S. Piérard and M. Van Droogenbroeck. **A perfect estimation of a background image does not lead to a perfect background subtraction : analysis of the upper bound on the performance**. In *International Conference on Image Analysis and Processing (ICIAP), Workshop on Scene Background Modeling and Initialization (SBMI), Lecture Notes on Computer Science*, to appear, 2015.

# A perfect estimation of a background image does not lead to a perfect background subtraction: analysis of the upper bound on the performance

S. Piérard, M. Van Droogenbroeck

INTELSIG Laboratory, Department of Electrical Engineering and Computer Science, University of Liège, Belgium

**Abstract.** The quest for the "best" background subtraction technique is ongoing. Despite that a considerable effort has been undertaken to develop flexible and efficient methods, some elementary questions are still unanswered. One of them is the existence of an intrinsic upper bound to the performance. In fact, data are affected by noise, and therefore it is illusory to believe that it is possible to achieve a perfect segmentation. This paper aims at exploring some intrinsic limitations of the principle of background subtraction. The purpose consists in studying the impact of several limiting factors separately. One of our conclusions is that even if an algorithm would be able to calculate a perfect background image, it is not sufficient to achieve a perfect segmentation with background subtraction, due to other intrinsic limitations.

## 1 Introduction

The background subtraction (BGS) is a well studied problem [2,10] for which, despite the impressive amount of methods proposed in the literature so far, no satisfactory technique has been found yet (for all cases) [6,7]. This paper discusses the limits of pixel-based BGS methods. They aim at classifying, for each frame of a video sequence, pixels in the foreground (FG) or background (BG) classes by performing a motion analysis. It is an online unsupervised one-class classification problem, as the goal is to learn the distribution of the background colors on-the-fly, in an unsupervised way (no sample with known label FG/BG being provided *a priori*), and to classify new colors based on the representation (named the *model*) of this sole BG class. Note that even if other types of features are sometimes considered (*e.g. local binary patterns* in [9], *local binary similarity patterns* in [13], or gradients [8]), we limit the scope of this paper to colors.

Each classification problem has an intrinsic performance limit; this holds also for background subtraction. Indeed, a perfect classifier can only be theoretically obtained when the distributions of the samples (the colors in the case studied in this paper) of the two classes have disjoint supports. For all other cases, the class overlapping introduces an upper limit to the performance. In the context of BGS, a class overlapping originates from similar colors in the foreground and background. This can result, for example, from the noise in the images (this noise

coming from the sensor or from the compression of the video stream), varying lighting conditions, shadows, or camouflage. We are interested in working out whether, aside from the theoretical limit, there is some room left for improving the performance of BGS techniques, or if the limit has already been reached. In the case of BGS, the performance limit also originates from the way the decisions are taken (the per-pixel decisions have to be fast in order to work in real time, and memory constraints prevent from storing a lot of information per pixel). This should also be considered when we discuss the performance limit.

The outline of this paper is as follows. Section 2 presents how the decisions are traditionally taken in BGS algorithms. Dividing the processing pipeline into the initialization and updating parts on the one hand, and the segmentation part on the other hand, allows us to present results independent of any choice for a particular BGS algorithm. As the initialization and updating parts essentially aim at estimating the background image (more information about this topic can be found in the SBMI workshop [11]), we derive a performance bound that depends on the amount of noise affecting this image (that is the model). Our approach to compute this bound is presented in Section 3. Bounds computed for a few common decision rules (the segmentation processes) are presented and discussed in Section 4. In addition, we show how these bounds vary with respect to the main characteristics of the video sequences (amount of noise, quantity and magnitude of shadows, proportion of foreground). Finally, the conclusion is given in Section 5. Our results establish that being able to compute a perfect background image does not really help to raise the performance limit, due to the other intrinsic bottlenecks of the BGS problem (even if this is important for other applications such as video inpainting or computational photography).

## 2 The traditional processing pipeline of BGS algorithms

*The segmentation process.* Even if online learning, unsupervised learning, and one-class classification are studied by the machine learning community, most of the BGS algorithms do not leverage machine learning techniques. The trend is to directly threshold the distances between colors stored in the model, as in a nearest neighbors analysis, or to test if the observed color follows the distribution encoded in the model. The difference originates from the constraints of the BGS: the per-pixel classifiers have to run in real time with low memory, and to adapt themselves with only a few observed samples of the (supposed) BG class.

