

Conditioning Analysis of Missing Data Estimation for Large Sensor Arrays

Hairong Qi
ECE Department, Ferris Hall
The University of Tennessee, Knoxville
Knoxville, TN 37996
hqi@utk.edu

Wesley E. Snyder
CACC, Box 7914
North Carolina State University
Raleigh, NC 27695-7914
wes@eos.ncsu.edu

Abstract

Optimal missing data estimation algorithms including deblurring and denoising are designed to restore images captured from large CCD sensor arrays using butting technique, where 1 to 2 columns of data are missed at the butting edge. We developed consistency method with separable deblurring to estimate the missing data. This method converts an ill-posed restoration problem into a well-posed one by making few assumptions based on regularization theory. Under the condition that no noise is inserted, and the separable blur kernel is exactly known, the consistency method can deblur the original image and at the same time estimate the missing column(s) exactly. However, this algorithm becomes unstable when large noise is inserted or inaccurate estimation of blur kernel is made. Conditioning analysis is used to quantify the amount of ill condition of the blur kernel when the assumptions are relaxed to different levels, which provides a solid measurement on how stable the system will remain knowing the signal-to-noise ratio and the inaccuracy of the blur kernel estimation. Experimental results from different approaches are compared.

1. Introduction

Missing data estimation has become one of the increasingly important problems in the vision community. Interesting applications include modeling the “filling-in” of the blind-spot in human vision, fixing bad pixels in CCD sensor arrays, estimating missing data for large sensor arrays, etc. The discussion in this paper focuses on missing data estimation for large CCD sensor arrays, but the algorithm developed should be general enough to be able to solve similar problems.

Among all the image sensors, the charge-coupled-device (CCD) is the dominant technology for both user and industry applications. CCD sensors provide higher resolution, wider dynamic range, and higher sensitivity, with which

conventional photography can not compete. However, CCD sensors also cost more. How to achieve large area and high resolution, while at the same time remaining cost-effective is an issue that draws a lot of concern.

One of the economic solution to large area is to use several CCDs side by side. This method is technically called *butting*. In our project, we use devices from EEV, Inc. EEV developed a light-guide system to butt CCD arrays together. The light-guide system uses fiber optics where a bunch of fibers connect each point on the scintillator with a corresponding cell on the CCD detector. In this way, light from an X-ray scintillator may be guided down to a CCD array preserving the alignment of all the pixels except that the gap between the two detectors is still on the order of 1 to 2 pixel(s) which is about 50 to 100 micron (Fig. 1). An array of 1×2 CCDs (1152 rows and 1242 columns each) are used for experimental purpose. However, there is no limits as how many CCD arrays can be butted together in this way.

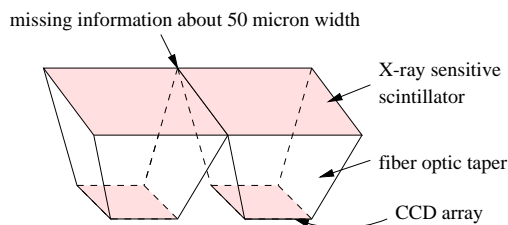


Figure 1. Two scintillator/fiber/CCD combinations butted together.

2. Problem Formulation

The missing data estimation problem can be formulated as Fig. 2. An original image f is slightly blurred by a point spread function (PSF) h from the X-ray source. The blurred image is further corrupted by fixed noise (n) and defects (d) from the CCD detector. g is the so called *measured image*.

The image formation process can be expressed by Eq. 1,

$$g = f \otimes h + n - d \quad (1)$$

where \otimes stands for the convolution operator. We assume that the PSF is already known, the noise distributes as independent Gaussian, and the positions of missing data can be identified.

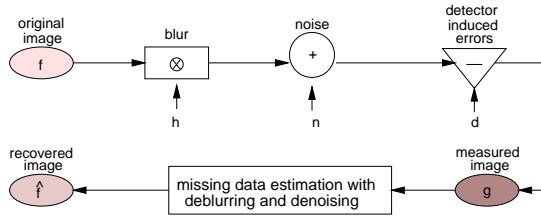


Figure 2. System model.

