



Bayesian Learning of Infinite Asymmetric Gaussian Mixture Models for Background Subtraction

Ziyang Song, Samr Ali^(✉), and Nizar Bouguila

Concordia University, Montreal, Canada

{zi_on,al_samr}@encs.concordia.ca, nizar.bouguila@concordia.ca

Abstract. Background subtraction plays an important role in many video-based applications such as video surveillance and object detection. As such, it has drawn much attention in the computer vision research community. Utilizing a Gaussian mixture model (GMM) has especially shown merit in solving this problem. However, a GMM is not ideal for modeling asymmetrical data. Another challenge we face when applying mixture models is the correct identification of the right number of mixture components to model the data at hand. Hence, in this paper, we propose a new infinite mathematical model based on asymmetric Gaussian mixture models. We also present a novel background subtraction approach based on the proposed infinite asymmetric Gaussian mixture (IAGM) model with a non-parametric learning algorithm. We test our proposed model on the challenging Change Detection dataset. Our evaluations show comparable to superior results with other methods in the literature.

Keywords: Infinite asymmetric gaussian mixture model · Gibbs sampling · MCMC · Metropolis-Hastings · Background subtraction

1 Introduction

Background subtraction is an active area of research in computer vision [1,2]. It is performed by using a reference image that contains only the background without any moving objects then subtracting the new frame from this image and thresholding to obtain the foreground [3]. Hence, it mainly involves the automatic segmentation of foreground regions in video sequences from the background [4]. This has many consequent applications such as in video surveillance [5], object detection [6], and anomaly detection [7].

Gaussian mixture models (GMM) is a commonly deployed approach that has been proposed for statistical modeling of foreground [8,9]. However, GMMs are not necessarily the ideal model since the background and foreground pixels are not always distributed according to a Gaussian distribution [10]. Hence, we

investigate the employment of the asymmetric Gaussian mixture (AGM) which has two variance parameters for the left and right sides of each of the mixture components [11].

Parameter learning is one of the challenges required for the use of mixture models. A variety of algorithms may be deployed to achieve this purpose. For instance, the expectation maximization algorithm for the maximum likelihood method is one of the famous learning approaches [12]. Nonetheless, this deterministic approach suffers from several drawbacks that include over-fitting and high dependency on initialization [8,9]. This compromises the efficiency of the learning algorithm and negatively impacts the accuracy of the model. The author in [13] recently studied Bayesian learning of AGM. Such sampling-based approaches include the Markov Chain Monte Carlo (MCMC) introduces prior and posterior distributions in order to address the over-fitting issues whereby the dependency between the mixture parameters and mixture components is eradicated [14,15].

In this paper, we propose a novel mathematical model by extending the AGM to the infinity [16]. This addresses another important challenge that arises when applying mixture models; choosing the correct number of mixture components. We use the Dirichlet process [17,18] for precise allocation of the observations in and determining the number of components for the proposed model, the infinite asymmetric Gaussian mixture (IAGM). We apply a hierarchical Bayesian learning approach for the proposed model.

The contributions of this paper are then summarized as: (i) proposal and derivation of the IAGM model; (ii) hierarchical Bayesian learning of the IAGM model; (iii) testing and evaluation of the IAGM model for the background subtraction application. The rest of this paper is organized as follows. In Sect. 2, we outline the IAGM model and present a complete learning algorithm. We evaluate our proposed model for the background subtraction application and compare it to three other models in Sect. 3. Finally, we conclude the paper in Sect. 4.

2 Infinite Asymmetric Gaussian Mixture Model

2.1 Mathematical Model

A finite AGM model is denoted by:

$$p(\mathcal{X} | \theta) = \prod_{i=1}^N \sum_{j=1}^M p_j p(X_i | \xi_j) \tag{1}$$

where $\mathcal{X} = (X_1, \dots, X_N)$ is the N observations dataset, each observation $X_i = (X_{i1}, \dots, X_{iD})$ could be represented as D -dimensional random variable. $M \geq 1$ is the number of mixture components, $\theta = (p_1, \dots, p_M, \xi_1, \dots, \xi_M)$ defines the complete set of parameters fully characterizing the mixture model where $\vec{p} = (p_1, \dots, p_M)$ is the mixing weights which must be positive and sum to one, and ξ_j is the set of parameters of mixture component j .

