

Parametric Characterization of Multimodal Distributions with Non-Gaussian Modes.

Ashutosh Tewari, Michael J. Giering
Decision Support and Machine Intelligence
United Technologies Research Center
East Hartford, CT, USA
tewaria@utrc.utc.com, gierinmj@utrc.utc.com

Arvind Raghunathan
System Dynamics and Optimization
United Technologies Research Center
East Hartford, CT, USA
raghunathan@merl.com*
*(currently at Mitsubishi Electric Research Labs.)

Abstract—In statistics, mixture models are used to characterize datasets with multimodal distributions. A class of mixture models called Gaussian Mixture Models (GMMs) has gained immense popularity among practitioners because of its sound statistical foundation and an efficient learning algorithm, which scales very well with both the dimension and the size of a dataset. However, the underlying assumption, that every mixing component is normally distributed, can often be too rigid for several real life datasets. In this paper, we introduce a new class of parametric mixture models that are based on Copula functions. The goal is to relax the assumption about the normality of mixing components. We formulate a class of functions called Gaussian Mixture Copula functions for the characterization of multi-modal distributions. The parameters of the proposed Gaussian Mixture Copula Model (GMCM) can be obtained in a Maximum-Likelihood setting. For this purpose, an Expectation-Maximization (EM) and a Gradient-based optimization algorithm are proposed. Owing to the non-convex log-likelihood function, only locally optimal solutions can be obtained. We also provide experimental evidence of the benefits of the GMCM over GMM using both synthetic and real-life datasets.

Keywords—Mixture Models; Copula Function; Gaussian Mixture Models; Image Segmentation; Non-convex Optimization

I. INTRODUCTION

Mixture model refers to a probabilistic model used for describing a population that consists of observations generated from different underlying distributions. Generally, these distributions are assumed to have the same parametric form, but are allowed to have different parameter values. For continuous and real valued random variables, Gaussian Mixture Model (GMM) is by far the most widely used mixture model. In a GMM, every mixing component is assumed to be normally distributed. For a given number of components, the parameters of a GMM can be efficiently learned using an Expectation-Maximization algorithm [1], which scales very well with both the dimension and size of a dataset. As a result, GMMs have found application in variety of problems such as unsupervised clustering of data for anomaly detection [2], statistical process control [3], [4], image retrieval and segmentation [5], [6], real time target tracking for surveillance [7] etc. Owing to their widespread popularity

and the availability of numerous computer software, GMMs are often misused by the practitioners in different domains. For several real-life datasets, the assumption of normally distributed components is violated, thus, making GMMs inappropriate for such cases. The focus of the present work is to introduce a new class of mixture models that relaxes the assumption of normally distributed components, while retaining some desirable properties of GMMs. To this end, we propose *Gaussian Mixture Copula Models* (GMCMs), whose foundation is laid on the theory of *Copula* functions. The Copula functions provide an elegant way to estimate the joint distributions by separating the estimation of marginal distributions of the random variables from the dependencies that exists between them [8], [9], [10]. These functions are used extensively in fields of finance and econometrics [11], [12], but are limited to only unimodal datasets. As per the knowledge of the author, there hasn't been any formal attempt to extend the theory of copula functions to multimodal datasets.

In section II, we setup the notation to be used in the paper and provide a brief introduction on the Copula functions. In section III, we present the formulation of GMCM and present two different algorithms for the parameter estimation. In section IV, we demonstrate the learning of a GMCM on a synthetic dataset and discuss some of the challenges and future work. We also show an application of GMCM for the task of image segmentation followed by concluding remarks in section V.

II. JOINT DENSITY ESTIMATION USING COPULA FUNCTION

A. Probability Distribution of Random Variables

In this section, we quickly revisit some basic concepts related to the joint distributions of the random variables, as it may be helpful in understanding the contents of the subsequent sections.