We propose the so-called *consistency method* using separable deblurring to estimate missing data along with deblurring and denoising. The basic idea behind this approach is to make use of the point spread function (PSF): before a pixel is missed, it has already distributed its information to its neighbors through the effect of blur. Therefore, if we can obtain a measurement of the PSF, in theory, we can reconstruct the missing data.

3. Regularization Theory

Missing data estimation is basically an image restoration problem. Image restoration is well known to be *ill-posed* [1][5], since a little change in the input data can dramatically affect the solution. Ill-posedness comes from the *ill-conditioning* of the matrix constructed from the system PSF. The most popular solution to ill-posedness is *regularization*. Although many different kinds of image restoration algorithms have been proposed so far, they all share a common structure: the *regularization theory*.

Generally speaking, any regularization method tries to analyze a related *well-posed* problem whose solution approximates the original *ill-posed* problem. The well-posedness is achieved by implementing one or more of the following basic ideas [7]: (1) change of the concept of a solution; (2) restriction of the data; (3) change of the space and/or topologies; (4) modification of the operator itself; (5) the concept of regularization operators; and (6) well-posed stochastic extensions of ill-posed problems.

4. The Consistency Method with Separable Deblurring

By putting some restrictions on input data and the blur operator, we are able to transform the missing data estima-

tion problem into a well-posed one. Several assumptions are made to simplify the problem: (1) the blur kernel is separable; (2) the blur kernel is exactly known; (3) noise is not considered for the time being; (4) only one column of data is missed in the original image; and (5) the original image is of integer type. Possible relaxation of these assumptions are further discussed in Sec. 5.

The image formation is modeled as Eq. 2,

$$f \otimes h = g - n \quad (2)$$

where g is the measured image with one column of data missing, f is the original image, and n , the noise, is regarded as small perturbation of g and is neglected in this section. The convolution kernel h is discrete and of finite support. Both f and g are $M \times N$ matrices.

The consistency approach concerns three issues: separable deblurring, Householder triangularization based QR factorization (HHQR), and consistency in missing data estimation [9][8].

4.1. Separable deblurring

Since we assume the convolution kernel $h_{m \times n}$ is separable, the convolution can then be performed separately as Eq. 3, where $hy_{m \times 1}$ and $hx_{1 \times n}$ are the separated vertical and horizontal components of h . $*$ denotes the matrix multiplication.

$$f \otimes h = f \otimes (hy * hx) = (f \otimes hy) \otimes hx = (f \otimes hx) \otimes hy \quad (3)$$

Assume f is an $M \times N$ image and h is an $m \times n$ Gaussian kernel, the convolution of image f with the horizontal kernel component hx can be equally achieved by a matrix multiplication between f and an $N \times N$ ($\lfloor \frac{n-1}{2} \rfloor, \lceil \frac{n-1}{2} \rceil$)-band matrix Dx (Eq. 4), generated from hx ; and the convolution with hy can be similarly achieved by a matrix multiplication with an $M \times M$ ($\lfloor \frac{m-1}{2} \rfloor, \lceil \frac{m-1}{2} \rceil$)-band matrix Dy (Eq. 5), generated from hy . A general band matrix has its nonzero elements arranged uniformly near the diagonal. A band matrix with m lower band widths and n upper band widths is denoted as (m, n) -band matrix.

$$Dx = \begin{cases} hx(\lfloor \frac{n+1}{2} \rfloor - i + j) & -\lceil \frac{n-1}{2} \rceil \leq i - j \leq \lfloor \frac{n-1}{2} \rfloor \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$Dy = \begin{cases} hy(\lfloor \frac{m+1}{2} \rfloor - i + j) & -\lceil \frac{m-1}{2} \rceil \leq i - j \leq \lfloor \frac{m-1}{2} \rfloor \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

