

A method for restoration of low-resolution document images

Paul D. Thouin¹, Chein-I Chang²

¹ Department of Defense, Fort Meade, MD 20755, USA; e-mail: pthouin@afterlife.ncsc.mil

² Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA; e-mail: cchang@umbc.edu

Received September 1, 1999 / Revised November 23, 1999

Abstract. Image restoration using resolution expansion is important in many areas of image processing. This paper introduces a restoration method for low-resolution text images which produces expanded images with improved definition. This technique creates a strongly bimodal image with smooth regions in both the foreground and background, while allowing for sharp discontinuities at the edges. The restored image, which is constrained by the given low-resolution image, is generated by iteratively solving a nonlinear optimization problem. Low-resolution text images restored using this technique are shown to be both quantitatively and qualitatively superior to images expanded using the standard methods of linear interpolation and cubic spline expansion. Experimental results demonstrate that text images created by this new algorithm improve optical character recognition accuracy more than images obtained by existing expansion methods.

Key words: Bimodal distribution – Modeling – Restoration – Optimization – Text

1 Introduction

Text image resolution expansion has become increasingly important in a number of areas of image processing. Optical character recognition (OCR) of document images continues to be of great importance as we attempt to become a paperless society. Restoring text from video surveillance imagery is often crucial to law enforcement agencies. Digital video compression algorithms can benefit from successful text resolution expansion techniques. Video archival efforts also make use of OCR to assist with indexing and retrieval. Text observed in these types of images is often low-resolution and requires resolution expansion in order to improve OCR performance. Common interpolation methods, which were not designed specifically for text images, typically smooth over the important details and produce inadequate expansion. This

paper proposes a new nonlinear restoration technique for text images, which creates smooth foreground and background regions while preserving sharp edge transitions. There has been significant research in both the text enhancement and resolution expansion fields. Some text enhancement efforts focus on fixing broken or touching characters [18,23]. Other methods [5,24] use degraded samples to model the image and then solve for the restored image using a recursive technique. To remove the effects of noise, [11] uses a morphological filter and [1] uses pixel patterns to improve OCR results. A variety of methods have been proposed in order to improve contrast within text images including quadratic filters [13], soft morphological filters [8], non-linear mapping [16], and using a multi-resolution pyramid approach and fuzzy edge detectors [14]. Numerous resolution expansion methods have been published in the literature as well [2,3,6,7,9,12,15]. Linear interpolation tends to smooth the image data at transition regions and results in a high-resolution image that appears blurry. Cubic spline expansion allows for sharp transitions, but tends to produce a ringing effect at these discontinuities. The method proposed in [4] uses relative displacements in image sequences to improve resolution.

A number of research efforts investigated combining text enhancement with resolution expansion in order to improve low-resolution text images. Shannon interpolation is performed with text separation from the image background in [10] to improve the OCR accuracy of digital video. Deblurring is combined with linear interpolation in [19] to enhance document images obtained from digital cameras. Both of these methods essentially perform the text enhancement and resolution expansion as two separate steps. The proposed method computes resolution expansion using an algorithm specifically designed to enhance text images to perform both of these steps simultaneously.

The goal of resolution expansion is to create an expanded image with improved definition from observed low-resolution imagery. Acquisition of this low-resolution imagery can be modeled by averaging a block of pixels within a high-resolution image. Resolution expansion

sion is an ill-posed inverse problem. For a given low-resolution image, a virtually infinite set of expanded images can be generated by the observed data. To solve a high-resolution image that is optimal in some sense, a Bimodal-Smoothness-Average (BSA) for score is introduced to measure how well potential expanded images exhibit desirable text-like characteristics. The BSA score is defined as the weighted sum of separate bimodal, smoothness, and average measures. Minimization of this score is performed using a nonlinear optimization technique which results in a strongly text-like image. Text images typically have bimodal distributions with large black and white peaks, and images restored using this new method are strongly bimodal as well. Images of text are also usually smooth in both the foreground and background regions with sharp transitions only at the edges. In addition, expanded images are constrained so the average of a group of high-resolution pixels is close to the original value of the low-resolution pixel from which they were derived. Low-resolution text images restored using this new BSA score-based technique are shown to be superior, by measuring improvement in OCR accuracy, to images expanded using standard methods.

The remainder of this paper is organized as follows. Section 2 describes the problem of image resolution expansion. In Sect. 3, three text scoring functions are introduced which exploit properties of text images and set a foundation for the image restoration method proposed in this paper. Section 4 derives the technique used to iteratively solve for the functional minimum that results in the restored image. Section 5 presents experiments using this technique and quantitatively compares the proposed method to other methods of image resolution expansion. Finally, a summary of this proposed technique is given in Sect. 6.

2 Problem statement

The image acquisition process consists of converting a continuous image into discrete values obtained from a group of sensor elements. Each sensor element produces a value which is a function of the amount of light incident on the device. For 8-bit grayscale quantization, the allowable range of values for each sensor are integers from 0 (black) to 255 (white). The sensors are typically arranged in a non-overlapping grid of square elements, smaller elements result in higher resolution imagery. A high-resolution imaging system is shown in Fig. 1 where the number of sensors is adequate to represent the desired text image. The majority of pixels within the image are either white or black, with a small number of gray pixels occurring at the edges. Figure 2 illustrates a low-resolution imaging system where the number of sensors has been reduced by a factor of $q = 4$ in both the horizontal and vertical directions. This low-resolution acquisition results in significant blockiness and is insufficient to accurately represent this image. Each sensor element effectively averages the image within its section of the grid, resulting in an increased amount of gray pixels.

