TABLE III
ESTIMATED VALUES OF (β_1, β_2) FOR REALIZATIONS OF 1ST ORDER GMRFS DEFINED ON SQUARE LATTICES BY MAXIMIZING (1) THE TRUE LIKELIHOOD, (2) THE APPROXIMATE LIKELIHOOD, AND (3) THE PSEUDO-LIKELIHOOD

Pixel number	$\beta_1 = 0.20, \quad \beta_2 = 0.20$			$\beta_1 = 0.10, \quad \beta_2 = 0.15$		
	1	2	3	1	2	3
8×8	0.244053	0.243762	0.306052	0.191802	0.191399	0.199992
	0.202476	0.203267	0.154937	0.167056	0.166789	0.147094
16×16	0.221357	0.221372	0.245706	0.093866	0.093970	0.086112
	0.178942	0.179546	0.170371	0.188620	0.188378	0.208411
32×32	0.201495	0.201684	0.212936	0.065614	0.065632	0.072296
	0.186200	0.186556	0.183712	0.134186	0.134147	0.148668
64×64	0.185070	0.185481	0.188647	0.099275	0.099298	0.098364
	0.210321	0.210412	0.208736	0.133404	0.133393	0.131875

TABLE IV
SAMPLE MEAN AND STANDARD DEVIATION OF THE ESTIMATES OF (β_1, β_2) USING 1000 32×32 SUBIMAGES EXTRACTED FROM REALIZATIONS OF GMRFS GENERATED WITH $\beta_1 = \beta_2 = 0.2$ BY MAXIMIZING (1) THE TRUE LIKELIHOOD, (2) THE APPROXIMATE LIKELIHOOD, AND (3) THE PSEUDO-LIKELIHOOD

	Sample mean			Sample standard deviation		
	1	2	3	1	2	3
$\beta_1 = 0.2$	0.199446	0.1997910	0.200814	0.0231955	0.02292836	0.0255177
$\beta_2 = 0.2$	0.200204	0.2005310	0.201290	0.0233693	0.02309547	0.0260651

values using the following experiment. For a given set of parameters (β_1, β_2), we repeat the above parameter estimation procedure for 1000 GMRF realizations, and then compute the sample mean and sample variance for the resulting estimates of the β's. For example, Table IV shows the sample means and standard deviations of the resulting estimates of (β_1, β_2) using 32×32 subimages for GMRF's generated with $\beta_1 = \beta_2 = 0.2$. Note that the variance for the pseudo-likelihood method is over 20% larger than the true ML method and the approximate ML method, while all three result in sample means very close to the true (β_1, β_2).

When the true β's approach the boundary of their valid domain, in the asymptotic case $|\beta_1| + |\beta_2| \to 0.5$, both the approximate ML and pseudo-likelihood methods quickly become ineffective. For example, for $(\beta_1, \beta_2) = (0.35, 0.15)$ both methods result in a nonpositive definite matrix $\hat{B}$ with β's estimated from a 64×64 subimage.

REFERENCES

[1] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc., Series B*, vol. 36, pp. 192–236, 1974.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, Nov. 1984.

[3] R. Chellappa and R. L. Kashyap, "Texture synthesis using 2-D noncausal autoregressive models," *IEEE Trans. on Acoust. Speech Signal Processing*, vol. 33, pp. 194–203, Feb. 1985.

[4] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images and Gibbs random fields," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 9, pp. 39–55, Jan. 1987.

[5] J. F. Silverman and D. B. Cooper, "Bayesian clustering for unsupervised estimation of surface and texture models," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 10, pp. 482–495, July 1988.

[6] K. Ord, "Estimation methods for models of spatial interaction," *J. Amer. Stat. Soc.*, vol. 70, pp. 120–126, Mar 1975.

[7] J. Besag and P. A. P. Moran, "On the estimation and testing of spatial interaction in Gaussian lattice processes," *Biometrika*, vol. 62, pp. 555–562, 1975.

[8] J. Besag, "Statistical analysis of nonlattice data," *The Statistician*, vol. 24, pp. 179–195, 1975.

[9] Y. Ogata and M. Tanemura, "Likelihood analysis of spatial point patterns," *J. Roy. Stat. Soc., Series B*, vol. 46, pp. 496–518, 1984.

[10] G. Potamianos and J. Goutsias, "Stochastic simulation techniques for partition function approximation of Gibbs random field images." Rep. JHU/ECE 91-02, Johns Hopkins University, Dept. of Electrical and Computer Engineering, 1991.

