

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Automated system for restoration of low-resolution document and text images

Paul D. Thouin

Department of Defense
Fort Meade, Maryland 20755

Chein-I Chang

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

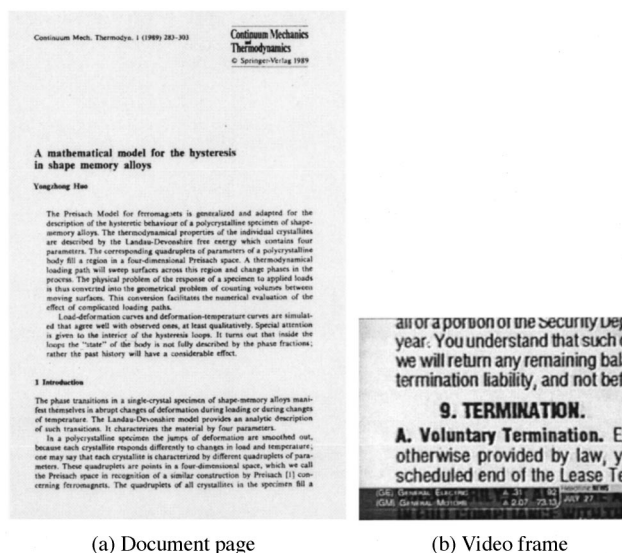
Abstract. Restoration of document and text images has become increasingly important in many areas of electronic imaging. This paper presents an automated system to restore low-resolution document and text images. It makes use of resolution expansion to enhance low-resolution images for optical character recognition accuracy as well as to improve the quality of degraded images. Several approaches have been proposed in the past for resolution expansion such as linear interpolation and cubic spline expansion. The proposed system implements a bimodal-smooth-average (BSA) scoring function as an optimal criterion for image quality. The BSA approach is very different from existing methods in the sense that it uses three measures: bimodal, smooth, and average to produce a scoring function as an optimal criterion. Its idea is to create for a given image a strongly bimodal image with smooth regions in both the foreground and background, while allowing for sharp discontinuities at the edges. Then the desired resolution-expanded image is obtained by solving a nonlinear optimization problem subject to a constraint that the average of expanded resolution must be equal to the original unexpanded resolution. The system can be used to restore both binary and grayscale images as well as video frames. Its capability is demonstrated experimentally to be quantitatively and qualitatively superior to standard interpolation methods. © 2001 SPIE and IS&T. [DOI: 10.1117/1.1351822]

1 Introduction

Enhancement of text images continues to be an important research area in both the document and video recognition fields. Automatic conversion from image to text using optical character recognition (OCR) is widely used to process documents. Video archiving efforts also make use of OCR to assist with indexing and retrieval. The state of the art in OCR algorithms continues to improve, with current systems achieving a 99% character accuracy rate on machine-printed laser quality documents with standard fonts,^{1,2} but do not perform as well on video frames or low-resolution document images.

This paper describes an automated system to enhance low-resolution document and video text images using resolution expansion so as to improve OCR accuracy. Significant research has been performed in both the text enhancement and resolution expansion fields.³ Some methods^{4,5} use degraded samples to model the image and then solve for the restored image using a recursive technique. Other efforts focus on fixing broken or touching characters^{6,7} within the text image. A morphological filter is implemented in Ref. 8 to remove the effects of noise. A number of approaches have been proposed to improve contrast within text images including nonlinear mapping,⁹ quadratic filters,¹⁰ and soft morphological filters.¹¹ Numerous resolution expansion methods have been published in the literature as well including cubic convolution,¹² frequency domain interpolation,¹³ interpolation based on variational principles,¹⁴ and hyperspace approximation.¹⁵ Unfortunately, none of these enhancement techniques was designed specifically for text images. Two of the most frequently used generic resolution expansion techniques are linear interpolation and cubic spline expansion. 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 ringing effects at these discontinuities.¹⁶ This paper describes a system specifically developed to restore text in both document and video images that overcomes these limitations and produces superior images with expanded resolution.