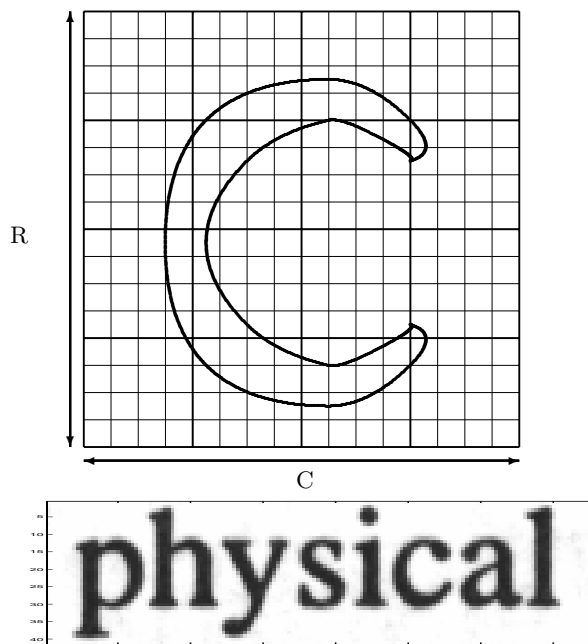


Fig. 1. High-resolution imaging system with $R \times C$ elements

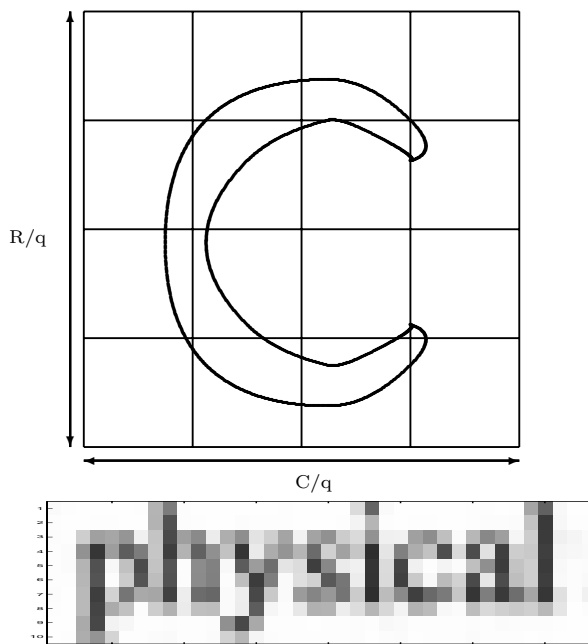


Fig. 2. Low-resolution imaging system with $R/q \times C/q$ elements

Low-resolution imaging can therefore be thought of as block-averaging high-resolution images.

The effects of this block-averaging can be further observed by examining image histograms. Figure 3 plots the histograms of an original text image and its block-averaged image. The original image's histogram is bimodal, which is common to text images. The peak occurs at background (white) values, since the majority of pixels on a text page is background. There is a secondary peak representing the black letters, which corresponds

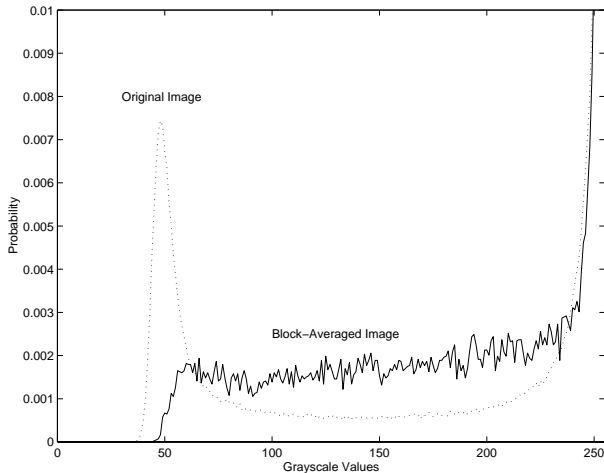


Fig. 3. Effect of block-averaging on histograms

to a grayscale value of approximately 50 for this image. Additionally, there are a small number of gray values occurring between the two peaks, which represent the gray pixels that exist at transitions from white to black. The histogram of the block-averaged low-resolution image contains a greatly reduced number of black pixels and an increased number of gray pixels caused by the pixel averaging. The problem addressed in this paper is to restore the high-resolution image $HI_{qr, qc}$ given only the low-resolution image $LI_{r, c}$, where r and c are the number of rows and columns in the low-resolution image and q is the resolution expansion factor. The image acquisition process of obtaining $LI_{r, c}$ from $HI_{qr, qc}$ is given by

$$LI_{r, c} = \frac{1}{q^2} \sum_{s=qr}^{(qr+q-1)} \sum_{t=qc}^{(qc+q-1)} HI_{s, t}. \quad (1)$$

The value of $LI_{r, c}$ is the average of the high-resolution pixels within the $q \times q$ neighborhood. Equation (1) represents a typical image restoration problem where we are required to restore the $HI_{qr, qc}$ based on the observed $LI_{r, c}$ via the relationship described by this equation. Since there are a great number of high-resolution images which may satisfy the constraint of the observed low-resolution image given by (1), image restoration is generally an ill-posed inverse problem.

3 The BSA algorithm scoring functions

The scoring function introduced in [20] is designed to measure how well a group of pixels within an image represents the desired properties of text. This function, referred to as the BSA scoring function, is expressed as the weighted sum of a bimodal score B , a smoothness score S , and an average score A , each of which will be discussed in detail in the remainder of this section. The BSA score is defined as

$$BSA(x) = \lambda_1 B(x) + \lambda_2 S(x) + \lambda_3 A(x) \quad (2)$$

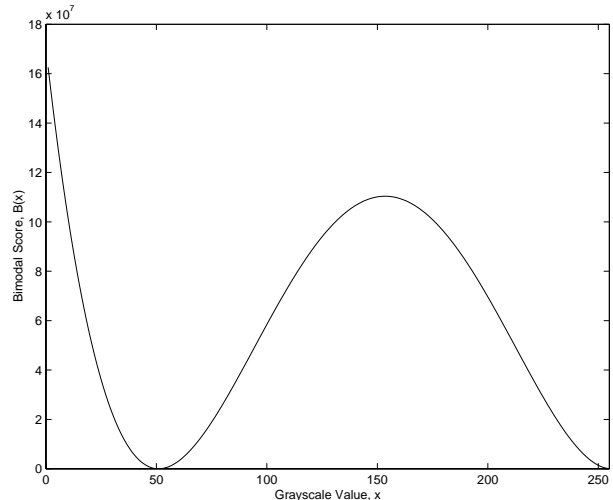


Fig. 4. Bimodal score for $\mu_{black} = 50$ and $\mu_{white} = 255$

where x is a block of pixels and λ_1 , λ_2 , and λ_3 are Lagrange multipliers. Our goal is to design a BSA-based algorithm which can iteratively solve for the block of pixels x that minimizes the $BSA(x)$ score given by (2). The BSA score is a function of the bimodal, smoothness, and average scores which are discussed in detail in the following three subsections.

3.1 The bimodal score

The typical distribution of a text image contains two peaks, a large one at μ_{white} , which normally represent the page's background, and a secondary peak at μ_{black} representing the foreground text as shown in Fig. 3. From the histogram of the given low-resolution text image, estimates of the means for the black and white distributions are calculated. These means are used to compute the bimodal score $B(x)$, which measures how far an image block x is from bimodal. The bimodal score used in this paper is defined by

$$B(x) = \sum_{r, c} (x_{r, c} - \mu_{black})^2 (x_{r, c} - \mu_{white})^2 \quad (3)$$

where r and c are the row and column indices within the block being evaluated.

