International Journal of Computer Vision 71(2), 143–160, 2007

DOI: 10.1007/s11263-006-8575-4

Pedestrian Detection via Periodic Motion Analysis

YANG RAN, ISAAC WEISS, QINFEN ZHENG AND LARRY S. DAVIS

Center for Automation Research, University of Maryland at College Park, College Park, MD 20742-3275

rany@cfar.umd.edu

weiss@cfar.umd.edu

qinfen@cfar.umd.edu

lsd@cfar.umd.edu

Received March 24, 2005; Revised March 29, 2006; Accepted March 29, 2006

First online version published in June, 2006

Abstract. We describe algorithms for detecting pedestrians in videos acquired by infrared (and color) sensors. Two approaches are proposed based on gait. The first employs computationally efficient periodicity measurements. Unlike other methods, it estimates a periodic motion frequency using two cascading hypothesis testing steps to filter out non-cyclic pixels so that it works well for both radial and lateral walking directions. The extraction of the period is efficient and robust with respect to sensor noise and cluttered background. In order to integrate shape and motion, we convert the cyclic pattern into a binary sequence by Maximal Principal Gait Angle (MPGA) fitting in the second method. It does not require alignment and continuously estimates the period using a Phase-locked Loop. Both methods are evaluated by experimental results that measure performance as a function of size, movement direction, frame rate and sequence length.

Keywords: pedestrian detection, periodic motion, frequency estimation, cyclic gait pattern, phase-locked loop

1. Introduction

1.1. Research Motivation

Pedestrian detection is critical in many surveillance systems in order to detect intrusions. Recently there has been a growing interest in using infrared cameras for human detection by robot vision techniques because of the sharply decreasing prices of such cameras and their ability to work in low light condition. Sensor noise and target pose variations present challenges. Previous systems for automatic detecting people by shape and/or motion using thermal infrared cameras are sometimes confused by moving foliage, building windows, traffic signs and more.

For close range pedestrian detection such as driving assistance systems, shape information can be reliably extracted. But for targets in middle or far range, it is no longer dependable and motion cue has to be integrated. The goal of this work is to develop a general motion based pedestrian detector in low or regular lighting situations, where previous shape based methods exhibit high false alarm rate.

1.2. Algorithm Overview

Object motions that repeat themselves are common in both nature and man-made environments. Many real-life motions are periodic such as the wings of flying birds, a rotating fan and so on. Most human locomo-

tory motions (e.g., walking, running, skipping, shuffling) are also periodic in the frame of reference that moves with the person. Knowing that an object's motion is periodic is a strong cue for object and action classification at a distance. Natural repeating motions tend not to be perfectly regular, i.e., the period varies slightly from one cycle to the next, or from one body part to another. For human gait, different parts of the human body share approximately the same period.

The two methods proposed here are based on periodic motion. One is a bottom-up approach based on hypothesis testing over periodograms and the other relies on global gait fitting. The first one is designed to be computationally efficient by identifying periodicity in pixels. The second is based on a gait feature called Maximal Principal Gait Angle (MPGA). It is insensitive to alignment error and does not require segmentation but it is more computationally expensive. The two methods can be used individually or in a combined manner.

We initialize target candidate by independent motion detection (Zheng and Chellappa, 1991) or by a method that uses a hierarchal shape structure, reported in Nanda and Davis (2002), and track them with a method reported in Zhou et al. (2004). The targets are specified by bounding boxes in each frame. We focus on periodic motion because gait characterizes pedestrians and is more reliable when the targets are observed at a distance. Whether using only period of motion or both motion and shape, the core task is how to efficiently use them. We design our methods to detect human:

1. In different poses and from various distances.
2. With a stable statistical performance.
3. With efficient implementation.

The outline of this paper is as follows. Section 2 discusses related work on pedestrian detection. In Sections 3 and 4 we discuss the pixel level based and Maximal Principal Gait Angle based methods with preliminary experiments. Section 5 provides a detailed analysis of sensitivity to several key factors. Finally Section 6 compares our new methods with other algorithms and provides conclusion.

2. Related Work

In recent years, automatic pedestrian detection in video has become an active research area in computer vision (Collins et al., 2000; Maybank and Tan, 2000). This task is especially difficult for video from moving platforms and low quality sensors for situation awareness applications. Some of the difficulties in these applications are (1) non-rigid kinematics of pedestrians; (2) targets pose and distance change; (3) cluttered backgrounds and low video quality; and (4) arbitrary camera motion. Reviews of some of the prior research on this topic can be found in Gavrila (1999) and Wang et al. (2003). Useful criteria for classifying pedestrian detectors are the cues they use such as shape or motion.

Examples of algorithms in the first category can be found in Gavrila (1999) with learning tools such as wavelets (Oren et al., 2003), neural networks (Zhao and Thorpe, 2000) and others. Nanda and Davis (2002) builds a probabilistic shape hierarchy to achieve efficient detection at different scales. The method in Hogg (1983) and Rohr et al. (1994), uses handcrafted human models for pedestrian detection, but requires segmentation into body parts which is very difficult. A system (Pai et al., 2004) proposed by Pai et al. recognizes pedestrians by measuring the distance between leg silhouette after background subtraction, which is not very applicable to moving platforms. Lipton et al. (1998) uses a skeleton based 'star' model to identify humans, which also depends on the extraction of a foreground mask. Another approach involves extracting low-level features such as edges or responses to filter banks, and using standard pattern classification techniques to determine the presence of a pedestrian as in Papageorgiou et al. (1998), where the authors extract wavelet features and then use a SVM to classify. Fang et al. (2003) compares the multi dimensional features between visible and infrared images and uses vertical projections of bright pixels specifically for infrared sensors. Objects in the background clutter such as windows, traffic signs and moving foliage often confuse shape based methods leading to high false alarm rates at acceptable detection rates. Besides, shape based detectors work better for close range targets than those far away.

In the motion based category, the gait feature is analyzed based on pixel-wise or region-based oscillations. The overall statistical periodic behavior provides classification. For example, Little and Boyd (1998) use a Discrete Fourier Transform (DFT) to measure pixel oscillations. Tsai et al. also describes a similar method using DFT to extract pixel period in Tsai et al. (1994). Efros et al. (2003) identifies the cyclic motion in the optical flow domain. Liu, and Picard (1998) examine the pixel oscillations over the XT plane to extract the fundamental frequency

of gait. Seitz and Dyer (1997) present a novel notion for repeating motion: periodtrace to detect motion trends. Boyd (2004) uses vPLLs (video Phase-Locked Loops) to measure the period contained in every pixel due to gait. Allmen and Dyer (1991), propose an approach to measure periodicity using a curvature scale space at each pixel. Polana and Nelson (1997), show that the recognition of human or animal locomotion can be done using low-level, non-parametric representations and matching against a spatio-temporal template of motion features. The main limitations of prior approaches in this second category are the sensitivity to alignment as well as to changing background. For videos acquired from moving platforms, accurate alignment is hard to achieve and hence pixel-wise periodicity can be corrupted. A method that is closely related to this paper and motivates our work can be found in Cutler and Davis (2000). The authors look for the gait period by calculating a similarity matrix for every image pair in a sequence. The approach is computationally expensive and sensitive to background clutter. Furthermore, video sensors in infra-red band contain higher noise levels than in the visible band, which makes the similarity calculation easier to corrupt and good alignment harder to achieve.

Another Major trend in previous methods is to combine shape and motion. Some of them directly train the detector over shape and motion information simultaneously. For example, Viola's Adaboost detector cascade in Viola et al. (2003) is a real-time pedestrian detection algorithm for a static camera. It was trained using patterns of frame difference as well as the static shape features. Because of the static camera, those regions which have human-like shapes such as windows, stop signs and trees etc., are filtered out as non-moving background by preprocessing and do not enter the classifier cascade.

There are many multi-stage systems to detect pedestrians by using different cues at different steps. One cue (shape or contour) is used for initial detection and others (motion, gait) are used as verification. For instance, Curio et al. (2000) proposed a method for the detection, tracking, and final recognition of pedestrians crossing a moving observer's path. The initial detection process is based on texture analysis and geometric features. The classification is obtained by a temporal analysis of the walking process. However their algorithm "is restricted to the detection of pedestrians that cross the road" (Curio et al., 2000) and hence is not general enough for robot's situation awareness such as intrusion detection.

Another class of methods tried to fit a 3D human model to 2D image to find the articulation. For example, A. Broggi et al. (2000) compare several approaches relying on the matching between image features and model features stored in a predefined or dynamically updated database. A challenge to using model based fitting is that the complicated nature of human gait and variations of pose requires a large number of DOF. Hence it is very difficult to map the non-rigid dynamics. The authors in Broggi et al. (2000) also concluded that "it is difficult to obtain an exhaustive model set that gives good results on very different scenes".

3. Pedestrian Detection by Periodogram

Although gait period is widely used to analyze walking, few of the proposed method are suitable for detecting pedestrians. The reason lies in two facts. One is the high complexity and the other is the sensitivity to pose change. The state-of-art still lacks of a simple, reliable and efficient method.