[11] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, pp. 434–449, 1954.

[12] X. Guyon, "Parameter estimation for a stationary process on a d-dimensional lattice," *Biometrika*, vol. 69, pp. 95–105, 1982.

[13] K. V. Mardia and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, vol. 71, pp. 135–146, 1984.

[14] B. D. Ripley, *Statistical Inference for Spatial Processes*. Cambridge University Press, 1988.

[15] J. Besag, "Efficiency of pseudolikelihood estimation for simple Gaussian fields," *Biometrika*, vol. 64, pp. 616–618, 1977.

Causal and Semicausal AR Image Model Identification Using the EM Algorithm

Yücel Yemez, Emin Anarim, and Yorgo Istefanopulos

Abstract—**In this study, we extend the method presented in a recent paper [1], which considers the problem of the semicausal autoregressive (AR) parameter identification for images degraded by observation noise. We propose a new approach to identify both the causal and semicausal AR parameters without *a priori* knowledge of the observation noise power. We decompose the image into 1-D independent complex scalar subsystems resulting from the vector state-space model by using the unitary discrete Fourier transform (DFT). Then, by applying the expectation-maximization (EM) algorithm to each subsystem, we identify the AR parameters of the transformed image. The AR parameters of the original image are then identified by using the least squares (LS) method. The restored image is obtained as a byproduct of the EM algorithm.**

Manuscript received July 2, 1992; revised February 2, 1993. The associate editor coordinating the review of this paper and approving it for publication was Dr. M. Ibrahim Sezan. This work was presented in part at ICCTD-93.

The authors are with the Department of Electrical and Electronics Engineering, Boğaziçi University, 80815 Bebek, Istanbul, Turkey.

IEEE Log Number 9210838.

I. INTRODUCTION

Processing of images generally requires some statistical knowledge about the image. The concern of image identification is the estimation of this statistical knowledge prior to the image processing application. The type of the statistical knowledge is determined by assuming a suitable image model. The most common approach is to model the image by using a 2-D autoregressive (AR) model. Then, the problem becomes the identification of the AR parameters. Usually, due to the imperfections in the electronic, photographic, or transmission medium, the image model identification has to be performed in the presence of observation noise. The classical approaches in the literature assume that the power of the observation noise is known *a priori*, and only with this assumption does identification become possible. In the recent studies of Katayama [1] and Lagendijk [2], the necessity for the *a priori* knowledge of the observation noise has been overcome by making use of the EM algorithm.

In [2], the aim is actually to restore the images against blur and noise. In this respect, the identification of 2-D causal AR model parameters is required. By making use of the EM algorithm, the method can identify only causal AR models without *a priori* knowledge of the observation noise; however, the computational load is excessive due to the 2-D structure of images.

In [1], Katayama presents a computationally efficient method to identify the semicausal AR model parameters in the presence of noise, again by using the EM algorithm. The idea in his method is to reduce the 2-D problem to a 1-D problem by using the unitary discrete sine transform (DST), as proposed originally by Jain [3].

In this paper, we present a new approach, based on [1], that can identify both causal (quarter-plane or nonsymmetric half-plane) and semicausal AR model parameters in the presence of noise. Our method makes use of the EM algorithm and reduces the dimension by using the DFT, which makes the approach flexible for different AR models.

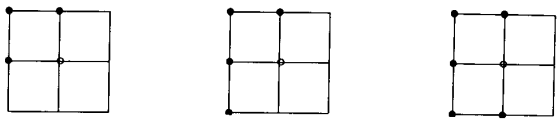
The paper is organized as follows. In Section II, we describe the image and the observation model to identify and derive a vector state-space model from the causal and semicausal models. We solve the boundary problem by assuming a circulant image model, and consequently, the image model coefficient matrices become circulant. In this respect, in Section III, we briefly review the DFT diagonalization results for circulant matrices. In Section IV, using these results, we decompose the image into 1-D uncorrelated subsystems by using DFT. In Section V, we apply the EM algorithm to each subsystem separately to restore and identify the AR parameters in the transform domain. The AR parameters of the original image are then identified by using the least squares (LS) method. The simulation results of our approach are given in Section VI and concluding remarks are considered in Section VII.

