Spatially-Coherent Segmentation Using Hierarchical Gaussian Mixture Reduction Based on Cauchy-Schwarz Divergence

Adama Nouboukpo and Mohand Said Allili[✉]

University of Quebec in Outaouais,
283 Boul Alexandre-Tache, Gatineau, QC J8X 3X7, Canada
{noua06,mohandsaid.allili}@uqo.ca

Abstract. Gaussian mixture models (GMM) are widely used for image segmentation. The bigger the number in the mixture, the higher will be the data likelihood. Unfortunately, too many GMM components leads to model overfitting and poor segmentation. Thus, there has been a growing interest in GMM reduction algorithms that rely on component fusion while preserving the structure of data. In this work, we present an algorithm based on a closed-form Cauchy-Schwarz divergence for GMM reduction. Contrarily to previous GMM reduction techniques which a single GMM, our approach can lead to multiple small GMMs describing more accurately the structure of the data. Experiments on image foreground segmentation demonstrate the effectiveness of our proposed model compared to state-of-art methods.

Keywords: Gaussian mixture models · Mixture reduction · Image segmentation · Cauchy-Schwarz divergence

1 Introduction

GMMs are semi-parametric density functions, which are represented as a weighted sum of Gaussian densities called components. They can practically approximate any density shape using a finite number of components. They are very useful for data clustering in particular for image segmentation where they have demonstrated a good performance [4,5]. Although using too many components increases the approximation accuracy, it often results in complex models which overfit data. To obtain less complex models, two approaches are used. The first approach (horizontal) generates multiple GMMs with different numbers of components, and choose the best one according an information theoretic criterion (e.g., AIC, BIC) [2,11]. The second approach (vertical), which is the interest of our work, starts from a GMM with a high number of components and reduce it to a less complex GMM by iteratively fusing the different components into Gaussians [8,13].

A simple method to reduce a GMM is to eliminate components that do not contribute much to the mixture. Alternatively, instead of deleting components

of the GMM or re-estimating it, one could merge components while preserving the structure of the data [13]. The idea is to merge similar components that "belong together" in a clustering sense. Therefore, there is a need to formalize "component similarity". Among the used measures, Kullback-Leibler divergence (KLD) is the most popular [17]. However, there is no closed-form expression for the KLD between two GMMs [16]. Because of this limitation, several existing methods use an approximation of the KLD using sampling methods [13,14,17]. However, such approaches can be computationally expensive, especially when the dimensionality of the data is very high. On the other hand, mixture reduction is achieved by fusing two components at a time in the GMM into a single Gaussian. This can end up in the long run to bold components having large variance. In other words, merging two or more Gaussians is not necessarily a Gaussian-distributed as in the case, for example, in foreground image segmentation where both foreground (resp. background) regions are not necessarily Gaussian [3,5].

In this paper, we introduce a new technique for GMM reduction into one or multiple small GMMs describing the different structures of the data. More specifically, an initial GMM is reduced iteratively by clustering its components into separate small GMMs describing the structure of the data in a hierarchical way. To facilitate components clustering, we use the Cauchy-Schwarz divergence (CSD) which has a closed-form expression and its computational complexity varies linearly with the dimension of data. We apply our reduction algorithm to foreground segmentation in still images, namely in salient object and skin lesion segmentation. Comparison with state-of-the art methods have shown the performance of our method.

The rest of this paper is organized as follows: Sect. 2 briefly describes the theoretical background of our algorithm. Section 3 presents details of our proposed method for mixture reduction problem. Section 4 describes experimental results such as qualitative and quantitative results. We end the paper with a conclusion and future work perspectives.

2   Background

Given a random vector $x \in R^D$, we consider two GMMs with K and M their respective numbers of mixture components. Let $f(x)$ and $g(x)$ denote respectively their probability density functions (PDF) as follows:

$$f(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \mathbf{\Sigma}_k), \qquad g(x) = \sum_{m=1}^{M} \omega_m \mathcal{N}(x|\nu_m, \mathbf{\Lambda}_m) \qquad (1)$$

where $(\pi_k, \mu_k, \mathbf{\Sigma}_k)$, $k = 1, ..., K$ are the weights, mean vectors and covariance matrices of the Gaussians composing mixture f, and $(\omega_m, \nu_m, \mathbf{\Lambda}_m)$, $m = 1, ..., M$, are the same parameters composing mixture g. $\mathcal{N}(x|\mu_k, \mathbf{\Sigma}_k)$ and $\mathcal{N}(x|\nu_m, \mathbf{\Lambda}_m)$ are the multivariate Gaussian distributions.

