A Novel Approach For Object Detection And Tracking

Ajit Ranjan[1], Manisha Chaple[2]

Abstract

Video and image processing has been used for traffic surveillance, analysis and monitoring of traffic conditions in many cities and urban areas. Motion tracking is one of the most active research titles in computer vision field. This paper aims to Detect and track the vehicle from the video frame sequence The vehicle motion is detected and tracked along the frames using Optical Flow Algorithm and Background Subtraction technique. The distance travelled by the vehicle is calculated using the movement of the centroid over the frames. Many proposed motion tracking techniques are based on template matching, blob tracking and contour tracking. A famous motion tracking and estimation technique, optical flow, however, is not being widely used and tested for the practicability on traffic surveillance system. Thus, to analyze the reliability and practicability of it, this research project proposed the idea of implementing optical flow in traffic surveillance system, and will evaluate its performance. The results of tracking using optical flow is proving that optical flow is a great technique to track the motion of moving object, and has great potential to implement it into traffic surveillance system.

1. Introduction

Video surveillance systems have long been in use to monitor security sensitive areas. The making of video surveillance systems "smart" requires fast, reliable and robust algorithms for moving object detection, classification, tracking and activity analysis. Moving object detection is the basic step for further analysis of video. It handles segmentation of moving objects from stationary background objects. This not only creates a focus of attention for higher level processing but also decreases computation time considerably. Commonly used techniques for object detection are background subtraction, statistical models, temporal differencing and optical flow. Due to dynamic environmental conditions such as illumination changes, shadows and waving tree branches in the wind object segmentation is a difficult and significant problem that needs to be handled well for a robust visual surveillance system.

Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, clutter, etc. It is necessary to distinguish objects from each other in order to track and analyze their actions reliably. Currently, there are two major approaches towards moving object classification, which are shape-based and motion-based methods [1]. Shape-based methods make use of the objects 2D spatial information whereas motion-based methods use temporal tracked features of objects for the classification solution. Detecting natural phenomenon such as fire and smoke may be incorporated into object classification components of the visual surveillance systems. Detecting re and raising alarms make the human operators take precautions in a shorter time which would save properties, forests and animals from catastrophic consequences.

The next step in the video analysis is tracking, which can be simply defined as the creation of temporal correspondence among detected objects from frame to frame. This procedure provides temporal identification of the segmented regions and generates cohesive information about the objects in the monitored area such as trajectory, speed and direction. The output produced by tracking step is generally used to support and enhance motion segmentation, object classification and higher level activity analysis.

The final step of the smart video surveillance systems is to recognize the behaviours of objects and create high-level semantic descriptions of their actions. It may simply be considered as a classification problem of the temporal activity signals of the objects according to pre-labelled reference signals representing typical human actions.

The outputs of these algorithms can be used both for providing the human operator with high level data to help him to make the decisions more accurately and in a shorter time and for online indexing and searching stored video data effectively. The advances in the development of these algorithms would lead to breakthroughs in applications that use visual surveillance .

