Evaluation of Background Subtraction Techniques for Video Surveillance

Sebastian Brutzer, Benjamin Höferlin, Gunther Heidemann
Intelligent Systems Group, Universität Stuttgart, Germany
{hoeferlin,ais}@vis.uni-stuttgart.de

Abstract

Background subtraction is one of the key techniques for automatic video analysis, especially in the domain of video surveillance. Although its importance, evaluations of recent background subtraction methods with respect to the challenges of video surveillance suffer from various shortcomings. To address this issue, we first identify the main challenges of background subtraction in the field of video surveillance. We then compare the performance of nine background subtraction methods with post-processing according to their ability to meet those challenges. Therefore, we introduce a new evaluation data set with accurate ground truth annotations and shadow masks. This enables us to provide precise in-depth evaluation of the strengths and drawbacks of background subtraction methods.

1. Introduction

A common method for search-space reduction and focus of attention modelling in video analysis is *background subtraction* (BS). The term BS covers a set of methods that aim to distinguish between foreground and background areas in video sequences utilizing a *background model*.

Over the past years, various BS methods have been developed (cf. surveys [24, 9, 4, 2, 3]), with each of them having its own characteristic, strength and weakness. Evaluation allows to identify those characteristics and helps to focus on the remaining problems. Although its importance, literature lacks of comprehensive evaluation of recent BS methods. One reason might be the huge effort involved in generating qualitatively high *ground truth* (GT) data of natural video sequences. Hence, some evaluations only use a few labeled frames or judge the performance at object-level, which is considerably easier. However, evaluation at pixel-level provides more insight into strengths and weaknesses. There exist several techniques to overcome manual GT annotation. Aside labor-intensive and highly subjective judgement of segmentation results by human experts, various approaches have been developed that do not depend on GT data [5, 10] or automatically generate them [12]. Unfortunately, such methods are not applicable to evaluate BS performance with respect to the challenges arising in video surveillance (cf. Section 3). To address the problem of GT data acquisition, we propose the use of artificial data. In order to cope with the problem "that the synthetic data will probably not faithfully represent the full range of real data" [9], we use a typical video surveillance scenario, high quality 3D-models and modern raytracing technology with global illumination to account for realistic image synthesis. Hence, we are able to generate high quality pixel-level GT data and evaluate the challenges separately from the others. Based on the introduced data set, we evaluate the performance of nine BS methods with post-processing. This includes a couple of *multi-modal* methods, able to cope with dynamic background. Such methods were barely compared in existing evaluations. Further, we are able to address challenges which have little cover in recent evaluation literature, such as shadows and noise. Figure 2 illustrates the typical structure of BS which serves as foundation for subsequent processing steps in many video surveillance applications (e.g., [14]). As illustrated, the main task of BS is to compare an input frame against a background model. Note that we neglect any pre-processing of the input such as image registration, color conversion, etc.. The model describes the background areas of the scene and is often represented by distribution of features, such as color information. The process of *foreground detection* determines which areas of an image belong to the foreground class with respect to the similarity of an input frame and the background model. The result of this classification is a binary *foreground mask*.

Figure 1. Exemplary frames of the artificial data set.

Since BS algorithms have to cope with several chal-

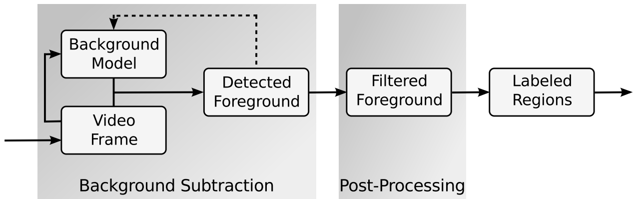


Figure 2. Typical process of background subtraction with post-processing in surveillance applications.

lenges in the context of video surveillance, extensions to naive BS were introduced, such as *post-processing* of the foreground mask or adaption to changes of the scenery (e.g., gradual illumination changes). Further processing steps often include *region labeling* combined with *region thresholding* or the feedback of tracking assumptions. This work focusses on BS methods and *post-processing* techniques that result in binary foreground masks. Hence, further processing steps are neglected in our evaluation.

2. Related Work

Over the past years, several papers have been published addressing the evaluation of BS methods. However, evaluation of recent background subtraction methods with respect to the problems occurring in video surveillance are missing, out-dated, or of low quality.

Toyama *et al.* [27] identified several challenges BS algorithms typically have to cope with. They generated a set of specific videos which cover some of the challenges they found, others like shadows, noise or compression were neglected. Their evaluation data set is well known in literature. However, the GT is only available for a single frame per sequence and was manually labeled.