The Cauchy-Schwarz divergence is based on the Cauchy-Schwarz inequality for inner products, and it is formulated between f and g as follows [16]:

$$CSD(f,g) = -\log \left(\frac{\int f(x)g(x)dx}{\sqrt{\int f(x)^2 dx \int g(x)^2 dx}} \right) \tag{2}$$

It is clear that the term inside the parentheses takes its values in the interval $[0,1]$ with equality to 1 if and only if $f(x) = g(x)$, such that CSD is always nonnegative and symmetric. The closed-form expression for CSD of a pair of GMMs can be derived using the Gaussian multiplication and identities, which forms the basic building block of (2). We obtain a closed-form expression for the CSD, which does not depend on x:

$$CSD(f,g) = -\log \left(\sum_{k=1}^{K} \sum_{m=1}^{M} \pi_k \omega_m \mathcal{N}(\mu_k | \nu_m, (\boldsymbol{\Sigma}_k + \boldsymbol{\Lambda}_m)) \right)$$
$$+ \frac{1}{2} \log \left(\sum_{k=1}^{K} \sum_{k'=1}^{K} \pi_k \pi_{k'} \mathcal{N}(\mu_k | \mu_{k'}, (\boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_{k'})) \right)$$
$$+ \frac{1}{2} \log \left(\sum_{m=1}^{M} \sum_{m'=1}^{M} \omega_m \omega_{m'} \mathcal{N}(\nu_m | \nu_{m'}, (\boldsymbol{\Lambda}_k + \boldsymbol{\Lambda}_{m'})) \right) \tag{3}$$

This expression has computational complexity order of $O(K^2)$ and not integral computation is required as in the case of the KLD. Therefore, the CSD is chosen in this work because it can be computed in closed-form which results in a faster algorithm.

3   Proposed Mixture Reduction Method

The proposed method, coined CSGMR (Cauchy-Schwarz divergence for Gaussian mixture reduction), is formulated as follows. Let us consider a GMM with K multivariate Gaussian densities of the form:

$$f(x|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \boldsymbol{\Sigma}_k), \tag{4}$$

where $\mu_k, \boldsymbol{\Sigma}_k$ represent the mean vector and covariance matrix of the kth component, $\boldsymbol{\Theta} = \{\pi_k, \mu_k, \boldsymbol{\Sigma}_k\}_{k=1}^{K}$ is the set of all mixture parameters, with π_k are the component weights that satisfy the constraints: $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$. Generally, the Expectation-Maximization (EM) algorithm [9] based on the maximum likelihood estimation (MLE) is used to calculate the GMM parameters.

Given the mixture f in (4), we wish to find a reduced mixture model g M having components ($M << K$) by collapsing the components of f in those of g.

Note that a similar challenge has been addressed in [5] using finite mixtures of generalized Gaussians. However, the method is hard to generalize to multiple dimensions of data.

Contrarily to traditional mixture reduction methods which output a single mixture [8,13,14], our method outputs several mixture models describing different clusters in the structure of data with arbitrary dimension. This characteristic of our method will reveal its importance for clustering problems such as image segmentation, where data structure comprises several non-Gaussian groups. In addition, the reduction is carried out without causing a strong deviation from the structure of the original GMM one according to CSD measure. Suppose that the density of reduced model is:

$$g(x|\boldsymbol{\Phi}) = \sum_{m=1}^{M} \omega_m p_m(x|\boldsymbol{\Psi}_m), \tag{5}$$

where $p_m(x|\boldsymbol{\Psi}_m) = \sum_{j=1}^{K_m} \pi_{jm}\mathcal{N}(x|\mu_{jm}, \boldsymbol{\Sigma}_{jm})$, with $\boldsymbol{\Psi}_m = \{\pi_{jm}, \mu_{jm}, \boldsymbol{\Sigma}_{im}\}_{j=1}^{K_m}$ and $\boldsymbol{\Phi} = \{\omega_m, \boldsymbol{\Psi}_m\}_{m=1}^{M}$. The reduction is made in such a way that $\sum_{m=1}^{M} K_m <<$ K. Note that when $K_m = 1$, $m = 1, ..., M$, we are in the classical mixture reduction scheme where each reduced component in g is a Gaussian, as proposed in [8,13].

3.1   Proposed Algorithm

Our mixture reduction method is similar to classical agglomerative clustering, where initial groups are Gaussian components which are iteratively merged into clusters. Therefore, the input of our algorithm is an initial mixture model $f(x|\boldsymbol{\Theta})$ having a large number of components, and the output is a reduced (super) mixture model $g(x|\boldsymbol{\Psi})$ composed of several small GMMs. For this purpose, we maintain *similarity matrix* indicating CSD distances between formed GMMs (clusters) at each merging iteration.