For a set of d real valued random variables, X_1, X_2, \dots, X_d the joint cumulative distribution function

(or simply distribution) is defined as,

$$F(x_1, x_2, \dots, x_d; \Theta) = \Pr[X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d] \quad (1)$$

where, x_j is an instantiation of the random variable X_j . The symbol Θ denotes the set of parameters that define the joint distribution. The joint distribution maps the multivariate random variable, X , to a scalar between 0 and 1 i.e. $F : \mathbf{R}^d \rightarrow (0, 1)$. A marginal distribution can be obtained from the joint distribution by marginalizing out other dimensions i.e. $F_j(x_j) = \lim_{x_i \rightarrow \infty \forall i \neq j} F(x_1, x_2, \dots, x_d; \Theta)$, $j = 1 \dots d$. The marginal distribution values are uniformly distributed in interval $(0, 1)$. As a result, an inverse distribution function can be uniquely defined as $F_j^{-1}(t : t \sim U(0, 1)) \rightarrow \inf[x_j : F_j(x_j) \geq t]$, which allows transformation of a scalar between 0 and 1 to a unique value of the random variable X_j . If the joint distribution is sufficiently differentiable one can derive, $f : \mathbf{R}^d \rightarrow (0, \infty)$, the probability density function (or density) as $f(x_1, x_2, \dots, x_d) = \frac{\partial^d F(x_1, x_2, \dots, x_d)}{\partial x_1 \partial x_2 \dots \partial x_d}$.

B. Joint Distribution and Copulas

Copula functions allow splitting the problem of joint distribution estimation into two parts, 1) estimation of the marginal distributions and 2) estimation of dependencies between the random variables. Due to this decoupling, it is possible for a joint distribution, to have marginal distributions from different parametric/non-parametric families. The theorem by Sklar [13] is the key to the copula based joint distribution/density estimation.

Sklar's Theorem (1959): A joint distribution $F(x_1, x_2, \dots, x_d)$ of random variables X_1, X_2, \dots, X_d , can be expressed as a function of the marginal distributions, $F_i(x_i)$, as shown in equation 2.

$$F(x_1, x_2, \dots, x_d; \Theta) = C(F_1(x_1; \lambda_1), F_2(x_2; \lambda_2), \dots, F_d(x_d; \lambda_d); \theta) \quad (2)$$

The parameter set Θ for the joint distribution is divided into the dimension specific parameters, λ_j , and the parameters encoding the dependence structure, θ . The function C is called a *Copula* function. Since the marginal distributions are distributed uniformly in the interval $(0, 1)$, the copula function is essentially defined on a unit hypercube i.e. $C : U(0, 1)^n \rightarrow (0, 1)$. The detailed proof of Sklar's theorem and additional properties of the copula function are provided elsewhere [8-11]. If the copula function, C , and the marginal distributions, F_j , are sufficiently differentiable then the joint density can be obtained as a product of individual marginal densities, f_j , and the copula density, c , as shown in equation 3. The marginal distribution values, $F_j(x_j; \lambda_j)$, have been denoted as u_j .

$$\begin{aligned} f(x_1, x_2, \dots, x_d; \Theta) \\ = f_1(x_1; \lambda_1) \times f_2(x_2; \lambda_2) \times \dots \times f_d(x_d; \lambda_d) \times c(u_1, u_2, \dots, u_d; \theta) \end{aligned} \quad (3)$$

Because of this separation of the marginal densities from the dependence structure (encoded by the copula density), equation 3 allows a way to estimate a joint density with arbitrary marginal densities. When the random variables are independent, the copula density should assume a value of 1. Equation 3 can be rearranged as shown in equation 4.

$$c(u_1, u_2, \dots, u_d; \theta) = \frac{f(x_1, x_2, \dots, x_d; \Theta)}{f_1(x_1; \lambda_1) \times f_2(x_2; \lambda_2) \times \dots \times f_d(x_d; \lambda_d)} \quad (4)$$

There is a subtle difference between equations 3 and 4 that needs to be emphasized here. The equation 3 is primarily used for estimating unknown joint densities, if the forms of the marginal and the copula densities are known. On the other hand, equation 4 allows deriving a copula density from a known multivariate joint density. One such example is the *Gaussian Copula* which is derived from a multivariate Gaussian density and has been used extensively in the areas of finance and econometrics [11], [12]. Briefly, a Gaussian copula density is derived from a standard normal multivariate Gaussian as shown in equation 5.

$$\begin{aligned} c(u_1, u_2, \dots, u_d; \theta) \\ = \frac{\phi(\Phi_1^{-1}(u_1), \Phi_2^{-1}(u_2), \dots, \Phi_d^{-1}(u_d); \theta)}{\phi_1(\Phi_1^{-1}(u_1)) \times \phi_2(\Phi_2^{-1}(u_2)) \times \dots \times \phi_d(\Phi_d^{-1}(u_d))} \end{aligned} \quad (5)$$

where, ϕ is the multivariate Gaussian density with standard normal marginal densities ϕ_j , with zero mean and unit variance along all dimensions. Φ_j^{-1} represents the inverse function of a univariate standard normal distribution. The Gaussian copula is perhaps the most widely used type of copula function, mainly because of its ability to model non-elliptical distributions and its analytical tractability in higher dimensions. However, they are not sufficiently equipped to handle multi-modal datasets, which provides the motivation for this work.

