Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **3**, May **2018**

Gaussian Mixture Model and Its Applications in Semantic Image Analysis

Dongping Tian

Institute of Computer Software
Baoji University of Arts and Sciences
No.1 Hi-Tech Avenue, Hi-Tech District, Baoji, Shaanxi 721013, P.R. China

Institute of Computational Information Science
Baoji University of Arts and Sciences
No.44 Baoguang Road, Weibin District, Baoji, Shaanxi 721007, P.R. China
tiandp@ics.ict.ac.cn, tdp211@163.com

Received September, 2017; revised December, 2017

Abstract. *Semantic image analysis is an active topic of research in computer vision and pattern recognition. In the last two decades, a large number of works on semantic image analysis have emerged, among which Gaussian mixture model (GMM) is one of the most commonly used density models due to its potential flexibility and precision in modeling the underlying distributions of sub-band coefficients. However, compared with various GMM models and their corresponding applications in semantic image analysis, there is almost no review research and analysis about GMM related studies. So the current paper, to begin with, elaborates the basic principles of GMM, subsequently summarizes GMM with applications to image annotation, image retrieval, image classification and several other applications comprehensively. Finally, we develop a novel Gaussian mixture model fitted by the rival penalized expectation maximization (RPEM) algorithm for the task of automatic image annotation and retrieval. Conducted experiments on Corel5k dataset reveal that the proposed GMM can yield better results in terms of effectiveness and efficiency by using both the robust RPEM algorithm and the visual feature normalization method.*

Keywords: GMM, RPEM, Image annotation, CBIR, Image classification, Image retrieval

1. **Introduction.** With the advent and popularity of world wide web, the number of accessible digital images for various purposes is growing at an exponential speed. To make the best use of these resources, people need an efficient and effective tool to manage them. In such context, content-based image retrieval (CBIR) was introduced in the early 1990s. It heavily depends on the low-level features to find images relevant to the query concept, which is represented by the query example provided by the user. However, in the field of computer vision and multimedia processing, the semantic gap between low-level visual features and high-level semantic concepts is a major obstacle to CBIR related tasks. As a result, automatic image annotation (AIA) has appeared and become an active topic of research in computer vision for decades due to its potentially large impact on both image understanding and web image search. Specifically, AIA refers to a process to automatically generate textual keywords to describe the content of a given image, which plays a crucial role in semantic based image retrieval.

As the representative work of AIA, Li et al.[1] presented the automatic linguistic index for pictures. Duygulu et al.[2] put forward the translation model to treat AIA as a process of translation from a set of blob tokens to a set of keywords. Jeon et al.[3] proposed cross-media relevance model (CMRM) to annotate image, assuming the blobs and words were mutually independent given a specific image. Subsequently CMRM was improved through continuous space relevance model (CRM)[4] and multiple-Bernoulli relevance model (MBRM)[5]. In addition, Monay et al.[6] came up with the PLSA-WORDS model which allowed modeling of an image as a mixture of latent aspects that was defined by its text captions for which the conditional distributions over aspects were estimated only from the textual modality. In recent work [7], a supervised PLSA (S-PLSA) was constructed to improve image segmentation by using the classification results with an integrated framework based on PLSA and S-PLSA to accommodate segmentation and annotation procedures. A more recent work by Tian [8] constructed an extended PLSA for automatic image annotation through improving the traditional bag-of-visual-words model and applying the rival penalized competitive learning based method. Besides, the standard PLSA was extended to higher order for image indexing by treating images, visual features and tags as three observable variables of an aspect model [9] so as to learn a space of latent topics that incorporated semantics of both visual and tag information. Luo et al.[10] presented a new method for AIA based on Gaussian mixture model by region-based color and coordinate of matching to take into account the spatial relation among objects. In the meanwhile, GMM was employed to extract the texture based images and the query point movement technique was served as the relevance feedback for content-based image retrieval [11]. As briefly reviewed above, a large number of methods involving various models have been proposed for semantic image analysis, among which it should be noted that GMM is one of the most commonly used density models in that its potential flexibility in modeling the underlying distributions of sub-band coefficients. However, there is almost no review research and analysis about GMM related studies compared to various improved GMM models and their corresponding applications in the area of semantic image analysis. So we provide a comprehensive survey of GMM model that related to the semantic image analysis in the last decade. The primary purpose of this paper is to illustrate the effectiveness of GMM and how to further improve its applications in the field of computer vision and pattern recognition.

The remainder of this paper is organized as follows. Section 2 summarizes GMM with applications to automatic image annotation, image retrieval, image classification and several other applications, respectively. In Section 3, the basic principle of GMM is first introduced, followed by a novel GMM fitted by the rival penalized expectation-maximization algorithm is formulated for automatic image annotation and retrieval. Experimental results are reported and analyzed in Section 4. Finally, this paper is ended with a summary of some important conclusions and potential research directions of GMM in semantic image analysis for the future in Section 5.

2. GMM for Semantic Image Analysis.

In this section, Gaussian mixture model for semantic image analysis will be summarized from the aspects of image annotation, image retrieval, image classification and several other applications, respectively. More details can be gleaned from the following subsections.

