Proceedings of the Workshop on Motion and Video Computing (MOTION'02)
0-7695-1860-5/02 $17.00 © 2002 **IEEE**

IEEE COMPUTER SOCIETY

A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information

Omar Javed, Khurram Shafique and Mubarak Shah
Computer Vision Lab,
School of Electrical Engineering and Computer Science,
University of Central Florida
E-mail: {ojaved,khurram,shah}@cs.ucf.edu

Abstract

We present a background subtraction method that uses multiple cues to robustly detect objects in adverse conditions. The algorithm consists of three distinct levels i.e pixel level, region level and frame level. At the pixel level, statistical models of gradients and color are separately used to classify each pixel as belonging to background or foreground. In region level, foreground pixels obtained from the color based subtraction are grouped into regions and gradient based subtraction is then used to make inferences about the validity of these regions. Pixel based models are updated based on decisions made at the region level. Finally frame level analysis is performed to detect global illumination changes. Our method provides the solution to some of the common problems that are not addressed by most background subtraction algorithms such as quick illumination changes, repositioning of static background objects, and initialization of background model with moving objects present in the scene.

1. Introduction

All Automated surveillance systems require some mechanism to detect interesting objects in the field of view of the sensor. Such a mechanism serves as a form of focus of attention. Once objects are detected, the further processing for tracking and activity is limited in the corresponding regions of the image. In vision based systems, such detection is usually carried out by using background subtraction methods. These methods build a model of the scene background, and for each pixel in the image, detect deviations of pixel feature values from the model to classify the pixel as belonging either to background or to foreground. This pixel based information is then grouped to make a similar classification of regions in the image. Though, pixel intensity or color are the most commonly used features for scene modelling, recently some effort has been made to combine this information with edges [7].

The Background differencing methods have to deal with several problems in realistic environments. These problems have been discussed in detail by Toyama et.al [15]. Here we briefly describe some of the important problems which have not been addressed by most background subtraction algorithms.

- Quick illumination changes: Quick illumination changes completely alter the color characteristics of the background, and thus increase the deviation of background pixels from the background model in color or intensity based subtraction. This results in a drastic increase in the number of falsely detected foreground regions and in the worst case, the whole image appears as foreground. This shortcoming makes surveillance under partially cloudy days almost impossible.

- Relocation of the background Object: Relocation of a background object induces change in two different regions in the image, its newly acquired position and its previous position. While only the former should be identified as foreground region, any background subtraction system based on color variation detects both as foreground.

- Initialization with moving objects: If moving objects are present during initialization then part of the background is occluded by moving objects. Thus many algorithms require a scene with no moving objects during initialization. This puts serious limitations on systems to be used in high traffic areas.

- Shadows. Objects cast shadows that might also be classified as foreground due to the illumination change in the shadow region.

In this paper we propose solutions to the first three problems. We have already presented a method to remove cast shadows [9] from a scene. Please see [13] for a detailed review of shadow removing algorithms.

Color based background systems are susceptible to sudden changes in illumination. Gradients of image are relatively less sensitive to changes in illumination and can be combined with color information effectively and efficiently to perform quasi illumination invariant background subtraction. We also note that only pixel level processing is not sufficient for the extraction of foreground from an image sequence. Thus, higher level processing is required to build upon the information obtained from pixel level processing. We use a bottom-to-top hierarchical processing that consists of three different levels, i. Pixel level, ii. Region level and iii. Frame level. In the first level, two features separately are used for background modelling, i.e. color and gradient. In the second level, foreground pixels obtained from the color based subtraction are grouped into regions. Each region is tested for the presence of gradient based foreground pixels at its boundaries. At this level, spurious regions caused by illumination changes are removed. False foreground regions (uncovered background) caused by repositioning of background objects are also detected during this stage. Finally global illumination changes are handled at the third level.

The organization of the paper is as follows. In the next section, we present a brief survey of related work. In Section 3- 5, we give details of each of the three levels of our system hierarchy respectively. In Section 6, we demonstrate the results of our system on a variety of indoor and outdoor sequences. Section 7 concludes the paper.

