Adaptive Patch-Based Background Modelling for Improved Foreground Object Segmentation and Tracking

Vikas Reddy, Conrad Sanderson, Andres Sanin, Brian C. Lovell
NICTA, PO Box 6020, St Lucia, QLD 4067, Australia
The University of Queensland, School of ITEE, QLD 4072, Australia

Abstract

A robust foreground object segmentation technique is proposed, capable of dealing with image sequences containing noise, illumination variations and dynamic backgrounds. The method employs contextual spatial information by analysing each image on an overlapping patch-by-patch basis and obtaining a low-dimensional texture descriptor for each patch. Each descriptor is passed through an adaptive multi-stage classifier, comprised of a likelihood evaluation, an illumination robust measure, and a temporal correlation check. A probabilistic foreground mask generation approach integrates the classification decisions by exploiting the overlapping of patches, ensuring smooth contours of the foreground objects as well as effectively minimising the number of errors. The parameter settings are robust against wide variety of sequences and post-processing of foreground masks is not required. Experiments on the difficult Wallflower and I2R datasets show that the proposed method obtains considerably better results (both qualitatively and quantitatively) than methods based on Gaussian mixture models, feature histograms, and normalised vector distances. Further experiments on the CAVIAR dataset (using several tracking algorithms) indicate that the proposed method leads to considerable improvements in object tracking accuracy.

1. Introduction

Segmentation of objects of interest from an image sequence is a primary and critical task in most intelligent surveillance applications such as object identification, tracking and analysis. Typical approaches for segmentation of foreground objects from image sequences employ the idea of comparing each frame against a model of the background, followed by selecting the outliers (*i.e*. pixels or areas that do not fit the model). In general, the pixels are selected in one of two ways: **(i)** pixel-by-pixel, where an independent decision is made for each pixel, possibly taking into account information from neighbouring pixels; **(ii)** region-based selection, where a decision is made on an entire group of spatially close pixels.

The majority of the algorithms described in the literature belong to the first category. Wren *et al*. modelled each pixel using a single Gaussian whose parameters were updated recursively [18]. To accommodate multi-modal characteristics of the background, Stauffer and Grimson proposed to use Gaussian mixture modelling [15]. Since then numerous variants and improvements over this method have been proposed. For example, Zivkovic's method can adaptively change the number of Gaussians per pixel [19], and Lee proposed a learning procedure that improves the segmentation accuracy and model convergence rate [7].

Prediction filters were employed to adapt to the changes in the background. Ridder *et al*. used Kalman filtering [13], and Toyama *et al*. applied a Wiener filter [17]. Non-parametric approaches via kernel density estimation were also proposed. For example, Elgammal *et al*. used a Gaussian kernel [3] while Tanaka *et al*. proposed a fast approach using Parzen windows [16]. Kim *et al*. modelled each background pixel by a set of code words [6], Li *et al*. constructed a histogram of features per pixel [8], and Heikkila *et al*. modelled each pixel using local binary pattern histograms [5].

As most of these algorithms do not analyse the contextual spatial information effectively, they are sensitive to varying illumination, cast shadows, dynamic backgrounds and inherent image noise. They also often require ad hoc post-processing (*e.g*. morphological operations) to remove incorrectly classified and scattered pixels from the foreground mask.

In the region-based category, each frame is typically split into blocks (or patches) and the classification is made at the block-level. The usage of contextual spatial information mitigates, to a certain extent, the influence of above mentioned problems on the segmentation. Differences between blocks from a frame and the background can be measured by, for example, edge histograms [10] and normalised vector distances [11]. Both of the above methods handle the problem of varying illumination but do not address dynamic backgrounds. Furthermore, as adjacently located blocks are used, the generated foreground masks exhibit "blockiness" artefacts, *i.e*. rough foreground object contours (see Fig. 2(e) for an example).

In this paper we propose a robust algorithm that has qualities of both the pixel-based and region-based categories. It is capable of dealing with image noise, illumination variations and dynamic backgrounds (often witnessed in sequences captured in outdoor environments), while obtaining smooth object contours. Specifically, each frame is analysed on an overlapping block-by-block basis, with a low-dimensional texture descriptor obtained for each block. Each descriptor is passed through an adaptive multi-stage classifier, where each stage analyses the descriptor from different perspectives before classifying it as belonging to the foreground.

Unlike conventional methods where a pixel is classified as foreground/background based on its statistics collected over time, our approach classifies a pixel based on how many overlapping blocks containing that particular pixel have been classified as foreground/background, eliminating the need to do any post-processing.

