Robust texture features for still-image retrieval

P. Howarth and S. Rüger

Abstract: A detailed evaluation of the use of texture features in a query-by-example approach to image retrieval is presented. Three radically different texture feature types motivated by i) statistical, ii) psychological and iii) signal processing points of view are used. The features were evaluated and tuned on retrieval tasks from the Corel collection and then evaluated and tested on the TRECVID 2003 and ImageCLEF 2004 collections. For the latter two the effects of combining texture features with a colour feature were studied. Texture features that perform particularly well are identified, demonstrating that they provide robust performance across a range of datasets.

1 Introduction

Texture is a key component of human visual perception. Like colour, this makes it an essential feature to consider when querying image databases. Everyone can recognise texture, but it is more difficult to define. Unlike colour, texture occurs over a region rather than at a point. It is normally perceived by intensity levels and as such is orthogonal to colour. Texture has qualities such as periodicity and scale; it can be described in terms of direction, coarseness, contrast and so on [1]. It is this that makes texture a particularly interesting facet of images and results in a plethora of ways of extracting texture features. To enable us to explore a wide range of these methods we chose three very different approaches to computing texture features: the first takes a statistical approach in the form of co-occurrence matrices, next the psychological view of Tamura's features and finally signal processing with Gabor wavelets.

Our study is the first to focus on evaluation of texture features on heterogeneous everyday images, and to tailor features for optimum retrieval performance in this context. The majority of original papers devising or evaluating texture features used classification or segmentation tasks to measure performance [2–5]. The VisTex database [6] contains both reference textures and texture scenes and has been used in several studies. Although its images are real-world scenes they are subsequently divided into homogeneous patches and used in a classification context. Segmentation and classification tasks are significantly different to the problems faced in image retrieval, where one looks at generic queries for an entire picture. Real pictures are made up of a patchwork of differing textures rather than the uniform texture images often used in studies, such as the ones taken from Brodatz's photographic book [7]. To that effect we suggest encoding texture in terms of joint histograms of low dimensional texture characteristics over the image in the same way 3-D colour histograms are computed, we have called this a Tamura image.

IEE Proceedings online no. 20045185

doi: 10.1049/ip-vis:20045185

Paper first received 24th September 2004 and in revised form 8th April 2005

The authors are with the Multimedia Information Retrieval Group, Department of Computing, South Kensington Campus, Imperial College London, London SW7 2AZ, UK

E-mail: peter.howarth@imperial.ac.uk

2 Texture features

2.1 Co-occurrence

Statistical features of grey levels were one of the earliest methods used to classify textures. Haralick [8] suggested the use of grey level co-occurrence matrices (GLCM) to extract second order statistics from an image. GLCMs have been used very successfully for texture classification in evaluations [2].

Haralick defined the GLCM as a matrix of frequencies at which two pixels, separated by a certain vector, occur in the image. The distribution in the matrix will depend on the angular and distance relationship between pixels. Varying the separation vector allows the capturing of different texture characteristics. Once the GLCM has been created, various features can be computed from it. These have been classified into four groups: visual texture characteristics, statistics, information theory and information measures of correlation [3, 8]. We chose the four most commonly used features, listed in Table 1, for our evaluation. Note that $P(a, b)$ is the frequency that quantised grey level a co-occurs with quantised grey level b, separated by a vector v.

Figure 1 is a diagram showing the construction of GLCM. The rows of the co-occurrence matrix represent the value at the start of the separation vector and the columns the value at the end. The circled entry in the GLCM therefore represents the number of occurrences of the separation vector [1, 0] that have a start value of 3 and an end value of 1. The three occurrences of these in the quantised image are shown by the arrows representing the vectors.

2.2 Tamura

Tamura *et al.* took the approach of devising texture features that correspond to human visual perception [1]. They defined six textural features (coarseness, contrast, directionality, line-likeness, regularity and roughness) and compared them with psychological measurements for human subjects. The first three attained very successful results and are used in our evaluation, both separately and as joint values.

