

# GMM Background Modeling Using Divergence-Based Weight Updating

Juan D. Pulgarin-Giraldo<sup>1(✉)</sup>, Andres Alvarez-Meza<sup>1</sup>,  
David Insuasti-Ceballos<sup>1</sup>, Thierry Bouwmans<sup>2</sup>,  
and German Castellanos-Dominguez<sup>1</sup>

<sup>1</sup> Signal Processing and Recognition Group,  
Universidad Nacional de Colombia, Manizales, Colombia  
jdpulgarin@unal.edu.co

<sup>2</sup> Laboratoire MIA - Université de La Rochelle, La Rochelle, France

**Abstract.** Background modeling is a core task of video-based surveillance systems used to facilitate the online analysis of real-world scenes. Nowadays, GMM-based background modeling approaches are widely used, and several versions have been proposed to improve the original one proposed by Stauffer and Grimson. Nonetheless, the cost function employed to update the GMM weight parameters has not received major changes and is still set by means of a single binary reference, which mostly leads to noisy foreground masks when the ownership of a pixel to the background model is uncertain. To cope with this issue, we propose a cost function based on Euclidean divergence, providing nonlinear smoothness to the background modeling process. Achieved results over well-known datasets show that the proposed cost function supports the foreground/background discrimination, reducing the number of false positives, especially, in highly dynamical scenarios.

**Keywords:** Background modeling · GMM · Euclidean divergence

## 1 Introduction

Intelligent video surveillance systems, devoted to detect and track moving objects, can accomplish unsupervised results using background modeling methodologies, where a representation of the background is estimated and the regions that diverge from it are subtracted and labeled as moving objects named foreground [1]. Afterwards, surveillance systems interpret the activities and behaviors from the foreground objects to support computer vision analysis (e.g. object classification, tracking, activity understanding, among others) [2]. To achieve proper results, background modeling approaches focus on the elaboration of a background model, which suitably represents the pixel dynamics from real-world scenarios [3]. Among developed background modeling approaches, the most used are the ones derived from the conventional pixel-wise Gaussian Mixture Models (GMM) since they provide a trade-off between robustness to real-world video conditions and computational burden [4]. To date, several adapted

versions have been proposed. In fact, Authors in [5] provide a survey and propose the classification of the most salient GMM-based background modeling approaches.

Regarding the updating rules of the GMM parameters, improvements have been mainly reported related to the learning rate parameter, which aims to adapt the background model by observing the pixel dynamics through time. Originally, from the derivation of the GMM parameter updating rules, a Gaussian kernel term is attained, providing smoothness to the updating rules of the mean and variance parameters, nonetheless, the cost function of the weights is set using a binary ownership value. This updating rule may lead to noisy foreground masks, especially, when the pixel labels are uncertain like in dynamical scenarios. Zivkovic et al. proposed an improvement for the original GMM, which uses Dirchilet priors to update the weight parameters. Nonetheless, this improvement was mainly made to decrease the computational cost and the foreground/background discrimination performance remains similar to the original GMM.

Here, we propose a new cost function for the GMM weights updating. Using Euclidean divergence (ED), we compare the instant and cumulative probabilities of each Gaussian GMM model fitting the pixel input samples. Then, employing Least Mean Squares (LMS), we minimize the ED of obtained probabilities to adjust the weights values through time. By doing so, we provide non-linear smoothness to the whole GMM parameter updating rules, reducing the number of false positives in the obtained foreground masks. The proposed cost function is coupled into the traditional GMM approach, producing a new background modeling approach named ED-GMM, which improves the foreground/background discrimination in the case of real-world scenarios, especially in dynamical environments.