We continue as follows. In Section 2 the proposed algorithm is described in detail. Performance evaluation and comparison with three other algorithms is given in Section 3, followed by the main findings in Section 4.

This paper is an extended and revised version of our earlier work [12], with the differences being in the algorithm as well as in the evaluation. The algorithm has an added model reinitlisation step and the foreground mask generation step is now probabilistic instead of using hard decisions. The evaluation uses two additional datasets (Wallflower and CAVIAR) and contains further experiments showing that the algorithm leads to increased tracking performance.

2. Proposed Segmentation Technique

The proposed technique has four main components:

1. Division of a given image into overlapping blocks (patches), followed by generating a low-dimensional descriptor for each block.

2. Classification of each block into foreground or background, where each block is sequentially processed by up to three classifiers. As soon as one of the classifiers deems that the block is part of the background, the remaining classifiers are not consulted. In sequential order of processing, the three classifiers are:

   (a) a probability measurement using a location specific Gaussian model of the background;

   (b) an illumination robust similarity measurement through a cosine distance metric;

   (c) a temporal correlation check, where blocks & decisions from the previous image are considered.

3. Model reinitialisation to address scenarios where a sudden and significant scene change can make the current model inaccurate.

4. Probabilistic generation of the foreground mask, where the classification decisions for all blocks are integrated. The overlapping nature of the analysis is utilised to minimise the number of errors (both false positives and false negatives) and produce smooth contours.

Each of the components is explained in more detail in the following sections.

2.1. Blocking and Generation of Descriptors

Each image is split into blocks with a size of 8×8 pixels, with each block overlapping its neighbours by 7 pixels (*i.e.* 87.5%) in both the horizontal and vertical directions. 2D Discrete Cosine Transform (DCT) decomposition is employed to obtain a relatively robust and compact description of each block [4, 14]. Image noise and minor variations are effectively ignored by keeping only several low-order DCT coefficients, which reflect the average intensity and low frequency information. Specifically, for a block located at (i, j), four coefficients per colour channel are retained (based on preliminary experiments), leading to a 12 dimensional descriptor:

$$\boldsymbol{d}_{(i,j)} = \left[c_0^{[r]}, \cdots, c_3^{[r]},\ \ c_0^{[g]}, \cdots, c_3^{[g]},\ \ c_0^{[b]}, \cdots, c_3^{[b]}\right]^T \quad (1)$$

where $c_n^{[k]}$ denotes the n-th DCT coefficient from the k-th colour channel, with $k \in \{r, g, b\}$.

2.2. Multi-Stage Block Classifier

Each block's descriptor is analysed sequentially by three classifiers (listed as (a), (b) and (c), below), with each classifier using location specific parameters. A block is deemed to belong to the background as soon as its descriptor is classified as such by any of the three classifiers.

The first classifier handles dynamic backgrounds (such as waving trees, water surfaces and fountains), but fails when illumination variations exist. The second classifier analyses if the anomalies in the descriptor are due to illumination variations. The third classifier exploits temporal correlations (that naturally exists in image sequences) to partially handle changes in environment conditions.

(a) The first classifier employs a multivariate Gaussian model for each of the background blocks. The likelihood of descriptor $\boldsymbol{d}_{(i,j)}$ belonging to the background class is found via:

$$p\left(\boldsymbol{d}_{(i,j)}\right) = \frac{\exp\left\{-\frac{1}{2}\left[\boldsymbol{d}_{(i,j)} - \boldsymbol{\mu}_{(i,j)}\right]^T \boldsymbol{\Sigma}_{(i,j)}^{-1}\left[\boldsymbol{d}_{(i,j)} - \boldsymbol{\mu}_{(i,j)}\right]\right\}}{(2\pi)^{\frac{12}{2}}\left|\boldsymbol{\Sigma}_{(i,j)}\right|^{\frac{1}{2}}} \quad (2)$$

where $\boldsymbol{\mu}_{(i,j)}$ and $\boldsymbol{\Sigma}_{(i,j)}$ are the mean vector and covariance matrix for location (i,j), respectively. If $p(\boldsymbol{d}_{(i,j)}) \geq T_1$ (where T_1 is an empirically determined threshold), the corresponding block is classified as background. If a block has been classified as background, the corresponding Gaussian model is updated using the adaptation technique proposed by Wren *et al*. [18].

To obtain the initial parameters, we train the background models using the first few seconds of the sequence. To allow the training sequence to contain moving foreground objects, a robust estimation strategy is employed instead of directly obtaining the parameters.