Table 1: Features calculated from the normalised co-occurrence matrix $P(a, b)$

Feature	Formula
Energy	$\sum_a \sum_b P^2(a, b)$
Entropy	$\sum_a \sum_b P(a, b) \log P(a, b)$
Contrast	$\sum_a \sum_b (a-b)^2 P(a, b)$
Homogeneity	$\sum_a \sum_b \frac{P(a,b)}{1+\lvert a-b\rvert}$

Coarseness has a direct relationship to scale and repetition rates and was seen by Tamura *et al.* as the most fundamental texture feature. An image will contain textures at several scales; coarseness aims to identify the largest size at which a texture exists, even where a smaller micro texture exists. Computationally one first takes averages at every point over neighbourhoods the linear size of which are powers of 2. The average over the neighbourhood of size $2^k \times 2^k$ at the point (x, y) is

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j)/2^{2k} \quad (1)$$

where $I(i, j)$ is the grey level at the image pixel coordinates (i, j). Then at each point one takes differences between pairs of averages corresponding to non-overlapping neighbourhoods on opposite sides of the point in both horizontal and vertical orientations. In the horizontal case this is

$$E_{k,h}(x, y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \quad (2)$$

At each point, one then picks the size of k that maximises E in either the horizontal or vertical direction; this is k_{opt}. The coarseness measure is then the average of $S_{\text{opt}}(x, y) := 2^{k_{\text{opt}}}$ over the picture.

Contrast aims to capture the dynamic range of grey levels in an image, together with the polarisation of the distribution of black and white. The first is measured using the standard deviation of grey levels and the second the kurtosis α_4. The contrast measure is defined as

$$F_{con} = \sigma/(\alpha_4)^n \quad (3)$$

The kurtosis is calculated by dividing the fourth moment about the mean, μ_4, by the variance squared,

$$\alpha_4 = \mu_4/\sigma^4 = \frac{\sum_i \sum_j (I(i,j) - \mu)^4}{N\sigma^4} \quad (4)$$

where μ is the mean intensity and $N = ij$ (the number of samples).

Experimentally, Tamura found $n = 1/4$ to give the closest agreement to human measurements. Our initial retrieval experiments confirmed this choice, and this is the value we used in our following experiments.

Directionality is a global property over a region. The feature described does not aim to differentiate between orientations or patterns, but measures the total degree of directionality. Two simple masks are used to detect edges in the image. At each pixel the angle and magnitude are calculated. A histogram, H_d, of edge probabilities is then built up by counting all points with magnitude greater than a threshold an quantising by the edge angle. The histogram will reflect the degree of directionality. To extract and measure from H_d the sharpness of the peaks are computed from their second moments. An example image and its direction histogram is shown in Fig. 2. This shows a main peak in the first bin which represents the horizontal direction in the image and a smaller peak in the ninth bin representing the vertical component.

Tamura image is a notion where we calculate a value for the above three features at each pixel and treat as a spatial joint Coarseness-coNtrast-Directionality (CND) distribution, in the same way as images can be viewed as spatial joint RGB distributions. The regional nature of texture means that the values at each pixel are computed over a window. We then extract colour histogram style features from the Tamura CND image, creating both marginal and 3-D histograms. A similar 3-D histogram feature is used by MARS [9].

