ISSN 2250-0987
Tapan kumar das et al, UNIASCIT, Vol 1 (3), 2011, 107-113

A Survey of Techniques for Background Subtraction and Traffic Analysis on Surveillance Video

U.Chandrasekhar [1] ,Tapan kumar Das [1]

[1]*School of Information Technology & Engineering , VIT University, Vellore,India*

Abstract—**Identifying stationary background from moving objects in a video is a critical task in many video processing and computer-vision applications. A basic approach is to perform background subtraction to isolate moving objects from a significant background model. The main challenge in this task is to devise an algorithm to work like a human would, under different scenarios and context. Technically, robustness of such algorithm will be tested by varying illumination, identifying non-stationary background objects like swinging leaves, rain, snow and shadow of moving objects. Also context wise, the algorithm should react quickly to changes in background like parked vehicles and background objects brought in later into the scene. Since we are interested in analyzing traffic frequency in a surveillance video, other challenges include identifying non vehicular moving objects like pedestrians etc. In this paper we compare various simple background verification techniques like frame differencing, Gaussian methods to few complex techniques like probabilistic models. Though complex techniques produce superior and context specific results, our experiments show that, for simple tasks like traffic analysis, simple techniques like adaptive median filtering produce good accuracy with low processing time. Video surveillance of traffic happens from a stationary camera having constant field of view. Generally the resolution and frame rate of these videos are low for high end algorithms to produce good results. Also traffic analysis is not required on live video stream. Except for spontaneous high security alerts, processing can be done offline by buffering the stream data. Hence, the paper concentrates on methods which are robust enough to handle noise, changing climatic conditions and issues with segmentation of moving objects.**

Keywords- *Background subtraction, Traffic video analysis*

I. INTRODUCTION

Today there are an ever growing number of cameras being used for scene analysis. Many of these are applied to traffic monitoring because it is a low cost and passive method for data collection. Research in several fields of traffic applications has resulted in the wealth of video processing and analysis methods [1]. Automatically detecting and tracking vehicles in video surveillance data is a challenging problem in computer vision with important practical applications, such as traffic analysis and security. Manually reviewing the large amount of data they generate is often impractical. Thus, algorithms for analyzing video which require little or no human input are an attractive solution and have been an area of active research for over a decade. In addition to correctness in detecting and tracking vehicles, the computational complexity of a tracking system is important. For many applications, real-time or near real-time tracking capability is desired.

There are various algorithms for object recognition in video. Sample areas include object tracking, traffic monitoring and analysis, Human tracking and gesture recognition etc. Almost all the projects involve identifying background from foreground objects. Also this is the most challenging part. In cases like object tracking, background is continuously changing and in cases like video surveillance background is constant and foreground objects are in motion. This work describes a single stationary camera feed on which the processing techniques are carried out. A comparison of the techniques and their accuracies, merits and demerits are also given.

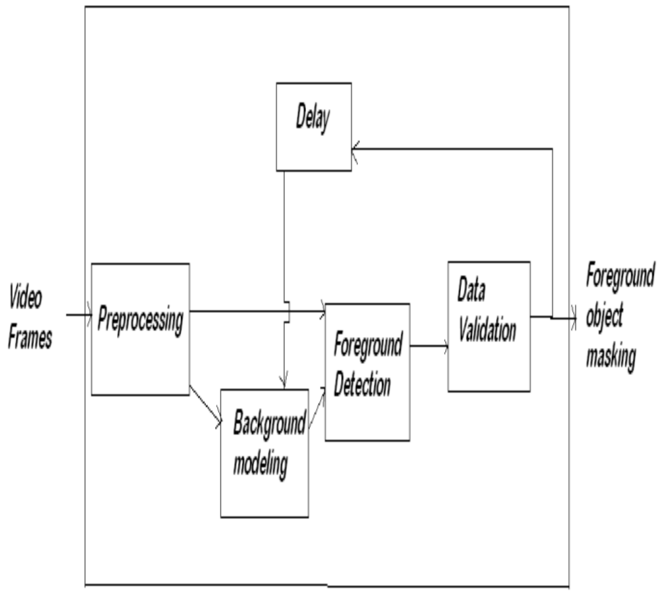


Fig.1. Flow diagram of a general background estimation process

II. BACKGROUND SUBTRACTION

A common approach to identifying the moving objects is background subtraction, where each video frame is compared against a reference or background model. Pixels in the current frame that deviate significantly from the background are considered to be moving objects. These foreground pixels are further processed for object localization and tracking. Since background subtraction is often the first step in many computer vision applications, it is important that the extracted foreground pixels accurately correspond to the moving objects of interest. Most of the algorithms follow simple flow as shown in Fig. 1.