2.1. GMM for image annotation.

As is well known, the performance of CBIR system is heavily dependent on the semantic annotation whether they can correctly describe the image content or not. Due to the traditional manual image annotation is time-consuming and labor-intensive. In contrast, automatic image annotation is a promising solution to enable the semantic-based image retrieval via keywords. AIA is desirable that images can be automatically labeled with linguistic terms so that both computers and human beings can be brought to the same ground of visual perception and the intrinsic semantic gap [12], to some extent, can be reduced or even eliminated. As the representative work, AIA was formulated as a supervised multi-class labeling problem by Yang et al.[13]. They exploited color and texture features to form two separate vectors, for which two independent Gaussian mixture models were estimated from the training set as the class densities by means of the EM algorithm in conjunction with a denoising technique. Wang et al.[14] proposed to build an effective visual vocabulary by using hierarchical GMM instead of the traditional clustering methods. Besides, PLSA was utilized to explore semantic aspects of visual concepts and discover topic clusters among documents and visual words. Subsequently a novel image annotation method was constructed by embedding GMM into the max-min posterior pseudo-probabilities [15]. It is generally believed that the spatial relation among objects is very important for image understanding and recognition. To this end, a recent work by Luo et al.[10] exploited a GMM based method for automatic image annotation via region-based color and coordinate of matching to take into account this factor. In more recent work [16], Jiu and Sahbi put forward a nonlinear deep multiple kernel learning (MKL) method for AIA. Extensive experiments on several challenging image collections validated its effectiveness and efficiency. Alternatively, it should be noted that the performance of AIA heavily depends on the features extracted, the normalization methods and the image segmentation approaches adopted. Table 1 summarizes some GMMs for automatic image annotation, including their sources, models adopted and image datasets exploited.

Table 1. Summary of GMM for image annotation

Sources	**Models adopted**	**Image datasets applied**
Luo et al.[10]	GMM	COREL Dataset
Sudhir et al.[11]	GMM, QPM	COREL Dataset
Yang et al.[13]	GMM, Bayesian classifier	TRECVID 2005
Wang et al.[14]	HGMM, PLSA, k-NN	TRECVID 2005
Wang et al.[15]	GMM, Bayesian classifier	COREL Dataset
Jiu et al.[16]	MKL,SVM	COREL/Banana Datasets

2.2. **GMM for image retrieval.** Gaussian mixture model has been extensively applied in many fields from image pattern recognition to text independent speaker recognition. There is no doubt that GMM has also been used in the field of image retrieval. Due to the explosive spread of digital devices, the amount of digital content grows rapidly. In CBIR system, images are usually indexed by their visual content, such as color, texture and shapes, etc. Sayad et al.[17] applied a multi-layer PLSA for image retrieval, in which the edge context descriptor was extracted by GMM as well as a spatial weighting scheme was constructed based on GMM to reflect the information about the spatial structure of the images. Based on the color histogram approximation, Piatek et al.[18] used GMM to retrieve all images whose color structure was similar to that of the given query image. Subsequently the generalized GMM was employed for CBIR [19]. Besides, GMM was introduced as a descriptor of the image color distribution for image indexing [20]. The main advantage was that it could overcome the problems associated with the high dimensionality of standard color histograms. Followed by they extended their previous work [21] by utilizing GMM working on color histograms built with weights delivered by the bilateral filter scheme. The proposed scheme enabled the retrieval system not only considered the global distribution of the color image pixels but also took into account their spatial arrangement. In addition, Sahbi [22] presented a new method for data clustering based on a particular GMM, etc.

Alternatively, GMM was also employed to model the feature distribution of the relevant images [23]. The statistical information of the GMM for a query was gathered during the relevance feedback process, and this information was exploited to estimate the probability density, i.e., relevance of an image to the query. This probability function was used to determine which images were more relevant to the query in the retrieval process. Recently, GMM was also exploited to model the target distribution of query where a novel idea to estimate the components of GMM was proposed based on the connected component analysis (CCA)[24]. Later on, Wan et al.[25] proposed a clustering based indexing approach called GMM-cluster forest to support multi-features based similarity search in high-dimensional spaces. In more recent work [11], GMM was leveraged to extract the texture based images as well as the query point movement technique was served as the relevance feedback in the task of CBIR. Note that Table 2 summarizes some GMMs aforementioned for image retrieval.

Table 2. Summary of GMM for image retrieval

Sources	**Models adopted**	**Image datasets applied**
Sayad et al.[17]	GMM, PLSA	Caltech101 Dataset
Luszczkiewicz [20]	GMM	Wang's Database
Luszczkiewicz [21]	GMM	WebMuseum Database
Sahbi [22]	GMM	Olivetti & Columbia Databases
Qian et al.[23]	GMM, Relevance feedback	COREL Dataset
Methre et al.[24]	GMM, CCA	COREL Dataset
Wan et al.[25]	GMM	COREL Dataset

2.3. **GMM for image classification.** Image classification refers to the procedure of labeling images into one of a number of predefined categories, which is a basic problem in many applications such as image annotation and object recognition. Image classification is still a challenging problem in computer vision although it has been studied for many years. Over the last two decades, a substantial amount of researches have been devoted to the problem of image classification [26-33]. In [26], Permuter et al. introduced GMM models of structure and color features so as to classify colored textures in images with a view to the retrieval of textured color images from databases. Wu et al.[27] put forward an image texture classification method based on finite GMM of sub-band coefficients, in which the Gaussian component parameters were estimated by an EM+MML algorithm, and the earth mover's distance (EMD) was used to measure the distributional similarity based on the Gaussian components. Meanwhile, GMM was also used as a supervised classifier for remote sensing multi-spectral images [28]. In recent years, a more general formulation of Bayesian adaptation was proposed in [30], which targeted class adaptation and could be applicable to generative and discriminative strategies for the problem of image classification. In both cases, a global GMM was first adapted to each class by using a Bayesian extension of the EM algorithm. In a more recent work [31], a GMM-based approach was developed to estimate the traffic speeds and to classify the length-based vehicle volume data by using single-loop outputs. Besides, an adaptive EM algorithm was proposed for Gaussian mixture model to segment an image according to local color and texture features extracted from discrete cosine transform coefficients [32]. In literature [33], Wei et al. came up with a novel hypotheses-CNN-pooling (HCP) framework to address the multi-label image classification problem, etc. As can be seen from the above reviews, GMM is a promising approach that has been widely applied in image classification, and most of them are able to achieve encouraging performance. In the following, several GMMs for image classification involved in this paper are succinctly summarized in Table 3, including their methods and test datasets employed in the corresponding literatures.

