Detecting Moving Pedestrians and Vehicles in Fluctuating Lighting Conditions

Stephen Nuske
ITEE, University of Queensland, Brisbane, Australia.
and
Autonomous Systems Lab, ICT Centre, CSIRO
P.O. Box 883, Kenmore, Australia, 4069.
Email: stephen.nuske@csiro.au

Manuel Yguel
INRIA, Grenoble, France
Manuel.Yguel@inrialpes.fr

Abstract

Detecting moving pedestrians and vehicles with foreground segmentation algorithms is problematic during fluctuating lighting conditions. Edge-based approaches are more robust to lighting than the conventional intensity-based ones. The issue with edge-based approaches though is segmenting the internal foreground areas. In this work a strategy is developed to detect complete foreground areas. Firstly, edge-extraction is performed at multiple-scales which increases the initial area detected. To complete the detection of object areas, edge-motion-history-images are introduced. The final segmentation is achieved with a region growing algorithm in the edge-motion-history-image. Examples are shown of the successful extraction of foreground objects through changing lighting conditions.

1 Introduction

Foreground segmentation is an active area of research with a primary application of detecting moving pedestrians and vehicles using static surveillance cameras. Foreground segmentation algorithms should be ideally unaffected by fluctuating lighting, but this is a difficult problem.

Traditional foreground segmentation approaches are based on models of raw intensity measurements [Piccardi, 2004] and are susceptible to error in fluctuating lighting. Any change in pixel intensity can either be from changes in lighting or changes in object reflectance, and it is not feasible to competently differentiate between the two causes of intensity changes, using only one pixel. There are recent foreground segmentation emerging [Davis and Sharma, 2006],[Javed *et al.*, 2002],[Yokoyama and Poggio, 2005] which are not intensity-based but instead are edge-based. Edges are derived from relative-intensity information, which has the advantage of being intrinsically robust to uniform lighting changes.

Although edge-based foreground segmentation can reliably detect the boundaries of objects, detecting the central regions of foreground objects is difficult. The three existing edge-based approaches [Davis and Sharma, 2006],[Javed *et al.*, 2002],[Yokoyama and Poggio, 2005] present different techniques to fill-in in internal object areas, however an adequate solution is still elusive.

The aim of this work is use an edge-based foreground segmentation to detect whole foreground areas without creating information that is not directly available in the input images. The first stage of the proposed method is to increase the areas initially detected by using relative-intensity information at multiple scales. This, as will be seen, is not enough to detect complete foreground areas, and therefore the second stage of the approach is to introduce edge-motion-history-images as a means to complete the segmentation. Motion history images (MHI) were developed by Davis and Bobick in [Davis and Bobick, 1997] for intensity images, and this paper presents a novel application to edge images. The final stage of the approach is a region growing algorithm designed to select only internal foreground areas and not select trailing areas of motion history behind moving objects.

This paper will first discuss the related work, and then introduce the unique solution with results showing successful segmentation of moving pedestrians and vehicles.

