Image Retrieval Using One-Sided Linear Prediction Based Texture Modelling

Prabin K. Bora and Navajit Saikia

Department of Electronics and Communication Engineering, Indian Institute of Technology, Guwahati – 781039, India.

Email: prabin_kb@yahoo.com, navajit_saikia@yahoo.com

Abstract - This paper presents a new technique to index and retrieve texture image. The image is modeled with the two-dimensional one-sided linear prediction (OSP) model. The prediction coefficients, normalized mean-square prediction error and a set of seven invariant moments of the prediction error are used as features to index the image. The proposed method is highly successful in retrieving perceptually similar textures.

Index Terms – Content-based image retrieval (CBIR), texture modelling, one-sided linear prediction (OSP), linear-prediction error.

1. Introduction

Traditionally textual features such as file names, captions and key words have been used to annotate and retrieve images. But there are several drawbacks of these methods and the inadequacy of the textual descriptions have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as colour, texture and shape – a technology now generally referred to as *Content-Based Image Retrieval* (CBIR) [1], [2]. A CBIR technique automatically extracts primitive visual features from the image and retrieve images on the basis of these features. Human uses texture, color and shape to understand and recall the content of an image. Therefore features based on these attributes are used for image retrieval. CBIR is relatively a new research area and its main objective is to retrieve images from image database based on sample or query image and similarity measurement criteria.

Texture refers to the visual patterns that have properties of homogeneity, which do not result from the presence of any single color or intensity. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Natural textures are generally random, whereas artificial textures are often periodic. Texture is a very useful attribute for image matching and retrieval [3]. In absence of color and shape information, texture can act as a vital cue for image classification. Scenes containing pictures of wood, grass, etc. can be easily classified using texture rather than shape or color [2].

K. Deguchi and I. Morishita [4] used the one-sided linear prediction (OSP) model, popularly known as the autoregressive (AR) model, to derive texture descriptors in terms of the prediction coefficients. F. Liu and R.W. Picard [5] presented a new Wold-decomposition based texture model for image retrieval. The deterministic component in 2-D Wold decomposition was orthogonally decomposed into two fields: harmonic and evanescent. The "harmonicity test" on an image provided a measure of the confidence that the image can be characterized as highly structured (with dominant harmonic component) or relatively unstructured (with dominant indeterministic component). Based on this measure, either harmonic-peak feature extraction or Multi-Resolution Simultaneous Auto-Regressive (MRSAR) method or both were applied. The final Wold representation of the image contained the harmonic confidence measure and the corresponding harmonic-peak features and MRSAR features. B.S. Manjunath and W.Y. Ma [6] represented a texture image by Gabor features. Given an image, Gabor Wavelet transform coefficients were calculated. The mean and standard deviation of the magnitudes of the transform coefficients were used for indexing and retrieval. J.R. Smith and S-F. Chang [7] used Wavelet transform to characterize a texture and the statistics (mean and variance) extracted from the Wavelet sub-bands were used as features. In [8], R. Stoika, J. Zerubia and J.M. Francos decomposed a texture into purely indeterministic, harmonic and evanescent field components. These field parameters were used to index the image.

The objective of this paper is to improve upon the OSP-based texture descriptors [9], [10] by including additional features derived from the prediction errors. The prediction coefficients along with the additional features will be used to index texture images.

2. Image Representation

In order to retrieve images, we must be able to efficiently compare two images to determine if they have similar contents. An efficient matching scheme further depends upon the discriminatory information contained in the feature vectors of the images. Let $\mathbf{F}=\{f_1, f_2, \cdots\cdots\cdots, f_N\}$ be the sequence of all gray-level images in an image database defined over a finite domain $\mathbf{D}=\{[n_1, n_2] \| n_1 = 1,2,\cdots\cdots, N_1; n_2 = 1,2,\cdots\cdots, N_2\}$. Thus $f_i[n_1, n_2]$ denotes the gray-scale intensity value at a pixel $[n_1, n_2] \in \mathbf{D}$ for the image f_i.

Let z represent a mapping from the image set $\mathbf{F}$ onto a $n-$dimensional feature space $\mathbf{A}$ given by

$$z : \mathbf{F} \rightarrow \mathbf{A}$$

where

$$\mathbf{A}=\left\{\begin{bmatrix}\alpha_1^i \\ \alpha_2^i \\ \vdots \\ \vdots \\ \alpha_n^i\end{bmatrix}, \alpha_j^i \in \mathbf{R}; i=1,2,\cdots\cdots, N; j=1,2,\cdots\cdots, n\right\}$$

and $\mathbf{R}$ is the set of real numbers.

