Image Texture Classification Based on Finite Gaussian Mixture Models

Yiming Wu, Kap Luk Chan, Yong Huang
School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore, 639798
Email: {pg03875203, eklchan, p146457665}@ntu.edu.sg

***Abstract*— A novel image texture classification method based on finite Gaussian mixture models of sub-band coefficients is proposed in this paper. In the method, an image texture is first decomposed into several sub-bands, then the marginal density distribution of coefficients in each sub-band is approximated by Gaussian mixtures. The Gaussian component parameters are estimated by an "EM+MML" algorithm which performs parameter estimation and model selection automatically. The Earth Mover's Distance (EMD) is used to measure the distributional similarity based on the Gaussian components. Thus, classification can be done by matching the marginal density distributions. Extensive experiments show that the proposed method achieved an overall improved classification accuracy compared to nonparametric representation of sub-band distributions.**

I. Introduction

Texture is an important regional characteristics of images. Image texture classification refers to the classification of image textures according to the feature properties extracted from them. Tuceryan and Jain [1] surveyed four major categories of methods for texture analysis, namely, statistical, geometrical, model based and signal processing approach. Over the last few years, more and more texture classification methods based on filtering theory have been reported, benefitting from the latest development in signal processing, such as multi-resolution or scale-space analysis. The basic assumption of these methods is that the energy distribution in the frequency domain identifies a texture. Hence, if the frequency spectrum of a texture is decomposed into a sufficient number of sub-bands, the spectral signatures of different textures will be different enough to give an accurate classification.

In spatial-frequency based texture analysis, multichannel filtering is often employed to produce filtered sub-band images from which statistics of sub-band coefficients can be derived. These statistics are usually used as features forming a composite feature vector. Although numerous decomposition schemes have been proposed for the purpose of texture characterization, numerical description is mainly achieved through the second order (or higher order) statistics of the sub-band coefficients [2] [3], often with implicit assumption of Gaussian distributions. However, many natural image textures are known to give rise to non-Gaussian sub-band marginal densities [4] [5]. The second-order (or higher order) statistics are insufficient to approximate the marginal densities of the sub-band coefficients, thus they are insufficient to describe various image textures accurately. As a remedy, non-parametric representations have been proposed [6] [7]. Histograms of the sub-band coefficients are considered a better representation which can capture the distributions of any form. However, the choice of quantization intervals in generating histograms is crucial and it is often decided empirically. Normalization of the histogram ranges may not be desirable because some ranges may be coarsely quantized and some ranges will be sparse, either loosing the precision of representation or wasting the histogram bins causing unnecessary computation overheads.

In this paper, we propose a new texture description using the Finite Gaussian Mixture (FGM) model to approximate the marginal densities of decomposed sub-band coefficients. The FGM model parameters are estimated by the "EM+MML" algorithm and they are used as features for texture discrimination. Since the FGM model parameters can approximate the marginal densities sufficiently well, the features extracted by this method achieve higher capacity of image texture representation. The classification is accomplished by using Earth Mover's Distance(EMD) to measure the distributional similarity by sets of the Gaussian components representing texture classes.

II. Decomposing images into sub-bands

In the past decade, the multichannel decomposition approach has been widely used for analyzing image textures. Some popular methods for decomposing an image into sub-bands are multichannel filtering using Gabor filter bank, wavelet transform and steerable pyramid. The Gabor filter bank produces filtered images corresponding to the scales and orientations of the corresponding Gabor filters [8]. In wavelet transform [9], an image is transformed into approximate and detailed coefficients at multiple scales with respect to the selected wavelet basis. A steerable pyramid combines the advantages of both Gabor wavelet and wavelet transform, producing a multiscale version of the original image in a pyramid hierarchy using directional derivatives [10] [11]. Figure 2 illustrates a steerable pyramid decomposition of an image texture (shown in Figure 1) with 4 orientational subbands at 3 scales. Figure 3 shows the histograms of the 14 filtered images in Figure 2. The distributions seen in these histograms are typical and an assumption of Gaussian for all of them would be too restrictive. In our work, a Finite Gaussian Mixture(FGM) model is proposed to approximate the marginal statistics due to its simplicity and ease of computation, although other mixtures can also be used.

III. Approximating marginal densities using FGM model

A. Finite Gaussian mixture model

Gaussian mixture model is a type of density model which comprises a number of component Gaussian functions. These component functions are combined with different weights to result in a multi-modal density. Gaussian mixture models are a semi-parametric alternative to non-parametric histograms (which can also be used to approximate densities) and it has greater flexibility and precision in modelling the underlying distribution of sub-band coefficients. As illustrated in Figure 3, the sub-band coefficient histograms in a steerable pyramid decomposition are not all Gaussians, even though many of them are Gaussian-like. Thus, we can use a Finite Gaussian Mixture model to approximate these densities more accurately.