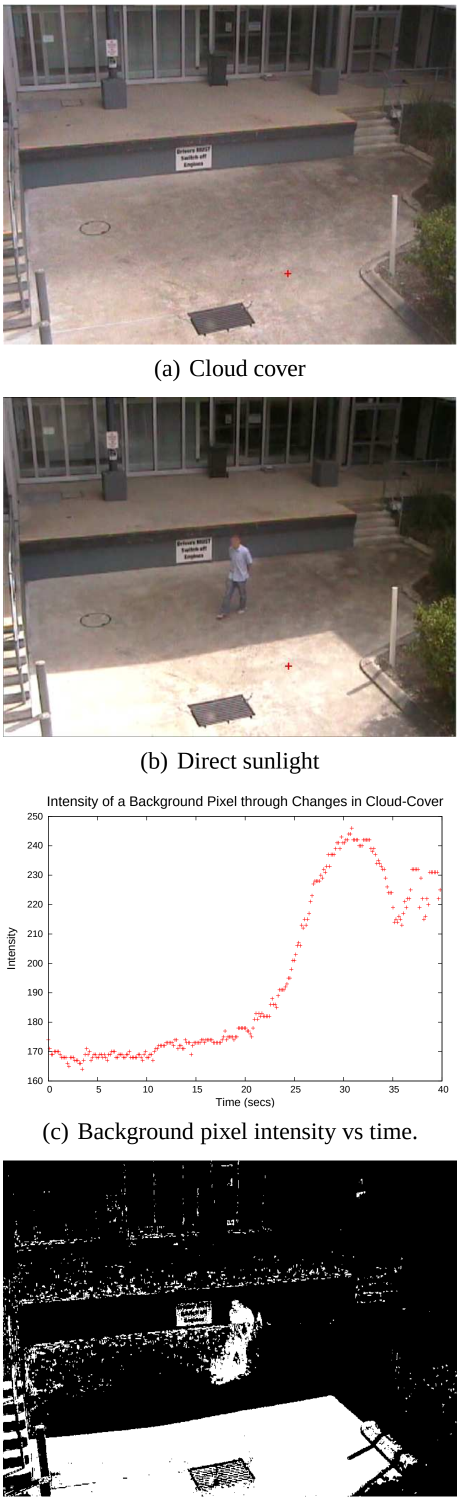
2 Related Work

Foreground segmentation approaches are traditionally based on intensity measurements, and Piccardi [Piccardi, 2004] provides a survey of the predominant intensity-based techniques. Image intensity measurements are a fusion of illumination source *L*, object reflectance *R* and camera sensitivity *C*, over light wavelength λ:

$$I(p) = \int_{\lambda} L(\lambda)R(\lambda)S(\lambda)d\lambda \quad (1)$$

Ignoring the sensitivity of the photo-receptors and the wavelength distribution the intensity formation equation simplifies to:

$$I(p) = LR \quad (2)$$



(a) Cloud cover

(b) Direct sunlight

(c) Background pixel intensity vs time.

(d) Erroneous output of intensity-based foreground segmentation.

Figure 1: Example images of a bay where trucks load/unload items. The two camera images were taken 40 seconds apart. A plot is shown of intensity vs time of a background pixel marked by a cross in the camera images. The erroneous output of an intensity-based foreground segmentation is also shown, where both lighting changes and a pedestrian are detected.

Assuming a static camera, a change in intensity will be due to either a change in the object reflectance R or a change in the illumination source L. The factor of interest is the object reflectance, R, considering the task of detecting foreground objects. However obtaining direct observations of R from I during fluctuating lighting L is problematic. This can be seen in Figure 1, showing a typical change due to a cloud cover, where the two camera images were 40 seconds apart in the video sequence. However the plot of the intensity of pixel intensity from the road through the 40 second sequence shows a drastic change in intensity of 30 percent. Understandably any intensity-based foreground extraction will create false detections due to such lighting changes. This can be seen in the output of an intensity-based algorithm where both sunshine and a person are detected. To avoid this problem more and more complex adaptive background models have been introduced, but fundamentally any intensity-based approach will suseptible to errors such as seen in Figure 1.

The three recent edge-based papers [Davis and Sharma, 2006][Javed *et al.*, 2002][Yokoyama and Poggio, 2005] avoid such dramatic errors during fluctuating lighting, because edge information is intrinsically robust to global lighting levels. Most edge-filters consider neighbourhoods of pixels, but just two pixels can be used (a and b) to illustrate the robustness to lighting levels of an edge image, E:

$$E(p) = |I(a) - I(b)| \tag{3}$$

$$E(p) = |L_a R_a - L_b R_b| \tag{4}$$

Assuming uniform lighting; $L_a = L_b$, the edge image will have the lighting invariant property of being zero in areas of homogenous object reflectance;

$$E(p) = \begin{cases} 0, & \text{if } R_a = R_b \\ > 0, & \text{if } R_a \neq R_b \end{cases} \tag{5}$$

Changes in E, from zero to non-zero values, can therefore be attributed to a foreground object, with certainty that the change is not due to lighting levels. Relative-intensity can reliably detect the edge of objects, the challenge is to detect the internal object areas. Each of the existing edge-based techniques provide a different means to *fill-in* foreground areas.