III. GAUSSIAN MIXTURE COPULA (GMC) FUNCTION

Equation 5 is the expression of a Gaussian copula, where the dependence structure is obtained from a unimodal, multivariate Gaussian density, ϕ . When a dataset has multiple modes with different dependence structure, the applicability of Gaussian copula gets severely limited. In such scenarios, the copula function needs to be defined using a density function that has the ability to encode dependencies present in different modes. To this end, we propose *Gaussian Mixture Copula* functions, wherein the dependence structure

is obtained from a GMM. Equation 6 is the expression for a GMM (ψ) with M Gaussians.

$$\psi(x_1, x_2, \dots, x_d; \Theta) = \sum_{k=1}^M \alpha^{(k)} \phi(x_1, x_2, \dots, x_d; \theta^{(k)}) \quad (6)$$

where, $\alpha^{(k)} \geq 0$ s.t. $\sum_{m=1}^M \alpha^{(k)} = 1$ are the mixing proportions of different modes. $\theta^{(k)}$ denotes the parameters associated with the mean vector, $\mu^{(k)}$, and the covariance matrix, $\Sigma^{(k)}$, of the k^{th} component. The parameter set, Θ , combines the mixing proportions, mean vectors and the covariance matrices of all the modes. Similar to a Gaussian copula, the GMC function can be derived as shown in equation 7.

$$c_{gmc}(u_1, u_2, \dots, u_d; \Theta) = \frac{\psi(y_1, y_2, \dots, y_d; \Theta)}{\prod_{j=1}^d \psi_j(y_j)} \quad (7)$$

where, $y = (y_1, y_2, \dots, y_d)$ are the inverse distribution values i.e. $y_j = \Psi_j^{-1}(u_j)$ and ψ_j and Ψ_j^{-1} denote the marginal density and inverse distribution function of the GMM along the j^{th} dimension.

Unlike a Gaussian copula, whose maximum likelihood parameter estimates can be obtained uniquely [14], obtaining the same for GMC function is intractable because of the non-convex form of the log-likelihood function. As a result, an iterative estimation algorithm, such as an Expectation-Maximization or a Gradient-based method is needed to obtain a locally optimal solution. Next section covers this topic in a greater detail.

A. Estimation of GMCM Parameters

In this section, we propose two disparate algorithms for the estimation of GMCM parameters. The first algorithm is motivated from the Expectation-Maximization algorithm for the estimation of GMM parameters. In the second approach, we pose the problem of GMCM parameter estimation as a maximization problem of a non-convex log-likelihood function subject to linear constraints. We also highlight the pros and cons of the two approaches.

1) *Expectation-Maximization Algorithm:* Given N i.i.d sample $\{x^{(i)}\}_1^N$, we would like to estimate the parameter set Θ that maximizes the log likelihood function given by equation 8.

$$\ell(\Theta | \{u^{(i)}\}_1^N) = \sum_{i=1}^N \log(c_{gmc}(u_1^{(i)}, u_2^{(i)}, \dots, u_d^{(i)}; \Theta)) \quad (8)$$

where, $u_j^{(i)}$ is the CDF value of the i^{th} sample along the j^{th} dimension. The assumption made here is that the marginal distributions have already been obtained beforehand. For conciseness, the term $\{u^{(i)}\}_1^N$ have been omitted from the expression of this log-likelihood function, in rest of the paper. According to equation 7, the GMC function share the same parameter set as the Gaussian mixture density