Let *x* be a sub-band coefficient in the steerable pyramid. Associated with it is an unobserved hidden state variable $S_i \in S_1, \cdots, S_k$. The value of S_i dictates which of the *k* components in the mixture model that generate *x*. It is said that *x* follows a *M*-component finite mixture distribution if its

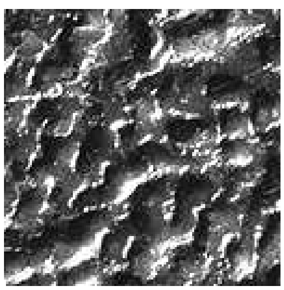


Fig. 1. An example of image texture

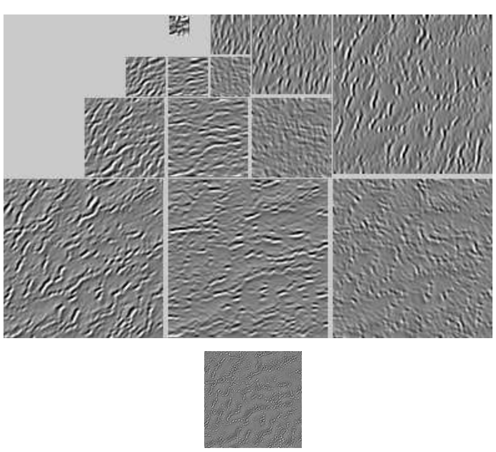


Fig. 2. The decomposition of Figure1 by using steerable pyramid. The bottom one is the residual high-pass sub-band.

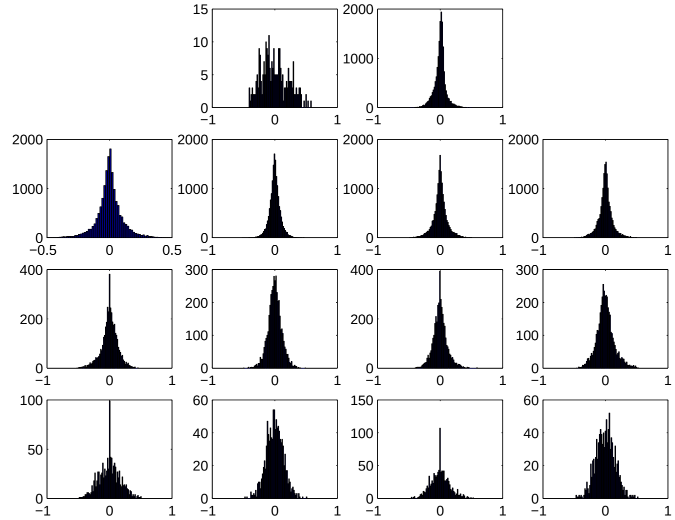


Fig. 3. Histograms of coefficients in each sub-band from the steerable pyramid decomposition(normalized). (The sub-band images are shown in Figure 2; The figures on the first row are the histograms of the low-pass band and the high-pass band)
.

probability density function can be written as

$$p(x|\Theta) = \sum_{m=1}^{M} \alpha_m p(x|\theta_m), \forall x \in \Re^n \quad (1)$$

where $\alpha_1, \cdots, \alpha_M$ are the mixing proportions, such that $\sum_{m=1}^{M} \alpha_m = 1$ and each θ_m is the set of parameters defining the *m*-th component, and $\Theta = (\alpha_1, \cdots, \alpha_M, \theta_1, \cdots, \theta_M)$ is the complete set of parameters needed to specify the mixture.

In the case of mixture components S_i being Gaussian densities with parameters $\theta_m = (\mu_m, \Sigma_m)$, eq. (1) can be expressed as

$$p(x|\Theta) = \sum_{m=1}^{M} \alpha_m N(x|\mu_m, \Sigma_m) \quad (2)$$

The complete set of parameters in the FGM model is $\Theta = (\alpha_1, \cdots, \alpha_M, \mu_1, \cdots, \mu_M, \Sigma_1, \cdots, \Sigma_M)$.

