USING COLOUR GABOR TEXTURE FEATURES FOR SCENE UNDERSTANDING

C J Setchell, N W Campbell

University of Bristol, UK

Gabor filters have been used extensively as a model of texture for image interpretation tasks. This paper demonstrates that when a bank of Gabor filters is applied to an image, there are strong relationships between the outputs of the different filters. These relationships are used to devise a new texture feature which is capable of describing texture information in a concise manner. Information about the distributions of filter responses is also encoded in the new feature. Performance of the feature is assessed by applying it to an image region classification task and comparing results to those obtained using features which do not utilise the relationships between filter outputs. It is shown that the distribution information aids the classification task. The new feature performs comparably with the other features whilst yielding a significantly smaller feature vector. We then describe how the feature may be applied to colour images. It is shown that the inclusion of colour information is beneficial to the classification task and also that the choice of colour space is important. The classification results are then compared to those obtained using a 28 element feature encoding colour, position, shape, size, context and also texture. The new colour Gabor feature outperforms the more intuitive 28 element feature. We conclude by suggesting that the Gabor based feature may be capable of implicitly encoding some shape and context information.

INTRODUCTION

Encoding textural information into a feature vector that can be used for image understanding tasks is a non-trivial problem. Early investigations focused on statistical approaches, Conners and Harlow (1) and Gotlieb and Kreyszig (2). Simple first order statistics, however, are severely limited. Image regions that are visually different can have the same first order statistics. For example, a region containing a chequer pattern of alternating black and white squares would have the same mean and standard deviation of grey value as a region divided into a black left half and a white right half. First order statistics are not able to distinguish between such textures. The use of second order statistics enables the situation to be improved by taking into account not just the grey levels of pixels but also the spatial relationships between them. Other researchers have made use of Markov Random Fields for texture modelling, Cross and Jain (3). More recently, researchers have investigated various filtering approaches. Such approaches employ a set of filters to decompose the image into components having different spatial frequencies, and perhaps different orientations, Malik and Perona (4) and Chang and Kuo (5).

In this paper we present a new texture feature based on the outputs of filters from a Gabor filter bank. Gabor filters have been used extensively as a model of texture for image interpretation tasks. Daugman (6) utilised Gabor filters in an automatic system for verifying a person's identity using images of the iris. Jain *et al* (7) used Gabor filters to perform segmentation of objects (tanks, cars, musical instruments, etc.) from complex backgrounds. Many other researchers have investigated Gabor filters for performing texture segmentation; Dunn and Higgins (8), Jain and Farrokhnia (9) and Weldon and Higgins (10).

A typical 'daisy-petal' Gabor filter bank will contain in excess of a dozen filters. The filter bank used in this study contains 15 filters, 5 frequencies and 3 orientations, each filter having a frequency bandwidth of 1.8 octaves and an angular bandwidth of 1.04 radians. The response (in the spatial domain) of such a filter bank to an input image is a set of 15 filtered images. When designing a feature to capture the response of all 15 filters over a region of pixels, a naive approach will lead to a large feature vector. In this paper we establish that relationships exist between the responses of filters at different frequencies and orientations. By exploiting these relationships it is possible to capture the characteristics of a region of texture in a small, concise feature vector. A detailed description of this feature is given in a later section.

The paper describes a number of experiments which have been carried out to assess the performance of the new feature, initially using grey-scale and then colour images. These experiments involve applying the feature to an image region classification task. The feature is used as the input to a neural network classifier which is trained and then tested on previously unseen image regions. The classification performance of the feature is compared to features which do not utilise the relationships between filter outputs. These results are discussed in the final conclusions section.

THE GABOR FILTER BANK

In 1946, Gabor (11) derived a set of filters which (in 1D) jointly optimise uncertainty in time and frequency. Gabor filters were extended by Daugman (12) in 1985 from 1D to 2D, providing a set of filters which jointly optimise uncertainty in 2D spatial resolution, spatial frequency and orientation. It is believed that early stages of the human visual system may be explained by a number of channels, each tuned to a different frequency and orientation (4). The characteristics of these channels may be modelled by a 'daisy-petal' Gabor filter bank as shown in Figure 1.

When considered in the 2D spatial domain, the Gabor filter is a Gaussian modulated sinusoid. It has parameters to determine the frequency and orientation of the sinusoid, and two parameters governing the spatial extent of the modulating Gaussian. In the frequency domain the Gabor filter is a 2D Gaussian which can be configured to pass any elliptical region. There are three types of filter: *odd-symmetric*, *even-symmetric* and *complex*. In an even-symmetric filter, the sinusoid is a cosine function, for the odd-symmetric it is a sine function, whereas for the complex filter the sinusoid is also complex, having both sine and cosine components.