*The variety of models.* In the simple case of the frame difference algorithm, the model is the color observed at the same location in the previous frame. To the contrary, conservative approaches try to build a model describing only the background. Strictly speaking, the model encodes the distribution of colors *predicted as being* in the background, instead of encoding the distribution of colors *being* in the background. This nuance should not be overlooked as it is intrinsic to any unsupervised one-class classification approach. A large family of BGS methods use probabilistic models of the background, often parametric

ones. For example Wren *et al.* [15] supposed a Gaussian distribution, and adapt only the mean and variance of this distribution on a pixel basis. Other well-known methods of this family include the mixture of a fixed amount of Gaussians proposed by Stauffer and Grimson [14], and the mixture of Gaussians with an adaptive amount of components introduced by Zivkovic [16]. As an alternative to these probabilistic models, sample-based ones have been developed. They represent the background distribution with a set of samples drawn from this distribution. This is the case for the KDE [4] and ViBe [1] algorithms. In this study, we assume that the model is an estimated background image (this is the topic of the SBMI workshop [11]).

*Motivation for discussing the limits of pixel-based BGS methods.* Aside from the particular technical details of BGS algorithms, three elements of background subtraction should be considered:

1. [**Initialization**] The initialization aims at learning a good model from as few frames as possible from the video. Typically, when foreground objects are present in the first frames, the model needs time to erase the foreground objects, leading the appearance of so-called ghosts.
2. [**Updating**] The model has to be maintained to deal with temporal changes in the scene. Note that periodic or quasi-periodic modifications with high frequency are often considered as giving rise to a distribution of background colors that is a mixture, instead of a varying distribution.
3. [**Segmentation**] The result of the classification process is a segmentation map, with identified foreground or background pixels. Spatial coherence can be enhanced by post-processing the segmentation masks, or by propagating the neighboring distributions into the pixel's model as done by ViBe. Both techniques can also be combined. Note that post-processing techniques are known to always increase the performance [3,12]. In this paper, we study the performance of the BGS without any post-processing.

While authors propose sophisticated methods to increase the performance, it is interesting to understand if there are limitations, and where they originate from. To our knowledge, this question has not been explicitly studied in the literature. In order to discuss this theoretical question, we focus on the simple, but often encountered case, of a fixed background. It follows that the initialization and updating problems then become the problem of estimating the background image, which is the main focus of the SBMI workshop [11]. In this paper, we assume that the background image can be estimated, and we discuss the existence of theoretical upper bounds on the performance for pixel-based BGS algorithms.

## 3 Methodology

For some video sequences, foreground and background colors are so different that obtaining a perfect segmentation is trivial once the background is perfectly known. Considering such a video leads to a highly optimistic upper bound on the

performance. To the contrary, in case of the *camouflage* effect, the foreground and background colors are so close that is impossible to distinguish between the two classes and that the BGS algorithms have a performance limited to that of a random classifier. Such a video leads to a very pessimistic upper bound. The upper bound of the performance is therefore video specific. The difficult question consists to determine the expected upper bound for a video sequence with unknown characteristics.

It is hard to determine the upper bound directly from state-of-the-art datasets for evaluating BGS algorithms, such as *changedetection.net* [6,7]. The reason is that they contain only a few dozens of video sequences. Due to the large variations in the characteristics of the video sequences, the small size of these datasets prevents from estimating a statistically significant averaged upper bound by averaging the performances measured experimentally on each sequence. Moreover, considering the pixels contained in a few video sequences as the test set is suboptimal as there is a natural spatial and temporal coherence in videos, leading to poor diversity and most probably to a biased estimation of the upper bound.

For studying the performance of BGS techniques taking their decisions on the pixel values, it is not necessary to have a video sequence. The reason is twofold.