B. Parameter estimation by "EM+MML" algorithm

The commonly used approach for determining the parameters Θ of a Gaussian mixture model from a given data set is by the Maximum Likelihood (ML) estimation. In this paper, the well-known Expectation-Maximum (EM) algorithm is used to estimate the Gaussian mixture parameters [12]. However, an important issue in mixture modelling is the selection of the number of components. With too many components, the mixture may over-fit the data, while a mixture with too few components may not be accurate enough to approximate the true underlying density. Many algorithms have been proposed to solve this problem, such as Split and Merge EM [13], Greedy EM [14] and the component-wise EM + MDL algorithm proposed by Figueiredo and Jain [15]. In our work, the Minimum Message Length (MML) is adopted as the model selection criterion, achieving the same objective as

- **Input**: Data vector X;
- **Params**: Maximum and minimum number of components M_{max}, M_{min}
- **Procedures**:

  Set the initial Message Length $L_{min} = +\infty$;
  for m = M_{max}:-1:M_{min}
  Initialize Θ_0 = Θ_{k_means};
  $\widehat{\Theta}_m$ = $\widehat{\Theta}_{EM}$;
  Compute L_m;
  if $L_m \leqslant L_{min}$, *then*
  L_{min}= L_m; $\widehat{\Theta}_{best}$ = $\widehat{\Theta}_m$;
  end if
  end for
- **Output**: Mixture model parameters $\widehat{\Theta}_{best}$ and the number of mixture components M_{best} determined by MML.

Fig. 4. EM+MML algorithm

MDL. Different from [15], we used a standard EM algorithm instead of the component-wise EM. The algorithm is called the "EM+MML" algorithm which integrate parameter estimation and model selection in a single algorithm. The Message Length L is expressed as

$$\boldsymbol{L}(\widehat{\theta}, y) = \frac{N}{2}\sum_{m=1}^{M}\log(\frac{n\widehat{\alpha_m}}{12}) + \frac{M}{2}\log\frac{n}{12} + \frac{M(N+1)}{2} - \log p(y|\widehat{\theta}) \quad (3)$$

where N is the number of parameters specifying each component, n is the number of observed data, M is the number of components, and $\widehat{\alpha}$, $\widehat{\theta}$ are the estimated Gaussian mixture parameters. The EM+MML algorithm is presented in Figure 4.

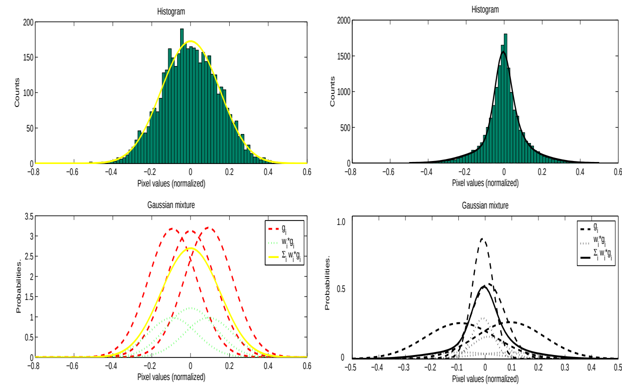
In the "EM+MML" algorithm, we run EM algorithm iteratively from M_{max} to M_{min}. In each iteration, the mixture parameters are initialized by the k-means algorithm. The number of components and mixture parameters are selected according to the Minimum Message Length expressed in eq. (3).

Considering the computation cost, in our experiment, we set M_{max}=6 and M_{min}=1, i.e., the number of mixture components is between 1 and 6. Experimental results show that mixture model using a maximum of 6 components can approximate the marginal density well enough for our purpose.

IV. Texture Classification Using FGM Model Parameters

A. Earth Mover's Distance

The Earth Mover's Distance is proposed to measure the distribution similarity between a set of Gaussian components representing an input texture and sets of such components representing all texture classes in the database. The Earth Mover's Distance was first proposed by Rubner *et. al.* for



(a) 3 components (b) 4 components

Fig. 5. Matching two marginal histograms using different number of Gaussian components.

content-based image retrieval in a large database [16]. It is used to compute the minimum amount of work that must be performed to transform one distribution into the other. Intuitively, given two discrete, finite distributions,

$$X = \{(x_1, w_{x1}), (x_2, w_{x2}), \cdots, (x_m, w_{xm})\}$$
$$Y = \{(y_1, w_{y1}), (y_2, w_{y2}), \cdots, (y_n, w_{yn})\}.$$

where x_i, y_j are two distributions, and w_{xi}, w_{yj} are corresponding weights of the distribution, finding a $m \times n$ matrix F where f_{ij} is the amount of weight w_i matched to w_{yj}, that will minimize the following function:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij} \quad (4)$$

and subject to the following constraints:

$$f_{ij} \geq 0, 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n \quad (5)$$
$$\sum_{j=1}^{n} f_{ij} = w_{xi}, 1 \leqslant i \leqslant m \quad (6)$$
$$\sum_{i=1}^{m} f_{ij} = w_{yj}, 1 \leqslant j \leqslant n \quad (7)$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min(w_x, w_y) \quad (8)$$

where $w_x = \sum_{i=1}^{m} w_{xi}$, $w_y = \sum_{j=1}^{n} w_{yj}$, the Earth Mover's Distance is defined by the normalized distance between *X* and *Y*

$$EMD(X,Y) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}}{min(w_x, w_y)} \quad (9)$$