With Dx and Dy , we can rewrite Eq. 3 into Eq. 6, which gives us another interpretation on how the original image is blurred. Eq. 7 is the inverse problem to Eq. 6, where if matrices Dx and Dy are both well-conditioned, the solution f

should be stable. The conditioning of Eq. 6 will be analyzed in detail in Sec. 5.

$$g = (f \otimes hx) \otimes hy = Dy * (Dx * f^T)^T \quad (6)$$

$$f = (Dx^{-1} * (Dy^{-1} * g)^T)^T \quad (7)$$

4.2. HHQR

To actually solve the system of Eq. 6, we use Householder triangularization based QR factorization (HHQR) algorithm [10]. HHQR has been proven to be *backward-stable* which means if the linear system is well-conditioned, this algorithm can provide the accurate solution.

In our problem, HHQR algorithm is used twice to solve the two linear systems derived from Eq. 6:

1. solve $g = Dy * f_y$ for f_y , and
2. solve $f_y^T = Dx * f^T$ for f .

4.3. Missing data estimation by consistency

Assume the original image has only integer brightness values, and these values are within $[0, 255]$, in particular, the missing pixels should have their original integer brightness values between 0 and 255, the two linear systems are solved as follows:

1. Solve $g = Dy * f_y$ for f_y using HHQR. The process is to deblur image g along the vertical direction. We show in Sec. 5 that matrix Dy is well-conditioned such that the inverse problem is well-posed.
2. For the second linear system $f_y^T = Dx * f^T$, HHQR can not be used directly, since there is one column of missing pixels in f_y . Instead, we solve each row of the image separately, i.e. solve $f_y^T(i) = Dx * f^T(i)$, where i represents the i th row in that image. For example, if we have an image f of dimension 5×4 , and a separable blur kernel h of dimension 4×3 , the second linear system can be written as Eq. 8,

$$\begin{bmatrix} f_y(i, 1) \\ f_y(i, 2) \\ f_y(i, 3) \\ f_y(i, 4) \end{bmatrix} = \begin{bmatrix} hx_2 & hx_3 & 0 & 0 \\ hx_1 & hx_2 & hx_3 & 0 \\ 0 & hx_1 & hx_2 & hx_3 \\ 0 & 0 & hx_1 & hx_2 \end{bmatrix} * \begin{bmatrix} f(i, 1) \\ f(i, 2) \\ f(i, 3) \\ f(i, 4) \end{bmatrix} \quad (8)$$

Assume the missing pixel is in the 3rd column, we then have 5 unknowns ($f(i, 1)$, $f(i, 2)$, $f(i, 3)$, $f(i, 4)$ and $f_y(i, 3)$) but only 4 equations in the linear system. To solve this problem, we need to find another condition. Since we already know that $f(i, 3)$ should be an integer between 0 and 255, an exhaustive search can be carried on by assuming $f(i, 3)$ is each one of them, which provides 4 unknowns

and 4 equations. Reconstruct vectors f_y , f , and matrix Dx in Eq. 8, so that all the unknowns are at one side of the equation, like Eq. 9. We denote the matrix reconstructed from Dx as Drx , and show in Sec. 5 that it is also well-conditioned.

$$\begin{bmatrix} f_y(i, 1) \\ f_y(i, 2) - hx_3 \cdot f(i, 3) \\ -hx_2 \cdot f(i, 3) \\ f_y(i, 4) - hx_1 \cdot f(i, 3) \end{bmatrix} = \begin{bmatrix} hx_2 & hx_3 & 0 & 0 \\ hx_1 & hx_2 & 0 & 0 \\ 0 & hx_1 & -1 & hx_3 \\ 0 & 0 & 0 & hx_2 \end{bmatrix} * \begin{bmatrix} f(i, 1) \\ f(i, 2) \\ f_y(i, 3) \\ f(i, 4) \end{bmatrix} \quad (9)$$

Solve the new linear system by HHQR, and check the solution to see if it satisfies the *consistency criterion*, that is, the solutions ($f(i, 1)$, $f(i, 2)$, $f(i, 4)$) are all between 0 and 255 and are all integers. If they do, then the process can continue to the next row; otherwise, next value of $f(i, 3)$ is attempted. The well-conditioned problem and the backward-stable algorithm assure that the solution is unique and accurate.

