Calculating the Similarity of Textures using Wavelet Scale Relationships

Andrew Busch, Wageeh W. Boles & Sridha Sridharan
Research Centre in Speech, Audio, Image and Video Technology
School of Electrical and Electronic Systems Engineering
Queensland University of Technology
a.busch@qut.edu.au, w.boles@qut.edu.au, s.sridharan@qut.edu.au

Abstract

Texture is an important property of many types of images, and has been widely used in a number of applications to describe, classify and index images. In this paper we propose a novel technique for characterising texture by modelling the relationships between the high and low frequency bands of its wavelet decomposition, which have been experimentally shown to provide information which is not available when considering such bands independently. A similarity measure using the recently proposed earth mover's distance is formulated, and the results of using this measure for retrieval of textures from a database presented.

1 Introduction

The wavelet transform has emerged over the last two decades as a powerful new theoretical framework for the analysis and decomposition of signals and images at multiple resolutions [1]. One of the most common forms of the transform used for image analysis applications is the separable two dimensional wavelet transform, defined as

$$A_j = [H_x * [H_y * A_{j-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \quad (1)$$

$$D_{j1} = [G_x * [H_y * A_{j-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \quad (2)$$

$$D_{j2} = [H_x * [G_y * A_{j-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \quad (3)$$

$$D_{j3} = [G_x * [G_y * A_{j-1}]_{\downarrow 2,1}]_{\downarrow 1,2} \quad (4)$$

where G and H are the high and low-pass filters along the subscripted axis, $*$ is the convolution operator, j is the resolution level, and $\downarrow a, b$ represents downsampling along the x and y axes by factors of a and b respectively. The resulting images A_j and $D_{ji},\ i \in \{1, 2, 3\}$ are known as the approximation and detail coefficients respectively.

The coefficients of the wavelet transform have been shown to provide an excellent basis for identifying and segmenting textured images, and have been used in many applications to date [2, 3, 4]. The first and simplest of the features extracted from the wavelet coefficients were the so-called wavelet energy signatures, which were a representation of the energy contained within each band of the decomposition. Extending this, the mean deviation and other higher order moments have also been used for the purposes of texture classification and segmentation. Such features have been shown to provide good characterisation of textures in certain environments, and typically outperform single resolution techniques such as grey-level co-occurrence matrix features.

More recently, a number of new algorithms for extracting features from the coefficients of the wavelet transform have been proposed in the literature. Kim and Udpa propose a new non-separable set of wavelet filters for characterising texture which are shown to outperform the more commonly used separable DWT [5]. Tabesh uses the zero-crossings of a wavelet frame representation to extract texture features, and has shown experimentally that these features contain information not contained within the energy signatures [6]. By combining these two feature sets, overall accuracy is improved by up to 70%. Van de Wouwer *et. al.* have proposed a set of features based on second-order statistics of the wavelet coefficients, calculated using co-occurrence matrices [7]. Numerous methods of extracting texture features using the wavelet packet transform have been proposed [8, 9, 10].

In each of the feature extraction techniques listed above, the coefficients of each band are analysed separately, with the correlations between bands of the same and different resolution levels ignored, even though it is well-known that strong relationships between neighbouring bands exists. Portilla and Simoncelli have shown that without knowledge of these correlations accurate reconstruction of the texture is not possible, indicating that this information is significant for characterising the texture [11].

This paper proposes a novel method for indexing and retrieving textures based on *wavelet scale co-occurrence matrices*, which capture information about the relationships between each band of the transform and the low frequency approximation at the corresponding level. A theoretical description of scale co-occurrence matrices is first presented, and is then used to formulate a metric for measuring the similarity of two texture samples. The *earth mover's distance* (EMD), a recently proposed metric for measuring the distance between two distributions, is used for this purpose, as it has been shown to provide a more robust measure in many applications. Results from retrieval experiments conducted using this metric on a set of textures from the Brodatz album [12] are presented.