Javed *et al.* [Javed *et al.*, 2002] propose to combine edge and intensity-based segmentation, by rejecting intensity-segments that do not have edges at the boundary. However Javed *et al.* state this approach is not valid in situations such as Figure 1, where large areas are falsely detected using intensity-based techniques. Therefore they suggest to just use a bounding-box formed by connected edge segments. This solution would lead to large false detections when a large shadow edge appears, as in Figure 1.

Yokoyama and Poggio [Yokoyama and Poggio, 2005] present an approach of clustering detected edge pixels using similarity in optical flow vectors, and selecting foreground areas by forming a convex hull around the clustered edges. The clustering phase is acknowledged in their work to be inappropriate for cluttered scenes. Also the clustering would be valid only for rigid objects, because an object such as a walking human would provide a range of different optical flow vectors.

Davis and Sharma [Davis and Sharma, 2006] develop an edge-based foreground segmentation approach using edges extracted from visible and thermal imagery. The method Davis and Sharma propose to *fill-in* foreground object areas is one of clustering edge components using the k-means algorithm. A significant issue for the k-means algorithm is that the number of objects present, k, needs to be known before clustering. Davis and Sharma finish the process by closing clustered edges using an A* search algorithm. Which would be troubled by objects with internal contours.

Another edge-based image segmentation approach is presented by Smith et al. [Smith *et al.*, 2004] which does not use a background model. In their method edges are extracted and motion models for each edge are assigned using Expectation-Maximization. Edges are then clustered, layered and segmented, considering the estimated motion model, using a probabilistic approach.

The edge-based segmentation approaches discussed above, all attempt to cluster edge pixels into distinct objects, which is very difficult at the pixel-level. Furthermore foreground areas are created without directly available information in the image indicating the presence of a foreground object.

3 Edge-based Background Model

The background model describes the edges of the stationary objects in the scene, it enables moving foreground objects to be detected. It is learnt using a temporal-filter from multiple-time-scales. The initial foreground area detected is increased by using a multi-resolution pyramid background model.

3.1 Edge-Filter

The first stage of forming the background model is to compute edge image, E, from the input intensity images I, using a standard operator:

$$E(x,y) = \frac{\sum_{i=x-n}^{x+n} \sum_{j=y-n}^{y+n} |I(i,j) - I(x,y)|}{n^2} \tag{6}$$

Where n determines the size of the edge kernel, in this paper $n = 1$ and therefore the edge kernel is 3×3. It is possible to use a larger neighbourhood to create the edge image, such as 5×5, however in this paper edge-image pyramids are used to incorporate different spatial scales, as discussed in Section 3.3.

Traditional Median	**Cucchiara *et al.***	**Multiple Temporal Scales**
I_{t-1}	I_{t-10}	I_{t-1}
I_{t-2}	I_{t-20}	I_{t-2}
I_{t-3}	I_{t-30}	I_{t-4}
I_{t-4}	I_{t-40}	I_{t-8}
I_{t-5}	I_{t-50}	I_{t-16}
I_{t-6}	I_{t-60}	I_{t-32}
I_{t-7}	I_{t-70}	I_{t-64}
I_{t-8}	I_{t-80}	I_{t-128}

Figure 2: Traditionally temporal median filters store a set of n previous sequential images.