Table 3. Summary of GMM for image classification

Sources	**Models adopted**	**Image datasets applied**
Permuter et al.[26]	GMM	VisTex Database
Wu et al.[27]	GMM, EMD	Brodatz Texture Album
Akbas et al.[29]	GMM, SVM	Oliva & Torralba Dataset
Dixit et al.[30]	GMM, SVM	UIUC-sports,LabelMe,15-scenes

Note:

1. COREL Dataset: *http : //vision.sista.arizona.edu/kobus/research/data/eccv* 2002*/index.html*.
2. TRECVID 2005: *http : //www − nlpir.nist.gov/projects/trecvid/*.
3. Caltech101 Dataset: *http : //www.vision.caltech.edu/Image_Datasets/Caltech*101*/*.
4. Wang's Database: *http : //wang.ist.psu.edu/docs/related/*.
5. Web-museum Database: *http : //www.ibiblio.org/wm/*.
6. NIST99 & NIST02: *http : //www.nist.gov/speech/tests/spk/*.
7. Multi-temporal Dataset 1: *http : //earth.esa.int/cgi−bin/satimgsql.pl?show_url* = 1738&*startframe* = 0.
8. Multi-temporal Dataset 2: *http : //change.gsfc.nasa.gov/alaska.html*.
9. TRECVID2010: *http : //www − nlpir.nist.gov/projects/tv*2010*/tv*2010*.html*.
10. TRECVID2011: *http : //www − nlpir.nist.gov/projects/tv*2011*/tv*2011*.html*.
11. NIST1999: *http : //www.nist.gov/speech/tests/spk/*.
12. Oliva & Torralba Dataset: *http : //people.csail.mit.edu/torralba/code/spatialenvelope/*.
13. LabelMe: *http : //www.csail.mit.edu/node/*127.
14. UIUC-sports: *http : //vision.stanford.edu/lijiali/event_dataset/*.&
15. UCID Database: *http : //vision.cs.aston.ac.uk/datasets/UCID/ucid.html/*.
16. Coil100 Database: *http : //www.cs.columbia.edu/CAVE/software/softlib/coil−*100*.php*.

2.4. **GMM for other applications.** Except for the content aforementioned, GMM has also been applied to deal with other problems. Representative work includes the hierarchical Gaussian mixture model (HGMM) for speaker verification [34]. In particular, Bredin et al.[35] proposed to apply support vector machine directly in the GMM space for face recognition. Celik [36] exploited GMM and genetic algorithm (GA) for unsupervised change detection in multi-temporal satellite images of the same scene. In the meantime, GMM was also used for automatic selection of regions of interest for functional brain images [37] that could relieve the so-called small size sample problem in the classification of functional brain images for the diagnosis of Alzheimer's disease. In particular, Beecks et al.[38] modeled image similarity between Gaussian mixture models by making use of the signature quadratic form distance (SQFD). More recently, the combination of GMM super-vector and SVM was proposed for event detection [39,40]. Additionally, GMM has also been applied in many other fields [41-43] and for more details of them please refer to the corresponding literatures.

TABLE 4. Summary of GMM for other related applications

Sources	**Models adopted**	**Image datasets applied**
Liu et al.[34]	HGMM	NIST99 & NIST02 Corpus
Bredin et al.[35]	GMM, SVM	BANCA Database
Celik [36]	GMM, GA	Multi-temporal Datasets
Segovia et al.[37]	GMM, SVM	SPECT & PET Images
Beecks et al.[38]	GMM, SQFD	UCID, Coil100, Wang's Database
Inoue et al.[39]	GMM, SVM	TRECVID2010
Kamishima et al.[40]	GMM, SVM	TRECVID2010/2011
Si et al.[41]	GMM, PLSA	NIST1999

3. The Proposed GMM Model.

Note that the basic principle of GMM is first introduced in this section. Subsequently the proposed GMM that is fitted by the rival penalized expectation-maximization algorithm is elaborated.

3.1. Gaussian mixture model.

Gaussian mixture model has been proposed as a general model for estimating an unknown probability density function or simply density. In general, A GMM is a parametric statistical model which assumes that the data originates from a weighted sum of several Gaussian sources. More formally, a GMM is a weighted sum of M component Gaussian densities as given by the following equation.

$$p(x|\lambda) = \sum_{i=1}^{M} \omega_i g(x|\mu_i, \Sigma_i) \tag{1}$$

where x is a D-dimensional continuous-valued data vector, w_i, $i = 1, 2, \cdots, M$, denote the mixture weights, $g(x|\mu_i, \Sigma_i)$, $i = 1, 2, \cdots, M$, are the component Gaussian densities. Each component density can be represented as a D-variate Gaussian function:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right\} \tag{2}$$

with mean vector μ_i and covariance matrix Σ_i. The mixture weights satisfy the constraint, i.e., sum to 1. The component number of the GMM M is determined by the minimum description length principle. And the parameter set λ of the GMM is estimated by means of the maximum likelihood estimation which is performed via the expectation maximization algorithm due to its simplicity and effectiveness.