Our system uses resolution expansion to create an expanded image with improved resolution from an observed low-resolution image. Acquisition of this low-resolution image can be modeled by averaging a block of pixels within a high-resolution image. So, resolution expansion 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 for a high-resolution image that is optimal in some sense, the



(a) Document page

(b) Video frame

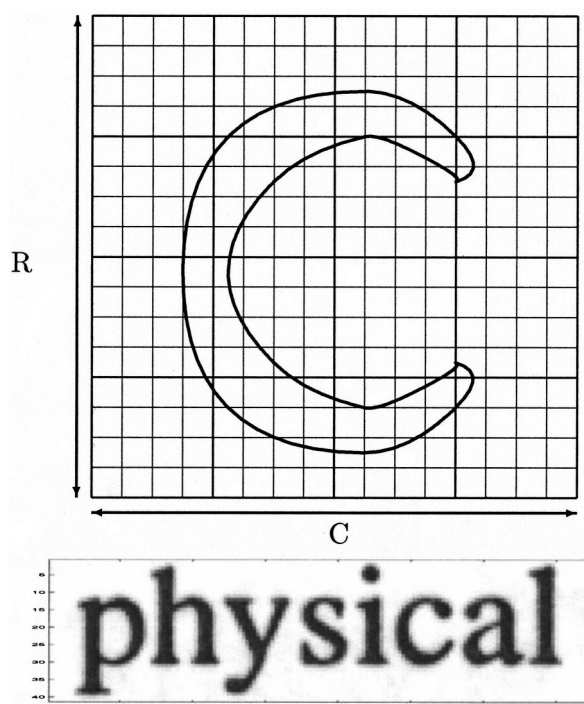
Fig. 1 Sample low-resolution images.

bimodal-smooth-average (BSA) score introduced in Ref. 17 is used 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. Minimizing this score results in a strongly text-like image. Due to the fact that text images typically have bimodal distributions with large black and white peaks, images restored using this new method are strongly bimodal as well. Since images of text are also usually smooth in both the foreground and background regions with sharp transitions only at the edges, the image restored by the BSA method also has the same properties. In addition, expanded images are constrained so 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. As a consequence, text images restored by our system using this BSA score as an optimal criterion are shown to be both quantitatively and qualitatively superior to images expanded using standard methods.

The remainder of this paper is organized as follows. Section 2 states the problem of resolution expansion from low-resolution document and video images, then describes our overall system architecture. In Sec. 3, the grayscale restoration process used to enhance text images is given in detail. The restoration system for binary document images is further proposed in Sec. 4. Section 5 presents experimental results and a quantitative study of comparing our system to other methods of image resolution expansion that have been used in the past. Finally, our text image resolution expansion system is summarized in Sec. 6.

2 System Architecture for Resolution Expansion

There are a wide variety of low-resolution text images, including both documents and video frames, that require resolution expansion. Two samples are displayed in Fig. 1. A binary document image is shown in Fig. 1(a) where every pixel is either black or white. Processing of document images has traditionally been performed using this binary format due to the savings in disk storage and computer


Fig. 2 High-resolution imaging system with $R \times C$ elements.

processing. As computer performance and storage capacity have dramatically increased recently, grayscale scanning and processing of document images have become more prevalent. Text images traditionally have been obtained from documents. However due to the current increase in digital video a significant amount of text is found in this type of imagery as well. A typical single video frame of a text image is shown in Fig. 1(b). The text restoration system developed in this paper is capable of enhancing a variety of text images including both the low-resolution video frame as well as the document image shown in Fig. 1.

Document images are typically acquired by a scanner, whereas video frames are most often captured by a digital camera. The acquisition process in both cases consists of converting an image that is continuous in both the spatial domain and intensity domain into discrete values. A scanner performs sampling in the spatial domain via sensor elements and quantization of the intensity domain from incident light. The sensors are typically arranged in a nonoverlapping grid of square elements; smaller elements result in higher resolution imagery. For 8-bit grayscale quantization, the allowable range of values for each sensor are integers from 0 (black) to 255 (white). Shown in Fig. 2 is a high-resolution imaging system 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 3 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 an image.

Our text resolution expansion system attempts to restore a high-resolution image $HI_{qr,qc}$ given only a low-resolution

