TWO-DIMENSIONAL LINEAR PREDICTIVE ANALYSIS OF ARBITRARILY-SHAPED REGIONS(*)

Petros A. Maragos, Russell M. Mersereau, and Ronald W. Schafer

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

This paper is concerned with the use of 2-D linear prediction for image segmentation. It begins with a brief summary of the mathematics involved in 2-D linear predictive analysis of arbitrarily-shaped regions. Then, it introduces a 2-D LPC distance measure based on the error residual of 2-D linear prediction. Finally, it describes how the above results can be applied to image segmentation using a simple cluster seeking algorithm. The results indicate that arbitrarily-shaped image regions can be well identified and clustered using as features their 2-D LPC parameters.

INTRODUCTION

One-dimensional linear prediction has been successfully used by Itakura [1] and others for extracting speech parameters and for deriving a LPC distance measure in speech classification and recognition. However, it appears that there has been no similar approach in pictorial feature extraction and in image segmentation by clustering [2,3]. Hence, it is the purpose of this paper to introduce the use of 2-D linear prediction and the resulting LPC distance for image segmentation. Because features in images typically are irregularly shaped, we begin by first formulating the problem of estimating the optimal LPC parameters for an arbitrarily-shaped image segment. Such a segment may be simply- or multiply-connected.

LINEAR PREDICTION

Let $x(m,n)$ represent a 2-D spatially-discrete array of intensity image samples. According to the autoregressive image model introduced in [4] for use in predictive image coding,

$$x(m,n)=\sum_k \sum_\ell a(k,\ell)x(m-k,n-\ell)+a_0+e(m,n) \quad (1)$$

We can view the 2-D prediction error sequence $e(m,n)$ together with the coefficients $\{a(k,\ell), a_0\}$ as an alternative exact characterization of the image signal $x(m,n)$. The bias coefficient a_0 accounts for the fact that the intensity image samples are explicitly biased since they are always nonnegative. The set $\{a(k,\ell),a_0\}$ can be seen as a set of features containing information about the specific image segment.

Suppose that $x(m,n)$ has support on the region Ω in the (m,n)-plane. Inside Ω we identify several homogeneous regions $D_\nu, \nu=1,\ldots,L$ as illustrated in Fig. 1. The general linear prediction problem is to find a set of optimal coefficients $\{a(k,\ell), a_0\}$ which minimize a mean-squared error

$$E=\sum_m \sum_n e^2(m,n) \quad (2)$$

where $e(m,n)$ is defined by Eq. (1). The array $a(k,\ell)$ is shown in Fig. 2 to possess a rectangular region of support which in the general case may include any other desired shape. The total number of prediction coefficients is $P=(Q_2-Q_1+1)x(R_2-R_1+1)-1$, and the number of our unknowns is $P+1$. We can distinguish two cases depending on whether the region is simply ($L=1$) or multiply-connected ($L > 1$):

a) One simply-connected region D_ν

We overcome the fact that D_ν has an irregular shape by considering a one-dimensional ordering of the greater rectangular region Ω of the (m,n)-plane; i.e., if Ω is an $N \times N$ region, then a rowwise ordering would be $O(m,n)=mN+n+1=j$. This ordering maps every pair (m,n), such as $0 \leq m, n \leq N-1$, onto an integer j belonging to the ordered set $Z_\Omega = \{1,2,3,\ldots,N^2\}$. If the information about the rowwise scanning of Ω is available, then $O(\cdot)$ is a reversible mapping of the region Ω onto the set Z_Ω, and we can recover (m,n) from j. Now

(*) This work was supported by the Joint Services Electronics Program under Contract #DAAG29-81-K-0024.

CH1841-6/83/0000-0104 $1.00 © 1983 IEEE

the regions D_ν are defined by the sets of integers $Z_\nu=\{z_\nu(j),\ j=1,2,\ldots,N^2\}$, $\nu=1,2,\ldots,L$ where

$$z_\nu(j)=\begin{cases}1, & (m,n)=O^{-1}(j)\ \epsilon\ D_\nu\\ 0, & \text{otherwise}\end{cases} \qquad (3)$$