Different known forms of frequency detection are studied. Phase-Locked Loop and autoregressive moving average models (ARMA) have been traditionally used as a convenient tool for estimating frequencies of sinusoidal time series data. In Quinn and Hannan (2001), it is estimated from a second order ARMA model in an efficient fashion. Other frequency estimation techniques are parametric minimum entropy and subspace methods such as multiple signal classification (MUSIC) method. The challenge here is how to use them efficiently to address the very specific problem of pedestrian detection.

3.1. Pixel Periodicity Extraction

Objects with periodic motion are similar in many aspects, including appearance, motion flow, and shape (Cutler and Davis, 2000). However, environmental conditions (such as lighting, shadows, cluttered backgrounds) and internal variations (pose, shape) pollute the signal and adversely affect detection. Besides, periodic behavior may only exist in some portion of an object. We carefully studied and implemented (Cutler and Davis, 2000) and found that focusing on the overall similarity between images may fail due to the large number of non-periodic pixels. In this section,

we describe our new algorithm which first tests the periodicity on a pixel-wise level and then analyzes the overall distribution of periods.

We start from a sequence of bounding boxes. Preprocessing is carried out to adjust for small alignment errors and to normalize for size. Assuming that the intensity at a pixel (i, j) is a sum of a periodic signal $M(i, j)(t)$ and additive noise $n(t)$ while a non-periodic pixel contains only noise, we expand the signal in a Fourier domain.

$$\begin{aligned} x_t(i, j) &= M_t(i, j) + n(t) \\ &= \mu(i, j) + \sum_{k=1}^{\infty} [\alpha_k \cos(k\omega t) \\ &\quad + \beta_k \sin(k\omega t)] + n(t) \end{aligned} \tag{1}$$

where $n(t)$ is noise, $M_{(i,j)}(t)$ is the oscillatory signal and t is time.

To simplify the equations we linearize them and only use the first 3 coefficients to approximate the original signal, hence:

$$\begin{aligned} x_t(i, j) \approx \mu(i, j) &+ \alpha(i, j)\cos(\omega t) \\ &+ \beta(i, j)\sin(\omega t) + n(t), \end{aligned} \tag{2}$$

This approximation enables convenient estimation with low computation cost. With N observations at $t = 0, 1, \ldots, N-1$, we have N linear equations and 3 unknown parameters to estimate. We rewrite the N equations in a matrix form as:

$$A(\omega) \begin{bmatrix} \mu(i, j) \\ \alpha(i, j) \\ \beta(i, j) \end{bmatrix} = b \tag{3}$$

where

$$A(\omega) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \cos(\omega) & \sin(\omega) \\ \vdots & & \vdots \\ 1 & \cos((N-1)\omega) & \sin((N-1)\omega) \end{bmatrix}$$

and

$$\begin{aligned} b = [&x_0(i, j) - n(0), x_1(i, j) - n(1), \ldots, x_{N-1}(i, j) \\ &- n(N-1)]^T. \end{aligned}$$

For a given period, or frequency ω, the least square estimator for the parameters is given by:

$$\begin{bmatrix} \hat{\alpha}(i, j) \\ \hat{\beta}(i, j) \\ \hat{\mu}(i, j) \end{bmatrix} = (A(\omega)^T A(\omega))^{-1} A(\omega)^T b. \tag{4}$$

The residual sum of squares, for a given ω, is calculated as:

$$\begin{aligned} &RSS(\omega) \\ &= \sum_{t=0}^{N-1} (x_t(i, j) - \hat{x}_t(i, j))^2 &(5)\\ &= \sum_{t=0}^{N-1} (x_t(i, j) - \hat{x}_t(i, j)) x_t(i, j) &(6)\\ &= \sum_{t=0}^{N-1} x_t^2(i, j) - b^T A(\omega)(A(\omega)^T A(\omega))^{-1} A(\omega)^T b &(7)\\ &= \sum_{t=0}^{N-1} [x_t(i, j) - \bar{x}]^2 - \left\{ [A(\omega)b]^T \begin{bmatrix} \hat{\alpha}(i, j) \\ \hat{\beta}(i, j) \\ \hat{\mu}(i, j) \end{bmatrix} - \bar{x}^2 \right\}. \end{aligned}$$

The second derivation above results from the orthogonality between the original signal and the estimation error under the Gaussian noise assumption (Quinn and Hannan, 2001). Thus, the estimate of the period for the object sequence should be that ω which contributes a minimal $RSS(\omega)$ over all possible frequencies (periods). Notice that minimizing $RSS(\omega)$ is equivalent to maximizing the second term in (7) with respect to ω. This enables us to estimate the period directly. Moreover, $(A(\omega)^T A(\omega))^{-1}$ could be approximated as Quinn and Hannan (2001):

$$(A(\omega)^T A(\omega))^{-1} \approx \frac{1}{N} \begin{bmatrix} 1 & o(1) & o(1) \\ o(1) & 2 & o(1) \\ o(1) & o(1) & 2 + o(1) \end{bmatrix}, \tag{8}$$

where $o(1)$ denotes terms tending to zero, so the cost function is simplified to:

$$I_{i,j}(\omega) = \sum_{t=0}^{N-1} \left\{ [A(\omega)b]^T \begin{bmatrix} \hat{\alpha}(i, j) \\ \hat{\beta}(i, j) \\ \hat{\mu}(i, j) \end{bmatrix} - \bar{x}^2 \right\} \tag{9}$$

$$\approx b^T \frac{1}{N} \begin{bmatrix} 1+2\cos 0\omega & \dots & 1+2\cos(N-1)\omega \\ 1+2\cos\omega+2\sin\omega & \dots & 1+2\cos(N-1)\omega\cos\omega+2\sin(N-1)\omega\sin\omega \\ \vdots & \vdots & \vdots \\ 1+2\cos(N-1)\omega+2\sin(N-1)\omega & \vdots & 1+2\cos^2(N-1)\omega+2\sin^2(N-1)\omega \end{bmatrix} b - N * \bar{x}^2 \tag{10}$$

Substituting b and expanding the result we obtain:

$$\begin{aligned} I_{i,j}(\omega) &= \frac{2}{N}\left\{\sum_{t=0}^{N-1} x_t(i,j)\cos(\omega t)\right\}^2 \\ &\quad + \frac{2}{N}\left\{\sum_{t=0}^{N-1} x_t(i,j)\sin(\omega t)\right\}^2 \\ &= \frac{2}{N}\left\|\sum_{t=0}^{N-1} x_t(i,j)e^{i\omega t}\right\|^2 \end{aligned} \tag{11}$$

The quantity $I_{i,j}(\omega)$ is referred to as the periodogram (Quinn and Hannan, 2001) of the pixel. It has been proved in Quinn and Hannan (2001) that the maximizer of the periodogram over all frequencies cannot be improved on, in terms of asymptotic variance, by any other technique without extensive knowledge of the distribution of the noise $n(t)$.

3.2. *Periodicity Verification*

The periodogram can be regarded as the signal response of the system at different frequencies. We verify the existence of a period via hypothesis testing to confirm the existence of a well-pronounced peak in the periodogram, i.e. in the strength as a function of the frequency, for each pixel. By filtering out non-periodic or stationary pixels, we are able to focus on periodic pixels only. Given the signal model as Eq. (2), where the noise is Gaussian, we perform the following statistical hypothesis test for every pixel:

$$\mathcal{H}_0 : \lambda(i,j) = 0 \ \text{ vs. } \ \mathcal{H}_\lambda : \lambda(i,j) > 0. \tag{12}$$

where λ is the oscillatory amplitude for function $M_{(i,j)}(t)$ at the pixel (i, j). It could be approximated from Eq. (2) as $\lambda \approx (\alpha^2+\beta^2)^{1/2}$. H_0 stands for non-periodic pixels (amplitude is zero) and H_λ for the periodic pixels. We use a Bayesian decision rule based on the posteriori probability

$$P(H_\lambda|X) <> P(H_0|X) \tag{13}$$

Using Bayes theorem, the decision rule can be transformed as:

$$P(X|H_\lambda)P(H_\lambda) <> P(X|H_0)P(H_0) \tag{14}$$

Under a Gaussian noise assumption, the test rejects the null hypothesis $\mathcal{H}_0$ for large values of the ratio of the maximized likelihood under $\mathcal{H}_\lambda$ to the maximized likelihood under $\mathcal{H}_0$, i.e.:

$$-\frac{N}{2}\log\left(2\pi\sigma_\lambda^2\right) - \frac{N}{2}\log\left(2\pi\sigma_0^2\right) - \frac{N}{2}, \tag{15}$$

where

$$\sigma_\lambda^2 = \frac{1}{N}\sum_{t=0}^{N-1}[x_t(i,j)-\bar{x}]^2 - \max_{\omega_k\in\Theta} I_{i,j}(\omega_k)$$

and

$$\sigma_0^2 = \frac{1}{N}\sum_{t=0}^{N-1}[x_t(i,j)-\bar{x}]^2.$$

We reject $\mathcal{H}_0$ if $\sigma_\lambda^2/\sigma_0^2$ is too small, or equivalently,

$$r = \frac{\max_{\omega_k\in\Theta} I_{i,j}(\omega_k)}{\frac{1}{N}\sum_{t=0}^{N-1}[x_t(i,j)-\bar{x}]^2}$$

is too large (Cutler and Davis, 2000).

After performing the hypothesis testing, we are left with only the periodic pixels and the most likely periods for them. We then compute the histogram of these periods and look for the maximum period in this histogram. This is done using the same hypothesis testing method used above, giving us the period of the global object.