In 2004, Cheung and Kamath [6] evaluated the performance of six BS methods on urban traffic videos under different weather conditions. Although, many BS methods exploit color information to constrain effects of camouflage, their data set consists of gray-scale images only.

In the work of Karaman *et al.* [15], nine BS approaches (spreading from 1982 to 2004) were compared with each other by means of segmentation performance, processing time, and memory consumption. Evaluation was performed on 5 test videos with manually labeled GT data, stemming from different sources. The data set contains videos with different illumination, some of them were compressed, some uncompressed, though distinction between BS challenges was neglected.

Parks and Fels [23] considered besides several BS approaches (only one of them state-of-the-art) various algorithms for foreground validation and post-processing. For evaluation they used videos from different sources (e.g. Wallflower data set [27]), too.

Recently, Herrero and Bescós presented a systematic evaluation of BS techniques [13] using an extensive data set of artificially combined foreground and background objects. However, this data set does not cover challenges like scene illumination and shadows. In their evaluation, they account for indoor/outdoor environment and dynamic backgrounds.

In contrast to the enumerated evaluations, we account for a broad variety of challenges for BS arising in the field of video surveillance. Therefore, we propose for evaluation an artificially created data set with accurate and objective GT data and shadow masks for every frame. This enables comprehensive evaluation of BS methods.

3. Challenges of BS for Video Surveillance

Background subtraction algorithms have to cope with several challenges arising from the nature of video surveillance. Besides the canonical problems defined by Toyama *et al.* [27] many different enumerations of BS challenges exist in literature. We refer to the work of Bouwmans *et al.* [4] for a comprehensive list. For evaluation we restrict ourselves to typical challenges of BS arising within the first two stages of Figure 2. Hence, we ignore challenges that are addressed in other stages (e.g. pre-processing) as well as problems stemming from a specific algorithmic approach. For evaluation we consider the following challenges:

Gradual illumination changes: It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day.

Sudden illumination changes: Sudden *once-off* changes are not covered by the background model. They occur for example with sudden switch of light, strongly affect the appearance of background, and cause false positive detections.

Dynamic background: Some parts of the scenery may contain movement, but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees).

Camouflage: Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications.

Shadows: Shadows cast by foreground objects often complicate further processing steps subsequent to BS. Overlapping shadows of foreground regions for example hinder their separation and classification. Hence, it is preferable to ignore these irrelevant regions.

Bootstrapping: If initialization data which is free from foreground objects is not available, the background model has to be initialized using a bootstrapping strategy.

Video noise: Video signal is generally superimposed by noise. BS approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artifacts.

Table 1. Overview and classification of the evaluated BS methods. FD stands for *Frame difference* and DM for *Difference to model*.

Method (first author)	Year	Model representation	Features	Model scale	Adaptive	Foreground detection
McFarlane [20]	1995	unimodal (median)	color	pixel	yes	DM
Stauffer [26]	1999	multimodal (Gaussian)	color	pixel	yes	DM
Oliver [22]	2000	linear subspace	pixel correlation (color)	frame	no	DM
McKenna [21]	2000	unimodal (Gaussian)	chromaticity, gradient	pixel	yes	DM
Li[2] [18]	2003	non-parametric (discretized)	color, co-occurrence	pixel	yes	FD, DM
Kim [16, 17][3]	2004	multimodal (codeword)	color, luminance	pixel	yes	DM
Zivkovic[4] [29]	2006	multimodal (Gaussian)	color	pixel	yes	DM
Maddalena[5] [19]	2008	multimodal (mean)	color	pixel[7]	yes	DM
Barnich[6] [1]	2009	non-parametric/non-recursive	color	pixel[7]	yes	DM

4. Experimental Setup

4.1. Evaluation Data Set

For performance evaluation we provide nine different test scenarios of a typical surveillance setting (available online[1]), which each covers a specific challenge (cf. Section 3 - note, this does not mean that a scenario is not influenced by other challenges, but one challenge is predominant). Figure 7 provides two exemplary frames of the data set. The following enumeration introduces the scenarios:

Basic: This is a basic surveillance scenario combining a multitude of challenges for general performance overview.

Dynamic Background: Uninteresting background movements are considered using a detail of the *Basic* sequence, containing moving tree branches and changing traffic light.

Bootstrapping: This scenario contains no training phase, thus subtraction starts after the first frame.

Darkening: Gradual scene change is simulated by decreasing the illumination constantly. Thus, background and foreground darkens and their contrast decreases.

Light Switch: *Once-off* changes are simulated by switching off the light of the shop (frame 901) and switching it on again (frame 1101).

