A Comparative Study of Rotation Invariant Classification and Retrieval of Texture Images

S. R. Fountain[†], T. N. Tan[‡] and K. D. Baker[†]

[†]Computational Vision Group, Department of Computer Science, The University of Reading, Reading RG6 6AY.

[‡]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing.
`S.R.Fountain@Reading.ac.uk`

Abstract

This paper presents a detailed comparative study of 4 rotation invariant texture analysis methods. Human subjects are included as a benchmark for the computational methods. Experiments are conducted on two databases of 450 and 1320 images taken from 10 and 44 Brodatz texture classes respectively. Classification and content based retrieval experiments are used in the comparison, which includes the effect of Gaussian noise on each method (including the human subjects). The methods tested are: the multichannel Gabor filtering method; an edge attribute processing method; the circular simultaneous autoregressive (CSAR) model method and a method based on hidden Markov models with wavelet decomposition. The test conditions in the study are kept both general and constant so direct comparisons of the results obtained can be made.

1. Introduction

Texture analysis is important for many applications such as content based image retrieval and scene analysis. Research on the subject has been active for decades. The vast majority of techniques developed to date (see [6] for a recent review) assume that images are acquired from the same viewpoint. This is an unrealistic assumption for most practical applications. For example, if images are obtained from scanning photographs they are often subject to random skew angles. Texture analysis methods should ideally be invariant to viewpoints in such applications. Obtaining truly viewpoint invariant texture features is a very difficult task [1]. Rotation (e.g. due to skew angles) and scale (e.g. change in focal length) invariance are important aspects of the general viewpoint invariance problem.

Rotation invariant texture analysis can be obtained either via the learning of rotation invariance during a training phase or through the extraction of rotation invariant properties from the input textures. The former method includes model based techniques such as the hidden Markov model (HMM) method adopted by Chen and Kundu [9], where rotation invariance is obtained by training the HMM on texture samples from a

wide variety of angles. The latter includes the Hough transform [8] and complex log mapping [11]. Model based (e.g. [4]), statistical (e.g.[12]) and Fourier (e.g. [2,3]) methods have all been applied successfully to the problem for the selected test images in each case (see [1] for further examples).

A system named RAIDER (Retrieval and Annotation of Image Databases) [15] is currently under development at The University of Reading. It requires a rotation invariant texture analysis method to assist image analysis for the content based annotation and retrieval stages. In order to determine which method is most suitable for the task a comparison of the classification and retrieval capabilities of several rotation invariant texture analysis methods was performed. The results obtained are appropriate to any application requiring rotation invariant texture analysis such as scanned document processing and OCR [10].

Four methods covering all the aforementioned approaches were selected for inclusion in the comparison. Two model based techniques were selected namely the successful wavelet decomposition and HMM method [9] and the early circular simultaneous autoregressive (CSAR) method [4]. A statistical method which exploits edge attribute information [12] was included along with the popular multichannel filtering method [7]. Similar experiments were performed on humans as a benchmark for the four computerised methods. A method is certainly worth considering if it can outperform a collection of human subjects.

The paper is organised as follows. Outlines of the 4 methods are given in Section 2. Classification and noise experiments are presented for a small database in Section 3 and for a large database in Section 4. CBR results for the large database are given in Section 5 and the overall conclusions are drawn in Section 6.

2. Outlines of The Selected Methods

Kashyap and Khotanzad were amongst the first to recognise the importance of rotation invariant texture analysis. The studies presented in [4] address just seven orientations. In this paper the effects of random orientations on this early technique are explored and a comparison with more recent methods given.

The comparison between the wavelet and HMM method and the multichannel filtering method is interesting as both methods are based on sub-band decomposition. The edge attribute method is extremely simple and is computationally very efficient. Until recently its accuracy as compared to other methods was unknown. It is interesting to see how this simplistic method compares with the other more complicated and time consuming techniques.

Due to space limitiations the four methods can only be outlined in the following paragraphs. Further details can be found in the respective references.