Specifically, for each block location a two-component Gaussian mixture model is trained, followed by taking the absolute difference of the weights of the two Gaussians. If the difference is greater than 0.5 (based on prelim. experiments), we retain the Gaussian with the dominant weight. The reasoning is that the less prominent Gaussian is modelling moving foreground objects and/or other outliers. If the difference is less than 0.5, we assume that no foreground objects are present and use all available data for that particular block location to estimate the parameters of the single Gaussian.

(b) If block (i,j) has not been classified as part of the background, the second classifier employs a cosine distance metric:

$$\text{cosdist}(\boldsymbol{d}_{(i,j)}, \boldsymbol{\mu}_{(i,j)}) = 1 - \frac{\boldsymbol{d}_{(i,j)}^T \, \boldsymbol{\mu}_{(i,j)}}{\|\boldsymbol{d}_{(i,j)}\| \, \|\boldsymbol{\mu}_{(i,j)}\|} \qquad (3)$$

where $\boldsymbol{\mu}_{(i,j)}$ is from Eqn. (2). Block (i,j) is deemed as background if $\text{cosdist}(\boldsymbol{d}_{(i,j)}, \boldsymbol{\mu}_{(i,j)}) \leq T_2$.

Empirical observations suggest the angles subtended by descriptors obtained from a block exposed to varying illumination are almost the same. A similar phenomenon was also observed in RGB colour space [6].

(c) For each block, the third classifier takes into account the current descriptor as well as the corresponding descriptor from the previous image, denoted as $\boldsymbol{d}_{(i,j)}^{[\text{prev}]}$. Block (i,j) is labelled as part of the background if the following two conditions are satisfied:

(i) $\boldsymbol{d}_{(i,j)}^{[\text{prev}]}$ was classified as background;

(ii) $\text{cosdist}(\boldsymbol{d}_{(i,j)}^{[\text{prev}]}, \boldsymbol{d}_{(i,j)}) \leq T_3$.

Condition (i) ensures the cosine distance measured in (ii) is not with respect to a descriptor classified as foreground. As the sample points are consecutive in time and should be almost identical if $\boldsymbol{d}_{(i,j)}$ belongs to background, we use $T_3 = 0.5 \times T_2$.

Thresholds T_1 and T_2 are deliberately tuned such that they result in slightly more false positives than false negatives. This ensures a low probability of misclassifying foreground objects as background. The surplus false positives are eliminated during the creation of the foreground mask (Section 2.4).

2.3. Model Reinitialisation

A scene change might be too quick and/or too severe for the adaptation and classification strategies used above (*e.g*. severe illumination change due to lights being switched on in a dark room). As such, the existing background model can wrongly detect a very large portion of the image as foreground.

Model reinitialisation is triggered if more than 70% of each image is consistently classified as foreground for a period of $\frac{1}{2}$ second. The corresponding images are accumulated and are used to rebuild the statistics of the new scene. Due to the small amount of retraining data, the covariance matrices are kept as is, while the new means are obtained as per the estimation method described in Section 2.2(a).

2.4. Probabilistic Foreground Mask Generation

In typical block based classification methods, misclassification is inevitable whenever a given block has foreground and background pixels (examples are illustrated in Fig. 1). We exploit the overlapping nature of the block-based analysis to alleviate this inherent problem. Each pixel is classified as foreground only if a significant proportion of the blocks that contain that pixel are classified as foreground.

Formally, let the pixel located at (x,y) in image I be denoted as $I_{(x,y)}$. Furthermore, let $B_{(x,y)}^{\text{fg}}$ be the number of blocks containing pixel (x,y) that were classified as foreground (fg), and $B_{(x,y)}^{\text{total}}$ be the total number of blocks containing pixel (x,y). We define the probability of foreground being present in $I_{(x,y)}$ as:

$$P\left(\text{fg} \mid I_{(x,y)}\right) = B_{(x,y)}^{\text{fg}} \;/\; B_{(x,y)}^{\text{total}} \qquad (4)$$

If $P\left(\text{fg} \mid I_{(x,y)}\right) \geq 0.90$, pixel $I_{(x,y)}$ is labelled as part of the foreground.

In other words, a pixel that was misclassified a few times prior to mask generation can be classified correctly in the generated foreground mask. This decision strategy, similar to majority voting, effectively minimises the number of errors in the output.

This approach is in contrast to conventional methods, such as those based on Gaussian mixture models [15], kernel density estimation [3] and codebook models [6], which do not have this built-in "self-correcting" mechanism. These methods can be prone to errors since their classification is based on a single decision and the models are built solely using temporal pixel statistics.