3.2. Our proposed GMM model.

There are several techniques available for estimating the parameters of a GMM. By far the most popular and well-established method is the maximum likelihood (ML) estimation, whose aim is to find the model parameters that maximize the likelihood of the GMM given the training data. But in general, the EM algorithm is used to fit GMM due to the infeasibility of direct maximization for ML. However, there is no penalty for the redundant mixture components based on the EM algorithm, which means that the number of components in a GMM cannot be automatically determined and has to be assigned in advance. To this end, the rival penalized expectation-maximization [44] is utilized to determine the number of components as well as to estimate the model parameters. Since RPEM introduces unequal weights into the conventional likelihood, the weighted likelihood can be written as below:

$$Q(\lambda, \mathrm{x}) = \frac{1}{N}\sum_{i=1}^{N}\log p(\mathrm{x}_i|\lambda) = \frac{1}{N\zeta}\sum_{i=1}^{N}\ell(\mathrm{x}_i;\lambda) \tag{3}$$

with

$$\ell(\mathrm{x}_i;\lambda) = \sum_{j=1}^{M} g(j|\mathrm{x}_i,\lambda)\log[\omega_j p(\mathrm{x}_i|\mu_j,\Sigma_j)] - \sum_{j=1}^{M} g(j|\mathrm{x}_i,\lambda)\log h(j|\mathrm{x}_i,\lambda) \tag{4}$$

where $h(j|\mathrm{x}_i,\lambda) = \omega_j p(\mathrm{x}_i|\mu_j,\Sigma_j)/p(\mathrm{x}_i|\lambda)$ is the posterior probability that x_i belongs to the j-th component in the mixture, λ is a positive constant, $g(j|\mathrm{x}_i,\lambda)$, $j = 1, 2, \cdots, M$, are designable weight functions, satisfying the following constraints:

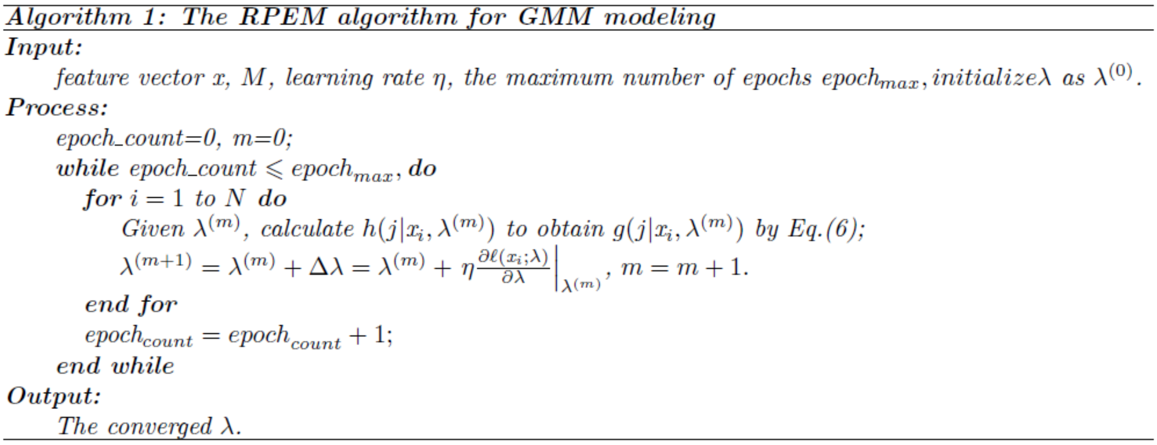
$$\sum_{j=1}^{M} g(j|\mathrm{x}_i,\lambda) = \zeta, 1 \leqslant i \leqslant N \tag{5}$$

in which $g(j|\mathrm{x}_i,\lambda)$ can be further constructed as follows:

$$g(j|\mathrm{x}_i,\lambda) = (1+\varepsilon_i)I(j|\mathrm{x}_i,\lambda) - \varepsilon_i h(j|\mathrm{x}_i,\lambda) \tag{6}$$

where $I(j|\mathrm{x}_i,\lambda)$ equals to 1 if $j = argmax_{1\leqslant i\leqslant M} h(i|\mathrm{x},\lambda)$ and 0 otherwise. ε_i is a small positive quantity.

So far, the major steps of RPEM algorithm can be summarized as below:



```
Algorithm 1: The RPEM algorithm for GMM modeling
Input:
    feature vector $x$, $M$, learning rate $\eta$, the maximum number of epochs $epoch_{max}$, initialize$\lambda$ as $\lambda^{(0)}$.
Process:
    epoch_count=0, m=0;
    while epoch_count $\leqslant epoch_{max}$, do
      for $i = 1$ to $N$ do
        Given $\lambda^{(m)}$, calculate $h(j|x_i, \lambda^{(m)})$ to obtain $g(j|x_i, \lambda^{(m)})$ by Eq.(6);
        $\lambda^{(m+1)} = \lambda^{(m)} + \Delta\lambda = \lambda^{(m)} + \eta\frac{\partial\ell(x_i;\lambda)}{\partial\lambda}\Big|_{\lambda^{(m)}}$, $m = m + 1$.
      end for
      $epoch_{count} = epoch_{count} + 1$;
    end while
Output:
    The converged $\lambda$.
```

Based on the Gaussian mixture model and RPEM algorithm described above, GMM is first trained and utilized to characterize the semantic model of the given concepts by Eq.(1). Assume that the training image is represented by both a visual feature $X = \{x_1, x_2, \cdots, x_m\}$ and a keyword list $W = \{w_1, w_2, \cdots, w_n\}$, where $x_i (i = 1, 2, \cdots, m)$ denotes the visual feature for region i and $w_j (j = 1, 2, \cdots, n)$ is the j-th keyword in the annotation. For a test image I represented by its visual feature vector $X = \{x_1, x_2, \cdots, x_m\}$, according to Bayesian rule, the posterior probability $p(w_i|I)$ can be calculated based on the conditional probability $p(I|w_i)$ and prior probability $p(w_i)$ as follows:

$$p(w_i|X) \infty \prod_{j=1}^{m} p(x_j|w_i)p(w_i) \tag{7}$$

From Eq.(7), the top n keywords can be selected as the annotations for unseen image X.