In our work, the weight w_{xi}, w_{yj} is the Gaussian component weight of each sub-band, so w_x = 1, w_y = 1. The distance d_{ij} is defined as the Kullback-Leibler (KL) divergence between two components. The KL divergence between two densities f and g is expressed as [17]:

$$D(f_i\|g_j) = \int_{\Omega} f_i(x)\log[f_i(x)/g_j(x)]dx \quad (10)$$

For two Gaussian distributions $f_i \sim N(\mu_i, \Sigma_i)$, $g_j \sim N(\mu_j, \Sigma_j)$, the KL divergence becomes,

$$D(f_i\|g_j) = \frac{1}{2}\log\frac{|\Sigma_j|}{|\Sigma_i|} + \frac{1}{2}tr[\Sigma_i(\Sigma_j^{-1} - \Sigma_i^{-1})] + \frac{1}{2}tr[\Sigma_j^{-1}(\mu_i - \mu_j)(\mu_i - \mu_j)^T] \quad (11)$$

where $tr[\cdot]$ denote the trace of the matrices.

So, the EMD is expressed as

$$EMD_{sub} = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} D(f_i\|g_j) \quad (12)$$

where$D(f_i\|g_j)$ is defined by (11).

For a texture which is decomposed as K sub-bands, the total EMD is the sum of that of each sub-band,

$$EMD = \sum_{k=1}^{K} EMD_{sub} \quad (13)$$

The classification is accomplished by computing the EMD between the input texture and each texture class, and the input texture is classified into the class with which the EMD is smallest.

V. Experimental Results

A. Database of texture images

The textures in these experiments are from the Brodatz texture album [18]. All 112 original textures were scanned and saved as 512×512 images. 100 image textures of size 128×128 are sampled from each 512×512 image, and the image textures are divided into balanced training and test sets. Each training set consists of 50 images sampled from the top half of the original 512×512 images, another 50 images sampled from the bottom half of the texture image for testing. Since the Brodatz textures contain both uniform and non-uniform textures, we selected 78 classes of uniform textures (see Fig.7). Other non-uniform textures in the Brodatz album are not used because the sampled textures will be very different even though they come from the same original Brodatz texture. Although including the non-uniform textures in the experiment will make the problem more challenging, the degradation of performance is caused by the difference in the sampled texture rather than due to the limitation of the proposed method.

B. Experimental results

In order to evaluate the performance of our classification method, we also experimented with histogram method proposed by Puzicha *et. al.* [7]. The reason for choosing this method for comparison is that the histogram method can achieved good classification result among the existing texture classification methods. For a fair comparison, the preprocessing of images for both methods are the same, i.e., we use steerable pyramid to decompose the texture into several sub-bands first. In the histogram method, the density distributions are represented using histograms of different number of bins, and use the histogram entries as features for classification. While in our proposed method, we use FGM model to fit the density distributions and use EMD to measure the distribution similarity.

In the experiment, we choose 20, 40, 60, 78 texture classes respectively. The experimental results are shown in Table I. Figure 6 is a plot of the classification rates. It is shown that when the class number is small, both the FGM and the histogram method achieve good classification rates, however, when the number of classes increases, the FGM method is better than the histogram method.

However, a disadvantage of the FGM algorithm lies in its computation. Since the FGM method involves the estimation of mixture parameters and Earth Mover's Distance computation, compared with the histogram method, it is more computation intensive.

TABLE I
Classification result on database1

Number of classes	10	20	40	60	78
FGM	98.50	97.80	97.40	97.70	95.03
Histogram(bin 32)	99.40	98.70	94.60	94.30	90.10
Histogram(bin 64)	100.00	98.20	96.05	94.13	87.51
Histogram(bin 128)	99.70	98.50	97.30	92.80	91.69

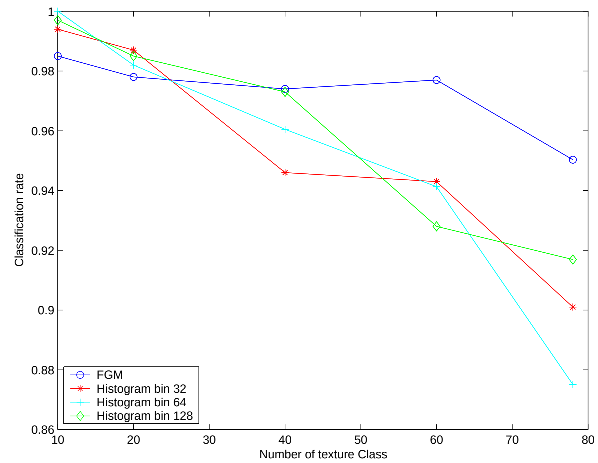


Fig. 6. Classification accuracy with respect to the number of texture classes.