from which the GMC function is derived. Therefore, for the estimation of GMCM parameters, one may be tempted to use the EM algorithm for the GMM. However, the EM algorithm for GMM parameter estimation is not directly applicable because of a fundamental difference between the two models. The inputs, (x_1, x_2, \dots, x_d) , in a GMM, remain fixed (equation 6), as the parameters are updated iteratively by a sequence of alternating *Expectation* and *Maximization* steps. On the other hand, the inputs to a GMCM are the marginal CDF values, (u_1, u_2, \dots, u_d) , which are used to obtain the inverse distribution values, (y_1, y_2, \dots, y_d) . Since the inverse distribution functions along the margins change with every parameter update, so does the inverse distribution values. As a result, the assumption of fixed observations, made by the EM algorithm for GMM, is violated, hence it cannot be used in its native form.

```

Set:  $\ell_{max} = -\infty$ 
Initialize:  $\Theta(1)$  randomly, ensuring mixing weights sum to unity and covariance matrices are positive definite.
repeat
  Step 1
   $\ell(\Theta(t)) \leftarrow \sum_{i=1}^N \log(c_{gmc}(u^{(i)}; \Theta(t)))$ 

  Step 2
  if  $\ell(\Theta(t)) > \ell_{max}$  then
     $\Theta(final) \leftarrow \Theta(t)$ 
     $y_j^{(i)} \leftarrow \Psi_j^{-1}(u_j^{(i)}; \Theta(t))$ 
     $\ell_{max} = \ell(\Theta(t))$ 
  end if

  Step 3: Update  $\Theta(t+1)$  using EM Steps
  E-Step:
  
$$h_k^i(t) \leftarrow \frac{\alpha^{(k)} \phi(y^{(i)} | \theta^{(k)}(t))}{\sum_{k=1}^M \alpha^{(k)} \phi(y^{(i)} | \theta^{(k)}(t))}$$


  M-Step:
  
$$\alpha^{(k)}(t+1) \leftarrow \frac{\sum_{i=1}^N h_k^i(t)}{N}$$

  
$$\mu^{(k)}(t+1) \leftarrow \frac{\sum_{i=1}^N h_k^i(t) y^{(i)}}{\sum_{i=1}^N h_k^i(t)}$$

  
$$\Sigma^{(k)}(t+1) \leftarrow \frac{\sum_{i=1}^N h_k^i(t) (y^{(i)} - \mu^{(k)}(t+1))^T (y^{(i)} - \mu^{(k)}(t+1))}{\sum_{i=1}^N h_k^i(t)}$$

until  $\max(\Theta(t+1) - \Theta(t)) < \epsilon$ 

```

Figure 1. An EM algorithm for the estimation of GMCM parameters. Step 1 computes the logarithm of observed data likelihood at current parameter estimates. Step 2, ensures that the algorithm outputs a solution, $\Theta(final)$, that gives highest likelihood of the observed data. Step 3, consists of standard E and M steps for updating the parameters of a GMM.

In an attempt to overcome the above mentioned problem, we propose a modified EM algorithm as outlined in figure 1. Although the obtained solution, $\Theta(final)$, is not guaranteed to be a local optimum, it serves as an excellent initial guess

for the gradient-based algorithm given in the next section. In step 1 of the EM algorithm, the observed data log-likelihood (equation 8) is obtained given the current parameter estimates. If this value exceeds the maximum log-likelihood value seen so far, (ℓ_{max}), the inverse values and ℓ_{max} are updated as shown (step 2). Finally, in step 3, the parameters are updated using the given EM update equations. The iterations are terminated when the relative change in the parameter values is less than a predefined threshold. Step 2 is critical as it ensures that the algorithm outputs a solution, $\Theta(final)$, at which the highest log-likelihood value was observed. The author understands the ad-hoc nature of the proposed EM algorithm, hence a gradient-based optimization method is proposed that guarantees convergence to a local maximum of equation 8.

2) *Gradient-based Algorithm:* An alternate way for parameter estimation in mixture models is via the gradient-based optimization algorithms. In [15], it is shown that the EM algorithm for GMM is a unique case of gradient descent algorithm, where the gradient is regularized by multiplication with a certain positive definite matrix. Here, we formulate the GMCM parameters estimation as an optimization problem, which aims to maximize the log likelihood function given by equation 9 subject to some linear constraints given by equation 10. Using the expression for GMC function from equation 7 and expanding the expression for Gaussian density, we obtain an objective function and a set of linear constraints as shown in equation 9 and 10 respectively. $V^{(k)}$ denotes the Cholesky factor of the k^{th} mode's covariance matrix i.e. $\Sigma^{(k)} = V^k V^{kT}$. The non-negativity constraints on the diagonal elements of these lower triangular matrices ensures the positive definiteness of the corresponding covariance matrices. $\bar{y}^{(i)} = y^{(i)} - u^{(k)}$ is the column vector of the mean adjusted inverse values. In the absence of a closed form expression of the inverse functions, Ψ_j^{-1} , we obtain the inverse values empirically using linear interpolation. This empirical estimation is carried out at every evaluation of the objective function, which adds on to the computational cost of the optimization routine. This is a major challenge, going forward, in order to have an optimization based parameter estimation of the GMCM.