The bimodal function for $\mu_{black} = 50$ and $\mu_{white} = 255$ is plotted in Fig. 4. When a pixel value within x is close to either μ_{black} or μ_{white} , its contribution to $B(x)$ is minimal. The bimodal minimum score of $B(x) = 0$ means that the image is perfectly bimodal, the value of every pixel is equal to either μ_{white} or μ_{black} . Solving for the block of pixels x that minimizes $B(x)$ produces a strongly bimodal image, which is one of the desired properties of this proposed text restoration technique. To minimize $B(x)$, both the first and second derivatives of the bimodal score will be computed. The score given by (3) is differentiable, the partial derivative of $B(x)$

with respect to pixel $x_{r,c}$ is

$$\begin{aligned} \frac{\partial B(x)}{\partial x_{r,c}} = & 4x_{r,c}^3 - 6(\mu_{white} + \mu_{black})x_{r,c}^2 \\ & + 2(\mu_{white}^2 + 4\mu_{white}\mu_{black} + \mu_{black}^2)x_{r,c} \\ & - 2\mu_{white}\mu_{black}(\mu_{white} + \mu_{black}) \end{aligned} \quad (4)$$

and the second partial derivative for the bimodal score is

$$\begin{aligned} \frac{\partial^2 B(x)}{\partial x_{r,c}^2} = & 12x_{r,c}^2 - 12(\mu_{white} + \mu_{black})x_{r,c} \\ & + 2(\mu_{white}^2 + 4\mu_{white}\mu_{black} + \mu_{black}^2). \end{aligned} \quad (5)$$

Partial derivatives of the bimodal score $B(x)$ are independent of neighboring pixels. The estimated means of the bimodal distribution, μ_{black} and μ_{white} , are known a priori and their appropriate constants can be pre-computed.

3.2 The smoothness score

With the exception of edges, text images tend to be very smooth in both the foreground and background regions which results in neighbors with similar values. A smoothness score, which is computed for each block of pixels, is introduced to measure this feature. For this proposed algorithm, a simple statistic using only the four nearest neighbors of each pixel is used. Other more sophisticated smoothness measures could be implemented as well. The smoothness score $S(x)$ used by this technique is given by

$$\begin{aligned} S(x) = \sum_{r,c} [& (x_{r-1,c} - x_{r,c})^2 + (x_{r,c-1} - x_{r,c})^2 \\ & + (x_{r,c+1} - x_{r,c})^2 + (x_{r+1,c} - x_{r,c})^2] \end{aligned} \quad (6)$$

where r and c are the row and column indices within the block being evaluated. The minimum value of $S(x) = 0$ occurs when all pixels have identical values.

The first and second partial derivatives are calculated in a straightforward manner. The first partial derivative of this smoothing score with respect to pixel $x_{r,c}$ is

$$\frac{\partial S(x)}{\partial x_{r,c}} = 8x_{r,c} - 2(x_{r-1,c} + x_{r,c-1} + x_{r,c+1} + x_{r+1,c}). \quad (7)$$

The second partial derivatives are nonzero only when

$$\begin{aligned} \frac{\partial^2 S(x)}{\partial x_{r,c}^2} = 8 \quad & \frac{\partial^2 S(x)}{\partial x_{r,c} \partial x_{r\pm 1,c}} = -2 \\ & \frac{\partial^2 S(x)}{\partial x_{r,c} \partial x_{r,c\pm 1}} = -2 \end{aligned} \quad (8)$$

and are equal to zero for all other combinations. The first partial derivative of the smoothness score $S(x)$ is a function of the four neighboring pixels. The second partial derivatives are non-zero only with respect to their corresponding neighbors.

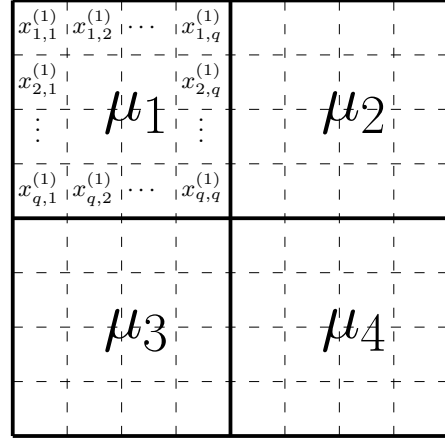


Fig. 5. Average score computation

3.3 The average constraint score

It is reasonable to require that the average of a group of high-resolution pixels is close to the original value of the low-resolution pixel from which they were derived. For each block of low-resolution pixels, an average score $A(x)$ is used to measure how well the restored high-resolution pixels meet the average constraint imposed by their corresponding low-resolution pixels. Figure 5 shows four low-resolution pixels whose values are $\mu_1, \mu_2, \mu_3,$ and μ_4 . The $q \times q$ group of high-resolution pixels that are being restored from pixel μ_1 are represented by $\{x_{r,c}^{(1)}, 1 \leq (r,c) \leq q\}$. The average score for this 2×2 block is expressed by

$$A(x) = \sum_{i=1}^4 [\mu_i - \frac{1}{q^2} \sum_{r=1}^q \sum_{c=1}^q x_{r,c}^{(i)}]^2 \quad (9)$$

where i is the index for the low-resolution pixels, μ_i is the value of each low-resolution pixel, and $x_{r,c}^{(i)}$ are the restored high-resolution pixels corresponding to pixel μ_i . The initial high-resolution image formed by using pixel replication always has an average score of zero because it satisfies the constraint.

The first partial derivative for the group of high-resolution pixels corresponding to pixel μ_i is equal to

$$\frac{\partial A(x)}{\partial x_{r,c}^{(i)}} = \frac{2}{q^2} [\frac{1}{q^2} \sum_{r=1}^q \sum_{c=1}^q x_{r,c}^{(i)} - \mu_i] \quad (10)$$

and the second partial derivatives are simply

$$\frac{\partial^2 A(x)}{\partial x_{r,c}^{(i)} \partial x_{r_1,c_1}^{(i)}} = \frac{2}{q^4} \quad \forall (r,c,r_1,c_1) \in x^{(i)}. \quad (11)$$

Both the first and second partial derivatives of the average score $A(x)$ are non-zero only with respect to pixels within the $q \times q$ group of high-resolution pixels corresponding to the low-resolution pixel from which they were expanded.