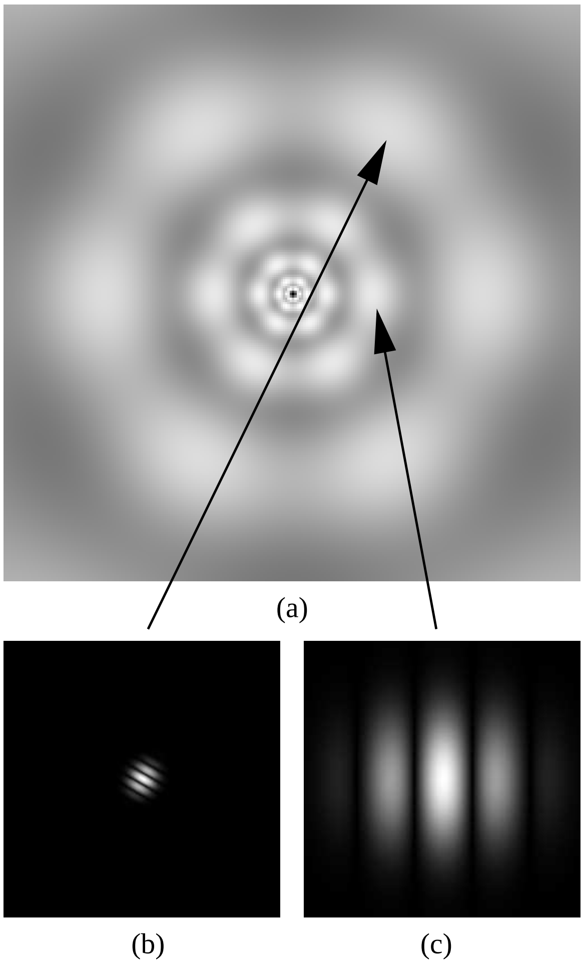

Figure 1: (a) daisy-petal Gabor filter bank in frequency domain, (b) high frequency filter in spatial domain, (c) low frequency filter in spatial domain

THE GABOR TEXTURE FEATURE

To apply the filters in the 'daisy-petal' Gabor filter bank to an image, the first step is to calculate the FFT of the image. This transforms the image from a spatial domain representation to a frequency domain representation. For each filter in the filter bank, the frequency domain representation of the filter is multiplied, on a pixelwise basis, with the frequency domain representation of the image. Complex filters have been used throughout this work. Applying the inverse FFT takes the filtered image from the frequency domain back to the complex spatial domain. The magnitude operator is then applied to each image, phase information being ignored. The result of applying all the filters in the Gabor filter bank is thus a set of filtered images.

The original image is then segmented using a K-means clustering algorithm, yielding a set of image regions. The use of regions enables higher level concepts such as size and shape to be considered. Naively describing the responses of all filters over a particular image region will lead to a large feature vector. For example, if the filter bank contains 15 filters, taking the mean response over an image region for each filter will lead to a 15 element feature vector.

The feature vector which has been derived during this work utilises (linear) relationships between filter responses resulting in a concise encoding of textural information. The feature vector comprises two components (Figure 2). The first is derived from the mean responses of each of the 15 filters across a region. Principal component analysis was applied to these mean filter responses from many thousands of image regions. It was found that the top 5 components explains over 94% of the variance within the data set, confirming that strong linear relationships do exist between the various filter outputs. The response to these 5 principal components constitute the first 5 elements of the texture feature vector.

The remaining 5 elements of the texture feature are associated not with the mean response of each fil-

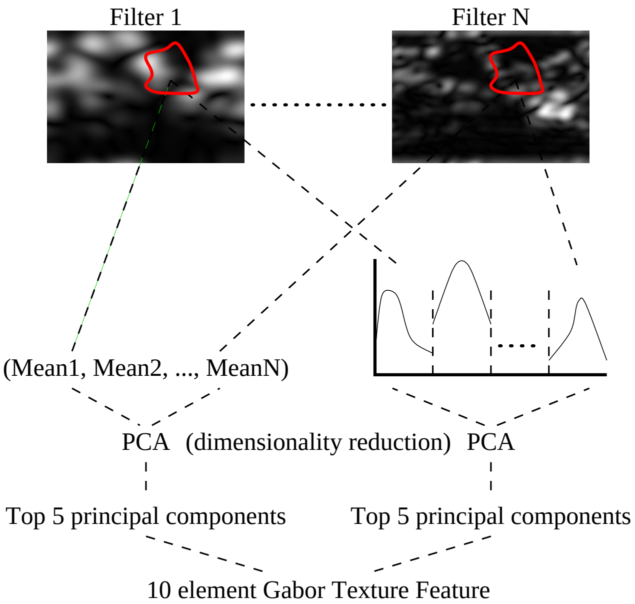

Figure 2: The Gabor Texture Feature

Mode	% variance	Cumulative % variance
1	66.9	66.9
2	14.3	81.2
3	5.6	86.8
4	4.4	91.2
5	2.9	94.1

