Robust Change-Detection by Normalised Gradient-Correlation

Robert O'Callaghan
Mitsubishi Electric ITE
Visual Information Laboratory
Guildford, GU2 7YD, United Kingdom

Tetsuji Haga
Mitsubishi Electric Corporation
Advanced Technology R&D Center
Amagasaki, Hyogo 661-8661, Japan

Abstract

A novel algorithm for robustly segmenting changes between different images of a scene is presented. This computationally efficient algorithm is based on a non-linear comparison of gradient structure in overlapping image-regions and offers intrinsic invariance to changing illumination, without recourse to background-model adaptation. High accuracy is demonstrated on test video data with and without illumination changes. The technique is applicable to motion-segmentation as well as measuring longer-term object-changes.

1. Introduction

Identifying significant changes between different images of the same scene is a fundamental operation in many computer vision applications. In this paper, we focus on the application of change detection to video surveillance, although the techniques developed are general.

CCTV is an example of an application where the volume of video data captured is usually far greater than the capacity for manual viewing and analysis. Much of this data may be "uninteresting" and so robust automated methods are needed to filter the content. Change detection can be seen as fundamental, because it enables analysis functions such as:

- Motion-detection
- Detection of object appearance / disappearance
- Intrusion-detection / Sterile-zone monitoring

It is also the foundation for more complex analysis, such as feature extraction for recognition or content-based retrieval, or providing object-initialisation for motion-tracking algorithms.

Given the utility of change-detection, it is unsurprising that a great many techniques can be found in the literature. A selection of the most important will now be reviewed in Section 2, followed by an analysis of their drawbacks in Section 3. The proposed algorithm is described in Section 4 and results of experimental evaluation in Section 5.

2. Review of Existing Methods

The relevant existing work could be categorised in several ways. In the following, we consider three modes of image-comparison: pixel-wise intensity (Section 2.1), pixel-wise gradient (Section 2.2) and neighbourhood intensity (Section 2.3). Section 2.4 describes a separate method of preprocessing for illumination compensation. It is also useful to note that the methods of Sections 2.2 and 2.3 can be viewed as texture-based, in contrast to the intensity methods referenced in Section 2.1.

2.1. Background Subtraction

Among the most elementary of techniques is the idea of subtracting a reference or "background" image from the current frame—*i.e.*, frame differencing or background subtraction. In the most basic application of this approach, the reference image at each point in time is simply the previous image. More complex background models can lead to improved performance: various methods have been proposed, which integrate the history of previous frames, including linear and non-linear filtering, prediction and statistical modelling. A review and evaluation of several of these is given by Cheung and Kamath [4].

One very well-known method is the so-called "Mixture-of-Gaussians" by Stauffer and Grimson [15]. In this algorithm, a statistical model of background intensity/colour is built for each pixel. This is expressed as a mixture model of a small number of Gaussian distributions, with each characterised by mean and standard deviation. The idea is that dynamic backgrounds (such as waving foliage), which give rise to a multi-modal intensity distribution, can be better captured by several mixture-components than by a single Gaussian distribution. For each new frame, each pixel is compared to every component in its mixture model and a match declared if it lies within 2.5 standard deviations of

the mean. The parameters of the model for each pixel are then updated, before proceeding to the next frame.

Elgammal *et al*. [7] replace the parametric mixture model with non-parametric kernel density estimation. This is a more general model, which (compared to [15]) avoids the need to estimate the model order and parameters of the mixture components.

Another multi-modal background model is used in the "W4" system, from Haritaoglu *et al*. [9]. In this case, a training period is first used to learn the background model from stationary pixels in the video sequence. This model is characterised by the minimum, maximum and maximum-difference values over the training period.

2.2. Gradient methods

Javed *et al*. [10] combine colour and edge-based background subtraction. They build Gaussian models not only of the pixel colour statistics but also of the gradient magnitude and direction.

2.3. Neighbourhood Change-Detection

Several authors have observed that the pixel-wise subtraction approach, although useful, has a number of drawbacks—especially in the context of changing scene illumination. In response, they have proposed a variety of local, neighbourhood-based methods.