An example of periodicity extraction is shown in Fig. 1. After the testing, 'good pixels' are filtered out

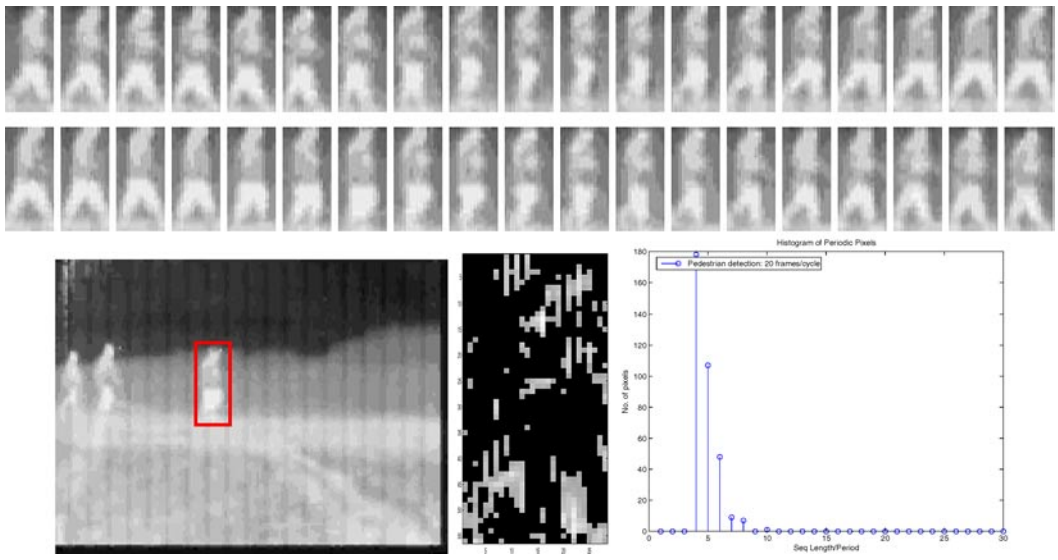

Figure 1. Left: illustration of periodicity extraction for an infrared surveillance video. Middle: filtered periodic mask for the pixels in the (red) bounding box. Right: a distinct peaks shows frequency corresponding to human gait.

as the non-black pixels in the middle mask image for the human on the left image. The higher intensity stands for a stronger periodicity for that pixel. The histogram is shown on the right, with a well-pronounced peak representing gait rate. We also display a complete cycle of walking sequence as reference at the top of the figure. Because of the low resolution of the infrared camera, only a small portion of it shows periodicity. Most of the periodicity comes from the region around the lower part (legs) of the human body, which captures most of the cyclic motion for human gait. In spite of the difficulties, the algorithm still correctly detects a distinct peak at the periodic frequency in Fig. 1.

As an extension of such two-stage testing method, we apply shape constraints to lower the false positives. For example, the symmetry and relatively fixed location of periodic pixels for human legs could help us discriminate pedestrian walk from other periodic motion.

3.3. Experimental Results: Pixel-Wise Method

Two datasets were tested. One was obtained using infrared cameras and the other employed color ground-based sensors. The infrared data (HONDA dataset, UMD dataset I) consists of 40 sequences ranging from 3 minutes to 7 minutes with more than 80 objects (55 pedestrians and 35 vehicles) from both static and moving sensors. The other dataset (UMD dataset II) contains 55 color/gray sequences with 90 pedestrians and is also captured from moving and static platforms. They include typical scenes such as parking lots, roads and other urban settings containing pedestrians varying in terms of size, speed, clothes and poses. We claim a successful detection for a object sequence only if the bounding boxes cover major portion of a human body and the motion based classification is correct.

3.3.1. Infrared Sensor. In Fig. 2, we illustrate the detection process for a more challenging sequence captured by a very low quality interlaced thermal sensor. The sensor blurs foreground target's contour and appearance with background, which makes most shape and motion based methods fail. For instance, the similarity based method such as Cutler and Davis (2000) yields a weak correlation matrix as shown in the left bottom image in Fig. 2. Each pixel in that matrix represents correlation between two frames. If the contrast is high enough, we will observe darker lines parallel to the diagonal, which is caused by the similarity between two images in the same gait phase. Although no periodicity is observed in this matrix, our method successfully filters out the 'good' periodic pixels (even a very small portion of the whole image sequence) for the correct gait rate.

To evaluate the accuracy of the method we compute the ROC (Receiver Operating Characteristics) curve.

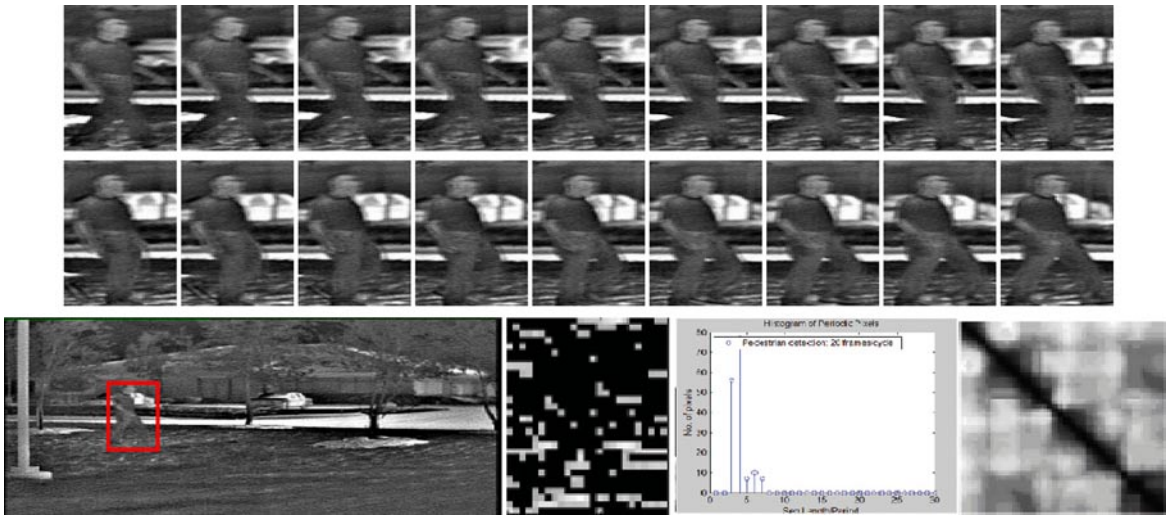

Figure 2. Detection pedestrian in low contrast infrared sequence. Top row: original object sequence; bottom row (from left to right): original image; mask generated by Hypothesis Testing; A peak in histogram showing gait rate; similarity matrix using (Cutler and Davis, 2000).

The ROC curve plots the false positive rate against the detection rate when the classification criterion is varied. The false positive rate is defined as the total number of false positive detections divided by the total number of objects in all sequences; the true detection rate is the ratio between total number of correct detections and total number of detections in all sequences. Since we use a whole sequence for each classification, we do not divide the above rates by frame number. Our classification criterion is the likelihood ratio (Eq. (12)). This criterion depends on other parameters, namely the frame length N and the noise variance σ_A^2, σ_0^2. We adjust only one of these parameters at a time to get different pairs of detection and false positives rate.

For infrared videos, our method maintains the detection rate above 80% with a false positives rate lower than 10% for both the static and the moving platforms.

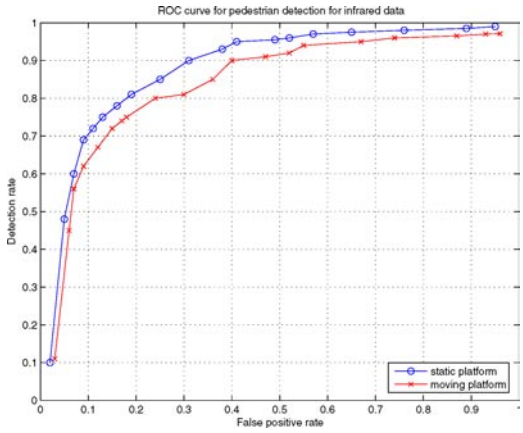

Figure 3. ROC analysis for infrared dataset from static/moving sensors.

3.3.2. Color/Gray Sensor. Figure 4 shows the the pixel-wise classification results for three representative pedestrians from the data set. Two are from the same sequence (static camera) with different views: lateral (across the camera) and radial (towards or away from the camera), and the other is from a moving sensor in lateral view. In all cases, the method detects the correct period and classifies the objects as pedestrians.

We plot the ROC curves for the static and moving platform for the UMD dataset II in Fig. 5. Compared to the results from the infrared data, we obtain higher performance in terms of detection rate at the same false negative rate. This is due to the higher contrast and lower sensor noise level.