Under this mapping, each image f_i can be represented by a unique feature vector

$$\mathbf{A}^i = \begin{bmatrix}\alpha_1^i \\ \alpha_2^i \\ \vdots \\ \vdots \\ \alpha_n^i\end{bmatrix}, \alpha_j^i \in \mathbf{R}; j=1,2,\cdots\cdots, n . \tag{1}$$

3. Texture Feature Extraction

3.1 One-Sided Linear Prediction

The linear prediction model with its attractive features has been extended to two dimensions for modeling the stationary random field represented by an image and used in different applications like two-dimensional spectral estimation [11], [12] and texture coding [4], [10], [13]. The one-sided linear prediction (OSP) or causal linear prediction model is given by the non-symmetric half-plane (NSHP) model [13], [14], where the predicted value is a function only of the already scanned pixel values.

Consider a two-dimensional function $f[n_1, n_2]$ defined over a finite domain $\mathbf{D}$.

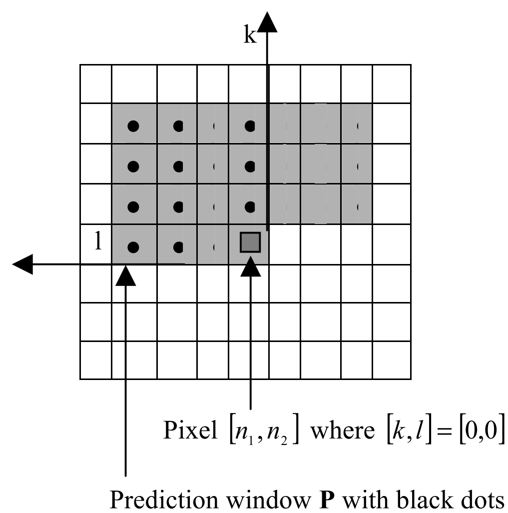

Figure-1: Support region P for two-dimensional one-sided linear prediction.

The predicted signal $f_p[n_1, n_2]$ at any location $[n_1, n_2] \in \mathbf{D}$ is given by

$$f_p[n_1, n_2] = \sum_{\substack{(k,l)\in \mathbf{P} \\ (k,l)\neq 0}} \sum a_{k,l} f[n_1 - k, n_2 - l] \tag{2}$$

where **P** is the NSHP prediction window defined at $[n_1, n_2]$ as shown in Figure-1.

The prediction coefficients $\{a_{k,l} \mid [k,l] \in \mathbf{P}, [k,l] \neq [0,0]\}$ are determined by minimizing the mean-square error of prediction. Given a prediction window and an index set $\{[k,l] \| [k,l] \in \mathbf{P}\}$, the optimal prediction coefficients are given by the following Wiener-Hopf equation [15].

$$\sum_{\substack{[k,l]\in\mathbf{P} \\ [k,l]\neq[\mathbf{0},\mathbf{0}]}} \sum a_{k,l} R_{ff}[k-m_1, l-m_2] = R_{ff}[m_1, m_2] \qquad (3)$$
$$\forall\, [m_1, m_2] \in \mathbf{P} \text{ and } [m_1, m_2] \neq [0,0]$$

where $\mathrm{R}_{ff}[k_1, k_2] = \mathbf{E} f[n_1+k_1, n_2+k_2] f[n_1, n_2]$ is the auto-correlation function at lag $[k_1, k_2]$ and $\mathbf{E}$ denotes the expectation operator.

It can be shown [16] that the maximum likelihood estimates of the prediction coefficients under a fairly general condition satisfy the Wiener-Hopf equation with the estimated auto-correlation values replacing corresponding true auto-correlation values.

The auto-correlation function at lag $[k_1, k_2]$ is estimated using the relation

$$\hat{\mathrm{R}}_{ff}[k_1, k_2] = \frac{1}{N_1 \times N_2} \sum_{n_1=1}^{N_1-k_1} \sum_{n_2=1}^{N_2-k_2} f[n_1, n_2] f[n_1-k_1, n_2-k_2] \qquad (4)$$

Hence the Equation (3) can be rewritten as

$$\sum_{\substack{[k,l]\in\mathbf{P} \\ [k,l]\neq[\mathbf{0},\mathbf{0}]}} \sum a_{k,l} \hat{R}_{ff}[k-m_1, l-m_2] = \hat{R}_{ff}[m_1, m_2] \qquad (5)$$
$$\forall\, [m_1, m_2] \in \mathbf{P} \text{ and } [m_1, m_2] \neq [0,0]$$