2.1 Edge Attribute Processing

In this method a Sobel edge operator is used to generate gradient direction and magnitude images of the input texture. The gradient directions (α) at all pixels are then histogrammed and weighted by the corresponding gradient magnitudes. The resulting histograms are spiky; spurious spikes are removed by smoothing. Normalisation is

required to remove the undesirable effects of different illuminations. The following equation defines the normalisation technique used:

$$B(\alpha) = \frac{b(\alpha)h}{m}$$

(1)

where h is the maximum height of the histograms, m is the largest histogram value of $b(\alpha)$, $B(\alpha)$ and $b(\alpha)$ are the normalised and original values at a histogram bin α.

The direction histogram so formed is cyclic and can be regarded as a periodic function of α with period 2π, where a rotation of the image results in a translation of this function. The Fourier expansion of the periodic function is taken and the magnitudes of the Fourier coefficients are invariant to rotations. The first n magnitudes can be represented in an n-dimensional feature vector for use in classification.

2.2 Multichannel Gabor Filtering

A multichannel filtering technique based on Gabor filters is used to acquire rotation invariant texture features. The definition of a complex Gabor filter is given in Equation 2.

$$h(x,y) = g(x,y)e^{-2\pi j(u_o x + v_o y)}$$

(2)

where $g(x,y)$ is a 2-D Gaussian (here assumed to be isotropic):

$$g(x,y) = e^{-\frac{1}{2}\left[\frac{x^2+y^2}{\sigma^2}\right]}$$

(3)

The complex function $h(x,y)$ can be split into two parts, the even and odd filters $h_e(x,y)$ and $h_o(x,y)$ which are also known as the symmetric and antisymmetric filters respectively. These filter pairs are given in Equation 4 and are used in the multichannel method of rotation invariant texture analysis:

$$h_e(x,y) = g(x,y)\cos(2\pi f(x\cos\theta + y\sin\theta))$$
$$h_o(x,y) = g(x,y)\sin(2\pi f(x\cos\theta + y\sin\theta))$$

(4)

where: $f = \sqrt{u_o^2 + v_o^2}$ and $\theta = \arctan(v_o/u_o)$

The filtering operation is carried out in the frequency domain. The Fourier transforms of the input image and filters are taken and the output images obtained via inverse Fourier transform. For example, the output from the even filter is computed by:

$$q_e(x,y) = FFT^{-1}\left[P(u,v)H_e(u,v)\right]$$

(5)

where $P(u,v)$ is the Fourier transform of the input image $p(u,v)$ and $H_e(u,v)$ that of the filter $h_e(u,v)$. The outputs of the two filters are combined using the following equation to obtain a single value at each pixel (see [2, 9] for a justification of this combination):

$$q(x,y) = \sqrt{q_e^2(x,y) + q_o^2(x,y)}$$

(6)

The two important parameters required to position a Gabor filter (or more precisely an odd and even Gabor filter pair) are the radial frequency (f) and orientation (θ) with respect to the horizontal axis. For each radial frequency f, multiple filters are obtained by sampling around the circle of radius f with sampling interval $\Delta\theta$. A total of $180/\Delta\theta$ filters are required as conjugate symmetry is exploited.

The resulting sequence of $180/\Delta\theta$ filtered images at each frequency is used to obtain rotation invariant texture features. The energies of the filtered images form a periodic function of θ with period π. The magnitudes of the periodic function's Fourier coefficients are invariant to image rotations (since a rotation of the image corresponds to a translation of the periodic function and Fourier magnitudes are invariant to translations). The first n magnitudes are used as n rotation invariant texture features. Thus for a set of M radial frequencies, we obtain a total of n*M rotation invariant texture features. The following parameters are used in the experiments (see [7] for details): f=2,4,8,16,32,64, n=3 and $\Delta\theta=10°$.

2.3 Wavelet Decomposition and HMM

In this method a quadrature mirror filter (QMF) bank is used as the wavelet transform to decompose the image into sixteen subbands. The multirate digital filter bank is composed of analysis banks (decimators) which split the signal into consecutive bands in the frequency domain and synthesis banks (interpolators) which reconstruct the signal without loss of information. Chen and Kundu use a 2-D (tree structure) FIR QMF bank in their method, and decomposition is accomplished via repeated FIR filtering and (row/column) decimation by a factor of 2.

Features are extracted from each of the 16 image bands (LLLL, LLLH, ..., HHHH). Two feature sets are employed. The first consists of third- and fourth-order central moments both of which are normalised with respect to the second order central moments. The second feature set requires histogram equalisation of the images and consists of normalised entropy and energy.