To compare this method to the predecessor in Cutler and Davis (2000), we implement a version of the latter. The first advantage is that ours is more robust to the background clutter. For a situation-aware intelligent sentry or surveillance camera, background texture is inevitable for middle or near range object classification. Correlation over all pixels for an image pair includes background. When the background intensity is no longer uniform, the correlation score for image pairs in the same gait phases will decrease,

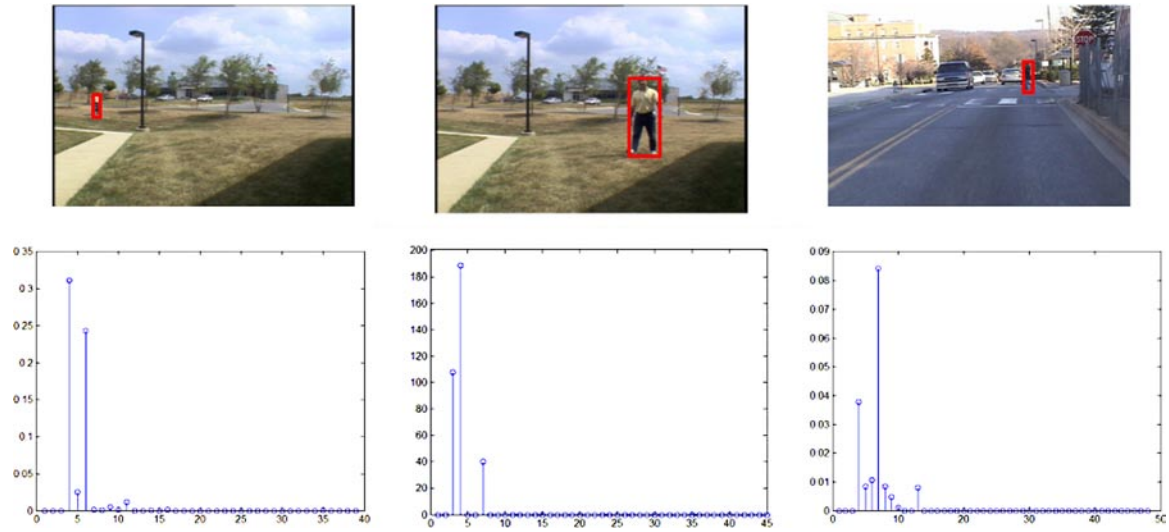

Figure 4. Period detection in Color/gray sensor dataset. top: first frames of original sequences; bottom: corresponding histogram.

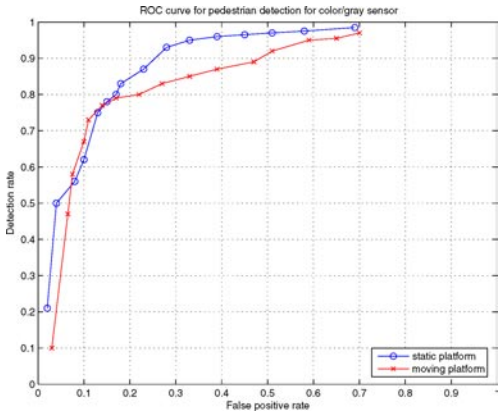

Figure 5. ROC analysis for color/gray dataset from static and moving sensors.

which is demonstrated in Table 1. We calculate the maximum-to-variance ratio of one row in the correlation matrix for a sequence in Fig. 4. By systematically increasing the bounding box size, we include more and more background clutter into correlation calculation. The same ratio in the histogram from our new method is also calculated. In Table 1, with the increase of the box sizes and hence the portion of background clutter, the correlation score ratio decreases sharply for a method like (Cutler and Davis, 2000), while ours successfully filters out the periodic pixels and extracts the gait rate correctly.

The second advantage is that ours saves computational power and is amenable for parallel hardware implementation. Assuming an object sequence with normalized size X *by* Y and frame number N, the major operation in periodogram based estimation is FFT and hence is $N\log(N)$. The total number of operations will be $X \cdot Y \cdot N\log(N)$. In order to have a robust detection, our version of the method in Cutler and Davis (2000) calculates the full correlation matrix to detect the lattice pattern. Each correlation between two images adds up to $X \cdot Y$ operations. The overall correlation matrix requires $X \cdot Y \cdot C_2^N = X \cdot Y \cdot 2/N(N-1)$. The computational saving is the ratio of the above two: $N\log(N)/C_N^2$, or $(N-1)/(\log N \cdot 2)$ times faster.

3.3.3. Alignment. This method works better when we have accurate alignment of the frames since it uses pixel-wise temporal information. Current detection and tracking algorithms cannot provide error-free alignment. We selected a subset sequences (more than 150) with a length of 64 frames and a box size around $40 * 80$. To obtain a quantitative estimate of the error in periodicity estimation resulting from misalignment, every box of the probing sequence is shifted in both directions by a quantity (dx, dy), which obeys a zero-mean uniform distribution $U(0, \sigma)$. The periodicity is re-calculated with various σ. We list in Table 2 the detected period for one sequence as a function of the shift parameter σ.

With the increase of alignment error, the hypothesis testing deteriorates. It classifies the pedestrian as

Table 1. Comparison of sensitivity to background clutter

Object size increasing ratio (%)	0	2.5	5	10	15
Correlation score ratio for Cutler and Davis (2000)	8.63	4.32	3.50	1.93	2.02
Maximum-to-mean ratio for our method	12.9	11.3	14.5	10.0	9.8

Table 2. Comparison of classification at different alignment error levels

σ	0	2.0	4.0	6.0	8.0	10.0
Period	34	32	35	30	–	–
Classification	Human	Human	Human	Human	Non-Human	Non-Human

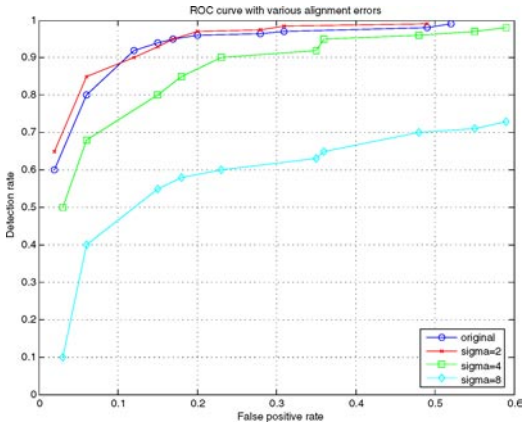

Figure 6. Pedestrian detection with various alignment error in color/gray sensor data.

non-human when the shift range goes beyond 6 pixels. Furthermore, we plot the ROC curves in Fig. 6 for a subset of our data.

The result shows that the method is susceptible to large alignment errors and only works fine when the shift level (or alignment error) is within a reasonable range. In order to reduce the sensitivity to alignment, we present another method based on an alignment-free measurment of the cyclic property.

4. Periodicity Analysis by Model Fitting

4.1. Cyclic Motion

Beside the speed of the motion (such as walking or running), the concept of gait also describes the style or manner in which a human moves. Periodicity differentiates a pedestrian from other non-periodic motions such as moving vehicles or foliage in winds, while gait differentiates humans from other cyclic moving objects such as machines or animals. By comparing different

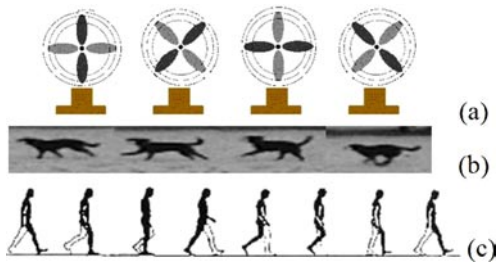

Figure 7. Different cyclic motion (a) rotating fan, (b) running dog, (c) walking pedestrian.

kinds of periodic motion extracted from various objects illustrated in Fig. 7, we can identify a distinctive pattern that applies only to pedestrians. In particular, the swing of the two legs characterizes this pedestrian-specific oscillation.

We start by investigating the kinematics of human gait from a synthesized sequence as in Fig. 8. The figure displays a complete cycle of a pedestrian's legs. We develop a computationally efficient human motion analysis algorithm based on the twin-pendulum model introduced in Aggarwal and Cai (1999) and Curio et al. (2000). The twin-pendulum model has a very simple form that captures the inherent nature of gait. It focuses on the motion of the legs. Each leg is represented by two jointed cylinders. The diameters of the cylinders are constant but the lengths of the cylinders are changing over time.

4.2. Extraction of Motion Pattern

The discussion above suggests that we classify a moving object as a human by features related to cyclic motion pattern. But change of appearance, non-rigid deformation of human body and arbitrary motion of camera present challenges. What is a good feature to help us extract cyclic motion pattern unique to human? The answer involves two issues. First, good features should

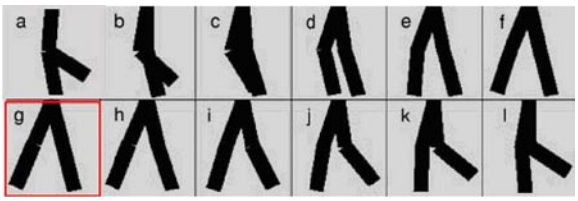

Figure 8. Motion signature in synthesized sequences (Curio et al., 2000). The one with (red) boundary represents the best fit using a twin-pendulum model.

be global and shape-based rather than pixel-based to reduce sensitivity to temporal alignment. Second, since precise shape extraction (segmentation) is very difficult, we favor a feature derived from the human contour instead of the contour itself. A closer look at Figs. 8 and 9 reveals that the relative angle between the two thighs can be used as such a feature. The Principle Gait Angle is defined as the angle between two legs during walking. But the non-rigid deformation and self-occlusion of two legs as well as the arbitrary pose makes it difficult to continually observe this angle in a complete stride. Instead we focus on a special case in which the two legs are maximally separated as in (g) in Fig. 8, enclosed in a box. We refer to it as the Maximal Principal Gait Angle (MPGA). This corresponds to a unique phase in the cycle in which the toe-to-toe distance approaches a maximum. The periodicity of the angle is a strong cue for detection. An example is given in Fig. 9. We apply an edge detector to pedestrians at different principle gait angles of walking. Only those with MPGA exhibit a salient angle pattern.