II. IMAGE MODELING

We consider a monochromatic image of size $N \times N$ and denote the gray levels of the original image and the observed image by $x(n,m)$ and $y(n,m)$, respectively, where n is the vertical and m is the horizontal position variable. To model the image, we use the following 2-D AR model driven by a zero mean random model noise $w(n,m)$:

$$x(n,m) = \sum_{k,l \in S} a_{kl} x(n-k, m-l) + w(n,m) \tag{1}$$

where $\{a_{kl}\}$ denote the image model coefficients which are determined by minimizing the variance of the noise $\sigma_w^2 = E[w^2(n,m)]$. In (1), S stands for the image model support. Some common model supports for first-order QP (quarter-plane) causal, NSHP (nonsymmetric half-plane) causal, and semicausal models are shown in Fig. 1.



(a) (b) (c)

Fig. 1. Examples of some commonly used model supports for various first-order AR models. (a) QP causal. (b) NSHP causal. (c) Semicausal.

We assume that the observed image is degraded by zero mean additive Gaussian white noise. Therefore, the observation can be modeled as

$$y(n,m) = x(n,m) + v(n,m) \tag{2}$$

where $v(n,m)$ denotes the observation noise with unknown variance σ_v^2.

Alternatively, (1) and (2) can be written as

$$\boldsymbol{x}(m) = \sum_{j=1}^{p} \boldsymbol{A}_0^{-1} \boldsymbol{A}_j \boldsymbol{x}(m-j) + \boldsymbol{A}_0^{-1} \boldsymbol{w}(m) \tag{3}$$

$$\boldsymbol{y}(m) = \boldsymbol{x}(m) + \boldsymbol{v}(m) \tag{4}$$

where $\boldsymbol{x}(m)$, $\boldsymbol{w}(m)$, $\boldsymbol{y}(m)$, and $\boldsymbol{v}(m)$ are the mth column vectors of the $N \times N$ original image, model noise, observed image, and observation noise, respectively. In (3), $\boldsymbol{A}_j, j = 0, 1, \cdots, p$, are the $N \times N$ model coefficient matrices determined by the 2-D AR model coefficients $\{a_{kl}\}$. For all types of AR models, the coefficient matrices become circulant if we solve the boundary problem by assuming that the image is circulant [4]. Below, we list the first rows of the circulant coefficient matrices for QP, NSHP, and semicausal AR models:

1. QP causal AR model:
   $\boldsymbol{A}_0 : [1 \quad 0 \cdots 0 \; -a_{p0} \cdots -a_{10}]$
   $\boldsymbol{A}_j : [a_{0j} \quad 0 \cdots 0 \quad a_{pj} \cdots a_{1j}]$
2. NSHP causal AR model:
   $\boldsymbol{A}_0 : [1 \quad 0 \cdots 0 \; -a_{p0} \cdots -a_{10}]$
   $\boldsymbol{A}_j : [a_{0j} \quad a_{-1,j} \cdots a_{-p,j} \quad 0 \cdots 0 \quad a_{pj} \cdots a_{1j}]$
3. Symmetric[1] semicausal AR model:
   $\boldsymbol{A}_0 : [1 - a_{10} \cdots -a_{p0} \quad 0 \cdots 0 - a_{p0} \cdots -a_{10}]$
   $\boldsymbol{A}_j : [a_{0j} \quad a_{1,j} \cdots a_{p,j} \quad 0 \cdots 0 \quad a_{pj} \cdots a_{1j}]$

For causal models, if we determine the AR model coefficients $\{a_{kl}\}$ by minimizing the variance of the model noise, then the resulting model noise is necessarily white [5]. Therefore, we can write the covariance matrix of the model noise $\boldsymbol{w}(m)$ as

$$E[\boldsymbol{w}(m)\boldsymbol{w}^T(k)] = \sigma_w^2 \boldsymbol{I} \delta_{mk} \tag{5}$$

where δ_{mk} is the Kronecker delta. For semicausal models, however, minimizing the variance of the model noise does not assure that the resulting model noise is white [5]. It is possible to assume the covariance matrix of the model noise in this case as

$$E[\boldsymbol{w}(m)\boldsymbol{w}^T(k)] = \sigma_w^2 \boldsymbol{A}_0 \delta_{mk}. \tag{6}$$

III. DIAGONALIZATION OF THE COEFFICIENT MATRICES

We recall that the coefficient matrices in (3) are circulant matrices. Any circulant matrix can be diagonalized by using unitary DFT so that [6]

$$\boldsymbol{F}\boldsymbol{A}_j\boldsymbol{F}^* = \text{diag}(f_j(W_N^0), f_j(W_N^1), \cdots, f_j(W_N^{N-1})) \tag{7}$$