2. REVIEW OF PREVIOUS WORK

Many applications have been developed for monitoring public areas such as offices, shopping malls or traffic highways.
We classify these tracking techniques into four categories:
Tracking based on a moving object region. This method identifies and tracks a blob token or a bounding box, which are calculated for connected components of moving objects in 2D space. The method relies on properties of these blobs such as size, color, shape, velocity, or centroid. A benefit of this method is that it time efficient, and it works well for small numbers of moving objects. Its shortcoming is that problems of occlusion cannot be solved properly in "dense" situations. Grouped regions will form a combined blob and cause tracking errors. For example, presents a method for blob tracking. Kalman filters are used to estimate pedestrian parameters. Region splitting and merging are allowed. Partial overlapping and occlusion is corrected by defining a pedestrian model.
Tracking based on an active contour of a moving object. The contour of a moving object is represented by a snake, which is updated dynamically. It relies on the boundary curves of the moving object. For example, it is efficient to track pedestrians by selecting the contour of a human's head. This method can improve the time complexity of a system, but its drawback is that it cannot solve the problem of partial occlusion, and if two moving objects are partially overlapping or occluded during the initialization period, this will cause tracking errors. For example, proposes a stochastic algorithm for tracking of objects. This method uses factored sampling, which was previously applied to interpretations of static images, in which the distribution of possible interpretations is represented by a randomly generated set of representatives. It combines factored sampling with learning of dynamical models to propagate an entire probability distribution for object position and shape over time. This improves the mentioned drawback of contour tracking in case of partial occlusions, but increases the computational complexity.
Tracking based on a moving object model. Normally model based tracking refers to a 3D model of a moving object. This method defines a parametric 3D geometry of a moving object. It can solve partially the occlusion problem, but it is (very) time consuming, if it relies on detailed geometric object models. It can only ensure high accuracy for a small number of moving objects for example, solved the partial occlusion problem by considering 3D models. The definition of parameterized vehicle models makes it possible to exploit the a-priori knowledge about the shape of typical objects in traffic scenes.
Tracking based on selected features of moving objects. Feature based tracking is to select common features of moving objects and tracking these features continuously. For example, corners can be selected as features for vehicle tracking. Even if partial occlusion occurs, a fraction of these features is still visible, so it may overcome the partial occlusion problem. The difficult part is how to identify those features which belong to the same object during a tracking procedure (feature clustering). Several papers have been published on this aspect. For example, extracts corners as selected

3. OBJECT DETECTION TECHNIQUE

Each application that benefit from smart video processing has different needs, thus re-quires different treatment. However, they have something in common: moving objects.

Thus, detecting regions that correspond to moving objects such as people and vehicles in video is the first basic step of almost every vision system since it provides a focus of attention and simplifies the processing on subsequent analysis steps. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process reliably. Frequently used techniques for moving object detection are background subtraction, statistical methods, temporal differencing and optical flow whose descriptions are given below.

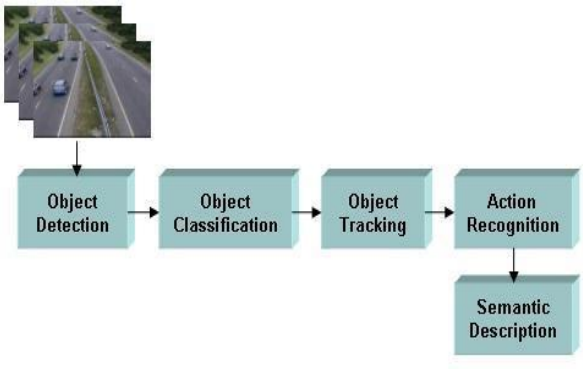


Figure 2.1: A generic framework for smart video processing algorithms.

A. *Background Subtraction*

Background subtraction is particularly a commonly used technique for motion segmentation in static scenes [2]. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes.

There are different approaches to this basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing. In [3] Heikkila and Silven uses the simple version of this scheme where a pixel at location (x, y) in the current image It is marked as foreground if

$$|I_t(x, y) - B_t(x, y)| > a \qquad (2.1)$$

is satisfied where it is a predefined threshold. The background image B_t is updated by the use of an In Infinite Impulse Response(IIR) filter as follows:

$$B_{t+1} = b * I_t + (1 - b)B_t \qquad (2.2)$$

Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop, they are usually sensitive to dynamic changes when, for instance, stationary objects uncover the back-ground (e.g. a parked car moves out of the parking lot) or sudden illumination changes occur.

Fig 2.1: Background Subtraction of Car from parking lot

B. *TEMPORAL DIFFERENCING*

Temporal differencing attempts to detect moving regions by making use of the pixel
by-pixel difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic scene changes, however, it generally fails in detecting whole relevant pixels of some types of moving objects. A sample object for inaccurate motion detection is shown in Figure 2.2. The mono coloured region of the human on the left hand side makes the temporal differencing algorithm to fail in extracting all pixels of the humans moving region. Also, this method fails to detect stopped objects in the scene. Additional methods need to be adopted in order to detect stopped objects for the success of higher level processing.