Noisy Night: *Basic* sequence at night, with increased sensor noise accounting for high gain level and low background/foreground contrast resulting in more camouflage.

Shadow: We use a detail of the street region to measure shadow pixels classified as foreground.

Camouflage: Detail of the street region is used, too. We compare performance between a sequence with persons wearing dark clothes and gray cars, and a sequence containing colored foreground objects significantly differing from the background.

Video Compression: *Basic* sequence compressed with different bitrates by a standard codec often used in video surveillance (H.264, 40-640 kbits/s, 30 frames per second).

The sequences were rendered by *Mental Ray*, a raytracer provided by *Autodesk Maya*, while the GT data was generated by *Maya Vector*. The sequences have a resolution of 800×600 pixels and are captured from a fixed viewpoint. For some scenarios we consider a detail of a sequence to focus on regions with high impact to a specific problem. To obtain realistic footage, sensor noise is simulated by adding Gaussian distributed ($\mu = 0.0$, $\sigma = 0.0001$) noise to each frame. Only for *Noisy Night* sequence we use a broader noise distribution ($\mu = 0.0$, $\sigma = 0.0025$). Besides GT masks, we provide *shadow masks* indicating the luminance change introduced by foreground objects. In this context shadows do not necessarily mean a decrease of illumination, but describe the color difference of a background pixel in the presence of foreground objects. We split the data set into two parts, one for *training-phase* (in general without foreground objects) and one for *subtraction-phase* (with foreground objects). Except for *Bootstrapping* sequence, the training phase consists of 801 frames. The subtraction phase consists of 600 frames with the exception of *Darkening* and *Bootstrapping* both having 1400 frames.

4.2. Performance Measure

We judge the performance of BS methods on pixel-level. Thus, we consider foreground detection as binary classification of each pixel, resulting in a segmentation mask. The correctness of this classification is expressed by means of *recall*, *precision* and their harmonic mean, the *F-Measure*:

$$recall = \frac{\#\text{correctly classified foreground pixels}}{\#\text{foreground pixels in GT}}$$

[1] http://www.vis.uni-stuttgart.de/~hoeferbn/bse/

[2] OpenCV BS algorithm: http://opencv.willowgarage.com

[3] Codebook algorithm [16] with layer extension [17]

[4] Implementation from author: http://www.zoranz.net/

[5] Implementation from author: http://www.na.icar.cnr.it/~maddalena.l/MODLab/SoftwareSOBS.html

[6] Implementation from author: http://www2.ulg.ac.be/telecom/research/vibe/

[7] Per pixel model, but with regional diffusion in update step.

Table 2. Parameters used for evaluation on *Basic* sequence. ROC-optimized parameters are shown in boldface, parameters taken from literature are represented by plain text. Threshold varied to plot precision-recall charts marked with *[plotted]*.

Method/Parameter	Model representation	Adaptivity	Foreground detection
McFarlane	-	update-step $= \pm 1$ (per color)	threshold = [plotted]
Stauffer	$K = 5$	$\alpha = \mathbf{0.0025}$, $T = 0.8$	std-factor = [plotted]
Oliver	$\mathbf{M = 15}$	-	t = [plotted]
McKenna	-	$\alpha = \mathbf{0.005}$	std-factor = [plotted]
Li	c: $N_1 = 15$, $N_2 = 25$, cc: $N_1 = 25$, $N_2 = 40$	$\alpha_\mathbf{1} = \mathbf{0.01}$, $\alpha_\mathbf{2} = \mathbf{0.005}$, $T = 90\%$	δ = [plotted]
Kim	-	$\mathbf{T}_{\mathscr{H}} = \mathbf{50}$, $\mathbf{T_{add}} = \mathbf{50}$, $\mathbf{T_{delete}} = \mathbf{75}$	$\alpha = \mathbf{0.55}$, $\beta = \mathbf{1.3}$, ϵ_2 = [plotted]
Zivkovic	$M = 4$	$\alpha = \mathbf{0.0013}$, $\sigma_\mathbf{0} = \mathbf{15}$	T_b = [plotted]
Maddalena	$n = 3$	$c_1 = 1$, $\mathbf{c_2} = \mathbf{0.005}$, $\epsilon_\mathbf{1} = \mathbf{0.15}$	ϵ_2 = [plotted]
Barnich	$n = 20$	subsampling-factor $= 16$	$\#_{min} = 2$, R = [plotted]

$$precision = \frac{\#\text{correctly classified foreground pixels}}{\#\text{pixels classified as foreground}}$$

$$F_1 = 2\frac{recall \cdot precision}{recall + precision}$$

We provide most of the results as precision-recall charts (varying the threshold parameter), since this presentation is favored over the commonly used ROC (Receiver Operating Characteristic) curves in the case of large skew in the class distribution [8]. Sometimes, the F-Measure of maximal performance (averaged over sequence) is illustrated to expose correlation to another variable (e.g., compression). Finally, we provide foreground masks for visual inspection in Figure 7. Further results are provided on our website[1].