2. Related Work

A large number of background subtraction methods have been proposed in recent years. Many background subtraction algorithms for fixed cameras work by comparing color or intensities of pixels in the incoming video frame to a reference image. Jain et. al. [8] used simple intensity differencing followed by thresholding. Significant differences in intensity from the reference image were attributed to motion of objects. Azerbyjani et. al. [16] used color images and a statistical model of the background instead of a reference image. The color intensity at each pixel was modelled by a single Gaussian. Stauffer and Grimson [14] extended the uni-modal background subtraction approach by using an adaptive multi-modal subtraction method that modelled the pixel color as a mixture of Gaussians (MOG). This method could deal with slow changes in illumination, repeated motion from background clutter and long term scene changes. The model in Haritouglu et. al. [4] is a simplification of the Gaussian models, where the absolute maximum, minimum and largest consecutive difference values are used. All the above mentioned models use only color or intensity information for background differencing and are susceptible to sudden illumination changes. Moreover, these methods do not attempt to resolve the problem of motion of background object.

Gao et.al [2] compare the assumption of a single vs. mixture of Gaussians to model the background color. They determine that mixture of Gaussian approach is indeed a better representation of backgrounds even in static scenes. Harville [5] presents a framework to update the mixture of Gaussians at each pixel based on feedback from other modules, for example tracking module, in a surveillance system. Pentland et. al. [12] used an eigen space model for background subtraction. The eigen background model can not deal with relocation of a background object.

Our work is most closely related to Jabri et.al [7]. They have used fusion of color and edge information for background subtraction. However the algorithm uses a pixel based fusion measure, such that either a large change in color or edges will result in foreground regions. Therefore their method can not deal with sudden changes in illumination. The background edges are not modelled statistically. Moreover, this algorithm doesn't present a solution to the relocation of background object problem.

Horprasert et. al. [6] use brightness distortion and color distortion measures to develop an algorithm invariant to illumination changes. Li and Leung [10] use the fusion of texture and color to perform background subtraction. The texture based decision is taken over a small neighborhood. Ohta [11] defines a test statistic for background subtraction using the ratio of illumination intensities. Greiffenhagen et. al. [3] propose the fusion of color and normalized color information to achieve shadow invariant change detection. All these algorithms don't use regional information to validate local results. Also these algorithms do not attempt to solve the problem of relocation of background object.

Toyama et. al. [15] propose a three tiered algorithm to deal with the background subtraction problem. The algorithm uses only color information at the pixel level. The region level deals with the background object relocation problem. Global illumination changes are handled at the frame level. This algorithm is able to handle sudden changes in illumination only if the model describing the scene after the illumination changes is known a priori.

3. Pixel Level Processing

At the pixel level, background modelling is done in terms of color and gradient.

3.1. Color based subtraction

We use a mixture of Gaussians method, slightly modified from the version presented by Stauffer and Grimson [14] to perform background subtraction in the color domain. In this method, a mixture of K Gaussian distributions adaptively models each pixel color. The pdf of the kth gaussian at pixel location (i,j) at time t is given as

$$N(x^t_{i,j}|m^{t,k}_{i,j},\Sigma^{t,k}_{i,j}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma^{t,k}_{i,j}|^{\frac{1}{2}}}$$
$$\exp -\frac{1}{2}(x^t_{i,j}-m^{t,k}_{i,j})^T(\Sigma^{t,k}_{i,j})^{-1}(x^t_{i,j}-m^{t,k}_{i,j}), \quad (1)$$

where $x^t_{i,j}$ is the color of pixel i,j, $m^{t,k}_{i,j}$ and $\Sigma^{t,k}_{i,j}$ are the mean vector and the covariance matrix of the kth Gaussian in the mixture at time t respectively. Each Gaussian has an associated weight $\omega^{t,k}_{i,j}$ (where $0 < \omega^{t,k}_{i,j} < 1$)

in the mixture. The covariance matrix is assumed to be diagonal to reduce the computational cost i.e. $\Sigma_{i,j}^{t,k} = diag((\sigma_{i,j}^{t,k,R})^2, (\sigma_{i,j}^{t,k,G})^2, (\sigma_{i,j}^{t,k,B})^2)$ where R,G and B represent the three color components.