Durucan and Ebrahimi [6] propose to model each pixel by concatenating pixel values in the surrounding neighbourhood into a vector. To measure change, the authors develop two different measures to assess whether a pair of such vectors are linearly independent. The underlying assumption is that changing illumination can be modelled by linear scaling of the vectors, so that measuring the degree of linear independence will be more robust to such changes than simple subtraction.

Nagaya *et al*. [14] have a similar insight. They construct a vector from a "slit" or linear slice of pixels in an image and observe that illumination changes will modify the magnitude of this vector, but not its direction. They therefore define a normalised vector distance, in which the vectors are first projected onto the unit sphere and then compared. This distance is invariant to scaling and hence the algorithm should be robust to illumination change. Aoki *et al*. [2] construct their vectors from blocks of pixels and note that an equivalent normalised distance can be calculated by using the inner product to directly measure the angle between two vectors.

Matsuyama *et al*. [13] further develop the concept of normalised vector distance, by recording spatial and temporal statistics of the distance values. In so doing, they aim to better adapt the decision function to the content of the image sequence.

Kondo *et al*. [11] use normalised correlation to compare vectors of pixel values in a neighbourhood. This distance measure achieves invariance not only to linear scaling, but also to additive factors between the two images under consideration.

Li and Leung [12] combine the concept of neighbourhood-based comparison with the use of image gradients, to measure differences in image texture. However, the distance function they propose remains susceptible to changes in illumination (as modelled by linear scaling).

2.4. Illumination Compensation

Bassman *et al*. [3] also model illumination change by linear scaling of pixel values. However, they treat it as a global effect and hence estimate the scaling factor and offset over the whole frame. These parameters are used to modify the reference image to compensate for the change, before performing regular pixel-wise subtraction.

3. Limitations

It is very important that automated visual analysis algorithms are robust to visual effects in real scenes, caused by a changing environment. This is particularly true in security-critical applications such as visual surveillance. The system must be able to distinguish between events (and potential threats) of interest and image changes due to benign variations. Changing lighting conditions are one such benign, but ubiquitous variation.

Most methods based on background-subtraction are sensitive to illumination change, because they measure intensity change over time. Sometimes, the effects are mitigated by a background update mechanism, as in [15]; however, this is not the original purpose of such adaptation and it is itself inherently fragile. Illumination compensation by background-update relies on a fundamental assumption that object dynamics are much faster than illumination dynamics. This may hold some of the time, but fast lighting changes are possible both indoors and outdoors. In the extreme case, we may consider time-lapse video sequences, where even "slow" illumination changes become "fast".

Several of the methods reviewed above do seek to offer robustness to the effects of illumination; however, they are based almost exclusively on linear models - either in a local neighbourhood or globally across the image. The global assumption quickly breaks down in natural, complex scenes: it is rare for a three-dimensional scene to have either uniform illumination or uniform changes in illumination throughout. The local assumption may hold in some parts of the image, such as across a single object or surface but inevitably, in the presence of inter-reflection, shadowing and inhomogeneous illumination, there will be discontinuities (both object boundaries and illumination boundaries)

across which the linear model will fail.

4. Robust Change Detection

The proposed Normalised Gradient-Correlation method uses a more flexible model of illumination change for real-world scenes in order to improve robustness. This model is based on the spatial derivatives of image intensity and includes a non-linear component, to deal with illumination changes at object boundaries. Its inherent robustness means that the model can avoid over-reliance on a background-update mechanism to accommodate changing illumination.

4.1. Normalised Gradient-Correlation

The idea behind the method is that the partial derivatives of image intensity form a better basis for achieving invariance to illumination change than using intensity values directly. The algorithm is therefore designed to detect and segment objects, by measuring the changes in gradient structure that occur between a pair of images.

Firstly, consider the following definition of the normalised correlation of two vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$, which is the cosine of the angle, θ, between them:

$$\Gamma(\mathbf{v}_1, \mathbf{v}_2) = \cos\theta = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \, \|\mathbf{v}_2\|} \tag{1}$$

where $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. This measure is invariant to linear scaling of the form $\mathbf{v}' = \alpha\mathbf{v}$. Now consider what happens if the vectors are formed from the partial derivatives $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial y}$ of each image, $f(x, y)$. The expanded correlation calculation is given by (2), where the bounds of the summations over x and y reflect a choice of neighbourhood for comparison (as described in the next section).

