MODEL-BASED NON-LINEAR ESTIMATION FOR ADAPTIVE IMAGE RESTORATION

Xiaolin Wu and Xiangjun Zhang

Dept. of Electrical & Computer Engineering, McMaster University, Hamilton, Canada L8S 4K1

ABSTRACT

We propose a new image restoration algorithm that is driven by an adaptive piecewise autoregressive model (PAR). The strength of the new algorithm is its ability to preserve spatial structures better than its predecessors. The high adaptability is achieved by locally fitting 2D image waveform to the PAR model in moving windows. The problem is posed as one of nonlinear least-square estimation of both PAR parameters and original pixels, constrained by the degradation function. Robust solutions of the underlying underdetermined inverse problem are obtained by an innovative use of multiple PAR models that circumvent the issue of model overfitting, and by applying a structured total least-square technique.

Index Terms— Image restoration, autoregressive process, structured total least squares.

1. INTRODUCTION

Restoring or estimating an image from a degraded form is inherently an ill-posed inverse problem. The performance of an image restoration algorithm largely depends on how well it can employ regularization conditions or constraints when numerically solving the problem. The solution can be greatly improved if a good adaptive image model can be integrated into the estimation process, because the model can regulate estimated pixels according to useful prior statistical knowledge. Since most natural images have varying second-order statistics, an image model for restoration has to be adaptive. The model class of piecewise two-dimensional autoregressive process is capable of being fit to image signals of varying covariance matrix. Let $x(i,j)$ be an image, the piecewise autoregressive model (PAR) is defined by

$$x(i,j) = \sum_{(i,j)\in W} \sum_{u,v} a_{u,v} x(i-u, j-v) + e_{i,j} \qquad (1)$$

where W is a local window, and $a_{u,v}$'s are the parameters of the PAR model. The term $e_{i,j}$ is a random perturbation due to both fine scale chaotic details and measurement noises.

The power of the PAR model class comes from the flexibility of adjusting the model parameters $a_{u,v}$ to local pixel structures. The fact that any semantically meaningful image constructs, such as edges and surface textures, are formed by spatially coherent contiguous pixels, suggests piecewise statistical stationarity of the image signal. In other words, in the class of the PAR models, the parameters $a_{u,v}$ remain (nearly) constant in a small locality, although they may and often do vary significantly in different image segments. This property of natural images makes it possible to estimate the PAR model parameters using sample statistics in a moving window.

Model-based image restoration was also studied by Acton and Bovik [1]. They used two classes of image models, one is called piecewise linear and the other locally monotonic, to regulate restored pixel values. However, these models are not as versatile as the 2D piecewise autoregressive model. The validity of the PAR model class with locally adaptive parameters is corroborated by the success of this modeling technique in lossless image compression. Among all known lossless image coding methods, including CALIC, TMW [2], invertible DWT and DCT, those driven by the PAR model achieved the lowest lossless bit rates [3, 4]. In the principle of Kolmogorov complexity, the true model of a stochastic process is the one that yields the minimum description length. Thus we have strong empirical evidence to support the appropriateness and usefulness of the PAR model for natural images. Another advantage of the PAR model class is computational. The estimation of the model parameters and the restoration of the degraded pixels can be posed as a unified problem of non-linear least-squares, and solved by the total least-squares technique. Such a continuous optimization approach is more amenable and efficient than a combinatorial optimization approach that has to examine through a very large class of models [1].

The rest of the paper is structured as follows. Sec. 2 poses the problem of non-linear estimation for adaptive restoration (NEAR), which can be formulated as one of model-based block estimation. The selection of the image models is discussed. The risk of model overfit is assessed and a remedy is proposed. Sec. 3 develops a structured total least squares (STLS) solution for the NEAR problem. Experimental results are reported and compared with those of some popular image restoration methods in Sec. 4.

2. MODEL-BASED NONLINEAR ESTIMATION

Let $\boldsymbol{x}$ be the original image and $\boldsymbol{y}$ be a degraded form of $\boldsymbol{x}$,

$$y(i,j) = \sum_{m,n} h(m,n)x(i-m, j-n) + \epsilon, \qquad (2)$$

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

where $\boldsymbol{h}(m,n)$ is the degradation function and ϵ is an additive environment noise. First, we state an easily provable relation between $\boldsymbol{x}$ and $\boldsymbol{y}$, assuming $\boldsymbol{x}$ is piecewise autoregressive.