Let $f_1, f_2, ..., f_{N_c}$, with $M \leq N_c \leq K$, be intermediate models resulting from fusions of initial GMM components. A similarity matrix S of size $N_c \times N_c$ is computed, where $S_{rs} = CSD(f_r, f_s)$, $\forall r \neq s$ and $r, s = 1, ..., N_c$. At each iteration of the reduction algorithm, we select the most similar groups to fuse, thus causing the smallest overall change in the initial mixture structure. More specifically, let f_r and f_s be two candidate models to be merged, where K_r and K_s are their number of components, respectively. To estimate the optimal number of components of their fusion, we use the Akaike (AIC) criterion [4] that will search the minimum of the following function:

$$K_{rs} = \underset{l \in [max(K_r, K_s), ..., K_r + K_s]}{\operatorname{argmin}} \left\{ -\log(\boldsymbol{\Psi}_l) + 2\lambda_l \right\}, \tag{6}$$

where $\log(\boldsymbol{\Psi}_l)$ is the log-likelihood corresponding to a GMM model with the number of components l, and $\boldsymbol{\Psi}_l$ is the MLE estimate. Finally, λ_l is the number of

the free parameters in that model. The following script summarizes our reduction algorithm:

Algorithm 1. Proposed Algorithm: CSGMR

Input: f: mixture of K components
Output: g: mixture of M mixtures ($M << K$)
1. Initialize the GMM f and number of clusters $N_c \leftarrow K$;
2. Initialize the similarity matrix S;
3. **while** $N_c > M$ **do**
   * Find the nearest pair of groups (f_r, f_s) according to S;
   * Merge f_r and f_s into f_{rs} and remove f_r and f_s;
   * Estimate the number of components of f_{rs} using AIC;
   * Update S; $N_c \leftarrow N_c - 1$;
   end
4. Build the final GMM g by the produced M GMMs.

3.2   Foreground Image Segmentation

We apply the proposed method for foreground segmentation in natural and medical images (see illustration in Fig. 1). First, we generated an initial GMM by over-segmenting the image into superpixels using the SLIC method [1]. The superpixel itself is not enough to provide a robust image descriptor for segmentation, since the consistency of its neighborhood is not considered [10]. To take advantage of the superpixel structure and our method for segmentation, let K be the number of produced superpixels. We model each superpixel as a multivariate Gaussian with parameters given by the mean vector and covariance matrix of the RGB color of its composing pixels. We then use our mixture reduction algorithm to form the foreground and background regions ($M = 2$).
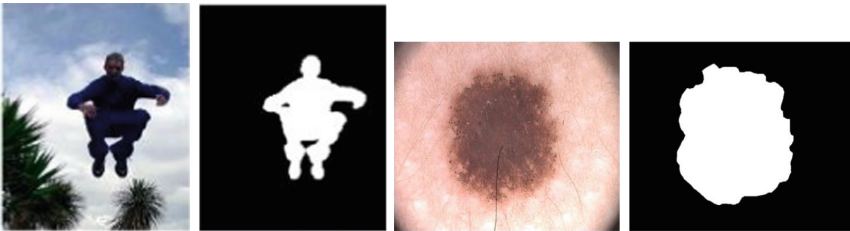

Fig. 1. Example of foreground segmentation.

To obtain compact regions for our segmentation, we enforce our reduction algorithm by imposing that two segments S_i, and S_j can be merged together only if they are spatially adjacent. To this end, we construct and maintain a region adjacency graph G in the clustering process, where the adjacency set of S_i is

denoted by G_i. Since our mixture reduction algorithm can output any number of clusters, we can tune the segmentation algorithm to produce any number of desired regions constituting the important structures of the image. For example, in binary segmentation where the objective is to segmentation the foreground from background, the output is $M = 2$.

4 Experiments and Results

We conducted several tests to assess the effectiveness and the usefulness of the proposed method. First, we compare the effectiveness of the proposed method against two of the best algorithms in the literature, namely COWA [6] and Runnalls [17]. We randomly generate a GMM with 10 components and 4 dimensions and search for an optimal reduced model comprising 5 components. We also conducted a binary classification of 20 Gaussians drawn from a sample of 3000 observations. As shown in Table 1, we can see that [17] has the lowest execution time followed by our method. However, our approach surpasses [17] and [6] in term of the reduction quality.

Table 1. Mixture simplification results.