In addition, it is worth noting that during the course of image feature extraction, different kinds of features may have different magnitudes. How to appropriately normalize these features plays a crucial role in the subsequent image processing. Based on this recognition, we propose to employ the Gaussian normalization method [45] for image feature normalization. Let $F_i = (f_{i1}, \cdots, f_{ik}, \cdots, f_{iq})$ be the feature vector representing the i-th image region. The mean μ_k and standard deviation σ_k of the k-th feature dimension can be easily calculated. Subsequently the feature vectors can be normalized to $N(0,1)$ according to:

$$F_i = \left(\frac{f_{i1} - \mu_1}{k\sigma_1}, \cdots, \frac{f_{ik} - \mu_k}{k\sigma_k}, \cdots, \frac{f_{iq} - \mu_q}{k\sigma_q} \right) = (f_{i1}^{'}, \cdots, f_{ik}^{'}, \cdots, f_{iq}^{'}) \tag{8}$$

In Eq.(8), assume that each feature is normally distributed and $k = 3$. According to the 3-σ rule, the probability of an entry's value being in the range of $[-1, 1]$ is approximately 99%. By defining the following Eq.(9), namely, a simple additional shift embedded can guarantee that 99% of the feature values will be within [0,1].

$$F_i = \left(\frac{f_{i1}^{'} + 1}{2}, \cdots, \frac{f_{ik}^{'} + 1}{2}, \cdots, \frac{f_{iq}^{'} + 1}{2} \right) \tag{9}$$

where each $f_{i1}^{'}$, $f_{ik}^{'}$, $f_{iq}^{'}$ represents a normalized feature vector within [-1,1].

4. **Experimental Results and Analysis.** To evaluate the performance of the proposed GMM with RPEM algorithm (abbreviated as GMM-RPEM), we test it on the Corel5k dataset which is broadly adopted as basic comparative data for recent research work in the image annotation community. Corel5k consists of 5000 images from 50 Corel Stock Photo CD's. Each CD contains 100 images with a certain theme (e.g. polar bears), of which 90 are designated to be in the training set and 10 in the test set, resulting in $4,500$ training images and a balanced 500-image test collection. Besides, the dictionary contains 260 words that appear in both the training and testing set. As for image segmentation, it is worth noting that the normalized cuts algorithm (Ncuts) [46] rather than JSEG [47] is applied to segment images into a number of meaningful regions. The main reason lies in that JSEG only focuses on local features and their consistencies while Ncuts aims at extracting the global impression of an image data. So Ncuts, to some extent, can get a better segmentation result than that of JSEG (As can be seen from Fig. 1). For each image at most the 10 largest regions are selected and 809-dimensional visual features (color, texture, shape and saliency)[1] are extracted for each region, which include 81-dimensional grid color features, 59-dimensional local binary pattern texture features, 120-dimensional Gabor wavelets texture features, 37-dimensional edge orientation histogram features and 512-dimensional GIST features, respectively.

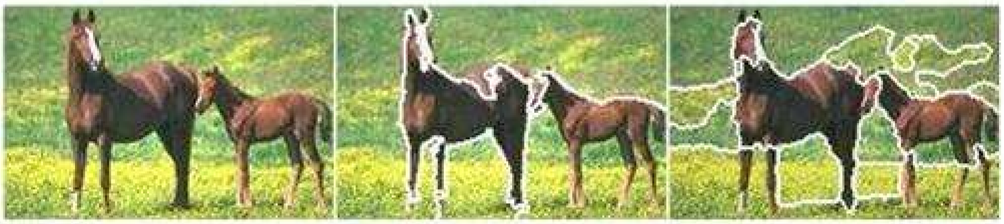


FIGURE 1. The segmentation results using Ncuts (mid) and JSEG (right)

To make a fair comparison with other AIA methods, the most commonly used metrics precision, recall and F-value of every word in the test set are calculated and the

[1] http://appsrv.cse.cuhk.edu.hk/~jkzhu/felib.html.

mean of these values is applied to summarize the model performance: recall=B/C and precision=B/A, where A is the number of images automatically annotated with a given keyword in the top 5 returned word list, B is the number of images correctly annotated with that keyword in the top 5 returned word list, and C is the number of images having that keyword in ground truth annotation. F-value=$2 \times \mathit{precision} \times \mathit{recall}/(\mathit{precision} + \mathit{recall})$. Besides, the mean average precision (mAP) is used to evaluate the retrieval performance of our model.

$$mAP = \frac{1}{N_w} \sum\nolimits_{w=1}^{N_w} AP(w) \tag{10}$$

with

$$AP(w) = \frac{\sum_{i \in relevant} precision(i)}{rel(w)} \tag{11}$$

Note that the AP of a query w is defined as the sum of the precisions of the correctly retrieved images at rank i divided by the total number of relevant images rel(w) for this query.