If we think of Z_Q and Z_ν as N^2-dimensional vectors, the r-th nonzero element (j_r) of their component-wise multiplication will give us the r-th pair (m,n) from the M_ν pairs which make up the region D_ν. Thus, we can consider a one-dimensional indexing for the region D_ν :

$$r=IS_\nu(m,n)=IS_\nu[O^{-1}(j_r)],\ r=1,2,\ldots,M_\nu \qquad (4)$$

The initials IS mean "indexing" for the "signal" x(m,n). Now, the restriction of x(m,n) or its translate x(m-k,n-ℓ) to D_ν can be thought of as a M_ν-dimensional vector:

$$s_q=[s_q(r):\ s_q(r)=x(m-k,n-\ell),\ r=IS_\nu(m,n)]^T \qquad (5)$$

where $q=IP(k,\ell)$, and $IP(0,0)=0$ is understood. The indexings $IS(\cdot)$, $IP(\cdot)$ need not be the same. At this point we can express the 2-D correlation lags [4] as inner-products of known vectors:

$$R(k,\ell:i,j)=\langle s_{q_1},\ s_{q_2}\rangle \qquad (6)$$

where $q_1=IP(k,\ell)$, $q_2=IP(i,j)$. Similarly, the 2-D shift lags $S(k,\ell)$ [4] are equal to the sum of the components of the vector s_q, $q=IP(k,\ell)$.

The optimal coefficients which minimize the squared error E over the region D_ν are the solution to a system of normal equations:

$$\mathbf{C}\cdot\boldsymbol{\alpha}=\mathbf{r} \qquad (7)$$

where $\mathbf{C}$ is a (P+1)x(P+1) matrix whose entries are equal either to R(k, ℓ:i,j) or to S(k, ℓ).

$$\boldsymbol{\alpha}=[a(IP^{-1}(1)),\ldots,a(IP^{-1}(P)),\ a_0]^T \qquad (8)$$

$$\mathbf{r}=[\langle s_0,s_1\rangle,\ldots,\langle s_0,s_P\rangle,S(0,0)]^T \qquad (9)$$

All the above analysis refers to the covariance method [4] which minimizes E only over the region D_ν. Alternatively, we could modify our approach to include also the autocorrelation method, which assumes that x(m,n) is zero outside D_ν and minimizes E over the entire (m,n)-plane. In the covariance method the matrix $\mathbf{C}$ is symmetric and positive-definite, except for degenerative cases where it is positive-semidefinite. In the autocorrelation method the matrix $\mathbf{C}$ is a symmetric block-Toeplitz matrix and is always positive-definite, because then R(k, ℓ:i,j) equals R(|k-i|, |ℓ-j|) or R(|k-i|, -|ℓ-j|).

b) Multiple disjoint regions D_ν , $\nu=1,2,..,L$

The problem here is to obtain a set of common coefficients $\{a(k,\ell),\ a_0\}$ which minimize the error E over all the regions D_ν, $\nu=1,2,..,L$ simultaneously. It can be easily shown that the optimal coefficients are the solution to the following system

$$\Big[\sum_{\nu=1}^{L}\mathbf{C}_\nu\Big]\cdot\boldsymbol{\alpha}=\sum_{\nu=1}^{L}\mathbf{r}_\nu \qquad (10)$$

where $\mathbf{C}_\nu$, $\mathbf{r}_\nu$ are the correlation matrix and correlation vector of the region D_ν. The approach to obtain the correlation and shift lags is almost the same as in part (a). The only difference is that in order to find the indexing IS(m,n) for the ensemble of all the regions, one has to multiply the vector Z_Q by the sum of all the vectors Z_ν defined in (3). However, if one has already precomputed $\mathbf{C}_\nu$ and $\mathbf{r}_\nu$, it is easier simply to add them component-wise.

2-D LPC DISTANCE

Let us consider the augmented coefficient vector $\mathbf{b}=[1,\ -\boldsymbol{\alpha}]^T$ and the augmented correlation matrix

$$\mathbf{A}=\left[\begin{array}{c|c} R(0,0:0,0) & \mathbf{r}^T \\ \hline \mathbf{r} & \mathbf{C}\end{array}\right] \qquad (11)$$

where R(0,0:0,0) is obviously the energy of x(m,n) over the analysis region. The matrix $\mathbf{A}$ may refer to a simply-connected region or to disjoint regions. It can be proven that the squared error E can be expressed as the positive-(semi) definite quadratic form

$$E=\mathbf{b}^T\mathbf{A}\,\mathbf{b} \qquad (12)$$

Having reduced the problem to our one-dimensional one by using the one-dimensional indexing for the arrays a(k, ℓ) and x(m,n) over the regions of interest, we could use a 2-D LPC distance similar to the one used by Itakura [1] in the 1-D case. Thus, over an analysis region possessing augmented correlation matrix $\mathbf{A}$, we define the distance between two sets ($\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$) of coefficients as

$$d_{\mathbf{A}}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2)=|\log\ (\boldsymbol{\alpha}_1^T\mathbf{A}\boldsymbol{\alpha}_1/\boldsymbol{\alpha}_2^T\mathbf{A}\boldsymbol{\alpha}_2)| \qquad (13)$$