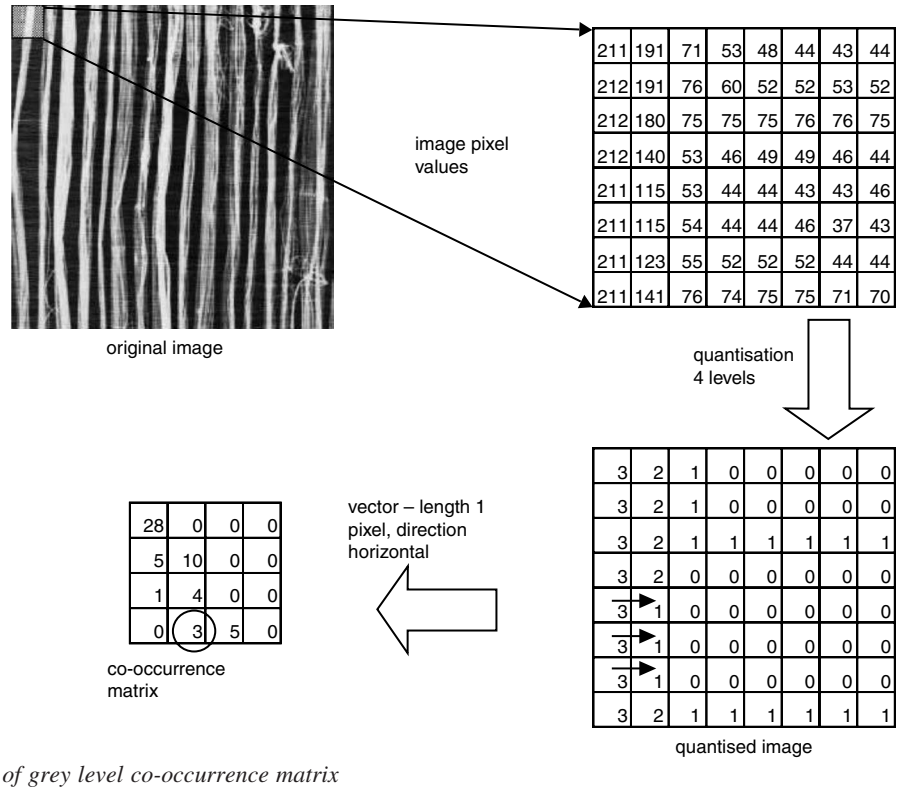

Fig. 1 *Construction of grey level co-occurrence matrix*

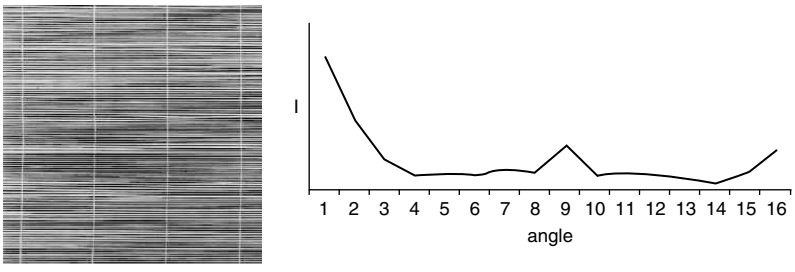

Fig. 2 *Example image and direction histogram*

2.3 Gabor

One of the most popular signal processing based approaches for texture feature extraction has been the use of Gabor filters. These enable filtering in the frequency and spatial domain. It has been proposed that Gabor filters can be used to model the responses of the human visual system. Turner [10] first implemented this by using a bank of Gabor filters to analyse texture. A bank of filters at different scales and orientations allows multichannel filtering of an image to extract frequency and orientation information. This can be used to decompose the image into texture features.

Our implementation is based on that of Manjunath *et al.* [11, 12]. The feature is computed by first filtering the image with a bank of orientation and scale sensitive filters and then computing the mean and standard deviation of the output in the frequency domain.

In the spatial domain a Gabor function is a sinusoid modulated by a Gaussian with an aspect ration of σ_x/σ_y. A two-dimensional Gabor function can therefore be expressed as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jFx\right) \quad (5)$$

We can generate a bank of self-similar filters by appropriate dilations and translations of the parent wavelet $g(x, y)$. For this we Manjunath's method [11] using the generation function:

$$\begin{aligned} g_{mn}(x,y) =& a^{-m}g(x',y') \text{ with } a>1 \text{ and } m,n \text{ both integers,} \\ & x' = a^{-m}(x\cos\theta + y\sin\theta) \text{ and} \\ & y' = a^{-m}(-x\sin\theta + y\cos\theta) \end{aligned} \quad (6)$$

where $\theta = n\pi/K$ and K is the total number of orientations. The scale factor a^{-m} is meant to ensure that the energy is independent of m.

We can now generate our bank of filters following a design strategy that ensures that the half-peak magnitude supports of the filter responses in the frequency domain touch each other [11]. An example of a filter bank generated in this way is shown in Fig. 3.

Filtering an image I with Gabor filters g_{mn} results in its Gabor wavelet transform:

$$W_{mn}(x,y) = \int I(x_1,y_1)g^*_{mn}(x-x_1,y-y_1)\,dx_1dy_1 \quad (7)$$

The mean and standard deviation of the magnitude $|W_{mn}|$ are used for the feature vector. The outputs of filters at different scales may be over differing ranges. For this reason each element of the feature vector is normalised using the standard deviation of that element across the entire database.