The asymmetric Gaussian density for each component, $p(X_i | \xi_j)$, is then given by:

$$p(X_i | \xi_j) \propto \prod_{k=1}^D \frac{1}{(S_{l_{jk}})^{-\frac{1}{2}} + (S_{r_{jk}})^{-\frac{1}{2}}} \times \begin{cases} \exp \left[-\frac{S_{l_{jk}}(X_{ik} - \mu_{jk})^2}{2} \right] & \text{if } X_{ik} < \mu_{jk} \\ \exp \left[-\frac{S_{r_{jk}}(X_{ik} - \mu_{jk})^2}{2} \right] & \text{if } X_{ik} \geq \mu_{jk} \end{cases} \quad (2)$$

where $\xi_j = (\mu_j, S_{l_j}, S_{r_j})$ is the parameter set for the asymmetric Gaussian distribution with $\mu_j = (\mu_{j1}, \dots, \mu_{jd})$, $S_{l_j} = (S_{l_{j1}}, \dots, S_{l_{jd}})$ and $S_{r_j} = (S_{r_{j1}}, \dots, S_{r_{jd}})$. μ_{jk} , $S_{l_{jk}}$ and $S_{r_{jk}}$ are the mean, the left precision and the right precision of the k th dimensional distribution. In this paper, we assume independence so that the covariance matrix of X_i is diagonal matrix. This assumption allows us to avoid costly computation during deployment.

We introduce the latent indicator variable $Z = (Z_1, \dots, Z_N)$, Z_i for each observation X_i to indicate which component it belongs to. $Z_i = (Z_{i1}, \dots, Z_{iM})$ where hidden label Z_{ij} is assigned as 1 if X_i belongs to component j otherwise will be set to 0. The likelihood function is then defined by:

$$p(\mathcal{X} | Z, \Theta) = \prod_{i=1}^N p(X_i | \xi_j)^{Z_{ij}} \quad (3)$$

Given the mixing weights \vec{p} with $p_j = Z_{ij} = 1$, for $j = 1, \dots, M$, Z is given a Multinomial prior:

$$p(Z | \vec{p}) = Multi(\vec{p}) = \prod_{j=1}^M p_j^{n_j} \quad (4)$$

where n_j is the number of observations that are associated with component j . The mixing weights are considered to follow a symmetric Dirichlet distribution with a concentration parameter α/M :

$$p(\vec{p} | \alpha) \sim Dirichlet\left(\frac{\alpha}{M}, \dots, \frac{\alpha}{M}\right) = \frac{\Gamma(\alpha)}{\Gamma(\frac{\alpha}{M})^M} \prod_{j=1}^M p_j^{\frac{\alpha}{M} - 1} \quad (5)$$

We then integrate out the mixing weights \vec{p} to obtain the prior of Z :

$$p(Z | \alpha) = \int p(Z | \vec{p})p(\vec{p} | \alpha)d\vec{p} = \frac{\Gamma(\alpha)}{\Gamma(N + \alpha)} \prod_{j=1}^M \frac{\Gamma(\frac{\alpha}{M} + n_j)}{\Gamma(\frac{\alpha}{M})} \quad (6)$$

The conditional prior for a single indicator is then denoted by:

$$p(Z_{ij} = 1 | \alpha, Z_{-i}) = \frac{n_{-i,j} + \frac{\alpha}{M}}{N - 1 + \alpha} \quad (7)$$

Where the subscript $-i$ defines all indexes except i , $Z_{-i} = (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_N)$, $N_{-i,j}$ is the number of observations excluding X_i in component j .

Next, we extend the model to infinity by updating the posteriors of indicators in Eq. (7) with $M \rightarrow \infty$:

$$p(Z_{ij} = 1 \mid \alpha, Z_{-i}) = \begin{cases} \frac{n_{-i,j}}{N-1+\alpha}, & \text{if } n_{-i,j} > 0 \\ \frac{\alpha}{N-1+\alpha}, & \text{if } n_{-i,j} = 0 \end{cases} \quad (8)$$

where $n_{-i,j} > 0$ occurs only when component j is represented. Thus, an observation X_i is associated with an existing component by a certain probability proportional to the number of observations already allocated to this component; while a new (when unrepresented) component is proportional to α and N . Given the conditional priors in Eq. (7), the conditional posteriors are obtained by multiplying the priors with Eq. (3) resulting in:

$$p(Z_{ij} = 1 \mid \dots) = \begin{cases} \frac{n_{-i,j}}{N-1+\alpha} \prod_{k=1}^d p(X_{ik} \mid \xi_{jk}), & \text{if } n_{-i,j} > 0 \\ \frac{\alpha}{N-1+\alpha} \int p(X_i \mid \xi_j) p(\xi_j \mid \lambda, r, \beta_l, \beta_r, w_l, w_r) d\xi_j, & \text{if } n_{-i,j} = 0 \end{cases} \quad (9)$$

where the hyperparameter α is defined by an inverse Gamma prior with shape parameter a and mean b chosen as follows:

$$p(\alpha^{-1}) \propto \alpha^{-\frac{3}{2}} \exp(-\frac{1}{2\alpha}) \quad (10)$$

Given the likelihood of α in Eq. (6), the posterior is then:

$$p(\alpha \mid M, N) \propto \frac{\alpha^{\frac{M-3}{2}} \exp(-\frac{1}{2\alpha}) \Gamma(\alpha)}{\Gamma(N + \alpha)} \quad (11)$$

The conditional posterior for α depends only on number of observations, N , and the number of components, M . The logarithmic representation of posteriors is log-concave, so we can sample α by using the Adaptive Rejection Sampling (ARS) method [19].

2.2 Bayesian Learning

In this section, we describe an MCMC-based approach for learning the proposed IAGM model as shown in Fig. 1. The means of the component μ_{jk} are given Gaussian priors with hyperparameters λ and r as follows:

$$p(\mu_{jk} \mid \lambda, r) \sim \mathcal{N}(\lambda, r^{-1}). \quad (12)$$

where the mean, λ , and precision, r , hyperparameters are common to all components in a specific dimension. λ is given Gaussian priors with mean e and variance f , and r is given Gaussian priors and inverse Gamma priors with shape parameter g and mean parameter h respectively:

$$p(\lambda) \sim \mathcal{N}(e, f) \quad (13)$$

$$p(r) \sim \Gamma(g, h) \quad (14)$$

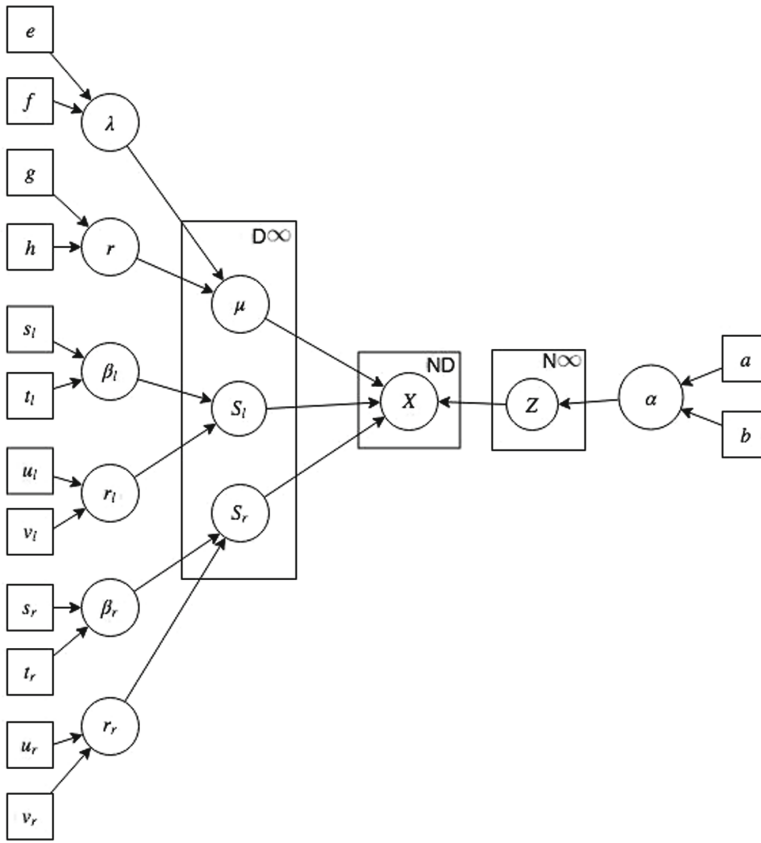


Fig. 1. Graphical model representation of the IAGM model. The random variables are in circles, and the model parameters in squares. The number mentioned in the right upper corner of the plates indicates the number of repetition of the contained random variables. The arcs describe the conditional dependencies between variables.

where e and f will be μ_y and σ_y^2 , the mean and variance of the observations which are used for the parameters of the Gaussian priors. The Gamma priors use constant values 1 as shape g and σ_y^2 as mean h to represent.