## 2 Methods

*Background Modeling Based on GMM:* The probability that a query input pixel  $\mathbf{x}_t \in \mathbb{R}^C$  ( $C \in \mathbb{N}$  is the color space dimension), at time  $t \in T$ , belongs to a given GMM-based background model is as:

$$p(\mathbf{x}_t | \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) = \sum_{m=1}^M w_{m,t} \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{m,t}, \boldsymbol{\Sigma}_{m,t}), \quad (1)$$

where  $M \in \mathbb{N}$  is the number of Gaussian models of the GMM,  $w_{m,t} \in \mathbb{R}$  the weight related to the  $m$ -th Gaussian model,  $\mathcal{N}\{\cdot, \cdot\}$  with mean value,  $\boldsymbol{\mu}_{m,t} \in \mathbb{R}^{1 \times C}$ , and covariance matrix  $\boldsymbol{\Sigma}_{m,t} \in \mathbb{R}^{C \times C}$ . For computational burden alleviation, all elements of the color representation set are assumed as independent and having the same variance value [4]:  $\sigma_{m,t}^2 \in \mathbb{R}^+$ ,  $\forall m$ , that is,  $\boldsymbol{\Sigma}_{m,t} = \sigma_{m,t}^2 \mathbf{I}$ , being  $\mathbf{I} \in \mathbb{R}^{C \times C}$  the identity matrix. Afterwards, each query pixel,  $\mathbf{x}_t$ , is evaluated until it matches a Gaussian model of the GMM. Here, the match occurs whenever a pixel value ranges within 2.5 standard-deviation interval of

the Gaussian model. However, if  $\mathbf{x}_t$  does not match any Gaussian model, the least probable model is replaced by a new one having low initial weight, large initial variance, and mean  $\boldsymbol{\mu}_{m,t} = \mathbf{x}_t$  [4]. In the positive case that the  $m$ -th model matches a new input pixel, its parameters are updated as follows:

$$w_{m,t} = w_{m,t-1} + \alpha(o_{m,t} - w_{m,t-1}) \quad (2a)$$

$$\boldsymbol{\mu}_{m,t} = \boldsymbol{\mu}_{m,t-1} + \rho_{m,t} o_{m,t} (\mathbf{x}_t - \boldsymbol{\mu}_{m,t}) \quad (2b)$$

$$\sigma_{m,t}^2 = \sigma_{m,t-1}^2 + \rho_{m,t} o_{m,t} ((\mathbf{x}_t - \boldsymbol{\mu}_{m,t})^\top (\mathbf{x}_t - \boldsymbol{\mu}_{m,t}) - \sigma_{m,t-1}^2) \quad (2c)$$

where  $\alpha \in \mathbb{R}^+$  is the weight learning rate,  $o_t \in \{0, 1\}$  is a binary number indicating the membership of a sample to a model, and  $\rho_{m,t} \in \mathbb{R}^+$  is the mean and variance learning rate set as a version of the  $\alpha$  parameter smoothed by the Gaussian kernel  $g(\mathbf{x}_t; \cdot, \cdot)$ , i.e.:  $\rho_{m,t} = \alpha g(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t})$ . Lastly, the derived models are ranked according to the ratio  $w/\sigma$  to determine the most likely produced by the background, making suitable the further foreground/background discrimination [6].

*Enhanced GMM-Based Background Using Euclidean Divergence (ED-GMM):* The updating rules of the GMM parameters, used in Eq. (2), can be derived within the conventional Least Mean Square (LMS) formulation framework as follows:

$$\theta_t = \theta_{t-1} - \eta_{\theta_{t-1}} \partial_{\theta_{t-1}} \{\varepsilon_{\theta_{t-1}}^2\}, \quad (3)$$

where  $\theta_t \in \{w_t, \boldsymbol{\mu}_t, \sigma_t\}$  is each one of the estimated parameters by the corresponding learning rate:

$$\eta_{\theta_{t-1}} \in \begin{cases} \eta_{w_{t-1}} = \alpha/2 \\ \eta_{\boldsymbol{\mu}_{t-1}} = \alpha \sigma_{t-1}^2 / 2 \\ \eta_{\sigma_{t-1}} = \alpha g(\mathbf{x}_t; \boldsymbol{\mu}_t, \sigma_{m,t}) / 2 \end{cases} \quad (4)$$

and the following cost functions, respectively:

$$\varepsilon_{\theta_{t-1}} \in \begin{cases} \varepsilon_{w_{t-1}} = o_t - w_{t-1} \\ \varepsilon_{\boldsymbol{\mu}_{t-1}} = g(\mathbf{x}_t; \boldsymbol{\mu}_t, \sigma_{m,t}) \\ \varepsilon_{\sigma_{t-1}} = |\mathbf{x}_t - \boldsymbol{\mu}_t|^2 - \sigma_{t-1}^2 \end{cases} \quad (5)$$