To show the effectiveness of GMM-RPEM proposed in this paper, we compare it with several previous approaches [2-4,48]. The experimental results listed in Table 5 are based on two sets of words: the subset of 49 best words and the complete set of all 260 words that occur in the training set. From Table 5, it is clear to see that GMM-RPEM outperforms all the others, especially the first three approaches. Meanwhile, it is also superior to CRM and GMM-EM (GMM with EM algorithm) by the gains of 2 and 4 words with non-zero recall, 11% and 5% mean per-word recall together with 13% and 13% mean per-word precision on the set of 260 words respectively. In addition, compared to GMM-EM on the set of 49 best words, we can also get improvement in mean per-word precision despite the mean per-word recall of GMM-RPEM is the same as that of GMM-EM.

Table 5. Performance comparison on Corel5k dataset

Models	Co-occurence	TM	CMRM	CRM	GMM-EM	GMM-RPEM
#words with recall>0	19	49	66	107	105	109
Results on 49 best words						
Mean per-word recall	-	0.34	0.48	0.70	0.71	0.71
Mean per-word precision	-	0.20	0.40	0.59	0.58	0.61
F-value	-	0.252	0.436	0.640	0.638	0.656
Results on all 260 words						
Mean per-word recall	0.02	0.04	0.09	0.19	0.20	0.21
Mean per-word precision	0.03	0.06	0.10	0.16	0.16	0.18
F-value	0.024	0.048	0.095	0.174	0.178	0.194

Fig. 2 shows some annotation results (only four cases are listed here due to the limited space) generated by GMM-EM and GMM-RPEM models, respectively. Note that the reranked and new words compared to the annotations yielded by GMM-EM and the ground truth are underlined and italicized respectively. In addition, to validate the retrieval performance of our model proposed in this paper, mean average precision (mAP) is also applied as a metric to evaluate the performance of single word retrieval, which has been a standard measure for the retrieval of text document for years and it has the ability to summarize the retrieval performance in a meaningful way. Here, we only compare our model with CMRM, CRM and PLSA-WORDS due to mAP of other methods cannot be accessed directly. As shown in Table 6, GMM-RPEM is obviously superior to CMRM. Compared with CRM and PLSA-WORDS, it can get 8% and 18% improvements on 260 words as well as 11% and 15% on words with positive recall, respectively, which further demonstrates the effect of the GMM-RPEM model for the task of image retrieval.

Table 6. Ranked image retrieval results on Corel5k dataset

Models	All 260 words	Words with recall >0
CMRM	0.17	0.20
CRM	0.24	0.27
PLSA-WORDS	0.22	0.26
GMM-RPEM	0.26	0.30

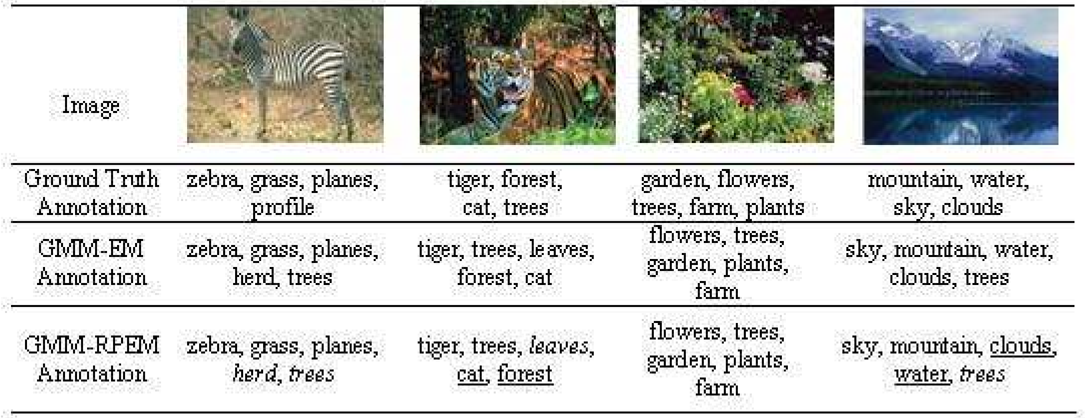


Image				
Ground Truth Annotation	zebra, grass, planes, profile	tiger, forest, cat, trees	garden, flowers, trees, farm, plants	mountain, water, sky, clouds
GMM-EM Annotation	zebra, grass, planes, herd, trees	tiger, trees, leaves, forest, cat	flowers, trees, garden, plants, farm	sky, mountain, water, clouds, trees
GMM-RPEM Annotation	zebra, grass, planes, *herd*, *trees*	tiger, trees, *leaves*, cat, forest	flowers, trees, garden, plants, farm	sky, mountain, clouds, water, *trees*

Figure 2. Annotation comparison with GMM-EM and GMM-RPEM on Corel5k dataset

5. **Conclusions and Future Work.** In this paper, we first present a comprehensive survey on GMM related studies in semantic image analysis from the aspects of image annotation, image retrieval, image classification and several other applications, respectively. Followed by we develop a novel GMM fitted by the RPEM algorithm for image annotation, especially the introduced Gaussian normalization method for image feature normalization. Conducted experiments validate its effectiveness and efficiency. The primary purpose of this paper is to illustrate the pros and cons of GMM combined with a great deal of existing researches as well as to point out the promising research directions of GMM for semantic image analysis in the future.