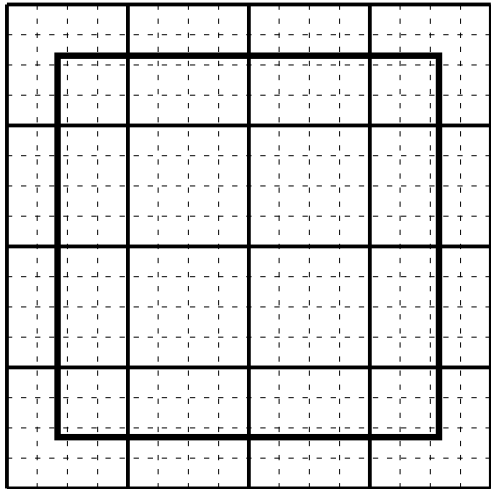


Fig. 6. Restored section of a 4×4 block of low-resolution pixels

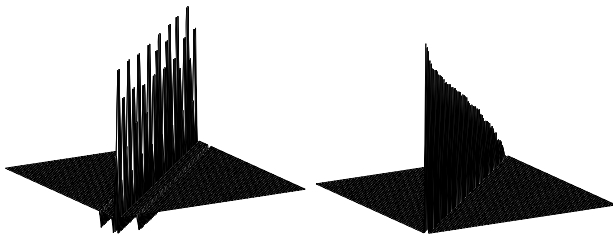


Fig. 7. Hessian matrix and diagonalized Hessian

4 Solving for the restored image

The goal of the restoration algorithm is to solve for the image block that minimizes the scoring function $BSA(x)$ introduced in 2. Throughout this paper, a block x is defined both as a group of 4×4 low-resolution pixels and as the $4q \times 4q$ high-resolution pixels that are derived from them. The 4×4 size was specifically chosen because it contains enough pixels to adequately measure text characteristics but is not too large to be computationally burdensome. The goal of resolution enhancement is to create a restored image with improved resolution. This is illustrated in Fig. 6 where a 4×4 block of low-resolution pixels is expanded by a factor of $q = 4$ to create a 16×16 block of high-resolution pixels. Pixel replication, where every value within a $q \times q$ neighborhood is identical to the corresponding low-resolution pixel, is used for the initial expansion. Each $4q \times 4q$ block of high-resolution pixels is restored independently using iterative optimization techniques described in this section to solve for the block which minimizes the BSA score. At each iteration, the first and second partial derivatives of the BSA scoring function are used to determine the image update. To avoid block boundary discontinuities only the center $3q \times 3q$ pixels, which are highlighted in Fig. 6, are updated. The entire image is therefore divided into blocks that overlap by one quarter, or $q \times 4q$ pixels, and can be restored independently. This iterative minimization of the BSA score continues until convergence is reached resulting in the restored image.

Initially, each $4q \times 4q$ block of pixels x in the expanded image is converted to a $(4q)^2$ -long vector \mathbf{x} using raster scanning,

$$\mathbf{x}[4q(r-1) + c] = x(r, c) \quad \text{for} \quad 1 \leq r, c \leq 4q. \quad (12)$$

A small distance away from \mathbf{x} the BSA function can be represented by its second order Taylor series approximation [17],

$$BSA(\mathbf{x} + \boldsymbol{\delta}) \approx BSA(\mathbf{x}) + [\nabla BSA(\mathbf{x})]\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T H \boldsymbol{\delta} \quad (13)$$

and the change in BSA is given by

$$\Delta BSA = [\nabla BSA(\mathbf{x})]\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T H \boldsymbol{\delta} \quad (14)$$

where $\boldsymbol{\delta}$ is the small change to the image vector \mathbf{x} , $\nabla BSA(\mathbf{x})$ is the gradient, and H is the Hessian matrix. The $(4q)^2 \times (4q)^2$ Hessian is the symmetric matrix of mixed partial second derivatives, which shows how a change in two variables affects the BSA function.

To maximize this function, the variables in the Hessian matrix are diagonalized via eigen-decomposition. Each eigenvector of the Hessian matrix is placed in a separate column to form a unitary eigenmatrix E . That is, the product of the eigenmatrix with its transpose is equal to the identity matrix $EE^T = I$. When the Hessian matrix is pre-multiplied by the transposed eigenmatrix and post-multiplied by the eigenmatrix, the resulting matrix $E^T H E$ is diagonal. Because the Hessian is real and symmetric, it is always diagonalizable. As an example, the Hessian matrix for an 8×8 block of pixels is shown in Fig. 7. Because of the neighborhood dependence of the scoring functions, their contributions to the Hessian are near the main diagonal. The similarity transform results in the diagonalized Hessian, $E^T H E$, which is shown at right. The Taylor series approximation to the change in the scoring function ΔBSA can now be expressed in terms of the $(4q)^2 \times (4q)^2$ Hessian matrix H , its $(4q)^2 \times (4q)^2$ eigenmatrix E , the $1 \times (4q)^2$ gradient of the scoring function $\nabla BSA(\mathbf{x})$, and the $(4q)^2 \times 1$ small change in the image vector $\boldsymbol{\delta}$,

$$\Delta BSA = ([\nabla BSA(\mathbf{x})]E)(E^T \boldsymbol{\delta}) + \frac{1}{2}(\boldsymbol{\delta}^T E)(E^T H E)(E^T \boldsymbol{\delta}). \quad (15)$$

With the following substitutions,

$$\nabla BSA'(\mathbf{x}) = [\nabla BSA(\mathbf{x})]E \quad (16)$$

$$\boldsymbol{\delta}' = E^T \boldsymbol{\delta} \quad (17)$$

$$H' = E^T H E. \quad (18)$$

Equation (15) can be simplified to

$$\Delta BSA = [\nabla BSA'(\mathbf{x})][\boldsymbol{\delta}'] + \frac{1}{2}\boldsymbol{\delta}'^T H' \boldsymbol{\delta}'. \quad (19)$$