The derivative operations (which can be approximated using a linear operator such as the Sobel filter) provide invariance to additive constants, with the cosine measure (1) providing invariance to linear scaling. Therefore, variations of the form $\mathbf{v}' = \alpha\mathbf{v} + \beta$ can be accommodated. This is a reasonable model for effects such as brightness and contrast change; however, illumination change is likely to be more complex.

Figure 1 illustrates one of the more complex effects of varying illumination. It represents an image region containing a three-dimensional object-edge, illuminated from the right and the left, in Figures 1(a) and 1(b), respectively. Obviously, this is an idealised version of light and shadow in real scenes, but it explains how a gradient may switch polarity under illumination change. Note that, although the sign of the gradient is switched, the orientation of the edge is preserved.

One way to deal with this in the comparison measure is to use the absolute value of the cosine measure—meaning that only the magnitude of the correlation is of interest, not its sign. Returning to our vector notation, this would mean invariance to the transformation $\mathbf{v}' = \pm\alpha\mathbf{v}$. However, a further problem arises if some (but not all) of the components of the vector change sign. This corresponds to the case where gradients at only some points in a neighbourhood have their polarity reversed (*e.g.* in the vicinity of an illumination-discontinuity or object-boundary):

(a) (b)

Figure 1. Directional effects of illumination

$$\begin{aligned}
\frac{\partial f'(x_1,y_1)}{\partial x} &= -\alpha\frac{\partial f(x_1,y_1)}{\partial x}, \\
\frac{\partial f'(x_1,y_1)}{\partial y} &= -\alpha\frac{\partial f(x_1,y_1)}{\partial y}, \\
\frac{\partial f'(x_2,y_2)}{\partial x} &= \alpha\frac{\partial f(x_2,y_2)}{\partial x}, \\
\frac{\partial f'(x_2,y_2)}{\partial y} &= \alpha\frac{\partial f(x_2,y_2)}{\partial y}
\end{aligned} \tag{3}$$

A trivial solution is to take the absolute values of the individual components of the gradient, but this also allows half-plane symmetric variations in local gradient orientations (rather than the simple polarity switch we wish to admit).

A better (*i.e.* more selective) way to accommodate such local variations is to modify the numerator of the correlation function (2) to include a non-linear absolute value operation in the inner product:

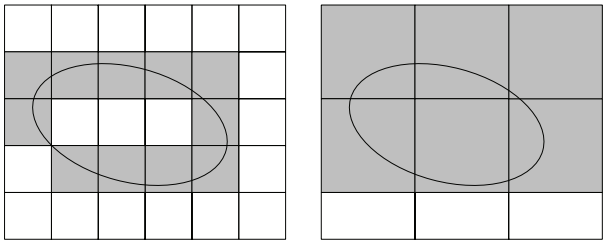
$$\Gamma(\mathbf{v}_1, \mathbf{v}_2) = \frac{\sum_{x,y} \left| \frac{\partial f_1(x,y)}{\partial x}\frac{\partial f_2(x,y)}{\partial x} + \frac{\partial f_1(x,y)}{\partial y}\frac{\partial f_2(x,y)}{\partial y} \right|}{\sqrt{\sum_{x,y}(\ldots)\sum_{x,y}(\ldots)}} \tag{4}$$

This provides specific invariance to polarity switches of the gradient direction at individual pixels in either of the images compared.

4.2. Multi-scale Processing

The modified normalised correlation defined above provides a means to detect changes in gradient structure in image neighbourhoods—corresponding to the extent of the summation over pixels, (x, y). Obviously, changes (*e.g.*, objects) can occur with a range of sizes; moreover, a neighbourhood comparison of this type is subject to an "aperture effect", whereby changes will not be detected if measured

$$\Gamma(\mathbf{v}_1, \mathbf{v}_2) = \frac{\sum_{x,y} \left(\frac{\partial f_1(x,y)}{\partial x} \frac{\partial f_2(x,y)}{\partial x} + \frac{\partial f_1(x,y)}{\partial y} \frac{\partial f_2(x,y)}{\partial y} \right)}{\sqrt{\sum_{x,y} \left(\frac{\partial f_1(x,y)}{\partial x}^2 + \frac{\partial f_1(x,y)}{\partial y}^2 \right) \sum_{x,y} \left(\frac{\partial f_2(x,y)}{\partial x}^2 + \frac{\partial f_2(x,y)}{\partial y}^2 \right)}} \tag{2}$$



