Texture Recognition through Modal Analysis of Spectral Peak Patterns

Marco Carcassoni, Eraldo Ribeiro and Edwin R. Hancock
Department of Computer Science
University of York
York, Y010 5DD, UK
email: marco,erh@cs.york.ac.uk

Abstract— **In this paper we investigate how texture recognition can be achieved through the modal analysis of the pattern of peaks in the spectral density function. We commence from a texture characterisation which is based on the positions of peaks in the power spectrum. Our aim is to use the modal structure of the pattern of peaks to perform texture retrieval from an image data-base. We explore two different approaches to the problem. First, we use a variant of the Shapiro and Brady method to perform recognition by comparing the modal structure of the proximity matrix for peak cluster centres. Second, we perform latent semantic indexing on vectors representing the polar distribution of frequency peaks. We provide an experimental evaluation of these two methods on a data-base of fabric and wrapping paper patterns.**

I. Introduction

Texture indexation is a key process for retrieving images from large image data-bases [16]. There are several well documented studies in the literature. Some of the pioneering work was done by Minka and Picard [15] who explored the use of texture for image annotation. This method was later incorporated into the PicHunter image retrieval system [12]. Other work has focused on finding good features for texture retrieval. For instance Gimel'farb and Jain [5] have used Gabor filter-banks to compute texture feature vectors for the purposes of texture recognition. Albanesi, Ferretti and Giancane [1] have reported a hierarchical approach to the problem using the wavelet transform. Rather than using frequency domain methods, Lin, Wang and Yang [9] adopt a structural approach to the retrieval process.

In this paper we explore a different approach to the problem. Our aim is to investigate whether the modal structure of the pattern of frequency peaks in the texture power spectrum can be used for the purposes of recognition. The reason for adopting this approach is that eigenvector and singular value methods have proved to be highly effective in the retrieval of both text and images from large databases. We explore two different approaches to the problem, which have been used effectively in the modal analysis of point-patterns and text retrieval from data-bases. The first of these is a variant of the modal correspondence algorithm of Shapiro and Brady [14]. Here we locate the centres of clusters of spectral peaks using information provided by the modal structure of their frequency domain peak proximity matrix. The modal structure of the cluster-centre proximity matrix is used to gauge the similarity of different power-spectra and perform texture matching. The second approach uses latent semantic indexing. This is a text retrieval technique that uses vectors of key-word word-frequencies. Here we adapt the approach by using vectors of frequency peak counts in a polar binning of the power spectrum. The peak count vectors for the different texture images are used as the columns of a pattern matrix. The singular vectors of this matrix form the projection basis for the original peak count vectors. Recognition is achieved by finding the transformed pattern vector from the data-base that is closest to the query vector.

II. Spectral Analysis for Textures

We are interested in computing the spectral density function, or power spectrum, for textured images [8], [6]. A comprehensive coverage of spectral estimation and its application can be found in [10], [7]. The power spectrum representation of a two-dimensional image $f(x,y)$ is defined to be the Fourier transform of its autocorrelation. Due to the discrete nature of the images under study and the limited number of sample points available for analysis, we use the Blackman-Tukey estimator [13]. This estimator is the frequency response of the windowed autocorrelation function of an image. The general form of this estimator is given by:

$$P_{BT}(u,v) = \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} r_x(m,n) w(m,n) exp\left[-j2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)\right] \quad (1)$$

where u and v are the frequency variables, m and n are the spatial coordinates, and $r_x(m,n)$ is the discrete version of the autocorrelation function of the image $f(m,n)$, which is given by:

$$r_x(l,k) = \frac{1}{MN} \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} f^*(m,n) f(m+l, n+k). \quad (2)$$

The spectrum represents the frequency content for several directions at the location in the image. The peaks in the spectrum correspond to dominant spatial frequencies in the textures. Here we summarise the structure of the power spectrum using the frequency vectors of the principal peaks. We denote the frequency vector of the ith peak by $\mathbf{U}_i = (u_i, v_i)^T$. The frequency vectors can be concatenated according to the energy ordering of the peaks to form the long-vector of peak frequencies $\mathbf{V} = (u_1, v_1, u_2, v_2, u_3, v_3, ...)^T$.