2 Limitations of Independent Wavelet Features

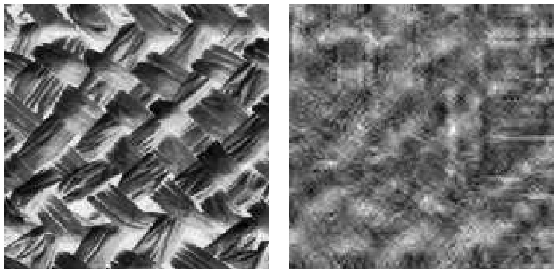
Initial experiments conducted by Julesz in the field of visual texture led him to conjecture that such images could be completely characterised by their second-order statistics [13]. Eventually, this was shown to be false, with many counter-examples presented showing visually distinct textures with identical second-order statistics. Recently, the main focus of much texture analysis research has centered around multi-scale filtering, with Gabor filters and the WT used to good effect. Common WT-based texture analysis techniques extract features from each band of the wavelet decomposition, measuring statistical information or modelling these coefficients via some parametric form. In this section, we show that it is possible for markedly different textures to have identical such statistics, indicating that they do not completely characterise texture. Examples of synthetic textures are provided to illustrate this point.

Wavelet energy signatures are one of most commonly used texture features in many applications, and can be calculated from the coefficients of the separable FWT by

$$E_{jl} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} D_{jl}(m,n)^2 \quad (5)$$

where D_{jl} are the wavelet detail coefficients are resolution level j, $l \in \{1,2,3\}$ indicates which of the detail images is being analysed, and M and N are the dimensions of the coefficient matrix. Such features have been shown to perform well in various texture analysis applications, however they are not sufficient to fully characterise a texture, and the performance of these features is generally not for any segmentation, classification or retrieval task.

In order to improve upon the performance of the first-order energy features, a method of modelling textures using both the first and second order statistics of wavelet detail coefficients has been proposed by Van de Wouwer *et. al.* [7]. Co-occurrence matrices are generated from each band of a redundant wavelet frame decomposition, and a set of eight standard co-occurrence features extracted to represent the



(a) (b)

Figure 1. Example showing the limitations of second-order statistics of wavelet coefficients as a texture descriptor. The natural texture (a) and synthesised texture (b) have identical first and second-order wavelet statistics of wavelet coefficients, yet are clearly distinguished by human observers.

second-order statistics of the texture. While this set of features improves performance compared to energy signatures, figure 1 shows that such a representation is still insufficient to completely characterise a texture. The synthetic image is generated having equal first and second order statistics of wavelet coefficients for the first 4 decomposition levels, and using the same low resolution approximation image. Clearly, these two textures are easily distinguishable by any observer, proving that the second order statistics of wavelet coefficients are insufficient to fully characterise texture. Without information to describe the relationships between each band of the WT, visual artifacts of the texture which contain elements at numerous scales are not adequately represented.

3 Wavelet Scale Co-occurrence Matrices

From the examples shown in the previous section, it is clear that features obtained independently from each band of the wavelet decomposition are not sufficient to fully characterise textured images. Portilla and Simoncelli have shown that relationships between direction and scale bands of the wavelet transform are in many cases critical for adequate reconstruction of a textured image [11, 14], and use the correlations between the coefficients of a complex steerable pyramid decomposition to characterise texture for the purpose of synthesis. These features quantitatively measure the correlation between each orientation band at a given resolution level j, as well as between each detail image and the detail images at neighbouring resolution resolution levels. These parameters are then used, along with autocorrelation features of both the magnitude and raw coefficient values and various statistics of the grey-level values, to generate synthetic texture images by restraint enforcement.