We next describe how to extract the critical Principle Gait Angle from a cluttered background. We first apply an edge operator (i.e. Canny operator) to the image. Then a Hough line detector scans the edge map and generates a list of the candidate lines in the image with length above a pre-specified threshold. We categorize these lines into pairs by checking symmetry and slope and choose the pair with longest average length to be our candidate in that frame. Finally we use a Bayes classifier to identify the occurrence of the MPGA. Intuitively MPGA should arise from line segments with sufficient length forming the gait angle and the angle should be related to the model and the pose. The distance between segments should fall into a narrow range. We formulate the detection of the MPGA in a Bayes linear classifier framework. An observation vector X is defined as

$$X = \{l, d, \alpha\} \tag{16}$$

where l is the average length of the twin-line, d is the center distance and α is the angle formed by that pair.

Figure 9. Principle Gait Angle in original and gradient images. Only those with MPGA exhibit a salient angle pattern.

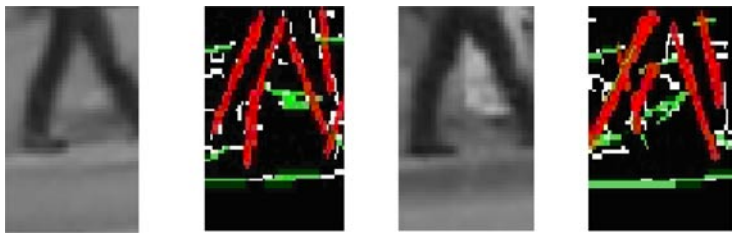

Figure 10. Illustration of twin-pendulum model fitting. White pixels: Edges; Green Pixels: detected lines by Hough Transform; Red Pixels (line segment pairs along legs): Fitted lines forming Principle Gait Angle.

The first two are normalized by the height of the object bounding box. However, in order to construct the likelihood ratio, the conditional probability must be in closed form for each class. In applications like ours, we have to estimate such distribution using samples from a training set. Since this is impractical, we use an approximation, namely a linear classifier. We are looking for variable V and v_0 such that $y = V^T X <> -v_0$ is the classifying function in this two class problem. When X is normally distributed, y is also normal and we outline the process to generate a linear Bayesian classifier from training set as follows (adapted from Fukunaga, 1990).

1. Compute the sample mean $\hat{M}_i$ and covariance matrix $\hat{\Sigma}_i$ of vector $\{l, d, \alpha\}$ from a manually labeled training set;
2. Calculate V for a given weight s by $V = [s\hat{\Sigma}_1 + (1-s)\hat{\Sigma}_2]^{-1}(\hat{M}_2 - \hat{M}_1)$;
3. Using the V computed as above, obtain $y_j^{(i)} = V^T X_j^{(i)}, i = 1, 2, 3...N.X_j^{(i)}$ is the j th i-class sample
4. the $y_j^{(1)}$ and $y_j^{(2)}$, which do not satisfy $y_j^{(1)} < -v_0$ and $y_j^{(2)} > -v_0$, are counted as errors.
5. Vary v_0 to find the v_0 which gives the smallest error
6. Vary s from 0 to 1 and repeat Step 2–5; choose the s giving smallest error as well as the corresponding V and v_0 as the discriminant function

We give some sample fitting results in a sequence of the lower part of a Pedestrian in Fig. 10.

Such a Bayesian classifier gives us a binary sequence representing the classification decision for each frame in the video sequence. That is, for an image with a positive detection we have a 1 in the binary sequence and 0 otherwise. Intuitively, the sequence should be quasi-periodic and its instantaneous frequency should be the gait rate. In fact, even with false alarms, we still can observe a strong periodic oscillation in such a sequence and a more accurate solution will be provided in the next section.

4.3. *Estimation of Period*

The motivation for this section is to integrate shape and appearance with motion, which is expected to give higher detection rate and fewer false alarms. Having detected the presence or absence of the Maximal Principal Gait Angle for each frame in a sequence, we can test the periodicity by the hypothesis testing methods used in the previous section. Phase Locked Loops (PLL) are a useful tool for this problem (Lindsey and Chie, 1986).

4.3.1. Phase-Locked Loop (PLL). A PLL, or Phase-Locked Loop, is basically a close-loop feedback control system, whose operation is based on the detection of the phase difference between the input and output signals of a voltage controlled oscillator (VCO). Phase-locked loops are widely used in communications. An introduction to PLL can be found in Blanchard (1976).

We use a software version of PLL as in Blanchard (1976). Fig. 11 shows the classic configuration. The phase detector is a device that compares two input frequencies, generating an output approximately proportional to their phase difference (if, for example, they differ in frequency, it gives a periodic output whose frequency is the difference frequency). Let's denote the reference signal frequency and the output of VCO frequency as f_{IN} and f_{VCO}. If f_{IN} doesn't equal f_{VCO}, the phase-error signal causes the VCO frequency to deviate in the direction of f_{IN}. If conditions are right, the VCO will quickly "lock" to f_{IN}, maintaining a fixed relationship with the input signal.

4.3.2. Recursive Period Estimation. We use the output of the previous stage, namely the binary sequence provided by fitting the Principle Gait Angles, as the input to the dPLL module. The method estimates both the frequency of the gait and its phase.

In more detail, the method for pedestrian detection has a cascade structure. It operates on the output of the model fitting algorithm. The input is a 0-1 sequences representing the critical phases corresponding to presence or absence of maximum toe-to-toe distances. This signal is passed through a low pass filter to remove high

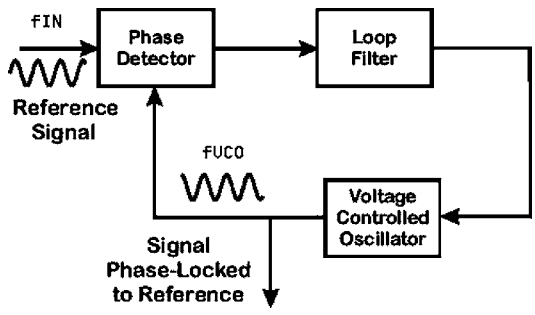

Figure 11. Diagram of a Digital PLL.

frequency components:

$$V_i(t) = Lowpass(V_1(t)) \tag{17}$$

Without loss of generality, we write the input signal and the output signal from the VCO as

$$V_i(t) = A \cdot \sin(\omega_i(t) + \theta_i),\ V_o(t) = \cos(\omega_o(t) + \theta_o) \tag{18}$$

If we use a multiplier as the phase detector, the signal after multiplication will be

$$V_{\mathrm{PD}}(t) = K \cdot A \cdot \sin(\omega_i(t) + \theta_i) \cdot \cos(\omega_o(t) + \theta_o) \tag{19}$$

where K is the gain of the phase detector (multiplier in our case). Furthermore, we could write it as:

$$\begin{aligned} V_{\mathrm{PD}}(t) = 1/2 \cdot A \cdot\ K \cdot + \theta_o]\{\sin[(\omega_i(t) + \omega_o(t)) + \theta_i \\ + \cos[(\omega_o(t) - \omega_o(t)) + \theta_i - \theta_o]\} \end{aligned} \tag{20}$$

When $\omega_o \approx \omega_i$, the first item in the above representation is attenuated by the low pass filter (inside the loop filter) in Fig. 11. The input of the VCO after low pass filtering can be approximated as

$$V_{\mathrm{VCO,IN}}(t) = 1/2 \cdot A \cdot K \cdot \sin(\theta_i - \theta_o) \tag{21}$$

When the phase difference is small enough, this equation can be simplified to

$$V_{\mathrm{VCO,IN}}(t) = 1/2 \cdot A \cdot K \cdot (\theta_i - \theta_o) \tag{22}$$

$\mathcal{V}_{\mathrm{VCO,IN}}$ is proportional to $\theta_i - \theta_o$. We can now explain how the dPLL locks the gait period. Suppose at first that the object's period is unknown. The initial frequency of the VCO output is set to a gait frequency guess ω_0 (20 frames/cycle). When the gait period (frequency ω_t of V_i) changes, the difference between V_o and V_i is detected by the phase detector which controls the $V_{\mathrm{VCO,IN}}$, causing the VCO frequency to deviate in the direction of ω_t. Hence, the period is estimated. Only when the rate falls into an interval representing a normal gait range will the object be classified as a pedestrian.

In practice, the dPLL loop is activated by the initial hypothesis test described in Sec. 3. Only when the testing detects a period for the first time will the dPLL module set to work on the following frames with the initial VCO frequency set to the detected period.

4.4. Experimental Results: Model-Based Method

We use a module reported in Nanda and Davis (2002) to initialize detection and the method reported in Zhou et al. (2004) to track bounding boxes for targets. We test the system over the Infrared and Color/gray datasets as for the pixel-based method.