The functional minimum is achieved by stepping in the direction

$$\boldsymbol{\delta}' = H'^{-1} \nabla BSA'(\mathbf{x}) \quad (20)$$

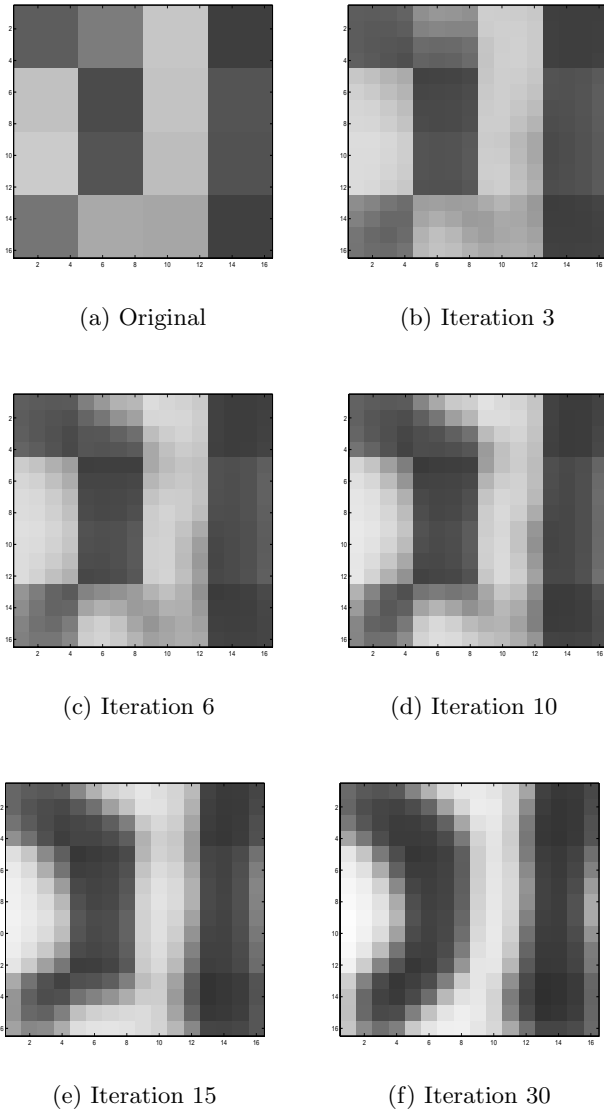
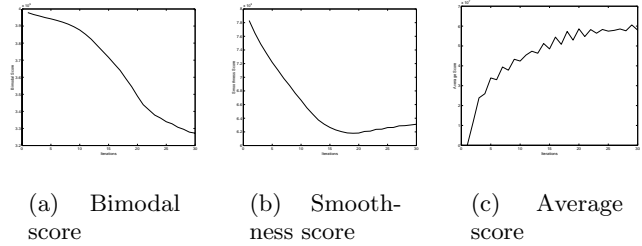
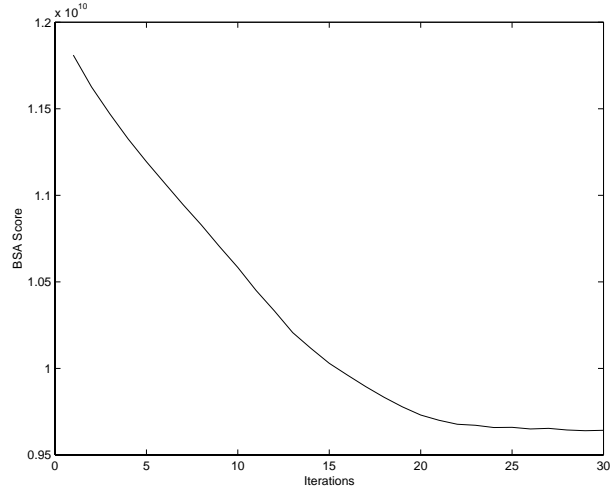


Fig. 8a–f. Iterative text restoration example

in the transformed domain, which is simply $\delta = E\delta'$ in the pixel domain. For each iteration, the image update δ' is determined. The iterations continue until convergence is reached, resulting in a desired restored image. An example of this iterative image restoration process is shown in Fig. 8. The original 4×4 block of pixels is expanded by a factor of $q = 4$ using pixel replication to produce a 16×16 high-resolution image shown in Fig. 8a. As the iterative restoration process proceeds in Figs. 8b–f, the image becomes more bimodal and smooth resulting in a greatly improved image. The majority of gray pixels that occur between characters are replaced with either black or white values, resulting in a strongly bimodal distribution. The resulting image is also smooth in both the foreground and background regions while maintaining the constraint that the average of each 4×4 block of high-resolution pixels is close to the original value of each corresponding low-resolution pixel. The minimiza-



(a) Bimodal score (b) Smoothness score (c) Average score



(d) BSA score

Fig. 9a–d. BSA score minimization

tion procedure is completed in 30 iterations for this image and the BSA score is plotted in Fig. 9d. The individual bimodal, smoothness, and average scores are shown separately in Figs. 9a–c. Initially, both the bimodal and smoothness scores are high because the image scores poorly for these measures. As the iterations proceed, the image becomes more bimodal and smooth and these two scores are reduced. The average score plotted in Fig. 9c originally has the minimum value at zero because the constraint is met exactly. The average score increases as the restoration proceeds, but this score is still significantly smaller than the bimodal and smoothness scores. Minimization of the BSA score produces a restored image that is the optimal combination of these bimodal, smoothness, and average measures.

5 Experimental results

The proposed BSA restoration algorithm was compared to several common expansion methods, including pixel replication, linear interpolation, and cubic spline expansion. In linear interpolation, a linear fit is calculated between all pixels within each column, and then repeated for all pixels within each row. These images naturally tend to be smooth, without sharp discontinuities, producing blurry results. Cubic spline expansion [3] approximates the given discrete low-resolution pixels as a smooth

continuous curve obtained from the weighted sum of cubic spline basis functions and resamples the curve to obtain the high resolution image. This method allows for sharp edges but often overshoots at these discontinuities, producing a ringing effect. The BSA text restoration technique creates smooth foreground and background regions and permits sharp edges at transition regions, while maintaining the low-resolution average constraint. Images restored with this technique are shown to be both qualitatively and quantitatively superior to other common resolution expansion methods. Two experiments were conducted to numerically compare the restoration methods. The first experiment involved creating low-resolution images from high-resolution originals, expanding the low-resolution imagery, and then measuring the distance to the originals. The second experiment involved scanning low-resolution document images, expanding the images with the various techniques, and using OCR accuracy to measure the success of the restoration.

The values for the three Lagrange multipliers, which are weights in the BSA scoring function,

$$BSA(x) = \lambda_1 B(x) + \lambda_2 S(x) + \lambda_3 A(x)$$

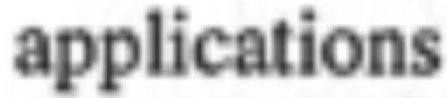
were determined experimentally. The rough order of magnitude for the bimodal score given in (3) is $q^2 d^4$ where q is the expansion factor and d is the difference between two pixels. The smoothness score defined by (6) is of the order $4q^2 d^2$ and the average score defined by (9) is approximately of the order $4d^2$. Throughout these experiments the relative weights used were $\lambda_1 = 1$, $\lambda_2 = 10,000$, and $\lambda_3 = 1,000,000$. The restored image is solved for iteratively until convergence is achieved. For the experiments described in this section, convergence occurs when the average pixel change within a $4q \times 4q$ block of pixels is less than 0.5.