In order to capture the relationships between bands of the wavelet transform, we propose the *scale co-occurrence matrix* $S(k, l)$, which is defined as the probability of a detail coefficient $D(x, y)$ having a quantised value k while the approximation coefficient $A(x, y)$ at the same spatial position has a quantisedvalue of l. For an image of size $N \times M$, this can be expressed as

$$S_{ji}(k, l) = \frac{|\{(u, v) : q_1(D_{ji}(u, v)) = k, q_2(A_j(u, v)) = l\}|}{NM} \quad (6)$$

where A_j is the approximation image at resolution level j, D_{ji} are the three detail images, and $q_1(x)$ and $q_2(x)$ are the quantisation functions for the detail and approximation coefficients respectively. An overcomplete wavelet frame decomposition is used in our experiments in order to provide translation invariance and a higher robustness against nose, give higher spatial resolution, and avoid an overly sparse matrix at the lower resolutions.

The scale co-occurrence captures first order statistics of both the detail and approximation coefficients, and can be seen to fully encompass the wavelet mean deviation signatures. More importantly, by defining the relationships between the low and high frequency information at each scale, much information regarding structural components of the texture such as lines and edges can be extracted. The scale co-occurrence matrices overcome many of the limitations of features modelling the first and second order statistics of wavelet coefficients, as can be seen in figure 2, which shows two of the scale co-occurrence matrices extracted from the textures of figure 1, which have identical first and second order statistics. These matrices are clearly distinguishable, and provide the discrimination power lacking in other wavelet texture features. Previous work has shown that features extracted from the scale co-occurrence representation of texture are extremely powerful when used for classification and segmentation tasks [15].

4 Scale Co-occurrence Matrices for Similarity Measure

Using the scale co-occurrence matrices defined previously, it is possible to calculate a similarity measure between two textures

$$SM = \frac{1}{\sum_{j=1}^{J} \sum_{i=1}^{3} w_{ji} d(S1_{ji}, S2_{ji})} \quad (7)$$

where $d(x)$ is a distance metric used to determine the difference of the two distributions $S1$ and $S2$, and w_{ji} are weighting constants satisfying $\sum_j \sum_i w_{ji} = 1$. Such a similarity measure can then be used in image retrieval tasks, and also in classification and segmentation problems, where a candidate image is assigned to the class with the highest similarity measure. A number of choices for the distance metric $d(x)$ are available, ranging from the computationally inexpensive mean-squared error to more sophisticated techniques. We have chosen the earth mover's distance for this application as it has been experimentally shown to give a more accurate and stable measure of the difference between two scale co-occurrence representations that other metrics.

4.1 Earth Mover's Distance

The earth mover's distance (EMD) is a relatively new metric for representing the distance between two distributions in which the minimum amount of *work* required to transform one distribution to the other is calculated, given a set of ground distances between each point of the matrix [16]. Calculating this minimum amount can be viewed as a case of the *network transportation problem*, where one distribution is considered as a set of suppliers, and the other as a set of consumers. For the case of the scale co-occurrence matrix, each element of the matrix corresponds to a supplier or consumer.

Formally, the earth mover's distance can then be expressed as the minimum of the cost function of a set of weighted flows f_{ij} given by

$$EMD = \min \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} c_{ij} f_{ij} \quad (8)$$

where $\mathcal{I}$ and $\mathcal{J}$ are the sets of suppliers and consumers, and c_{ij} is the ground distance between bins i and j. To ensure a valid solution, the following restraints are also applied:

$$f_{ij} \geq 0 \quad (9)$$

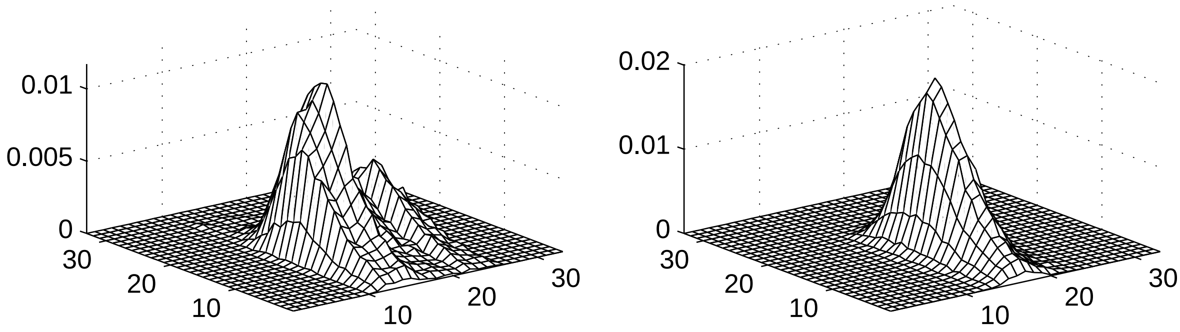
$$\sum_{i \in \mathcal{I}} f_{ij} = y_j \quad (10)$$