$$\begin{aligned} & \underset{\{\alpha^{(k)}, \mu^{(k)}, V^{(k)}\}_{k=1}^M}{\operatorname{argmax}} \\ & \left[\sum_{i=1}^N \log \sum_{k=1}^M \frac{\alpha^{(k)}}{\sqrt{V^{(k)} V^{(k)T}}} \exp \left(-\frac{\bar{y}^{(i)T} \times (V^k V^{(k)T})^{-1} \times \bar{y}^{(i)}}{2} \right) \right. \\ & \left. - \sum_{i=1}^N \sum_{j=1}^d \log \sum_{k=1}^M \frac{\alpha^{(k)}}{\sqrt{(V^{(k)} V^{(k)T})_{j,j}}} \exp \left(\frac{-\left(\bar{y}^{(i)}\right)^2}{2 \times (V^{(k)} V^{(k)T})_{j,j}} \right) \right] \end{aligned} \quad (9)$$

s.t.

$$\begin{aligned} & V_{j,j}^{(k)} > 0 \\ & \sum_{k=1}^M \alpha^{(k)} = 1 \\ & \alpha^{(k)} \geq 0; k = 1 : M \end{aligned} \quad (10)$$

IV. EXPERIMENT

In this section, we present some experimental studies showing the benefits of GMCMs over GMMs, when the data exhibit multimodality with non-Gaussian modes.

A. Synthetic Dataset

We delineate the steps involved in the estimation of joint density by a GMCM on a synthetic bivariate dataset with 6000 points. The dataset is so generated that it had three modes, two of which were non-Gaussian with high degree of nonlinear dependency between the two dimensions as shown in figure 2.

Step 1: Estimation of the Marginal Densities: In this step, the marginal densities were estimated. We chose to fit non-parametric densities with Gaussian kernels, using a diffusion based method proposed in [16]. There are no strict guidelines on choosing the best way to characterize the marginal densities. However, if the data shows multimodality with non-Gaussian components, non-parametric methods usually provide a reliable fit. The red curves on the histograms, in figure 2 represent the best fit densities.

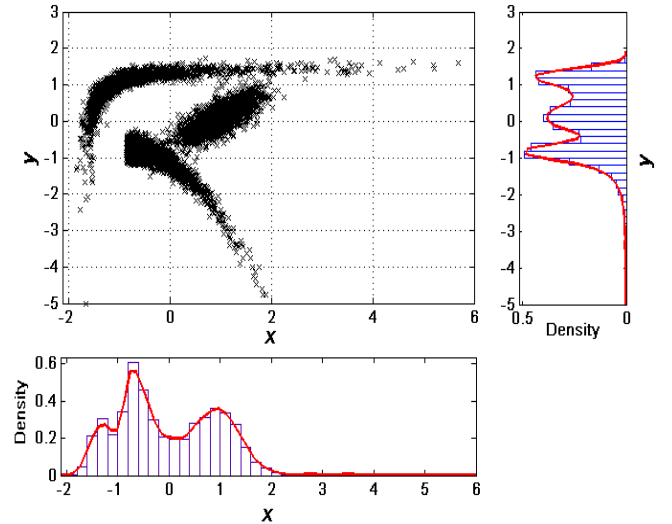


Figure 2. Scatter plot of the synthetic two dimensional data. The histograms on the sides represent the marginal distribution. The red curves are the best fit density for the two dimensions.

Step 2: Data transformation: The marginal distributions learnt in the previous step were used to transform the original data to the marginal CDF values (figure 3).

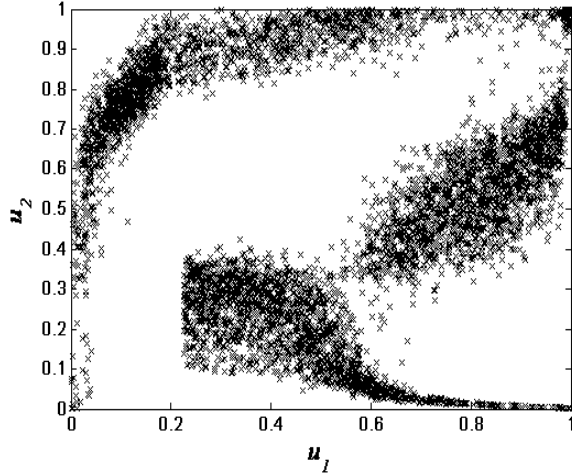


Figure 3. Scatter plot of the marginal distribution values of the synthetic data shown in figure 2. The marginal densities are learnt non-parametrically.