(a) Small block-size (b) Large block-size

Figure 2. Aperture problem. The shaded blocks denote detector-activation.

at the wrong scale. This is illustrated in Figure 2, which represents an image of an elliptical object of uniform intensity against a background of different (but also uniform) intensity. It is assumed that the reference image contains only the uniform background. In the first case, shown in Figure 2(a), the block-wise comparison of structure detects change in the hatched blocks around the periphery of the object but not in the interior. The central blocks are not activated by the presence of the object, because the change in intensity can be modelled by an illumination change. In Figure 2(b), the larger blocks result in complete coverage. Of course, a penalty is paid in terms of reduced localisation accuracy, because the overall area detected is increased. For this reason, it is important to integrate measurements over a range of scales (*i.e.*, neighbourhood sizes).

The missed detections of Figure 2(a) occur because the presence of a uniform object on a uniform background does not modify the gradient structure. Note that this problem is not exclusive to gradient-correlation: the intensity change would also be interpreted by intensity-correlation as a linear scaling due to illumination change; however, pixel-based differencing methods such as [15] would obviously detect correctly in this case.

To mitigate aperture problems, the current method is implemented in a multi-scale form. The correlation is computed for square blocks at three scales: 8x8, 16x16 and 32x32. At each scale, adjacent blocks have a 50% overlap in area and the correlation results are assigned to a smaller, central output-block so that they tightly fill the image. The results are summed across the scales, to produce a final correlation map. It is then straightforward to detect changes by comparing the map to an appropriate threshold, T (which may be chosen with reference to application-specific priors on the frequency of target appearance, or the relative cost of misses compared to false alarms).

Note that it is also possible to implement a multi-resolution version of the algorithm, in which the images are first sub-sampled to a variety of sizes before gradients are extracted. The size of the correlation blocks can then be fixed (across resolutions). However, in experiments, this approach was not found to produce any improvement, despite the added computational complexity entailed; therefore, in the current implementation, the gradients are extracted at a single scale and processed in blocks of different sizes, as described.

4.3. Efficient Implementation

Under the proposed block-shift pattern, each pixel is included in the correlation calculation of four blocks, at each of three scales. In order to minimise the computational burden, an efficient implementation has been devised. In the first stage, the following product-images are calculated:

$$p(x,y) = \left| \frac{\partial f_1(x,y)}{\partial x} \frac{\partial f_2(x,y)}{\partial x} + \frac{\partial f_1(x,y)}{\partial y} \frac{\partial f_2(x,y)}{\partial y} \right| \tag{5}$$

$$e_1(x,y) = \left(\frac{\partial f_1(x,y)}{\partial x}^2 + \frac{\partial f_1(x,y)}{\partial y}^2 \right) \tag{6}$$

$$e_2(x,y) = \left(\frac{\partial f_2(x,y)}{\partial x}^2 + \frac{\partial f_2(x,y)}{\partial y}^2 \right) \tag{7}$$

The use of product-images means that the multiplications at each pixel are not repeated. Calculation of each correlation value will now rely only on selection and summation of elements from these product-images, followed by a single multiplication, square root and division.

The second stage is to efficiently implement the selection and summation operations. This can be achieved by the method of integral images [16] (also known as summed area tables [5]). Transformation of the data into the integral representation provides fast computation, with only four pixel references (the co-ordinates of corners of the rectangles) and three addition operations per summation. Importantly, this makes the computational cost of the summations constant, regardless of the block size.

Following this approach, we generate integral product

images as follows:

$$P(x,y) = \sum_{a \le x} \sum_{b \le y} p(a,b) \tag{8}$$

$$E_1(x,y) = \sum_{a \le x} \sum_{b \le y} e_1(a,b) \tag{9}$$

$$E_2(x,y) = \sum_{a \le x} \sum_{b \le y} e_2(a,b) \tag{10}$$

This can be effected, in each case, by a single pass over the image, with two addition operations per pixel.