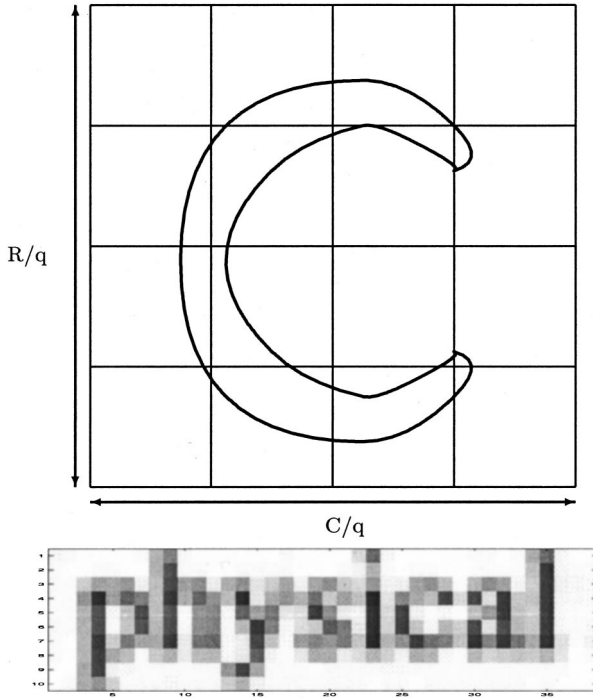


Fig. 3 Low-resolution imaging system with $R/q \times C/q$ elements.

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 carried out by averaging q^2 pixels in $HI_{qr,qc}$

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

For the decimation model in Eq. (1), the constraint value $LI_{r,c}$ is the average value of the high-resolution image pixels located within the corresponding $q \times q$ neighborhood. This equation 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. Low-resolution images typically are blocky and do not possess the strong bimodal distribution observed in high-resolution text images. The goal of this research is to create a high-resolution output from the given low-resolution input image using resolution expansion. Image resolution expansion is an ill-posed inverse problem because a nearly infinite set of high-resolution images may exist that satisfy the given constraint from the original data.

The system presented in this paper is the result of an effort to automatically enhance text within document and video images prior to performing automatic character recognition. Its block diagram is shown in Fig. 4 and comprises five modules, the preprocessing module, the bimodal estimation module, the resolution expansion module, the BSA restoration module, and the postprocessing module. A low-resolution image is initially input to the preprocessing module where the intensities of binary and color images are converted to low-resolution grayscale images. The image

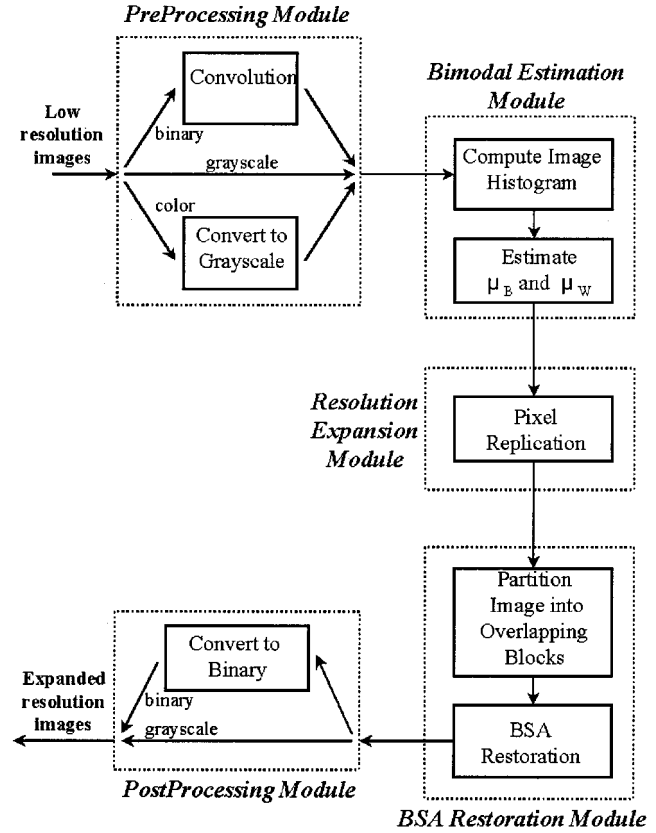


Fig. 4 Block diagram of the text resolution expansion system.

histogram is computed by the bimodal estimation module to estimate the means of the black and white pixel distributions. The initial spatial domain image expansion is performed by the resolution expansion module which uses pixel replication to create a high-resolution image from the low-resolution original. This high-resolution image is divided into overlapping blocks of pixels whose intensities are restored independently by the BSA algorithm in the BSA restoration module. When a binary expanded image is desired, the postprocessing module performs a grayscale-to-binary conversion. Section 3 details the image expansion system for grayscale images, followed by a section on binary image expansion.