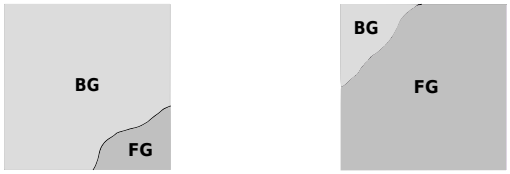


Figure 1. Misclassification is inevitable at the pixel level whenever a given block has both foreground (FG) and background (BG) pixels. Classifying Block A as background results in a few false negatives (foreground pixels classified as background) while classifying Block B as foreground results in a few false positives (background pixels classified as foreground).

3. Experiments

The proposed algorithm was compared with segmentation methods based on Gaussian mixture models (GMMs) [15], feature histograms [8], and normalised vector distances (NVD) [11]. We used the OpenCV v2.0 [2] implementations for the first two algorithms with default parameters. The first two methods classify individual pixels into foreground or background, while the last method makes decisions on groups of pixels.

For evaluations of the methods, we conducted two sets of experiments: **(i)** subjective and objective evaluation of foreground segmentation efficacy, using datasets with available ground-truths; **(ii)** comparison of the effect of the different foreground segmentation methods on tracking performance. The details of the experiments[1] are described in Sections 3.1 and 3.2, respectively.

3.1. Evaluation by Ground-Truth Similarity

For standalone evaluation of the methods, we used the I2R and Wallflower datasets. The I2R dataset[2] has sequences captured in diverse and challenging environments. It contains nine sequences, and for each sequence there are 20 randomly selected images for which the ground-truth foreground masks are available. The Wallflower dataset [3] has seven sequences, with each sequence being a representative of a distinct problem encountered in background modelling (see [17] for details). Each sequence has only one ground-truth foreground mask.

In our experiments the same parameter settings were used across all sequences (*i.e*. they were not optimised for any particular sequence). Post-processing using morphological operations was required for the foreground masks obtained by the GMM and feature histogram methods, in order to clean up the scattered error pixels. For the GMM method, opening followed by closing using a 3×3 kernel was performed, while for the feature histogram method we enabled the built-in post-processor (using default parameters in the OpenCV implementation). We note that the proposed method does not require any such ad hoc post-processing. We present both qualitative and quantitative analysis of the results.

For quantitative evaluation we adopted the similarity measure used by Maddalena and Petrosino [9], which quantifies how similar the obtained foreground mask is to the ground-truth. The measure is defined as:

$$similarity = \frac{tp}{tp + fp + fn} \tag{5}$$

where $similarity \in [0, 1]$, while tp, fp and fn are total number of true positives, false positives and false negatives (in terms of pixels), respectively. The higher the $similarity$ value, the better the segmentation result.

Figs. 2 and 3 show qualitative results on three sequences from the I2R and Wallflower datasets, respectively.

In Fig. 2, the AP sequence (top row) has significant cast shadows of people moving at an airport. The FT sequence (middle row) contains people moving against a background of fountain with varying illumination. The MR sequence (bottom row) shows a person entering and leaving a room where the window blinds are non-stationary and there are significant illumination variations caused by the automatic gain control of the camera.

In Fig. 3, the *time of day* sequence (top row) has a gradual increase in the room's illumination intensity over time. A person walks in and sits on the couch. The *waving trees* sequence (middle row) has a person walking against a background consisting of the sky and strongly waving trees. In the *camouflage* sequence (bottom row), a monitor has a blue screen with rolling bars. A person in a blue coloured clothing walks in and occludes the monitor.

We note that output of the GMM based method (column **c** in Figs. 2 and 3) is sensitive to reflections, illumination changes and cast shadows. While the histogram based method (column **d**) overcomes these limitations, it has a lot of false negatives. The NVD based method (column **e**) is largely robust to illumination changes, but fails to handle dynamic backgrounds and produces "blocky" foreground masks. The results obtained by the proposed method (column **f**) are qualitatively better than those obtained by the other three methods, having low false positives and false negatives. However, we note that due the the block-based nature of the analysis, objects very close to each other tend to merge.