An alternative to the integral-image approach is to use a linear filter to average the product-images. This has the potential advantages of producing a dense output (a correlation calculated independently for each pixel) and allowing the pixels to be weighted, by means of the filter coefficients. Centre-weighting, with a Gaussian kernel, for example, leads to improved localisation of changes. The drawback, of course, is the increase in computational complexity.

4.4. Effect of Noise

Recorded image-intensities are subject to fluctuations due to noise. This can affect a system based on gradient-correlation more than some others, since derivatives are naturally more susceptible to noise. This becomes a particular problem in regions of uniform intensity: the mean-removing effect of the derivative will, in the worst case, leave only noise remaining. This noise is unlikely to be correlated in time (*i.e.*, between frames) and so the gradient-correlation results for such regions will be very low—often giving rise to false alarms.

Spatio-temporal smoothing can be used to suppress noise in the input images, but measures can be taken in the correlation operation itself. This relies on setting to zero gradient components whose magnitude falls beneath some threshold. The magnitude of the gradient is defined in the usual way, as:

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2\right)} \tag{11}$$

The idea is to ignore "small" gradient values based on the assumption that these are more likely to be generated by noise than large responses. A threshold value of $|\nabla f| = 45$ was used in all the experiments of Section 5 (where ∇f is calculated using the Sobel filter, as described previously).

4.5. Extension for Background Modelling

While the sparseness of gradient responses has drawbacks such as the aperture problem and sensitivity to noise, it also has additional advantages. One of these is in modelling the background. Many background-subtraction algorithms use the mean of a set of historical frames as the reference image for change detection; however, this does not translate well to illumination-invariant comparison, since the mean of the intensity-images will be corrupted by illumination change. The gradient components, on the other hand, have a significant degree of innate robustness, since they are invariant to additive constants. By accumulating the derivative images and taking the median in the temporal dimension, reference gradient images with an even greater degree of robustness can be obtained. Indeed this property of gradient images has been used in the past to extract the "intrinsic" reflectance images from scenes under varying illumination [17]. The use of the median will also act to reduce the influence of noise, as noted in Section 4.4 above.

5. Evaluation

The results of two experiments, on different datasets, are presented. Both test the medium-term robustness to illumination-change in a typical object detection scenario. The Ground Truth (GT) is in the form of bounding boxes for each human target. Performance metrics proposed in [8] are adopted. These are defined in terms of Good Detections (GD), False Detections (FD) and Missed Detections (MD). For blobs (*i.e.*, connected regions of detected pixels):

- GD: GT-boxes having sufficient overlap with blobs
- MD: GT-boxes having insufficient overlap with blobs

For the purposes of the current evaluation, the threshold for "sufficient overlap" for detection was set at 20% of GT-box area. For pixels:

- GD: pixels in both the GT set and the blob set
- FD: pixels in the blob set but not the GT set

The Precision (P) and Recall (R) are now defined as:

$$P_{\text{pixels}} = \frac{GD_{\text{pixels}}}{GD_{\text{pixels}} + FD_{\text{pixels}}} \tag{12}$$

$$R_{\text{blobs}} = \frac{GD_{\text{blobs}}}{GD_{\text{blobs}} + MD_{\text{blobs}}} \tag{13}$$

