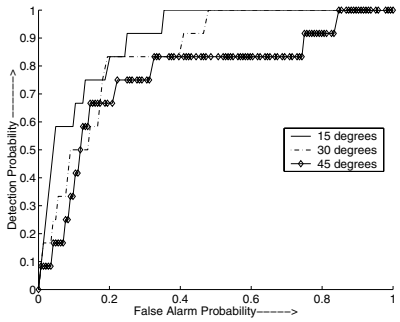


Figure 8: ROC curves for $\theta = 15, 30$ and 45 degrees for the case of leg and height fusion.

than the dotted ones (representing full body). This means that recognition performance is better when only the leg dynamics, instead of the whole body, is used. The reason for this is the incorrect broadening of the torso region. The performance in the case where height information is fused with leg dynamics is even better. Interestingly, [15] notes that the lower 20 % of the silhouette accounts for roughly 90% of the recognition. Similarly [2] showed that the gait recognition in the case when the subject was carrying a ball, viz. no upper body dynamics, the recognition rates were better. The fact that the gait recognition results are encouraging upto angles of 45 degrees allows us to hypothesize that it is possible to do reasonable human identification using gait with only two cameras (installed perpendicular to each other).

In order to study the efficiency of gait recognition with synthesized views, we compute the Receiver Operating Characteristic (ROC), which is a plot of the probability of detection (i.e. correct recognition), p_D, vs. the probability of a false alarm (i.e. false acceptance), p_F, for azimuth angles $\theta = 15, 30$ and 45 degrees. The plots are shown in Figure 8. The performance degradation with increasing θ can be understood from these plots. The ROC curves indicate that the proper detection threshold should vary with θ, so as to obtain a performance characteristic with small p_F and large p_D.

5 Conclusion and Future Work

In this paper, we have proposed a method for synthesizing arbitrary views of planar objects, and applying the synthesized views for gait recognition whenpeople are walking at any arbitrary angle to the camera. Our method used a perspective projection model and an optical flow based structure from motion model for estimating the azimuth angle of the original view from monocular video data. Thereafter, a video sequence at the new view was synthesized. The entire process was done in 2D, though 3D structure of the scene played an implicit role. A simple, yet accurate, camera calibration procedure was also proposed. Examples of synthesized views are presented. Preliminary results of gait recognition on a database of people was reported using these synthesized views. Development of appropriate gait recognition algorithms for people walking at arbitrary angles is one of our future research directions. Though the method has been explained from the motivation of the gait recognition problem, it has important applications in other areas too, like multimedia and video processing. That, too, forms a part of our future research into this problem.

Appendix

Proof of Equation (2):

Consider Equation (1) and the perspective projection model. Then,

$$x = f\frac{X}{Z} = f\frac{X}{Z_0 + Tan(\theta)X}, \tag{16}$$

$$y = f\frac{Y}{Z} = f\frac{Y}{Z_0 + Tan(\theta)X}. \tag{17}$$

Dividing Equation (17) by (16) we get (except for the few degenerate points where the denominator is zero),

$$\frac{y}{x} = \frac{k}{X}. \tag{18}$$

Now consider Equation (16):

$$\begin{aligned} & Z_0 x + Tan(\theta)Xx = fX, \\ \text{i.e.} \quad & Z_0 x = -(Tan(\theta)x - f)X, \\ \text{i.e.} \quad & \frac{x}{X} = \frac{f - Tan(\theta)x}{Z_0}. \end{aligned} \tag{19}$$

Substituting Equation (19) in (18), we get (2).

Proof of Equation (10):