It is worth noting that the  $\boldsymbol{\mu}_t$  and  $\sigma_t$  updating rules, grounded on kernel similarities  $g(\mathbf{x}_t; \cdot, \cdot)$  (Eqs. (4) and (5)), provide smoothness to encode the uncertainty of a pixel belonging whether to the background or foreground. In contrast, the cost function of the weights is set using a binary reference (i.e., membership  $o_t$ ). This updating rule may lead to noisy foreground masks when the ownership value is uncertain, especially, in environments holding dynamical sources like trees waving, water flowing, snow falling, etc. To cope with this, we propose to set the cost function of  $w_t$  using the ED as follows:

$$\varepsilon_{w_{t-1}} = \bar{g}(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t}) - w_{t-1}. \quad (6)$$

The ED allows measuring the difference between two probabilities [7]. So, back into the LMS scheme we aim to minimize the ED between an instant probability determined by the Gaussian kernel  $\bar{g}(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t})$  and a cumulative probability encoded by  $w_{t-1}$ . This is grounded by the fact that, if a model has a high cumulative probability  $w_{t-1}$ , means that such model has suitably adapted to the pixel dynamics through time, then,  $\bar{g}(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t})$  should have a high value too. Since the difference between both is expected to be low, the following updating rule is introduced:

$$w_{m,t} = w_{m,t-1} + \alpha(\bar{g}(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t}) - w_{m,t-1}) \quad (7)$$

where the kernel term,  $\bar{g}(\mathbf{x}_t; \boldsymbol{\mu}_{m,t}, \sigma_{m,t}) = \mathbb{E} \{g(\mathbf{x}_t^c; \boldsymbol{\mu}_{m,t}^c, \sigma_{m,t}) : \forall c \in C\}$ , measures the average similarity along color channels. Also, since we aim to incorporate the information about each new input sample into all the GMM Gaussian models, we exclude the ownership  $o_t$  from the weight updating.

### 3 Experimental Set-Up

Aiming to validate the proposed cost function, the ED-GMM approach is compared against the traditional GMM (GMM1) and the Zivkovic GMM proposed (ZGMM) in [8], which uses Dirichlet priors into the weight updating rules to automatically set the number of Gaussians  $M$ . The following three experiments are performed: (i) Visual inspection of the temporal weight evolution to make clear performance of background model and the foreground/background discrimination through time. (ii) Foreground/background discrimination over a wide variety of real-world videos that hold ground-truth sets. (iii) Robustness against variations of the learning rate parameter in foreground/background discrimination tasks. The following datasets are employed for the experiments:

*DBa- Change Detection:* (at <http://www.changedetection.net/>) Holds 31 video sequences of indoor and outdoor environments, where spatial and temporal regions of interest are provided. Ground-truth labels are background, hard shadow, outside region of interest, unknown motion, and foreground.