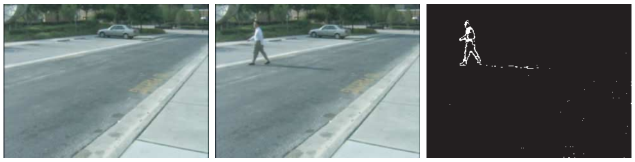
A K-means approximation of the EM algorithm is used to update the mixture model. Each new pixel color value, $x_{i,j}^t$, is checked against the existing K Gaussian distributions, until the pixel matches a distribution. A match is defined if $x_{i,j}^t$ is within a Mahalanobis distance, D, from the distribution. If $x_{i,j}^t$ does not match any of the distributions, the lowest weight distribution is replaced with a distribution having $x_{i,j}^t$ as its mean, a fixed value as initial variance and low prior weight. The model parameters i.e. the weights,means and covariance matrices are updated using an exponential decay scheme with a learning factor.

In the original approach [14], the weights are sorted in decreasing order and the first B distributions are selected as belonging to the background i.e. $B = \arg\min_b(\Sigma_{k=1}^b \omega > T')$. Note that if a higher order process changes the weight of one distribution, it could affect the selection of other distributions as belonging to the background. Since we use input from the region and global levels to update our background, we adopt a method in which changing the weight of one distribution doesn't affect the selection of other distributions into the background. This is achieved by putting a threshold on individual distributions, rather than on their sum. Any distribution with the weight greater than a threshold, T_w, is incorporated in the set of distributions belonging to background.

A connected component algorithm is applied to group all the color foreground pixels into regions. Morphological filtering is performed to remove noise.

3.2. Gradient based subtraction

We use $\Delta = [\Delta_m, \Delta_d]$ as a feature vector for gradient based background differencing, where Δ_m is the gradient magnitude i.e. $\sqrt{f_x^2 + f_y^2}$ and Δ_d is the gradient direction i.e. $tan^{-1}\frac{f_y}{f_x}$. The gradients are calculated from the gray level image.



(a) (b) (c)

Figure 1. Gradient based subtraction results, (a) first image in the sequence (b)150th image (c) gradient based subtraction. There is some noise but it can easily be removed by size based filtering

In order to model the gradient of background intensities, we need to compute the distribution of Δ. We achieve this by initially assuming that for a given pixel (i, j), the highest weighted Gaussian distribution, say kth distribution, models the color background at time t. Let $x_{i,j}^t = [R, G, B]$ be the latest color value that matched the kth distribution at pixel location (i, j), then $g_{i,j}^t = \alpha R + \beta G + \gamma B$ will be its gray scale value. Since we assumed independence between color channels, $g_{i,j}^t$ will be distributed as

$$g_{i,j}^t \sim N(\mu_{i,j}^t, (\sigma_{i,j}^t)^2), \tag{2}$$

where

$$\begin{aligned}\mu_{i,j}^t &= \alpha m_{i,j}^{t,k,R} + \beta m_{i,j}^{t,k,G} + \gamma m_{i,j}^{t,k,B}, \\ (\sigma_{i,j}^t)^2 &= \alpha^2(\sigma_{i,j}^{t,k,R})^2 + \beta^2(\sigma_{i,j}^{t,k,G})^2 + \gamma^2(\sigma_{i,j}^{t,k,B})^2.\end{aligned}$$