$$\sum_{j \in \mathcal{J}} f_{ij} = x_i \quad (11)$$

where x_i is the total supply the supplier i and y_j the total capacity of consumer j, which in our case are represented by the values of each element of the scale co-occurrence matrices.

A solution to the transportation problem of finding f_{ij} can be achieved using the simplex algorithm [17], an iterative method which will eventually converge to a local minimum. Russell has proposed an algorithm to determine a near-optimal starting point for this algorithm, which is used to ensure that the final value is close to the global minimum.

Using the earth mover's distance on the full scale co-occurrence matrices of two textures involves solving a transportation optimisation problem, which is of computational complexity $O(N^2)$, for more than 1000 suppliers and



(a) (b)

Figure 2. Examples of wavelet scale co-occurrence matrices for each of the textures of figure 1, showing considerable differences. (a) shows the horizontal and scale co-occurrence matrices respectively for the first level decomposition of the texture of figure 1(a), while (b) shows the same for figure 1(b).

consumers. Even for a relatively low number of iterations, this computational time can easily become excessive in all but the most trivial of applications. To improve this performance, it is necessary to significantly reduce the size of the flow optimsation problem without adversely affecting the accuracy of the distance metric. This can be accomplished by using *signatures* to represent the scale co-occurrence data rather than the traditional matrix form. Signatures, rather than representing fixed intervals, model a distribution using a set of clusters, and are defined as [16]

$$\{\vec{s}_i = (\vec{m}_i, v_i)\} \tag{12}$$

where $\vec{m}_i$ are the means of each cluster, and v_i the weightings. If sufficient clusters are used, it is possible to represent any distribution with arbitrary accuracy. It can also be shown that the histogram or matrix representation is actually a special case of a signature in which the clusters are set at fixed equidistant intervals in the underlying space. In fact, because of the possibility the each bin mean is not the mean of the distribution within it, a signature in this case will give a more accurate representation than the corresponding histogram form.

Calculating the signature of a distribution can be easily done by any one of a number of data clustering techniques. Using the k-means clustering algorithm has shown to produce acceptable results using approximately 50 clusters. More sophisticated techniques for approximating the modality of a distribution can be used for this purpose, but are beyond the scope of this paper.

The optimal flows f_{ij} and thus the final value of the EMD is highly dependent on the set of ground distances c_{ij}. Generally, these values are expressed as a function of (i, j), which in our case are the indices (k, l) of the scale co-occurrence matrices S_1 and S_2, and thus a two-dimensional vector. One commonly used metric for the ground distance between 2D points is the Euclidean distance

$$d(k_1, l_1, k_2, l_2) = \sqrt{(k_1 - k_2)^2 + (l_2 - l_2)^2} \tag{13}$$

where (k_1, l_1) and (k_2, l_2) are the indices of the scale co-occurrence matrices S_1 and S_2 respectively. Other metrics include the city block or Manhattan distance

$$d(k_1, l_1, k_2, l_2) = |k_1 - k_2| + |l_2 - l_2| \tag{14}$$

which is a summation of the distance of each dimension, and the maximum distance

$$d(k_1, l_1, k_2, l_2) = \max(|k_1 - k_2|, |l_2 - l_2|) \tag{15}$$

which considers only the greatest of the differences over all dimensions. Experimentally, a weighted Euclidean distance defined by

$$d(k_1, l_1, k_2, l_2) = \sqrt{a_k(k_1 - k_2)^2 + a_l(l_2 - l_2)^2} \tag{16}$$

where a_k and a_l are the weights for each dimension, has been found to give the most robust distance metric. Values of $a_k = 2$ and $a_l = 1$ are used in our experiments, meaning that differences in the detail coefficients are considered more important than a similar difference in the approximation coefficients.