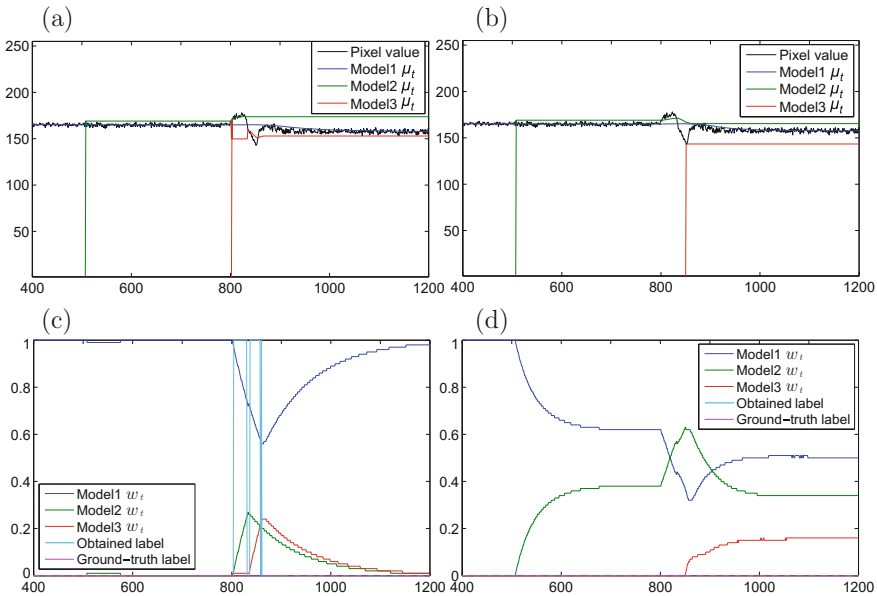
*DBb- A-Star-Perception:* (at <http://perception.i2r.a-star.edu.sg>) Recorded in both indoor and outdoor scenarios, contains nine image sequences with different resolution. The ground-truths are available for random frames in the sequence and hold two labels: background and foreground.

*Measures:* The foreground/background discrimination is assessed only for two ground-truth labels (*foreground* and *background*) by supervised pixel-based measures: Recall,  $r = t_p / (t_p + f_n)$ , Precision,  $p = t_p / (t_p + f_p)$ , and  $F_1 = 2pr / (p + r)$ . Here,  $t_p$  is the number of true positives,  $f_p$  is the false positives, and  $f_n$  is the false negatives. These values are obtained comparing against the ground-truth. Measures range within  $[0,1]$ , where the higher the attained measure – the better the achieved segmentation.

*Implementation and Parameter Tuning:* The ED-GMM algorithm is developed using as basis the C++ BGS library [9]. Parameters are left as default for all the experiments except for task three requiring to vary the learning rate  $\alpha$ . We set three mixing models  $M = 3$  (noted as *Model1*, *Model2*, and *Model3*). The GMM1 and ZGMM algorithms are also taken from the BGS library.

### 4 Results and Discussion

*Temporal Analysis:* We conduct a visual inspection of the temporal evolution of estimated parameters to make clear the contribution of the proposed weight cost function. Testing is carried out on the video *DBa-snowFall* for which a single pixel in the red color channel is tracked as seen in Fig. 1, showing temporal evolution of  $\mu_t$  (top row) and  $w_t$  (bottom row). Also, the inferred foreground/background labels and the ground-truth are shown in subplots Fig. 1(c) and (d) ('1': foreground and '0': background). It can be observed that the estimated  $\mu_t$  parameter by either GMM1 (see subplot 1(a)) or ED-GMM (subplot 1(b)) is similar for the three considered mixing models. However, the weights estimated by ED-GMM are updated better along time. Particularly, the ED-GMM weight increases around the 500th frame, where the *Model2* (in green) is generated (see subplot 1(d)). Then, the model properly reacts to the pixel change occurring close to the 800th



**Fig. 1.** Temporal evolution of  $w_t$  and  $\mu_t$  for pixel (150,250,1) from DBa-snowFall video. (a) GMM1  $\mu_t$ , (b) ED-GMM  $\mu_t$ , (c) GMM1  $w_t$  and label, (d) ED-GMM  $w_t$  and label (Color figure online)