Proposition 1. *If an image is a 2D autoregressive (AR) process and it has gone through a degradation $\boldsymbol{h}$, then the degraded image is a 2D autoregressive moving average (ARMA) process. Its AR part has the same parameters as those of the original image, and its MA part is determined by $\boldsymbol{h}$.*

Although $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same AR parameters, one cannot directly estimate the PAR model parameters $a_{u,v}$'s from the observed image $\boldsymbol{y}$ using conventional linear least-squares method, because $\boldsymbol{y}$ has the signal-dependent MA part. But estimating $a_{u,v}$'s from $\boldsymbol{x}$ poses a problem of chicken and egg, because our original problem is to estimate $\boldsymbol{x}$ using $a_{u,v}$'s. To resolve this dilemma, we propose a new image restoration technique that jointly estimates the parameters of the PAR model and the pixels of the original image. We cast the joint estimation task as a constrained nonlinear least-squares problem having $a_{u,v}$'s and $\boldsymbol{x}$ both as variables. Our task of nonlinear estimation for adaptive restoration (NEAR) is to solve the following constrained optimization problem:

$$\begin{aligned}&\min_{\boldsymbol{x},\boldsymbol{a}} \sum_{(i,j)\in W}\left(x(i,j)-\sum_{u,v} a_{u,v}x(i-u,j-v)\right)^2\\ &\text{subject to}\quad \|\boldsymbol{x}*\boldsymbol{h}-\boldsymbol{y}\|=\sigma,\end{aligned} \tag{3}$$

where σ is variance of the noise. Here we assume that in local window W the original image $\boldsymbol{x}$ is a stationary 2D autoregressive process. This assumption is valid because common image constructs, such as edges and surface textures, tend to have consistent second order statistics in a locality.

However, one should exercise caution when applying (3), and be aware of the risk of data overfitting, or curse of dimensionality. If a PAR model of order t (the length of parameter vector $\boldsymbol{a}$ in (3)) is used, then the number of variables in (3) is $|W|+t$ ($|W|$ is the size of window W). But the number of equations between these variables is only $2|W|$, because the regularization term $\boldsymbol{x}*\boldsymbol{h}=\boldsymbol{y}$ generates $|W|$ equality constraints. Ideally, we want to make the order t of the PAR model high enough to fit $\boldsymbol{x}$ well and at the same time make the window size $|W|$ small enough such that $\boldsymbol{x}$ remains stationary in W. These two criteria may not be met simultaneously without risking data overfitting, because $|W|+t$ gets too close to $2|W|$ for large t and small $|W|$.

On a second reflection, fortunately, the two dimensions of the image signal offer ways to circumvent the problem of data overfitting. One way to increase the number of equations or constraints on pixels $\boldsymbol{x}\in W$ is the use of multiple PAR models that associate pixels in different directions. Among many possibilities, we introduce two PAR models of order 4, called the diagonal model $\mathcal{A}R_\times$ and the axial model $\mathcal{A}R_+$, which act on two disjoint neighborhoods of $x(i,j)$:

$$\begin{aligned}\mathbf{s}_+(i,j)&=(x(i,j-1),x(i-1,j);x(i,j+1),x(i+1,j))^T\\ \mathbf{s}_\times(i,j)&=(x(i-1,j-1),x(i-1,j+1),\\ &\qquad x(i+1,j+1),x(i+1,j-1))^T\end{aligned} \tag{4}$$

Vectors $\mathbf{s}_\times(i,j)$ and $\mathbf{s}_+(i,j)$ consist of four 8-connected neighbors and four 4-connected neighbors of $x(i,j)$ in the HR image, respectively, explaining our terminology of diagonal and axial models. Incorporating these two PAR models into the original nonlinear estimation framework, we modify the objective function (3) to the following:

$$\begin{aligned}\min_{\boldsymbol{x},\chi,\tau,\hat{w},\dot{w}}\Big\{&\sum_{(i,j)\in W}\hat{w}(x(i,j)-\chi^T\mathbf{s}_\times(i,j))^2+\\ &\sum_{(i,j)\in W}\dot{w}(x(i,j)-\tau^T\mathbf{s}_+(i,j))^2\Big\}\\ \text{subject to}\quad &\|\boldsymbol{x}*\boldsymbol{h}-\boldsymbol{y}\|=\sigma\end{aligned} \tag{5}$$

where $\chi=(\chi_1,\chi_2,\chi_3,\chi_4)$ and $\tau=(\tau_1,\tau_2,\tau_3,\tau_4)$ are parameters of the two PAR models $\mathcal{A}R_\times$ and $\mathcal{A}R_+$, and $\hat{w}$ and $\dot{w}$ are optimal weights of the two models. From (3) to (5) the number of equations is increased by $|W|$, whereas the number of unknown variables increases only by 2. Although we adopt two PAR models $\mathcal{A}R_\times$ and $\mathcal{A}R_+$ of order 4 only for the sake of preventing model overfit, the block estimation process of NEAR has the net effect of an adaptive non-separable two dimensional inverse filter that has as many as $|W|$ taps.