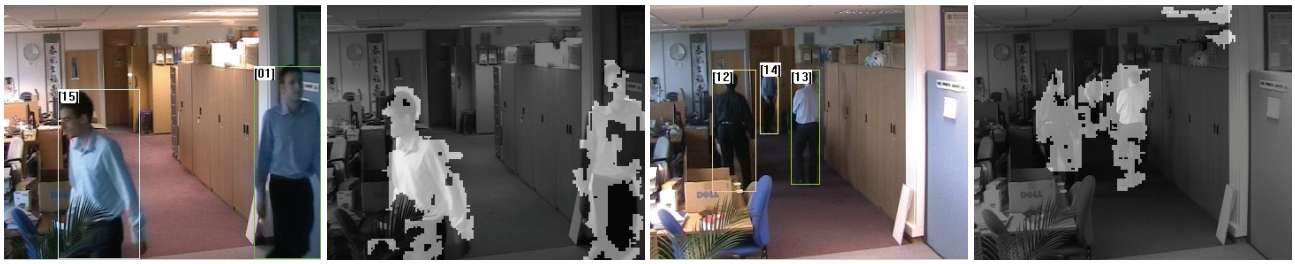
The reason for mixing pixels and blobs in this way is that neither P_{blobs} nor R_{pixels} are reliable metrics: pixel-recall is unfair, in that targets never fully occupy their bounding boxes; blob-precision can be drastically reduced by a small number of false positives, even if these are tiny (in area) compared to the correctly detected targets.

5.1. Office Sequences

The test video sequences for the first experiment represent a typical office environment. Results are reported for two challenging sequences, with complex backgrounds and sudden, dramatic illumination changes, due to lights

(a) "Reference" illumination (b) Output for (a) (c) Changed illumination (d) Output for (c)

Figure 3. Example of changing illumination in Office Sequence I; (a) and (c) show the GT bounding boxes



(a) "Reference" illumination (b) Output for (a) (c) Changed illumination (d) Output for (c)

Figure 4. Example of changing illumination in Office Sequence II; (a) and (c) show the GT bounding boxes

Table 1. Office Sequence I

	Precision (%)	Recall (%)
Proposed Method	91	95
Mixture-of-Gaussians	46	51
Intensity Correlation	82	94

Table 2. Office Sequence II

	Precision (%)	Recall (%)
Proposed Method	84	87
Mixture-of-Gaussians	38	76
Intensity Correlation	64	87

switching on and off (during daylight hours). Sample images of the two scenes and their segmentation results using the proposed method are given in Figures 3 and 4, respectively. Unfortunately, the authors are not aware of any public dataset containing a significant density of sudden illumination changes, therefore the private video data for this experiment was chosen out of necessity. The GT was manually annotated. Results are given in Tables 1 and 2

Two baseline methods were used to compare the performance of the Normalised Gradient-Correlation algorithm: an implementation of the Mixture-of-Gaussians background-subtraction algorithm [15] and a version of the normalised correlation approach using intensity images in place of gradients (representative of the normalised vector-distance methods [2, 6, 11, 13, 14]). Each of the correlation algorithms used a single background reference frame, empty of targets. The Mixture-of-Gaussians approach relies on adaptation to changes in the scene; therefore, it generates its own background model, on-line (although this model is initialised by the empty reference frame). The Mixture-of-Gaussians and Normalised Gradient-Correlation algorithms were used with unchanged parameter values for both sequences ($\alpha = 0.0001$ and $T = 0.58$, respectively); however, for the intensity correlation baseline algorithm only, it was necessary to adjust the threshold to rebalance the precision/recall bias for each sequence, in order to make the results directly comparable.

The results show that Normalised Gradient-Correlation has the highest performance in the presence of illumination changes, followed by normalised intensity-correlation. As expected, both outperform the non-invariant Mixture-of-Gaussians algorithm—particularly in terms of precision. The first sequence has several temporarily-static targets, which also reduces the recall of Mixture-of-Gaussians, because these static targets tend to be absorbed into the background. The second sequence has more complex and frequent illumination changes, reducing precision across the board; however, there are very few static targets, so the recall of Mixture-of-Gaussians is relatively higher.

The false-alarm blobs in the top-right of Figure 4(d) reveal a limitation of the illumination-model used by Normalised Gradient-Correlation: the fundamental assumption

is that of objects reflecting ambient incident illumination. There is no allowance for luminous elements, like the ceiling lights visible in the scene of Figure 4, which cause atypical structure changes (including sharp shadow-edges) when they are switched on and off.

5.2. CAVIAR Sequences

The second experiment uses the CAVIAR Shopping Center clips[1] [1]. The purpose of this experiment was to measure any degradation in performance due to the use of an illumination-invariant method on a non-varying dataset. We might expect that the built-in invariance would sacrifice recall compared to a "standard" method in this situation. Comparative results are shown in Figures 5(a) and 5(b). Each point on the plots encodes the (Precision,Recall) of one clip.