In Fig. 12, we show the results based on tracking two pedestrians for 200 frames. The period for the first object is locked at a frequency of 32 frames/cycle, which corresponds to the gait rate. The second is locked at 24.0 frames/cycle. We plot the PLL VCO output vs. locking time in the second row to illustrate how fast the method adapts to the real signal. In Fig. 13, we present a sequence from the color/gray set. We track pedestrians for 150 frames and the PLL locks to the period after 40 frames.

In order to evaluate the performance of the detector in fitting the Maximal Principal Gait Angle, we plot the ROC curves in Fig. 14. We draw the curves for infrared and color/gray data set respectively using the shape matching method reported in Nanda and Davis (2002) and compare the performance improvement when cascading with the our new fitting method. In this experiment, the training set is composed of 827 positive samples (boxes containing a pedestrian with maximum toe-to-toe distance) and 3270 negative samples (images containing pedestrians in other gait phases, other objects or background). For each data set, two cases are compared. One is the direct results obtained purely by using a shape hierarchy (Nanda and Davis, 2002) in every frame and the other is cascading matching and cyclic motion verification as initialization and verification modules respectively. As we can observe, the cascaded detectors successfully use the gait angle to identify true pedestrians with a higher detection rate under the same false positive rate with the help of MPGA fitting and PLL gate rate estimation.

5. Sensitivity Analysis

Two methods have been proposed in the paper. In this section we study the sensitivity of detection accuracy to several important independent variables. The variables considered here are object size (determined by the distance to the camera), signal length,

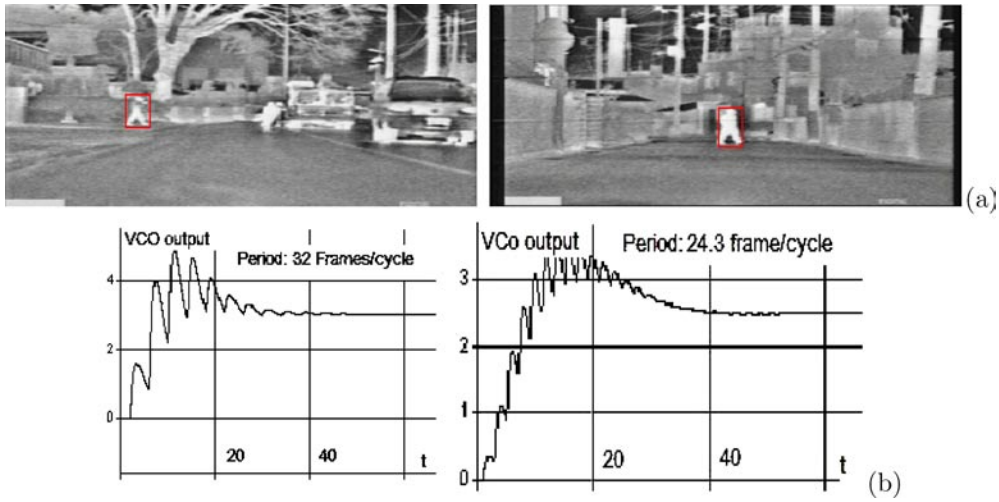

Figure 12. PLL VCO output voltage vs. locking time for infrared data (a) Representative frames. (b) VCO output.

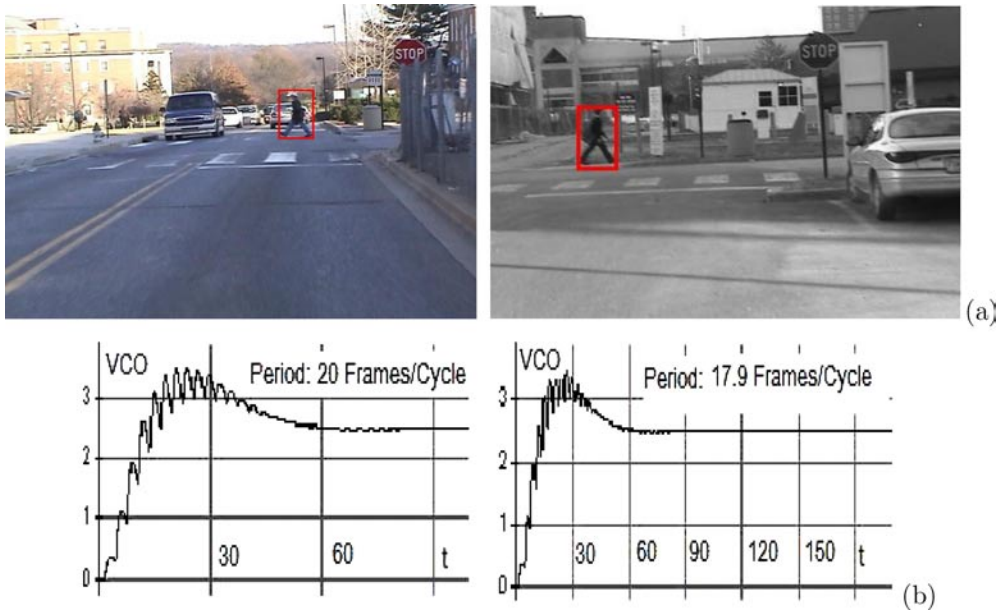

Figure 13. PLL output voltage vs. locking time for color/gray sensor. (a) Representative frames. (b) VCO output.

frame rate and movement directions (camera viewing angles).

5.1. Object Size

Sensitivity to object size is important for judging a system's ability to detect targets in various distances. We present the result for a subset of the two data sets with different down-sample ratios in Fig. 15. The object size from this subset of sequences is greater than 3200 pixels (based on a bounding box of 40∗80). We obtain consistently correct results for the first method even when the target size is gradually reduced to 10×20. Notice that during the down-sampling, the detected period does not change. This demonstrates that the pixel based method exhibits only a weak dependence on object size. But the second method works on relatively larger objects, which is not surprising since the MPGA is extracted from edges.

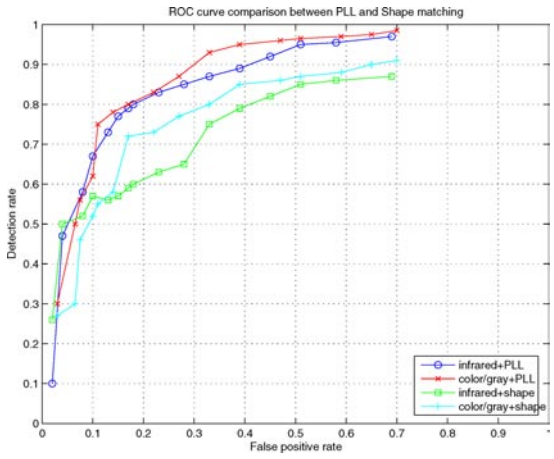

Figure 14. ROC comparison between PLL + shape and shape matching for both datasets.

This results also give us a promising way to reduce the computational cost when using the pixel-based method. By reducing the object size by 2 or even 4, it saves time while maintaining the performance.

5.2. *Video Sequence Length*

An interesting issue is the minimal sequence length needed to reliably extract the gait rate. We would like the detector to make a decision with minimal delay and to keep a reasonable detection rate. For shorter lengths, the signal may not be long enough to exhibit periodicity. When the length increases, tracking is harder and the change of external variables such as pose, size, lighting etc will corrupt the cyclic signal.

Suppose we estimate the frequency directly from the periodogram output without any further processing (Quinn and Hannan, 2001). If the true period for a pixel is ω, and it falls into two adjacent bins: k and $k+1$,

$$\omega \in [k \times 2\pi F_{\text{sample}}/N, (k+1) \times 2\pi F_{\text{sample}}/N] \tag{23}$$

where F_{sample} is the sample frequency and N is the signal length, we will have a bias up to the width of the bin. Hence this method requires longer sequences for higher resolution. But it is not always easy to achieve long tracking of small objects in low quality video from a moving platform. We test the first method for object sequences with various lengths and display the ROC curves in Fig. 16. Using the first method, for a typical human, we need only about two to three stride cycles (30–40 frames for a 30 fps video) to estimate the correct period.

For the MPGA fitting based method, a more meaningful measure will be the PLL locking time. Given the initial guess of the internal oscillator to be a regular gait rate (for example, 1Hz or 30 frames/cycle in color sensor and 2 Hz or 15 frames/cycle in IR video for full frame rate), we draw the figure of the locking time (in terms of frames) vs. number of lockings in Fig 17. Left image is for the infrared sensor and right is for visible spectrum sensor, together with the approximate Gaussian distributions. The histogram for the color sensor has a clear peak at 36 frames with a narrow bandwidth,

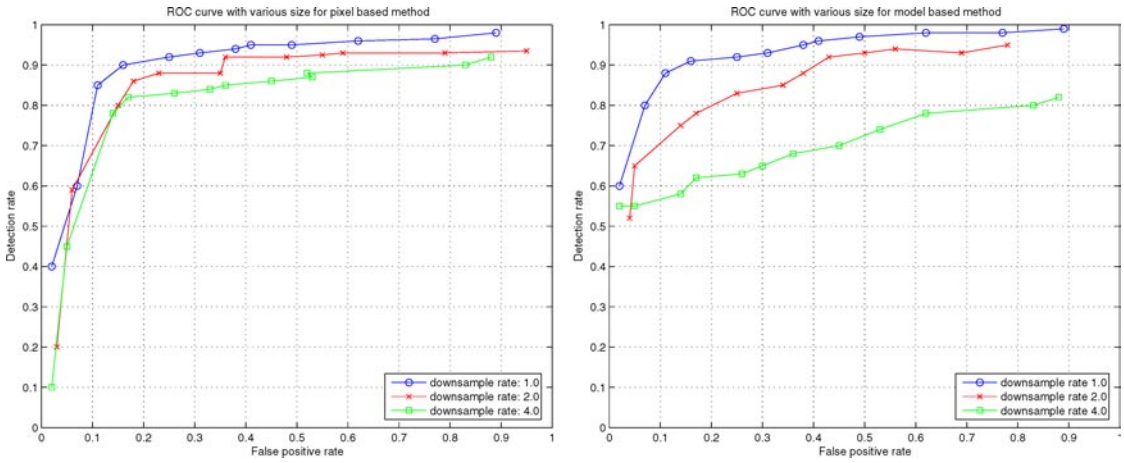

Figure 15. ROC analysis for the two methods at different object sizes.