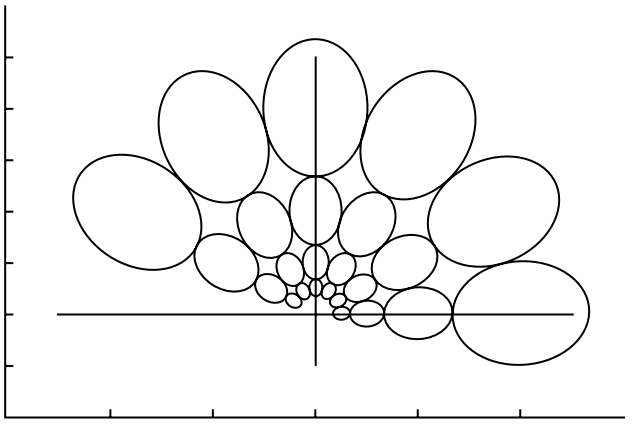

Fig. 3 *Filter responses of a Gabor filter bank with 4 scales and 6 orientations*

3 Experimental setup

We used a three-stage approach for our work. Initial evaluation and modification to the features were tested using a carefully selected subset of the Corel image library. We then ran larger tests on the TRECVID 2003 [13] data to confirm our findings and completed a submission to the image track of the Cross Language Evaluation Forum [14, 15] (ImageCLEF).

3.1 Image collections

3.1.1 Corel: We selected 6,192 images from the Corel collection to give 63 categories that were visually similar internally, but different from each other [16]. A set of 630 single-image category queries was executed to test performance across all categories. Relevance judgements on the retrieved images were based on the categorisation. The results shown in Section 4 are the corresponding mean average precision (m.a.p) as defined in [17].

3.1.2 TRECVID 2003: A second, larger image collection was used to give a more realistic performance comparison. This comprised of 32,318 key-frames from the TRECVID 2003 collection [13]. The search task specified for TRECVID 2003 consisted of 25 topics; for each topic several example images were given as a query. The published relevance judgements for these topics were used to evaluate the retrieval performance for different features and combinations of features.

3.1.3 ImageCLEF: The collection for this evaluation comprised of 8,725 medical images and 26 single image queries. Medical images have very specific characteristics. The majority are monochrome images, such as X-rays and CT scans, with a very formulaic layout.

3.2 k-nearest neighbours

Distances between feature vectors were calculated using the Manhattan metric. The resultant distances were then median normalised to give even weighting when combined. For multiple image queries with the TRECVID collection we used a version of the distance weighted k-nearest neighbour measure (k-nn) from Mitchell [18]. We have a baseline for evaluation from previous work with the TREC dataset for which k-nn has consistently proved the best retrieval method [16].

k-nn is calculated using positive image examples (P) supplied as the query and negative examples (N) randomly selected from the collection. To rank an image i in the collection we identify those images in P and N that are amongst the set K of k-nearest neighbours of i. Using these neighbours we determine the dissimilarity as

$$D(i) = \frac{\sum_{n \in N \cap K} (d(i,n) + \epsilon)^{-1}}{\sum_{p \in P \cap K} (d(i,p) + \epsilon)^{-1} + \epsilon} \quad (8)$$

where $d()$ is the distance function and ϵ is a small positive number to avoid division by zero. For our experiments we used a value of $k = 40$, which was found to give good performance in [16].

4 Evaluation and results

For each feature we evaluated performance in the configuration described in Section 2. Ideas to improve performance were devised and evaluated. The general themes considered were how best to represent an entire image, how to accommodate differing sizes and scale of images and how to cope with the regional qualities of textures. These evaluations were run on the Corel data. Paired t-tests were carried out to check whether results were statistically significant at $\alpha = 0.05$.

The best performing features from the initial evaluation were then tested on the TRECVID 2003 data set. Tests were run with each texture feature combined with a high performing colour feature.

4.1 Image tiling

As part of the evaluation we carried out experiments to determine the effect of applying features to image tiles. The method was to split the image into non-overlapping tiles and calculate the feature for each tile. The resulting features for each tile are then concatenated into a single vector. Tilings up to 9×9 were evaluated. The effect of tiling is to add location information to the feature. The distance between features will be the sum of the distances between corresponding tiles.

Tiling was evaluated for all three texture features. Retrieval performance increased with tiling, peaking at 7×7 or 9×9 depending on the feature. The graph in Fig. 4 shows the average retrieval for different tilings with the Tamura features. The detailed results can be seen in Table 3.