III. Matching

We aim to perform texture recognition by performing modal analysis on the frequency vectors for peaks extracted from the power-spectra. We explore two different routes. The first of these involves locating correspondences between the spectral peaks. This is done using a variant of the Shapiro and Brady method. The second approach applies latent semantic indexing to the long-vectors representing the polar distribution of spectral peaks in the frequency domain.

A. Spectral clustering

The idea underpinning our modal correspondence method is as follows. We commence by searching for the main structure in the arrangement of spectral peaks. To do this, we use a pairwise clustering method. From the positions of the main cluster-centres, i.e. the centres of spectral activity, we extract a modal representation of the power-spectrum. We match the query texture by measuring the similarity of its modal matrix to those of texture patterns in the data-base.

Suppose the texture indexed k is composed of the peaks whose frequency vectors are drawn from the set $S_k = \{\mathbf{U}_i^k, i = 1, ...N_k\}$. For this set of frequency vectors we commence by computing the spectral peak proximity matrix. Following our earlier work on point-pattern matching [4], the element with row i and column j of this matrix is computed using the sigmoidal weighting function,

$$H_k(i,j) = \frac{2}{\pi ||U_i^k - U_j^k||} \log\cosh\left[\frac{\pi}{s}||U_i^k - U_j^k||\right]. \quad (3)$$

We characterise the modal-structure of the set of spectral peaks by solving the eigenvalue equation $det[H_k - \lambda I] = 0$ together with the associated eigenvector equation $H_k \phi_l^k = \lambda_l^k \phi_l^k$. The eigenvectors are sorted according to the magnitude order of the eignevalues, i.e. $|\lambda_1^k| > |\lambda_2^k| > ... > |\lambda_{N_k}^k|$ and are stacked in eigenvalue order to construct the modal matrix $\Phi_k = (\phi_1^k|\phi_2^k|....|\phi_{N_k}^k)$. We use the M largest eigenvectors to represent clusters of peaks in the power spectrum. For the cluster indexed m, the cluster centre is given by

$$C_m^k = \frac{\sum_{i=1}^{N_k} |\Phi_k(i,m)| \mathbf{U}_i^k}{\sum_{i=1}^{N_k} |\Phi_k(i,m)|}$$

The cluster-centres are used to compute a cluster-centre proximity matrix. Repeating the procedure above, the element with row l and column m of this matrix is

$$H_k^C(l,m) = \frac{2}{\pi ||C_l^k - C_m^k||} \log\cosh\left[\frac{\pi}{s}||C_l^k - C_m^k||\right]. \quad (4)$$

We then apply the modal analysis described above to the matrix H_k^C to compute an $M \times M$ cluster-centre model matrix, which we denote by Ψ_k.

Our aim is to find the image in the data-base whose power-spectrum most closely resembles that of a query image. For the query image, we hence repeat the analysis described above to obtain the spectral-peak modal matrix Ψ_q. Recognition is achieved by comparing the query modal matrix Ψ_q, with the modal matrices Ψ_k, $k = 1, ..., D$ for each of the images in the data-base. We measure similarity by comparing the rows of the modal matrices. We perform the similarity assessment on an element-by-element basis using a Gaussian error kernel. The identity of the retrieved pattern is the one that satisfies the condition

$$\omega_q = \arg\max_k \sum_{l=1}^{M}\sum_{m=1}^{M} \exp\left[-k \sum_{n=1}^{M}\Big(\Psi_q(l,n) - \Psi_k(m,n)\Big)^2\right]. \quad (5)$$

Hence, elements in the rows of the modal matrices contribute insignificantly if they differ greatly.