The conditional posterior distribution for the mean μ_{jk} is then computed by multiplying the likelihood from Eq. (3) by the prior Eq. (12) as follows:

$$p(\mu_{jk} | X_k, S_{l_{jk}}, S_{r_{jk}}, \lambda, r) \propto \mathcal{N}\left(\frac{S_{l_{jk}} \sum_{i: X_{ik} < \mu_{jk}} X_{ik} + s_{r_{jk}} \sum_{i: X_{ik} \geq \mu_{jk}} X_{ik} + r\lambda}{r + ps_{l_{jk}} + (n_j - p)s_{r_{jk}}}, \frac{1}{r + ps_{l_{jk}} + (n_j - p)s_{r_{jk}}}\right), \quad (15)$$

Where X_k is the k th dimensional observation allocated to component j . n_j is the count of X_k and p is the count of X_k which are less than μ_{jk} . $\sum_{i: X_{ik} < \mu_{jk}} X_{ik}$ and $\sum_{i: X_{ik} \geq \mu_{jk}} X_{ik}$ are the sums of the observations which are less than and greater

than μ_{jk} respectively. For the hyperparameters λ and r , we use hyperposteriors to update parameters. Equation (12) plays the role of the likelihood function. As such, we combine Eq. (12), Eq. (13) and Eq. (14) to obtain the following posteriors:

$$p(\lambda \mid \mu_{1k}, \dots, \mu_{Mk}, r) \propto \mathcal{N}\left(\frac{\mu_y \sigma_y^{-2} + r \sum_{j=1}^M \mu_{jk}}{\sigma_y^{-2} + Mr}, \frac{1}{\sigma_y^{-2} + Mr}\right) \quad (16)$$

$$p(r \mid \mu_{1k}, \dots, \mu_{Mk}, \lambda) \propto \Gamma(M + 1, \frac{M + 1}{\sigma_y^2 + \sum_{j=1}^M (\mu_{jk} - \lambda)^2}) \quad (17)$$

The component precision $S_{l_{jk}}$ and $S_{r_{jk}}$ are given Gamma priors with common hyperparameters β and w^{-1} as follows:

$$p(S_{l_{jk}} \mid \beta, w) \sim \Gamma(\beta, w^{-1}), \quad p(S_{r_{jk}} \mid \beta, w) \sim \Gamma(\beta, w^{-1}) \quad (18)$$

where β is given inverse Gamma priors with shape parameter s and mean parameter t , and w is given Gamma priors with shape parameters u and v :

$$p(\beta^{-1}) \sim \Gamma(s, t) \quad (19)$$

$$p(w) \sim \Gamma(u, v) \quad (20)$$

Where we set both of mean and shape parameters of hyperprior β as constant value 1, and mean and shape parameters of hyperprior w are defined as 1 and σ_y^2 respectively. The conditional posterior distribution for left precision $S_{l_{jk}}$ and right precision $S_{r_{jk}}$ are obtained by multiplying the likelihood from Eq. (3) by the prior Eq. (18) as follows:

$$p(S_{l_{jk}} \mid X_k, \mu_{jk}, S_{r_{jk}}, \beta, w) \propto (S_{l_{jk}}^{-\frac{1}{2}} + S_{r_{jk}}^{-\frac{1}{2}}) S_{l_{jk}}^{\frac{\beta}{2} - 1} \exp\left[-\frac{S_{l_{jk}} \sum_{i: X_{ik} < \mu_{jk}} (x_{ik} - \mu_{jk})^2}{2} - \frac{w\beta S_{l_{jk}}}{2}\right] \quad (21)$$