Table 1: First component of feature: Percentage of variance within data set captured by each of the top 5 modes

Mode	% variance	Cumulative % variance
1	28.7	28.7
2	10.5	39.2
3	5.0	44.2
4	3.9	48.1
5	3.1	51.2

Table 2: Second component of feature: Percentage of variance within data set captured by each of the top 5 modes

ter but with the distribution of each filter's response across a region. Previously, standard deviation has been used to characterise the filter's distribution (Jain and Farrokhnia (9) actually used mean absolute deviation) but Dunn and Higgins (8) has suggested these distributions to be Ricean rather than Gaussian, indicating that standard deviation is not necessarily appropriate. Rather than using the Rice distribution's parameters, an alternative model of the distribution has been derived. This model was created by concatenating the distributions (centred about the mean) of each filter into a single vector and performing principal component analysis. The top 5 principal components are used as the remaining 5 elements in the texture feature vector. This approach elegantly models the shape of the filter distributions as well as extracting the relationships between the individual filters.

Applying the feature to colour images is simply a matter of extracting the feature from each colour band. An RGB image would therefore result in a $3 \times 10 = 30$ element feature vector.

RESULTS

Table 1 shows the results of the PCA analysis of the filter responses for the first component of the feature, based on the mean filter responses. Results for the second component of the feature, based on the distributions of the filter responses, can be found in Table 2. It is clear from Table 1 that there are indeed strong relationships between the mean responses of the different filters in the filter bank. The PCA analysis identifies these linear relationships and enables the feature to concisely describe the textural information using only a small number of elements. The relationships between the distributions of the filter responses are less strong. However, it will be shown later that this component of the feature does still capture important textural information.

In order to assess the performance of the feature vector, it has been applied to the task of classifying regions from images taken from the Bristol Image Database, Campbell *et al* (13). This consists of ≈ 300 high quality colour images for which a ground truth labelling exists in which every image region has been manually classified into one of 11 generic classes (sky, road, vegetation, etc.). The ground truth labelling enables classifiers to be trained and their performance assessed quantitatively. Each image has been segmented into regions using a K-Means algorithm, a value of $K = 4$ giving a slight over-segmentation. These regions were divided between a training set, a validation set and a test set in the ratio 70:15:15. An MLP neural network classifier was then trained using regions from the training set and the Scaled Conjugate Gradient descent training algorithm. After each iteration of training, the classifier error is calculated using regions from the validation set. Training of the network is terminated when this error begins to rise as it signifies the network is becoming over trained. The performance of the network is then assessed using the previously unseen regions from the test set. Networks with a single hidden layer, having a range of hidden units from 4 to 60 are trained in order to find an optimum, which

Feature	No. of elements	% of regions correctly classified
Mean	15	73.4
Mean & s.d.	30	77.6
New feature	10	75.5

Table 3: Classification results for feature based on mean, feature based on mean & standard deviation and the new texture feature

No. of modes	% of regions correctly classified
5 + 2	72.7
5 + 4	73.3
5 + 5	75.5

Table 4: Classification results for new feature using 5 modes for first component of feature and 2, 4 & 5 modes for the second component

is typically ≈ 20 hidden units.

For the first set of experiments, the colour images in the database were converted to grey-scale by calculating the mean of the red, green and blue components of each pixel.

As a benchmark, the performance of two feature vectors which do not utilise the relationships between filter responses were assessed. The first of these features was based solely on the mean response of each filter over a region. This feature had 15 elements, one for each filter in the filter bank. The second feature contained both the mean and the standard deviation of the filter response over a region, yielding a 30 element feature vector. These two features, together with our new feature, were then applied to the image region classification task outlined above. The classification accuracy of each feature is shown in Table 3. These results demonstrate that the new feature is capable of capturing a comparable amount of textural information to the naive approaches but in a significantly smaller, more concise vector.

Regarding the second component of the feature and its contribution to capturing the texture information, Table 4 shows that if fewer modes are used in this component the classification accuracy decreases. This confirms that the distributions of the filter responses are important to the classification task and that the 5 modes used to describe them are significant to the performance of the feature.

For the final experiments, colour images were used. Three different colour spaces were investigated;

Colour space	% Classification accuracy
RGB	79.6
$LC_{rg}C_{yb}$	83.9
L*u*v*	85.5

Table 5: Classification accuracy for RGB, $LC_{rg}C_{yb}$ and L*u*v* colour spaces

RGB, $LC_{rg}C_{yb}$ and L*u*v*, Glassner (14). $LC_{rg}C_{yb}$ is based on the opponent colour model of the human visual system while L*u*v* is a perceptually linear colour space. Table 5 shows the classification accuracy achieved by the feature when applied to colour images represented in each of the three colour spaces.