B. Latent semantic indexing

Latent semantic indexing [2], [11] was first developed to retrieve text-documents from large databases (for example pages from internet). Each document is represented by a vector of key-word frequencies. These vectors are then stacked to form a so-called term-document matrix. The structure of this matrix is extracted using singular value decomposition. Document retrieval is effected by projecting the query-vectors of word frequencies onto the space spanned by the singular vectors, and locating the closest document in the data-base. Obviously, the main computational bottleneck arises from the need to perform singular value decomposition on a large matrix. However, this is a preprocessing operation that needs only to be performed once.

Here we aim to apply this technique to a representation of the peaks of the texture power-spectra. Our vector is an encoding of the polar distribution of frequency of peaks about the centroid of the power spectrum. To obtain this polar representation we construct bins by sampling the frequency plane at 10 degree angle intervals and dividing the span of the radius into 10 uniform length intervals. We count the frequency of spectral peaks in each bin and normalise the total bin contents to unity. For the texture power spectrum indexed k, the normalised bin contents are encoded in a long-vector $\mathbf{V}_k = (h_{r_1,\theta_1}^k, h_{r_1,\theta_2}^k, ..., h_{r_1,\theta_{10}}^k, h_{r_2,\theta_1}^k, ..., h_{r_{10},\theta_{10}}^k)^T$, where h_{r_i,θ_j}^k is the normalised number of spectral peaks contained in the polar frequency bin with radial index i and angular index j. Each vector has $L = 100$ components.

Latent semantic indexing of the texture images commences by stacking the polar peak frequency vectors to form the $L \times D$ matrix $\mathbf{A} = (\mathbf{V}_1|\mathbf{V}_2|.....|\mathbf{V}_k|....\mathbf{V}_D)$. We proceed by performing the singular value decomposition $\mathbf{AA^T} = \mathbf{S\Delta T^T}$ on the $L \times L$ matrix $\mathbf{AA}^T$. Suppose that $\mathbf{AA}^T$ is of rank R, i.e. it has R non-zero eigenvalues. In the decomposition $\mathbf{S}$ and T are $L \times L$ orthogonal matrices of rank R. The matrix $\mathbf{\Delta}$ is diagonal; its non-zero diagonal elements are the R non-zero eigenvalues of $\mathbf{AA}^T$. The first h columns of the matrix $\mathbf{S}$ are taken to represent the best projection basis for the original vectors. The projection of the vector $\mathbf{V}_k$ onto this basis is $\hat{\mathbf{V}}_k = \mathbf{V}_k^T \cdot \mathbf{S_h} \cdot \mathbf{\Delta_h^{-1}}$. where $\mathbf{S_h}$ and $\mathbf{\Delta_h}$ are the $L \times h$ matrices obtained by respectively taking the first h rows of $\mathbf{S}$ and $\mathbf{\Delta}$.

To retrieve textures we proceed as follows. First, we compute the polar peak count vector $\mathbf{V}_q$ for the query image. This vector is projected onto the singular value basis and gives a transformed vector $\hat{\mathbf{V}}_q$. We measure texture similarity using the angle between the transformed query vector and the texture-vectors. The identity of the retrieved texture is the one that satisfies the condition

$$\omega_q = \arg\min_k \arccos \frac{\hat{\mathbf{V}}_k.\hat{\mathbf{V}}_q}{|\hat{\mathbf{V}}_k| \times |\hat{\mathbf{V}}_q|} \tag{6}$$

IV. Experiments

In this section we describe our experimental evaluation and we compare these two texture image recognition methods. We use both synthetic and real world data. We commence with a sensitivity study on synthetic data. This is aimed at measuring the effectiveness and robustness of the method when the power spectra and their associated spectral peaks under study are subject to disruption. The second part of the study focuses on real world data. Here we investigate the recognition performance of the methods when applied to a data-base of images of regularly textured patterns, provided by samples of fabric and wrapping paper.