As for future work, GMM should be applied in wider ranges to deal with more multimedia related tasks, such as speech recognition, video recognition, action recognition, music information retrieval and other multimedia event detection tasks, etc. At the same time, it is worth noting that the parallelization of GMM model to very large-scale multimedia datasets is also an important issue to be further studied, especially in the current circumstances of cloud computing, cloud services, hadoop, smartwatch, fingerprint password, web of things, 3D printing and deep learning techniques, etc. In addition, it should be noted that the following several issues remain to be investigated. First, due to the classic GMM has limitation in its modeling abilities as all data points of an object are required to be generated from a pool of mixtures with the same set of mixture weights. In other words, GMM assumes that all data points are generated from a set of Gaussian models with the same set of mixture weights. So how to determine the weight factors of GMM more appropriately is a worthy research direction. Second, how to speed up the GMM estimation with EM algorithm is also an important work for large-scale multimedia processing tasks. Third, due to the complementary performance of hybridizing two or more machine learning techniques together, which can usually make them benefit from each other. Based on this recognition, how to efficiently integrate GMM with other methods based on the trade-off between computational complexity and model reconstruction error is a also valuable research direction in the future. Fourth, it is worthwhile to find studies of the influence of the order of the GMM on the quality of the classification. Last but not the least, since the assumptions made to build a GMM include: priori probabilities for each class are known, samples come from a known number of classes, the forms of the class-conditional probability densities are known for all classes, and the unknowns are the values of the several parameter vectors. So how to relax these assumptions of GMM to a certain extent as well as to introduce the semantic contexts of images and keywords is well worth exploring and pursuing.

Acknowledgment. The author would like to sincerely thank the editor and anonymous reviewers for their valuable comments and insightful suggestions that have helped to improve the paper. This work is partially supported by the National Program on Key Basic Research Project (No.2013CB329502), National Natural Science Foundation of China (No.61035003, No.61202212), and Key Research Project of Baoji University of Arts and Sciences (No.ZK2018061).

REFERENCES

[1] J. Li and J. Wang, Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[2] P. Duygulu, K. Barnard, N. de Freitas, et al., Object recognition as machine translation: learning a lexicon for a fixed image vocabulary, *Proc. of the European Conf. on Computer Vision (ECCV'02)*, pp. 97–112, 2002.

[3] L. Jeon, V. Lavrenko and R. Manmantha, Automatic image annotation and retrieval using cross-media relevance model, *Proc. of the 26th Int'l Conf. on Research and Development in Information Retrieval (SIGIR'03)*, pp. 119–126, 2003.

[4] V. Lavrenko, R. Manmatha and J. Jeon, A model for learning the semantics of pictures, *Advances in Neural Information Processing Systems 16 (NIPS'03)*, pp. 553–560, 2003.

[5] S. Feng, R. Manmatha and V. Lavrenko, Multiple Bernoulli relevance models for image and video annotation, *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition (CVPR'04)*, pp. 1002–1009, 2004.

[6] F. Monay and D. Gatica-Perez, Modeling semantic aspects for cross-media image indexing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1802–1817, 2007.

[7] Q. Guo, N. Li, Y. Yang, et al., Integrating image segmentation and annotation using supervised PLSA, *Proc. of the 20th Int'l Conf. on Image Processing (ICIP'13)*, pp. 3800–3804, 2013.

[8] D. Tian, Extended probabilistic latent semantic analysis for automatic image annotation, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 4, pp. 903–915, 2017.

[9] S. Nikolopoulos, S. Zafeiriou, I. Patras, et al., High order PLSA for indexing tagged images, *Signal Processing*, vol. 93, no. 8, pp. 2212–2228, 2013.

[10] X. Luo and K. Kita, Region-based image annotation using Gaussian mixture model, *Proc. of the 2nd Int'l Conf. on Information Technology and Software Engineering (ITSE'13)*, pp. 503–510, 2013.

[11] R. Sudhir and S. Baboo, An efficient content based image retrieval system using GMM and relevance feedback, *International Journal of Computer Applications*, vol. 72, no. 22, pp. 50–61, 2013.

[12] Y. Liu, D. Zhang, G. Lu, et al., A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.

[13] F. Yang, F. Shi and Z. Wang, An improved GMM-based method for supervised semantic image annotation, *Proc. of the IEEE Int'l Conf. on Intelligent Computing and Intelligent Systems (ICIS'09)*, pp. 506–510, 2009.

[14] Z. Wang, H. Yi, J. Wang, et al., Hierarchical Gaussian mixture model for image annotation via PLSA, *Proc. of the 5th Int'l Conf. on Image and Graphics (ICIG'09)*, pp. 384–389, 2009.

[15] Y. Wang, X. Liu and Y. Jia, Automatic image annotation with cooperation of concept-specific and universal visual vocabularies, *Proc. of the 16th Int'l Conf. on Multimedia Modeling (MMM'10)*, pp. 262–272, 2010.

[16] M. Jiu and H. Sahbi, Nonlinear deep kernel learning for image annotation, *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1820–1832, 2017.

[17] I. Sayad, J. Martinet, T. Urruty, et al., Toward a higher-level visual representation for content-based image retrieval, *Multimedia Tools and Applications*, vol. 60, no. 2, pp. 455–482, 2012.

[18] M. Piatek and B. Smolka, Effective color image retrieval based on the Gaussian mixture model, *Proc. of the 3rd Int'l Computational Color Imaging Workshop (CCIW'11)*, pp. 199–213, 2011.

[19] L. Raju, K. Vasantha and Y. Srinivas, Content based image retrievals based on generalization of GMM, *International Journal of Computer Science and Information Technologies*, vol. 3, no. 6, pp. 5326–5330, 2012.

[20] M. Luszczkiewicz and B. Smolka, Gaussian mixture model based retrieval technique for lossy compressed color images, *Proc. of the 4th Int'l Conf. on Image Analysis and Recognition (ICIAR'07)*, pp. 662–673, 2007.