The preprocessing step involves removal of noise, smoothening, and removing transient environmental noise such as rain or snow. We may need to reduce frame rate or frame size to reduce computation. Also in case multiple cameras are used, we need to register the images with respect to each other.

The most common approaches for identifying the moving objects include: optical flow, temporal differencing and background subtraction based methods [8]. Among them, background subtraction is the mostly adopted one. An estimate of the background (often called a background model) is computed and evolved frame by frame, and thus moving objects can be detected by the difference between the current frame and the previous frame.

There are many challenges in developing a good background subtraction algorithm [2]. First, it must be robust against changes in illumination. Second, it should avoid detecting non-stationary background objects such as swinging leaves, rain, snow, and shadow cast by moving objects. Finally, its internal background model should react quickly to changes in background such as starting and stopping of vehicles.

There are papers on comparisons of background subtraction algorithms for detecting moving vehicles and pedestrians in urban traffic video sequences [2]. In another work, they addressed the problem of moving object segmentation using background subtraction [6]. All these applications require as a first step, the detection of moving objects in the observed scene before applying any further technique for object recognition and activity identification. Some have proposed a reliable foreground segmentation algorithm that combines temporal image analysis with a reference background image [6]. Many different approaches have been proposed for the generation of background models.

However, if the dynamics of the scene are more complex, with objects stopping and starting their motion, standard techniques would suffer from significant errors, due to the absence of the object-level knowledge [4]. For traditional background subtraction, the standard intensities in interested images are compared to those in the reference image [5]. In this paper, background subtraction based on logarithmic intensities is proposed. Experimental results show that background subtraction based on logarithmic intensities is superior to traditional background subtraction in producing images with better quality. As background subtraction is the basis for moving object extraction, the improvement in background subtraction can lead to performance enhancement in moving object extraction. The problems of threshold selection [4] have to be pushed to a lower level so that it doesn't affect the background model much. In particular, back grounding methods, using an opportune threshold procedure on the difference between each image of the sequence and a model image of the background, are recognized by the scientific community as those that provide the best compromise between performance and reliability. In addition, they produce the most complete feature data and allow the recovery of the most reliable shapes of the segmented moving objects [6].

Optical flow has also recently arisen as a contender for a good object detection method [9]. Optical flow in a video can be detected using the Horn-Schunck method [3]. This method has recently sprung into wide use but the problems of deleting extraneous movement still remain. The computational complexity of most algorithms is linear in the size of a video frame and the number of vehicles tracked. Also the memory requirements and the speed of the algorithms matter however effective the algorithms are because vehicle tracking is a real time application. Hence, accuracy with speed of computational and robustness is required [7].

III. FOREGROUND DETECTION

We have worked on various techniques, which may require color space conversion. Mixture of Gaussians method however can work on RGB space. For others we work on grey scale image by forming a weighted sum of RGB in the following ratio.

$$0.2989 * R + 0.5870 * G + 0.1140 * B$$

Background subtraction is performed on this image. What ever remaining now is foreground plus noise. We used Thresholding to convert the grey level image to binary image, so that it detects objects free from noise. Otsu's method is used for this. It chooses a threshold to minimize the intra-class variance of the black and white pixels.

Next follows morphological analysis. The binary image output has a lot of noise if the thresholding is low and the contour and blob details are lost if the thresholding is high. This leads to irregular blobs in the binary image. To validate these blobs, the binary image is dilated and eroded ("closed") to remove all the noise and to validate the blobs. This is used to depress false positives and retrieve missing foreground. This is done using a structuring element usually a disk or some other shape. Results of each step are shown in Fig 2.