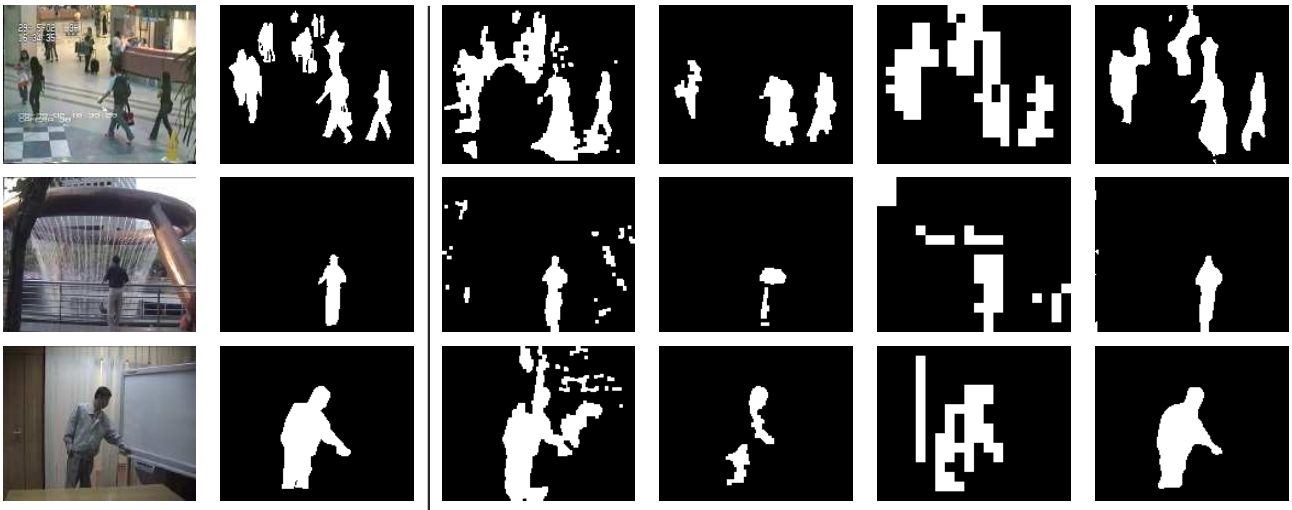
The quantitative results (using the similarity metric) obtained on the I2R and Wallflower datasets, shown in Figs. 4 and 5, respectively, largely confirm the visual results[4].

[1] The experiments were performed with the aid of the Armadillo C++ linear algebra library, available from http://arma.sourceforge.net

[2] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

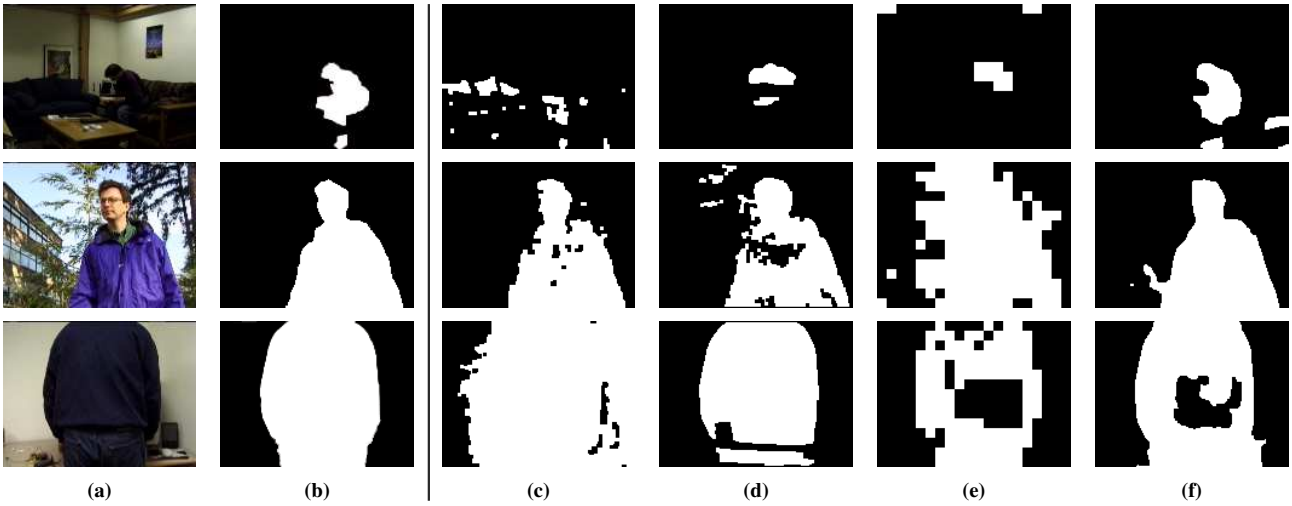
[3] http://research.microsoft.com/en-us/um/people/jckrumm/WallFlower/TestImages.htm

[4] The similarity value of *moved object* sequence from the Wallflower dataset is zero for all algorithms and is therefore not shown in Fig. 5. This is due to the absence of true positives in its ground-truth.