Random samples of posteriors can be drawn by using the MCMC method. In our work, we use Metropolis–Hastings algorithm to sample precision parameters. For the hyperparameters β and w , the equation Eq. (18) plays the role of the likelihood function. Combining Eq. (12), Eq. (19), and Eq. (20), we obtain the following posteriors:

$$p(\beta_l \mid S_{l_{1k}}, \dots, S_{l_{Mk}}, w_l) \propto \Gamma\left(\frac{\beta_l}{2}\right)^{-M} \exp\left(-\frac{1}{2\beta_l}\right) \left(\frac{\beta_l}{2}\right)^{\frac{M\beta_l - 3}{2}} \prod_{j=1}^M (w_l S_{l_{jk}})^{\frac{\beta_l}{2}} \exp\left(-\frac{\beta_l w_l S_{l_{jk}}}{2}\right) \quad (22)$$

$$p(w_l \mid S_{l_{1k}}, \dots, S_{l_{Mk}}, \beta_l) \propto \Gamma(M\beta_l + 1, \frac{M\beta_l + 1}{\sigma_y^{-2} + \beta_l \sum_{j=1}^M S_{l_{jk}}}) \quad (23)$$

where we only show the left side of β and w parameters with similar posteriors for the right side parameters. The posterior distribution of precision β is not a standard form, but its logarithmic posterior is log-concave. Therefore, we can sample from the distribution for $\log(\beta)$ using ARS technique and transform the resultant to get values for β .

The proposed complete algorithm can be summarized by the following:

Algorithm 1. IAGM algorithm

- 1: **procedure**
 - 2: initialize assignments and parameters
 - 3: *loop*:
 - 4: Update mixture parameters μ_j , $S_{l_{jk}}$ and $S_{r_{jk}}$ from posteriors in Eq. (15) and Eq. (21).
 - 5: Update hyperparameters λ , r , β , w and Dirichlet process concentration parameter α from posteriors in Eq. (16), (17), (22), (23) and (11).
 - 6: Update the indicators conditioned on the other indicators and the hyperparameters from Eq. (9).
 - 7: The convergence criteria is reached when the difference of the current value of joint posteriors and the previous value is less than 10^{-4} . Otherwise, repeat above procedures until convergence.
-

3 Experimental Setup

3.1 Background Subtraction Application

In this section, we employ the proposed IAGM model for video background subtraction with a pixel-level evaluation approach as in [8]. The background subtraction methodology starts off by constructing the model using the proposed IAGM model. After applying the learning algorithm for the model, we discriminate between the mixture components for the representation of foreground and background pixels for each of the new input frames.

Assume that each video frame has P number pixels such that $\vec{\mathcal{X}} = (\mathcal{X}_1, \dots, \mathcal{X}_P)$ then each pixel \mathcal{X} is assigned as a foreground or background pixel according to the trained IAGM model $p(\mathcal{X} | \Theta) = \prod_{i=1}^N \sum_{j=1}^M p_j p(\mathcal{X}_i | \xi_j)$. Components that occurs frequently, i.e. high p value, and with a low standard deviation $S^{-\frac{1}{2}}$) are modeled as the background.

Accordingly, the value of $p_j / (\|S_{l_j}^{-\frac{1}{2}}\| + \|S_{r_j}^{-\frac{1}{2}}\|)$ is used to order the mixture components, where p_j is the mixing weight for component j , $\|S_{l_j}^{-\frac{1}{2}}\|$ and $\|S_{r_j}^{-\frac{1}{2}}\|$ are the respective norms of left and right standard deviations of the j component [8]. The first B number of components are chosen to model the background, with B estimated as:

$$B = \arg \min_b \sum_{j=1}^b p_j > T \quad (24)$$

where T is a measure of the minimum proportion of the data that represents the background in the scene, and the rest of the components are defined as foreground components.

3.2 Results and Discussion

We apply the proposed algorithm to the Change Detection dataset [20]. The dataset consists of six categories with a total of 31 videos totaling 90,000 frames. Each of the categories (baseline, dynamic background, camera jitter, shadows, intermittent object motion, and thermal) contains around 4 to 6 different videos sequences from low-resolution IP cameras.