4.3. Evaluated Approaches

4.3.1 Background Subtraction Techniques

For evaluation, we have selected a wide range of BS methods including recent methods and more conventional approaches, in order to point out potential directions for further research. There exist different classifications of BS methods in literature, of which none seems to be fully accepted. In Table 1 we present a consistent classification of the methods which we evaluate in this paper with respect to different classification schemes proposed in literature.

4.3.2 Post-Processing Methods

As already mentioned, we restrict ourselves to the evaluation of BS with post-processing methods up to level of *foreground mask* representation (see Figure 2). We further focus on most commonly used approaches, in particular: median filter, morphological operations (*Opening*, *Closing*, and their combination), and shadow removal. For the task of shadow removal, we apply the algorithm *Cucchiara* [7] proposed by Prati *et al.* [25] for general-purpose shadow detection with minimal assumptions. Based on the observation that several BS methods utilize post-processing to increase their performance, we expect post-processing to generally narrow performance differences between BS approaches.

4.3.3 Parameter Selection

The selection of appropriate parameters is critical to the evaluation of BS methods. There exist several parameter adaption methods in literature (e.g. [11, 28]), among them the common ROC analysis, which we use to set the parameters of the evaluated BS methods. For simplicity, we only search for optimal parameters around initial parameters published by the authors of the particular BS methods, since we assume only small adaption in order to optimally fit to our test set. Further, we only adapt parameters that are critical to the success of a BS method with respect to a particular experiment (e.g., model adaption parameter has to be separately selected for *Basic* and *Bootstrapping* sequence). Also, the parameters are optimized one by one, applying the previously adapted parameters or initial parameters from literature. Additionally, we pay attention to keep the power of the model representation comparable across different methods (e.g., methods should maintain about the same number of modes, if possible). Table 2 shows the parameters used for the *Basic* experiment.

5. Experimental Results

First, we compare the different BS methods neglecting any post-processing to evaluate the performance of their model representation and model update. Additionally, we introduce a *don't care* border around the GT foreground objects of 1 pixel to account for background/foreground blending due to image discretization. Pixels on this border are not considered in evaluation. Note that for *Kim* no results are available for *Light Switch* and *Noisy Night* due to extensive generation of codewords in dark and noisy regions, which result in heavy memory footprint.

Basic: Before we address the challenges defined in Section 3, *Basic* sequence provides a first impression about the performance of the approaches in a typical surveillance scenario. Results are depicted in Figure 3 (a). Note that for *Kim*, only a fraction of precision-recall interval is covered by the applicable range of the threshold parameter ϵ_2. Also remarkable is that *Li* shows decreasing precision at low recall. This seems to stem from the combination of change detection by frame differences and background subtraction. Segmentation masks in Figure 7 exhibit that *Kim* has difficulties to properly handle dark regions (e.g. windows). This is due to a small initial luminance interval of codewords generated in almost black regions.

Dynamic Background: By using a detail of *Basic* sequence, we review the capability of each background representation to cope with uninteresting movement that has to be deemed as background. Figure 3(b) illustrates that hardly any approach is able to reach more than 50% precision at recall level higher than 90%. Segmentation masks (cf. Figure 7) provide insight into the shortcomings of each BS method. While most background models have problems with the proper representation of waving branches, some also show weaknesses with switching traffic light. Approaches using a *Gaussian Mixture* model, such as *Stauffer* and *Zivkovic*, show besides *Maddalena* the best representation of dynamic backgrounds.

Bootstrapping: Bootstrap training is the ability to adaptively create a proper model of the background even if training data contains foreground objects. We conduct the experiment of bootstrap model learning for all BS methods except for *Oliver*, since this approach does not support model updates. An alternative experiment (evaluation after a model was trained with 801 frames containing foreground objects) conducted for the *Oliver* method does only show decrease in precision of few percentage points. Also, most of the other approaches show only little decrease in performance when bootstrap training is applied (see Figure 3(c)). However, it is remarkable that *Li* does not reach recall values above 85%. Only *Maddalena* shows major performance loss within this experiment, which results from using only the second update weight c_2, even though optimized for this experiment.