HMMs are used for training and classification because underlying fundamental structures of the image (which may or may not be directly observable) are found. It is assumed that an image can be modelled by a statistical process whose states are not directly observable. Each state is associated with a probability distribution of an observable quantity, for example, the grey levels. Transition probabilities govern the transitions between states. An observation can be generated at a particular state according to the relevant probability distributions; the states remain hidden though the observations are visible.

The elements of an HMM are defined as follows: The number of states (N), the states $\{q_1...q_N\}$ (Q), the length of the observation sequence (T), the observation sequence $\{O_1...O_t\}$ (O), the initial state probabilities $\{\pi_1\}$ (Π), the probability of (q_i at t=1) (π_I), the state transition probability $\{a_{ij}\} = P(q_j$ at t+1 | q_i at t) (A), and the observation probability $\{b_j(O_t)\} = b_j(O_t) = f(O_t|q_j)$ (B), where f is a density function, x_t is a stochastic process and $f(x_t|x_{t-1} x_{t-2} ...) = f(x_t|x_{t-1}) = f(q_t|q_{t-1})$.

HMM (containing 9 states for this application) is created for each texture class. The segmental k-means algorithm [14] is used which converges to the state optimised likelihood function (Equation 7) for the assumed Gaussian density:

$$P\left(O,Q^*|\lambda\right) = \max P(O,Q|\lambda)$$

$$= \max_Q \pi_{q1} b_{q1}(O_1) \prod_{t=2}^{T} a_{qt-1qt}(O_t) \tag{7}$$

where $Q^* = \{q_1^*, q_2^*, \ldots, q_T^*\}$ is the optimal state sequence associated with the state optimised likelihood function.

During classification the unknown texture O (i.e., the feature sequence derived from the texture) is classified to one of the M models λ_m (m=1..M) via the calculation of the state optimised likelihood function using the Viterbi algorithm [13]. Texture O is classified to m^* by:

$$m^* = \arg \max_M P(O, Q^* | \lambda_m) \qquad (8)$$

2.4 The CSAR Method

A circular simultaneous autoregressive (CSAR) model was developed by Kashyap and Khotanzad for the extraction of rotation invariant texture features. Spatial interaction models such as this represent the grey level value at a pixel s (y(s)) as a linear combination of its neighbours plus a noise component. In this circular model the interpolation of a pixel's eight neighbours is required to obtain the values of N_c (the circular neighbour set).

The eight neighbours of a pixel are interpolated according to:

$$y(s) = \alpha \sum_{r \in N} g_r y(s \oplus r) + \sqrt{\beta} v(s) \qquad (9)$$

where g_r = interpolation kernel; r = neighbour of current pixel s, $s \in \Omega$; v(s) = a correlated sequence with zero mean and unit variance; and N = neighbour set {(0,1) (0,-1) (1,0) (-1,0) (1,1) (-1,-1) (-1,1) (1,-1)}.

The least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ of the model parameters α and β are used as two rotation invariant texture features in this method. The performance of $\hat{\alpha}$ and $\hat{\beta}$ on their own is relatively poor so a third feature ζ (which captures directionality measures) is required. Two SAR models (with neighbour sets N_a and N_b) are fitted to the image and the maximum likelihood (ML) estimation is used to estimate the model parameters θ from which ζ is calculated as follows:

$$\zeta = \max\left(|\theta_{(1,0)}^* - \theta_{(0,1)}^*|, |\theta_{(1,1)}^* - \theta_{(1,-1)}^*| \right) \qquad (10)$$

where the SAR model is defined by:

$$y(s) = \sum_{r \in N} \theta_r y(s \oplus r) + \sqrt{\rho} w(s) \qquad (11)$$

and $\theta_{(x,y)}^*$ the ML estimates of corresponding parameters: w(s) an identical and independently distributed sequence, $\theta_r = \theta_{-r}$, $N_a = \{(0,1)(0,-1)(1,0)(-1,0)\}$ and $N_b = \{(1,1)(1,-1)(-1,1)(-1,-1)\}$.