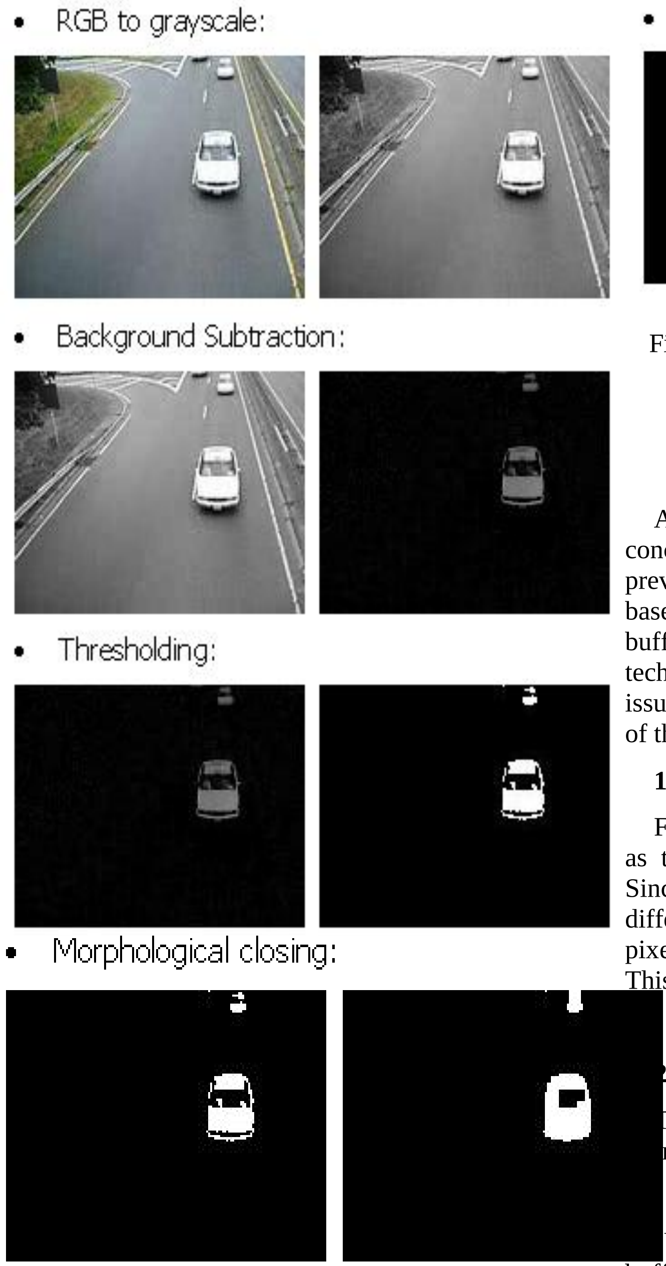


Fig.2. Steps involved in foreground detection.

IV. FOREGROUND OBJECT MASKING

The output of the segmentation is given to Region Filtering & Connected Component Analysis by which the objects are detected as “Blobs”. These blobs are labeled and detected using Blob Analysis. Blob analysis is used for computing various statistics and parameters for connected components of a region like area, centroid, Boundaries, etc.

As shown in Fig 3. , the Boundary boxes for each of the object regions are obtained from blob analysis and are drawn on the frames of the video as rectangles. There by tracking of the vehicles is done. An imaginary line is drawn on road perpendicular to traffic flow. Number of boxes crossing the line is counted.

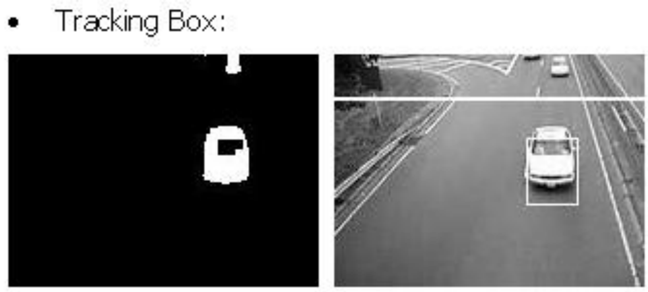


Fig.3. Vehicle detection.

V. VARIOUS BACKGROUND SUBTRACTION TECHNIQUES USED

A. *Non Recursive techniques*

A Non Recursive approach uses sliding window concept for background estimation. It has a buffer of previous video frames and estimates the background based on temporal variation of each pixel within the buffer. Thought the storage requirement is high, these techniques are highly adaptive. To solve storage issue, we can store frames at slow frame rate. Some of the techniques used are described below.