3. STRUCTURED TOTAL LEAST-SQUARES SOLUTION

This section presents an algorithm for NEAR. We make (5) an unconstrained nonlinear least square problem:

$$\begin{aligned}\min_{\boldsymbol{x},\chi,\tau,\hat{w},\dot{w}}\Big\{&\sum_{(i,j)\in W}\hat{w}(x(i,j)-\chi^T\mathbf{s}_\times(i,j))^2+\\ &\sum_{(i,j)\in W}\dot{w}(x(i,j)-\tau^T\mathbf{s}_+(i,j))^2+\lambda||\boldsymbol{y}-\boldsymbol{x}*\boldsymbol{h}||^2\Big\}\end{aligned} \tag{6}$$

The Lagrangian multiplier λ is adjusted such as the solution $\boldsymbol{x}$ satisfies $\|\boldsymbol{x}*\boldsymbol{h}-\boldsymbol{y}\|=\sigma$. For convenient representation we rewrite (6) in matrix form:

$$\min_{\boldsymbol{x},\chi,\tau,\hat{w},\dot{w}}\left\{\hat{w}\|\boldsymbol{x}-\mathbf{C}_1\boldsymbol{x}\|^2+\dot{w}\|\boldsymbol{x}-\mathbf{C}_2\boldsymbol{x}\|^2+\lambda\|\boldsymbol{y}-\mathbf{C}_3\boldsymbol{x}\|^2\right\} \tag{7}$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are $|W|\times|W|$ matrices. Matrix $\mathbf{C}_3$ corresponds to the convolution operation of $\boldsymbol{h}$.

Define the residue vector $\boldsymbol{r}(\boldsymbol{x},\chi,\tau),\hat{w},\dot{w}$ as

$$\boldsymbol{r}(\boldsymbol{x},\chi,\tau,\hat{w},\dot{w})=\begin{bmatrix}\boldsymbol{r}_1(\boldsymbol{x},\chi,\hat{w})\\ \boldsymbol{r}_2(\boldsymbol{x},\tau,\dot{w})\\ \boldsymbol{r}_3(\boldsymbol{x})\end{bmatrix} \tag{8}$$

where

$$\begin{aligned} \boldsymbol{r}_1(\boldsymbol{x}, \chi, \hat{w}) &= \sqrt{\hat{w}}(\mathbf{I} - \mathbf{C}_1) * \boldsymbol{x} \\ \boldsymbol{r}_2(\boldsymbol{x}, \tau, \dot{w}) &= \sqrt{\dot{w}}(\mathbf{I} - \mathbf{C}_2) * \boldsymbol{x} \\ \boldsymbol{r}_3(\boldsymbol{x}) &= \sqrt{\lambda}(\boldsymbol{y} - \mathbf{C}_3 * \boldsymbol{x}) \end{aligned} \tag{9}$$

and present (7) in the following quadratic form:

$$\min_{\boldsymbol{x}, \chi, \tau, \hat{w}, \dot{w}} \boldsymbol{r}(\boldsymbol{x}, \chi, \tau, \hat{w}, \dot{w})^T \boldsymbol{r}(\boldsymbol{x}, \chi, \tau, \hat{w}, \dot{w}) \tag{10}$$

The nonlinear least square problem (10) can be solved by an iterative algorithm of structured total least squares (STLS) [5]. First, we linearize the residue vector $\boldsymbol{r}(\boldsymbol{x}, \chi, \tau, \hat{w}, \dot{w})$. Let $\Delta\boldsymbol{x}$, $\Delta\chi$, $\Delta\tau$, $\Delta\hat{w}$, and $\Delta\dot{w}$ represent small changes in $\boldsymbol{x}$, χ, τ, $\hat{w}$, and $\dot{w}$ respectively, the residue vector $\boldsymbol{r}(\boldsymbol{x}, \chi, \tau, \hat{w}, \dot{w})$ can be linearized as following:

$$\begin{aligned} &\boldsymbol{r}(\boldsymbol{x} + \Delta\boldsymbol{x}, \chi + \Delta\chi, \tau + \Delta\tau, \hat{w} + \Delta\hat{w}, \dot{w} + \Delta\dot{w}) \\ &= \begin{bmatrix} \boldsymbol{r}_1(\boldsymbol{x} + \Delta\boldsymbol{x}, \chi + \Delta\chi, \hat{w} + \Delta\hat{w}) \\ \boldsymbol{r}_2(\boldsymbol{x} + \Delta\boldsymbol{x}, \tau + \Delta\tau, \dot{w} + \Delta\dot{w}) \\ \boldsymbol{r}_3(\boldsymbol{x} + \Delta\boldsymbol{x}) \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{r}_1(\boldsymbol{x}, \chi, \hat{w}) + \frac{\partial \boldsymbol{r}_1}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} + \frac{\partial \boldsymbol{r}_1}{\partial \chi}\Delta\chi + \frac{\partial \boldsymbol{r}_1}{\partial \hat{w}}\Delta\hat{w} \\ \boldsymbol{r}_2(\boldsymbol{x}, \tau, \dot{w}) + \frac{\partial \boldsymbol{r}_2}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} + \frac{\partial \boldsymbol{r}_2}{\partial \tau}\Delta\tau + \frac{\partial \boldsymbol{r}_2}{\partial \dot{w}}\Delta\dot{w} \\ \boldsymbol{r}_3(\boldsymbol{x}) + \frac{\partial \boldsymbol{r}_3}{\partial \boldsymbol{x}}\Delta\boldsymbol{x} \end{bmatrix} \end{aligned} \tag{11}$$

Therefore, given the current estimates of the HR pixels $\boldsymbol{x}$, the model parameters χ, τ, and the weights $\hat{w}$, $\dot{w}$, (10) reduces to

$$\min_{\Delta\boldsymbol{x}, \Delta\chi, \Delta\tau, \Delta\hat{w}, \Delta\dot{w}} \left\| \begin{bmatrix} \frac{\partial \boldsymbol{r}_1}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{r}_1}{\partial \chi} & \mathbf{0} & \frac{\partial \boldsymbol{r}_1}{\partial \hat{w}} & \mathbf{0} \\ \frac{\partial \boldsymbol{r}_2}{\partial \boldsymbol{x}} & \mathbf{0} & \frac{\partial \boldsymbol{r}_2}{\partial \tau} & \mathbf{0} & \frac{\partial \boldsymbol{r}_2}{\partial \dot{w}} \\ \frac{\partial \boldsymbol{r}_3}{\partial \boldsymbol{x}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\chi \\ \Delta\tau \\ \Delta\hat{w} \\ \Delta\dot{w} \end{bmatrix} + \begin{bmatrix} -\boldsymbol{r}_1(\boldsymbol{x}, \chi, \hat{w}) \\ -\boldsymbol{r}_2(\boldsymbol{x}, \tau, \dot{w}) \\ -\boldsymbol{r}_3(\boldsymbol{x}) \end{bmatrix} \right\|^2 \tag{12}$$

The resulting $\Delta\boldsymbol{x}$, $\Delta\chi$, $\Delta\tau$, $\Delta\hat{w}$ and $\Delta\dot{w}$ are the updates of the estimates of $\boldsymbol{x}$, the model parameters χ, τ, and the weights $\hat{w}$, $\dot{w}$ for the next iteration. In each iteration the least-squares problem (12) is linear with $\Delta\boldsymbol{x}$, $\Delta\chi$, $\Delta\tau$, $\Delta\hat{w}$ and $\Delta\dot{w}$ being variables, and hence it can be solved efficiently.

The remaining problem is the initialization of the STLS algorithm. By Proposition 1, the PAR model parameters can be initialized from $\boldsymbol{y}$:

$$\begin{aligned} \chi^{(0)} &= \arg\min_{\chi} \Big\{ \sum_{(i,j)\in W} \big(y(i,j) - \chi^T \mathbf{s}_\times(i,j)\big)^2 \Big\} \\ \tau^{(0)} &= \arg\min_{\tau} \Big\{ \sum_{(i,j)\in W} \big(y(i,j) - \tau^T \mathbf{s}_+(i,j)\big)^2 \Big\}, \end{aligned} \tag{13}$$

where $\mathbf{s}_\times(i,j)$ and $\mathbf{s}_+(i,j)$ are 8-connected and 4-connected neighborhood, respectively. The weights $\hat{w}$ and $\dot{w}$ are initialized as