(a) (b) (c) (d) (e) (f)

Figure 2. **(a)** Example frames from 3 video sequences from the I2R dataset. *Top:* people walking at an airport, with significant cast shadows. *Middle:* people moving against a background of fountain with varying illumination. *Bottom:* a person walks in and out of a room where the window blinds are non-stationary, with illumination variations caused by automatic gain control of the camera. **(b)** Ground-truth foreground mask, and foreground mask estimation using: **(c)** GMM based [15] with morphological post-processing, **(d)** feature histograms [8], **(e)** NVD [11], **(f)** proposed method.



(a) (b) (c) (d) (e) (f)

Figure 3. As per Fig. 2, but using the Wallflower dataset. *Top:* room illumination gradually increases over time and a person walks in and sits on the couch. *Middle:* person walking against a background of strongly waving trees and the sky. *Bottom:* a monitor displaying a blue screen with rolling bars is occluded by a person wearing blue coloured clothing.

On the I2R dataset the proposed method consistently outperforms the other methods. The next best method (GMM with morphological post-processing) obtained an average *similarity* value of 0.468, while the proposed method achieved 0.689, representing an improvement of about 47%.

On the Wallflower dataset the proposed method achieved considerably better results for the *foreground aperture* and *time of day* sequences. While for the remainder of the sequences the performance was roughly on par with the other methods, the proposed method nevertheless still achieved the highest average *similarity* value. The next best method (histogram of features) obtained an average value of 0.525, while the proposed method obtained 0.653, representing an improvement of about 24%.

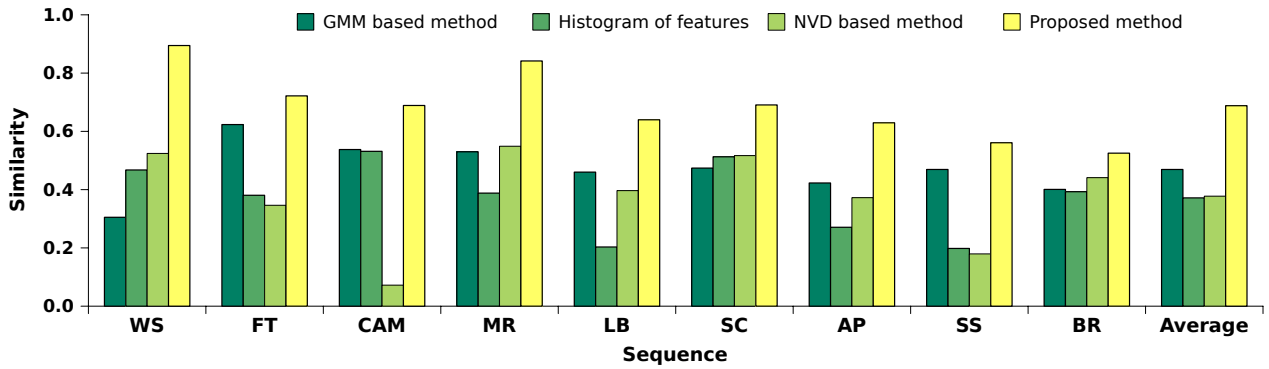


Figure 4. Comparison of *similarity* values (defined Eqn. 5) obtained on the I2R dataset using foreground segmentation methods based on GMMs [15], feature histograms [8], NVD [11] and the proposed method. The higher the *similarity* (*i.e.* agreement with ground-truth), the better the segmentation result.

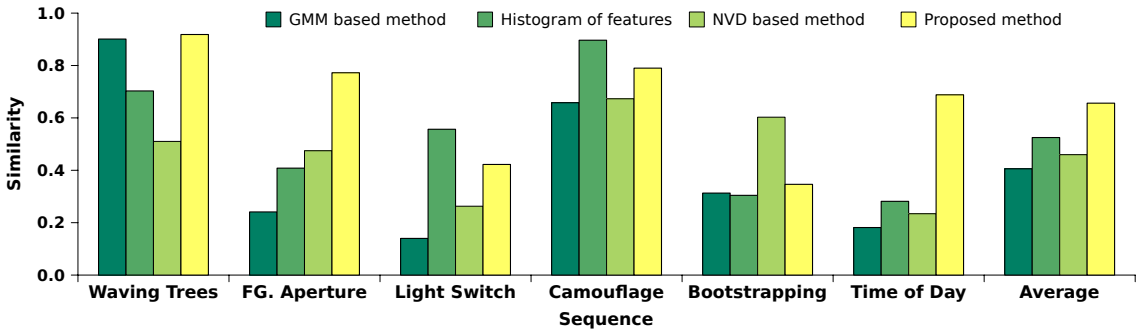


Figure 5. As per Fig. 4, but obtained on the Wallflower dataset.