Let us define $f_x = g_{i+1,j}^t - g_{i,j}^t$ and $f_y = g_{i,j+1}^t - g_{i,j}^t$. Assuming that gray levels at each pixel location are independent from neighbouring pixels, we observe that

$$f_x \sim N(\mu_{f_x}, (\sigma_{f_x})^2), \tag{3}$$

$$f_y \sim N(\mu_{f_y}, (\sigma_{f_y})^2). \tag{4}$$

where

$$\begin{aligned}\mu_{f_x} &= \mu_{i+1,j}^t - \mu_{i,j}^t, &(5)\\ \mu_{f_y} &= \mu_{i,j+1}^t - \mu_{i,j}^t, &(6)\\ (\sigma_{f_x})^2 &= (\sigma_{i+1,j}^t)^2 + (\sigma_{i,j}^t)^2, \\ (\sigma_{f_y})^2 &= (\sigma_{i,j+1}^t)^2 + (\sigma_{i,j}^t)^2.\end{aligned}$$

Note that even if gray values are assumed to be independent from each other, f_x and f_y are not independent. The covariance is given by,

$$\begin{aligned}Cov(f_x, f_y) &= Cov(g_{i+1,j}^t - g_{i,j}^t, g_{i,j+1}^t - g_{i,j}^t) \\ &= Cov(g_{i,j}^t, g_{i,j}^t) = (\sigma_{i,j}^{t,k})^2 \qquad (7)\end{aligned}$$

Knowing the distribution of f_x and f_y, and using standard distribution transformation methods [1], we determine the distribution of feature vector $[\Delta_m, \Delta_d]$:

$$F(\Delta_m, \Delta_d) = \frac{\Delta_m}{2\pi\sigma_{f_x}\sigma_{f_y}\sqrt{1-\rho^2}} \exp\left(-\frac{z}{2(1-\rho^2)}\right), \tag{8}$$

where

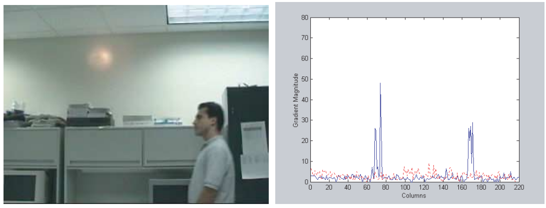
$$\begin{aligned}z = &\left(\frac{\Delta_m \cos\Delta_d - \mu_{f_x}}{\sigma_{f_x}}\right)^2 \\ &-2\rho\left(\frac{\Delta_m \cos\Delta_d - \mu_{f_x}}{\sigma_{f_x}}\right)\left(\frac{\Delta_m \sin\Delta_d - \mu_{f_y}}{\sigma_{f_y}}\right) \\ &+\left(\frac{\Delta_m \sin\Delta_d - \mu_{f_y}}{\sigma_{f_y}}\right)^2,\end{aligned}$$

$$\rho = \frac{\sigma_{i,j}^2}{\sigma_{f_x}\sigma_{f_y}}.$$

Note that for zero means, unit variances and $\rho = 0$, the above given distribution becomes a Rayleigh distribution. All the parameters in the above distribution can be calculated from the means and variances of the color distributions.

For each incoming frame, gradient magnitude and direction values are computed. If for a certain gradient vector, the probability of being generated from the background gradient distribution is less than T_g, then pixel belongs to foreground, other wise it belongs to the background. There is no need to explicitly update the parameters of the background gradient distribution since all the parameters can be computed from the updated color background parameters.

Now if multiple color Gaussians belong to the background model, then we can generate gradient distributions for all possible combination of the neighboring background Gaussian distributions. A pixel will belong to the gradient background model if it belongs to any of these gradient distributions. See figure 1 for an example of background gradient subtracted image.



(a) (b)

Figure 2. (a) Image with a person and highlight (generated by a flash light (b) Plot of gradient magnitudes at rows 60 (dashed) and 180 (solid). The plot shows that the person has high gradients at the boundaries. The highlight is diffused at the boundaries and therefore gradients are small

4. Region Level Processing

The edge and color information obtained from pixel level is integrated at the region level. The basic idea is that any foreground region that corresponds to an actual object will have high values of gradient based background difference at its boundaries. The idea is explained in detail in the following paragraphs.