1. In the absence of post-processing, the pixel-based nature ensures that the neighboring pixels do not have to be considered in order to study the behavior of the BGS for a pixel. Moreover, because we assume the background image can be computed, there is no need to consider the past of the video to study the behavior of the BGS: the model does not change over time as it represents the real background, regardless of the current frame.
2. The evaluation of the segmentation map occurs on a pixel base in most benchmarks such as *changedetection.net* [2,10], which indicates that the spatial coherence is not the primary concern. In fact, none of the 7 metrics computed on that website depends on the temporal or spatial order of pixels.

We argue that, in consequence, the expected upper bound can be obtained by choosing the test samples randomly in the space of pixels (which is of low dimension) instead of selecting them in the space of video sequences (which has an intractable dimension). Accordingly, we decide to simulate synthetic distributions at the pixel level, and to measure the upper bound experimentally. Note that our methodology could also be used to calculate upper bounds for region-based BGS methods, but in that case a statistical model of the spatial coherence would be necessary. In the absence of any prior information about the observed colors, we assume an uniform distribution of colors in the RGB space, both for the foreground and the background. All colors components are assumed to be real numbers between 0 and 1, but final RGB values are quantized over 8 bits in the input images as well as for the background image stored in the model. The noise statistic is supposed to be Gaussian, as assumed in several BGS techniques [14,15,16]. More precisely, we draw noise values randomly, independently for each channel, from a centered normal distribution truncated in such a way that the noisy color is still in the range of possible colors (*i.e.* all components between 0 and 1). Noise can affect the observed images as well as the background

(a) original image (b) noise with $\sigma = 0.06$ (c) value decreased by 0.3

Fig. 1: Illustration of the magnitude of simulated imperfections.

image (that is the model). A standard deviation of 0.06 is considered as realistic (see Figure 1). The noise level in the model can be lower than the one in the input images if the BGS method uses a temporal noise filtering technique. We also simulate the shadows affecting the background part of the input images due to the foreground elements in the scene. In our experiments, the shadows are not present in the model, but only in the observed image. Shadows decrease the value channel; we consider a decrease of 0.3 as being typical (see Figure 1).

We wrote a software for computing ROC curves [5] with a Monte-Carlo approach, given (1) the amount of noise corrupting the input image, (2) the quality of the background image estimation that is stored in the model (that is the quantity of noise affecting it), (3) the average proportion of shadowed pixels, and (4) the corresponding decrease of value. We also compare four segmentation rules, that correspond to different ways to build the value to be thresholded:

- **[V1]** $|C_{input} - C_{model}|$ with $C \in \{R, G, B\}$ for grayscale images;
- **[V2]** $\sum_{C \in \{R,G,B\}} |C_{input} - C_{model}|$,
- **[V3]** $\max_{C \in \{R,G,B\}} |C_{input} - C_{model}|$,
- **[V4]** and $\sum_{C \in \{R,G,B\}} (C_{input} - C_{model})^2$ for color images.

These ROC curves are our upper bounds on the performance limit of BGS algorithms performing per-pixel segmentation based only on the color information.

## 4 Results and discussion

The simulated ROC curves of a few upper bounds are given in Figure 2. The first observation is that ROC curves for V2, V3, and V4 are always very close, and always improve with respect to the segmentation rule V1 (based on grayscale values only). Working with one channel images is therefore suboptimal, and the performance does not depend much on the decision rule itself. The second observation, by comparing the ROC plots of the first and second rows, is that the performance is not very different when the model contains the perfect background image or a noisy version of it. *Being able to estimate a very precise background image does not help much for the BGS, due to other intrinsic limitations of the BGS problem* (at least under the working assumptions of this paper).