Lipton et al. presented a two-frame differencing scheme where the pixels that satisfy the following equation are marked as foreground.

$$|I_t(x, y) - I_{t-1}(x, y)| > \tau \qquad (2.4)$$

In order to overcome shortcomings of two frame differencing in some cases, three frame differencing can be used [4]. For instance, Collins et al. developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM project [5]. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction.

(a) (b)

Figure 2.2: Temporal differencing sample. (a) A sample scene with two moving objects. (b) Temporal
differencing fails to detect all moving pixels of the object on the left hand side since it is uniform coloured. The detected moving regions are marked with red pixels.

C. *OPTICAL FLOW*

- ***Vehicle Detection Phase***

The object detection is performed by extracting the features of each object. Based on the

dimension of every object it has its own specific feature. The feature extraction algorithm applied in this paper is optical flow which is used to detect and point out object in each frame sequence. In this method, the pixels are calculated based on the vector position and it is compared in frame sequences for the pixel position. In general the motion is correspond to vector position of pixels. Finding optic flow using edges has the advantage (over using two dimensional features) that two dimensional feature detection.

- ***Motion Estimation***

Optical flow is used to compute the motion of the pixels of an image sequence. It provides a dense (point to point) pixel correspondence. The problem is to determine where the pixels of an image at time t are in the image at time t+1. Large number of applications uses this method for detecting objects in motion

- ***Optical flow Algorithm***

Optical flow computation is based on two assumptions:
The experimental brightness of any object point is constant over time. Close to points in the image plane move in a similar manner (the velocity smoothness constraint). Suppose we have a continuous image; f{x,y,t) refers to the gray-level of (x,y) at time t. Representing a dynamic image as a function of position and time permits it to be expressed.
Assume each pixel moves but does not change intensity
Pixel at location (x, y) in frame1 is pixel at (x+Δx, y+Δy) in frame2.
Optic flow associates displacement vector with each pixel.
The optical flow methods try to calculate the motion between two image frames which are taken at times t and t + δt at every voxel position. These methods are called differential since they are based on local Taylor series approximations of the image signal; that is, they use partial derivatives with respect to the spatial and temporal coordinates.

Assume I (x, y, t) is the center pixel in a n×n neighborhood and moves by δx, δy in time δt to I (x+δx, y +δy, t+δt). Since I (x, y, t) and I (x + δx, y + δy, t + δt) are the images of the same point (and therefore the same) we have:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (1)$$

Solving Eq (1) gives

$$I_x V_x + I_y V_y = -I_t \quad (2)$$

Where, Ix,Iy,It are intensity derivative in x,y,t respectively and Vx,Vy are the x and y components of the velocity or optical flow of I(x,y,t). Eq(2) is then solved using Horn-Schunck method.

The video sequence is captured using a fixed camera. The Optical Flow block using the Horn – Schunck algorithm (1981) estimates the direction and speed of object motion from one video frame to another and returns a matrix of velocity components . Various image processing techniques such as thresholding, median filtering are then sequentially applied to obtain labeled regions for statistical analysis.

Thresholding is the simplest method of image segmentation. The process of thresholding returns a threshold image differentiating the objects in motion (in white) and static background (in black). More precisely, it is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. A Median Filter is then applied to remove salt and pepper noise from the threshold image without significantly reducing the sharpness of the image Median filtering is a simple and very effective noise removal filtering process and an excellent filter for eliminating intensity spikes.

- ***Vehicle Tracking using optical flow***

Object tracking refers to the process of tracing the moving object in progression of frames. The task of tracking is performed by feature extraction of objects in a frame and discovering the objects in sequence of frames. By using the location values of object in each frame, we can determine the position and velocity of the moving object.
They can detect motion in video sequences even from a moving camera, however, most of the optical flow methods are computationally complex and cannot be used real-time without specialized hardware [4].