In this paper, we have selected five videos from the Change Detection dataset to evaluate our proposed methodology. We initialize the IAGM by incrementally increasing the threshold multiple times and choosing the optimum parameter setting. We adopt the threshold factor $T = 0.9$ in our method. We set the maximum component number for the algorithm as 9 and the standard deviation factor $K = 2$. Evaluations of the proposed IAGM can be observed in the confusion matrices in Fig. 2. Moreover, Fig. 3 show visual results of our proposed method on samples frames in the Library and Street Light video sequences.

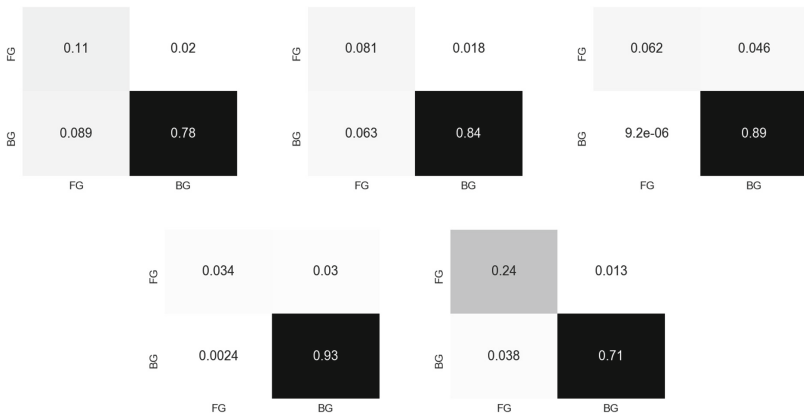


Fig. 2. Confusion matrices of the proposed method employed for background subtraction on the boulevard (top left), abandoned box (top center), street light (top right), sofa (bottom left), and library (bottom right) videos where FG denotes the foreground and BG denotes the background.

We also compare our results with three other methods from the literature. These include the Gaussian mixture model-based background subtraction algorithms by Sauffer et al. [8] and Zivkovic [9] as well as the finite asymmetric Gaussian mixture model by Elguebaly et al. [3]. We evaluate the performance of the algorithms in terms of the recall and the precision metrics. These are defined respectively as $\text{Recall} = TP / (TP + FN)$ and $\text{precision} = TP / (TP + FP)$ where



Fig. 3. A sample frame from Street Light (left) and Library (right) video sequences and the detected foreground object respectively.

Table 1. Experimental results for the background subtraction application.

	Stauffer et al. [8]	Zivkovic [9]	Elguebaly et al. [3]	IAGM (proposed)
<i>Boulevard</i>				
Recall	83.21%	79.77%	79.54%	84.72%
Precision	40.02%	43.79%	61.13%	55.80%
<i>Abandoned Box</i>				
Recall	45.74%	45.64%	45.18%	81.53%
Precision	65.52%	62.14%	67.41%	56.23%
<i>Street Light</i>				
Recall	32.25%	33.94%	30.33%	57.41%
Precision	89.16%	92.47%	97.56%	99.99%
<i>Sofa</i>				
Recall	51.62%	51.41%	59.90%	53.56%
Precision	85.92%	89.25%	92.52%	93.41%
<i>Library</i>				
Recall	28.00%	28.68%	31.33%	94.74%
Precision	84.76%	81.76%	94.66%	86.52%

TP is the number of true positive foreground pixels, FN is the number of false negatives, and FP is the number of false positives. The results can be seen in Table 1.

As can be observed in Table 1, the proposed IAGM model mostly outperforms the other methodologies in terms of the recall metric, while achieving comparable precision results. For instance, IAGM attains better recall results for the Street Light video sequence with a near perfect precision. This clearly demonstrates the effectiveness of our proposed model and approach.

In particular, our approach detects more foreground pixels; most of which are clustered correctly. This ensures comparable precision results compared with the other algorithms. Our method does not remarkably improve the precision metric due to the sensitivity of the proposed method to the change in environments. With higher number of detected foreground pixels, our approach shows significant improvement in the recall metric. This improvement is especially distinct for the Library video.

These improvements are due to the nature of the IAGM model that is capable of accurately capturing the asymmetry of the observations. This higher flexibility of asymmetric Gaussian distribution allows the incorporation of the different shape distributions of objects. Furthermore, the extension to the infinite mixture using the Dirichlet process with a stick-breaking construction increases adaptability of the proposed model. Hence, we addressed both the parameter learning and the mixture component number determination challenges. These advantages provide a more efficient model for background subtraction.