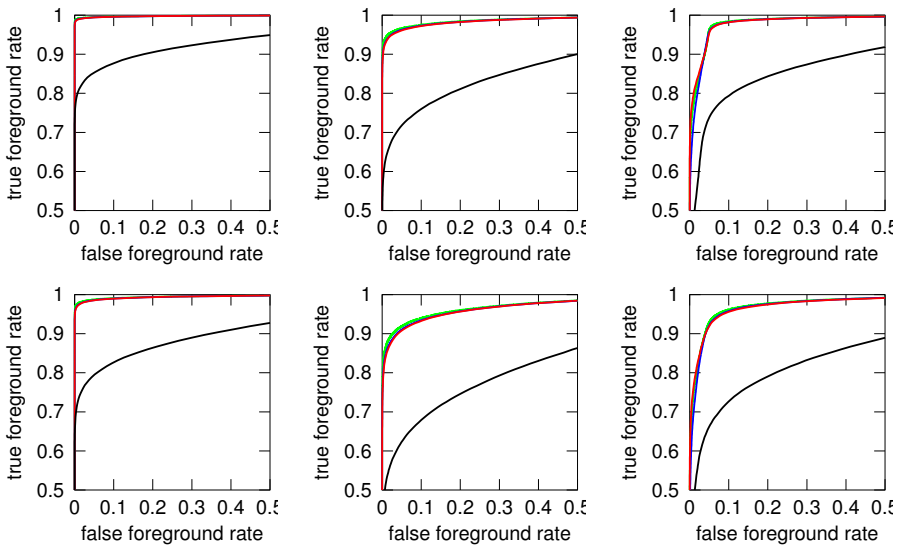


Fig. 2: Some upper bounds obtained for the performance of BGS methods, in the ROC space. In the left column, the input images have a low noise level ($\sigma = 0.04$) and no shadows. In the middle column, they have a high noise level ($\sigma = 0.08$), and no shadows. In the right column, they have a mid noise level ($\sigma = 0.06$) and 5% of their pixels in shadowed areas (decrease of value: 0.3). The model comprises a perfect background image in the upper row, and a noisy estimate of it in the lower row (the same amount of noise is added to the model and to the input images). The segmentation rules are: ■ V1, ■ V2, ■ V3, ■ V4.

In order to show how the performance evolves with the amount of noise affecting the estimate of the background image, we need to express the performance with a numerical value instead of a curve. Once the decision threshold is set, the performance is given by a single point in the ROC space. We argue that a good way of choosing the decision threshold is to force the classifier to be unbiased. In that case, the BGS method predicts the right proportion of foreground. Unbiased classifiers are, in the ROC space, on the line passing through the points $(\mathrm{TNR}, \mathrm{TPR}) = (1, 1)$ and $(1 - p^+, p^+)$, where TNR denotes the true negative rate, TPR the true positive rate, and $p^+$ the prior of the positive class. The average $p^+$ is 4.56 % in the dataset of [6]. The performance resulting from the threshold selection is the intersection between the ROC curve and this line. Many metrics could be used to measure it. We report the balanced accuracy $\frac{\mathrm{TNR}+\mathrm{TPR}}{2}$. As there is a mapping between this value and the other metrics (see Figure 3), reporting how the balanced accuracy of the unbiased classifier varies with the amount of noise affecting the estimated background image suffices. Figure 4 shows that a very large standard deviation of noise ($\simeq 0.1$) should be reached before observing a significant decrease of performance.

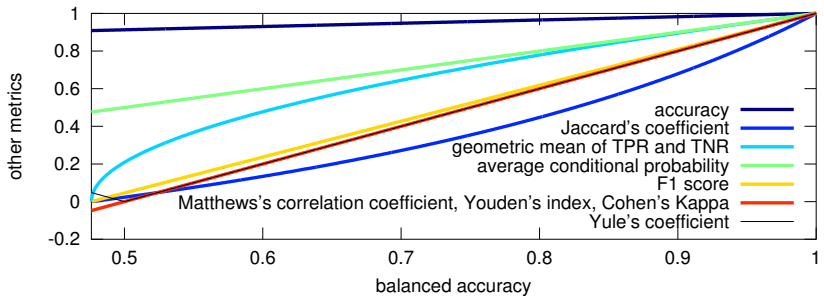


Fig. 3: Relationships between various metrics and the balanced accuracy that exist in the case of unbiased classifiers and $p^+ = 4.56\,\%$.

