ROBUST IMAGE SEGMENTATION WITH MIXTURES OF STUDENT'S *t*-DISTRIBUTIONS

Giorgos Sfikas   *Christophoros Nikou*   *Nikolaos Galatsanos*

University of Ioannina,
Department of Computer Science,
PO Box 1185, 45110 Ioannina, Greece,
{sfikas, cnikou, galatsanos}@cs.uoi.gr

ABSTRACT

Gaussian mixture models have been widely used in image segmentation. However, such models are sensitive to outliers. In this paper, we consider a robust model for image segmentation based on mixtures of Student's *t*-distributions which have heavier tails than Gaussian and thus are not sensitive to outliers. The *t*-distribution is one of the few heavy tailed probability density functions (pdf) closely related to the Gaussian, that gives tractable maximum likelihood inference via the Expectation-Maximization (EM) algorithm. Numerical experiments that demonstrate the properties of the proposed model for image segmentation are presented.

Index Terms— Image segmentation, clustering, Student's *t*-distribution, mixture model, EM algorithm, segmentation evaluation.

1. INTRODUCTION

Image segmentation is the process of grouping image pixels based on the coherence of certain attributes such as intensity, color or texture. Many approaches have been proposed to solve the image segmentation problem. For surveys on this topic the reader may refer to [1]. In this paper, we will focus our attention to image segmentation methods based on clustering. Clustering is the process of arranging data into groups having common characteristics and is a fundamental problem in many fields of science [2]. Thus, image segmentation can be viewed a special type of clustering. Usually, in image segmentation, our data, the image pixels have spatial locations associated with them. Thus, apart from the commonality of attributes such as intensity, color or texture, commonality of location is an important characteristic of the grouping that we are seeking in image segmentation.

More specifically, in this paper we will focus our attention on clustering methods based on the modeling of the probability density function (pdf) of the data via finite mixture models (FMM) [3, 4]. Modeling the pdf of data with FMM is a natural way to cluster data because it automatically provides a grouping of the data based on the components of the mixture that generated them. More specifically, FMM are based on the assumption that each datum originates from one component of the mixture according to some probability. Thus, this probability can be used to assign each datum to the component that has most likely generated it. Furthermore, the likelihood of an FMM is a rigorous measure for evaluating the clustering performance [4].

FMM based pdf modeling with Gaussian components has been used successfully in a number of applications ranging from bioinformatics [5] to image retrieval [6]. The parameters of Gaussian mixture models (GMM) can be estimated very efficiently through maximum likelihood (ML) estimation using the EM algorithm [7]. Furthermore, it can be shown that Gaussian components allow efficient representation of any pdf [4]. However, it is well known that GMM are sensitive to outliers and may lead to excessive sensitivity to small numbers of data points. The problem of providing protection against outliers in multivariate data is very difficult and increases with the dimensionality.

In this paper, we apply to the image segmentation problem mixture models with Student-*t* pdf components. This pdf has heavier tails as compared to the exponentially decaying tails of a Gaussian [8]. Thus, each component in the mixture originates from a wider class of elliptically symmetric distributions with an additional parameter called the degrees of freedom. Hence a more robust model is used than the classical normal mixture.

In the remainder of this paper, background on standard GMM is given in Section 2 and the mixture of multivariate t-distributions and the EM algorithm for parameter estimation are described in Section 3. Results on image segmentation and comparisons with the standard GMM are presented in Section 4 and conclusions are drawn in Section 5.

2. BACKGROUND ON STANDARD GAUSSIAN MIXTURE MODELS

Let X denote the vector of features representing an image spatial location (pixel). The GMM assumes that the pdf of

the observation x is expressed by

$$\phi(x; \mathbf{\Omega}) = \sum_{i=1}^{K} \pi_i f(x; \mu_i, \Sigma_i) \qquad (1)$$

where $\mathbf{\Omega}$ is the mixture parameter set $\mathbf{\Omega} = [\Omega_1, \Omega_2, ..., \Omega_K]$ with $\Omega_i = (\pi_i, \mu_i, \Sigma_i)$. For the mixing proportions of the ith component π_i, we have that

$$0 \le \pi_i \le 1, \ \ i = 1, 2, ..., K, \quad \sum_{i=1}^{K} \pi_i = 1 \qquad (2)$$

For each component of the model in (1), the Gaussian pdf is expressed by

$$f(x; \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{-\frac{d}{2}} |\Sigma_i|^{-\frac{1}{2}}} exp^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)} \quad (3)$$

where d is the dimensionality of the vector (e.g. intensity, location, texture features) and μ_i, Σ_i are the mean vector and covariance matrix respectively.