3.2 Temporal-Background-Filter

A temporal median filter is used to form the background model, which essentially is a stored set of previous images. Traditional median filters store every image in the sequence, which is expensive. Cucchiara *et al.* [Cucchiara *et al.*, 2003] propose to use a single temporal scale of every 10 frames to reduce the storage and processing cost. However with a single temporal scale, frame $t - 100$ holds as much weight as frame $t - 10$. This paper introduces a background model with increasing time scales. The background model B, at pixel p, at time t is calculated as:

$$B(x,y) = \frac{\sum_{l=0}^{m} E_{t-2^l}(x,y)}{m} \tag{7}$$

Where E_k is the input edge image at time k. m is the number of distinct images to include in the model and in this work $m = 7$ and therefore frame $t - 128$ is the furthest image into the past that is incorporated. In the background model the frames $t - 1$ to $t - 8$ will have the same weight as frames $t - 16$ to $t - 128$. This model places greater importance of events nearer in the past, whilst still incorporating events over a long period. Figure 2 illustrates this technique. To avoid the memory cost of storing 128 full images, only m images are stored at any time. The edge images, E_{2^k}, used to calculate B are only temporally approximate and updated only every $2^{(k-1)}$ frames, as follows:

$$E_{t-2^{(k-1)}} \rightarrow E_{t-2^k} \tag{8}$$

This reduces the memory overhead of image storage. The filter does require per-pixel processing of each of the m background images, however the processing is not as intensive in comparison with the updating steps of competing background models, such as the Gaussian mixture model [Stauffer and Grimson, 2000].

There are two stages to detect the foreground pixels. Firstly a set of binary images T are calculated from thresholded subtractions performed between the current edge image and the edge-images in the background model. The subtraction detects those pixels with an edge value a threshold, ψ, above the background image value:

$$T_m(x,y) = \begin{cases} \mathit{true}, & \text{if } E_t(x,y) - B_m(x,y) > \psi \\ \mathit{false}, & \text{otherwise} \end{cases} \tag{9}$$

Through empirical experimenting to find an appropriate threshold value, $\psi = 10$ provided the best results. The next

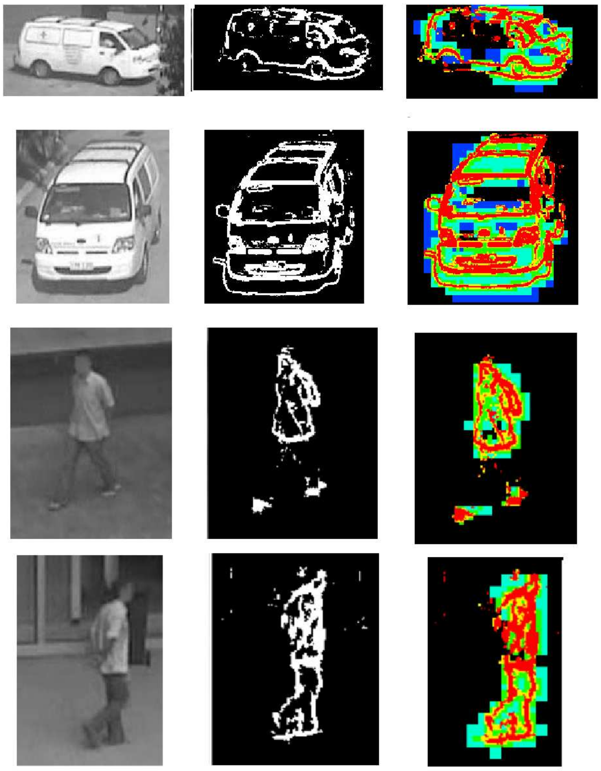


Figure 3: Images showing the results of the edge-based background modeling. Top: A moving van. Bottom: Person walking around a loading bay. Left: Raw video. Middle: Foreground at the finest scale. Right: Multiple-spatial-scale foreground extraction merged together. Even with multiple-scales not all of the object area is detected.

stage is to calculate the final foreground image F by selecting pixels with a majority of positives from the set T;

$$F(x,y) = \begin{cases} true, & \text{if } \sum_{l=0}^{m} T_l(x,y) > \frac{m}{2} \\ false, & \text{otherwise} \end{cases} \tag{10}$$

3.3 Multiple-Spatial-Scales

This section describes how to combine the extracted foreground at multiple-scales. If only one scale is used there is a lot of internal foreground area that is not detected, as seen in the middle column of Figure 3. Multiple-scales are therefore used to increase the area of foreground segmented.