Darkening: The quality of model adaption to gradually changing background appearance of all nine methods is depicted in Figure 3(d), even though, *Oliver* does not support model updates. Hence, it shows very low performance. However, it surprises that the performance of *Stauffer* is very low, too. This originates from the quick weight update of the Gaussian mode, while the Gaussian distribution itself is adapted very slowly (cf. [4]). In this experiment, the model update of *Zivkovic* - an extension of *Stauffer* - does a better job, but still is not fully satisfying. While dimming, the issue with dark areas of *Kim* becomes evident. The most robust approaches to gradual illumination changes are *Barnich*, *Li*, and *Maddalena*.

Light Switch: None of the tested BS methods was able to satisfactorily handle sudden *once-off* changes in illumination, as Figure 3(e) points out. Merely, the approaches that incorporate two different features (*Li*, *McKenna*) were able to stand somewhat out from the bad performance of the other methods. However, *Li* was not able to properly handle the *once-off* changes of this experiment, although it is equipped with two different detection modules.

Noisy Night: The most challenging experiment we have conducted includes a typical surveillance scenario at night. Signal to noise ratio is low, due to large sensor noise stemming from high gain settings, while foreground to background contrast is low which results in camouflage of foreground objects. Accordingly, the results of this experiment show quite low performance for all evaluated approaches (cf. Figure 3(f)). However, the three most recent BS methods (*Zivkovic*, *Maddalena*, and *Barnich*) make the best of it.

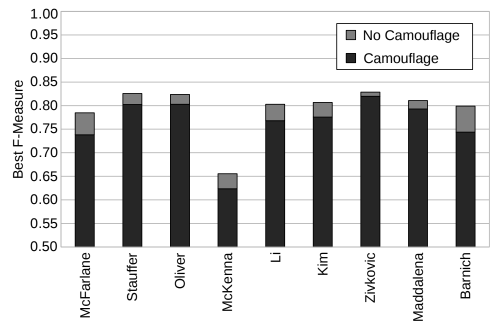
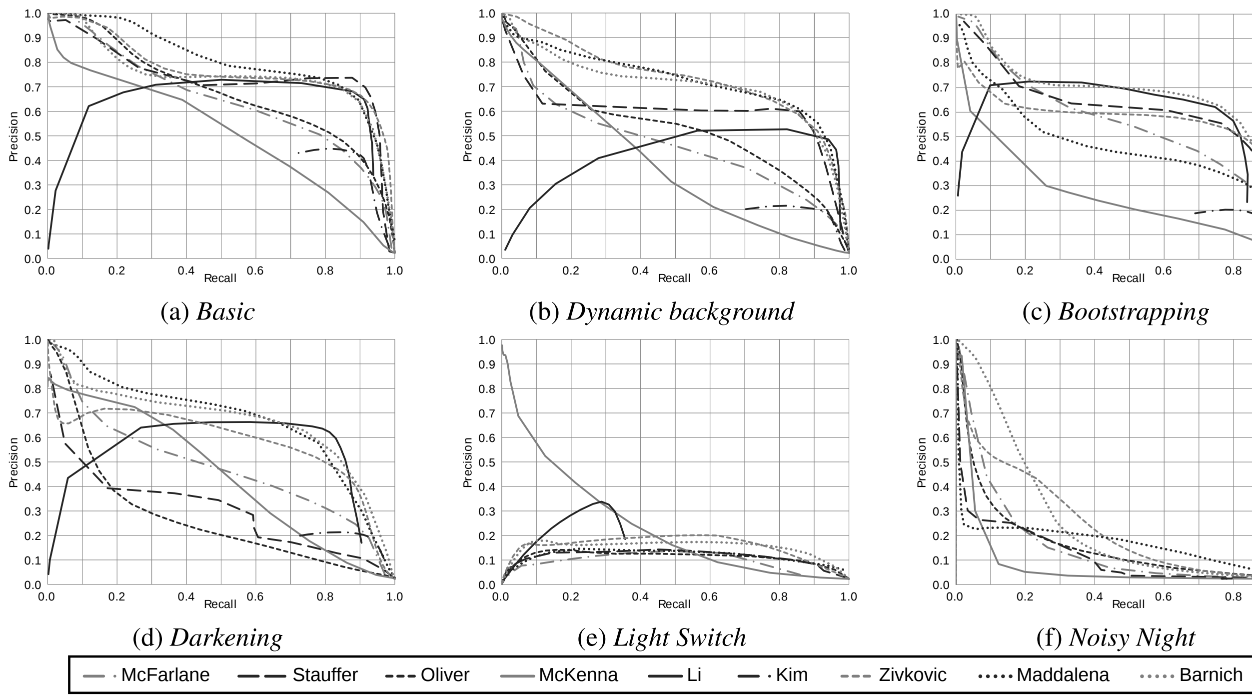


Figure 4. Maximal (average) F-Measure of sequence with and without camouflage of foreground objects.