Let I be the current frame and Δ be the gradient feature vector of its gray levels. Also, let $C(I)$ and $G(I)$ be the output of color based and gradient based subtraction respectively. $C(I)$ and $G(I)$ are binary images such that $C(i,j) = 1$ iff the pixel at location (i,j) is classified as foreground by color based subtraction. Similarly $G(i,j) = 1$ iff the pixel at location (i,j) is classified as foreground by gradient based subtraction. Let $1 \leq a \leq k$ be the k regions in $C(I)$ that are detected as foreground. For any region R_a such that R_a corresponds to some foreground object in the scene, there will be a high gradient at ∂R_a in the image I, where ∂R_a is the set of boundary pixels (i,j) of region R_a. Thus it is reasonable to assume that Δ will have high deviation from the gradient background model at ∂R_a, i.e. $G(I)$ must have a high percentage of ON pixels in ∂R_a. However, if a region R_b corresponds to a falsely detected foreground induced by local illumination changes, for example highlights, then there will be a smooth change in I at ∂R_b, see figure 2. Thus the gradient of I at ∂R_b is not much different than the gradient of background model, hence producing low percentage of ON pixels in $G(I)$. Following this intuition, boundaries ∂R_a of each detected region R_a in $C(I)$ are determined. If p_B percent of boundary pixels also show up in $G(I)$ then the object is declared valid. Otherwise it is determined to be a spurious object caused by an illumination change or noise.

Now, consider a background object that is repositioned to a new location in the scene thus inducing two regions say R_x and R_y in $C(I)$ corresponding to its newly acquired position and its previous position respectively. We want to classify R_x as a foreground region and R_y as a background region. Though R_x will usually have high percentage of boundary pixels that are ON in $G(I)$, the same is also true for R_y. This is due to the presence of edges in the background model at ∂R_y and the absence of edges in I at the same location. It follows that a foreground region R should not only have higher percentage of ON boundary pixels in $G(I)$ but these pixels should also lie on some edge of image I. Formally, we classify a region as a foreground region if

$$\frac{\sum_{(i,j)\in\partial R_a} (\nabla I(i,j) G(i,j))}{|\partial R_a|} \geq p_B. \qquad (9)$$

where ∇I denotes the edges of image I and $|\partial R_a|$ denotes the number of boundary pixels of region R_a.

Once a region R_a is identified as falsely detected, then for all pixels (i,j) in R_a, the weight of the color distribution that matched to $x^t_{i,j}$ is increased to have a value more than T_w.

5. Frame Level Processing

The Frame level process kicks in if more than 50 percent of the color based background subtracted image becomes a part of the foreground. The Frame level processor then ignores the color based subtraction results. Thus only gradient information is used for subtraction. Connected components algorithm is applied to the gradient based subtraction results i.e. $G(I)$. Only the bounding boxes belonging to the edge based results are considered as foreground regions. When the frame level model is active the region level processing is not done, since the color based subtraction is presumed to be completely unreliable at this point.

6. Results

The system was tested on a variety of indoor and outdoor sequences. The same thresholds were used for all the sequences. The values of important thresholds used were $T_g = 10^{-3}$, $p_b = .2$.

Proceedings of the Workshop on Motion and Video Computing (MOTION'02)
0-7695-1860-5/02 $17.00 © 2002 **IEEE**

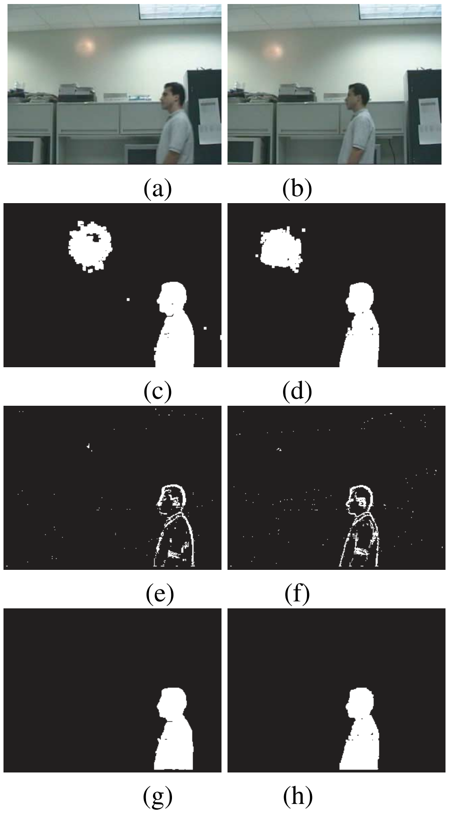


Figure 3. (a)(b) Two frames from a sequence with a person and a highlight in the foreground. (c)(d) color background subtraction using Mixture of Gaussians. Note that Illumination change is causing spurious foreground regions. (e)(f) edge based background subtraction (g)(h) edge and color combined results.