A standard pyramid approach is adopted and can be seen in Figure 4. It begins by recursively down-sampling each input edge image into a multiple spatial scale image pyramid. The image-pyramid is passed through the edge filter (Equation 6) to form an edge-pyramid. A background-model is recorded as a set of edge-pyramids using the temporal filter described in Section 3.2. Foreground pixels are detected from each level in the pyramid using Equation 10. The final foreground pyramid is merged into one high-resolution output image. Examples of the final merged output are shown in the right column of Figure 3.

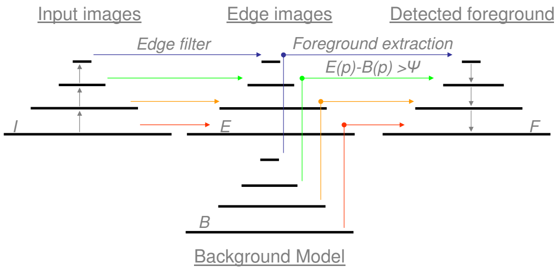


Figure 4: The input image is down-sampled recursively into a multiple spatial scale pyramid. The pyramid is passed through an edge-filter and foreground is extracted from each level and then merged into a final high-resolution output image.

4 Region Growing with Motion History

Even with the use of multiple-scales described in the previous section, there are still situations where there is not a complete detection of objects. This can be seen in Figure 3, where there are internal areas of the van that are not detected. Other relative-intensity approaches attempt to detect full object areas by first clustering edges into an object. But clustering is problematic as stated by Yokoyama and Poggio [Yokoyama and Poggio, 2005]. In light of this clustering is not used in this paper, instead edge-motion-history-images are introduced as a basis for segmenting entire foreground regions.

Motion-History-Images (MHI) were developed by Davis and Bobick [Davis and Bobick, 1997] and are traditionally intensity-based images. This paper presents a novel use edge information with MHI. Edge-MHI enables complete foreground areas to be segmented while also being robust to global lighting conditions.

4.1 Edge-MHI

The Edge-MHI describes the history of the the edge foreground extraction calculated as follows:

$$F_t^{MHI}(x,y) = \begin{cases} MAX(0, F_{t-1}^{MHI}(x,y) - 1), & \text{if } F_t(x,y) = false \\ \kappa, & \text{if } F_t(x,y) = true \end{cases} \tag{11}$$

where κ is the number of frames to record the motion history. κ is dependant on a number of factors; the frame rate, the image resolution, the speed of the moving objects and the distance the moving objects are from the camera. In this work $\kappa = 15$ and was selected through experimentation. Examples of Edge-MHI can be seen in the left column of Figure 5. The detected object area with the edge-MHI, is noticeably more complete when compared with the straight edge extraction in Figure 3.

4.2 Region Growing Algorithm

The area detected by the Edge-MHI is more complete, but there are trails in the Edge-MHI left behind objects, which need to be removed. Furthermore, there are areas surrounding the object due to the use of larger scales, that also need to be removed to give an accurate high-resolution segmentation. Examples of both these problems can be seen in Figure 5. A seeded region growing algorithm is developed to discard the unwanted areas and select the actual foreground areas.

A related edge-restricted growing method is proposed in [Smith *et al.*, 2004]. The algorithm presented in this paper restricts the growing at the fine-scale foreground edges and is summarized as follows:

1. Distance transform on the Edge-MHI to find central pixels.
2. Breadth-first region growing in Edge-MHI, seeded from pixels discovered in previous step. Stop growing when restricted by fine-scale foreground pixels or the boundary of the Edge-MHI. The middle column of Figure 5 shows examples of the regions grown.
3. Regions are selected as foreground when they have a significant count of boundary pixels that are either fine-scale foreground pixels or neighbour previously selected regions. As defined in Equation 12.