1. Frame Differencing

Frame differencing uses the video frame at time t-1 as the background model for the frame at time t. Since it uses only a single previous frame, frame differencing may not be able to identify the interior pixels of a large, uniformly-colored moving object. This is commonly known as the aperture problem.

| frame(i) – frame (i-1) | > Th

2. Median filtering.

This is most widely used technique for background estimation. The background estimate is defined to be the median at each pixel location of all the frames in the buffer. The assumption is that the pixel stays in the background for more than half of the frames in the buffer. Median filtering has been extended to color by replacing the median with the medoid [10]. The complexity of computing the median is O(Llog L) for each pixel.

3. Linear Predictive filter

It computes the current background estimate by applying a linear predictive filter on the pixels in the buffer [11]. The filter coefficients are estimated at each frame time based on the sample co variances, making this technique difficult to apply in real-time.

B. Recursive techniques

Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model. Compared with non-recursive techniques, recursive techniques require less storage, however any error in the background model can continue for a longer period of time.

1. Approximated median filter

Due to the success of non-recursive median filtering, McFarlane and Schofield propose a simple recursive filter to estimate the median [12]. This technique has also been used in back-ground modeling for urban traffic monitoring [13]. In this scheme, the running estimate of the median is incremented by one if the input pixel is larger than the estimate, and decreased by one if smaller. This estimate eventually converges to a value for which half of the input pixels are larger than and half are smaller than this value, that is, the median.

2. Kalman filter

Kalman filter is a widely-used recursive technique for tracking linear dynamical systems under Gaussian noise. Many different versions have been proposed for background modeling, differing mainly in the state spaces used for tracking.

3. Mixture of Gaussians (MoG)

The pixel distribution f(It = u) is modeled as a mixture of K Gaussians:

$$f(I_t = u) = \sum_{i=1}^{K} \omega_{i,t} \cdot \eta(u; \mu_{i,t}, \sigma_{i,t})$$

where $\eta(u; \mu_{i,t}, \sigma_{i,t})$ is the i'th Gaussian component with intensity mean $\mu_{i,t}$ and standard deviation $\sigma_{i,t}$. $\omega_{i,t}$ is the weight portion of the data accounted for by the i'th component. Typically, K ranges from 3-5. For each input pixel It, the first step is to identify the component i whose mean is closest to It. Component i is declared as the matched component if $|I_t - \mu_{\hat{i},t-1}| \leq D \cdot \sigma_{\hat{i},t-1}$, where D defines a small positive deviation threshold. The component with least variance is chosen as background. A Gaussian model is fit to the frames temporally so that the mean of the Gaussian gives the background of the video.

VI. OPTICAL FLOW METHOD

All the basic modules in this method are the same as that as the temporal median method with the only exception of the foreground extraction model. The foreground is extracted using the optical flow of the objects against the background.

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. Optical flow can arise from relative motion of objects and the viewer. Discontinuities in the optical flow can help in segmenting images into regions that correspond to different objects. In this implementation, optical flow is calculated using the Horn-Schunck method.

By assuming that the optical flow is smooth over the entire image, the Horn-Schunck method computes an estimate of the velocity field, [u v], that minimizes this equation:

$$E = \iint (I_x u + I_y v + I_t)^2 dxdy + \alpha \iint \left\{ \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \right\} dxdy$$

In this equation, the following values are represented:

• Ix,y,t are the spatiotemporal image brightness derivatives

• u is the horizontal optical flow

• v is the vertical optical flow

$$u_{x,y}^{k+1} = \bar{u}_{x,y}^{k} - \frac{I_x [I_x \bar{u}_{x,y}^k + I_y \bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}$$

$$v_{x,y}^{k+1} = \bar{v}_{x,y}^{k} - \frac{I_y [I_x \bar{u}_{x,y}^k + I_y \bar{v}_{x,y}^k + I_t]}{\alpha^2 + I_x^2 + I_y^2}$$

In this equation, $\begin{bmatrix} u_{x,y}^k & v_{x,y}^k \end{bmatrix}$ is the velocity estimate for the pixel at (x,y), and $\begin{bmatrix} \bar{u}_{x,y}^k & \bar{v}_{x,y}^k \end{bmatrix}$ is the neighborhood average of $\begin{bmatrix} u_{x,y}^k & v_{x,y}^k \end{bmatrix}$. For k=0, the initial velocity is 0. Iteratively solve for u and v.

u and v are the velocity vectors for the optical flow. The magnitude of the total velocity gives the velocity of the object i.e., the foreground. This foreground is velocity threshold in the next step.