VI. Conclusion

In this paper, a method to approximate the marginal distribution using Finite Gaussian Mixture model is proposed. Experimental results shows our proposed method can achieve better classification accuracy than the histogram method when there is a larger number of classes. The methods avoid

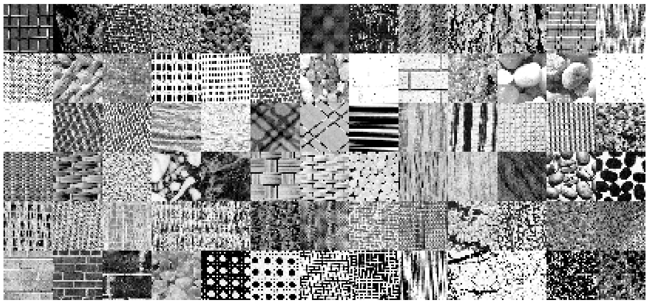


Fig. 7. 78 texture image database used in the classification experiment 1

the problems faced with the use of histograms. The use of EMD for distributional similarity also avoids the problem in comparing the Gaussian components for classification, such as inaccuracy of the parameter estimation by EM, ordering of the components and incompatible number of components due to MML. In our experiments, we use the steerable pyramid filter to decompose the texture. However, it is not necessary to use steerable pyramid to achieve this objective, Gabor filters, wavelet filters or other decomposition methods are also suitable for this purpose.

REFERENCES

[1] M. Tuceryan and A. K. Jain, "Texture Analysis," in *Handbook of Pattern Recongnition and Computer Vision*, C. Chen, L. Pau, and P. Wang, Eds. World Scientific, 1993, pp. 235–276.

[2] D. Dunn, W. E. Higgins, and J. Wakeley, "Texture Segmentation using 2D Gabor Elementary Functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 130–149, February 1994.

[3] A. K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation using Gabor Filters," *Pattern Recognition*, vol. 24, no. 12, pp. 1167–1186, 1991.

[4] M. J. Wainwright and E. P. Simoncelli, "Scale Mixtures of Gaussians and the Statistics of Natural Images," in *Advances in Neural Information Processing System*, vol. 12. MIT Press, 2000, pp. 855–861.

[5] X. W. Liu and D. L. Wang, "Texture Classification using Spectral Histograms," Department of Computer Science, Florida State University, Tech. Rep., 2002.

[6] J. Puzicha, T. Hofmann, and J. Buhmann, "Non-Parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval," in *Proc. of IEEE Conference Computer Vision and Pattern Recognition*, June 1997, pp. 267–272.

[7] J. Puzicha, Y. Rubner, C. Tomasi, and J. M. Buhmann, "Empirical Evaluation of Dissimilarity Measures for Color and Texture," in *Proc. IEEE International Conference on Computer Vision*, 1999, pp. 1165–1173.

[8] J. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters," in *J. Opical Soc. Am.*, vol. 2, no. 7, 1985, pp. 1160–1169.

[9] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: Society for Industry and Applied Math, 1992.

[10] E. P. Simoncelli and W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation," in *Second International Conference on Image Processing*, vol. 3, November 1995, pp. 444–447.

[11] W. T. Freeman and E. H. Adelson, "The Design and Use of Steerable Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

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

[13] N. Ueda, R. Nakano, Y. Ghahramani, and G. Hiton, "SMEM Algorithm for Mixture Models," *Neural Computation*, vol. 12, no. 10, pp. 2109–2128, 2000.

[14] J. Verbeek, N. Vlassis, and B. Krose, "Efficient Greedy Learning of Gaussian Mixture Models," *Neural Computation*, vol. 15, no. 2, pp. 469–485, 2003.

[15] M. A. T. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.

[16] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for Distributions with Applications to Image Databases," in *Proc. IEEE International Conference on Computer Vision*, 1998.

[17] Cover and T. Thomas, *Elements of Information Theory*. Wiley, 1991.

[18] P. Brodatz, *Texture: A Photographic Album for Designers and Artists*. Dover, 1966.