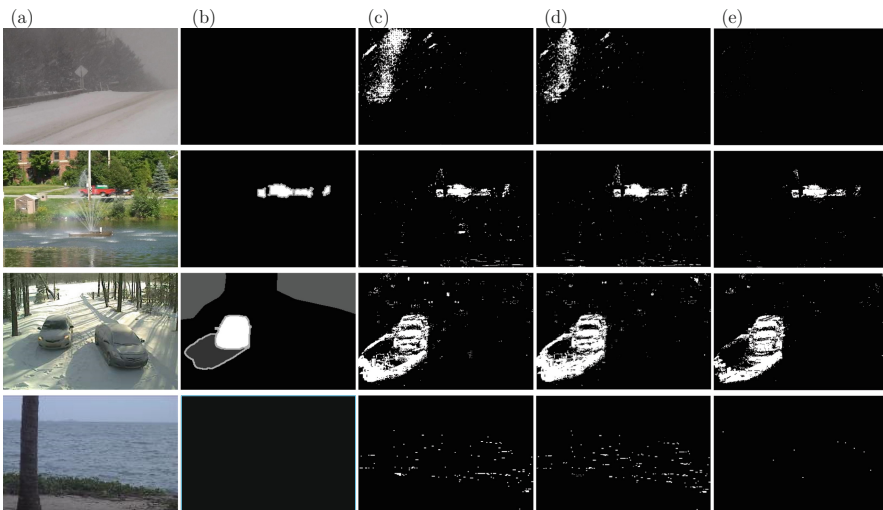
frame, obtaining labels corresponding to the ground-truth (background). In contrast, the GMM1 updating rule makes the  $w_t$  weight remain almost zero even if the *Model2* gets very close to the pixel value (see subplot 1(c)). As a consequence, this strategy infers wrongly foreground labels.

*Performance of the Foreground/Background Discrimination Task:* Aiming to check for the generalizing ability of the ED-GMM method, we test 25 videos embracing a wide variety of dynamics. The videos are grouped into two categories *a* and *b*. The former holds videos where the background is mostly static and the latter videos where the background exhibit highly dynamical variations. The total average seen in Table 1 shows that the ED-GMM reaches higher precision

**Table 1.** Foreground discrimination performance

	Video	GMM1			ZGMM			ED-GMM		
		$r$	$p$	$F_1$	$r$	$p$	$F_1$	$r$	$p$	$F_1$
Category <i>a</i>	DBa-abandonedBox	0.34	0.43	0.38	0.32	0.46	0.37	0.35	<b>0.60</b>	<b>0.44</b>
	DBb-bootstrap	<b>0.59</b>	0.56	0.57	0.55	0.60	0.58	0.55	0.61	0.58
	DBa-corridor	0.60	0.74	0.66	0.58	0.79	0.67	0.61	<b>0.88</b>	<b>0.72</b>
	DBa-diningRoom	0.45	0.88	0.60	<b>0.57</b>	0.94	<b>0.71</b>	0.45	0.93	0.61
	DBb-hall	0.51	0.65	0.57	<b>0.54</b>	<b>0.73</b>	<b>0.62</b>	0.51	0.69	0.59
	DBa-highway	0.90	0.90	0.90	0.82	0.91	0.86	0.89	0.93	0.91
	DBa-library	0.17	0.91	0.29	0.19	0.94	0.32	0.19	0.93	0.32
	DBa-office	0.29	0.63	0.40	0.28	0.63	0.38	<b>0.31</b>	<b>0.71</b>	<b>0.43</b>
	DBa-park	0.65	0.70	0.67	0.64	0.73	0.68	0.53	<b>0.80</b>	0.64
	DBa-parking	0.27	0.78	0.40	0.17	0.69	0.28	0.27	<b>0.84</b>	0.41
	DBa-pedestrians	0.94	0.51	0.66	0.93	0.46	0.61	0.90	<b>0.71</b>	<b>0.79</b>
	DBa-pets2006	<b>0.69</b>	0.54	0.61	0.67	0.57	0.62	0.63	<b>0.74</b>	<b>0.68</b>
	DBb-shoppingMall	0.64	0.53	<b>0.58</b>	0.63	0.48	0.55	0.54	<b>0.58</b>	0.56
	DBa-sofa	0.42	0.65	0.51	<b>0.46</b>	0.64	<b>0.54</b>	0.35	<b>0.84</b>	0.49
	DBa-streetLight	0.17	0.48	0.25	<b>0.23</b>	0.52	<b>0.32</b>	0.18	<b>0.72</b>	0.29
DBa-tramStop	0.37	0.87	0.52	0.38	0.87	0.53	0.33	0.88	0.48	
<i>Average</i>	0.50	0.67	0.53	0.50	0.69	0.54	0.47	<b>0.77</b>	<b>0.56</b>	
Category <i>b</i>	DBa-boats	0.43	0.24	0.31	<b>0.46</b>	0.18	0.26	0.43	<b>0.27</b>	<b>0.33</b>
	DBa-canoe	0.49	0.31	0.38	0.53	0.32	0.40	<b>0.63</b>	<b>0.35</b>	<b>0.45</b>
	DBa-fountain02	0.86	0.33	0.48	0.85	0.28	0.42	0.82	<b>0.52</b>	<b>0.64</b>
	DBa-overpass	0.73	0.76	0.74	<b>0.80</b>	0.77	0.78	0.78	<b>0.83</b>	0.80
	DBa-skating	0.76	0.80	0.78	0.77	0.77	0.77	0.78	<b>0.89</b>	<b>0.83</b>
	DBa-snowFall	0.73	0.35	0.47	0.71	0.48	0.58	0.66	<b>0.65</b>	<b>0.65</b>
	DBb-waterSurface	0.55	0.74	0.63	0.52	0.71	0.60	<b>0.61</b>	<b>0.79</b>	<b>0.69</b>
	DBa-wetSnow	<b>0.64</b>	0.76	0.69	0.58	0.80	0.67	0.59	0.81	0.68
	DBa-winterDriveway	0.59	0.40	0.48	0.58	0.43	0.49	0.60	<b>0.52</b>	<b>0.55</b>
	<i>Average</i>	0.64	0.52	0.55	0.64	0.53	0.56	0.66	<b>0.63</b>	<b>0.62</b>
Total average		0.55	0.62	0.58	0.55	0.63	0.59	0.54	<b>0.72</b>	<b>0.62</b>