[1]Symmetric modeling implies $a_{kl} = a_{-k,l}$.

where $\boldsymbol{A}_j$ denotes the jth coefficient matrix and $\boldsymbol{F}$ is the defining matrix of DFT, which is given by

$$\boldsymbol{F} = \left\{ \frac{1}{\sqrt{N}} W_N^{kn} \right\} \qquad 0 \le k, n \le N-1 \tag{8}$$

where by definition, $W_N = \exp \frac{-i2\pi}{N}$. In (7), the diagonal elements are the eigenvalues of the coefficient matrix $\boldsymbol{A}_j$ and are given by the DFT of its first column [6]:

$$f_j(W_N^k) = \sum_{l=0}^{N-1} \alpha_j(l) W_N^{kl} \qquad 0 \le k \le N-1 \tag{9}$$

where $\{\alpha_j(l)\}$ denote the elements of the first column of the coefficient matrix $\boldsymbol{A}_j$. It is important to note that although DST almost diagonalizes the coefficient matrices of the symmetric semicausal model [3], the use of DFT does not require any approximation on the diagonalization of the coefficient matrices of both causal and semicausal models.

IV. Reduction of Dimensionality

In this section, using the diagonalization results of Section III, we aim to decompose the 2-D image into 1-D independent scalar subsystems.

We define the DFT's of the vectors in (3) and (4) as

$$\begin{aligned} \boldsymbol{\theta}(m) &= \boldsymbol{F}\boldsymbol{x}(m) \qquad & \boldsymbol{\xi}(m) &= \boldsymbol{F}\boldsymbol{A}_0^{-1}\boldsymbol{w}(m) \\ \boldsymbol{\eta}(m) &= \boldsymbol{F}\boldsymbol{y}(m) \qquad & \boldsymbol{\zeta}(m) &= \boldsymbol{F}\boldsymbol{v}(m) \end{aligned} \tag{10}$$

and taking the DFT of (3) and (4) yields

$$\theta_k(m) = \sum_{j=1}^{p} (a_j)_k \theta_k(m-j) + \xi_k(m) \tag{11}$$

$$\eta_k(m) = \theta_k(m) + \zeta_k(m) \tag{12}$$

for $k = 0, \cdots, N-1$, where the subscript k denotes the kth elements of the vectors in (10) and

$$(a_j)_k = \frac{f_j(W_N^k)}{f_0(W_N^k)}. \tag{13}$$

We note that the scalars in (11) and (12) are all complex-valued.

Equations (11) and (12) together with (5) and (6) imply that the rows of the transformed image can be processed independently, and the covariance matrices of the transformed complex scalars $\xi(m)$ and $\zeta(m)$ are given by [4]

$$\mathrm{E}[\xi_k(m)\xi_l^*(t)] = \sigma_\xi^2(k)\delta_{kl}\delta_{mt} \tag{14}$$

$$\mathrm{E}[\zeta_k(m)\zeta_l^*(t)] = \sigma_\zeta^2(k)\delta_{kl}\delta_{mt} \tag{15}$$

$$\mathrm{E}[\xi_k(m)\zeta_l^*(t)] = 0 \tag{16}$$

where, for causal models,

$$\sigma_\xi^2(k) = \frac{\sigma_w^2}{|f_0(W_N^k)|^2} \tag{17}$$

while for semicausal modeling,

$$\sigma_\xi^2(k) = \frac{\sigma_w^2}{|f_0(W_N^k)|}. \tag{18}$$

Further reduction in computational complexity can be achieved after a closer examination (see Appendix), by introducing the following approximation which implies that the real and the imaginary parts of the transformed image are nearly uncorrelated:

$$a_{i,j} \simeq 0 \qquad j = 1, \cdots, p \tag{19}$$

Therefore, dropping the subscript k in (11) and (12) and considering the imaginary and real parts of the complex scalars separately for each row of the DFT image (11) and (12) can be written as

$$\begin{aligned} \theta_r(m) &= \sum_{j=1}^{p} a_j \theta_r(m-j) + \xi_r(m) \\ \eta_r(m) &= \theta_r(m) + \zeta_r(m) \end{aligned} \tag{20}$$

$$\begin{aligned} \theta_i(m) &= \sum_{j=1}^{p} a_j \theta_i(m-j) + \xi_i(m) \\ \eta_i(m) &= \theta_i(m) + \zeta_i(m) \end{aligned} \tag{21}$$

where a_j's are real scalars.