$$\hat{w}^{(0)} = \frac{e_+}{e_+ + e_\times}; \quad \dot{w}^{(0)} = \frac{e_\times}{e_+ + e_\times} \tag{14}$$

where $e_\times$ and e_+ are the squared errors associated with the solutions of (13). These weights are optimal in least squares sense if the fit errors of the two PAR models are independent. With the initialized PAR model parameters and the two weights, $\boldsymbol{x}$ can be initialized by

$$\boldsymbol{x}^{(0)} = \arg\min_{\boldsymbol{x}} \left\| \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \sqrt{\lambda}\boldsymbol{y} \end{bmatrix} - \begin{bmatrix} -\mathbf{I} + \mathbf{C}_1 \\ -\mathbf{I} + \mathbf{C}_2 \\ \sqrt{\lambda}\mathbf{C}_3 \end{bmatrix} \boldsymbol{x} \right\|^2 \tag{15}$$

4. EXPERIMENTAL RESULTS AND REMARKS

In our experiments the input images were generated by passing test images through a degradation process plus small Gaussian noise. The degraded images were restored by NEAR and three well-known methods: Wiener deconvolution, Regularized Least Square (RLS) algorithm and the Lucy-Richardson algorithm. Two types of degradation were tested: the point spread function (PSF) of a camera and linear camera motion. Table 1 lists the PSNRs of the restored images by different methods when the degradation is caused by a Gaussian PSF and Gaussian noise of $\sigma^2 = 4 \times 10^{-4}$. For all test images, NEAR achieves the highest PSNR.

Fig. 1 and Fig. 2 compare different methods in visual quality. In Fig. 1, the input image is degraded (blurred) by a linear motion of 5 pixels; Fig. 2 lists output images of different methods when the degradation function is a Gaussian PSF. Although Lucy-Richardson and RLS algorithms can deblur the image, they introduce visually annoying noises in the restored images. The proposed NEAR method restores the details cleanly and in particular preserves the edge structures. The superior visual quality of the NEAR method should be evident by observing reconstructed flower pedals in Fig. 1, cloth patterns (in image Barb) and feathers (in image Lena) in Fig. 2.

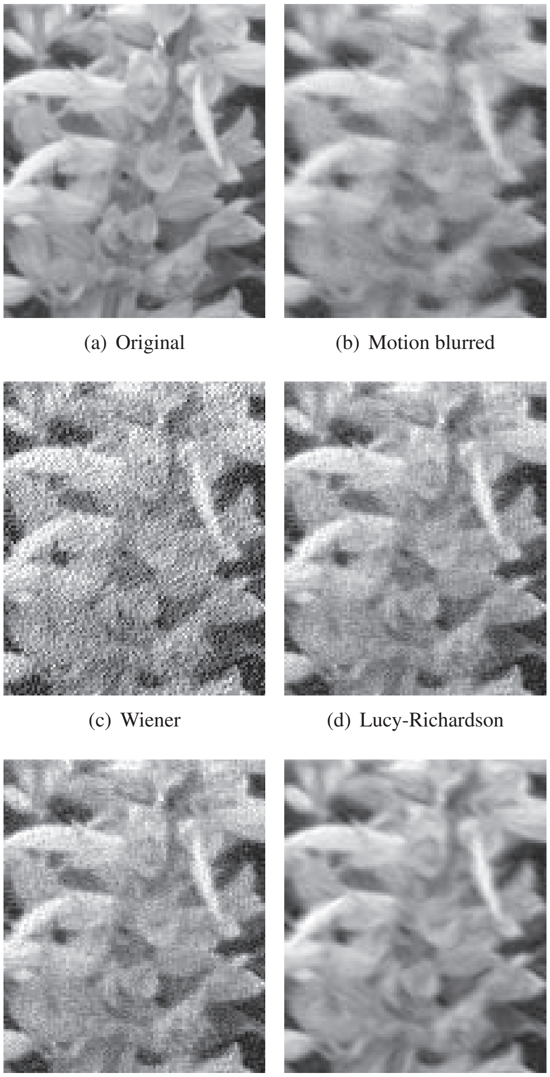
The proposed NEAR method can be applied when the degradation function or/and the noise level varies in the image. We are extending NEAR to situations where the degradation function is unknown.

Table 1. PSNR (dB) results of different methods.