Step 3: Maximum-Likelihood Parameter Estimation: These CDF values go in as the inputs to the EM algorithm outlined in figure 1. The parameters obtained by the EM algorithm were further fine tuned using the gradient-based optimization algorithm outlined in section III B 2. For optimization, we used the *Active-Set* algorithm. The Matlab implementation of the proposed algorithms can be obtained from here [REF]. We found that the initialization of the parameters by the EM algorithm resulted in faster convergence of the optimization routine.

Step 4: Estimating the overall density: The final step is to get the overall joint density function as a product of the non-parametrically estimated marginal densities and the GMC function.

Figure 4a shows the contour of GMCM joint density on the synthetic data. Clearly, the GMCM density is a better representative of the underlying distribution of the points compared to the GMM density (figure 4b). This observation is also corroborated by the observed data log-likelihood values, which was much higher for the GMCM compared to that of GMM. We also used the learnt mixture models to cluster the points using maximum a posteriori criterion. The results of the unsupervised clustering of the points using the two models are given in table I. Clearly, the clustering by GMCM resulted in significantly higher classification accuracy compared to the GMM based clustering. This observation is a direct consequence of a better fit to the data by a GMCM compared to a GMM.

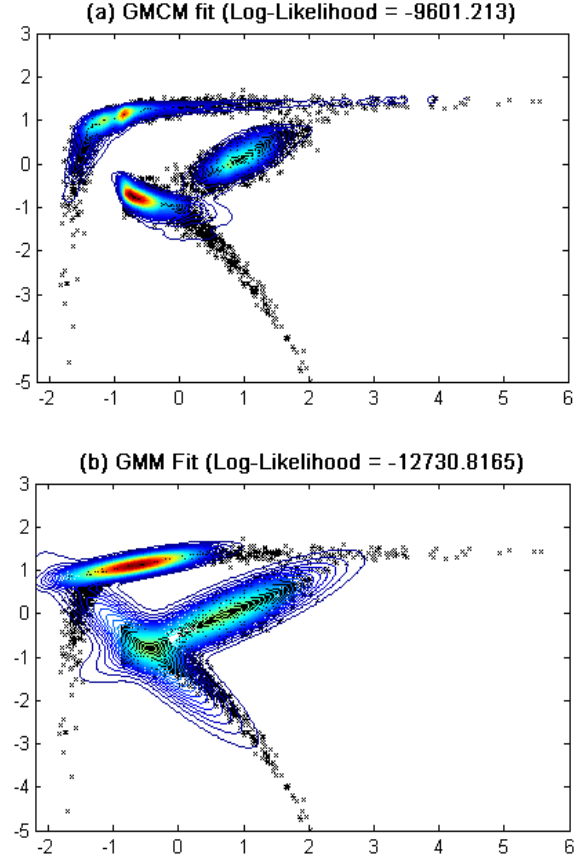


Figure 4. Contours of the GMCM (a) and GMM (b) fits on the scatter plot of the synthetic data. The visual inspection and the observed data log-likelihood values indicate that the GMCM fits the data much better than the GMM.

Table I
CONFUSION MATRIX RESULTING FROM THE UNSUPERVISED CLUSTERING PERFORMED USING (A) GMCM AND (B) GMM.

(a)

Predicted Actual	Cluster 1	Cluster 2	Cluster 3
Cluster 1	1999	1	0
Cluster 2	15	1975	10
Cluster 3	0	0	2000

(b)

Predicted Actual	Cluster 1	Cluster 2	Cluster 3
Cluster 1	1963	28	9
Cluster 2	15	2000	0
Cluster 3	0	536	1464