3 Grayscale Image Resolution Expansion

The image resolution expansion process for grayscale images, which includes both video frames and grayscale document images, is described in this section. It should be noted that our system depicted in Fig. 4 processes grayscale and color images in a very similar fashion. The only difference is in the preprocessing module where color images are initially converted to grayscale images by keeping only the luminance component Y from the RGB color model¹⁸

$$Y = 0.299R + 0.587G + 0.114B. \quad (2)$$

These formerly color images are processed throughout the system as grayscale and the resulting high-resolution expanded images are grayscale as well. Since the goal of our system is to improve automated OCR results, this loss of

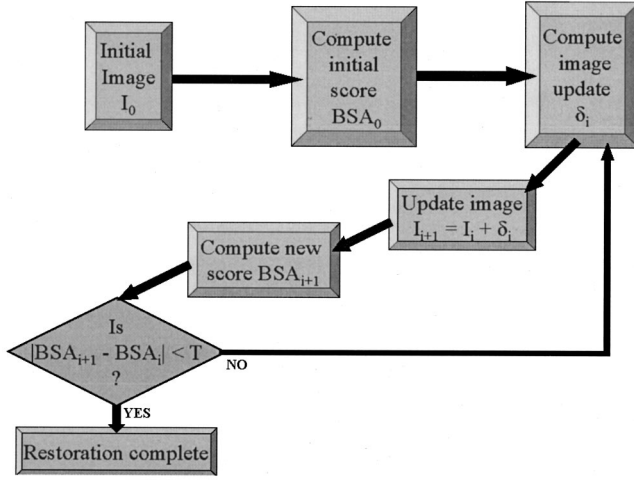


Fig. 5 BSA restoration algorithm.

color information is not a concern. Once the system has a grayscale image, its histogram is then computed by the bimodal estimation module. The distribution of a text image is typically bimodal with a large white peak corresponding to the background and a smaller dark peak corresponding to the text pixels. The peak of the white distribution μ_{white} and the peak of the black distribution μ_{black} are estimated from the calculated histogram. These values are required later by the system. Pixel replication, where the value of multiple high-resolution pixels are equal to its corresponding low-resolution pixel, is performed in the resolution expansion module to form the initial expanded image. The BSA restoration module partitions the image into overlapping blocks and restores the blocks independently. Finally, the postprocessing module converts the high-resolution grayscale image to a binary image if a binary expanded image is desirable. In what follows, the BSA restoration module will be described in detail.

A block diagram of the BSA restoration process is shown in Fig. 5 where I_i represents the image at iteration i , BSA_i is the score at iteration i , and δ_i is the change to the image at iteration i . The restoration process can be summarized as follows. The initial score BSA_0 is computed from the expanded original image I_0 . At each iteration, the image update δ_i and new score BSA_i are computed. The iterations continue to minimize the BSA score until a convergence is reached. At this stage, a restored image is produced.

The scoring function used by the BSA restoration process is designed to measure how well a group of pixels within an image represent the desired properties of a text image. 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 . More precisely, the BSA score is defined by

$$BSA(x) = \lambda_B B(x) + \lambda_S S(x) + \lambda_A A(x), \quad (3)$$

where λ_B , λ_S , and λ_A are Lagrange multipliers and x is a block of pixels to be restored. Throughout this paper, a block x is defined both as a group of 4×4 low-resolution

pixels or 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 the restoration process is to iteratively solve for the block of pixels x that minimizes the $BSA(x)$ score given by Eq. (3). Each of the three scores will be briefly discussed in the following subsections followed by a description of an iterative minimization procedure.

3.1 The Bimodal Score

The typical distribution of a grayscale text image contains two peaks: a large one at μ_{white} , which normally represents the text page's background, and a secondary peak at μ_{black} representing the foreground text. From the histogram of the given low-resolution text image, estimates of the means for the black and white distributions are calculated. The larger of the bimodal peaks is equal to the maximum value of the histogram and a weighted sum of the center band is used to determine the value of the secondary peak. These means are then used to compute the bimodal score $B(x)$, which measures how far an image block x is from a bimodal. The bimodal score is defined by

$$B(x) = \sum_{r,c} (x_{r,c} - \mu_{\text{black}})^2 (x_{r,c} - \mu_{\text{white}})^2, \quad (4)$$

where $x_{r,c}$ is the pixel value at row r and column c within block x . Solving for the block of pixels x that minimizes $B(x)$ produces a strongly bimodal image, that is one of the desired properties of the BSA text restoration technique.