Camouflage: The experiment is conducted on a detail of the scenery, in order to provide more significant results. Figure 4 illustrates the difference in F-Measure between foreground objects standing out from background and camouflaged ones. Besides the good results of *Zivkovic* and *Maddalena*, the linear subspace model of *Oliver* exhibits good performance, too. In contrast, the earliest (*McFarlane*) and the most recent approach (*Barnich*) we tested, show the strongest decrease in performance when faced with camouflaged foreground objects. We suppose this problem originates from the fixed threshold applied to classify pixels, instead of a distribution-depended threshold.

Video Compression: The results of BS on compressed video footage are surprising. Results show for many approaches hardly any noticeable decrease in performance (cf. Figure 5). Besides *Kim*, only methods using a *Gaussian* model begin to degrade noticeably around lower bitrates. Most approaches even benefit from some degree of com-



(a) *Basic* (b) *Dynamic background* (c) *Bootstrapping*

(d) *Darkening* (e) *Light Switch* (f) *Noisy Night*

Figure 3. Precision-recall charts of the performance of BS methods with varying threshold.

pression, probably because of the elimination of high frequency components (sensor noise) by the codec. Two BS methods (*McFarlane*, *Oliver*) even exhibit in this experiment a continuous improvement with decrease in bitrate.

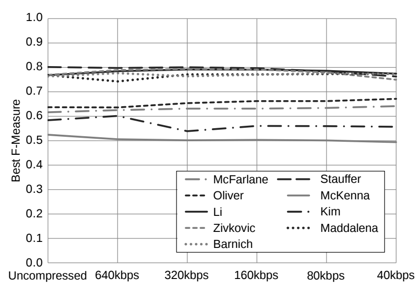
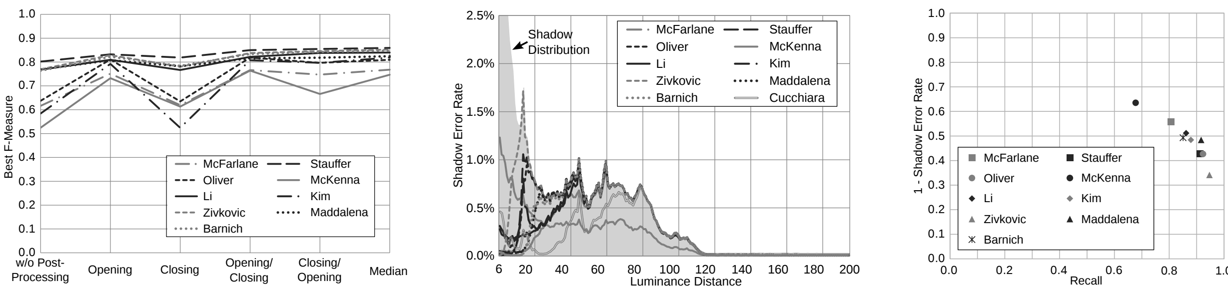


Figure 5. Maximal (average) F-Measure of *Basic* sequence compressed by H.264 with different bitrates.

Post-processing: The results of the experiments satisfy our expectation that post-processing of foreground masks is able to reduce the gap between stronger and weaker BS methods. Figure 6(a) illustrates the performance of different post-processing methods applied to the foreground masks obtained by several BS approaches. The results show F-Measure after post-processing with optimized parameters. However, only practical parameter ranges were tested (i.e. median mask size up to 20 pixels and square structure element size up to 7 pixels). We further expect that post-processing on higher abstraction levels (e.g. region thresholding) is able to reduce the remaining gap even more. *McKenna* and *Kim* benefit most from post-processing (performance increase > 24 percentage points) while stronger methods such as *Li* are also able to catch up to the top methods. There is no approach which does not profit from any of the post-processing methods. The most beneficial post-processing methods are median filter (mask sizes between 14 to 20 pixels) and a combination of *Opening* and *Closing* (small *Opening* and big *Closing* structure elements).