Training of a GMM, or in other words finding its ML solution, can be performed using the EM algorithm [7]. The EM algorithm is a well-known numerical method used in a variety of ML problems. In the case of a GMM, each image pixel x, is associated with a binary hidden variable z of dimension K, whose k^{th} component has a value of 1 if the observation (i.e. the pixel) was produced by that component and is zero otherwise. In the E-step of the algorithm, the expected value of the hidden variables conditioned on the observation is computed. These expected values give the probabilities that a given datum originates from a different component of the mixture. Thus, they provide a means for segmenting the data. In the M-step, the model parameters (mean, covariance and mixing proportions) are computed by maximizing the log-likelihood of the complete data (hidden variables and observations). This scheme is repeated iteratively until convergence is achieved.

3. MIXTURE OF STUDENT'S *t*-DISTRIBUTIONS AND THE EM ALGORITHM

A d-dimensional random variable X follows a multivariate t-distribution with mean μ, positive definite, symmetric and real $d \times d$ covariance matrix Σ and has $\nu \in [0, \infty)$ degrees of freedom when, given the weight u, the variable X has the multivariate normal distribution with mean μ and covariance Σ/u:

$$X|\mu, \Sigma, \nu, u \sim N(\mu, \Sigma/u),$$

and the weight u follows a Gamma distribution parameterized by ν:

$$u \sim Gamma(\nu/2, \nu/2).$$

Integrating out the weights from the joint density leads to the density function of the marginal distribution:

$$p(x; \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right) |\Sigma|^{-\frac{1}{2}}}{(\pi\nu)^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) [1 + \nu^{-1}\delta(x, \mu; \Sigma)]^{\frac{\nu+d}{2}}} \quad (4)$$

where $\delta(x, \mu; \Sigma) = (x - \mu)^T \Sigma^{-1}(x - \mu)$ is the Mahalanobis squared distance and Γ is the Gamma function. It can be shown that for $\nu \to \infty$ the Student's t-distribution tends to a Gaussian distribution with covariance Σ. Also, if $\nu > 1$, μ is the mean of X and if $\nu > 2$, $\nu(\nu - 2)^{-1}\Sigma$ is the covariance matrix of X. Therefore, the family of t-distributions provides a heavy-tailed alternative to the normal family with mean μ and covariance matrix that is equal to a scalar multiple of Σ, if $\nu > 2$ (fig. 1).

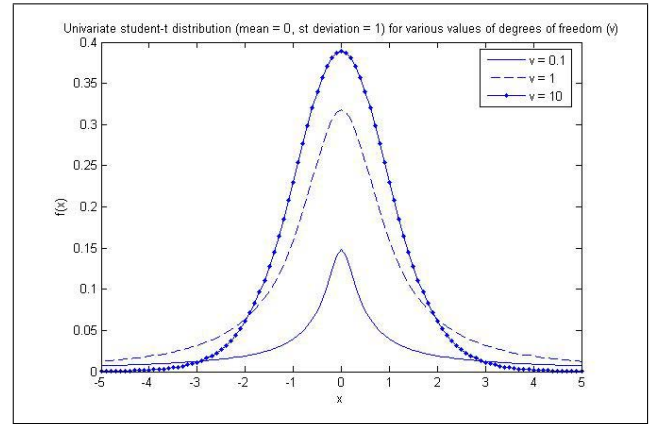**Fig. 1**. The Student's t-distribution for various degrees of freedom. As $\nu \to \infty$ the distribution tends to a Gaussian. For small values of ν the distribution has heavier tails than a Gaussian.

A Student's t-distribution mixture model (SMM) may also be trained using the EM algorithm [8]. A K-component mixture of t-distributions is given by

$$\phi(x, \Psi) = \sum_{i=1}^{K} \pi_i p(x; \mu_i, \Sigma_i, \nu_i) \qquad (5)$$

where $x = (x_1, ..., x_N)^T$ denotes the observed-data vector and

$$\Psi = (\pi_1, ..., \pi_K, \mu_1, ..., \mu_K, \Sigma_1, ..., \Sigma_K, \nu_1, ..., \nu_K)^T. \quad (6)$$

are the parameters of the components of the mixture.