The real and the imaginary parts of DFT are actually two separate linear transformations. Therefore, $\xi_r(m)$, $\xi_i(m)$, $\zeta_r(m)$, and $\zeta_i(m)$ are zero mean Gaussian white random processes with variances $\sigma_{\xi r}^2$, $\sigma_{\xi i}^2$, $\sigma_{\zeta r}^2$, and $\sigma_{\zeta i}^2$, respectively. It is also possible to write, dropping the subscript k, the variances of the transformed noise vectors defined in (14) and (15) as

$$\sigma_\xi^2 = \sigma_{\xi r}^2 + \sigma_{\xi i}^2 \tag{22}$$

$$\sigma_\zeta^2 = \sigma_{\zeta r}^2 + \sigma_{\zeta i}^2. \tag{23}$$

We note that (20) and (21) represent 1-D causal AR model equations for the real and the imaginary parts of each row of the transformed image separately and are suitable forms to apply 1-D Kalman filtering. Now that we have decomposed the $N \times N$ image to N independent 1-D complex scalar subsystems, we can apply the EM algorithm separately to each subsystem to restore the image and identify the AR parameters.

V. EM Algorithm

It has been shown that EM-like approaches lead to computationally efficient estimation algorithms in various signal processing applications [7]–[9].

Let Θ denote the unknown parameter vector of the system, which is given by

$$\Theta = [\{a_j\}, \sigma_{\xi r}^2, \sigma_{\xi i}^2, \sigma_{\zeta r}^2, \sigma_{\zeta i}^2]. \tag{24}$$

The problem is to estimate Θ based on the transformed observed image $\boldsymbol{\eta}(m)$ by using the EM algorithm. The EM algorithm has two steps, which are referred to as expectation step (E-step) and maximization-step (M-step). E-step finds the conditional expectation of the complete log-likelihood function $\mathcal{L}_c(\Theta) = p(\boldsymbol{Z}^N, \boldsymbol{Y}^N|\Theta)$ where $\boldsymbol{Z}^N = [z(-1)z(0)\cdots z(N-1)]$, $\boldsymbol{Y}^N = [\boldsymbol{\eta}(-1)\boldsymbol{\eta}(0)\cdots\boldsymbol{\eta}(N-1)]$ and $z(m) = [\theta(m)\theta(m-1)\cdots\theta(m-p+1)]^T$. Since $\xi(m)$ and $\zeta(m)$ are assumed to be Gaussian white, it is possible to write the complete log-likelihood function denoted by $\ell_c(\Theta) = \log p(\boldsymbol{Z}^N, \boldsymbol{Y}^N|\Theta)$ as

$$\ell_c(\Theta) = \ell_{cr}(\Theta) + \ell_{ci}(\Theta) \tag{25}$$

where $\ell_{cr}(\Theta)$ and $\ell_{ci}(\Theta)$ denote the complete log-likelihood function for the real and imaginary parts of the transformed image, respectively.

The conditional expectation computed in E-step is, then, maximized in M-step with respect to the parameter vector Θ so that Θ is updated. The algorithm alternates between the two steps iteratively until Θ converges to its optimum value. The convergence of the EM algorithm is assured under the conditions that the underlying AR model is causal and the variances of the model noise and the observation noise are positive [10]. These conditions are already satisfied in our case, recalling that (20) and (21) are 1-D causal AR model equations resulting from either causal or semicausal 2-D model.

Fig. 2. Original face image.

Maximization of the conditional expectation $-E[\ell_c(\Theta)|\boldsymbol{Y}^N, \Theta]$ with respect to the unknown parameter $\boldsymbol{f} = [a_1 \cdots a_p]$ yields the following update equation in M-step:

$$\hat{\boldsymbol{f}} = (\boldsymbol{S}_{01,r}^{(1)} + \boldsymbol{S}_{01,i}^{(1)})^T (\boldsymbol{S}_{11,r} + \boldsymbol{S}_{11,i})^{-1} \tag{26}$$

where $\boldsymbol{S}_{01}^{(1)}$ denotes the first column of $\boldsymbol{S}_{01}$. $\boldsymbol{S}_{01}$ and $\boldsymbol{S}_{11}$ are given by

$$\begin{aligned} \boldsymbol{S}_{01} &= \sum_{m=1}^{N} [\boldsymbol{P}(m, m-1|N) + \hat{z}(m|N)\hat{z}^T(m-1|N)] \\ \boldsymbol{S}_{11} &= \sum_{m=1}^{N} [\boldsymbol{P}(m-1|N) + \hat{z}(m-1|N)\hat{z}^T(m-1|N)] \end{aligned} \tag{27}$$

and to be computed separately for real and imaginary parts. The conditional mean estimate $\hat{z}(m|N)$ and the covariance matrices $\boldsymbol{P}(m|N)$, $\boldsymbol{P}(m, m-1|N)$ are given by the Kalman filter and the fixed interval smoother [11], [12].

