BMOG: Boosted Gaussian Mixture Model with Controlled Complexity

Isabel Martins[1,2], Pedro Carvalho[2,3] Luís Corte-Real[3,4], and José Luis Alba-Castro[1]

[1] University of Vigo, Vigo, Spain
[2] School of Engineering, Polytechnic Institute of Porto, Porto, Portugal
mis@isep.ipp.pt
[3] INESC TEC, Porto, Portugal
[4] Faculty of Engineering, University of Porto, Porto, Portugal

Abstract. Developing robust and universal methods for unsupervised segmentation of moving objects in video sequences has proved to be a hard and challenging task. The best solutions are, in general, computationally heavy preventing their use in real-time applications. This research addresses this problem by proposing a robust and computationally efficient method, BMOG, that significantly boosts the performance of the widely used MOG2 method. The complexity of BMOG is kept low, proving its suitability for real-time applications. The proposed solution explores a novel classification mechanism that combines color space discrimination capabilities with hysteresis and a dynamic learning rate for background model update.

Keywords: GMM · MOG · background subtraction · change detection

1 Introduction

Unsupervised segmentation of moving objects in video sequences, based on background subtraction (BS), is a fundamental step in many computer vision applications. However, there is no universal solution that successfully deals with all the many challenges presented, including poor lighting conditions, sudden illumination changes and parasitic background motion. Comprehensive reviews of BS approaches have been presented [1, 2]. Recent research has shown that BS methods appear to be complementary in nature [3], driving to a more complex, and time consuming, solution in general.

Gaussian Mixture Model (GMM), or Mixture of Gaussians (MoG), has been well explored and it is probably the most popular strategy to model the background. It is a parametric model capable of handling several modes in a pixel value [4]. It can deal with slow lighting changes, periodical motion in the cluttered background, slow moving objects, long-term scene changes, and camera noise. It is widely used due to its computational efficiency and good performance in a large number of applications. These traits inspired improvements and extensions, such as [5, 6], many at the cost of increased computational load.

This paper proposes a robust and computationally efficient method to address the problem of BS. It is based on an adaptive GMM background model proposed by Zivkovic [5], commonly known as MOG2, which achieves increased performance in multi-modal backgrounds without penalizing computational performance, a critical issue in real time applications. This efficiency makes it a common choice in real-world applications, despite the emergence of other methods with better performance. Our method, named "Boosted MOG"(BMOG), explores the characteristics of the color spaces and further adapts the algorithm using simple but efficient rules to boost the performance of MOG2. An exhaustive set of experiments was performed on public datasets. Results show that BMOG consistently outperforms MOG2, and that it approaches top ranking, but much more complex, algorithms. Its controlled complexity makes it a good choice for real-time applications.

2 Selection of Color Space

RGB is the most common color space used in BS. However, color spaces such as YUV, YCbCr and CIE L*a*b*, that separate the luminance component, have proven to be advantageous in image processing applications [7, 8]. CIE L*a*b* is a perceptual color space, where the non-linear relationships for the L*, a*, and b* components are intended to mimic the Human Visual System features. The advantages of using other color spaces than RGB in BS have been pointed out in terms of increased robustness to noise and shadows [9].

We performed a Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) analysis to assess the discrimination capability of different color spaces based on the color distance between the input and the background model pixel generated by MOG2, applied to a set of samples obtained from frames with moving objects, extracted from different videos from the CDnet 2014 Dataset [10], with the corresponding ground truth classification. The distance used was the squared Mahalanobis distance. This study showed the superior discrimination capability, for the task of BS, of alternative color spaces, particularly CIE L*a*b* (AUC 0.8) compared to RGB (AUC 0.52). Moreover, this analysis revealed that the distance of each of the color components has a discrimination capability similar to the color distance (AUC for independent channels R, G, B: 0.527, 0.527, 0.526, and L*, a*, b*: 0.742, 0.716, 0.777, respectively). These results led us to conclude that not only the adoption of the CIE L*a*b* color space could yield a significant improvement in the classification accuracy but also to define a new decision rule where the distance of each of the color components could be used as an independent classifier followed by the fusion of the independent decisions. The combination of these independent classifiers revealed, experimentally, better results than just one classifier based on the color distance.