Consider now the complete data vector

$$x_c = (x_1, ... x_N, z_1, ..., z_N, u_1, ..., u_N)^T \qquad (7)$$

where $z_1, ..., z_N$ are the component-label vectors and $z_{ij} = (z_j)_i$ is either one or zero, according to whether the observation x_j is generated or not by the ith component. In the light

I - 274

of the definition of the *t*-distribution, it is convenient to view that the observed data augmented by the $z_j, j = 1, ..., N$ are still incomplete because the component covariance matrices depend on the degrees of freedom. This is the reason that the complete-data vector also includes the additional missing data $u_1, ..., u_N$. Thus, the E-step on the $(t + 1)^{th}$ iteration of the EM algorithm requires the calculation of the posterior probability that the datum x_j belongs to the ith component of the mixture:

$$z_{ij}^{t+1} = \frac{\pi_i^t p(x_j; \mu_i^t, \Sigma_i^t, \nu_i^t)}{\sum_{m=1}^{K} p(x_j; \mu_m^t, \Sigma_m^t, \nu_m^t)} \quad (8)$$

as well as the expectation of the weights for each observation:

$$u_{ij}^{t+1} = \frac{\nu_i^t + d}{\nu_i^t + \delta(x_j, \mu_i^t; \Sigma_i^t)} \quad (9)$$

Maximizing the log-likelihood of the complete data provides the update equations of the respective mixture model parameters:

$$\pi_i^{t+1} = \frac{1}{N} \sum_{j=1}^{N} z_{ij}^t, \quad \mu_i^{t+1} = \frac{\sum_{j=1}^{N} z_{ij}^t u_{ij}^t x_j}{\sum_{j=1}^{N} z_{ij}^t u_{ij}^t}, \quad (10)$$

$$\Sigma_i^{t+1} = \frac{\sum_{j=1}^{N} z_{ij}^t u_{ij}^t (x_j - \mu_i^{t+1})(x_j - \mu_i^{t+1})^T}{\sum_{j=1}^{N} z_{ij}^{t+1}}. \quad (11)$$

The degrees of freedom for each component are computed as the solution to the equation:

$$\log \left(\frac{\nu_i^{t+1}}{2} \right) - \psi \left(\frac{\nu_i^{t+1}}{2} \right) + 1 - \log \left(\frac{\nu_i^t + d}{2} \right) +$$

$$+ \frac{\sum_{j=1}^{N} z_{ij}^t (\log u_{ij}^t - u_{ij}^t)}{\sum_{j=1}^{N} z_{ij}^t} + \psi \left(\frac{\nu_i^t + d}{2} \right) = 0 \quad (12)$$

where $\psi(x) = \frac{\partial(ln\Gamma(x))}{\partial x}$ is the digamma function. A detailed derivation of the EM algorithm for Student's *t*-mixtures is presented in [8].

4. EXPERIMENTAL RESULTS

In this paper, we employed an 8-dimensional vector as a feature for each image pixel [9]. The first three components of

Table 1. *Number of images (over 30) where the SMM provides lower quantization error than the GMM for $p = 1.2$.*

Noise type	$K = 3$	$K = 5$	$K = 7$
noise free	23	17	24
uniform 20 dB	20	19	21
uniform 14 dB	23	17	18
uniform 7 dB	26	15	17
salt pepper 10%	19	15	16

the feature vector are the *Lab* color coordinates, the next three components are texture descriptors, namely, the polarity, the anisotropy and the contrast as described in [9] and the remaining two coordinates are the horizontal and vertical pixel locations. Prior to model training, each feature vector component was separately normalized to ensure that no feature dominates the others.

In order to evaluate the proposed segmentation scheme and compare it to GMM segmentations we compute a quantization error for 30 images provided by the Berkeley image segmentation data base [10]. The quantization error, for each pixel location, is defined as the distance between the image feature and the mean of the mixture component that generated the measure (i.e. the component with the larger mixing proportion). This *p*-norme distance, between a *d*-dimensional feature vector x and the mean vector μ is defined as

$$D(x, \mu) = \left(\sum_{i=1}^{d} (x_i - \mu_i)^p \right)^{\frac{1}{p}} \quad (13)$$

We have experimented with different values of p, namely 0.7, 1.2 and 2 (Euclidean distance). It is well known that norms close to 1 measure a quantization error that better corresponds to human perceptual characteristics. These experiments were performed by degrading the images by uniform and salt-and-pepper noise of varying strength. Also, the predefined number of kernels varied ($K = 3, 5, 7$).