To qualitatively illustrate the differences between these resolution expansion techniques, Fig. 10 shows resulting images obtained from linear interpolation, cubic spline expansion, and the proposed BSA-based restoration technique. The word “applications” from an image scanned at 100 dots per inch (dpi) using 8-bit grayscale quantization is shown in Fig. 10a where significant blockiness is apparent. Linear interpolation by a factor of four was used to create the image in Fig. 10b which is very blurry and lacks good contrast. Figure 10c depicts the resulting image from cubic spline expansion which has better contrast but is still not sharp at the edges. The image obtained using BSA restoration in Fig. 10d has excellent contrast and sharp edges and is superior to the images obtained using other interpolation methods for this example.

The first experiment to quantitatively measure image restoration success involved creating low-resolution images by block-averaging images as described by (1). Restored images are then compared with the original to determine the success of restoration numerically. The mean squared error (MSE) was used to compare the various methods of image resolution expansion. The defini-



(a) Original Image



(b) Linear Interpolated Image



(c) Cubic Spline Interpolated Image



(d) BSA Restored Image

Fig. 10a–d. Text restoration results for low-resolution original

tion of mean squared error used in this paper is

$$MSE = \frac{1}{RC} \sum_{r=1}^R \sum_{c=1}^C (original_{r,c} - restored_{r,c})^2 \quad (21)$$

where R and C are the number of rows and columns in the images. Restoration results for a severely degraded 41×376 section of a 3300×2544 image scanned at 300 dpi are shown in Fig. 11. An averaging factor of $q = 4$ was used to create a blocky image with a significant amount of touching characters shown in Fig. 11b. The mean squared error between this blocky section and the original is 1129.1. Linear interpolation produces the severely blurred image in Fig. 11c which reduces the MSE only slightly by 14.5% to 965.4. The resulting image obtained from cubic spline interpolation in Fig. 11d is significantly improved with a 40.9% reduction in MSE to 667.3, but is still somewhat blurry. The BSA restoration produced the best image shown in Fig. 11e by reducing the MSE by 60.2% to 449.4.

These results clearly demonstrate several advantages of this technique that was designed specifically for text images. The BSA-restored image in Fig. 11e is strongly bimodal and has both smooth foreground and background regions. There are sharp discontinuities at the edges which are not observed in the linear or cubic spline expansion results. The significant reduction in the

solutions that conform

(a) Original Image

solutions that conform

(b) Block-Averaged Image

solutions that conform

(c) Linear Interpolated Image

solutions that conform

(d) Cubic Spline Interpolated Image

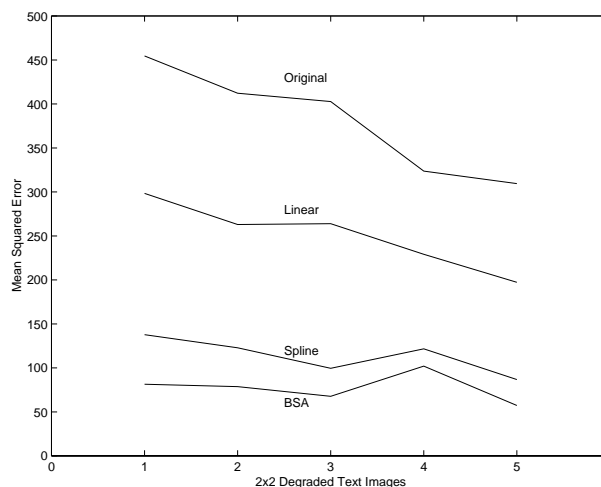
solutions that conform

(e) BSA Restored Image

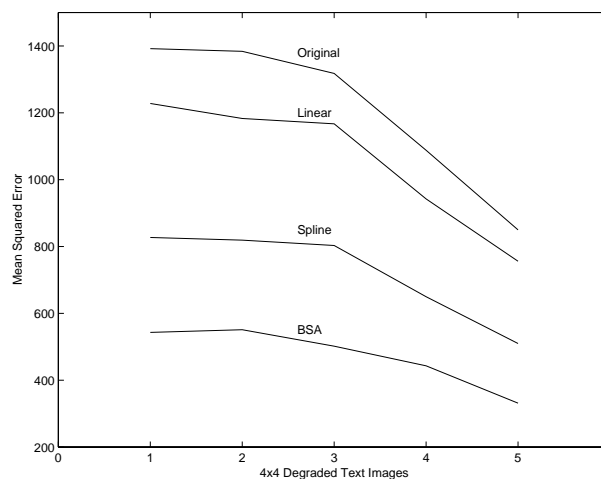
Fig. 11a–e. Text restoration results for 4×4 degradation

amount of gray pixels produces superior character separation evident in the word “conform”. This gray pixel decrease also frequently results in sharper contrast within a single character. The hole in the character “a” in the word “that” which is hardly apparent in the block-averaged image is vastly improved by the algorithm. A comparative study of the reduction in mean squared error for the various image expansion techniques is plotted in Fig. 12. The results for a group of five full-page images that were scanned at 300 dpi and degraded with block-averaging factor $q = 2$ are shown in Fig. 12a. The average MSE for these five block-averaged images was 380.5. Linear interpolation reduced the average MSE to 250.3 for these images for a 34.4% improvement. Cubic spline interpolation reduced the average MSE to 113.7 for an improvement of 70.3%. The proposed BSA restoration technique was the most accurate for these images and resulted in an average MSE of 77.4 for an 80.9% improvement. Figure 12b shows the restoration results for a second group of five images again scanned at 300 dpi and severely degraded by $q = 4$ block-averaging for an average MSE of 1206. Expansion using linear interpolation reduced the MSE to 1055 for a 12.4% reduction and cubic spline expansion resulted in an MSE of 722 for a 40.2% reduction. The best results were obtained using the BSA algorithm which reduced the MSE to 474 for an average reduction 60.7%.

As a final experiment, we used OCR accuracy to numerically compare our algorithm against existing reso-



(a) 2x2 degradation



(b) 4x4 degradation

Fig. 12a,b. Comparison of MSE reduction for various restoration techniques

lution expansion methods. A set of 122 full-page journal documents from the University of Washington English Document Image Database I CD-ROM [22] were used. These binary documents were initially printed by a laser printer with 300 dpi resolution. Each page was then scanned using 8-bit grayscale quantization at 75 dpi to create low-resolution original images. These 75 dpi resolution pages were then expanded using various resolution expansion methods by a factor of four to create 300 dpi images which were processed by Caere’s OmniPage Pro 7.0 commercial OCR package, the world’s best-selling desktop OCR software. For all cases within these experiments, both spline interpolation and BSA restoration produced superior results to linear interpolation, therefore the linear interpolation results will not be discussed further. The resulting text files were com-

The Preisach Model for ferromagnets is generalized and adapted for the description of the hysteretic behaviour of a polycrystalline specimen of shape-memory alloys. The thermodynamical properties of the individual crystallites are described by the Landau-Devonshire free energy which contains four parameters. The corresponding quadruplets of parameters of a polycrystalline body fill a region in a four-dimensional Preisach space. A thermodynamical loading path will sweep surfaces across this region and change phases in the process. The physical problem of the response of a specimen to applied loads is thus converted into the geometrical problem of counting volumes between moving surfaces. This conversion facilitates the numerical evaluation of the effect of complicated loading paths.