3 Background Model

The background model adopted is based on MOG2 [5]. We extended it to deal with the characteristics of color spaces with separate channels for luminance and chrominance. The proposed BMOG algorithm performs consistently better using L*a*b*, YUV or YCbCr color spaces than using RGB; from these three, the best results were achieved with L*a*b*. Separating luminance from chrominance allows to set meaningful variance thresholds for each channel. Apart from having different discrimination capabilities, large luminance variance threshold with smaller chrominance variance thresholds can also accommodate shadows without changing the state of the pixel. In our proposal, each channel component of L*a*b* is analyzed independently, and their decisions combined with the AND rule (that provided better results than a decision based on majority voting).

3.1 Pixel Classification with Hysteresis

For each Gaussian in the mixture, if the distance between one color component of the incoming pixel $x = \{x_L, x_a, x_b\}$ and the mean of the Gaussian component is above a pre-defined threshold, the overall match is rejected. Therefore, the decision rule for a sample being classified as belonging to the background (BG) becomes

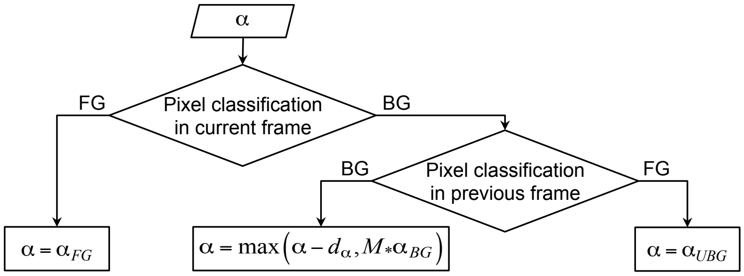
$$(x_L - \mu_L)^2 < (T_L \pm d_{th})\sigma_L^2 \wedge (x_a - \mu_a)^2 < (T_a \pm d_{th})\sigma_a^2 \wedge (x_b - \mu_b)^2 < (T_b \pm d_{th})\sigma_b^2 \quad (1)$$

where μ_L, μ_a, μ_b are the Gaussian means, σ_L, σ_a, σ_b are the Gaussian variances and T_L, T_a, T_b are the independent thresholds. As the probability of a pixel changing classification from the previous frame to the current frame is much lower than the probability of a pixel maintaining the same classification, a hysteresis mechanism has been implemented to prevent noisy pixels, whose color distance is very close to the decision threshold, from incorrectly changing the classification. Therefore, the threshold values in (1) depend on the classification of the same pixel in the previous frame. If the pixel was previously classified as FG, the threshold values for that pixel are decreased by d_{th} in order to make the change to BG more difficult; if the pixel was previously classified as BG, the threshold values are increased by d_{th} in order to hinder the change to FG.

3.2 Dynamic Learning Rate

The background model is updated by using the recursive equations of MOG2, described in [5], modified by embedding a conditional mechanism at the pixel level. The learning rate α is adapted independently for each pixel and depends on the change of classification decision, as shown in Fig. 1. In this context, a change from FG to BG is sub-classified as uncovered background (UBG) and is, therefore, treated differently from other BG samples. This approach is followed in order to avoid phantom images, by promoting a quick adaptation when the background is uncovered. Thus, if the pixel is classified as UBG a faster learning rate α_{UBG} is applied. This learning rate is then decreased for each successive frame by a fixed step size d_α until it reaches the dynamic BG learning rate. On the other hand, if the change is from BG to FG a slow learning rate α_{FG} is applied, in order to prevent foreground objects to be quickly absorbed by the background model, while assuring that pixels misclassified as FG in consecutive frames are not completely ignored.

As a MOG2 feature, dynamic areas of the scene are modeled using more gaussians than static areas. Hence, so that the learning rate is better adapted to the characteristics of the scene background, the dynamic BG learning rate is made dependent on the number of gaussians in the mixture M (M*α_{BG}). This ensures that the model adaptation is faster for dynamic areas and slower for static ones.