4.2 Computational Considerations

The wavelet scale co-occurrence signature representation is quite small, less than 1kb for an image, making

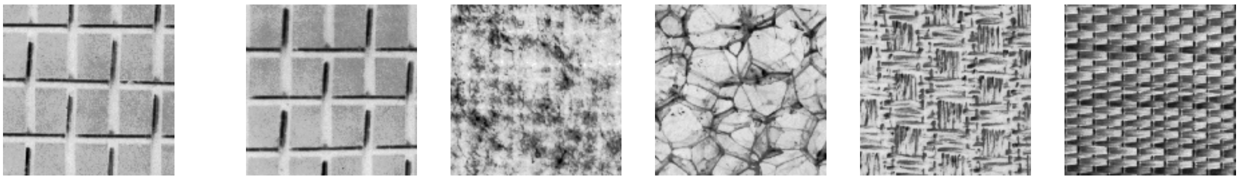


Figure 3. Results of a query using the scale co-occurrence similarity measure and the EMD. Query image (left) and top 5 matches, with distance measures of 42.7, 264, 274, 347.1 and 495.4 respectively.

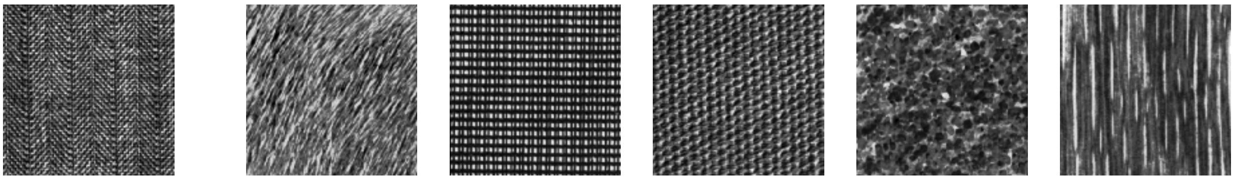


Figure 4. Retrieved images when using texture not present in the database. Query image (left) and top 5 matches, with distance measures of 196.8, 207.8, 226.7, 401.2 and 457.1 respectively.

it suitable for indexing large collections of textures. This size can be further reduced by traditional compression algorithms, which generally perform quite well given the relatively sparse nature of the data. Calculation of the similarity measure using the mean squared error is very fast, with more than 1000 comparisons performed per second on a 1700MHz workstation. Using the signatures rather than the full co-occurrence matrices provides a significant improvement to the efficiency of calculating this distance metric, reducing the computation time by many orders of magnitude, resulting in over 100 comparisons per second in most cases. On large databases or when computation speed is of critical importance, however, it may be necessary to further improve the computational efficiency of the search.

A suggested technique for pruning the search tree in large databases is to estimate a lower bound of the EMD, and use this estimate to remove unlikely match candidates. One such estimate of this lower bound is the distance between the centroids of the distributions, given using the notation of (8)-(11) as [16]

$$\min(EMD) = \left\| \sum_{i \in \mathcal{I}} x_i p_i - \sum_{j \in \mathcal{J}} y_j q_j \right\| \qquad (17)$$

where p_i and q_j are the coordinates of each cluster in the signatures $\mathcal{I}$ and $\mathcal{J}$ respectively. Such a lower bound is significantly faster to computer than the EMD, and by setting a suitable adaptive threshold a large proportion of the total candidates can be removed from the search without compromising the final result.