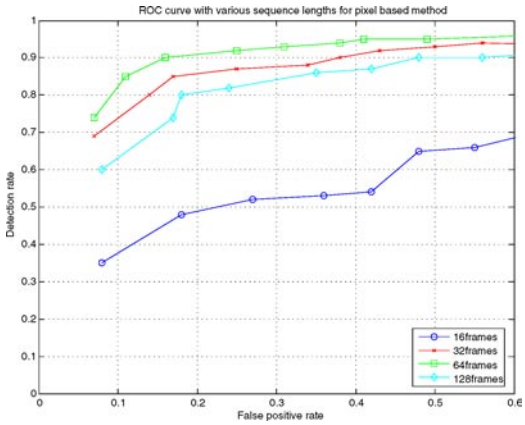

Figure 16. ROC analysis for the pixel-based method with various lengths.

showing the quick locking time. We observe two peaks for infrared data due to the fact that some low quality thermal sensor has the same response for left and right legs and so the real 'period' is half of the gait rate because of the symmetry of walking. After locking to the correct frequency, the module adapts to it without any re-initialization at a very high speed.

5.3. *Frame Rate*

Due to sensor limitation, it may be unable to always have the full frame rate (>25 fps). In addition, robustness to frame rate could be useful for reducing overall computational cost. We present results for the two data sets with different frame rates in Fig. 18. The original rate is 30 frames/second and we reduce it systematically to 6 frames/second. We still obtain good results for both of the methods even when the frame rate is about 10 frames/second.

Comparing it with the ROC curves in the sensitivity analysis for size, the results show that the pixel and model based methods are more sensitive to frame rate than to object size. At lower frame rates, longer sequence length could be used to compensate for the loss in periodic signal strength.

5.4. *Walking Direction*

The observed oscillation amplitude of walking in images varies with different walking directions. It will approach a minimum when the pedestrian is walking in radial direction and will increase gradually to a maximum when the walker is moving in lateral direction. The change of the amplitude of periodicity will directly affect the detector performance.

We divide part of the data sets into subgroups according to walking directions and compare the results for the pixel based detector in Fig. 19. The results show that the first method correctly classifies human under different poses ranging from radial to lateral. This could

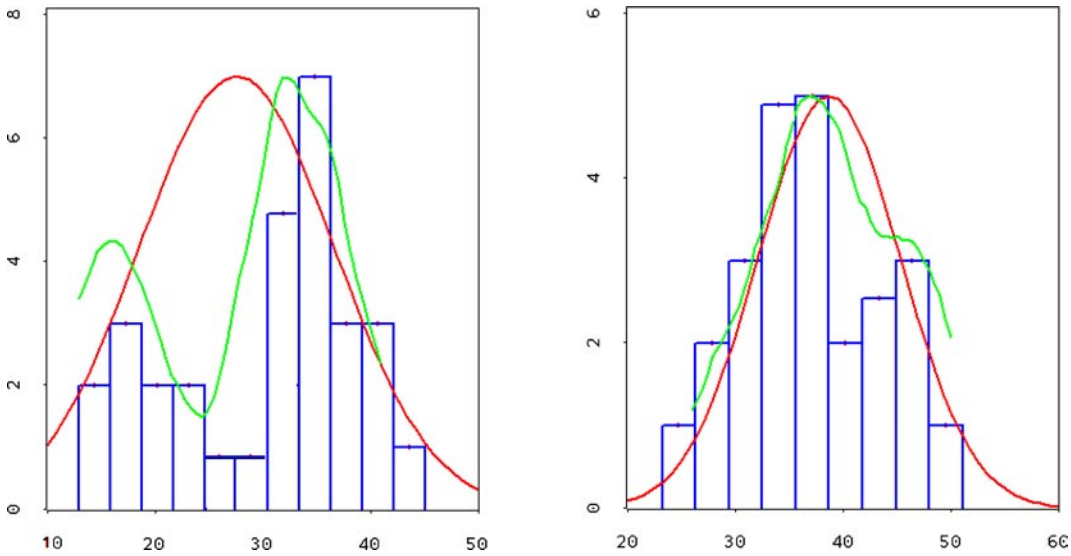

Figure 17. Locking time (in terms of frame number) vs. number of lockings for the MPGA based method. Left: infrared sensor; right: color/gray sensor.

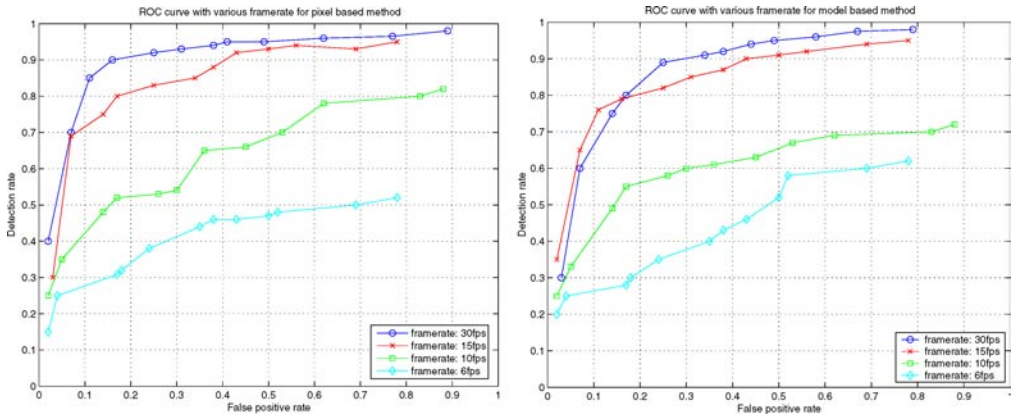

Figure 18. ROC analysis for the two methods at different frame rate.

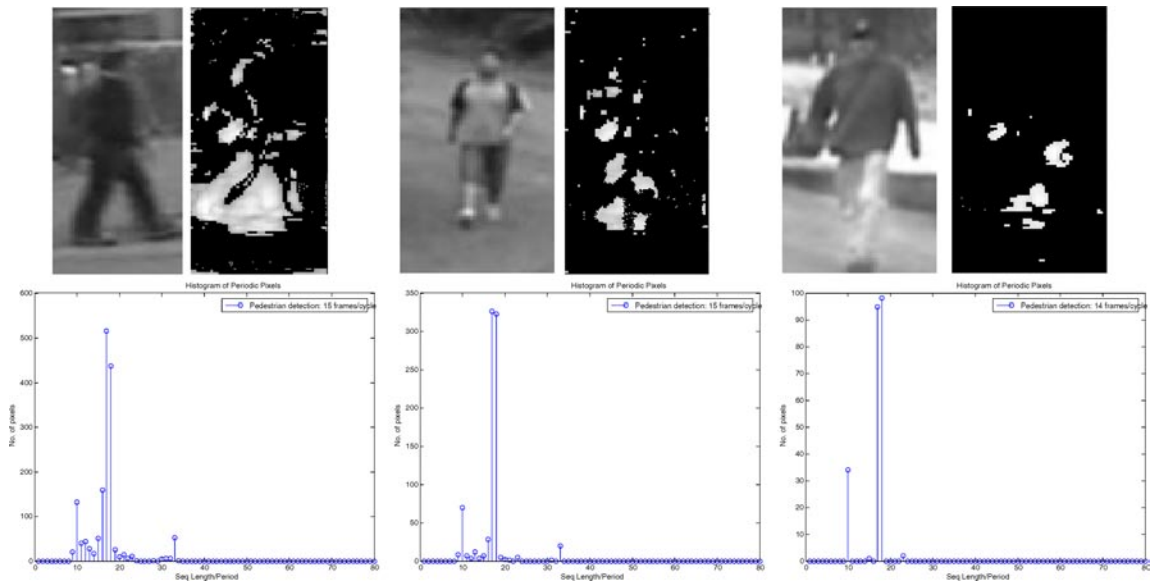

Figure 19. Detection with various walking directions.