Beyond this optimum level of tiling there was no significant gain. As the image tiles decrease in area the features have smaller regions to work on, thus reducing their effectiveness. At the limit the feature will approximate to comparing images at pixel level. We find that increasing tiling much beyond 9×9 starts to add noise to the feature for images of a typical linear size of 300 pixels.

4.2 Co-occurrence

The two main variables when creating a GLCM are the number of quantisation levels and the vector. We decided to use four vector angles: 0, 45, 90, 135 and four distances. This could be used to calculate up to sixteen GLCMs. However, as the statistics are not invariant under rotation we also tried summing the four angles at each distance into a single matrix. GLCMs can be made symmetrical by calculating the GLCM using both the specified vector and one in the opposite direction; symmetric and asymmetric matrices were tested. The number of quantisation levels dictate the size and density of the matrix. This may become a problem with small images or tiles. The effect of varying quantisation between 4 and 64 levels was tried. Features were calculated for whole and tiled images.

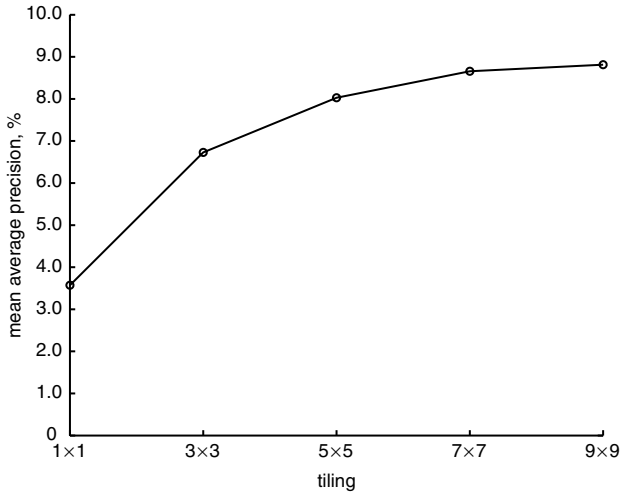

Fig. 4 *Effect of tiling for Tamura features*

Preliminary results showed that distances between 1 and 4 pixels gave the best performance. Tiling of the image gave a large increase in retrieval which flattened out by 9×9 tiles. The results in Table 2 are for 7×7 tiles. Results and implications of the evaluation are below:

- There was no significant difference between generating matrices symmetrically or asymmetrically.
- Increasing quantisation up to 5–6 bits improves performance. The optimum level depends on the feature.
- Concatenated matrices outperformed the rotationally invariant summed matrices.
- With this collection the homogeneity feature performed best with a m.a.p. of 12.2%.

4.3 Tamura

When calculating standard Tamura features for whole or tiled images the main variable is the largest k value for

Table 2: Co-occurrence features - mean average precision retrieval

	Quantisation				
Feature	4	8	16	32	64
Energy: cat	7.6%	8.1%	9.3%	9.9%	9.5%
Energy: sum	7.0%	7.8%	8.9%	9.2%	9.0%
Entropy: cat	8.1%	9.2%	10.4%	11.1%	11.4%
Entropy: sum	7.5%	8.8%	9.8%	10.4%	10.7%
Contrast: cat	8.5%	8.5%	8.4%	8.3%	8.3%
Contrast: sum	7.8%	7.9%	7.7%	7.6%	7.6%
Homogeneity: cat	9.2%	10.2%	11.2%	11.8%	12.2%
Homogeneity: sum	8.5%	9.5%	10.4%	10.9%	11.3%

cat = 16 concatenated matrices

sum = 4 rotationally invariant summed matrices

coarseness. The effect of varying this, and the number of tiles, can be seen in Table 3. The dashes in the table are where the image size resulting from tiling meant that the k value was too large to be used because of the border needed.

With the histogram features the main variable to evaluate was the window size. Coarseness can be calculated at a pixel level. However, both the directionality and contrast features operate over a region. A large window would smear the feature and lose resolution; conversely a small window may invalidate the statistical features, particularly if the directionality histogram is too sparsely populated. To evaluate this the features were run over several window sizes, creating a histogram for each feature, see Table 4.