5. Conditioning Analysis

Conditioning of a mathematical problem is measured by the sensitivity of output to changes in input. For a well-conditioned problem, a small change of input does not affect the output much; while for an ill-conditioned problem, a small change of input can change the output a great deal.

Condition number is the measurement of the conditioning of a problem. The conditioning of a linear system $Ax = b$ is determined by the condition number of matrix A . The relative condition number K is defined as Eq. 10,

$$K = \|A\| \|A^{-1}\| \quad (10)$$

where $\|\cdot\|$ usually indicates the 2-norm. K is in the range of $[1, \infty)$. When $K \gg 1$, the linear system is ill-conditioned.

While most of the literature on image restoration concentrates on smoothing methods to reduce the effect of noise, few [4][6] pay attention to quantify the amount of ill condition of the blur kernel. Since the whole idea of the consistency method is built upon the well-conditioning of the two linear systems, analyzing the amount of ill condition of matrices becomes an important issue.

In the missing data estimation problem, the conditioning of the two linear systems is measured by the condition number of the two matrices: Dy and Drx .

Fig. 3 shows the condition number of Dy calculated with different image sizes and different Gaussian blur kernel sizes. Fig. 3 (a) is the growth curve for 3×3 blur kernel. The curve increases steadily without converging to an

upper bound, which indicates the system is ill-conditioned. Fig. 3 (c) shows six curves with respect to six different kernel sizes: 5×5 , 7×7 , 9×9 , 11×11 , 13×13 , and 15×15 . All the curves behave similar: a sharp increase at the beginning and asymptotically converge to a constant (less than 11) after the image size is larger than 100×100 . Fig. 3 (b) displays three curves corresponding to different image sizes. It indicates that all the condition numbers fall into the area with an upper bound created by the largest image size (2000×2000) and a lower bound related to the smallest image size (15×15). The upper bound is around 10, which demonstrates the well-conditioning of the system.

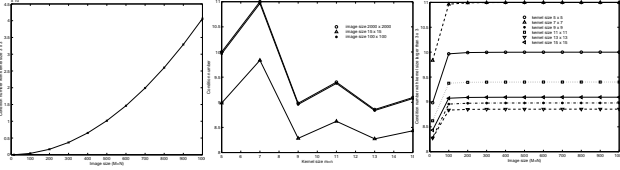


Figure 3. Condition number of matrix Dy with respect to different image and kernel sizes.

Matrix Drx is reconstructed from Dx . Its conditioning curves behave very similar to those of Dy , except that the upper bound is larger (less than 450) than that from Fig. 3 (b) and (c). This still indicates well-conditioning since this bound is not far greater than 1.

From the above analysis, we claim that when blur kernel is larger than 3×3 , both of the linear systems are well-conditioned. We also point out that this result is obtained from experiments, which has not been proved yet.

At the beginning of Sec 4, we made several assumptions to simplify the restoration problem. This section discusses the possible relaxation to these assumptions. Except that the blur kernel still needs to be separable, we show that other assumptions can all be relaxed to a certain degree, though by sacrificing certain amount of accuracy of the solution. We first analyze sensitivity of solution to perturbations in blur kernel (h), and perturbations in measured image (g) caused by the insertion of noise (n). We also evaluate the conditioning of problem when there are more than one column of missing data. Finally, we relax the assumption of integer-typed original image, and design a *neighbor least square error* criterion to select the optimal solution [8].