3. Comparison under a Small Database

A database of 450 images taken from ten Brodatz texture classes [5], shown with bold italic labels in Figure 1, was used as a testbed for initial experimentation. Each 512*512 pixel image was rotated by an arbitrary angle and randomly cropped to 128*128 pixels.

The resulting images were subjected to histogram equalisation to prevent bias towards images possessing similar grey levels.

For classification, half the database was used for training and the remaining half for testing. Unlike in previous work (e.g. [4]) the training and test images are at different orientations. For simplicity the Euclidean classifier is used where appropriate.

In the equivalent human classification tests a texture is presented to a subject who then returns a classification. To prevent bias due to strong / weak subjects and the learning element incurred during such experiments the following precautions were taken: A number of different subjects from a variety of backgrounds and ages were involved. Each subject classified an equal number of images from each Gaussian noise level. The order in which noise levels were presented to the subjects was randomised as was the order in which textures appeared. A time limit was not imposed on the subjects as the computational methods took various times to execute.
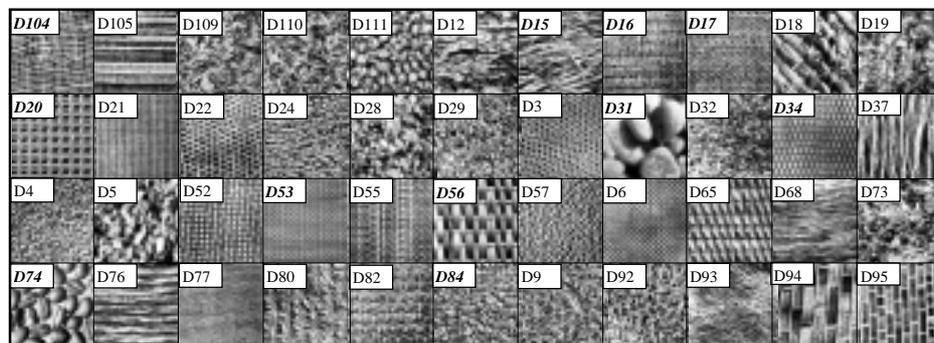

Figure 1 *The texture classes used for initial testing (bold italic labels) and final testing (all textures)*

In practice images are often noisy and the performance methods under noise conditions must be characterised. Gaussian noise at various levels ($\sigma=10,30,50,70,90$) was added to each image in the database. The resulting 2250 (450*5) noisy images were classified using the exemplar feature vectors derived from the original noise free images.

3.1 Comparison of Classification Accuracy

The classification results are presented in Figure 2. This database posed no problems for the human subjects who, along with the multichannel Gabor filtering method, attained a perfect classification rate. The edge attribute method performed surprisingly well achieving a recognition rate of 84%. It is interesting to note that the CSAR method gave similar results to the HMM method using feature set 1, while feature set 2 was slightly more accurate.

A breakdown of the results for each texture class is given in Table 1. Not surprisingly, the edge attribute method gave the best results for regular textures as the graphs produced are distinctive. Textures such as D31 are coarse (contain large patches of smooth variation), therefore, there are fewer edges to produce unique histograms. In

contrast textures such as D15 are the easiest textures to identify for humans as they appear very different from the remaining classes.

Texture	D104	D15	D16	D17	D20	D31	D34	D53	D56	D84
Edge Att.	100	54	83	100	100	39	100	67	98	100
HMM Set 1	44	78	91	78	48	74	78	48	76	15
HMM Set 2	63	87	83	37	72	100	80	76	67	46
CSAR	30	61	67	87	13	100	41	100	100	100

Table 1 Classification results for each texture class

3.2 Comparison of Noise Robustness

The experimental results (see Figure 2) show that humans are much more resilient to noise than the computational methods as the overall classification accuracy for humans dropped only 5% (from 100% to 95%) at a noise level of σ=90. Computational techniques require a strict definition of each texture class whereas humans rely mainly on fuzzy knowledge, therefore, deviations (e.g. due to noise) from the established texture class have more effect on the former.