A little surprisingly, initial results showed that increasing the largest considered k value for coarseness reduced the performance – the optimum value was 2. This may be due to the large borders necessary for higher values of k. However, it is more likely caused by the nature of textures in images and the way the algorithm averages the 2^k values. In normal images textures occur at much smaller scales than in Brodatz style texture images. Correspondingly, there are unlikely to be textures with a coarseness of 64 or 32 pixels in images with a typical diagonal of 450 pixels. The algorithm may still detect noise at this dimension, biasing the average value of the feature. A change to the algorithm was made so that it took the values of k rather than 2^k – effectively introducing a logarithmic scaling of the

Table 3: Standard Tamura features – mean average precision retrieval

	Tiling				
Feature	1 × 1	3 × 3	5 × 5	7 × 7	9 × 9
Contrast	3.2%	6.1%	7.2%	8.1%	8.0%
Directionality: peak finding	2.9%	4.2%	5.0%	5.8%	6.6%
Directionality: entropy	2.7%	5.4%	7.5%	8.9%	9.7%
Coarseness 2^k: max $k = 2$	4.4%	8.3%	9.5%	9.9%	9.9%
Coarseness 2^k: max $k = 3$	3.5%	7.6%	8.8%	9.2%	9.0%
Coarseness 2^k: max $k = 4$	3.5%	7.2%	7.7%	7.0%	–
Coarseness 2^k: max $k = 5$	3.3%	5.7%	–	–	–
Coarseness 2^k: max $k = 6$	2.9%	–	–	–	–
Coarseness k: max $k = 2$	4.4%	8.0%	9.3%	9.6%	9.6%
Coarseness k: max $k = 3$	3.9%	7.5%	8.9%	9.1%	8.9%
Coarseness k: max $k = 4$	3.4%	7.0%	7.7%	7.2%	–

Table 4: Histogram Tamura features – mean average precision retrieval

	Window size			
Feature	2	4	8	16
Contrast	6.0%	6.7%	7.0%	6.9%
Directionality: peak finding	5.4%	5.6%	5.6%	4.9%
Directionality: entropy	4.9%	4.4%	5.2%	5.4%
Coarseness 2^k : max $k = 2$	6.9%	6.0%	6.1%	6.0%
Coarseness 2^k : max $k = 3$	6.5%	5.9%	6.0%	5.8%
Coarseness 2^k : max $k = 4$	6.1%	5.7%	5.6%	5.4%
Coarseness k: max $k = 2$	6.4%	10.0%	9.8%	8.2%
Coarseness k: max $k = 3$	5.7%	10.1%	9.2%	7.9%
Coarseness k: max $k = 4$	8.8%	9.3%	8.1%	7.7%

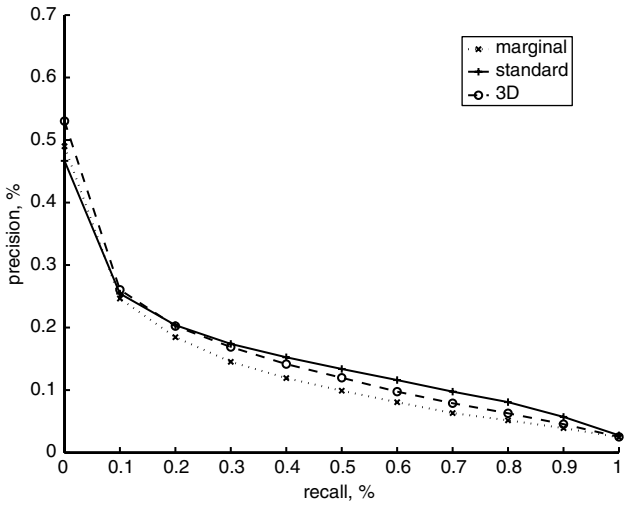

Fig. 5 *Precision-recall for combined Tamura features*

coarseness and giving less influence to the larger scales. This gave a significant increase in performance for the histogram, from 6.1% to 10.1%, but no improvement when applied to the standard feature.

Performance of the directionality feature was poor. A detailed look at the operation of the algorithm showed that this was largely a result of the sparse population of the histogram and subsequent difficulty in calculating valid variance of its peaks. Several options for improvement were tried including calculating global variance of the histogram and using entropy. The latter gave a substantial improvement, from 6.6% to 9.7%, for the standard feature but negligible effect on the histogram. The entropy of the normalised histogram H_d was computed using $\sum_{i=0}^{d} H_i \log H_i$.