From (13) it is inferred that the above distance is a semi-metric, in the sense that it satisfies all the properties of a metric except one; i.e., $d_{\mathbf{A}}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2)=0$ does not imply that $\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2$. Also, it is clear that this distance

relates α_1 and α_2 only indirectly through the matrix **A**.

CLUSTERING ALGORITHM

Let us suppose now that we are given a 2-D image data array x(m,n) defined over a greater region Ω. The starting point of our algorithm for image segmentation by clustering is to divide the entire image into N_p smaller disjoint regions which consist of more or less homogeneous pictorial texture. This homogeneity will be hopefully reflected in a stationarity of the prediction coefficients over one region and a similarity between coefficients of disjoint regions with similar texture. Then, we obtain the augmented correlation matrices **A** for each image subregion. This way, each analysis region can be thought of as a pattern whose features are the entries of the matrix **A**. From Eqs. (7) and (11) it is clear that the optimal LPC coefficients can be obtained from the matrix **A**. Having obtained the LPC characterization of each region (pattern) one could use any clustering algorithm which employs a distance measure. We have used a variation of the so-called K-means [5] clustering algorithm modified to use the LPC distance measure. Our approach is summarized below:

Step-1: Select K initial cluster centers (regions) $\mathbf{c}_j$, j=1,2,...,K. The selection may be either arbitrary, or automatic using a max-min algorithm [5] which finds the K LPC patterns which are farthest apart.
Step-2: Allocate each of the N_p LPC patterns (characterized by their correlation matrices $\mathbf{A}_i$ and/or by their optimal coefficients α_i) to one of the K cluster centers according to:

α_i belongs to cluster j if

$$d_{\mathbf{A}_i}(\alpha_i, \mathbf{c}_j) < d_{\mathbf{A}_i}(\alpha_i, \mathbf{c}_m),\ m=1,2,\ldots,K$$

for all i, i=1,2,...,N_p. Ties are solved arbitrarily.
Step-3: Update the cluster centers: Having found from step-2 that each cluster consists of N_j LPC patterns, we find a set of prototype coefficients for each cluster (its cluster center) by using linear predictive analysis of multiple disjoint regions (10); i.e., for each cluster j we sum up the N_j correlation matrices and vectors and solve (10).
Step-4: The algorithm terminates whenever the cluster centers do not change from the previous iteration. Otherwise, go back to step-2 and iterate again.

The above clustering algorithm is an unsupervised pattern recognition scheme. We have found that it always converges in about 3-10 iterations. A good choice of the initial cluster centers may affect considerably the speed of convergence. The performance of this clustering algorithm obviously depends upon the method used to extract the LPC parameters. Thus, if we use the same number of prediction coefficients and the same prediction mask (Fig. 2), the covariance and the autocorrelation method yield similar results. However, the correlation matrix **A** in the autocorrelation method has much fewer different entries because of its block-Toeplitz property. For instance, if P=8, the matrix **A** has only 15 different entries compared to (P+1)(P+2)/2=55 for the covariance method. The size of the analysis regions does not play an important role as long as one stays well inside homogeneous regions. For regions, however, which contain boundaries between different textures, smaller analysis regions are required. The shape of the prediction mask (Fig. 2) was found to be of paramount importance. We tried 3 different shapes: 1) $Q_1=R_1=-1$, $Q_2=R_2=1$ gives an all-plane symmetric mask, 2) $Q_1=0$, $Q_2=2$, $R_1=-1$, $R_2=1$ gives a half-plane mask and, 3) $Q_1=R_1=0$, $Q_2=R_2=2$ gives a quarter plane mask. All these different masks involve the same number of prediction coefficients P=8. In terms of the average normalized mean-squared error E, the first mask is the best and the third is the worst. However, in terms of clustering performance the third mask is the best whereas the first is the worst. The reason for this might lie in the fact that the quarter-plane mask is the deepest in both directions.