Let $N_{\mathbf{P}}$ be the cardinality of the set formed by the prediction coefficients $\{a_{k,l} \mid [k,l] \in \mathbf{P}, [k,l] \neq [0,0]\}$. These prediction coefficients can be used to form a $N_{\mathbf{P}}-$dimensional vector given as $\mathbf{b}^{\mathrm{OSP}} = \left[b_1^{\mathrm{OSP}}\ b_2^{\mathrm{OSP}} \cdots\cdots b_{N_{\mathbf{P}}}^{\mathrm{OSP}}\right]'$. Henceforward $\mathbf{b}^{\mathrm{OSP}}$ will be called the *prediction vector*. The number of terms used for prediction depends on the accuracy required for prediction. This prediction vector is the texture descriptor in [4].

3.2 Linear Prediction Error

For a two-dimensional function $f[n_1, n_2]$ defined over the finite domain **D**, the OSP $f_p[n_1, n_2]$ at any location $[n_1, n_2] \in \mathbf{D}$, is given by the Equation (3). The corresponding prediction error of linear prediction is obtained as [13]

$$\begin{aligned} e[n_1, n_2] &= f[n_1, n_2] - f_p[n_1, n_2] \\ &= f[n_1, n_2] - \sum_{\substack{(k,l)\in\mathbf{P} \\ (k,l)\neq(0,0)}} \sum a_{k,l} f[n_1-k, n_2-l] \end{aligned} \qquad (6)$$

The prediction coefficients $\{a_{k,l} | (k,l) \in \mathbf{P}\}$ along with the prediction-error vector constituted by the prediction errors $\{e[n_1, n_2] \| [n_1, n_2] \in \mathbf{D}\}$ completely describe the two-dimensional function $\{f[n_1, n_2] \| [n_1, n_2] \in \mathbf{D}\}$. This demands extraction of some meaningful descriptors from the prediction-error vector.

4. Feature Vector Design Principle

One of the difficult problems in the design of an efficient CBIR technique is to select a set of appropriate features to be extracted from the image [17]. The success of an image retrieval technique depends critically on this issue. There is no concrete theory to guide how to select the features of an image. The task of finding effective features for image database retrieval is still an open problem, but it is possible to state some desirable properties of features [2]:

- The features should be informative. The dimensionality of features should be as low as possible.
- The features should have high discriminating power.
- The features should have some meaning to human.

5. Selection of Features

The proposed method uses the following features to index a texture image:

- OSP coefficients of the texture image: Linear prediction coefficients characterize a random variable [15]. OSP is already used in texture image retrieval [4], [9].

- Scaled mean-square prediction error: The mean-square prediction error msq_e is the normalized square norm of the prediction-error vector. It is a measure of the magnitudes of the prediction-error vector components and hence chosen as a feature for texture image indexing. The mean square prediction error is given by [15]

$$msq_e = \mathbf{E}\, e^2[n_1, n_2] \tag{7}$$

An estimate $m\hat{s}q_e$ of msq_e is given by [13]

$$m\hat{s}q_e = \frac{1}{N_1 \times N_2} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} e^2[n_1, n_2] \tag{8}$$

- A set of seven invariant moments derived from the prediction-error vector components: For a two-dimensional function $f[n_1, n_2]$ defined over a finite domain $\mathbf{D}$, a set of seven invariant moments can be derived from the second and third order moments [9]. This moment set is used to represent the prediction-error vector and it carries information about the spatial variation of the individual prediction-error vector components. Simple moments could be used instead of the invariant moment set. The invariant moment set is chosen because of its extensive usage in CBIR based on shape information [9].

6. Similarity Measure

The difference between two images, f_1 and f_2, can be expressed as the distance 'd' between the respective feature vectors $\mathbf{A}^1$ and $\mathbf{A}^2$. The problem of retrieval can then be posed as follows:

Given a query image f_Q, retrieve an image f_T from the image database $\mathbf{F} = \{f_i \mid i = 1, 2, \cdots\cdots\cdots N\}$ such that

$$\begin{gathered}\mathrm{d}(\mathbf{A}^Q, \mathbf{A}^T) \leq d(\mathbf{A}^Q, \mathbf{A}^K),\\ \forall f_T \in \mathbf{F}, f_T \neq f_K\end{gathered} \tag{9}$$