In fact, the mean performance of Normalised Gradient-Correlation is better in terms of both precision and recall; however, the distribution of results is different for the two views tested. The CAVIAR dataset consists of two views ("front" and "corridor") of the same scene. The "front" view has a wide angle, but looks deep inside a clothes shop, in which targets may be small or partially visible in the complex background. The Normalised Gradient-Correlation algorithm can miss these targets, because of the presence of significant unchanged background structure, acting to raise the correlation value—particularly in the larger 16x16 and 32x32 regions of integration. This leads to a tendency toward decreased recall in the "front" clips.

Ironically, Mixture-of-Gaussians seems to suffer because of its background update. The scenes all have static backgrounds, with negligible variation, and so the update mechanism can only cause the algorithm to miss targets, as their statistics are learned by the background model. This occurs more often in the "corridor" sequences, because the targets are often moving toward or away from the camera and so their positions in the image plane change very slowly. The effect is to reduce the recall scores for these clips. Additional experiments were performed using a limited version of the Stauffer and Grimson algorithm, in which the mixture-model was restricted to a single Gaussian; these confirmed that recall is increased by preventing the model from learning new backgrounds.

Being based on pixel-wise measurements, Mixture-of-Gaussians also exhibits "salt & pepper" noise—not all of which can be removed by the morphological filtering proposed in [15] and which acts to reduce precision. By contrast, the false-positives of the proposed robust algorithm tend to occur as leakage around correctly detected objects, due to the use of neighbourhood measurements.

[1] From `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`

6. Conclusions

The proposed robust change-detection algorithm makes it possible to accurately segment changes in a scene, while discounting the effects of changing illumination (and related variations, such as camera adaptive gain-control). It also addresses other drawbacks of existing techniques:

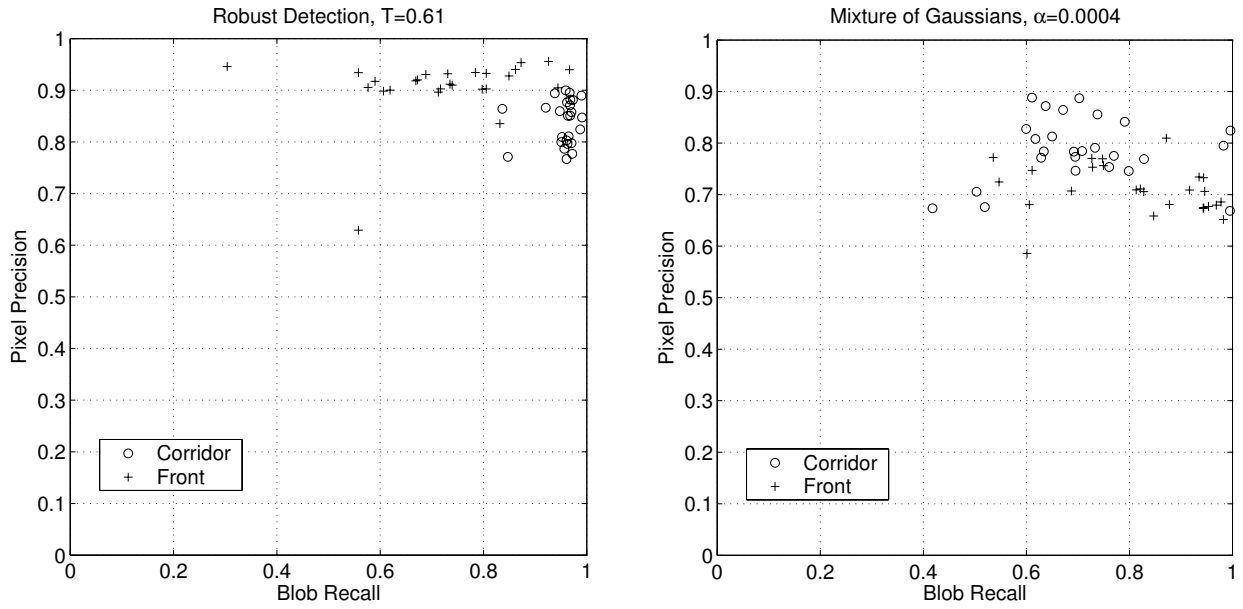
- The so-called "aperture problem", encountered by other local-neighbourhood methods, in which the interior of a moving region of uniform intensity cannot be detected over a uniform background, because the difference can be modelled by a local illumination change
- The need to maintain complex background models
- The dependence on the background update mechanism to absorb the effects of illumination variation—in practice, the methods based on background subtraction distinguish between object-motion and other changes based on their temporal dynamics. Fast illumination changes and slowly moving objects therefore cause errors. (It is important to note that this does not deny the usefulness of background model update—which is beneficial for adapting to gradual variations, outside the scope of the illumination-model)