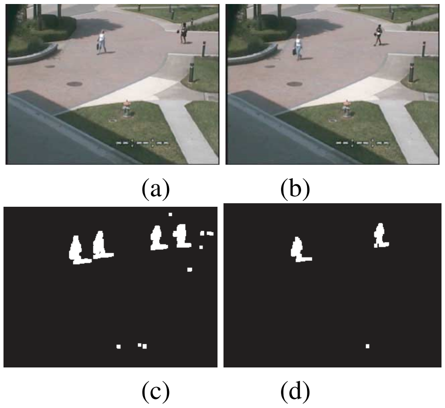


Figure 4. Initialization example (a) first image in the sequence (b)30th image (c) color based results. Ghosts are visible on uncovered background. (d)edge and color combined results. The ghosts were removed because their boundary pixels did not contain significant edges

In one particular sequence a flash light was used to direct a beam of light in the scene to create a local illumination change. In this case the color based algorithm generated regions representing the change, but the gradient based algorithm did not respond to the illumination change as anticipated (see figure 3).

In another test sequence, there were moving persons in the scene during initialization. As long as there was overlap between uncovered background and moving object, both areas were shown as foreground. However as soon as there was no overlap between the uncovered area and the moving objects, the region level process removed the uncovered areas from the foreground since edges did not delineate their boundaries. Please see figure 4.

In outdoor sequences we have observed that a sudden illumination change over the complete area under observation is a rarity. Quick Illumination changes are caused by movement of clouds in front of the sun. Therefore illumination change starts from one area of the image and then sweeps through the image. Since the background color model is continuously updated by forcing the distribution of false objects into the background, only those areas of image in which illumination changed in consecutive frames show up as foreground in color based subtraction. Background subtraction results during illumination change, for both mixture of Gaussians method ([14]) and our proposed algorithm, are presented in Figure 5.

More results are available on http://www.cs.ucf.edu/∼ vision/projects/Knight/background.html

7 Conclusion

We have presented a method for object detection in video sequences which uses both color and gradient information. The detection is performed at pixel, region and frame levels. We use the presence of high gradients on object boundaries to remove spurious objects and illumination changes . The pixel based models are updated based on the decisions made at the higher levels. We have compared our results with the widely used Mixture of Gaussian background subtraction method.

References

[1] G. Casella and R. Berger. *Statistical Inference*. Duxbury, 2 edition, 2001.

[2] X. Gao, T.E. Boult, F. Coetzee, and V. Ramesh. " Error analysis of background subtraction". In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2000.

[3] M. Greiffenhagen, V. Ramesh, and H. Nieman. " The systematic design and analysis of a vision system: A case study in video surveillance". In *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2001.

[4] I. Haritaoglu, D. Harwood, and L.S. Davis. "W4: Realtime surveillance of people and their activities". *IEEE Trans. on PAMI*, 22(8):809–830, Aug 2000.