5.1. Sensitivity of solution to perturbations in blur kernel

According to [10], in a linear system $Ax = b$, sensitivity of solution x to perturbations in b can be measured by

Eq. 11,

$$\frac{\|x - \tilde{x}\|}{\|x\|} = \frac{K(A) \|\delta b\|}{\eta \|b\|} \quad (11)$$

where $\eta = \frac{\|A\| \cdot \|x\|}{\|Ax\|}$. Sensitivity of x to perturbations in A can be estimated by Eq. 12.

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq K(A) \frac{\|\delta A\|}{\|A\|} \quad (12)$$

In our system, perturbations in blur kernel (h) are reflected in blur matrices Dy and Drx as shown in Eq. 13 and Eq. 14.

$$(Dy + \delta Dy) \tilde{f}_y = g \quad (13)$$

$$(Drx + \delta Drx) \tilde{f}^T = f_y^T \quad (14)$$

Based on Eq. 12, sensitivity of f with respect to perturbations in Drx can be expressed by Eq. 15. Based on Eq. 11 and Eq. 12, sensitivity of f to perturbations in Dy can be expressed by Eq. 16,

$$\frac{\|f - \tilde{f}\|}{\|f\|} \leq K(Drx) \cdot \frac{\|\delta Drx\|}{\|Drx\|} \quad (15)$$

$$\frac{\|f - \tilde{f}\|}{\|f\|} = \frac{K(Drx)}{\eta_x} \cdot \frac{\|\delta f_y\|}{\|f_y\|} \leq \frac{K(Drx)K(Dy)}{\eta_x} \cdot \frac{\|\delta Dy\|}{\|Dy\|} \quad (16)$$

where $\eta_x = \frac{\|Drx\| \cdot \|f\|}{\|Drx \cdot f\|}$, a number greater than 1, but very close to 1. If Drx and Dy have the same degree of perturbation, then Eq. 15 derives a smaller upper bound than Eq. 16. Therefore, sensitivity of f to perturbations in blur kernel h is measured more accurate by Eq. 15.

5.2. Sensitivity of solution to perturbations in measured image

Perturbations in measured image g can be interpreted as the effect of inserted noise. The two perturbed linear systems can be written as Eq. 17 and Eq. 18.

$$Dy \cdot \tilde{f}_y = g + \delta g \quad (17)$$

$$Drx \cdot \tilde{f}^T = f_y^T + \delta f_y^T \quad (18)$$

Sensitivity of f_y to perturbations in measured image g is defined by Eq. 19 based on Eq. 11,

$$\frac{\|f_y - \tilde{f}_y\|}{\|f_y\|} = \frac{K(Dy)}{\eta_y} \cdot \frac{\|\delta g\|}{\|g\|} \quad (19)$$

where $\eta_y = \frac{\|Dy\| \cdot \|f_y\|}{\|Dy \cdot f_y\|}$, a number greater than 1, but very close to 1. Sensitivity of f to perturbations in measured image g can be derived by Eq. 20.

$$\frac{\|f - \tilde{f}\|}{\|f\|} = \frac{K(Drx)}{\eta_x} \cdot \frac{\|\delta f_y\|}{\|f_y\|} = \frac{K(Drx)K(Dy)}{\eta_x \cdot \eta_y} \cdot \frac{\|\delta g\|}{\|g\|} \quad (20)$$

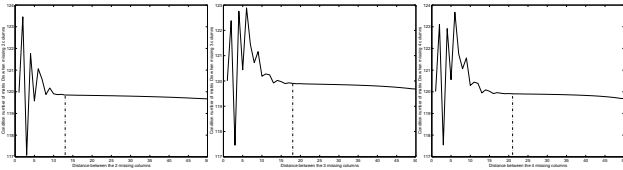


Figure 4. Condition number of Drx with different numbers of missing columns at kernel size 5×5 .

Compare Eq. 20 with Eq. 15, if Drx and g have same degree of perturbation, then solution f can be less sensitive to perturbations in blur kernel than to perturbations in measured image. We will verify this conclusion by experimental results in Sec. 6.