Let us also notice that the experiments were performed using a variation of the standard EM algorithm, called Greedy EM [11] providing a segmentation result independent of the model initialization. The performance of the model is presented in table 1. A comparison is shown in table 2 where one can see that the SMM has a slight yet better performance than the GMM. Some segmentation results are depicted in fig. 2 and 3 where it can be observed that SMM provide smoother segmentations than the standard GMM.

5. CONCLUSION

We have presented a methodology for image segmentation based on mixtures of Student's *t*-distributions. The model can account for outliers values and thus provides smoother

I - 275

Table 2. *Quantization error statistics for 30 images of the Berkeley segmentation data base for all the configurations of the uniform noise (see table 1).*

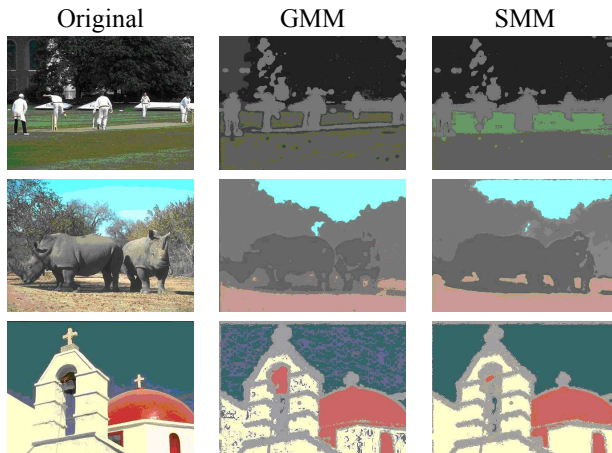|	K = 3		K = 5		K = 7	
	GMM	SMM	GMM	SMM	GMM	SMM
p = 0.7						
mean	11.54	11.54	10.27	10.23	9.59	9.51
s. d.	0.85	0.92	0.87	0.89	0.88	0.92
p = 1.2						
mean	3.85	3.83	3.48	3.47	3.27	3.25
s. d.	0.24	0.25	0.25	0.26	0.27	0.28
p = 2.0						
mean	2.26	2.25	2.07	2.07	1.96	1.96
s. d.	0.12	0.13	0.14	0.14	0.15	0.15

Original   GMM   SMM

Fig. 2. Segmentation examples using the GMM and the SMM methods for $K = 5$ components.
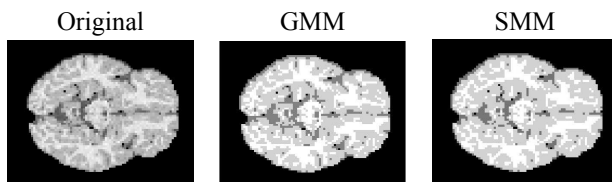

Original   GMM   SMM

Fig. 3. Segmentation of a MRI brain image into $K = 3$ classes (white matter, grey matter and cerebrospinal fluid).

segmentations than the standard GMM. However, important issues for mixture based clustering still need to be addressed. Such issues are how the number of model components can be selected automatically and which features should be used. These are open questions and are subject of current research.

6. REFERENCES

[1] N. Pal and S. Pal, "A review of image segmentation techniques," *Pattern Recognition*, vol. 26, pp. 1277–1294, 1993.

[2] R. Xu and D. Wunsch II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[4] G. McLachlan, *Finite mixture models*, Wiley-Interscience, 2000.

[5] K. Blekas, A. Likas, N. Galatsanos, and I. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.

[6] H. Greenspan, G. Dvir, and Y. Rubner, "Context-dependent segmentation and matching in image databases," *Computer Vision and Image Understanding*, vol. 93, no. 1, pp. 86–109, 2004.

[7] P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[8] D. Peel and G. J. McLachlan, "Robust mixture modeling using the *t*-distribution," *Statistics and Computing*, vol. 10, pp. 339–348, 2000.

[9] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.

[10] 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 *Proceedings of the 8th International Conference one Computer Vision*, July 2001, vol. 2, pp. 416–423.

[11] N. Vlassis and A. Likas, "A greedy EM algorithm for Gaussian mixture learning," *Neural Processing Letters*, vol. 15, pp. 77–87, 2002.