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 result in neighbors with similar values. A smoothness score computed for each block of pixels is used to measure this feature. The proposed smoothness score $S(x)$ is given by

$$S(x) = \sum_{r,c} [(x_{r-1,c} - x_{r,c})^2 + (x_{r,c-1} - x_{r,c})^2 + (x_{r,c+1} - x_{r,c})^2 + (x_{r+1,c} - x_{r,c})^2], \quad (5)$$

where r and c are the row and column indices within the block being evaluated. At block edges, the appropriate pixels from neighboring blocks are used in this calculation in order to improve smoothness between adjacent blocks. The minimum value of $S(x) = 0$ occurs when all pixels have identical values.

3.3 The Average Constraint Score

Upon performing image resolution expansion, 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 pix-

els meet the average constraint imposed by their corresponding low-resolution pixels. The average score for a $4q \times 4q$ block is expressed by

$$A(x) = \sum_{i=1}^{16} \left[\frac{(1/q^2) \sum_{r=1}^q \sum_{c=1}^q (x_{r,c}^{(i)} - \mu_i)^2}{128} \right], \quad (6)$$

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 .

3.4 Iterative BSA Minimization

The goal of the restoration algorithm is to solve for an image block that minimizes the scoring function $BSA(x)$ introduced in Eq. (3). The values for the three Lagrange multipliers, which are weights in the BSA scoring function, were determined experimentally. The rough order of magnitude for the bimodal score given in Eq. (4) is $q^2 d^4$, where q is the expansion factor and d is the difference between two pixels. The smoothness score defined by Eq. (5) is of the order $4q^2 d^2$ and the average score defined by Eq. (6) is approximately of the order $4d^2$. Throughout these experiments the relative weights used were $\lambda_B = 1$, $\lambda_S = 10\,000$, and $\lambda_A = 1\,000\,000$. 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.

Pixel replication, where every value within a $q \times q$ neighborhood is identical to the corresponding low-resolution pixel, is used for an initial expansion. Each $4q \times 4q$ block of high-resolution pixels is restored independently using the iterative optimization technique described by Eq. (10) to solve for the block that minimizes the BSA score. At each iteration, the gradient of the BSA scoring function is used to determine an image update. To avoid boundary discontinuities at block edges only the center $3q \times 3q$ pixels are updated. The entire image is therefore divided into 4×4 blocks that are overlapped by one quarter in both the horizontal and vertical directions. Each block can be restored independently. The process of iteratively minimizing the BSA score continues until a convergence is reached where a restored image is produced.

A mathematical description of this process is given as follows. Initially, each $4q \times 4q$ block of pixels x is converted to a $(4q)^2$ -long vector \mathbf{x} using raster scanning

$$\mathbf{x}[4q(r-1)+c] = x(r,c) \quad \text{for } 1 \leq r, c \leq 4q. \quad (7)$$

The BSA function can be represented by its Taylor series approximation¹⁹ a small distance away from \mathbf{x}

$$BSA(\mathbf{x} + \boldsymbol{\delta}) \approx BSA(\mathbf{x}) + [\nabla BSA(\mathbf{x})] \cdot \boldsymbol{\delta}, \quad (8)$$

where $\boldsymbol{\delta}$ is the small change to the image vector \mathbf{x} and $\nabla BSA(\mathbf{x})$ is the gradient. The image change at iteration i is computed using

$$\boldsymbol{\delta}_i = -\nabla BSA(\mathbf{x}_i). \quad (9)$$

The image is then updated using

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \boldsymbol{\delta}_i. \quad (10)$$

Overcorrection is prevented by enforcing a global maximum step parameter. For each iteration, if the maximum value within a $\boldsymbol{\delta}_i$ update vector is more than this global maximum step parameter, then each value within $\boldsymbol{\delta}_i$ is scaled down proportionally. The iterative process continues until $\boldsymbol{\delta}_i \approx \mathbf{0}$ or a predetermined maximum number of iterations has been reached.

An example of this iterative image restoration process is shown in Fig. 6. 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. 6(a). As the iterative restoration process proceeds in Figs. 6(b)–6(f), the image becomes more bimodal and smoother and this results in a greatly improved image. The majority of gray pixels that occur between characters are replaced with either black or white values so the histogram became a strongly bimodal distribution. As a consequence, 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 minimization of the BSA score is completed in 30 iterations for this image and produces a restored image that results in an optimal combination of these bimodal, smoothness, and average measures.