α
Pixel classification in current frame
FG
BG
Pixel classification in previous frame
BG
FG
$\alpha = \alpha_{FG}$
$\alpha = \max\left(\alpha - d_\alpha, M{*}\alpha_{BG}\right)$
$\alpha = \alpha_{UBG}$

Fig. 1. Conditional learning rate update at the pixel level.

3.3 Embedded Post-Processing

As we use the current BG/FG segmentation mask to make decisions when processing the next frame, the post-processing step is embedded in our algorithm to increase the success of decisions. The binary segmentation mask is filtered with an $N \times N$ median filter followed by filling of closed contours.

4 Experimental Setup

The experiments were conducted on the complete set of videos of the CDnet 2014 Dataset [10], consisting of 53 videos representative of a wide variety of challenges, grouped in 11 categories: Bad Weather (BW), Baseline (BL), Camera Jitter (CJ), Dynamic Background (DB), Intermittent Object Motion (IOM), Low Frame Rate (LFR), Night Videos (NV), Pan-Tilt-Zoom (PTZ), Shadows (SW), Thermal (TH) and Turbulence (TB). Evaluation was performed using the ground truth (GT) segmentation provided along with the videos. Pixels in the mask may have one of 5 labels: *Moving*, corresponding to foreground pixels;

Static, corresponding to background pixels; *Shadow* corresponding to moving shadows; $Non-ROI$ corresponding to regions outside the ROI; *Unknown* corresponding to pixels whose status is unclear.

These experiments involved the generation of all masks for our method, BMOG and, in a second step, the comparison with the results:

- for MOG2 [5],

- for a recent MoG based method RMoG [6], which claims to be computationally very efficient,

- and for a top-ranked state-of-the-art method, SuBSENSE [11],

using the results reported in the CDnet site [10] for all the methods. In the context of this comparison, the authors considered to be relevant to assess performance as a balance between complexity and segmentation quality. To this end, the OpenCV implementation of MOG2 was used as a reference. SuBSENSE was selected as representative of a state-of-the-art method whose algorithm implementation code is made available by the authors [12]. This allowed the comparison of the processing time running the algorithms in exactly the same conditions.

Only one set of parameters was used for all the videos. We set T_a=T_b=12, T_L=35, d_{th}=5, α_{FG}=0.0005, α_{BG}=0.001, α_{UBG}=0.01 and d_α=0.0005. The post-processing filter dimension $N \times N$ was set to 11×11. The maximum number of Gaussians allowed in the GMM was set to 5. These default values were determined empirically and worked well for many different scenarios as demonstrated by the results obtained with videos that incorporate a wide range of challenges.

The F-measure was used as an indicator of performance since, as reported in [3], it correlates more strongly with the rankings produced by evaluation algorithms. The processing time was used as a measure of complexity (it does not include image I/O operations). In real time applications, this feature can be critical. All algorithms were run in exactly the same conditions, using an Intel Core i7 2 GHz processor with 16 GB 1333 MHz DDR3 and OS X Yosemite 10.10.5. Our code has no low-level optimization.

5 Results and Discussion

The exhaustive set of experiments performed allowed us to assess the performance of BMOG. Fig. 2 shows the average F-measure for each video category, and across all categories for the overall set of videos, as reported in CDnet site [10], for MOG2 algorithm (GMM|Zivkovic in CDnet), the proposed BMOG method, RMoG, and SuBSENSE. It is clear that BMOG consistently outperforms MOG2, approaching SuBSENSE. RMoG slightly outperforms BMOG only for IOM and PTZ scenarios. For most categories, BMOG approaches SuBSENSE, but with a much faster solution. Mind that the best state-of-the-art methods, like SuBSENSE, are complex algorithms combining different approaches.

It must be highlighted that for 18 of the 53 videos (approximatelly 34%), belonging to 9 different categories, BMOG outperforms SuBSENSE. Only in two categories, PTZ and TH, BMOG does not perform better than SuBSENSE in any of the videos.