4 Conclusion

We propose a new infinite mixture model, IAGM, that is capable of modeling the asymmetry of data in contrast to the traditionally deployed GMM. Moreover, we address the challenges of parameter learning and choosing the right number of components through the employment of Bayesian learning and extension of the AGM model to infinity. Furthermore, we demonstrate the efficiency of our model by utilizing it for the background subtraction application. Our achieved results are comparable to three different methods in terms of precision, and superior in terms of the recall metric. Finally, our future plan considers Variational Inference to improve learning speed.

Acknowledgments. The completion of this research was made possible thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Sakpal, N.S., Sabnis, M.: Adaptive background subtraction in images. In: 2018 International Conference on Advances in Communication and Computing Technology (ICACCT), pp. 439–444, February 2018
2. Fan, W., Bouguila, N.: Video background subtraction using online infinite dirichlet mixture models. In: 21st European Signal Processing Conference, EUSIPCO 2013, Marrakech, Morocco, 9–13 September, pp. 1–5 (2013)
3. Elguebaly, T., Bouguila, N.: Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection. *Mach. Vis. Appl.* **25**(5), 1145–1162 (2014)
4. Tang, C., Ahmad, M.O., Wang, C.: Foreground segmentation in video sequences with a dynamic background. In: 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1–6. IEEE (2018)
5. Hedayati, M., Zaki, W.M.D.W., Hussain, A.: Real-time background subtraction for video surveillance: from research to reality. In: 2010 6th International Colloquium on Signal Processing its Applications, pp. 1–6, May 2010
6. Bouguila, N., Ziou, D.: Using unsupervised learning of a finite dirichlet mixture model to improve pattern recognition applications. *Pattern Recogn. Lett.* **26**(12), 1916–1925 (2005)

7. Bilge, Y.C., Kaya, F., Cinbis, N., Çelikcan, U., Sever, H.: Anomaly detection using improved background subtraction. In: 2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, May 2017
8. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), vol. 2, pp. 246–252. IEEE (1999)
9. Zivkovic, Z.: Improved adaptive gaussian mixture model for background subtraction. In: null, pp. 28–31. IEEE (2004)
10. Wang, D., Xie, W., Pei, J., Lu, Z.: Moving area detection based on estimation of static background. *J. Inf. Comput. Sci.* **2**(1), 129–134 (2005)
11. Fu, S., Bouguila, N.: A bayesian intrusion detection framework. In: 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1–8 (2018)
12. Bouguila, N., Ziou, D.: Unsupervised selection of a finite dirichlet mixture model: an mml-based approach. *IEEE Trans. Knowl. Data Eng.* **18**(8), 993–1009 (2006)
13. Elguebaly, T., Bouguila, N.: Simultaneous high-dimensional clustering and feature selection using asymmetric gaussian mixture models. *Image Vision Comput.* **34**, 27–41 (2015)
14. Carlin, B.P., Chib, S.: Bayesian model choice via markov chain monte carlo methods. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **57**(3), 473–484 (1995)
15. Elguebaly, T., Bouguila, N.: A nonparametric bayesian approach for enhanced pedestrian detection and foreground segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2011, Colorado Springs, CO, USA, 20–25 June, pp. 21–26 (2011)
16. Elguebaly, Tarek, Bouguila, Nizar: Infinite generalized gaussian mixture modeling and applications. In: Kamel, Mohamed, Campilho, Aurélio (eds.) ICIAR 2011. LNCS, vol. 6753, pp. 201–210. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21593-3_21
17. Bouguila, N., Ziou, D.: A countably infinite mixture model for clustering and feature selection. *Knowl. Inf. Syst.* **33**(2), 351–370 (2012)
18. Bouguila, N.: Infinite liouville mixture models with application to text and texture categorization. *Pattern Recogn. Lett.* **33**(2), 103–110 (2012)
19. Rasmussen, C.E.: The infinite gaussian mixture model. In: Solla, S.A., Leen, T.K., Müller, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 12, pp. 554–560. MIT Press, Cambridge (2000)
20. Wang, Y., Jodoin, P.M., Porikli, F., Konrad, J., Benezeth, Y., Ishwar, P.: CDnet 2014: an expanded change detection benchmark dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 387–394 (2014)