B. Image Segmentation

Initial experiments on some publically available images [17] show advantages of image segmentation using GMCs over GMMs. All the images were 481×321 in size resulting in total 154401 pixels. For simplicity, the segmentation was performed via clustering only in the RGB space. The

mixture models were learnt with multiple initializations and the model with the highest observed data likelihood was chosen to cluster the pixels. For the first image of a star fish, the GMCM (figure 5.1 b) was able to segment the circular pattern on the fish's body more clearly compared to a GMM (figure 5.1 c). Similar observation can be made for the second image, where the GMCM was able to distinguish between the vegetation and the mountain while the GMM grouped these pixels together into one class. Moreover, notice the over-segmentation of the sky, obtained using the GMM. The same region has been segmented into just one class by the GMCM. For the third image, the sky has been segmented cleanly and the patterns on the clothes of players are much more evident in the GMCM segmentation. Overall, a GMCM seems to identify intricate patterns in the images better than a GMM, which can be an important requirement for some image segmentation tasks. This quality is due to the ability of a GMCM to discriminate between the non-elliptical modes in a distribution, which a GMM may fail to do.

V. CONCLUSION

In this paper, we propose a new class of parametric mixture models based on copula functions. The main motivation is to characterize the probability distribution of a continuous, multivariate dataset with multiple non-Gaussian modes. For such datasets, Gaussian Mixture Models (GMMs) can prove to be too rigid because of the underlying assumption it imposes about the normality of the constituent modes. The proposed Gaussian Mixture Copula Model (GMCM), decouple the estimation of marginal distributions from the dependency structure present in different modes. This decoupling allows the marginal distributions to be estimated independently of each other, while borrowing the dependence structure from a GMM. We provide an Expectation-Maximization and a Gradient-based optimization algorithm for the estimation of GMCM parameters in the maximum likelihood setting. The benefits of GMCMs over GMMs are highlighted by preliminary experiments on a synthetic dataset and some real image datasets. Going forward, we intend to make the gradient-based optimization algorithm for GMCM parameter estimation more efficient. Currently, it suffers from the lack of a closed form expression of the objective function, which increases the computational overheads of the algorithm.

REFERENCES

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, pp. 15:1–15:58, July 2009. [Online]. Available: <http://doi.acm.org/10.1145/1541880.1541882>
- [3] J. Chen and J. Liu, "Mixture principal component analysis models for process monitoring," *Industrial and Engineering Chemistry Research*, vol. 38, no. 4, pp. 1478–1488, 1999. [Online]. Available: <http://pubs.acs.org/doi/abs/10.1021/ie980577d>
- [4] J. Yu and S. J. Qin, "Multimode process monitoring with bayesian inference-based finite gaussian mixture models," *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008. [Online]. Available: <http://dx.doi.org/10.1002/aic.11515>
- [5] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blob-world: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1026–1038, 1999.
- [6] M. S. Allili, N. Bouguila, and D. Ziou, "Finite general gaussian mixture modeling and application to image and video foreground segmentation," *J. Electron. Imaging*, vol. 17, pp. 013 005–1–013 005–13, 2008.
- [7] J. V. Arnaud and A. Doucet, "Maintaining multi-modality through mixture tracking," in *In ICCV*, 2003, pp. 1110–1116.
- [8] R. B. Nelsen, *An Introduction to Copulas (Lecture Notes in Statistics)*, 1st ed. Springer, Oct. 1998. [Online]. Available: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387986235>
- [9] A. Charpentier, J.-D. Fermanian, and O. Scaillet, "The estimation of copulas: Theory and practice," in *Copulas: From theory to application in finance*, J. Rank, Ed., 2007, pp. 35–62.
- [10] T. Schmidt, "Coping with copulas," in *Copulas: From theory to application in finance*, J. Rank, Ed., 2007, pp. 35–62.
- [11] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Manage. Sci.*, vol. 45, pp. 208–224, February 1999. [Online]. Available: <http://dx.doi.org/10.1287/mnsc.45.2.208>
- [12] D. X. Li. (2000) On default correlation: A copula function approach. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.8219>
- [13] A. Sklar, "Fonctions de rpartition n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959.
- [14] I. Kojadinovic and J. Yan, "Modeling multivariate distributions with continuous margins using the copula r package," *Journal of Statistical Software*, vol. 34, no. 9, pp. 1–20, 5 2010. [Online]. Available: <http://www.jstatsoft.org/v34/i09>
- [15] L. Xu and M. I. Jordan, "On convergence properties of the em algorithm for gaussian mixtures," *Neural Computation*, vol. 8, pp. 129–151, 1995.
- [16] Z. Botev, J. Grotowski, and D. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, vol. 38, pp. 2916–2957, 2010.