A. Real World Images

For our real-world experiments we have used a data-base of 90 images of regularly textured patterns. The images fall into 9 different texture classes. In figure 1 we show some examples of the images used in our study. Figure 2 shows the corresponding power spectra for the images. We have performed recognition experiments in which each texture in turn is withheld from the data-base and used as a query. We have then recorded the number of times that the best-matched pattern in the data-base is of the same class as the query pattern. For latent semantic indexing the fraction of correct matches is 91%. When the modal clustering method is used the fraction of correct matches is 93%. The main source of recognition error is the ambiguity in the structure of the power spectra. For instance, in Figure 1 the sixth and the seventh patterns, although quite distinct, have similar power spectra. In addition, in about 3-4% of the images the power-spectrum computation fails to converge.

B. Synthetic Data

For sensitivity study, we constructed a data-base of synthetic images. Here there were again 10 synthetic variants for each of 9 texture pattern. To the synthetic data, we have added three types of noise. Firstly, there is Gaussian blurring which has the effect of smoothing edge structure

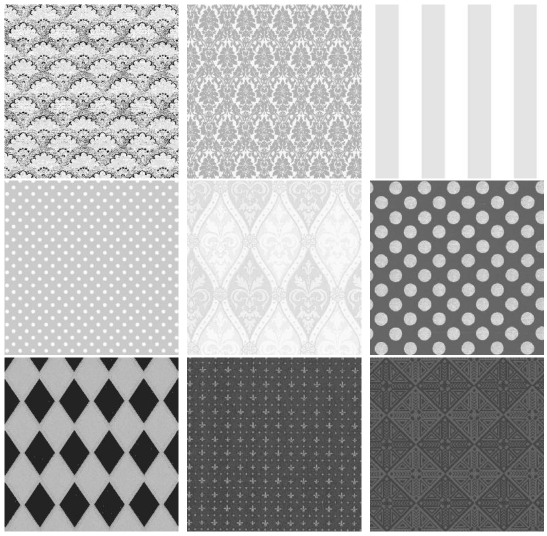

Fig. 1. Real data image set

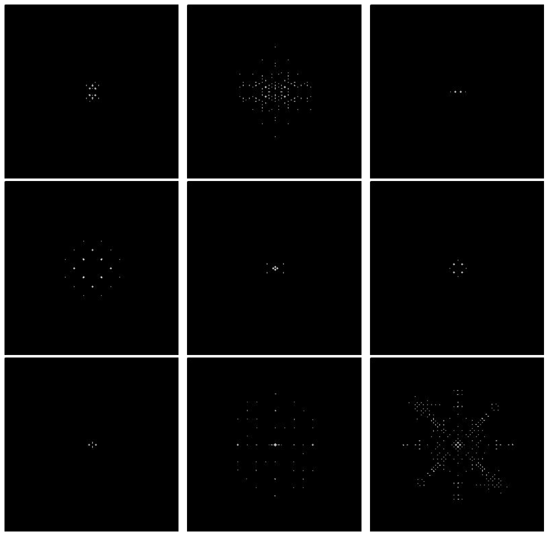

Fig. 2. Power spectra

in the images. Second, we have added a uniform clutter background. Finally, we have added randomly distributed clutter. These latter two noise processes have the effect of distorting and breaking the texture symmetries. The images used in our study are of dimension 256x256 pixels. On average they contain between 3 and 6 texture repetitions in the direction of each of the pixel lattice directions.

In Figure 3 we show the results of our sensitivity study. Here we show the fraction of correct matches as the fraction of added noise increases. The three different plots in the figure are for different noise processes(Gaussian-top, uniform-middle, irregular-bottom). The two different curves in each plot are for the modal clustering method (top), and latent semantic indexing (bottom). For each noise process, the modal clustering process outperforms latent semantic indexing.