[21] M. Luszczkiewicz and B. Smolka, Application of bilateral filtering and Gaussian mixture modeling for the retrieval of paintings, *Proc. of the 16th Int'l Conf. on Image Processing (ICIP'09)*, pp. 77–80, 2009.

[22] H. Sahbi, A particular Gaussian mixture model for clustering and its application to image retrieval, *Soft Computing*, vol. 12, no. 7, pp. 667–676, 2008.

[23] F. Qian, M. Li, L. Zhang, et al., Gaussian mixture model for relevance feedback in image retrieval, *Proc. of the Int'l Conf. on Multimedia and Expo (ICME'02)*, pp. 229–232, 2002.

[24] R. Methre and C. Bhagvati, Connected component method to find components of GMM in image retrieval, *Proc. of the Int'l Conf. on Computational Intelligence and Communication Networks (CICN'10)*, pp. 50–54, 2010.

[25] Y. Wan, X. Liu, K. Tong, et al., GMM-ClusterForest: a novel indexing approach for multi-features based similarity search in high-dimensional spaces, *Proc. of the 19th Int'l Conf. on Neural Information Processing (ICONIP'12)*, pp. 210–217, 2012.

[26] H. Permuter, J. Francos and H. Jermyn, Gaussian mixture models of texture and color for image database retrieval, *Proc. of the 28th Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03)*, pp. 569–572, 2003.

[27] Y. Wu, K. Chan and H. Wang, Image texture classification based on finite Gaussian mixture model, *Proc. of the IEEE Int'l Workshop on Texture Analysis and Synthesis*, pp. 1–5, 2003.

[28] A. Melo, R. Moraes and L. Machado, Gaussian mixture models for supervised classification of remote sensing multispectral images, *Proc. of the 8th Iberoamerican Congress on Pattern Recognition (CIARP'03)*, pp. 440–447, 2003.

[29] E. Akbas and N. Ahuja, Low-level image segmentation based scene classification, *Proc. of the 20th Int'l Conf. on Pattern Recognition (ICPR'10)*, pp. 3623–3626, 2010.

[30] M. Dixit, N. Rasiwasia and N. Vasconcelos, Adapted Gaussian models for image classification, *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, pp. 937–943, 2011.

[31] Y. Lao, G. Zhang, J. Corey, et al., Gaussian mixture model-based speed estimation and vehicle classification using single-loop measurements, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 16, no. 4, pp. 184–196, 2012.

[32] H. Al-Jubouri, H. Du, H. Sellahewa, Applying Gaussian mixture model on discrete cosine features for image segmentation and classification, *Proc. of the 4th Computer Science and Electronic Engineering Conference (CEEC'12)*, pp. 194–199, 2012.

[33] Y. Wei, W. Xia, M. Lin, et al., HCP: a flexible CNN framework for multi-label image classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2016.

[34] M. Liu, E. Chang and B. Dai, Hierarchical Gaussian mixture model for speaker verification, *Proc. of the 7th Int'l Conf. on Spoken Language Processing (ICSLP'02)*, pp. 1353–1356, 2002.

[35] H. Bredin, N. Dehak and G. Chollet, GMM-based SVM for face recognition, *Proc. of the 18th Int'l Conf. on Pattern Recognition (ICPR'06)*, pp. 1111–1114, 2006.

[36] T. Celik, Image change detection using Gaussian mixture model and genetic algorithm, *Journal of Visual Communication and Image Representation*, vol. 21, no. 8, pp. 965–974, 2010.

[37] F. Segovia, J. Grriz, J. Ramrez, et al., Classification of functional brain images using a GMM-based multi-variate approach, *Neuroscience Letters*, vol. 474, no. 1, pp. 58–62, 2010.

[38] C. Beecks, A. Ivanescu, S. Kirchhoff, et al., Modeling image similarity by Gaussian mixture models and the signature quadratic form distance, *Proc. of the 13th Int'l Conf. on Computer Vision (ICCV'11)*, pp. 1754–1761, 2011.

[39] N. Inoue and K. Shinoda, A fast and accurate video semantic-indexing system using fast MAP adaptation and GMM super-vectors, *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1196–1205, 2012.

[40] Y. Kamishima, N. Inoue and K. Shinoda, Event detection in consumer videos using GMM super-vectors and SVMs, *EURASIP Journal on Image and Video Processing*, vol. 51, pp. 1–13, 2013.

[41] L. Si and R. Jin, Adjusting mixture weights of Gaussian mixture model via regularized probabilistic latent semantic analysis, *Proc. of the 9th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'05)*, pp. 622–631, 2005.

[42] S. Mitra, Gaussian mixture models for human face recognition under illumination variations, *Applied Mathematics*, vol. 3, no. 12, pp. 2071–2079, 2012.

[43] Y. Wang, W. Chen, J. Zhang, et al., Efficient volume exploration using the Gaussian mixture model, *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 11, pp. 1560–1573, 2011.

[44] Y. Cheung, Maximum weighted likelihood via rival penalized EM for density mixture clustering with automatic model selection, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 750–761, 2005.

[45] Y. Rui, T. Huang, M. Ortega, et al., Relevance feedback: a power tool for interactive content-based image retrieval, *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, 1998.

[46] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[47] Y. Deng and B. Manjunath, Unsupervised segmentation of color-texture regions in images and video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, 2001.

[48] Y. Mori, H. Takahashi and R. Oka, Image-to-word transformation based on dividing and vector quantizing images with words, *Proc. of the 1st Int'l Workshop on Multimedia Intelligence Storage and Retrieval Management (MISRM'99)*, pp. 405–409, 1999.