(a) Spline Image

The Preisach Model for ferromagnets is generalized and adapted for the description of the hysteretic behaviour of a polycrystalline specimen of shape memory alloys. The thermodynamical properties of the individual crystallites are described by the Landau-Devonshire free energy which contains four parameters. The corresponding quadruplets of parameters of a polycrystalline body fill a region in a four-dimensional Preisach space. A thermodynamical loading path will sweep surfaces across this region and change phases in the process. The physical problem of the response of a specimen to applied loads is thus converted into the geometrical problem of counting volumes between moving surfaces. This conversion facilitates the numerical evaluation of the effect of complicated loading paths.

(b) Spline OCR Text

The Preisach Model for ferromagnets is generalized and adapted for the description of the hysteretic behaviour of a polycrystalline specimen of shape-memory alloys. The thermodynamical properties of the individual crystallites are described by the Landau-Devonshire free energy which contains four parameters. The corresponding quadruplets of parameters of a polycrystalline body fill a region in a four-dimensional Preisach space. A thermodynamical loading path will sweep surfaces across this region and change phases in the process. The physical problem of the response of a specimen to applied loads is thus converted into the geometrical problem of counting volumes between moving surfaces. This conversion facilitates the numerical evaluation of the effect of complicated loading paths.

(c) BSA Image

The Preisach Model for ferromagnets is generalized and adapted for the description of the hysteretic behaviour of a polycrystalline specimen of shape memory alloys. The thermodynamical properties of the individual crystallites are described by the Landau-Devonshire free energy which contains four parameters. The corresponding quadruplets of parameters of a polycrystalline body fill a region in a four-dimensional Preisach space. A thermodynamical loading path will sweep surfaces across this region and change phases in the process. The physical problem of the response of a specimen to applied loads is thus converted into the geometrical problem of counting volumes between moving surfaces. This conversion facilitates the numerical evaluation of the effect of complicated loading paths.

(d) BSA OCR Text

Fig. 13a–d. Sample OCR results for spline vs BSA

pared to the ground truth provided by the University of Washington CD-ROM using the OCR Accuracy Report Version 5.3 software [21] developed at UNLV-ISRI. There were a total of 339,575 characters in these 122 images. Cubic spline interpolation resulted in 36,959 character errors and the BSA-based method had 30,668 character errors for an overall improvement of 17.0% for this set of images.

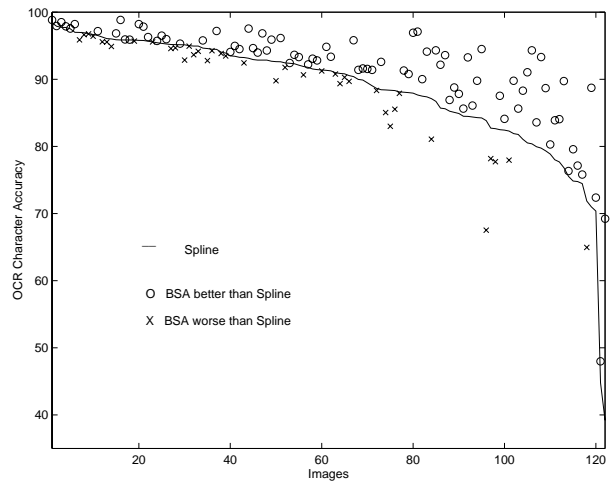


Fig. 14. OCR accuracy – spline vs BSA

A sample section of two restored images and their corresponding OCR results are shown in Fig. 13. Figure 13a shows the resulting high-resolution image obtained using cubic spline expansion. The text file created by OCR is shown in Fig. 13b where mistakes have been highlighted. For this example, the OCR results for the cubic spline image had 7 areas where mistakes were made. The same image paragraph was expanded using the proposed BSA method resulting in the image in Fig. 13c which has better contrast than the corresponding cubic spline image. OCR results are improved as well, only 4 mistake areas are highlighted in Fig. 13d. An example where cubic spline expansion produced better OCR results is in the word “Preisach.” The top and bottom of the “s” character were darker than the middle section in the original image. The BSA algorithm caused the dark top and bottom regions of the “s” to be separated and resulted in an “i” for this example.

OCR character accuracy for all 122 document images is plotted in Fig. 14 for cubic spline interpolation and BSA expansion. The original 75 dpi images were of such low resolution that the commercial OCR package, which performs optimally on 300 dpi images, frequently failed to recognize any characters within the low-resolution document images. The results are sorted based on the spline OCR results which are shown as a solid line. For each image, the BSA result is plotted as an “X” if the OCR accuracy is worse than spline and plotted as an “O” where the accuracy has been improved. The BSA restoration resulted in improved OCR accuracy for 72% of the images in this test set. Even in the cases where cubic spline resolution expansion improved OCR more than the BSA algorithm, the images produced by the BSA were typically more visually appealing. Cubic spline expansion from 75 dpi originals to 300 dpi images required approximately 10 s of CPU time for each document. The same expansion required about 5 min for BSA restoration. For the images in this experiment, the expansions produced by the BSA algorithm produce superior OCR accuracy results compared to other existing methods.

6 Conclusions

In this paper, we present a new resolution expansion technique for the restoration of low-resolution grayscale text images. Bimodal, smoothness, and average (BSA) scores that measure desired properties observed in text images were introduced and combined to form a single scoring function. The restored image obtained by solving this nonlinear optimization problem is one which is strongly bimodal and smooth, while satisfying the average constraint score. The proposed BSA restoration technique was shown to be both qualitatively and quantitatively superior to the existing linear interpolation and cubic spline expansion techniques.

Although this technique was specifically designed for text images, it has also proven effective at restoring low-resolution fingerprint images. Text images have much in common with fingerprint images. Both have strongly bimodal histograms, are smooth in both the foreground and background regions, and have sharp edge transitions. With no modifications, the BSA algorithm was able to successfully restore fingerprint images. In addition to fingerprints, it is expected that the proposed algorithm will accurately expand tables, charts, graphs, and other line drawings.