The grown regions, R, pass through a selection process based on its boundary pixels, β, to decide whether this region is an internal foreground region or a motion history trail. The decision metric is the ratio between the count of boundary pixels that are deemed valid and those that are deemed invalid. A valid boundary pixel, β_v, is one that is a fine-scale foreground pixel or is a pixel belonging to a region that has previously passed the selection process. An invalid boundary pixel, β_i, is one which is not a fine-scale foreground pixel or is a pixel belonging to a region that has previously failed the selection process. The decision is made by a threshold μ;

$$R = \begin{cases} foreground, & \text{if } \frac{\beta_v}{\beta_i} > \mu \\ background, & \text{otherwise} \end{cases} \tag{12}$$

During the experiments the threshold $\mu = 1$ proved to select correct regions in most of the cases. The middle column of Figure 5 shows the decisions made by the selection process and the right column shows regions that passed the process. The final foreground extraction is in Figure 6.

5 Discussion

The algorithm is tested on both moving pedestrians and vehicles during quick changes in lighting. The quick changes were a result of moving cloud cover, when the sun goes from being totally blocked by clouds to being completely uncovered within a space of around 30 seconds.

In the third row of the examples in the Figure 6, the pedestrian is detected without any false positives. This shows the improvement of this algorithm over a traditional intensity-based technique as shown in Figure 1. The trousers of the pedestrian are a similar shade to the dirty concrete background. Using $\psi = 10$ most of the pedestrian's body is detected, but the trousers are not. The sensitivity of the extraction can be increased, by decreasing ψ in Equation 10, which will enable the full figure of the pedestrian to be detected. However, this will also increase the potential for errors, especially as the algorithm processes compressed video streams. Furthermore, this may be considered a non-issue as it is bordering on a situation of camouflage.

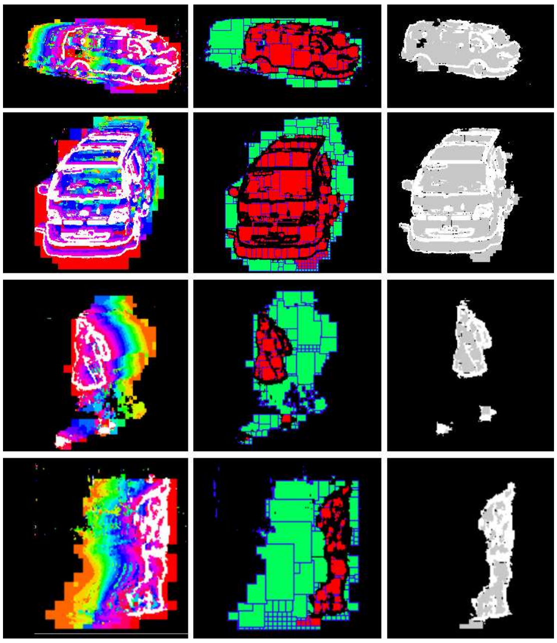


Figure 5: Examples of edge-MHI and the region growing process. Top: A moving van. Bottom: Person walking around a loading bay. Left: Edge-MHI, trails are left behind objects. Middle: Region growing. Right: Final output, trails are removed.

Another issue that this algorithm deals with is quickly appearing sharp shadows caused by buildings or other static objects. These sharp shadows cause changes in relative-intensity and therefore will be detected by a edge-based foreground segmentation. However, the edge-motion-history images and region growing approaches presented in this paper will ignore static edges. Only shadows that are cast by moving objects will result in false detections. Shadows cast by buildings are static and will not have motion history and therefore will not be segmented with the algorithm presented in this paper.

When processing 640×480 images at half-resolution the entire algorithm presented in this paper functions at 20Hz on a 3.2GHz CPU.

6 Conclusion

This paper presents a foreground segmentation technique which uses edge information. Traditional foreground extraction techniques are based on raw-intensity measurements which are sensitive to changes in lighting. Edge information is derived from relative-intensity which is intrinsically robust to global lighting levels.