Another computational improvement can be realised by using a tree structured search algorithm, whereby the lowest resolution matrices or signatures are first compared, and processing continued for only those samples with the highest partial similarity measure. By combining these two methods of searching, the computation time for a typical search is reduced by approximately 95% with no noticeable affect on the quality of the retrieved matches.

4.3 Texture Retrieval Results

Using the distance metrics shown above, texture retrieval experiments were conducted using a small database of 50 images from the Brodatz album. The scale co-occurrence signature representations in of each of these images is extracted to 4 levels of wavelet decomposition, and used to create an index into the database. A test set of 200 images was the selected from the same 50 images, 4 from each class. In all cases, the training and test images were extracted from separate parts of the image such that no overlap between the two sets of possible. The top 5 matches for each of these test cases were then found in the database using the proposed similarity measure using the EMD metric. Overall, an image of the same class as the test case was returned as the most similar image in 97% with only 8 samples returning another class of image as the most likely. The results of a typical query for such an image is shown in figure 3.

In order to show the robustness of the proposed measure, a selection of textures which were not present in the database were searched for, with the results for a few such examples shown in figure 4. It can be seen from these examples that the textures retrieved from the database are visually similar to the candidate image, indicating that the scale co-occurrence matrices provide a good characterisation of the visual appearance of texture.

5 Conclusions and Future Work

In this paper we have proposed a novel representation of texture which models the relationships between bands of the wavelet transform decomposition. Such a representation has been empirically shown to provide better characterisation of many textures than statistics extracted independently from each band. Using the earth mover's distance, a similarity measure for comparing two textures based on this representation has been presented, with results from texture retrieval experiments showing excellent performance on a set of natural textures. Future work will aim to further improve the modelling of the scale co-occurrence distributions in a parametric form, and to study the types of textures for which the proposed representation provides the best representation. The proposed representation is also being investigated for the purposes of texture classification using both the distance metric and via a set of features extracted from the scale co-occurrence matrices.

References

[1] M Vetterli and J Kovacevic. *Wavelets and Subband Coding*. Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[2] T. Chang and C. C. Kuo. Texture segmentation with tree-structured wavelet transform. In *Proceedings of IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, volume 2, page 577, 1992.

[3] S. Fukuda and H. Hirosawa. A wavelet-based texture feature set applied to classification of multifrequency polarimetric sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5):2282–2286, 1999.

[4] H. Greenspan, S. Belongie, R. Goodman, and P. Perona. Rotation invariant texture recognition using a steerable pyramid. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 2, pages 162–167, Jerusalem, Israel, 1994.

[5] Sam-Deuk Kim and S. Udpa. Texture classification using rotated wavelet filters. *IEEE Transactions on Systems, Man and Cybernetics*, 30(6):847–852, 2000.

[6] A. Tabesh. Zero-crossing rates of wavelet frame representations for texture classification. *Electronics Letters*, 38(22):1340–1341, 2002.

[7] G. Van de Wouwer, P. Scheunders, and D. Van Dyck. Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing*, 8(4):592–598, 1999.

[8] T. Chang and C. C. J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing*, 2:429–441, 1993.

[9] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1191, 1993.

[10] C.-M. Pun and M.-C. Lee. Rotation-invariant texture classification using a two-stage wavelet packet feature approach. *IEE Proceedings - Vision, Image and Signal Processing*, 148(6):422–428, 2001.

[11] Javier Portilla and Eero. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.

[12] Phil Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications Inc., New York, 1966.

[13] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch. Inability of humans to discriminate between visual texture that agree in second-order statistics - revisited. *Perception*, 2:391–405, 1973.

[14] E. P. Simoncelli and J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proceedings of International Conference on Image Processing*, volume 1, pages 62–66, 1998.

[15] A. Busch and W. W. Boles. Texture classification using wavelet scale relationships. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 3484–3487, 2002.

[16] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Proceedings of IEEE International Conference on Computer Vision*, pages 59–66, Bombay, India, 1998.

[17] F. S. Hillier and G. J. Liberman. *Introduction to mathematical programming*. McGraw-Hill, 1990.