The Euclidean distance is used to calculate the distance between the feature vectors of the query image f_Q and the target image f_T. Let $\mathbf{A}^Q = [\alpha_1^Q\ \alpha_2^Q \cdots \alpha_n^Q]'$ be the n-dimensional feature vector for the query image f_Q and $\mathbf{A}^T = [\alpha_1^T\ \alpha_2^T \cdots \alpha_n^T]'$ be the feature vector for the target image f_T. Then the Euclidean Distance between $\mathbf{A}^Q$ and $\mathbf{A}^T$ represents the similarity of the two images f_Q and f_T and is defined as [18]

$$\mathrm{d}(\mathbf{A}^Q, \mathbf{A}^T) = \left[\sum_{i=1}^{n} \left(\alpha_i^Q - \alpha_i^T\right)^2\right]^{\frac{1}{2}} \tag{10}$$

7. Experimental Results

A number of experiments were carried out to study the performance of the proposed image retrieval scheme. The image database was obtained from the Brodatz Album [19], [20] and consisted of 60 different texture classes. Four 200x200 non-overlapping sub-images were obtained from each texture class of size 640x640, thus creating a database of 240 images.

A query image in the following is any one of the 240 images in the database and each image is used once as a query image. A four-coefficient OSP model is chosen where four immediate neighbors of a pixel location of an image are considered. The Euclidean distance given by the Equation (10) is used as the similarity criterion and is applied to the feature vectors of the query image and each image in the database. For a query image f_Q, the top sixteen images are considered in hierarchical order as the retrieved set.

Quantitative performance was evaluated by calculating the *Percentage Retrieval Rate (%RR)* at a retrieved set size N_R [5], [6] defined as

$$RR = \frac{R_1}{R_2} \times 100 = \frac{1}{4} \times R_1 \times 100 \tag{11}$$

where R_2 is the total number of images in the texture class corresponding to the query image f_Q and R_1 represents the number of images of the same class appearing in the retrieved set. A 100% RR is reached by a search when all the R_2 matches are found in the retrieved set of images.

A number of experiments were performed with different combinations of the proposed features extracted from the prediction errors with the set of prediction coefficients. The following combinations were considered.

- Indexing with the OSP coefficients
- Indexing with the OSP coefficients and the normalized mean-square prediction error
- Indexing with the OSP coefficients and the set of seven prediction-error invariant moments
- Indexing with the OSP coefficients, the normalized mean-square prediction error and the set of seven prediction-error invariant moments

Table-1 shows the corresponding retrieval results.

METHOD ↓	% RETRIEVAL RATE			
	100	75	50	25
OSP	53	32	7	8
OSP COEFF. +ERROR MEAN SQUARE VALUE	68	13	10	9
OSP COEFF. +ERROR INVARIANT MOMENTS	65	09	13	13
OSP COEFF. +ERROR MEAN SQUARE VALUE +ERROR INVARIANT MOMENTS	82	12	3	3

Table-1: The Percentage Retrieval Rates for different CBIR techniques performed on a texture image database consisting of 240 images and 60 different texture classes. Each image was used in turn as a query image, and therefore 240 queries were performed. A retrieved set of 16 closely matched images was considered. The table should be interpreted as follow: for the OSP coefficient based method, 53% of the queries achieve RR of 100%, 32% achieve retrieval rate of 75%, etc.

From Table-1, it is seen that the performance of the OSP coefficient based retrieval method is very low. The retrieval rate substantially increases when the OSP coefficients and the normalized mean-square prediction error are used to index an image. As seen from the table, 68% queries achieve a RR of 100%, 13% queries achieve a RR of 75%, 10% queries achieve a RR of 50% and 9% queries only achieve a RR of 25%. The retrieval technique that uses the OSP coefficients and the set of seven invariant moments derived from the prediction-error vector components as indices of an image, retrieves 65% queries with 100% RR, 9% queries with 75% RR, 13% queries with 50% RR and 13% queries with 25% RR.

Much improved retrieval performance is obtained by using the OSP coefficients, normalized the mean-square prediction error and the set of seven invariant moments derived from the prediction-error vector components to form a twelve component feature vector representing an image. With this retrieval technique, 82% queries achieve a RR of 100%, 12% queries achieve a RR of 75%, 3% queries achieve a RR of 50% and 3% queries achieve a RR of 25%. This means, on the average 93.33% of the correct textures are in the retrieved set of 16 images. In [6], B.S. Manjunath, et.al., reported an average of 74.37% correct textures in the top 15 retrieved images with a different texture database.

A sample query image and the corresponding retrieved set is shown in Figure-2. The results show that the features extracted from the prediction errors have important roles in describing the textures. Further work is going on with higher order linear prediction models on larger texture databases.