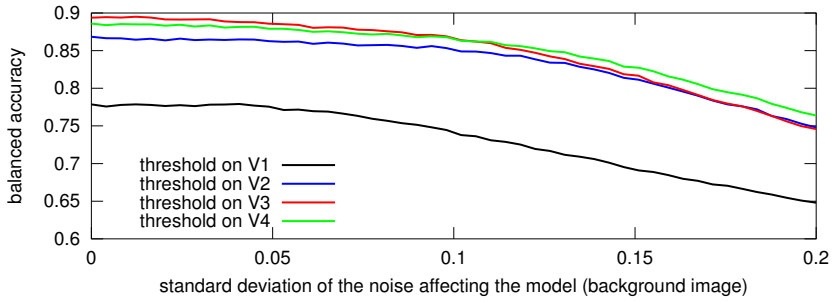


Fig. 4: The balanced accuracy of unbiased BGS methods, with respect to the amount of noise affecting the estimated background image. These results correspond to video sequences such that the noise has a standard deviation of $\sigma = 0.06$, and $5\%$ of pixels are in shadowed areas with a value decreased by 0.3.

## 5 Conclusion

The SBMI workshop [11] focuses on the estimation of a background image given a video sequence taken from a static viewpoint; the background is assumed to be unimodal. In this context, we have proposed an original methodology to simulate the expected upper bound on the performance of pixel-based BGS algorithms, when their model reduces to the estimated background image. In our simulations, the provided bounds depend on the amount of noise corrupting the video sequence, the quality of the background image estimation, and the proportion of shadowed pixels. One important conclusion is that the quality of the estimate of the background image helps for the BGS, but only marginally, because of other intrinsic limitations. This questions the need for a perfect estimation of the background in general. Note that the presented methodology could also be tuned to derive bounds for a given video sequence, or a family of them (*e.g.* BGS for video surveillance of roads), by adapting the distribution of colors.

## References

1. O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.*, 20(6):1709–1724, June 2011.
2. T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12:31–66, May 2014.
3. S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *IEEE Int. Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 1937–1944, Providence, Rhode Island, USA, June 2011.
4. A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *European Conf. Comput. Vision (ECCV)*, volume 1843 of *Lecture Notes in Comp. Science*, pages 751–767, London, UK, June-July 2000. Springer.
5. T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006.
6. N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. changedetection.net: A new change detection benchmark dataset. In *IEEE Int. Conf. Comput. Vision and Pattern Recognition Workshop (CVPRW)*, Providence, Rhode Island, USA, June 2012.
7. N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. A novel video dataset for change detection benchmarking. *IEEE Trans. Image Process.*, 23(11):4663–4679, Nov. 2014.
8. S. Gruenwedel, P. Van Hese, and W. Philips. An edge-based approach for robust foreground detection. In *Advanced Concepts for Intelligent Vision Syst. (ACIVS)*, volume 6915 of *Lecture Notes in Comp. Science*, pages 554–565. Springer, 2011.
9. M. Heikkilä and M. Pietikäinen. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(4):657–662, Apr. 2006.
10. P.-M. Jodoin, S. Piérard, Y. Wang, and M. Van Droogenbroeck. Overview and benchmarking of motion detection methods. In T. Bouwmans, F. Porikli, B. Hoferlin, and A. Vacavant, editors, *Background Modeling and Foreground Detection for Video Surveillance*, chapter 24. Chapman and Hall/CRC, July 2014.
11. L. Maddalena and T. Bouwmans. Scene background modeling and initialization (SBMI) workshop. http://sbmi2015.na.icar.cnr.it, Sept. 2015.
12. D. Parks and S. Fels. Evaluation of background subtraction algorithms with post-processing. In *IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, pages 192–199, Santa Fe, New Mexico, USA, Sept. 2008.
13. P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. SuBSENSE: A universal change detection method with local adaptive sensitivity. *IEEE Trans. Image Process.*, 24(1):359–373, Jan. 2015.
14. C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Int. Conf. Comput. Vision and Pattern Recognition (CVPR)*, volume 2, pages 246–252, Ft. Collins, USA, June 1999.
15. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):780–785, July 1997.
16. Z. Zivkovic. Improved adaptive gausian mixture model for background subtraction. In *IEEE Int. Conf. Pattern Recognition (ICPR)*, volume 2, pages 28–31, Cambridge, UK, Aug. 2004.