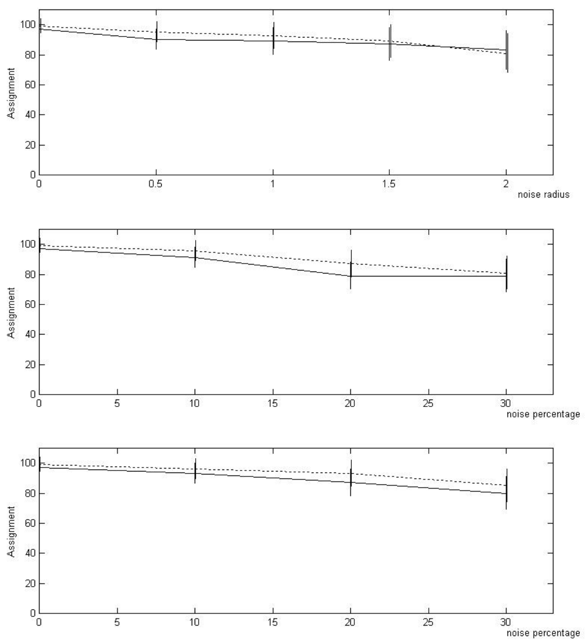

Fig. 3. Synthetic data study

V. Conclusions

In this paper we have considered how to perform texture recognition from image data-bases through the modal analysis of the pattern of peaks in the texture power spectrum. We explore two different approaches to the problem. The first of these uses a modal correspondence method to match using proximity information for the cluster-structure in the detected spectral-peaks. The second method applies latent semantic indexing to vectors representing the polar distribution of spectral peaks. On experiments involving both real world and synthetic data, the former method provides better recognition performance.

Our future plans involve using texture measures that allow for a degree of viewpoint invariance. Here we intend to exploit our recent work [13] which has shown that the direction of the leading eigenvector of the spectral distortion matrix is an affine invariant. We will also investigate methods for affine and perspective rectification of the texture images.

References

[1] MG. Albanesi, M. Ferretti, A. Giancane "Robust hierarchical indexing based on texture features", *J. of Vis. Lang. and Comp.*, **11**(4):383-404, Aug 2000.

[2] Y. Azar, A. Fiat, A.R. Karlin, F. McSherry and J. Saia, "Spectral analysis of data", *Symposium on Theory of Computing (ACM)*, 619-626, 2001.

[3] M. Carcassoni and E.R. Hancock, "Correspondence Matching using Spectral Clusters", Proceedings of the 12th Scandinavian Conference on Image Analysis, SCIA 2001, pp. 243–249 , 2001.

[4] M.Carcassoni and E.R.Hancock, "Point Pattern Matching with Robust Spectral Correspondence", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, pp. 649-655, 2000.

[5] G.L.Gimel'farb and A.K. Jain, "On retrieving textured images from an image database", *Pattern Recognition*, **29**(9):1461–1483,1996.

[6] Anil K. Jain, *Fundamentals of digital image processing*, Prentice Hall, 1989.

[7] S.M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, 1988.

[8] J.S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, 1990.

[9] H.C. Lin, L.L. Wang, S.N. Yang, "Regular-texture image retrieval based on texture-primitive extraction", *Image and Vision Computing*, **17**(1):51-63, Jan 1999.

[10] S.L. Marple, *Digital Spectral Analysis with Applications*,Prentice Hall, 1987.

[11] C.H. Papadimitriou, P. Raghavan, H. Tamaki and S. Vempala", "Latent Semantic Indexing: A Probabilistic Analysis", *J. of Comp. and Syst. Sci.*, **61**, 2000.

[12] T.V. Papathomas , I.J. Cox, P.N. Yianilos, et al. "Psychophysical experiments on the PicHunter image retrieval system", *Journal of Electronic Imaging*, **10**(1):170-180, Jan 2001.

[13] E. Ribeiro and E.R. Hancock, "Shape From Periodic Texture using the Eigenvectors of Local Affine Distortion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23** pp. 1459–1465, 2001.

[14] L.S. Shapiro and J.M. Brady, "Feature-based correspondence - an eigenvector approach", *Image and Vision Computing*, **10**, pp. 283–288, 1992.

[15] R. W. Picard and T. P. Minka, "Vision Texture for Annotation", *J. Multimedia Systems*, **3**:3–14, 1995.

[16] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years" *Trnsaction on Pattern Analysis and Machine Intelligence*, **22**(12): 1349-1380, Dec 2000.