In order to further improve the proposed BSA restoration technique, new measures may be added that make use of a priori knowledge. If the scanning resolution and font size within an image are known, the average stroke width in pixels can be computed. A new score could be created to measure how close a group of pixels is to this desired width. Additionally, different measures for the existing bimodal, smoothness, and average scores could be implemented. Another research area under exploration is optimally selecting the weights for the three scoring functions, by scanning the same document at multiple resolutions and restoring the low-resolution and comparing it to the corresponding scanned high-resolution image, which could potentially be of significant benefit. Additional research is also being considered to determine how to apply this method to images with non-uniform background such as video frames.

Acknowledgements. The first author wishes to thank G. H. Nickel of Los Alamos National Laboratory and C. S. Cumbee of the Department of Defense for technical discussions, and S. J. Dennis of the Department of Defense for his support of this research.

References

1. T. Akiyama, N. Miyamoto, M. Oguro, K. Ogura.: Faxed document image restoration method based on local pixel patterns. Proc. SPIE, 3305: 253–262, 1998
2. T. C. Chen, R. J. P. de Figueiredo.: Image decimation and interpolation techniques based on frequency domain analysis. IEEE Trans. on Commun., 32(4), 1984
3. H. S. Hou, H. C. Andrews.: Cubic splines for image interpolation and digital filtering. IEEE Trans. on Acoust., Speech, and Signal Process., 26(6), December 1978
4. M. Irani, S. Peleg.: Improving resolution by image registration. CVGIP: Graphical Models and Image Process., 53(3): 231–239, 1991
5. M. Y. Jaisimha, E. A. Riskin, R. Ladner., S. Werner.: Model-based restoration of document images for OCR. Proc. SPIE, 2660: 297–308, 1996
6. N. B. Karayiannis, A. N. Venetsanopoulos.: Image interpolation based on variational principles. Signal Process., 25: 259–288, 1991
7. R. G. Keys.: Cubic convolution interpolation for digital image processing. IEEE Trans. on Acoust., Speech, and Signal Process., 29(6): 1153–1160, 1981
8. L. Koskinen, H. Huttunen., J. T. Astola.: Text enhancement method based on soft morphological filters. Proc. SPIE, 2181: 243–253, 1994
9. A. D. Kulkarni, K. Sivaraman.: Interpolation of digital imagery using hyperspace approximation. Signal Process., 7: 65–73, 1987
10. H. Li, O. E. Kia., D. S. Doermann.: Text enhancement in digital video. Proc. SPIE, 3651: 2–9, 1999
11. J. Liang, R. M. Haralick., I. T. Phillips.: Document image restoration using binary morphological filters. Proc. SPIE, 2660: 274–285, 1996
12. V. S. Nalwa.: Edge-detector resolution improvement by image interpolation. IEEE Trans. Pattern Anal. Mach. Intell., 9(3): 446–451, 1987
13. G. Ramponi, P. Fontanot.: Enhancing document images with a quadratic filter. Signal Process., 33: 23–34, 1993
14. F. Sattar, D. B. H. Tay.: On the multiresolution enhancement of document images using fuzzy logic approach. ICPR98 Proc. Int. Conf. on Pattern Recognition, pp. 939–942, 1998
15. R. R. Schultz, R. L. Stevenson.: A Bayesian approach to image expansion for improved definition. IEEE Trans. on Image Processing, 3(3): 233–242, 1994
16. Y.C. Shin, R. Sridhar, V. Demjanenko, P. W. Palumbo., J. J. Hull.: Contrast enhancement of mail piece images. Proc. SPIE, 1661: 27–37, 1992
17. J. Skilling, R. K. Bryan.: Maximum entropy image reconstruction: general algorithm. Mon. Not. R. Astron. Soc., 211: 111–124, 1984
18. P. Stubberud, J. Kanai., V. Kalluri.: Adaptive image restoration of text images that contain touching or broken characters. ICDAR95 Proc. Int. Conf. Doc. Anal. Recognition, pp. 778–781, 1995
19. M. J. Taylor, C. R. Dance.: Enhancement of document images from cameras. Proc. SPIE, 3305: 230–241, 1998
20. P. D. Thouin, C. -I Chang.: A Method for Restoration of Low-Resolution Text Images. Symp. Doc. Image Understanding, pp. 143–148, 1999
21. OCR Accuracy Report Version 5.3.: University of Nevada, Las Vegas, NV, UNLV-ISRI

22. UW English Document Image Database I, Volume 1, Binary Images.: University of Washington, Seattle, WA, 1993
23. A. P. Whichello, H. Yan.: Linking broken character borders with variable sized masks to improve recognition. *Pattern Recognition*, 29(8): 1429–1435, 1996
24. M.Y. Yoon, S.W. J.S. Kim.: Faxed image restoration using Kalman filtering. *ICDAR95 Proc. Int. Conf. Doc. Anal. Recognition*, pp. 677–681, 1995



Paul D. Thouin received his B.S. degree in electrical engineering from the University of Michigan, Ann Arbor in 1987. In 1993, he obtained the M.S.E.E. degree from George Washington University in Washington, D.C. He is currently a Ph.D. candidate at the University of Maryland Baltimore County majoring in electrical engineering. Mr. Thouin has been employed by the U.S. Department of Defense since 1987 where he is a Senior Engineer currently

assigned to the Image Research Branch. His research interests include image enhancement, statistical modeling, document analysis, and pattern recognition. Mr. Thouin is a member of SPIE and Phi Kappa Phi.



Chein-I Chang received his B.S., M.S. and M.A. degrees from Soochow University, Taipei, Taiwan, 1973, the Institute of Mathematics at National Tsing Hua University, Hsinchu, Taiwan, 1975 and the State University of New York at Stony Brook, 1977, respectively, all in mathematics, and M.S., M.S.E.E. degrees from the University of Illinois at Urbana-Champaign in 1982 respectively and Ph.D. in electrical engineering from the University of Maryland,

College Park in 1987. He was a visiting assistant professor from January 1987 to August 1987, assistant professor from 1987 to 1993, and is currently an associate professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County. He was a visiting specialist in the Institute of Information Engineering at the National Cheng Kung University, Tainan, Taiwan from 1994-1995. He is an editor for *Journal of High Speed Network* and the guest editor of a special issue on *Telemedicine and Applications*. His research interests include automatic target recognition, multispectral/hyperspectral image processing, medical imaging, information theory and coding, signal detection and estimation, and neural networks. Dr. Chang is a senior member of IEEE and a member of SPIE, INNS, Phi Kappa Phi and Eta Kappa Nu.