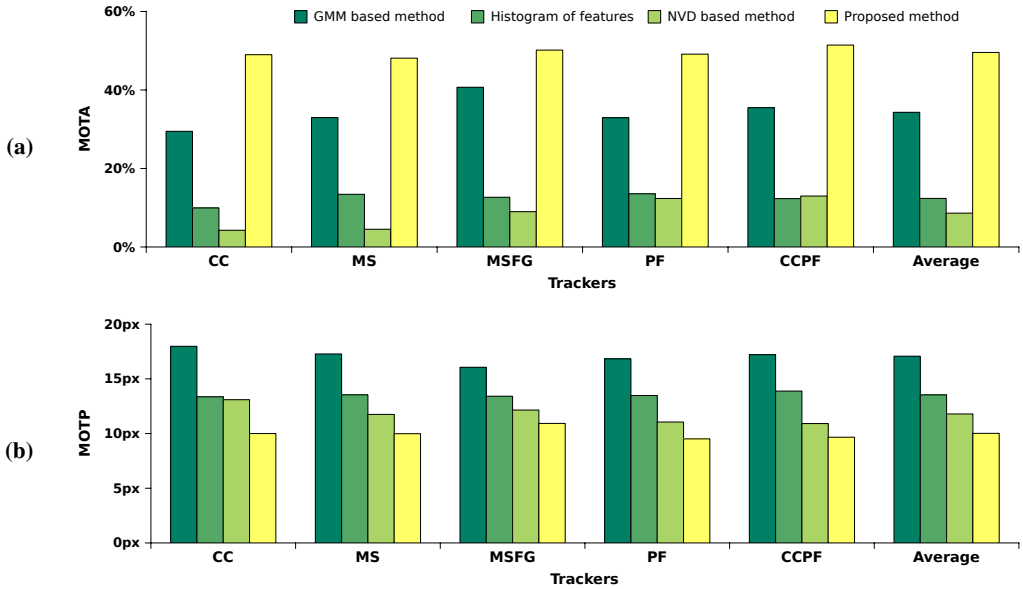


Figure 6. Effect of foreground detection methods on: **(a)** multiple object tracking accuracy (MOTA), where taller bars indicate better accuracy; **(b)** multiple object tracking precision (MOTP), where shorter bars indicate better precision (lower distance). Results are grouped by tracking algorithm: blob matching (CC), two mean shift trackers (MS and MSFG), particle filter (PF) and hybrid tracking (CCPF).

Block Size	Average *similarity*		
	I2R	Wallflower	mean
2×2	0.604	0.437	0.520
4×4	0.699	0.539	0.619
6×6	**0.702**	0.594	0.648
8×8	0.690	0.620	**0.655**
10×10	0.667	0.638	0.653
12×12	0.634	0.646	0.640
14×14	0.602	**0.658**	0.630
16×16	0.556	0.646	0.601

Table 1. Average *similarity* values obtained using various block sizes on the I2R and Wallflower datasets. The "mean" column indicates the mean of the values obtained for I2R and Wallflower.

The performance of the proposed algorithm for block sizes ranging from 2×2 to 16×16 is shown in Table 1. The optimal block size for the I2R dataset is 6×6, with the performance being quite stable from 4×4 to 8×8. For the Wallflower dataset the optimal size is 14×14, with similar performance obtained from 12×12 to 16×16. By taking the mean of the values obtained for each block size across both datasets, the overall optimal size appears to be 8×8, as used in the preceding experiments.

For other datasets, we expect the optimal block size to be sensitive to parameters such as frame resolution, field of view and size of foreground objects. Preliminary experiments with various real-life surveillance videos suggest that for frame resolutions around 352×288 (CIF), block size of 8×8 appears to be well suited. For significantly larger resolutions, block size of 16×16 works well.

3.2. Evaluation by Tracking Precision & Accuracy

We conducted a second set of experiments to evaluate the performance of the segmentation methods in more pragmatic terms rather than limiting ourselves to the traditional ground-truth evaluation approach. To this effect, we studied the influence of the different foreground detection algorithms on tracking performance. The foreground masks obtained from the detectors were passed as input to several tracking systems. We used the tracking systems implemented in the video surveillance module of OpenCV v2.0 [2] and the tracking ground truth data that is available for the 50 sequences in the second set of the CAVIAR[5] dataset. The tracking performance was measured with the two metrics proposed by Bernardin and Stiefelhagen [1], namely multiple object tracking accuracy (MOTA) and multiple object tracking precision (MOTP).

Briefly, MOTP measures the average pixel distance between the ground-truth locations of objects and their locations according to a tracking algorithm. The lower the MOTP, the better. MOTA accounts for object configuration errors, false positives, misses as well as mismatches. The higher the MOTA, the better.