It can be seen from Figure 2 that the Gabor method is the most robust computational technique to Gaussian noise. A level of approximately σ=50 is reached before the drop in classification accuracy becomes severe (the remaining methods have already reached random classification ($\approx$10%) by this point). Feature set 2 of the HMM method is greatly effected by low levels of noise. This is due to the fact that the first element of the feature vector is entropy which increases dramatically upon the addition of noise.
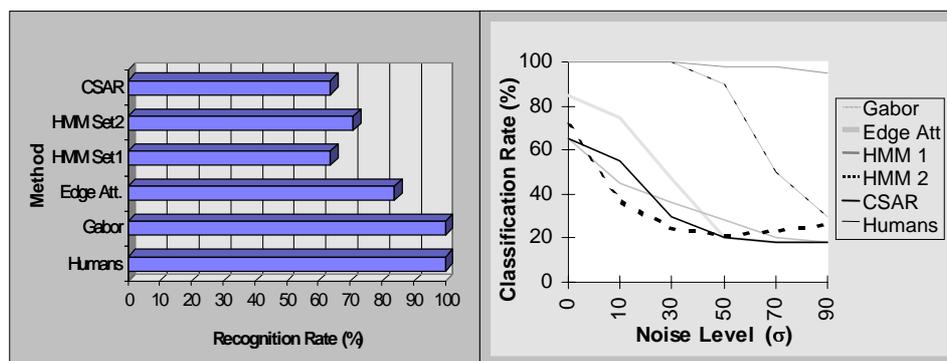

Figure 2 Classification results (left) and noise results (right) for the small database

The plots from the remaining methods are very similar. The edge attribute and CSAR curves follow the same pattern. A slight deterioration in classification accuracy is apparent at a noise level of σ=10 and a significant drop in recognition rate follows

between noise levels σ=10 and σ=50. The results then remain reasonably constant at an approximate level of 10% (random classification).

Feature set 1 of the HMM method possesses a more shallow and almost linear plot. Feature set 2 is greatly affected by low levels of noise. This is due to the fact that the first element of the feature vector is entropy which increases dramatically upon the addition of noise. The plot levels out as more noise is added and the classification accuracy rises slightly at a level of σ=50.

4. Comparison under A Large Database

A second large database was created in order to thoroughly test the capabilities of each method. 1320 images were generated, from the 44 texture classes shown in Figure 1, in the same manner as the small database (see Section 3). The noise tests were only performed for the Gabor and human methods (the remaining methods are not accurate to obtain interesting results for the noise experiments as the recognition rate is very low). The number of states for the HMM was raised from 9 to 20 as this gave significantly better results.

4.1 Classification Accuracy

The classification results are shown in Table 2. This investigation accentuates the difference in the results of similar experiments performed on the small database. The methods at both the upper and lower ends of the scale are ranked giving a clearer indication of their relative performances. It is interesting to note that the Gabor method gives superior results to humans on the large database.

Experiment	Database	HMM 1	CSAR	HMM 2	Edge	Humans	Gabor
CBR	Large		40		63	84	98
Classification	Large	20	26	45	53	85	94
Classification	Small	64	64	71	84	100	100

Table 2 Classification and CBR results for the six methods (%)

4.2 Noise Resistance

The noise results for the large database are consistent with those for the small database. It can be seen from Table 3 that humans are more resistant to noise than the Gabor method. A 30% drop in classification rate (from σ=0 to σ=90) is apparent for humans compared to a 65% drop using the Gabor method.

	σ=0	σ=50	σ=90
Humans	85	80	55
Gabor	94	45	29

Table 3 Classification rates at different noise levels using the large database (%)

5. Content Based Retrieval

Content based image retrieval is an important application of texture analysis. With the advent of digital cameras and rise in image databases the need for automatic content based search tools is increasing dramatically. The tests performed in this paper are based on similar image retrieval. A query image is presented to the retrieval system which returns the n (=5) closest images from the database.

The HMM method is not suited to this application as each database image forms a separate class and must be individually trained (as each texture class was in during classification). The result is a HMM consisting of just one state for each class which is obviously infeasible for texture analysis.

5.1 Retrieval Results

The results for CBR on the large database are given in Table 2. The Gabor method was the most successful as 98% of all images returned from a search belonged to the same texture class as the query image. Textures of a statistical nature (e.g. D57) proved problematic for the human subjects (e.g. mix with D4) causing the majority of mistakes as they appear very similar to the human eye (due to the lack of easily identifiable features).