	Execution time (sec)	Classification error
COWA	0.0390	0.0710
Runnalls	0.0048	0.0033
CSGMR	0.0054	0.0027

In order to illustrate the performance of the proposed method for image segmentation, we computed GMMs of different sizes ($K = 100, 125, 500$) for each
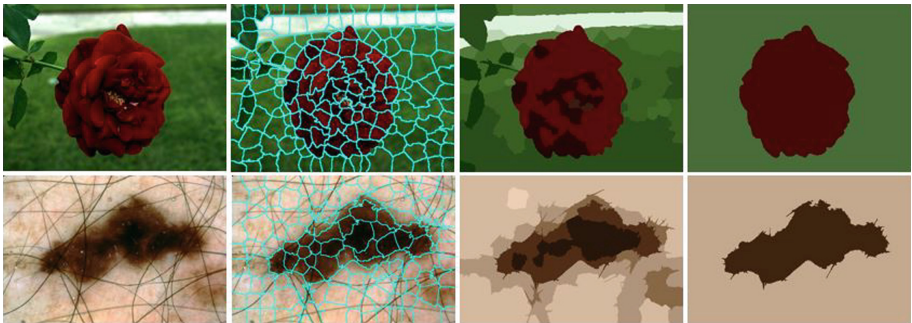

Fig. 2. Result of CSGMR on the ISIC and MSRA10K datasets: Original images (first column), initial superpixels (second column), reduction with ($M = 12$ top third column and $M = 6$ bottom third column), $M = 2$ (last column).

Table 2. Quantitative segmentation performance: (a) for ISIC and (b) for MSRA10K.

<table>
<tr><td rowspan="4">(a)</td><td></td><td>[13]</td><td>CSGMR</td></tr>
<tr><td>Accuracy</td><td>88.9326</td><td>97.0047</td></tr>
<tr><td>Precision</td><td>84.7651</td><td>97.0394</td></tr>
<tr><td>Recall</td><td>83.1526</td><td>91.9662</td></tr>
</table>

<table>
<tr><td rowspan="4">(b)</td><td></td><td>[13]</td><td>CSGMR</td></tr>
<tr><td>Accuracy</td><td>92.9503</td><td>98.7359</td></tr>
<tr><td>Precision</td><td>94.0669</td><td>96.5222</td></tr>
<tr><td>Recall</td><td>93.1526</td><td>95.9896</td></tr>
</table>

image and perform our mixture reduction algorithm. The experiments were performed using two different public datasets: ISIC dataset [15] and MSRA10K dataset [7]. All experiments were performed without any correction of the images. The visual results are shown in Fig. 2. In Fig. 2 (column two) with $K = 181$ and $K = 126$, the reduced mixture is, respectively, $M = 12$ and $K = 6$ (Fig. 2 column three). As presented in Fig. 2 (column 3), the foreground and the background are modeled by a mixture. In order to compare the results of the proposed method
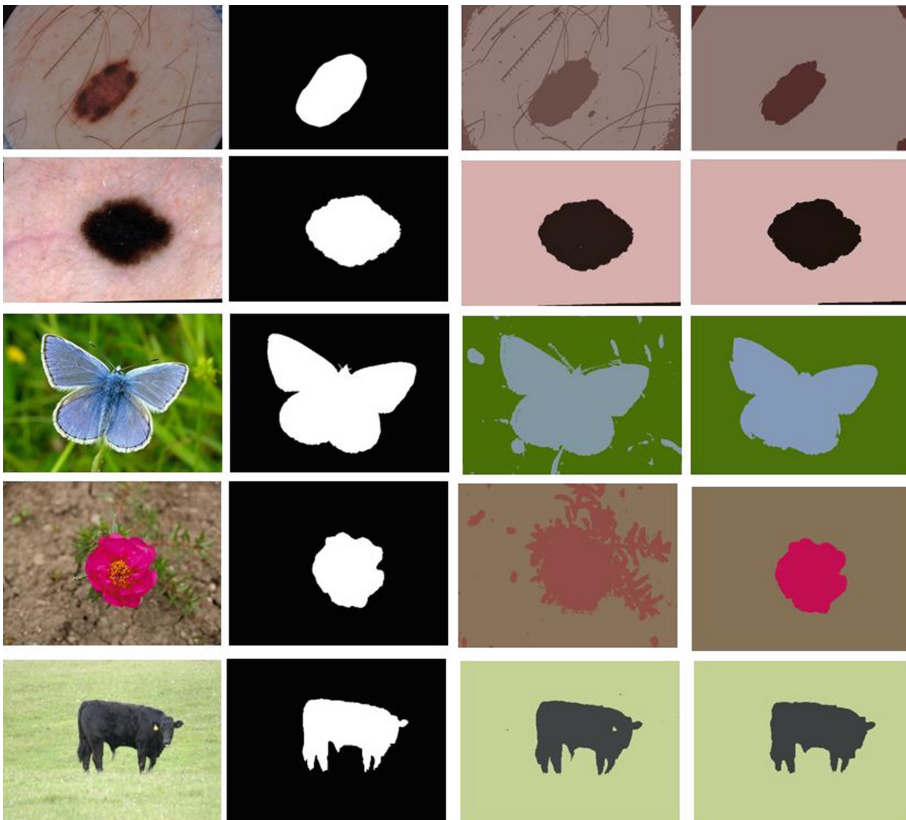