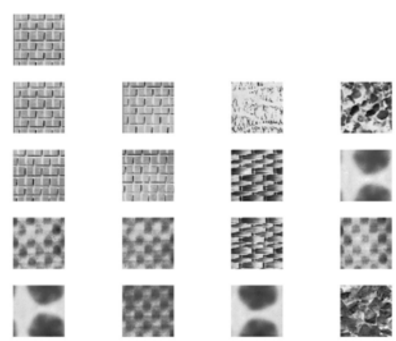

Figure-2: A sample retrieval result. The lone texture image in the top row is the query image and the following four columns show the retrieval set of sixteen texture images.

8. Conclusion

The existing OSP based texture image retrieval technique is improved by incorporating features extracted from the prediction errors. This new technique uses (i) the OSP coefficients (ii) the normalized mean-square prediction error and (iii) a set of seven invariant moments derived from the prediction-error vector components as the indices of an image. It is tested on a texture image database derived from the “Brodatz Texture Database” and it shows promising retrieval results.

References

[1] Y. Rui, T.S. Huang, S-F. Chang, “Image Retrieval: Current Techniques, Promising Directions and Open Issues, *Journal of Visual Communication and Image Representation*, Vol. 10, No. 1, pp. 39-62, 1999.

[2] Y. Rui, T.S. Huang and S-F. Chang, “Image Retrieval: Past, Present and Future”, *http://citeseer.nj.nec.com/192987html*.

[3] H. Xu and M. Liao, "Cluster Based Texture Matching for Image Retrieval", *Proc. IEEE Conf. Image Processing*, Vol. II, pp. 766-769, 1998.

[4] K. Deguchi and I. Morishita, “Texture Characterization and Texture-Based Partitioning Using Two-Dimensional Linear Estimation”, *IEEE Trans. Computers*, 27, pp. 739-749, 1978.

[5] F. Liu and R. W. Picard, “Periodicity, Directionality and Randomness: Wold Features for Image Modeling and Retrieval” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 7, pp. 722-733, July 1996.

[6] B.S. Manjunath and W.Y. Ma, “Texture Features for Browsing and Retrieval of Image Data”, *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp. 837-842, August 1996.

[7] J.R. Smith and S-F. Chang, “Transform Features for Texture Classification and Discrimination in Large Image Databases”, *Proc. IEEE Int. Conf. Image Processing*, Vol. 3, pp. 407-411, 1994.

[8] R. Stoika, J. Zerubia and J.M. Francos, "Image Retrieval and Indexing: A Hierarchical Approach in Computing the Distance Between Textured Images", *Proc. IEEE Conf. Image Processing*, Vol. II, pp. 794-798, 1998.

[9] M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Machine Vision*, PWS Publishing, 2nd Edition, 1999.

[10] D.G. Gost and L. Torres, “Efficient Coding of Homogeneous Textures Using Stochastic Vector Quantization and Linear Prediction”, *IEE Proc. Vision, Image, Signal Processing,* Vol. 146, No. 3, pp. 151-158, June, 1999.

[11] S. Raghunath and A.K. Jain, “Two-Dimensional Linear Prediction Models Part-I : Spectral Factorization and Realization”, *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 33, No. 1, pp. 280-299, Feb. 1985.

[12] W. Fang and A.E. Yagle, “Two-Dimensional Linear Prediction and Spectral Estimation in Polar Raster”, *IEEE Trans. Signal Processing*, Vol. 42, No. 3, pp. 628-641, Mar. 1994.

[13] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall of India Pvt. Ltd., 1997.

[14] Jin-Jen Hsue and Andrew E. Yagle, “Similarities and Differences Between One-Sided and Two-Sided Linear Prediction” *IEEE Trans. Signal Processing*, Vol. 43, No. 1, pp. 345-349, Jan. 1995.

[15] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1991.

[16] M.B. Priestly, *Spectral Analysis and Time Series*, Academic Press, 1981.

[17] Walter G. Kropatsch and Horst Bischof, *Digital Image Analysis: Selected Techniques and Applications*, Springer-Verlag Inc., 2001.

[16} X. Wan and C.C. Jay Kuo, “An Interactive Approach to Image Retrieval Using Multiple Seed Images” *Proc. IEEE Conf. Image Processing*, Vol. II, pp. 775-778, 1998.

[19] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, New York: Dover, 1966.

[20] Brodatz Textures, *http://www.ux.his.no/~tranden/brodatz.html.*