Image	Wiener	RLS	Lucy-Richardson	NEAR
Lena	16.72	32.72	26.82	33.46
Leaves	18.03	30.62	27.45	32.15
Flower	18.72	31.40	26.98	32.37
Barb	16.87	25.90	25.51	28.28
Hat	16.41	30.09	26.75	31.34
Average	17.44	30.40	26.75	31.72

5. REFERENCES

[1] S. T. Acton and A. C. Bovik, "Piecewise and local image models for regularized image restoration using cross-validation," *IEEE*



(a) Original (b) Motion blurred
(c) Wiener (d) Lucy-Richardson
(e) RLS (f) NEAR

Fig. 1. Comparison of different methods on image Flower.

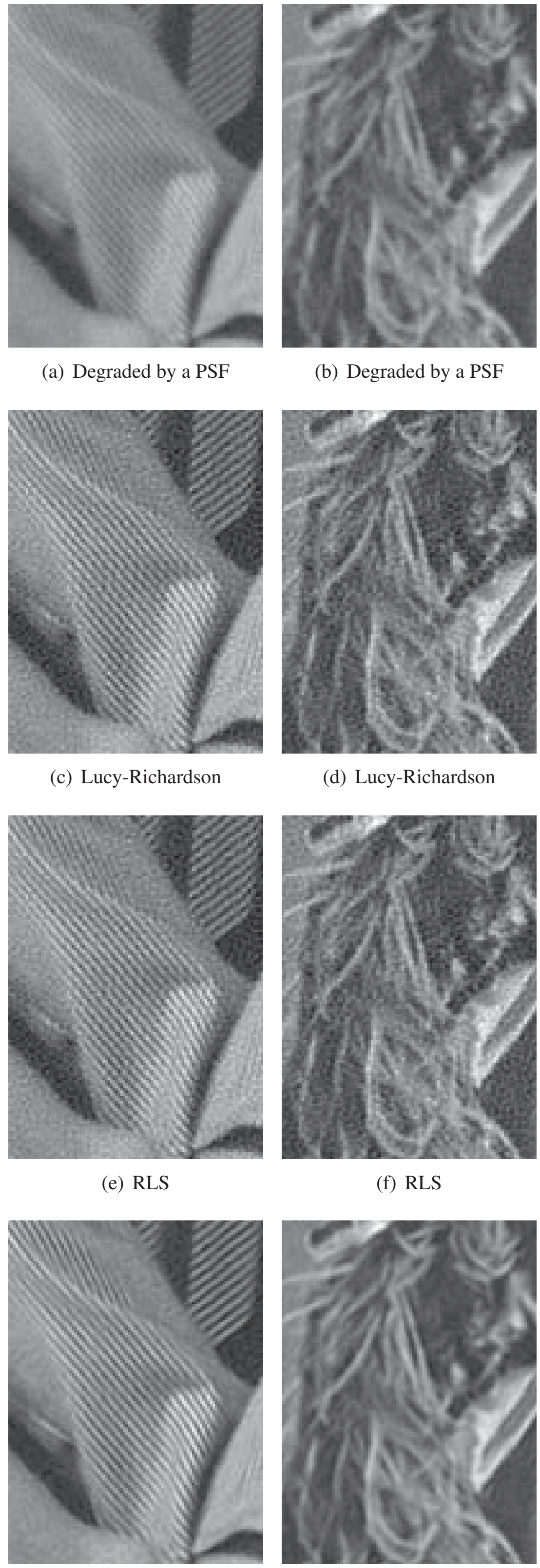
Trans. Image Processing, vol. 8, no. 5, pp. 652–665, May 1999.

[2] B. Meyer and P. Tischer, "Tmw – a new method for lossless image compression," in *Proceedings of 1997 Picture Coding Symposium*, Berlin, Germany, October 1997.

[3] X. Wu, K. U. Barthel, and W. Zhang, "Piecewise 2d autoregression for predictive image coding." in *Proceedings of the IEEE Int. Conf. on Image Processing*, vol. 3, Chicago, Oct. 1998, pp. 901–904.

[4] I. Matsuda, H. Mori, J. Maeda, and S. Itoh, "Design and evaluation of minimum-rate predictors for lossless image coding," *Trans. IEICE (DII)*, vol. J85-D-II, no. 3, pp. 448–456, March 2002, (in Japanese).

[5] J. Rosen, H. Park, and J. Glick, "Structured total least norm for nonlinear problem," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 1, pp. 14–30, Jan. 1998.



(a) Degraded by a PSF (b) Degraded by a PSF
(c) Lucy-Richardson (d) Lucy-Richardson
(e) RLS (f) RLS
(g) NEAR (h) NEAR

Fig. 2. Comparison of different methods on images Barb and Lena.