5.3. Conditioning analysis for more than one missing column

If there are more than one column of missing data in the measured image, the condition number of the vertical blur matrix Dy stays the same, but condition number of the reconstructed horizontal blur matrix Drx is changed. Here, two problems need to be considered: how many missing columns can be in the image, and how far away these missing columns should be apart for stable inverse.

Fig. 4 shows the condition number of Drx with blur kernel (h) at dimension 5×5 . The three plots are generated with different numbers of missing columns (2, 3, and 4). We can see that no matter how many columns are missed in the measured image, and how far away the two nearest missing columns are apart, all of the plots behave similarly - after an initial perturbation, all plots converge to a constant condition number, which is around 120.

Fig. 5 compares the condition number of Drx with different kernel size when two columns of data are missed. Both of the two curves increase steadily with respect to the kernel size.

These results substantiate our claim that when more than one column of data are missed, it is not the image size, or the distance between two nearest missing columns, or the number of missing columns that affect the condition number, it is the *kernel size* that plays the most important role. When the kernel size is less than 10×10 , the problem is still well-conditioned. The larger the kernel size, the worse the problem is conditioned.

5.4. Neighbor least square error consistency

The last assumption we made is that the original image needs to be of integer type. We relax this assumption to

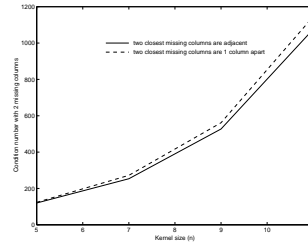


Figure 5. Condition number comparison of Drx , missing 2 columns of data, with different kernel size: 5×5 , 7×7 , 9×9 , and 11×11 .

float-type image and design a new consistency criterion to be used when selecting the optimal solution at the last step of the consistency method. We call it *neighbor least square error* (NLSE) consistency criterion. For each possible solution of $f(i)$ (the i th row of the restored image) solved by assuming the missing pixel ranges from 0 to 255, the optimal solution is the one that minimizes NLSE. NLSE is computed by Eq. 21, where d is the column position of the missing data, and δ is the number of neighbors that is involved in this least square error computation. It is important to choose close neighbors instead of the entire row for the computation.

$$\| f(i) \otimes hx - f_y(i) \|^2 = \sum_{j=d-\delta}^{d+\delta} [f(i) \otimes hx)_j - f_y(i, j)]^2 \quad (21)$$

6. Experimental Results

Based on the theoretical analysis in Sec. 5, this section evaluates the algorithm performance by experimental results using synthetic images. Five synthetic images (piecewise constant, piecewise linear, piecewise quadratic, sinusoid, and mammogram) are generated showing different degrees of smoothness and different image properties.

Experimental results from both the consistency method (using integer criterion and NLSE criterion) and the mean field annealing (MFA) method [3][2] are exhibited and compared.

With the complete set of assumptions (Sec. 4) satisfied, the consistency method using integer criterion can recover the missing column *exactly* from the measured image. A cross-correlation comparison between the original image and the restored one using different methods shows that the consistency method with integer criterion always has the highest cross-correlation coefficient, which is 1. The NLSE criterion shows better correlation than the MFA method most of the time except for sinusoid images. See Fig. 6.

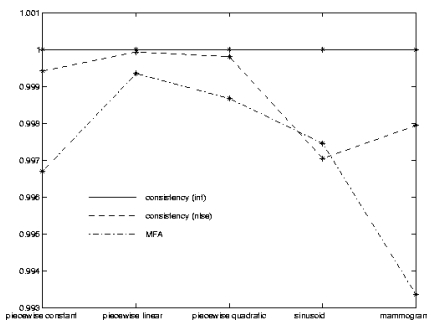


Figure 6. Cross-correlation comparison.

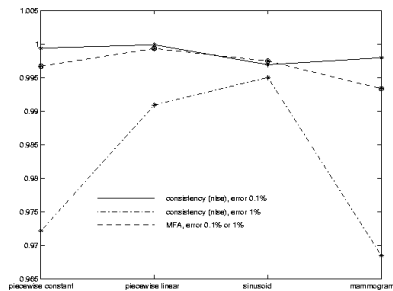


Figure 7. Cross-correlation comparison.