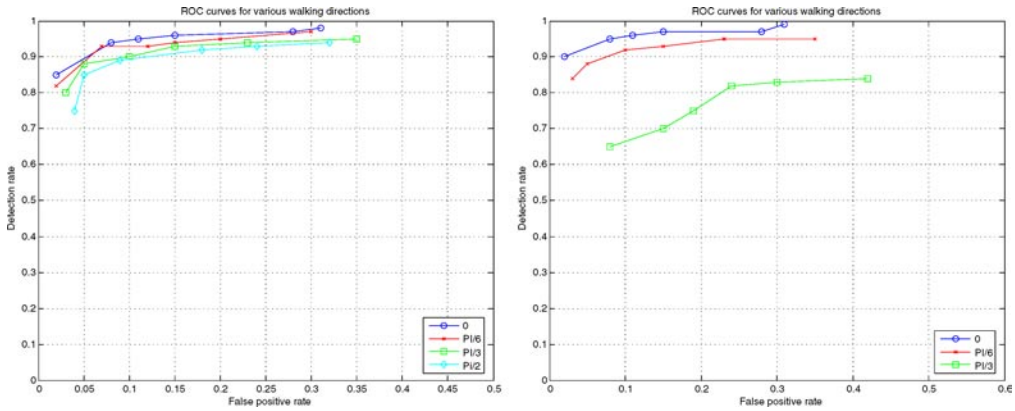

Figure 20. ROC analysis at different walking directions. Left: Pixel based; Right: MPGA fitting based.

be explained by the 2-step hypothesis testing. When a pedestrian is walking towards the camera, many locations no longer exhibit strong cyclic pattern. Yet subtle oscillations still exist around body parts such as limbs and shoulders. By filtering out the non-periodic pixels, a small number of 'good' pixels with reasonable high amplitude can be extracted and support correct estimation. In Fig. 19, with a box size of 40*80, the number of 'good' pixels decreased from thousands to hundreds and even to several tens with the movement direction changing from lateral to radial. A portion of good samples less than $1/16\ (= 200/(40*80))$ is used in the final case when the target is moving towards the sensor.

A comparison of the ROC curves of the sensitivity to walking directions of the two methods is shown in Fig. 20 for both detectors. The angle of walking direction is referred to the camera image plane. The results demonstrate that the pixel based method is more stable to walking directions due to the selectivity of periodic pixels. The MPGA fitting based method, as expected, drops performance sharply when the walking direction is over $\pi/3$.

6. Conclusion

We have developed two algorithms for pedestrian classification based on periodic motion. The first method is simple, efficient and robust to camera motion, sensor noise and walking directions but depends on good alignment accuracy. The second uses a global descriptor combining shape and motion, which is robust to alignment and recursively estimates the gait rate. Both methods can detect pedestrians within a short time period (less than 2 seconds). Sensitivity analysis shows the robust behavior of the proposed methods with respect to a number of important factors such as frame rate, walking directions and object size.

The pixel-periodicity based method monitors the oscillation at each pixel site and statistically extracts the overall frequency. It works better when alignment is reliable for both lateral and radial view and is computationally efficient. The model-fitting based method obtains classification cues from global shape and appearance and then examines periodic gait dynamics. It extracts the MPGA in special gait phases and uses a phase-lock loop to continuously classify targets. It does not require alignment between frames.

A promising future direction is to use a shape detector such as Viola's Adaboosting method (Viola et al., 2003) or Nanda's shape hierarchy (Nanda and Davis, 2002) followed by the cyclic motion detector or pixel periodic detector as a verification module to obtain higher performance. By doing so we do not need to search the whole image in every frame and hence it is more computationally efficient. As part of the results shown in Fig. 14, this will form an automatic pedestrian detection system with lower false alarm rate and faster speed.

Acknowledgments

This research effort was supported by General Dynamics Robotic Systems (GDRS) and the Collaborative Technology Alliance (CTA) under contract DAAD19-012-0012 ARL-CTA-DJH. We would like to thank Dr. Rama Chellappa and Dr. Kevin S. Zhou for providing code and discussion. We would also like to thank Dr. Wael Abd-Almageed and Mr. Feng Guo for their helpful comments.

References

Adelson, E.H. and Bergen, J.R. 1985. Spatiotemporal energy models for the perception of motion. *Journal Optical Society of America A*, 2(2).

Aggarwal, J.K. and Cai, Q. 1999. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3):428–440.

Allmen, M.C. 1991. Image sequence description using spatiotemporal flow curves: Toward motion-based recognition. Ph.D. Dissertation, Computer Sciences Department Technical Report 1040, University of Wisconsin-Madison.

Blanchard, A. 1976. *Phase-locked loops*. New York, NY: John Wiley and Sons.

Boyd, J.E. 2004. Synchronization of oscillations for machine perception of gaits. *Computer Vision and Image Understanding*, 96(1):35–59.

Broggi, A., Bertozzi, M., Fascioli, A., and Sechi, M. 2000.Shape-based pedestrian detection. In *Proc. IEEE Intell. Veh. Symp.*, pp. 215–220.

Collins, R.T., Lipton, A.J., and Kanade, T. 2000. Introduction to the special section on video surveillance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):745–746.

Cutler, R. and Davis, L.S. 2000. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):781–796.

Curio, C., Edelbrunner, J., Kalinke, T., Tzomakas, C., and von Seelen, W. 2000. Walking pedestrian recognition. In *IEEE Transactions on Intelligent Transportation Systems*, 1(3):155–163.

Efros, Berg, A.C., Mori, G., Malik, J. 2003. Recognizing action at A distance. In *Proceedings of IEEE International Conference on Computer Vision*, pp. 726–733.

Fang, Y., Yamada, K., Ninomiya, Y., Horn, B., and Masaki, I. 2003. Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection. *IEEE. Intelligent Vehicles Symposium*, pp. 505–510.

Fukunaga, K. 1990. *Introduction to Statistical Pattern Recognition*, 2nd ed. Boston: Academic Press,

Gavrila, D.M. 1999. The visual analysis of human movement: A survey. *Computer Vision and Image understanding*, 73(1):82–98.

Hogg. D. 1983. Model-based vision: A program to see a walking person. *Image and Vision computing*, 1(1):5–20.

Wang, L., Hu, W., and Tan, T. 2003. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601.

Liu, F. and Picard, R.W. 1998. Finding periodicity in space and time. In *Proceedings of the 6th International Conference on Computer Vision*, pp. 376–382.

Lindsey, W.C. and Chie, C.M. (eds.) 1986. *Phase-Locked Loops*. IEEE PRESS Selected Reprint Series, New York, NY: IEEE Press.

Lipton, J., Fujioshi, H., and Patil, R.S. 1998. Moving target classification and tracking from real-time video. In *Workshop on Applications of Computer Vision*, Princeton, NJ, pp. 8–14.

Maybank, S. and Tan, T. 2000. Introduction to special section on visual surveillance. *International Journal of Computer Vision*, 37(2):173–173.

Nanda, H. and Davis, L. 2002. Probabilistic template based pedestrian detection in infrared videos. In *IEEE Intelligent Vehicle Symposium*, Versailles, France.

Niyogi, S.A. and Adelson, E.H. 1994. Analyzing and recognizing walking figures in XYT. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 469–474.

Oren, M., Papageorgiou, C.P., Sinha, Osuna, E., and Poggio, T. 2003. Pedestrian detection using wavelet templates. In *IEEE Conference on Computer Vision and Pattern Recognition*, 193–199.

Pai, C.-J., Tyan, H-R., Liang, Y.-M., Liao, H.-Y. M., Chen, S.-W. 2004. Pedestrian detection and tracking at crossroads. *Pattern Recognition*, 37(5):1025–1034.

Papageorgiou, C., Evgeniou, T., and Poggio, T. 1998. A trainable pedestrian detection system. In *IEEE Int. Conf. on Intelligent Vehicles*, pp. 241–246.

Phillips, P.J., Sarkar, S., Robledo, I., Grother, P., and Bowyer, K.W. 2002. The gait identification challenge problem: Data sets and baseline algorithm. *International Conference on Pattern Recognition*, pp. 385–388.

Polana, R. and Nelson, C. 1997. Detection and recognition of periodic, nonrigid motion. *International Journal of Computer Vision*, 23(3):261–282

Quinn, B.G. and Hannan, E.J. 2001. *The Estimation and Tracking of Frequency*. Cambridge University Press, ISBN 0-521-80446-9.

Rohr, K. 1994. Towards model-based recognition of human movement in image sequences. *CVGIP: Image Understanding*, 59(1):94–115.

Seitz, S.M. and Dyer, C.R. 1997. View-invariant analysis of cyclic motion. *Int. J. Computer Vision*, 25(3):231–251.

Tsai, P., Shah, M., Keiter, K., and Kasparis, K. 1994. Cyclic motion detection. *Pattern Recognition*, 27(12).

Viola, P., Jones, M., Snow, D. 2003. Detecting pedestrians using patterns of motion and appearance. In *Ninth IEEE International Conference on Computer Vision*, pp. 734–782.

Little, J.J. and Boyd, J.E. 1998. Recognizing people by their gait: The shape of motion. *Videre: Journal of Computer Vision Research*, The MIT Press, 1(2):24–42.

Zhao, L. and Thorpe, C. 2000. Stereo and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):148–154.

Zheng, Q. and Chellappa, R. 1991. Automatic registration of oblique aerial images. In *IEEE International Conference on Image Processing*, pp. 218–222.

Zhou, S., Chellappa, R., and Moghaddam, B. 2004.Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Transactions on Image Processing*, 11:1434–1456.