4 Binary Image Resolution Expansion

Document images have traditionally been scanned using binary thresholding due to the inherent binary nature of text images. Binary image processing therefore still retains great significance in the document research community. The BSA restoration method described in Sec. 3 is well suited to process grayscale images but is not capable of restoring binary images. In order for our system to restore binary text images, each original binary image is first convolved with a 3×3 spatial mask in the preprocessing module shown in Fig. 4 to create a grayscale image. The bimodal estimation, resolution expansion, and BSA restoration modules used for binary images are identical to those used for grayscale images in our system. Once the high-resolution grayscale image has been created, a global threshold T_{binary} is used in the postprocessing module to convert the grayscale image back to a binary image. This threshold is computed to be halfway between the white and black bimodal peaks

$$T_{\text{binary}} = \frac{\mu_{\text{white}} + \mu_{\text{black}}}{2}. \quad (11)$$

This section will detail the conversion from a binary image to a grayscale image using a spatial convolution mask in the preprocessing module.

In order to illustrate the restoration process, a single binary character in Fig. 7 is used as an example. The original A character shown in Fig. 7(a) was obtained by scanning at 100 dots per inch (dpi) using a binary threshold. The first step in the restoration process was to convert this low-

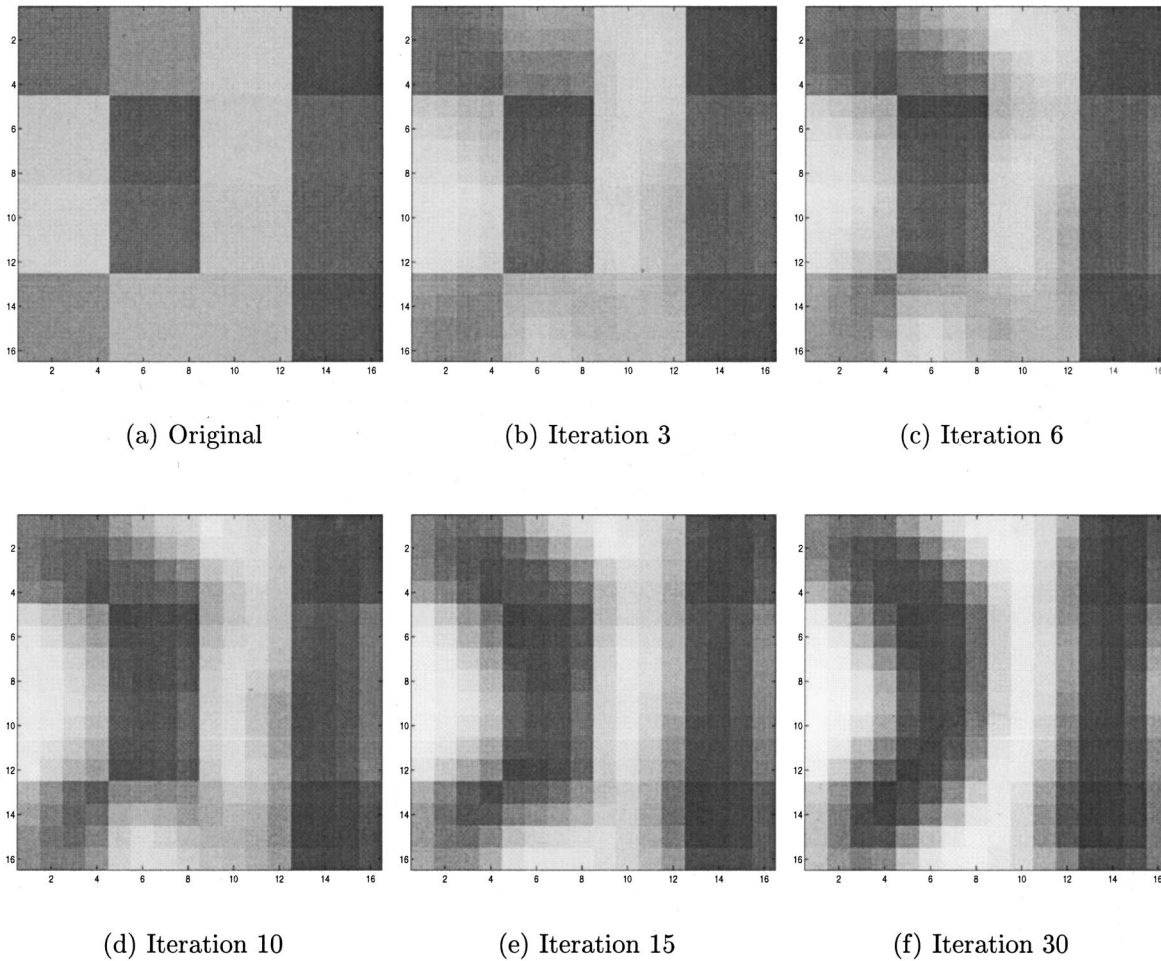


Fig. 6 Iterative BSA restoration example.