Maximization with respect to the unknown parameters $\sigma_{\xi r}^2$, $\sigma_{\xi i}^2$, $\sigma_{\zeta r}^2$, and $\sigma_{\zeta i}^2$ yields identical update equations to those in [1], which are valid and must be computed separately for both the real and imaginary parts. By iterating the conditional expectation and the update, we obtain the parameters for each row of the transformed image and the estimate of the transformed image itself. The original image is then obtained by taking the inverse transform of the estimated image.

1-D AR model parameters of the DFT image preserve all the information that is sufficient to identify 2-D causal (QP and NSHP) and semicausal AR model parameters of the original image. Having computed the 1-D model parameters, we solve the linear equation system resulting from (13) by making use of the well-known least squares method and obtain the parameters of the original image model [4], [1].

VI. Simulation Results

In this section, we present the simulation results of the proposed identification method for first-order AR models and compare them with those in [1]. As the original image, we have used the monochromatic 100 × 100 face image shown in Fig. 2. Before processing, we have normalized the observed image, i.e., we have corrected for its mean value in order to satisfy the zero mean assumption on the model noise.

TABLE I
Identification in DFT Domain

	Quarter plane		NSHP		Semicausal	
Noiseless	−0.488	0.795	−0.261	0.475	−0.135	0.286
	0.693		0.710		0.697	
			0.076		−0.135	0.286
SNR = dB	−0.556	0.779	−0.389	0.561	−0.226	0.341
	0.777		0.780		0.769	
			0.048		−0.226	0.341
SNR = 10 dB	−0.482	0.776	−0.225	0.418	−0.125	0.271
	0.706		0.725		0.709	
			0.081		−0.125	0.271

TABLE II
Identification in DST Domain

	0 dB		10 dB	
Semicausal	−0.208	0.308	−0.132	0.276
	0.801		0.712	
	−0.208	0.308	−0.132	0.276

TABLE III
Performance of Restoration in DFT and DST Domain: Improvements in SNR

	DFT	DST
SNR = 0 dB	9.611 dB	9.506 dB
SNR = 10 dB	4.518 dB	4.977 dB

Tables I and II show the identification results in DFT and DST domain, respectively. We measure the performance of AR model identification by comparing the identified coefficients with the coefficients found by directly applying the LS method to (20) and (21) using the noiseless original image. In order to improve the performance of identification, we have also used the constraint $\sum_{k,l\in S} a_{kl} = 1$, which is true for most of the real images.

We measure the performance of image restoration by the improvement in signal-to-noise ratio (SNR) which is defined as

$$\text{SNR} = 10 \log \frac{\sigma_x^2}{\sigma_v^2} \quad \text{(dB)} \tag{28}$$

where σ_x^2 is the variance of the original image and σ_v^2 is the mean squared error between the original image and the observed or restored image. The performances of restoration in DFT and DST domain at 0 dB and 10 dB noise, by the improvement in SNR, are displayed in Table III. The 0 dB noisy image and the images restored by DFT and DST are shown in Figs. 3, 4, and 5, respectively.Although the improvements in SNR are almost identical, in view of Figs. 3, 4, and 5, we observe that some important features of the original face image seem to be restored more successfully in DFT domain as compared to those restored in DST domain.

Since by using DFT we end up with a complex image in the transform domain with uncorrelated rows, the computational load in DFT approach increases by a factor of two, as compared to DST approach. However, the computation is still of the same order.

VII. Conclusion

In this study, we have proposed a new approach for the identification of the images degraded by noise, based on the method in [1]. The method presented in [1] makes use of the DST to decompose the image into scalar subsystems and can identify the semicausal AR parameters by applying the EM algorithm. In this study, we have used the DFT for decomposition, and in this way, it has become possible to

Fig. 3. Noisy image with SNR = 0 dB.

Fig. 4. Image restored by DFT.

identify both causal and semicausal AR parameters without *a priori* knowledge of the observation noise power.