Fig. 3. Visual comparative results: Original images (first column), Ground-truth (second column), [12] result (third column) and CSGMR result (last column).

with [12], we focus our experiment in foreground and background detection i.e. $M = 2$ (see Fig. 3, column 3). The evaluation performance between our approach and [12] is conducted with 500 images in each dataset. The metrics used to quantify the segmentation result are recall, precision, accuracy. Note that higher values of these metrics indicate better results. The segmented image is the result of the reduced mixture. We can see that the images are segmented more correctly by using our proposed method. A quantitative comparative study between different methods is depicted in Table 2. It represents the average metrics for the 500 images selected from each dataset. According to these qualitative and quantitative results, obtained results with CSGMR are very encouraging and useful.

5   Conclusion

We have proposed an efficient GMM reduction algorithm based on CSD hierarchical clustering. Our method enables to quickly compute a compact version of an initial GMM to an another GMM constituted of non-Gaussian clusters. Experimental results on the simulated and image segmentation showed the effectiveness of the proposed technique with in comparison to other techniques. Note that our model can be extended easily into other exponential family distributions.

References

1. Achanta, R., Shaji, A., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE TPAMI **34**(11), 2274–2282 (2012)
2. Allili, M.S., Ziou, D., et al.: Image and video segmentation by combining unsupervised generalized Gaussian mixture modeling and feature selection. IEEE TCSVT **20**(10), 1373–1377 (2010)
3. Allili, M.S.: Effective object tracking by matching object and background models using active contours. In: IEEE ICIP, pp. 873–876 (2009)
4. Allili, M.S., Ziou, D.: Automatic colour-texture image segmentation using active contours. Int. J. Comput. Math. **84**(9), 1325–1338 (2007)
5. Boulmerka, A., Allili, M.S., et al.: A generalized multiclass histogram thresholding approach based on mixture modelling. PR **47**(3), 1330–1348 (2014)
6. Chen, H.D., Chang, K.C., Smith, C.: Constraint optimized weight adaptation for Gaussian mixture reduction. In: Signal Processing, Sensor Fusion, and Target Recognition, SPIE, vol. 7697 (2010)
7. Cheng, M.-M., Mitra, N.J., et al.: Global contrast based salient region detection. IEEE TPAMI **37**(3), 569–582 (2015)
8. Crouse, D.F., Willett, P., et al.: A look at Gaussian mixture reduction algorithms. In: IEEE International Conference on Information Fusion, pp. 1–8 (2011)
9. Dempster, A.P., Laird, N.M., et al.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B **39**(1), 1–38 (1977)

10. Filali, I., Allili, M.S., et al.: Multi-scale salient object detection using graph ranking and global-local saliency refinement. Sig. Process. Image Commun. **47**, 380–401 (2016)
11. Figueiredo, M.T., Jain, A.K.: Unsupervised learning of finite mixture models. IEEE TPAMI **24**(3), 381–396 (2002)
12. Fu, Z., Wang, L.: Color image segmentation using Gaussian mixture model and EM algorithm. In: Wang, F.L., Lei, J., Lau, R.W.H., Zhang, J. (eds.) CMSP 2012. CCIS, vol. 346, pp. 61–66. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35286-7_9
13. Goldberger, J., Roweis, S.T.: Hierarchical clustering of a mixture model. In: NIPS, pp. 505–512 (2005)
14. Hershey, J.R., Olsen, P.: Approximating the Kullback-Leibler divergence between Gaussian mixture models. In: IEEE ICASSP, pp. 317–320 (2007)
15. ISIC: The International Skin Imaging Collaboration. https://challenge2018.isic-archive.com/. Accessed 24 May 2019
16. Kampa, K., Hasanbelliu, E., Principe, J.: Closed-form Cauchy-Schwarz PDF divergence for mixture of Gaussians. In: IEEE IJCNN, pp. 2578–2585 (2011)
17. Runnalls, A.R.: Kullback-Leibler approach to Gaussian mixture reduction. IEEE TAES **43**(3), 989–999 (2007)