D. SHADOW AND LIGHT ILLUMINATION CHANGE

The algorithms described above for motion detection perform well on indoor and outdoor environments and have been used for real-time

surveillance for years. However, without special care, most of these algorithms are susceptible to both local (e.g. shadows and highlights) and global illumination changes (e.g. sun being covered/uncovered by clouds). Shadows cause the motion detection methods fail in segmenting only the moving objects and make the upper levels such as object classification to perform inaccurate. The proposed methods in the literature mostly use either chromaticity [6,7,8,9,10] or stereo [2] information to cope with shadows and sudden light changes.

Figure 2.3: Sudden light change sample. (a) The scene before sudden light change (b) The same scene after sudden light change

Horprasert et al. present a novel background subtraction and shadow detection method [6]. In their method, each pixel is represented by a color model that separates brightness from the chromaticity component. A given pixel is classified into four different categories (background, shaded background or shadow, highlighted background and moving foreground object) by calculating the distortion of brightness and chromaticity between the background and the current image pixels. Like [6], the approach described by McKenna et al. in [7] uses chromaticity and gradient information to cope with shadows. They also use the gradient information in moving regions to ensure reliability of their method in ambiguous cases.

4. OBJECT CLASSIFICATION TECHNIQUE

Moving regions detected in video may correspond to different objects in real-world such as pedestrians, vehicles, clutter, etc. It is very important to recognize the type of a detected object in order to track it reliably and analyze its activities correctly.

Currently, there are two major approaches towards moving object classification which are shape-based and motion-based methods [4]. Shape-based methods make use of the objects 2D spatial information whereas motion-based methods use temporally tracked features of objects for the classification solution.

A. SHAPE BASED CLASSIFICATION

Common features used in shape-based classification schemes are the bounding rectangle, area, silhouette and gradient of detected object regions.

The approach presented in [5] makes use of the objects silhouette contour length and area information to classify detected objects into three groups: human, vehicle and other. The method depends on the assumption that humans are, in general, smaller than vehicles and have complex shapes. Dispersedness is used as the classification metric and it is defined in terms of objects area and contour length (perimeter).

$$\text{Dispersedness} = \text{perimeter}^2/\text{area} \qquad (2.3)$$

Classification is performed at each frame and tracking results are used to improve temporal classification consistency.

The classification method developed by Collins et al. [3] uses view dependent visual features of detected objects to train a neural network classifier to recognize four classes: human, human group, vehicle and clutter. The inputs to the neural network are the dispersedness, area and aspect ratio of the object region and the camera zoom magnification. Like the previous method, classification is performed at each frame and results are kept in a histogram to improve temporal consistency of classification. Saptharishi et al. propose a classification scheme which uses a logistic linear neural network trained with Differential Learning to recognize two classes: vehicle and people Papageorgiou et al. presents a method that makes use of the Support Vector Machine classification trained by wavelet transformed object features (edges) in video images from a sample pedestrian database .This method is used to recognize moving regions that correspond to humans.

B. MOTION BASED CLASSIFICATION

Some of the methods in the literature use only temporal motion features of objects in order to recognize their classes [8]. In general, they are used to distinguish non-rigid objects (e.g. human) from rigid objects (e.g. vehicles). The method proposed in [8] is based on the temporal self-similarity of a moving object. As an object that exhibits periodic motion evolves, its self-similarity measure also shows a periodic motion. The method exploits this clue to categorize moving object used periodicity.

Optical flow analysis is also useful to distinguish rigid and non-rigid objects. A. J. Lipton proposed a method that makes use of the local optical flow analysis of the detected object regions. It is expected that non-rigid objects such as humans will present high average residual flow whereas rigid objects such as vehicles will present little residual flow. Also, the residual flow generated by human motion will have a periodicity. By using this cue, human motion, thus humans, can be distinguished from other objects such as vehicles.

5. OBJECT TRACKING STRATEGIES

Tracking is a significant and difficult problem that arouses interest among computer vision researchers. The objective of tracking is to establish correspondence of objects and object parts between consecutive frames of video. It is a significant task in most of the surveillance applications since it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as activity analysis and behaviour recognition. Tracking has been a difficult task to apply in congested situations due to inaccurate segmentation of objects. Common problems of erroneous segmentation are long shadows, partial and full occlusion of objects with each other and with stationary items in the scene. Thus, dealing with shadows at motion detection level and coping with occlusions both at segmentation level and at tracking level is important for robust tracking.