Finally the combined marginal and 3-D histograms were evaluated using a window size of 8, largest k of 3 and entropy directionality. In addition, a combined feature vector of the 3 standard features was evaluated. The m.a.p. results were: marginal histogram 12.0%, 3-D histogram 13.7% and standard 14.3%. All gave a significant improvement over the single features. Precision recall graphs for these combined features are shown in Fig. 5.

4.4 Gabor

Section 2.3 describes the generation of this feature. However, there still remain questions over how to apply it to a heterogeneous set of images in which texture patches of varying size, scale and orientation occur. The evaluation in [11] was applied to fixed tiles extracted from the Brodatz album. In [12] the feature was used successfully with aerial photographs split into a large number of fixed size tiles and then querying to find individual tiles. We decided to evaluate the feature in two configurations across a range of scale and orientation values. The first scaled the filter dictionary to the size of the image or image tile. This should scale the response so that the same image of different size gives a similar value. The second approach was to use a fixed size filter and apply this to a sliding window over the image.

Initial results showed that scaling the filter size gave much superior results to the sliding window approach. Tiling increased performance in a similar manner to the other features. The results shown in Table 5 are for 7×7 tiling. The best performance is obtained from just 2 scales and 4 orientations. This was unexpected as most literature recommends 4 scales and 6 orientations. Looking at the filtered images indicated that, as for Tamura, this may be due to noise at coarser scales.

Table 5: Gabor wavelets – mean average precision retrieval

	Orientation		
Scale	3	4	6
2	13.1%	14.0%	13.9%
3	11.0%	11.4%	11.3%
4	10.8%	11.4%	11.2%

4.5 Evaluation using TRECVID 2003 video data

A range of the best performing features were run on the TRECVID 2003 data and evaluated using the published relevance judgements. The queries were run singly and then combined with a colour histogram feature, HSV [16]. The results are shown in Table 6. For comparison some features used for previous evaluations [16] gave m.a.p. of: HSV 1.9%, convolution 2.2% and variance 1.7%; random retrieval would give 0.26%.

In this evaluation the texture features performed extremely well in comparison with previous benchmarks. Gabor gave the best results, 3.9% or 15 times better than random retrieval. Of the Tamura features the best performing was the combined standard features. The top 3 performing texture features combined gave a m.a.p. of 4.22%.

Combining with the HSV feature improved average retrieval performance in all cases, but at an individual query level the benefits were both positive and negative. It is interesting that using a simple combination of features gives varying degrees of improvement; being able to choose the optimum combination based on the query would be beneficial.

4.6 ImageCLEF evaluation

Following the original evaluation work [19] an opportunity arose to evaluate the texture features on an additional collection within the image track of the cross language evaluation forum (CLEF); full details of this work are presented in [15].

The nature of the image collection, described in Section 3.1, makes texture a key discriminator. Table 7 shows the retrieval performance for a range of the texture features. The Gabor feature on its own outperformed all runs submitted to the evaluation, achieving a m.a.p. of 35.3%. We actually submitted a run which combined five features. This gave a map of 34.5% which was third in the evaluation, less than 1% behind the leader.

This evaluation confirmed that the work evaluating and modifying the features is applicable over a range of image collections. The choice of parameters has produced robust features that perform well.

Table 6: TRECVID evaluation – mean average precision retrieval

Feature	Single	Combined with HSV
Gabor-2-4	3.93%	4.31%
Co-occurrence homogeneity	2.85%	3.03%
Tamura standard all	2.57%	3.43%
Tamura CND	1.65%	2.72%
Tamura coarseness-2	0.97%	2.49%

Table 7: ImageCLEF retrieval results

Feature	Mean average precision
Gabor-2-4	35.3%
Co-occurrence homogeneity	19.8%
Tamura standard all	20.7%
Tamura CND	18.4%
Tamura coarseness-2	14.5%

5 Conclusions and recommendations

We selected from the literature three different texture features, implemented, modified and evaluated them. Both the evaluation and implementation focused on query-by-example image retrieval rather than the usual classification task.

This led to some novel modifications to the Tamura features. We found that looking for large scale coarseness degraded performance, so we limited the range and used a logarithmic scale. An improvement in directionality performance over small window sizes was achieved by using an entropy measure rather than taking the second moments of the peaks. We also encoded the features in terms of joint histograms, the overall performance of these was similar to the standard features.