We also conduct experiments by relaxing the blur kernel and noise free assumptions.

We use blur kernel h to blur the original image, but use $h + \varepsilon$ to restore the measured image, where ε represents the error added to a blur kernel. The accuracy of an estimated blur kernel is measured by correct number of digits. For example, $\varepsilon = 1\%$ or 0.1% means the 2nd or the 3rd digit of elements in the blur kernel is not correct. From the cross-correlation comparison in Fig. 7, we can see that the MFA method is not sensitive to very small errors in the blur kernel since the correlation coefficients do not change when increasing the error rate from 0.1% to 1%. However, the coefficients of the consistency method with NLSE criterion drops when increasing the error. When error is on the 3rd digit, the consistency method provides better correlation than the MFA method except for the sinusoid image.

Three degrees of Gaussian white noises are inserted to the blurred image with one column deleted. The column profiles of restoration results in Fig. 8 show us that when noise is very small, the consistency method performs better than the MFA method, it achieves a higher correlation coefficient than the MFA method. However, when increasing the standard derivation of the inserted noise, the consistency method is much more disturbed than the MFA method.

To summarize, the consistency method works better when the noise and errors in the blur kernel are small. Performance of the consistency method may be largely affected

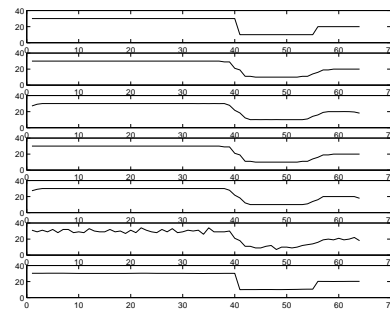


Figure 8. Column profiles comparison. The seven curves (from top to bottom) represents profiles from the original image, restored image by NLSE ($\sigma = 0.01$), by MFA ($\sigma = 0.01$), by NLSE ($\sigma = 0.1$), by MFA ($\sigma = 0.1$), by NLSE ($\sigma = 1$), by MFA ($\sigma = 1$)

by perturbations in the estimated blur matrix, or by noise, with perturbations in noise playing a more important role.

References

- [1] H. Andrews and H. B.R. *Digital Image Restoration*. Prentice-Hall, 1977.
- [2] G. Bilbro, R. Mann, T. Miller, W. E. Snyder, et al. Optimization by mean field annealing. In *Advances in Neural Information Processing Systems*. Morgan-Kaufman, San Mateo, 1989.
- [3] G. L. Bilbro and W. E. Snyder. Image restoration by mean field annealing. In *Advances in Neural Network Information Processing Systems*. N/A, 1988.
- [4] K. Forbes and V. Anh. Condition of system matrices in image restoration. *J. Opt. Soc. Am. A*, 11(6):1727–1735, June 1994.
- [5] A. Katsaggelos. *Digital Image Restoration*. Springer-Verlag, 1991.
- [6] F. Milinazzo, C. Zala, and I. Barrodale. On the rate of growth of condition numbers for convolution matrices. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-35(4):471–475, April 1987.
- [7] M. Z. Nashed. Aspects of generalized inverses in analysis and regularization. In M. Z. Nashed, editor, *Generalized Inverses and Applications*. Academic, New York, 1976.
- [8] H. Qi. *A High-Resolution, Large-Area, Digital Imaging System*. PhD thesis, North Carolina State University, Raleigh, June 1999.
- [9] H. Qi, W. E. Snyder, and G. L. Bilbro. Missing data estimation by separable deblurring. In *Proceedings for teh IEEE International Joint Symposia on Intelligence and Systems*, pages 348–353, Rockville, MD, May 1998.
- [10] L. Trefethen and B. D. III. *Numerical Linear Algebra*. SIAM, 1997.