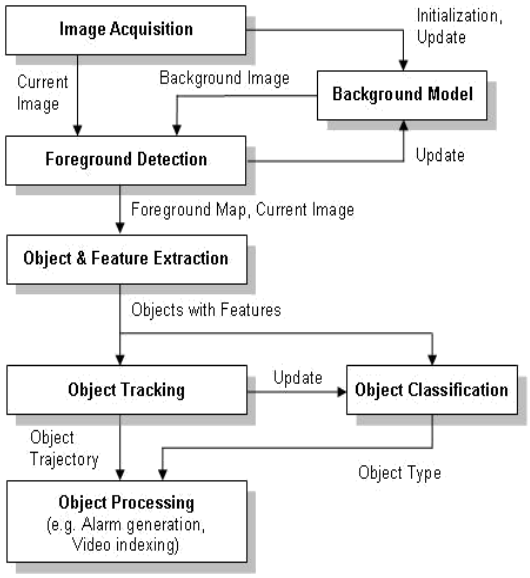


Fig 4.1: Real time Application System

There are two common approaches in tracking objects as a whole [11]: one is based on correspondence matching and other one carries out explicit tracking by making use of position prediction or motion estimation. On the other hand, the methods that track parts of objects (generally humans) employ model-based schemes to locate and track body parts. Some example models are stick figure, Cardboard Model [12], 2D contour and 3D volumetric models.

W4 [13]combines motion estimation methods with correspondence matching to track objects. It is also able to track parts of people such as heads, hands, torso and feet by using the Cardboard Model [12] which represents relative positions and sizes of body parts. It keeps appearance templates of individual objects to handle matching even in merge and split cases.

As an example of model based body part tracking system, Pathfinder [14] makes use
of a multi-class statistical model of colour and shape to track head and hands of people in real-time.

6. EXPERIMENTAL RESULTS

- *Noise Removal Technique*

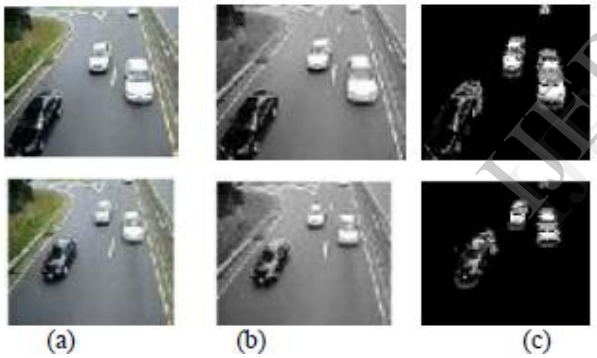
In this phase 2D median Filteris used to remove the nosie present in the video and the respective frames or images.Test was conducted on different filters among them the best suited filter are for Gaussian noise the wiener filter best suits , Salt and Pepper noise is effectively removed by Median

filter and for the periodic noise 2D FIR filter performs better than other filters. The results obtained are shown in the figures.

Figure7: The above figure uses the 2D median Filter to remove the salt pepper noise present in the image.

- *Segmentation Technique*

The segmentation technique is used to group the similar objects by performing background subtraction using Frame difference. This technique best suited for moving objects segmentation. The result shows the input image, the previous frame and after applying the frame difference and subtracting the background objects the foreground is alone displayed the result is displayed in the figures.



From the figure (a) shows the original video, (b) shows that the video is converted to grey scale and (c) shows the segmented output of the video in performing the frame difference of background subtraction.

- *Object Identification and Object Tracking*

Object tracking in video is performed by applying the optical flow method to set the motion vector of the moving objects then finding the threshold of each object and detecting and tracking the objects which exceeds the threshold value as moving objects. The experimental results are shown in the figure.