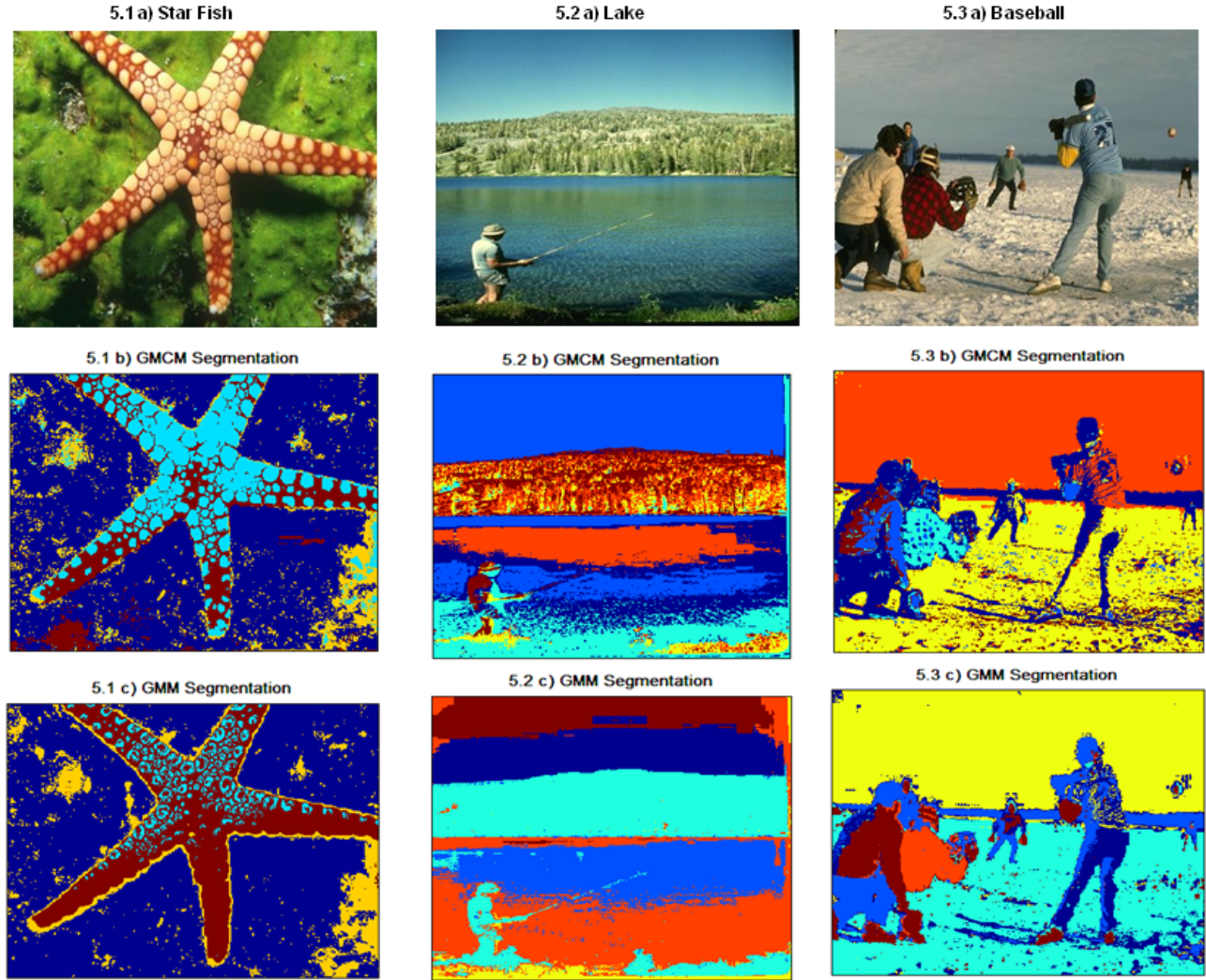


Figure 5. 5.1a) Star Fish (original image), 5.1b) GMCM segmentation ($M=4$), 5.1c) GMM Segmentation ($M=4$), 5.2a) Lake (original image), 5.2b) GMCM segmentation ($M=6$), 5.2c) GMM Segmentation ($M=6$), 5.3a) Baseball (original image), 5.3b) GMCM segmentation ($M=6$), 5.3c) GMM Segmentation ($M=6$)

- [17] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.