Using Equations (8) and (9), we get

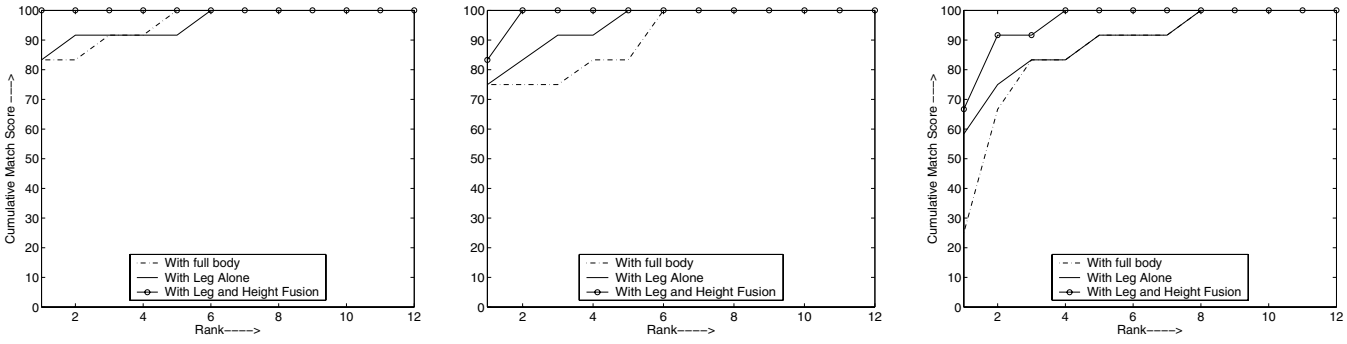


Figure 9: Cumulative Match Characteristics for Synthesized images (solid line represents the full body used for matching, dash-dotted line represents the case where only the leg is used for matching) for (a) $\theta = 15$ (b) $\theta = 30$ and (c) $\theta = 45$.

$$X_0 = X_\theta Cos(\theta) + Z_\theta Sin(\theta) \quad (20)$$

$$Y_0 = Y_\theta \quad (21)$$

$$Z_0 = -X_\theta Sin(\theta) + Z_\theta Cos(\theta). \quad (22)$$

Under perspective projection,

$$x_0 = f\frac{X_0}{Z_0}, y_0 = f\frac{Y_0}{Z_0} \quad (23)$$

$$x_\theta = f\frac{X_\theta}{Z_\theta}, y_\theta = f\frac{Y_\theta}{Z_\theta.} \quad (24)$$

Substituting from (20), (21) and (22) in (23), we get

$$x_0 = f\frac{X_\theta Cos(\theta) + Z_\theta Sin(\theta)}{-X_\theta Sin(\theta) + Z_\theta Cos(\theta)}, \quad (25)$$

$$y_0 = f\frac{Y_\theta}{-X_\theta Sin(\theta) + Z_\theta Cos(\theta)}. \quad (26)$$

Substituting for X_θ and Y_θ from (24) yields Equation(10).

References

[1] J. Cutting and L. Kozlowski, "Recognizing friends by their walk:gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, pp. 353–356, 1977.

[2] A. Kale, N. Cuntoor, and R. Chellappa, "A framework for activity-specific human recognition," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (Orlando, FL)*, May 2002.

[3] L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification," *Proceedings of the IEEE Conference on Face and Gesture Recognition*, pp. 155–161, 2002.

[4] P.S. Huang, C.J. Harris, and M.S. Nixon, "Recognizing humans by gait via parametric canonical space," *Artificial Intelligence in Engineering*, vol. 13, no. 4, pp. 359–366, October 1999.

[5] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," *Proceedings of IEEE Conference on Face and Gesture Recognition*, May 2002.

[6] O.D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[7] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[8] A.F. Bobick and A. Johnson, "Gait recognition using static activity-specific parameters," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[9] G.Shakhnarovich, L.Lee, and T.Darrell, "Integrated face and gait recognition from multiple views," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, December 2001.

[10] Vishwjit Nalwa, *A Guided Tour of Computer Vision*, Addison-Wesley, 1993.

[11] D. Cunado, J.M. Nash, M.S. Nixon, and J. N. Carter, "Gait extraction and description by evidence-gathering," *Proc. of the International Conference on Audio and Video Based Biometric Person Authentication*, pp. 43–48, 1995.

[12] M.P. Murray, A.B. Drought, and R.C. Kory, "Walking patterns of normal men," *Journal of Bone and Joint surgery*, vol. 46-A, no. 2, pp. 335–360, 1964.

[13] R.T. Collins Y. Liu and Y. Tsin, "Gait sequence analysis using frieze patterns," *Proceedings of the European Conference on Computer Vision*, vol. 2, pp. 657–671, 2002.

[14] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *FRAME-RATE Workshop, IEEE*, 1999.

[15] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. W. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," *Proc of the International Conference on Pattern Recognition*, 2002.