during the discrimination of the foreground/background labels, decreasing the amount of false positives. This fact is explicable since the proposed weight updating rule (see Eq. (7)) allows the ED-GMM models to adapt faster to changes of the pixel dynamics. The above can be even more remarked for videos with dynamical background sources as seen in the *Category b* in which the precision is improved by 10% comparing against the other two methods. By the other hand, GMM1 and ZGMM attain very similar results, since the main proposal of Zivkovic was focused to reduce computational cost. As a result, the foreground masks attained by ED-GMM have less false positives and are more similar to the ground truth masks as seen in Fig. 2 showing concrete scenarios with highly dynamical background sources relating to snow falling (DBa-snowFall, DBa-winterDriveway) and water flowing (DBa-fountain02, DBb-waterSurface).

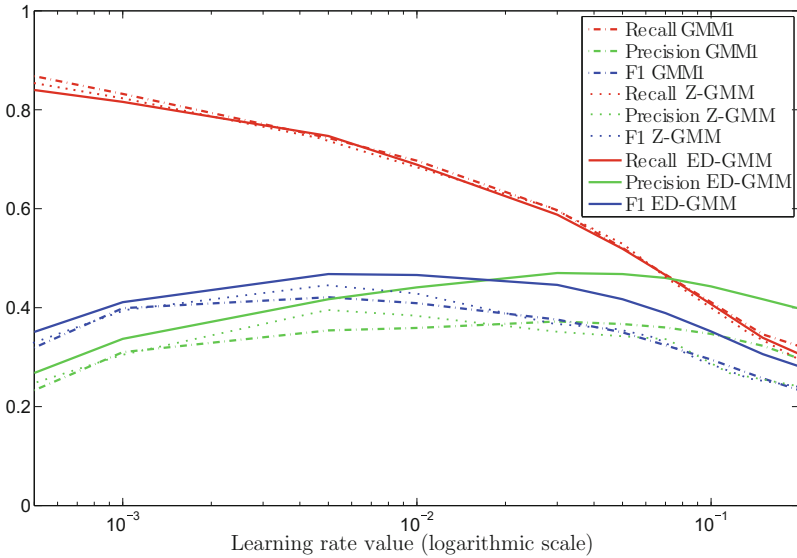


**Fig. 2.** Foreground masks of interest. (a) Original frame, (b) Ground-truth, (c) GMM1, (d) ZGMM, (e) ED-GMM.

*Robustness Against Variation of the Learning Rate Parameter:* The influence of the learning rate variation on the foreground/background discrimination is assessed through supervised measures, which are estimated from the videos of Category a: DBa-highway, DBa-office, DBa-pedestrians and DBa-pets2006 and Category b: DBa-boats, DBa-canoe, DBa-fountain02 and DBa-overpass.