The decomposition of images into uncorrelated scalar subsystems is based on the diagonalization of the coefficient matrices of the vector state-space model. While the DST almost diagonalizes the coefficient matrices of the symmetric semicausal model, the use of DFT does not require any approximation on the diagonalization of the coefficient matrices.

The approximation introduced by (19) in our method can be avoided; however, this would result in an unnecessary increase in the computational complexity without a significant improvement in the performance.

APPENDIX

Derivation of Equation (19)

Recalling (11), the complex coefficients $a_j = a_{r,j} + ia_{i,j}$ are found so as to minimize the log-likelihood function (omitting the constant terms)

$$\begin{aligned}\log p(\boldsymbol{Z}^N|\Theta) &= \frac{1}{\sigma_\xi^2}\sum_{m=1}^{N}|\theta(m)-\sum_{j=1}^{p}a_j\theta(m-j)|^2 \\ &= \frac{1}{\sigma_\xi^2}\sum_{m=1}^{N}[\theta_r(m)-\boldsymbol{a}_r^T\boldsymbol{\theta}_r(m-1)+\boldsymbol{a}_i^T\boldsymbol{\theta}_i(m-1)]^2 \\ &\quad+\frac{1}{\sigma_\xi^2}\sum_{m=1}^{N}[\theta_i(m)-\boldsymbol{a}_i^T\boldsymbol{\theta}_r(m-1) \\ &\quad-\boldsymbol{a}_r^T\boldsymbol{\theta}_i(m-1)]^2 \qquad (29)\end{aligned}$$

where $\boldsymbol{a}_r^T = [a_{r,1}\cdots a_{r,p}]$, $\boldsymbol{a}_i^T = [a_{i,1}\cdots a_{i,p}]$, $\boldsymbol{\theta}_r^T(m-1) = [\theta_r(m-1)\cdots\theta_r(m-p)]$, and $\boldsymbol{\theta}_i^T(m-1) = [\theta_i(m-1)\cdots\theta_i(m-p)]$, with respect to $\boldsymbol{a}_r$ and $\boldsymbol{a}_i$. Differentiating the log-likelihood function with respect to $\boldsymbol{a}_i$ and setting the derivative to zero yield

$$\begin{aligned}0 &= \left[\sum_{m=1}^{N}\boldsymbol{\theta}_r(m-1)\boldsymbol{\theta}_i^T(m-1)-\boldsymbol{\theta}_i(m-1)\boldsymbol{\theta}_r^T(m-1)\right]\boldsymbol{a}_r \\ &\quad+\left[\sum_{m=1}^{N}\boldsymbol{\theta}_r(m-1)\boldsymbol{\theta}_r^T(m-1)+\boldsymbol{\theta}_i(m-1)\boldsymbol{\theta}_i^T(m-1)\right]\boldsymbol{a}_i \\ &\quad+\sum_{m=1}^{N}\theta_r(m)\boldsymbol{\theta}_i(m-1)-\theta_i(m)\boldsymbol{\theta}_r(m-1). \qquad (30)\end{aligned}$$

Since edges appear infrequently in images, or, in other words, images contain relatively few high-frequency components, we introduce the following approximation:

$$\sum_{m=1}^{N}\theta_r(m)\theta_i(m-j)-\theta_i(m)\theta_r(m-j)\simeq 0 \qquad j=1,\cdots,p \quad (31)$$

Using (30) and (31), we get

$$\boldsymbol{H}\boldsymbol{a}_i \simeq 0 \qquad (32)$$

where $\boldsymbol{H} = \sum_{m=1}^{N}\boldsymbol{\theta}_r(m-1)\boldsymbol{\theta}_r^T(m-1)+\boldsymbol{\theta}_i(m-1)\boldsymbol{\theta}_i^T(m-1)$. Unless the whole image is zero, $\boldsymbol{H}$ is nonsingular, thus the unique solution to (32) is given by

$$\boldsymbol{a}_i \simeq 0.$$

Fig. 5. Image restored by DST.

REFERENCES

[1] T. Katayama and T. Hirai, "Parameter identification for noisy image via the EM algorithm," *Signal Process.*, vol. 20, no. 1, pp. 15–24, May 1990.

[2] R. L. Lagendijk, "Iterative identification and restoration of images," Ph.D. dissertation, Delft University of Technology, 1990.

[3] A. K. Jain, "An operator factorization method for restoration of blurred images," *IEEE Trans. Comp.*, vol. 26, no. 11, pp. 1061–71, 1977.