resolution binary image into a low-resolution grayscale image using a 3×3 spatial convolution mask. Other approaches such as Ref. 20, which uses area mapping to perform this conversion, have been successful as well. The resulting grayscale A is shown in Fig. 7(b). Pixel replication by a factor of 4 in both the horizontal and vertical directions was used to create an initial high-resolution image which was then restored by the grayscale BSA algorithm. The result is shown in Fig. 7(c). The final step in the restoration process converted this high-resolution grayscale back to a binary image using a global threshold. The resulting binary image is shown in Fig. 7(d). The following subsections will describe in detail the process to convert the binary images to grayscale.

4.1 Binary to Grayscale Conversion

The first step of the binary restoration process is to convert the original low-resolution binary image into a low-resolution grayscale image. In our system, 8-bit grayscale quantization is used so that the allowable range of values for each grayscale pixel are integers from 0 (black) to 255 (white). Convolution is performed to create gray values from a group of neighboring binary pixels. The discrete convolution of two images $f_{x,y}$ and $g_{x,y}$, denoted by $f_{x,y} * g_{x,y}$, is defined by

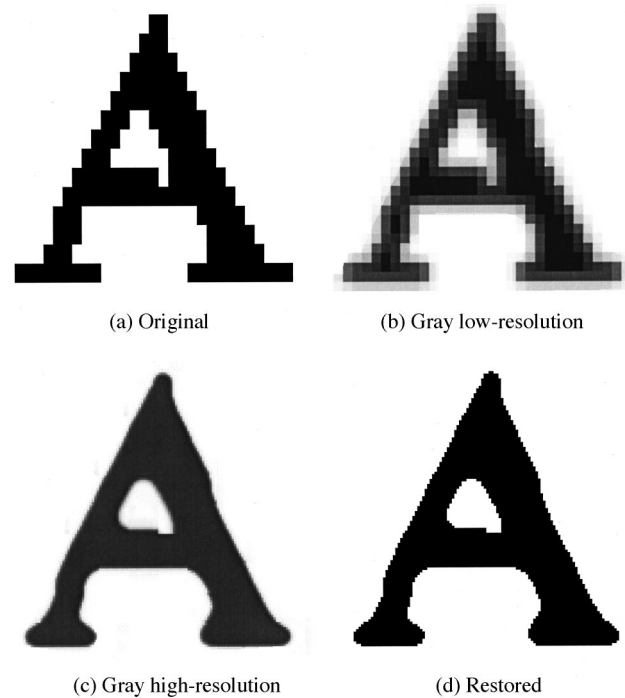


Fig. 7 Example of binary restoration process.

$$f_{x,y} * g_{x,y} = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} f_{r,c} g_{x-r,y-c}, \quad (12)$$

where R is the number of rows and C is the number of columns within each image.

The expression in Eq. (12) is a mathematical representation of a spatial convolution mask. Because convolution involves flipping $g_{x,y}$, it must be symmetric about its center. Throughout this paper the 3×3 mask $g_{x,y}$ defined by

$$g_{x,y} = \frac{1}{w_M + 4w_S + 4w_C} \begin{bmatrix} w_C & w_S & w_C \\ w_S & w_M & w_S \\ w_C & w_S & w_C \end{bmatrix} \quad (13)$$

is used to perform this convolution, where w_M is the relative weight for the middle pixel, w_S represents the four side pixels, and w_C corresponds to the weight for each of the four corner pixels. The values of the mask are equal to zero for all other possible locations. Spatial convolution can be thought of as a weighted averaging over a neighborhood of pixels. A constraint is enforced on these weights, based on the distance from the center of the mask

$$w_M \geq w_S \geq w_C \geq 0 \quad (14)$$

so pixels closer to the center are more heavily weighted. The value of the grayscale pixels x^{gray} are easily computed from the original binary pixel $x_{r,c}$ along with its eight neighbors