Shadow: The approaches vary in their capability of classifying shadow pixels as background. Note, for fair comparison, we deactivated built-in post-processing algorithms for shadow removal of *Maddalena* and *Zivkovic*. Figure 6(b) shows besides the shadow misclassification rate of the BS approaches also the performance of *Cucchiara*. *Cucchiara* is evaluated using an ideal background model and parameters optimized under the constraint of at most 10% rejection of foreground pixels ($\alpha = 0.2$, $\beta = 1$, $T_{sat} = 0.13$, $T_{hue} = 0.4$). The shadow error rate is depicted in relation to the luminance distance between the shadow pixel color and the color of the pixel in case of absence of any shadow casting object. Additionally, we introduce a *don't care* for shadow pixels with luminance distance equal or smaller than 5. Many approaches reject shadow pixels with small luminance distance, but shadows with large distance are often classified as foreground. Further, it is noticeable that BS methods which show low shadow misclassification errors over a large distance spectrum (e.g., *McKenna*), show bad recall, too. Figure 6(c) illustrates this strong correlation.



(a) (b) (c)

Figure 6. Performance of post-processing methods: (a) Best F-Measure for median filtering and morphological operations, (b) shadow misclassification rate (normalized to total number of shadow pixels) with respect to luminance distance between shadow pixel and unshaded background (includes *Cucchiara* shadow removal method and shadow distribution), (c) relationship between shadow error rate and recall.

6. Discussion and Conclusion

We have evaluated the performance of nine methods for BS including post-processing with respect to the challenges of video surveillance. Even though every of the tested BS approaches did exhibit drawbacks in some experiments, we found *Li*, *Zivkovic*, *Maddalena*, and *Barnich* to be the most promising ones. But it is also noticeable that especially weak approaches benefit most from pixel-level post-processing and thus are able to catch up with the top ones. We'd like to point out the promising trend of regional diffusion of background information in the update step, which is used by *Maddalena* and *Barnich*. Even though actual methods show good performance for many challenges, they fail in more demanding experiments such as *Noisy Night* and *Light Switch*. Here, we see further research requirement.

In this work we neglected further aspects of BS methods like time and space complexity as well as the number of parameters to be tweaked for a particular sequence. Considering these aspects, *Barnich* is a strong favorite, since it is simple and almost parameterless. But we suggest for *Barnich* to use a threshold in relation to the samples in the model for a better handling of camouflaged foreground.

Acknowledgements

The authors like to thank Johannes Kollmer for modelling the ground-truth scene. This work was funded by German Research Foundation (DFG) as part of the Priority Program "Scalable Visual Analytics" (SPP 1335).

References

[1] O. Barnich and M. Van Droogenbroeck. Vibe: A powerful random technique to estimate the background in video sequences. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 945 –948, 2009. 1939

[2] T. Bouwmans. Subspace learning for background modeling: A survey. *Recent Patents on Computer Science*, 2(3):223–234, 2009. 1937

[3] T. Bouwmans, F. El Baf, and V. B. *Statistical Background Modeling for Foreground Detection: A Survey*, volume 4 of *Handbook of Pattern Recognition and Computer Vision*, chapter 3. World Scientific Publishing, 2010. 1937

[4] T. Bouwmans, F. El Baf, and B. Vachon. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3):219–237, 2008. 1937, 1938, 1941

[5] T. Chalidabhongse, K. Kim, D. Harwood, and L. Davis. A perturbation method for evaluating background subtraction algorithms. In *Proc. of Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003. 1937

[6] S. Cheung and C. Kamath. Robust techniques for background subtraction in urban traffic video. *Video communications and image processing, SPIE Electronic Imaging*, 5308:881–892, 2004. 1938

[7] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati. *The Sakbot system for moving object detection and tracking*, chapter 12, pages 145–157. Video-Based Surveillance Systems—Computer Vision and Distributed Processing. Kluwer Academic Publishers, 2001. 1940

[8] J. Davis and M. Goadrich. The relationship between precision-recall and roc curves. In *Proc. of Int. Conf. on Machine Learning*, pages 233–240. ACM, 2006. 1940

[9] S. Elhabian, K. El-Sayed, and S. Ahmed. Moving object detection in spatial domain using background removal techniques-state-of-art. *Recent Patents on Computer Science*, 1(1):32–54, 2008. 1937

[10] C. Erdem, A. Tekalp, and B. Sankur. Metrics for performance evaluation of video object segmentation and tracking without ground-truth. In *IEEE Int. Conf. on Image Processing*, 2001. 1937

[11] X. Gao, T. Boult, F. Coetzee, and V. Ramesh. Error analysis of background adaption. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 503–510, 2000. 1940

[12] E. Grossmann, A. Kale, and C. Jaynes. Offline generation of high quality background subtraction data. In *Proc. of British Machine Vision Conference*, 2005. 1937

[13] S. Herrero and J. Bescós. Background subtraction techniques: Systematic evaluation and comparative analysis. In