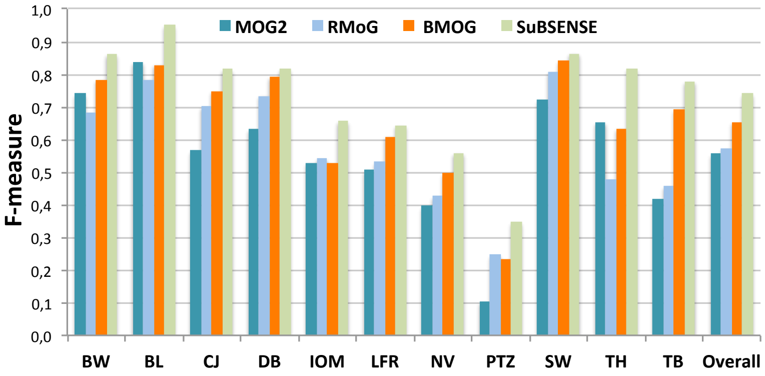


Fig. 2. Average F-measure for each category and across all categories.

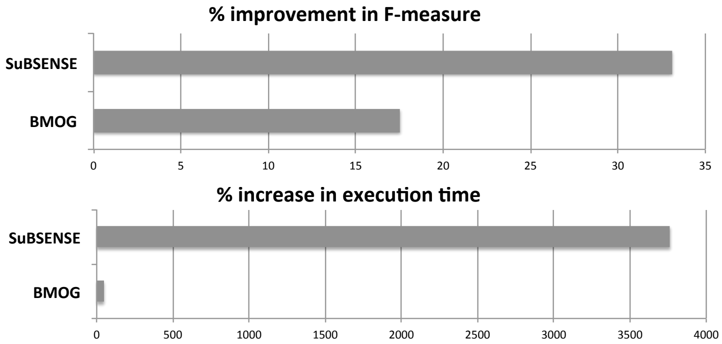


Fig. 3. Comparison of improvement in performance and increase in complexity for BMOG and a top-ranked method (reference: 0=MOG2).

The charts in Fig. 3 show that the proposed method, BMOG, achieves an excellent compromise in performance versus complexity when compared to MOG2 and SuBSENSE.

A comparative example of foreground masks obtained with BMOG, MOG2, RMoG and SuBSENSE is illustrated in Figure 4. This picture shows the original frame (Input), the ground truth (GT), and the segmentation masks for BMOG, MOG2, RMoG and SuBSENSE, with the pixels that are not labeled *Static* (BG) or *Moving* (FG) marked in gray. From left to right it pictures: frame 1138 of video highway (BL); frame 1389 of video bridgeEntry (NV); frame 3190 of video copyMachine (SW); and frame 2412 of video overpass (DB). These masks are all available at the CDnet site [10].

Results demonstrate that the proposed method, BMOG, achieves a good performance in a wide range of scenarios, both for indoor or outdoor scenes either in daytime or nighttime. Even in very difficult scenarios such as nighttime videos, one of the most difficult categories in the CDnet dataset [3], consisting of traffic scenes captured at night. In these videos, the main challenge is to deal with low-visibility of vehicles and their very strong headlights that cause halos and reflections on the street. BMOG shows very interesting results, outperforming SuBSENSE in 4 of the 6 nighttime videos.

6 Conclusion

This paper proposes a computationally efficient boosted GMM method, BMOG, for the unsupervised segmentation of moving objects in video sequences that proves to be more robust than a widely used approach (MOG2) and more recent proposals like RMoG. The choice of the color space, the decision criteria and a classification mechanism with hysteresis along with a dynamic learning rate for the background model update, proved to boost the overall detection accuracy while keeping complexity low, making it a good choice for real-time applications. It must be highlighted that, for each method, all the experiments in all the testing scenarios were run with the same set of parameters.