We performed 20 tracking simulations by evaluating four foreground object segmentation algorithms (GMM based, histogram of features, NVD and the proposed method) in combination with five tracking algorithms (blob matching, mean shift, mean shift with foreground feedback, particle filter, and blob matching with particle filter for occlusion handling). The performance result in each simulation is the average performance of the 50 test sequences.

The quantitative results, presented in Fig. 6, indicate that in all cases the proposed method led to the best precision and accuracy values. For tracking precision (MOTP), the next best method (NVD based) obtained an average pixel distance of 11.79, while the proposed method reduced the distance to 10, indicating an improvement of approximately 15%. For tracking accuracy (MOTA), the next best method (GMM based) obtained an average accuracy value of 0.343, while the proposed method achieved 0.495, representing a noteworthy improvement of about 44%.

4. Main Findings

In this paper we have proposed a new foreground object segmentation method that is robust to image sequences containing noise, illumination variations and dynamic backgrounds. The model initialisation strategy allows the training sequence to contain moving foreground objects.

Contextual spatial information is employed through analysing each frame on an overlapping block-by-block basis. The low-dimensional texture descriptor for each block alleviates the effect of image noise. The adaptive multi-stage classifier analyses the descriptor from different perspectives before classifying it as foreground. Specifically, it checks if the disparity is due to background motion or change in illumination. The temporal correlation check minimises the occasional false positives emanating due to background characteristics which were not handled by the preceding stages.

A probabilistic foreground mask generation approach integrates the block-level classification decisions by exploiting the overlapping nature of the analysis, ensuring smooth contours of the foreground objects as well as effectively minimising the number of errors.

Experiments conducted to evaluate the standalone performance (using the difficult Wallflower and I2R datasets) and the effect on tracking performance (using the CAVIAR dataset) show the proposed method obtains considerably better results (both qualitatively and quantitatively) than methods based on Gaussian mixture models, feature histograms and normalised vector distances. The parameter settings appear to be quite robust against wide variety of sequences and the method does not require explicit post-processing of the foreground masks.

[5] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

Acknowledgements

NICTA is funded by the Australian Government as represented by the *Department of Broadband, Communications and the Digital Economy*, as well as the Australian Research Council through the *ICT Centre of Excellence* program.

Support for this work was also provided by the Australian Government's National Security Science and Technology Branch within the Department of the Prime Minister and Cabinet. This support does not represent and endorsement of the contents or conclusions.

The experiments were implemented and evaluated using open-source software, including Linux-based operating systems. The authors would like to extend their thanks to all open-source developers for their efforts.

References

[1] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. *EURASIP Journal on Image Video Processing*, 2008.

[2] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008.

[3] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proc. European Conf. Computer Vision (ECCV)*, pages 751–767, 2000.

[4] R. Gonzales and R. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition, 2007.

[5] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(4):657–662, 2006.

[6] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground–background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, 2005.

[7] D. Lee. Effective gaussian mixture learning for video background subtraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 827–832, 2005.

[8] L. Li, W. Huang, I. Gu, and Q. Tian. Foreground object detection from videos containing complex background. In *Proc. ACM Int. Conf. Multimedia*, pages 2–10, 2003.

[9] L. Maddalena and A. Petrosino. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. *IEEE Trans. Image Processing*, 17:1168–1177, 2008.

[10] M. Mason and Z. Duric. Using histograms to detect and track objects in color video. In *Proc. Applied Imagery Pattern Recognition Workshop*, pages 154–159, 2001.

[11] T. Matsuyama, T. Wada, H. Habe, and K. Tanahashi. Background subtraction under varying illumination. *Systems and Computers in Japan*, 37(4):77, 2006.

[12] V. Reddy, C. Sanderson, and B. C. Lovell. Robust foreground object segmentation via adaptive region-based background modelling. In *International Conference on Pattern Recognition (ICPR)*, Turkey, 2010.

[13] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. In *Proc. Int. Conf. Recent Advances in Mechatronics*, pages 193–199, 1995.

[14] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *International Conference on Biometrics, Lecture Notes in Computer Science (LNCS)*, volume 5558, pages 199–208, 2009.

[15] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 246–252, 1999.

[16] T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. A fast algorithm for adaptive background model construction using parzen density estimation. In *Proc. Advanced Video and Signal Based Surveillance (AVSS)*, pages 528–533, 2007.

[17] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. Int. Conf. Computer Vision (ICCV)*, volume 1, pages 255–261, 1999.

[18] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[19] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In *Proc. Int. Conf. Pattern Recognition (ICPR)*, volume 2, pages 28–31, 2004.