VII. TEST OUTPUTS

These are all the successful results of the repeated common modules. The outputs of various background generation modules are given below.

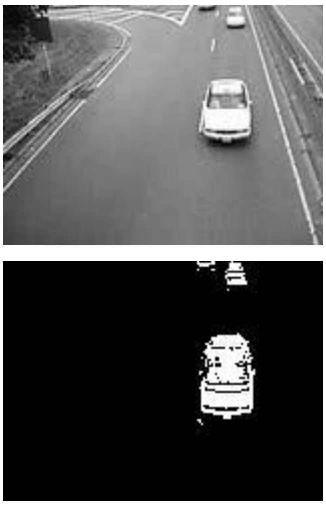
- Temporal median:

- Temporal Mean:

- Approximate median:

- Mixture of Gaussians

- Temporal Differencing



- Optical Flow

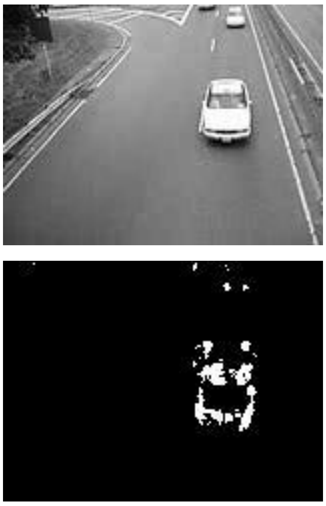
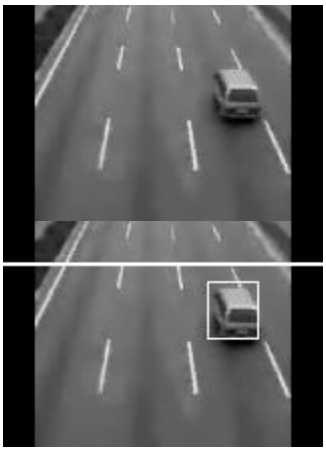


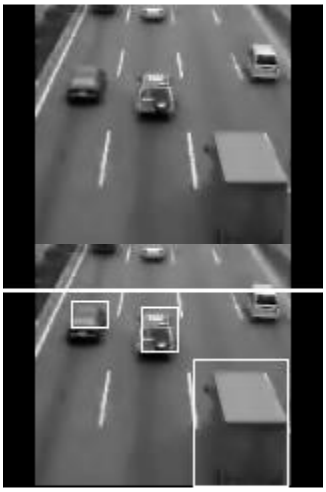
Fig 4. Results of Background subtraction techniques

Following are results of vehicular detection by applying various background estimation techniques. The systems are tested with various video sequences with distinct cases. The vehicles are tracked in the sequences with varying accuracy. The computation time varies with the algorithm and the resolution of the video. The memory requirement also increases linearly with resolution.

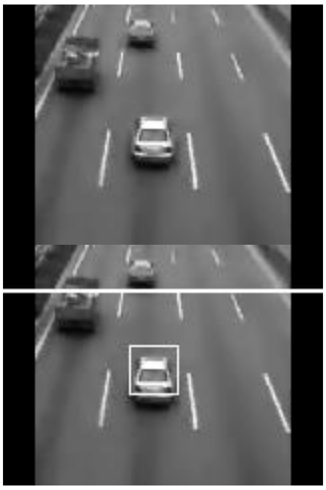
- Temporal Median:



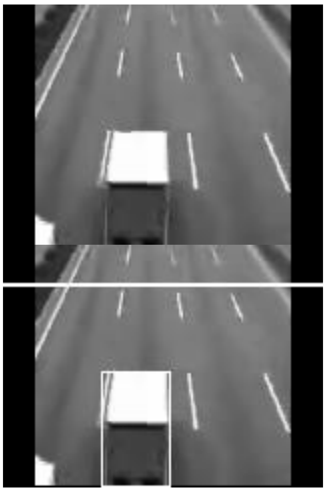
- Temporal Mean:



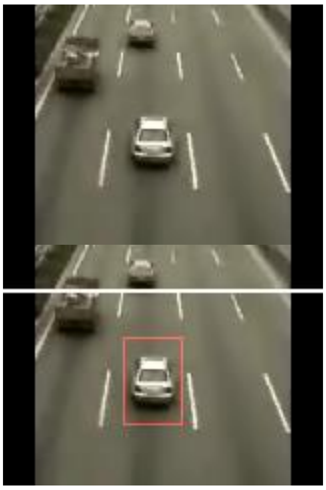
- Approximate median



- Gaussian Mixture Model:



- Temporal differencing:



- Optical Flow

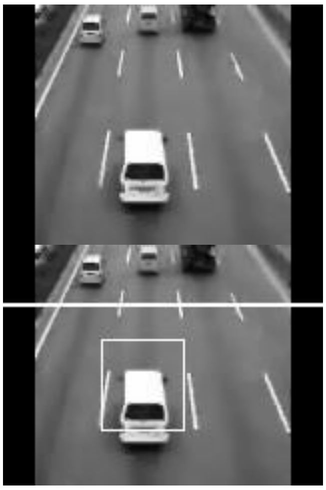


Fig 5. Results of Vehicle detection.

VII. RESULTS AND DISCUSSIONS

The first test case sequence is an overhead video of a highway traffic sampled at 25 fps. Due to some reason there is a sudden change in illumination every 15 frames. The total number of frames is 120. This change in illumination caused certain errors. But the algorithms have shown good results and hence have some vulnerability to illumination changes.

As is often the case, the simplest method is arguably the most robust. While it has major flaws, and is probably not suitable for most applications, frame differencing does the best job of subtracting out extraneous background noise such as waving trees and illumination changes. The median method is very good, better in some cases than the other but has a high memory buffer problem. The approximate median gives us significantly increased accuracy for not much more computation. It had a little trouble with quickly changing light levels, but handled them better than mixture of Gaussians. And Mixture of Gaussians, the most complex of the methods, gives us good performance, but presents a tricky parameter optimization problem. Optical flow likewise has an optimization problem in addition to the high memory buffer, illumination changes and noise vulnerability.

REFERENCES

[1] V. Kastrinaki, M. Zervakis, K. Kalaitzakis, 2003, "A survey of video processing techniques for traffic applications", Image and Vision Computing 21, pg:359–381
[2] Sen-Ching S. Cheung and Chandrika Kamath , 2004," Robust techniques for background subtraction in urban traffic video", Image and Vision Computing.
[3] Berthold K.P. Horn and Brian G. Rhunck, 1981, "Determining Optical Flow", Artificial Intelligence.
[4] Rita Cucchiara, Costantino Grana,Massimo Piccardi & Andrea Prati, 2003, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams", IEEE Transactions On Pattern Analysis And Machine Intelligence.
[5] Quen-Zong Wu & Bor-Shenn Jeng, 2002, "Background subtraction based on logarithmic intensities", Pattern Recognition 22.
[6] P. Spagnolo, T.D' Orazio, M. Leo & A. Distante, 2006, "Moving object segmentation by background subtraction and temporal analysis", Image Vision & Computing 24.
[7] Alper Yilmaz, 2001, "Object Tracking: A Survey".
[8] Alan M. McIvor, "Background Subtraction Techniques".
[9] M. Maire, C. Kamath, 2000, "Tracking Vehicles in traffic Surveillance Video".
[10] R. Cucchiara, M. Piccardi, and A. Prati, \Detecting moving objects, ghosts, and shadows in video streams," IEEE Transactions on Pattern Analysis and Machine Intelligence 25, pp. 1337{1342, Oct 2003.
[11] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, \Wall Ower: Principles and practice of background maintenance," in ICCV (1), pp. 255{261, 1999.
[12] N.McFarlane and C.Schoield, "Segmentation and tracking of piglets in images," Machine Vision and Applications 8(3), pp. 187{193, 1995.
[13] P.Remagnino et al., "An integrated traffic and pedestrian model-based vision system," in Proceedings of the Eighth British Machine Vision Conference, pp. 380{389, 1997.