The results show that the computational methods are better for content based image retrieval than for classification though this is not the case for humans. The largest difference between CBR and classification results is 24%, observed for the least accurate CSAR method. The smallest difference of 4% is for the most accurate Gabor method. These results are due to the fact that improvement is inversely proportional to the accuracy of the results.

6. Conclusions

A detailed comparison is given between four computational methods of rotation invariant texture classification (the multichannel Gabor filtering method, the HMM with wavelet decomposition method, the CSAR method and the edge attribute processing method) and humans. The experiments performed include classification, resistance to Gaussian noise and content based retrieval. Two databases are used, containing 450 and 1320 images taken from 10 and 44 Brodatz texture classes respectively.

The multichannel filtering method proves to be the most accurate, in terms of classification and CBR accuracy, of all the computational techniques thus validating their increasing usage in texture analysis. The edge attribute method followed leaving the two model based methods, HMM and CSAR, in third and fourth position respectively. This ordering is accentuated when the experiments are performed on the large database.

The computational techniques performed better at content based image retrieval than at classification problems on the same database, whereas similar results were obtained in both applications for humans. The Gabor method gave superior results to humans in all experiments except classification on the small database in which both methods achieved a 100% recognition rate. The human subjects performed significantly

better than the Gabor method under conditions of high level Gaussian noise ($\sigma > \approx 50$). More generally it was found that humans are more resistant to Gaussian noise than all the computational techniques tested.

The Gabor method was selected for inclusion in RAIDER due to its high accuracy in both classification and retrieval problems. The edge attribute method will also be used as it is reasonably accurate and is very efficient rendering it highly suited to applications such as image retrieval.

7. References

[1]   T. N. Tan, Geometric Transform Invariant Texture Analysis, *Proc. of SPIE*, Vol. 2488, pp475-485 (1995).

[2]   T. N. Tan, Noise Robust and Rotation Invariant Texture Classification, *Proc. of EUSIPCO-94*, pp1377-1380 (1994).

[3]   G. M. Hayley and B. M. Manjunath, Rotation Invariant Texture Classification using Modified Gabor Filters, *Proc. of IEEE ICIP95*, pp262-265 (1994).

[4]   R. Kashyap and A. Khotanzad, A Model Based Method For Rotation Invariant Texture Classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4), pp. 786-804 (1986).

[5]   P. Brodatz, Textures: A Photographic Album for Artists and Designer, Dover, New York (1966).

[6]   T. Reed and J. M. H. du Buf, A Recent Review of Texture Segmentation and Feature Extraction Techniques, *CVGIP: Image Understanding,* Vol. 57, pp359-372, (1993).

[7]   S. R. Fountain and T. N. Tan, Extraction of Noise Robust Invariant Texture Features via Multichannel Filtering, *Procs. of ICIP '97*, Vol 3, pp197-200, (1997).

[8]   G. Eichmann and T. Kasparis, Topologically Invariant Texture Descriptors, *CVGIP*, Vol. 41, pp267-281, (1988).

[9]   J-L. Chen and A. Kundu, Rotation and Greyscale Transform Invariant Texture Identification using Wavelet Decomposition and HMM, *IEEE Trans. PAMI*, Vol. 16, No. 2, (1994).

[10]   G. Peake and T. Tan, Script and Language Identification from Document Images, *Procs. of BMVC'97*, Vol. 2, pp610-619 (1997).

[11]   M. Tsatsanis and G. Giannakis, Translation, Rotation and Scaling invariant object and texture classification using polyspectra. *SPIE, Advanced Signal Processing Algorithms, Architectures and Implementations,* Vol. 1348, (1990).

[12]   S. Fountain and T. Tan, Efficient Rotation Invariant Texture Features For Content Based Image Retrieval and Annotation of Image Databases, *To appear in Pattern Recognition*.

[13]   L. Rabiner and B. Juang, An Introduction to Hidden Markov Models. *IEEE ASSP Mag*, pp4-16, June (1986).

[14]   L. Rabiner, J. Wilpon and B. Juang, A Segmental K-Means Training Procedure for Connected Word Recognition, *AT&T Technical Journal*, Vol.65, No.3, pp21-31, (1986).

[15]   S. Fountain and T. Tan, Rotation Invariant Retrieval and Annotation of Image Databases, *Procs. of BMVC'97*, Vol.2, pp390-399, (1997).