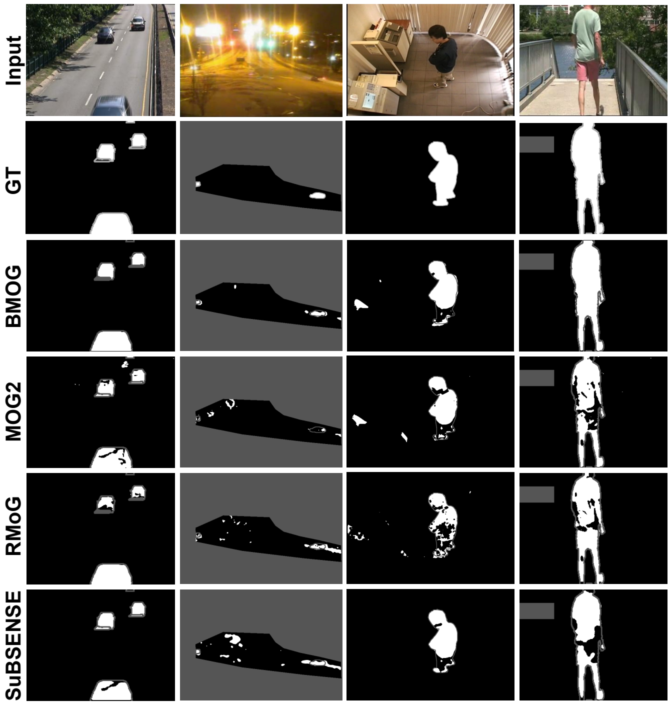


Fig. 4. Example of foreground masks for BMOG, MOG2, RMoG and SuBSENSE.

Remark The authors are willing to share the code if requested by e-mail and, in the meantime, the code will be prepared to be submitted to OpenCV as a candidate for inclusion in future updates of the library.

Acknowledgments Work supported by the Galician Regional Government under agreement for funding the Atlantic Research Center for Information and Communication Technologies (AtlantTIC) and research contract GRC2014/024 (Modalidade: Grupos de Referencia Competitiva 2014) and project "TEC4Growth - Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/ NORTE-01-0145-FEDER-000020", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

1. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: An overview. Comput. Sci. Review, 11, 31–66 (2014)
2. Sobral, A., and Vacavant, A.: A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput. Vis. Image Underst. 122, 4–21 (2014)
3. Wang, Y., Jodoin, P.-M., Porikli, F., Konrad, J., Benezeth, Y., Ishwar, P.: CDnet 2014: An Expanded Change Detection Benchmark Dataset. In Proc. IEEE Workshop on Change Detection (CDW-2014) at CVPRW-2014, pp. 393–400 (2014)
4. Stauffer, C., Grimson, E.: Adaptive background mixture models for real-time tracking. In Proc. IEEE Int. Comput. Soc. Conf. Comput. Vis. Patt. Recogn., vol. 2, pp. 246–252 (1999)
5. Zivkovic, Z., van der Heijden, F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. Patt. Recogn. Lett. 27 (7), 773–780 (2006)
6. Varadarajan, S., Miller, P., Zhou, H.: Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. Patt. Recogn. 8 (11), 3488–3503 (2015)
7. Lissner, I., Urban, P.: Toward a unified color space for perception-based image processing. IEEE Trans. Image Process. 21 (3), 115–1168 (2012)
8. Balcilar, M., Amasyali, M. F., Sonmez, A. C.: Moving object detection using lab2000hl color space with spatial and temporal smoothing. Appl. Math. Inf. Sci. 8 (4), 1755–1766 (2014)
9. Cucchiara, R.,Grana, C., Piccardi, M., Prati, A.: Detecting Moving Objects, Ghosts, and Shadows in Video Streams. IEEE Trans. Pattern Anal. Mach. Intell. 25 (10), 1337–1342 (2003)
10. ChangeDetection.NET, http://www.changedetection.net, accessed Nov. 2016
11. St-Charles, P.-L., Bilodeau, G.-A., Bergevin, R.: SuBSENSE : A Universal Change Detection Method with Local Adaptive Sensitivity. IEEE Trans. Image Process. 24 (1), 359–373 (2015)
12. SuBSENSE, https://bitbucket.org/pierre_luc_st_charles/subsense, accessed May 2016