The algorithm achieves robustness through a series of computationally simple processes, in order to maximise efficiency as well as accuracy. Several real-time implementations have been developed, including a C program on a 300MHz Renesas M32R embedded processor. Although not properly optimised for this architecture, the program runs at a frame rate of 2.5 frames per second, processing images at 160x120 pixels (including image capture and result-display routines). Real-time software tested on a PC-platform (Pentium 4, 3GHz) at 320x240 and 640x480 pixels is currently limited by camera frame rate.

Note that, although not tested here, we may expect the algorithm's performance to degrade in the presence of dynamic backgrounds (foliage waving in wind; flashing neon signs *etc*.), which will cause "genuine" structure changes. Methods based on probabilistic models of intensity/colour [15, 7, 9] are designed to learn the multiple modes of such backgrounds at each pixel and should therefore be robust to such continual variation. However, all such methods will behave similarly to the Mixture-of-Gaussians technique tested here, in the face of sudden brightness or contrast change: pixel values shift (instantaneously) outside the learned, historical modes and false detection occurs, until the model can learn the new distribution.

References

[1] The EC-funded CAVIAR project/IST 2001 37540.



(a) Normalised Gradient-Correlation

(b) Mixture-of-Gaussians [15]

Figure 5. Results on the CAVIAR dataset

[2] T. Aoki, O. Nakayama, M. Shiohara, and Y. Murakami. US patent application 2001/0004400: Method and apparatus for detecting moving object, 2000.

[3] R. G. Bassman, B. B. Bhatt, B. J. Call, M. W. Hansen, S. C. Hsu, G. S. van der Wal, and L. E. Wixson. US patent 6,044,166: Parallel-pipelined image processing system, 1996.

[4] S.-C. S. Cheung and C. Kamath. Robust techniques for background subtraction in urban traffic video. In *Proc. IS&T/SPIE Symposium on Electronic Imaging*, Jan. 2004.

[5] F. C. Crow. Summed-area tables for texture mapping. *SIGGRAPH Computer Graphics*, 18(3):207–212, 1984.

[6] E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra, Oct. 2001.

[7] A. M. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. In *Proc. 6th European Conference on Computer Vision*, volume II, London, UK, 2000. Springer-Verlag.

[8] ETISEO Project. Introduction to evaluation and metrics. Version 1.04, Mar. 2005.

[9] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Real-time surveillance of people and their activities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):809–830, 2000.

[10] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 5–6, Dec. 2002.

[11] T. Kondo, N. Fujiwara, J. Ishibashi, T. Sawao, T. Nagano, T. Miyake, and S. Wada. US patent application 2005/0259870: Image processing apparatus and method, and image pickup apparatus, 2002.

[12] L. Li and M. K. H. Leung. Integrating intensity and texture differences for robust change detection. *IEEE Trans. Image Processing*, 11(2), Feb. 2002.

[13] T. Matsuyama, T. Ohya, and H. Habe. Background subtraction for non-stationary scenes. In *Proc. 4th Asian Conference on Computer Vision*, pages 662–667, 2000.

[14] S. Nagaya, T. Miyatake, and T. Fujita. US patent application 2002/0030739: Moving object detection apparatus, 2001.

[15] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 246–252, 1999.

[16] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Dec. 2001.

[17] Y. Weiss. Deriving intrinsic images from image sequences. In *Proc. International Conference on Computer Vision*, volume 2, pages 68–75, 2001.