The three previous edge-based techniques [Davis and Sharma, 2006][Javed *et al.*, 2002][Yokoyama and Poggio, 2005], segment areas without direct information from the images. The edge-based method in this paper can detect complete foreground areas using information gathered from multiple-scales and edge-motion-history-images.

The approach has the inherent insensitivity to uniform lighting of an edge-based approach, with the ability to detect large foreground areas and without creating information and avoiding the problematic clustering process. Another notable advantage is the robustness to static lighting-related edges, which is a result of only considering foreground as connected areas of motion history.

References

[Cucchiara *et al.*, 2003] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25:1337–1342, 2003.

[Davis and Bobick, 1997] James W. Davis and Aaron F. Bobick. The representation and recognition of human movement using temporal templates. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 928, Washington, DC, USA, 1997. IEEE Computer Society.

[Davis and Sharma, 2006] J. Davis and V. Sharma. Background-subtraction in thermal imagery using contour saliency. *Int. Journal of Computer Vision (online)*, 71:161–181, 2006.

[Javed *et al.*, 2002] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Proceedings. Workshop on Motion and Video Computing*, 2002.

[Piccardi, 2004] M. Piccardi. Background subtraction techniques: a review. In *Proc. of IEEE SMC 2004 International Conference on Systems, Man and Cybernetics*, 2004.

[Smith *et al.*, 2004] Paul Smith, Tom Drummond, and Roberto Cipolla. Layered motion segmentation and depth ordering by tracking edges. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(4):479–494, 2004.

[Stauffer and Grimson, 2000] C. Stauffer and W. E. L. Grimson. Learning patterns of acitivty using real-time tracking. *IEEE Trans. on PAMI*, 2000.

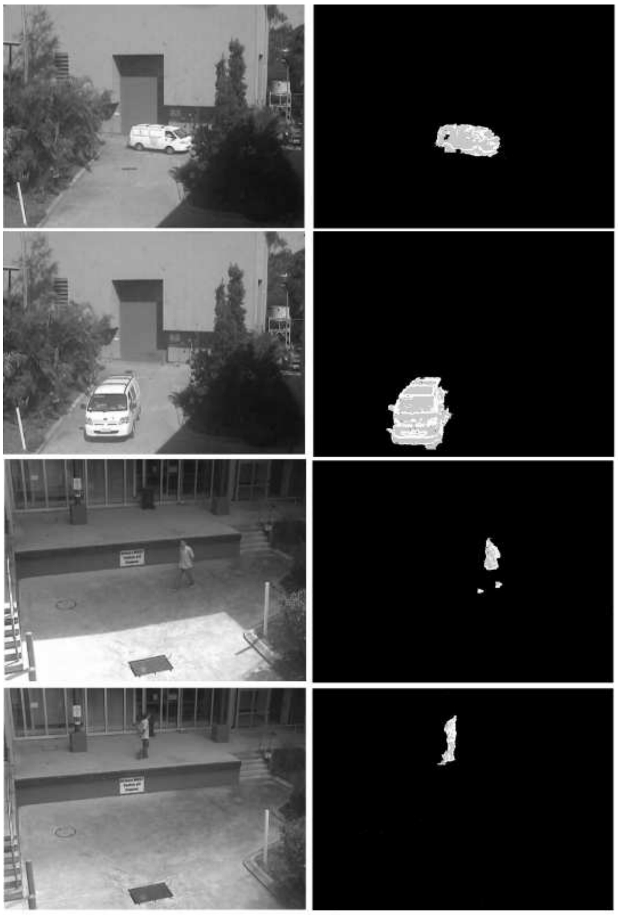


Figure 6: Top: Raw video. Bottom: Final output. There is no false-positive detection, showing an improvement of the algorithm in this paper over the intensity-based method shown in Figure 1.

[Yokoyama and Poggio, 2005] M. Yokoyama and T. Poggio. A contour-based moving object detection and tracking. In *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2005.