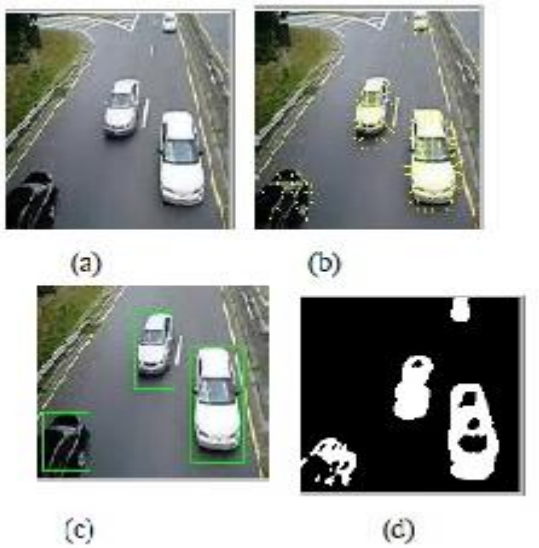


Figure: Object detection and tracking using Optical flow

From the figure (a) shows the original video, (b) shows optical flow method to set the motion vector, (c) shows the object and detecting and tracking the objects which exceeds the threshold value as moving objects and (d) Foreground of the moving object is detected.

7. CONCLUSION

Moving object tracking is a key task in video monitoring applications. The common problem is occlusion detection. In this case the selection of appropriate features is critical for moving object tracking and classification. This method proved to be easy and efficient, but it only works well on separated regions. So removing shadows is an important pre-processing task for the subsequent extraction of moving objects masks, because shadows merge otherwise separated regions. Future work will also apply 3D analysis (a binocular stereo camera system and an infrared camera), which allows a more detailed classification of cars. The intention is to identify the type of a vehicle. The height value of the car is, for example, easily to extract from the infrared picture.

8. REFRENCES

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

[2] A. M. McIvor. Background subtraction techniques. In Proc. of Image and Vision Computing, Auckland, New Zealand, 2000.

[3] J. Heikkila and O. Silven. A real-time system for monitoring of cyclists and pedestrians. In Proc. of

Second IEEE Workshop on Visual Surveillance, pages 74–81, Fort Collins, Colorado, June 1999.

[4] H. Wang and S.F. Chang. Automatic face region detection in mpeg video sequences. In Electronic Imaging and Multimedia Systems, pages 160–168, SPIE Photonics China, November 1996.

[5] R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.

[6] T. Horprasert, D. Harwood, and L.S. Davis. A statistical approach for realtime robust background subtraction and shadow detection. In Proc. of IEEE Frame Rate Workshop, pages 1–19, Kerkyra, Greece, 1999.

[7] S.J. McKenna, S. Jabri, Z. Duric, and H. Wechsler. Tracking interacting people. In Proc. of International Conference on Automatic Face and Gesture Recognition, pages 348–353, 2000.

[8] H.T. Chen, H.H. Lin, and T.L. Liu. Multi-object tracking using dynamical graph matching. In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 210–217, 2001.

[9] M. Xu and T. Ellis. Colour-Invariant Motion Detection under Fast Illumination Changes, chapter 8, pages 101–111. Video-Based Surveillance Systems.Kluwer Academic Publishers, Boston, 2002.

[10] P. KaewTraKulPong and R. Bowden. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection, chapter 11, pages 135–144. Video-Based Surveillance Systems. Kluwer Academic Publishers, Boston, 2002.

[11] A. Amer. Voting-based simultaneous tracking of multiple video objects. In Proc. SPIE Int. Symposium on Electronic Imaging, pages 500–511, Santa Clara, USA, January 2003.

[12] S. Ju, M. Black, and Y. Yaccob. Cardboard people: a parameterized model of articulated image motion. In Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, pages 38–44, 1996.

[13] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: A real time system for detecting and tracking people. In Computer Vision and Pattern Recognition, pages 962–967, 1998.

[14] C. R. Wren, A. Azarbayejani, T. J. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. IEEE Pattern Recognition and Machine Intelligence, 19(7):780–785, July 1997.