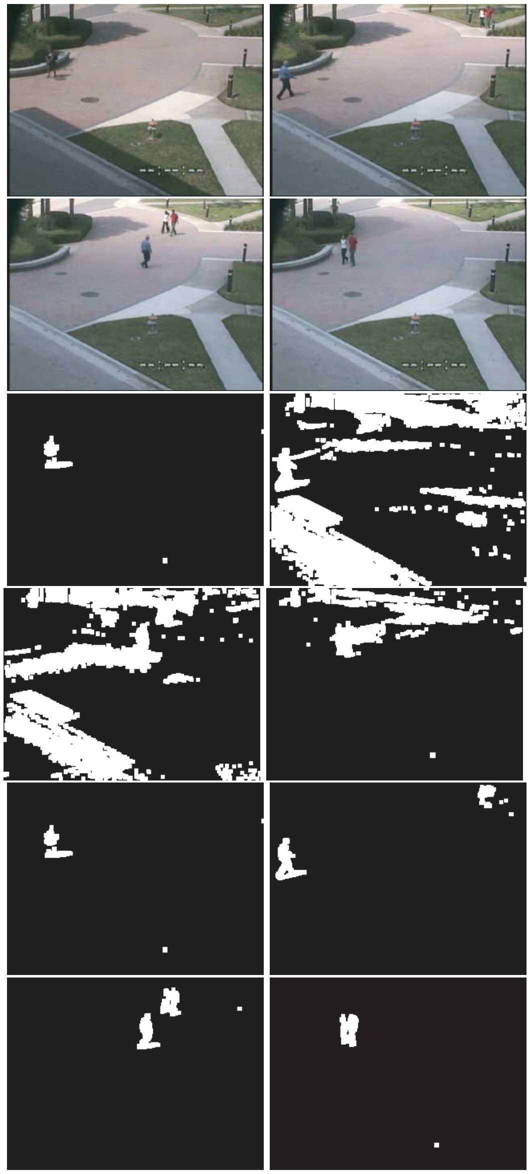


Figure 5. Illumination change in out door scenes. The first two rows show images at an interval of 70 frames from a sequence. In the first image there is a shadow on bottom left , but it disappears in subsequent frames. Note that there are three people in the second image, two people close together on the top right corner and one on the left. The 3rd and the 4th row show the color based results using the Mixture of Gaussians method ([14]) with no feedback from higher level. The last two rows shows the hierarchical background subtraction results. All the people have been successfully detected.

[5] Michael Harville. " A framework for high-level feedback to adaptive per-pixel mixture of gaussian models". In *Proceedings of European Conference on Computer Vision*, 2002.

[6] T. Horprasert, D. Harwood, and L. Davis. "A statistical approach for real time robust background subtraction and shadow detection ". In *IEEE Frame Rate Workshop*, 1999.

[7] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. " Detection and location of people using adaptive fusion of color and edge information". In *Proceedings of International Conference on Pattern Recognition*, 2000.

[8] R. Jain, D. Militzer, and H. Nagel. "Separating non-stationary from stationary scene components in a sequence of real world tv-images". *IJCAI*, pages 612–618, 1977.

[9] O. Javed and M. Shah. "Tracking and object classification for automated surveillance". In *The seventh European Conference on Computer Vision*, 2002.

[10] L. Liyuan and L. Maylor. "Integrating intensity and texture differences for robust change detection". *IEEE Trans. on Image Processing*, 11(2):105–112, Feb 2002.

[11] N. Ohta. "A statistical approach to background subtraction for surveillance systems ". In *International Conference on Computer Vision*, 2001.

[12] N.M. Oliver, B. Rosario, and A.P. Pentland. "A bayesian computer vision system for modeling human interactions". *IEEE Trans. on PAMI*, 22(8):831 –843, Aug 2000.

[13] A. Prati, R. Cucchiara, I. Mikic, and M.M. Trivedi. "Analysis and detection of shadows in video streams: a comparative evaluation". In *International Conference on Computer Vision and Pattern Recognition*, 2001.

[14] C. Stauffer and W. E. L. Grimson. "Learning patterns of acitivty using real-time tracking". *IEEE Trans. on PAMI*, 22(8):747–757, Aug 2000.

[15] K. Toyama, B. Brumitt J. Krumm, and B. Meyers. "Wallflower: Principles and practicle of background maintenance". In *Proceedings of International Conference on Computer Vision*, 1999.

[16] C. Wren, A Azarbayejani, T. Darrel, and A. Pentland. " Pfinder, real time tracking of the human body". *IEEE Trans. on PAMI*, 19(7), Aug 1997.