Figure 3 shows the obtained supervised measures, averaged over the 10 videos, where the  $x$  axis is the logarithm of the employed  $\alpha$  rate ranging within:  $\{0.0005, 0.001, 0.005, 0.01, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2\}$ . It can be seen that the proposed ED-GMM (continuous lines) behaves similar as the traditional GMM1 method (dashed lines) and ZGMM (pointed lines). However, the obtained Precision and  $F1$  measures are everywhere higher than the ones reached by GMM1 and

ZGMM. An interval of confidence is found within the interval  $\alpha \in 0.005 - 0.01$ , where the highest  $F1$  measure is reached.



**Fig. 3.** Supervised measures changing the learning rate value for GMM1, Z-GMM and ED-GMM.

## 5 Conclusions

We propose a cost function for the GMM weights updating using Euclidean divergence. The proposed cost function is coupled into the traditional GMM, producing a new approach named ED-GMM used to support the background modeling task for videos recorded in highly dynamical scenarios. The Euclidean divergence allows comparing the instant and cumulative probabilities of a GMM model fitting the pixel input samples. Then, employing LMS, we minimize the Euclidean divergence of such probabilities to adjust the weights values through time. Carried out experiments show that ED-GMM reduces the amount of false positives in the obtained foreground masks comparing against traditional GMM and Zivkovic GMM, especially, for videos holding dynamical background sources: water flowing, snow falling and trees waving. Additionally, the proposed cost function demonstrated to be robust when varying the learning rate parameter value, always achieving better results than traditional GMM. Consequently, the proposed cost function can be coupled into more complex GMM based background modeling approaches to improve the foreground/background discrimination. As future work, authors plan to test the proposed cost function



using selective updating strategies to improve the discrimination in scenarios holding motionless foreground objects.

**Acknowledgment.** This work was developed in the framework of the research project entitled “*Caracterización de cultivos agrícolas mediante estrategias de teledetección y técnicas de procesamiento de imágenes*” (36719) under the grants of “*Convocatoria conjunta para el fomento de la investigación aplicada y desarrollo tecnológico*” 2016, as well as by program “*Doctorados Nacionales convocatoria 647 de 2014*” funded by COLCIENCIAS and partial Ph.D. financial support from Universidad Autónoma de Occidente.

## References

1. Molina-Giraldo, S., Álvarez-Meza, A.M., García-Álvarez, J.C., Castellanos-Domínguez, C.G.: Video segmentation framework by dynamic background modelling. In: Petrosino, A. (ed.) ICIAP 2013. LNCS, vol. 8156, pp. 843–852. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-41181-6\\_85](https://doi.org/10.1007/978-3-642-41181-6_85)
2. Maddalena, L., Petrosino, A.: A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Trans. Image Process.* **17**(7), 1168–1177 (2008)
3. Alvarez-Meza, A.M., Molina-Giraldo, S., Castellanos-Dominguez, G.: Background modeling using object-based selective updating and correntropy adaptation. *Image Vis. Comput.* **45**, 22–36 (2016)
4. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE (1999)
5. Bouwmans, T., El Baf, F., Vachon, B., et al.: Background modeling using mixture of Gaussians for foreground detection - a survey. *Recent Pat. Comput. Sci.* **1**(3), 219–237 (2008)
6. Hayman, E., Eklundh, J.-O.: Statistical background subtraction for a mobile observer. In: Proceedings of Ninth IEEE International Conference on Computer Vision, vol. 1, pp. 67–74, October 2003
7. Principe, J.: *Information Theoretic Learning. Renyi’s Entropy and Kernel Perspectives*. Springer, New York (2010)
8. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2, pp. 28–31, August 2004
9. Sobral, A.: BGSLibrary: an OpenCV C++ background subtraction library. In: IX Workshop de Visão Computacional (WVC 2013), Rio de Janeiro, Brazil, pp. 38–43, June 2013