[4] Y. Yemez, "Image identification and restoration using EM algorithm," M.S. thesis, Boğaziçi University, 1992.

[5] A. K. Jain, "Advances in mathematical models for image processing," in *Proc. IEEE*, vol. 69, no. 5, pp. 502–528, 1981.

[6] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725–730, Nov. 1972.

[7] B. R. Musicus and J. S. Lim, "Maximum likelihood parameter estimation of noisy data," in *Proc. 1979 IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 224–227, 1979.

[8] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 95–103, 1988.

[9] A. Dembo, "Signal reconstruction from noisy partial information of its transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 1, pp. 65–72, 1989.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data," *J. Royal Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[11] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[12] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithms," *J. Time Ser. Anal.*, vol. 3, no. 4, pp. 253–263, 1982.

Vector Directional Filters—A New Class of Multichannel Image Processing Filters

P. E. Trahanias and A. N. Venetsanopoulos

Abstract— **Vector directional filters (VDF) for multichannel image processing are introduced and studied in this paper. These filters separate the processing of vector-valued signals into directional processing and magnitude processing. This provides a link between single-channel image processing, where only magnitude processing is essentially performed, and multichannel image processing where both the direction and the magnitude of the image vectors play an important role in the resulting (processed) image.**

VDF find applications in satellite image data processing, color image processing, and multispectral biomedical image processing. In this paper, results are presented for the case of color images, as an important example of multichannel image processing. It is shown that VDF can achieve very good filtering results for various noise source models.

I. INTRODUCTION

Although conventional approaches to multichannel image processing are based on processing the image channels separately, they fail to utilize the inherent correlation that is usually present in multichannel images. Consequently, *vector processing* of multichannel images is desirable [1]. Recently, this has been adopted by many researchers [2]–[4]. An important case of vector image processing operators are the vector median filters (VMF) that have been introduced as extension of scalar median filters [5]. VMF can be derived 1) as maximum likelihood estimates when the underlying probability densities are double-exponential or 2) using vector order statistics [6]. In the latter case, the vector median of a population is defined as the minimal vector according to the aggregate ordering technique [6]. Based on vector order statistics, extensions or modifications of VMF have also been proposed [2], [7].

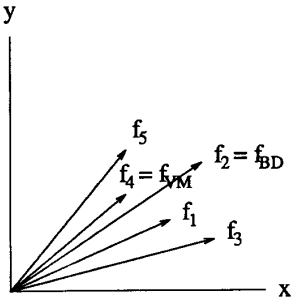


Fig. 1. A set of 2-dim vectors. The output of BVDF is always the middle vector (f_2); this is not necessarily the case for the VMF output (f_4).

The operation of the above-mentioned filters can be described according to some distance criterion that is applied to the set of vectors inside the processing window. However, the features that uniquely characterize a vector, namely *direction* and *magnitude*, are not considered by such an operation and this may produce erroneous results in certain cases. Such an example is shown in Fig. 1, where VMF is applied to the set of vectors $f_1, \cdots, f_5$. The output produced is vector f_4, although vector f_2 would be a better candidate to output.

This paper approaches the aforementioned problem by explicitly considering the vector features and separating the processing of vector-valued signals into two steps: *directional processing* and *magnitude processing*. A new class of filters is introduced, called *vector directional filters* (VDF). VDF perform the first step, namely directional processing. They operate on the direction of the image vectors aiming at eliminating vectors with *atypical* directions in the vector space. This is achieved by employing a novel vector ordering technique in which the *angle* between the image vectors serves as the ordering criterion. The term "directional processing" used here denotes the processing performed according to the vectors' direction in the vector space. This term has been adopted by other authors to denote processing in certain directions in the image plane [8]. Here it is used in the context of vector spaces and hence it should not bring any confusion. Similarly, the term "magnitude processing" denotes the processing of image data where only the vector magnitudes are taken into account.

The application of VDF results in the removal of vectors with atypical directions and a set containing vectors with approximately the same direction in the vector space is produced as the output set. Since the vectors in this set are approximately collinear, a magnitude processing operation (second step) can be applied to produce a single output vector at each image pixel. This step can be performed by any classical gray-level image processing filter.[1]

Manuscript received August 5, 1992; revised February 3, 1993. The associate editor coordinating the review of this paper and approving it for publication was Prof. R. Chellappa.

The authors are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4.

IEEE Log Number 9210837.

[1] This is obvious since gray-level image processing filters operate on the *magnitude* at each pixel location.