EXPERIMENTAL RESULTS

Fig. 3 shows a 192 × 256 pixels black and white image which consists of 64 × 64 regions with different texture. We used 32 × 32 and 16 × 16 analysis regions with P=8 in our clustering algorithm, and the results were similar in both cases. The 8 prediction coefficients for each analysis region were obtained by using the autocorrelation method with the quarter-plane 3 × 3 mask. Fig. 4 shows the resulting clusters where K=3. The analysis regions were 32 × 32 pixels, and each region is illustrated by a number j, j=1,2,...,K, corresponding to the number of that cluster which this region was assigned to. Similarly, Fig. 5 shows results from clustering the same 32 × 32 regions in K=5 different clusters. From Fig. 4 and Fig. 5 we see that the clustering algorithm on this simple image yielded perfect results which agree with our own perceptual classification of the different textures in the image of Fig. 3. The above good results were obtained by using analysis regions which were embedded well inside homogeneous textures. If, however the analysis regions contain more than one different textures, then one should think of reducing the size of the analysis regions and/or employing other techniques to isolate the boundaries between different textures.

REFERENCES

[1] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. ASSP, Vol. ASSP-23, No. 1, pp. 67-72, Feb. 1975.

[2] A. Rosenfeld and L. S. Davis, "Image Segmentation and Image Models," Proc. IEEE, pp. 764-772, May 1979.

[3] G. B. Coleman and H. C. Andrews, "Image Segmentation by Clustering," Proc. IEEE, pp. 773-785, May 1979.

[4] P. A. Maragos, R. M. Mersereau and R. W. Schafer, "Some Experiments in ADPCM Coding of Images," Proc. ICASSP, pp. 1227-1230, May 1982, Paris, France.

[5] J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles" Addison-Wesley Publishing Company, Inc., 1974.

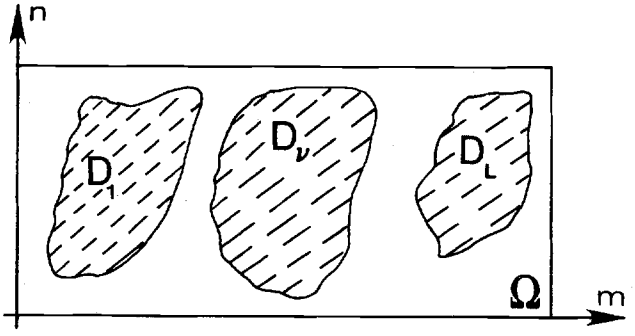

Fig. 1 — Multiple disjoint irregular regions of support

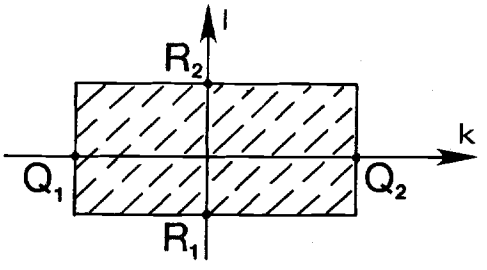

Fig. 2 — Region of support of the coefficient array a(k,l), a(0,0)=0

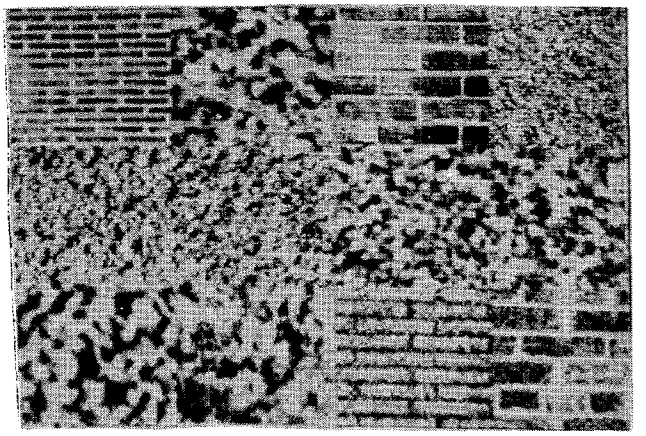

Fig. 3 — A 192 x 256 pixels texture image

1	1	2	2	1	1	3	3
1	1	2	2	1	1	3	3
3	3	3	3	2	2	2	2
3	3	3	3	2	2	2	2
2	2	2	2	1	1	1	1
2	2	2	2	1	1	1	1

Fig. 4 — Clustering results with K=3. Each number represents the cluster containing this 32x32 region.

1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
4	4	4	4	2	2	2	2
4	4	4	4	2	2	2	2
2	2	2	2	5	5	3	3
2	2	2	2	5	5	3	3

Fig. 5 — Clustering results with K=5