To improve the retrieval with Gabor we scaled the filter size to that of the image, rather than using a fixed size filter. Rather unintuitively we found that fewer scales gave higher retrieval rates. Our tests of co-occurrence matrices showed a solid performance, as was expected.

Although throughout our work we have considered how best to cope with varying image sizes, scales, formats and orientations, we have predominantly worked with images of roughly 10^5 pixels, i.e. 450 diagonal pixels. When working with larger images we suggest first down scaling them to roughly this size before extracting the texture features from this paper.

Our evaluation with TRECVID 2003 data showed that the top three texture features performed better than previously used colour features. Combination with a colour feature boosted retrieval performance in all cases. We carried out a further evaluation on a medical image collection, ImageCLEF. The Gabor feature performed particularly well. This further demonstrated the feature's robustness across differing collections and proved its effectiveness on a largely monochrome image library. Overall we have identified robust texture features for image retrieval.

6 Acknowledgment

This work was partially supported by the EPSRC, UK.

7 References

1 Tamura, H., Mori, S., and Yamawaki, T.: 'Textural features corresponding to visual perception', *IEEE Trans. Syst. Man Cybern.*, 1978, **8**, (6), pp. 460–472
2 Ohanian, P., and Dubes, R.: 'Performance evaluation for four classes of textural features', *Pattern Recognit.*, 1992, **25**, (8), pp. 819–833
3 Gotlieb, C., and Kreyszig, H.: 'Texture descriptors based on co-occurrence matrices', *Comput. Vis. Graph. Image Process.*, 1990, **51**, pp. 70–86
4 Jain, A., and Farrokhnia, F.: 'Unsupervised texture segmentation using gabor filters', *Pattern Recognit.*, 1991, **23**, (12), pp. 1167–1186
5 Randen, T., and Husøy, J.H.: 'Filtering for texture classification: A comparative study', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, **21**, (4), pp. 291–310
6 VisTex database. http://vismod.media.mit.edu/vismod/imagery/VisionTexture/1995
7 Brodatz, P.: 'Textures: A photographic album for artists & designers' (Dover, 1966)

8 Haralick, R.: 'Statistical and structural approaches to texture', *Proc. IEEE*, 1979, **67**, (5), pp. 786–804
9 Ortega, M., Rui, Y., Chakrabarti, K., Mehrotra, S., and Huang, T.: 'Supporting similarity queries in MARS'. Proc. 5th ACM Int. Conf. on Multimedia, Seattle, WA, USA, 1997, pp. 403–413
10 Turner, M.: 'Texture discrimination by Gabor functions', *Biol. Cybern.*, 1986, **55**, pp. 71–82
11 Manjunath, B., and Ma, W.: 'Texture features for browsing and retrieval of image data', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (8), pp. 837–842
12 Manjunath, B., Wu, P., Newsam, S., and Shin, H.: 'A texture descriptor for browsing and similarity retrieval', *Signal Process., Image Commun.*, 2000, **16**, (1), pp. 33–43
13 Smeaton, A., Kraaij, W., and Over, P.: 'TRECVID 2003 – An introduction', in TRECVID 2003 Proc.' 2003, pp. 1–10
14 Clough, P., Müller, H., and Sanderson, M.: 'The CLEF cross language image retrieval track (ImageCLEF) 2004', in Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004, Springer LNCS, 2005)
15 Howarth, P., Yavlinsky, A., Heesch, D., and Rüger, S.: 'Medical image retrieval using texture, locality and colour', in 'Fifth Workshop of the Cross-Language Evaluation Forum' (CLEF 2004, Springer LNCS, 2005)
16 Pickering, M., and Rüger, S.: 'Evaluation of key-frame based retrieval techniques for video', *Comput. Vis. Image Underst.*, 2003, **92**, (1), pp. 217–235
17 Voorhees, E.M., and Harman, D.: 'Overview of the eighth Text REtrieval Conference (TREC-8)', in Proc. TREC, 1999, A.17–A.18, pp. 1–33
18 Mitchell, T.: 'Machine learning' (McGraw Hill, 1997)
19 Howarth, P., and Rüger, S.: 'Evaluation of texture features for content-based image retrieval'. Int. Conf. on Image and Video Retrieval (CIVR, Dublin, Ireland, Springer LNCS 3115, Jul 2004), pp. 326–334