CONCLUSIONS

This work demonstrates that there are indeed strong relationships between the filter responses and the feature vector described here successfully utilises these inter-dependencies resulting in a concise description of textural information. Also, the distribution of filter responses is important and its inclusion in the new feature makes a significant contribution to its success at the image region classification task.

Clearly, the inclusion of colour information leads to a significant increase in classification accuracy over the use of solely grey-scale information. Table 5 shows also that the choice of colour space is extremely important. Although a neural network should be able to learn any colour space, it is clear that the learning task becomes easier in L*u*v*.

Comparing these results with those obtained during previous work by Campbell *et al* (15), in which a 28 element feature vector was applied to the same 11 class classification problem yields an interesting observation. The 28 element feature contained not just colour and texture but also size, position, shape and contextual information. This feature vector achieved a classification accuracy of 82.9% compared to the 85.5% achieved when the 30 element Gabor feature is applied in the L*u*v* colour space. This suggests one of two things. Either size, position, shape and contextual information play an insignificant role in image interpretation or they are partially encoded in the Gabor filter response. In a study by Clark *et al* (16), it has been shown that when humans perform this same 11 class classification task, little use is made of size and position information. This result seems intuitive as size and position are highly dependent upon view point. The importance of shape and contextual information were not thoroughly assessed by Clark *et al.* but the contribution of context at least is likely to be high. Consider an image region containing a road marking (white line). This is far more likely to be near areas of road than areas of sky. Context is captured by filters having large spatial extents, enabling them to bring in information from areas surrounding the region of interest. Shape is a little harder to understand, but perhaps the high frequency filters respond to the edge information which typically coincides with the region border. As the filters are oriented, this amounts to a basic shape descriptor. We therefore believe that the contribution of size and position to the classification task are negligible but the more important contextual, and possibly shape, information is partially captured implicitly by the Gabor filter.

In future work, the problem of choosing filter bank parameters will be tackled. So far, our work has used a filter bank consisting of 15 filters, 5 frequencies and 3 orientations, arranged in a 'daisy-petal' configuration. However, it is unlikely that this configuration is optimal. A method is required to choose filter bank parameters which are optimal for the particular problem in hand.

ACKNOWLEDGEMENTS

We thank British Aerospace PLC, Sowerby Research Centre, Bristol for their support with the development of the database.

REFERENCES

1. Conners R and Harlow C, 1980, "A theoretical comparison of texture algorithms", Trans. on Pattern Analysis and Machine Intelligence, 2, 204-222

2. Gotlieb C and Kreyszig H, 1990, "Texture descriptors based on co-occurrence matrices" Computer Vision, Graphics and Image Processing, 51, 70-86

3. Cross G and Jain A, 1983, "Markov random field texture models" Trans. on Pattern Analysis and Machine Intelligence, 5, 25-39

4. Malik J and Perona P, 1990, "Preattentive texture discrimination with early vision mechanisms" Journ. of the Optical Soc. of America, 7, 923-932

5. Chang T and Kuo C, 1993, "Texture analysis and classification with tree-structured wavelet transform" Trans. on Image Processing, 2, 429-441

6. Daugman J, 1993, "High confidence visual recognition of persons by a test of statistical independence" Trans. on Pattern Analysis and Machine Intelligence, 15, 1148-1161

7. Jain A, Ratha N and Lakshmanan S, 1997, "Object detection using Gabor filters" Pattern Recognition, 30, 295-309

8. Dunn D and Higgins W, 1995, "Optimal Gabor filters for texture segmentation" Trans. on Image Processing, 4, 847-964

9. Jain A and Farrokhnia F, 1991, "Unsupervised texture segmentation using Gabor filters" Pattern Recognition, 24, 1167-1186

10. Weldon T and Higgins W, 1996, "Integrated approach to texture segmentation using multiple Gabor filters" Proc. Int. Conf. Image Processing, 955-958

11. Gabor D, 1946, "Theory of communication" Journ. of Inst. Electrical Engineers, 93, 429-457

12. Daugman J, 1985, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters" Journ. of the Optical Soc. of America, 2, 1160-1169

13. Campbell N, Mackeown W, Thomas B and Troscianko T, 1995, "Automatic interpretation of outdoor scenes" Proc. of British Machine Vision Conference, 297-306

14. Glassner A, 1995, "Principle of digital image synthesis, Vol 1", Morgan Kaufmann

15. Campbell N, Mackeown W, Thomas B and Troscianko T, 1997, "Interpreting image databases by region classification" Pattern Recognition (special edition on image databases), 30, 555-563

16. Clark A, Troscianko T, Campbell N and Thomas B, 1996, "Image region labelling by humans and by an artificial neural network" Perception, 25