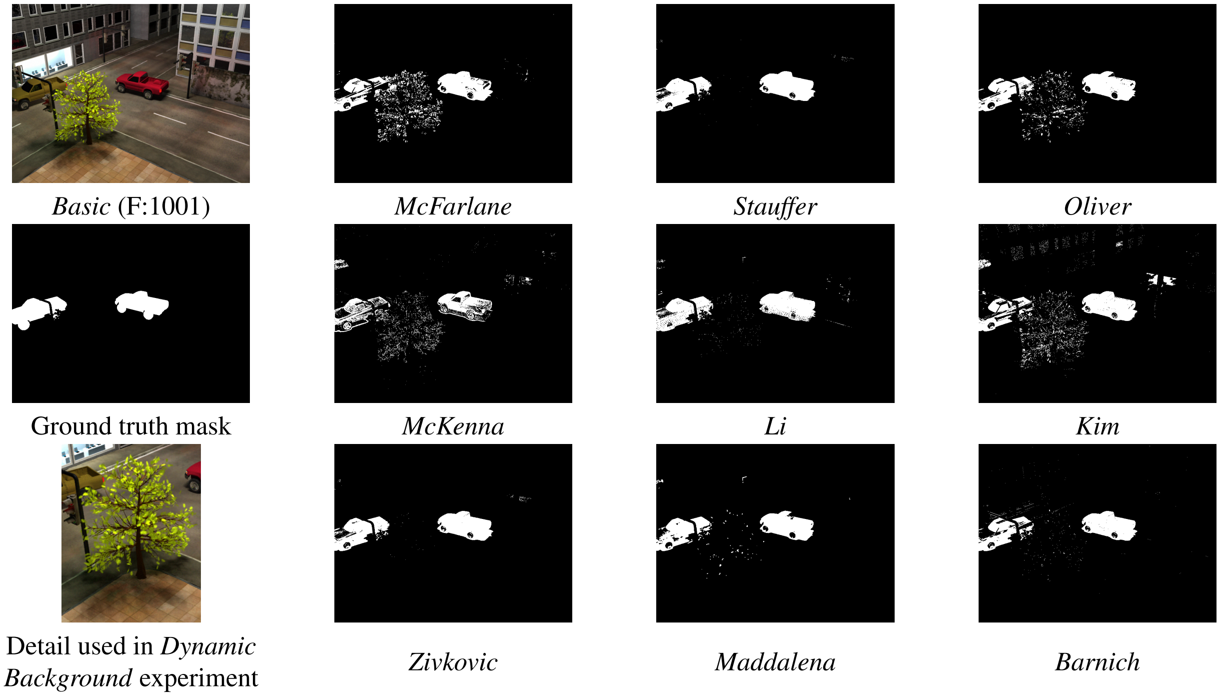


Figure 7. Representative foreground masks at best (averaged) F-Measure for *Basic* and *Gradual illumination change* sequence. Ground truth is depicted including *don't care* boundary pixels.

Advanced Concepts for Intelligent Vision Systems, pages 33–42. Springer, 2009. 1938

[14] O. Javed and M. Shah. *Automated Multi-Camera Surveillance: Algorithms and Practice*. Springer, 2008. 1937

[15] M. Karaman, L. Goldmann, D. Yu, and T. Sikora. Comparison of static background segmentation methods. In *Proc. SPIE*, volume 5960, pages 2140–2151. SPIE, 2005. 1938

[16] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. In *Int. Conf. on Image Processing*, pages 3061–3064, 2004. 1939

[17] 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. 1939

[18] L. Li, W. Huang, I. Gu, and Q. Tian. Foreground object detection from videos containing complex background. In *Int. Conf. on Multimedia*, pages 2–10. ACM, 2003. 1939

[19] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 17(7):1168–1177, 2008. 1939

[20] N. McFarlane and C. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3):187–193, 1995. 1939

[21] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *Computer Vision and Image Understanding*, 80(1):42 – 56, 2000. 1939

[22] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):831, 2000. 1939

[23] D. Parks and S. Fels. Evaluation of background subtraction algorithms with post-processing. In *Proc. of IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 192–199, 2008. 1938

[24] M. Piccardi. Background subtraction techniques: a review. In *IEEE Int. Conf. on Systems, Man and Cybernetics*, volume 4, 2004. 1937

[25] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):918–923, 2003. 1940

[26] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999. 1939

[27] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. of Int. Conf. on Computer Vision*, pages 255–261, 1999. 1938

[28] B. White and M. Shah. Automatically tuning background subtraction parameters using particle swarm optimization. In *IEEE Int. Conf. on Multimedia and Expo*, pages 1826–1829, 2007. 1940

[29] Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27:773–780, 2006. 1939