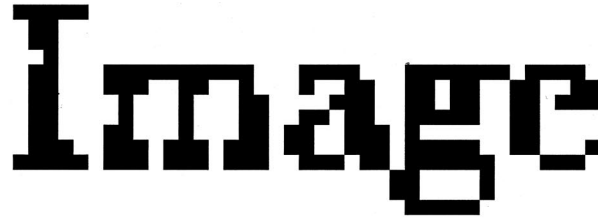
$$x_{r,c}^{\text{gray}} = \frac{1}{w_M + 4w_S + 4w_C} [w_M x_{r,c} + w_S (x_{r-1,c} + x_{r+1,c} + x_{r,c-1} + x_{r,c+1}) + w_C (x_{r-1,c-1} + x_{r-1,c+1} + x_{r+1,c-1} + x_{r+1,c+1})]. \quad (15)$$

To avoid boundary condition problems, when $x_{r,c}$ is an edge pixel only the existing image pixels are included in the computation.

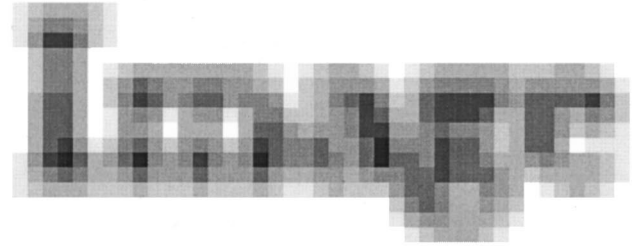
Figure 8 shows an original low-resolution binary image and two grayscale images obtained from convolving the original with this mask for various weights. The image in Fig. 8(b), resulting from equal values for all three weight coefficients which effectively averages all nine pixels within the 3×3 block, is severely blurred with a large amount of touching characters. The image in Fig. 8(c) resulting from the weights inversely proportional to the distance from the center pixel, $w_M = 4$, $w_S = 1$, and $w_C = 0.707$ creates a visually satisfactory grayscale image. These weights were used throughout this paper, and there was no attempt to optimize the relative weights for text images.

4.2 Constrained Binary to Grayscale Conversion

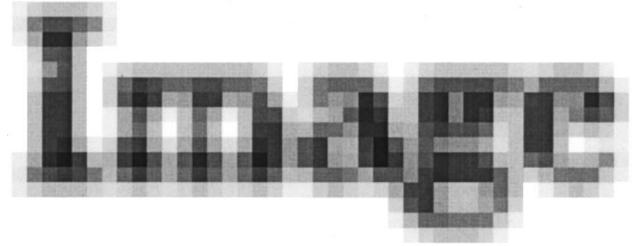
In our system, we impose a further constraint on the computed grayscale pixels. If the binary pixel is black, then the grayscale pixel should be dark as well. Specifically, if $x = 0$ then $0 \leq x^{\text{gray}} \leq 127$ and similarly if $x = 255$ then $128 \leq x^{\text{gray}} \leq 255$. However, this constraint is not always the



(a) Binary original



(b) $w_M = w_S = w_C = 1$



(c) $w_M = 4, w_S = 1, w_C = .707$

Fig. 8 Smeared images for various weights.

case by performing the straightforward 3×3 convolutions as shown in Figs. 8(b)–8(c) where the tail pixel of the character *e* which is black in Fig. 8(a) has a very light value of 198 in Fig. 8(b) and 144 in Fig. 8(c).

In order to compute each grayscale pixel while enforcing this constraint, a parameter Δ is introduced to measure the effect of the surrounding eight pixel neighbors on pixel $x(r,c)$

$$\Delta = \frac{1}{4w_S + 4w_C} [w_S (x_{r-1,c} + x_{r+1,c} + x_{r,c-1} + x_{r,c+1}) + w_C (x_{r-1,c-1} + x_{r-1,c+1} + x_{r+1,c-1} + x_{r+1,c+1})]. \quad (16)$$

The range of this measure is $0 \leq \Delta \leq 255$. The computed grayscale pixel x^{gray} is then determined based on the original value of the binary image x in the following manner:

$$\begin{aligned} \text{if } x = 0 \quad \text{then} \quad x^{\text{gray}}(r,c) &= \left\lfloor \frac{\Delta}{2} \right\rfloor \\ \text{if } x = 255 \quad \text{then} \quad x^{\text{gray}}(r,c) &= 128 + \left\lceil \frac{\Delta}{2} \right\rceil \end{aligned} \quad (17)$$

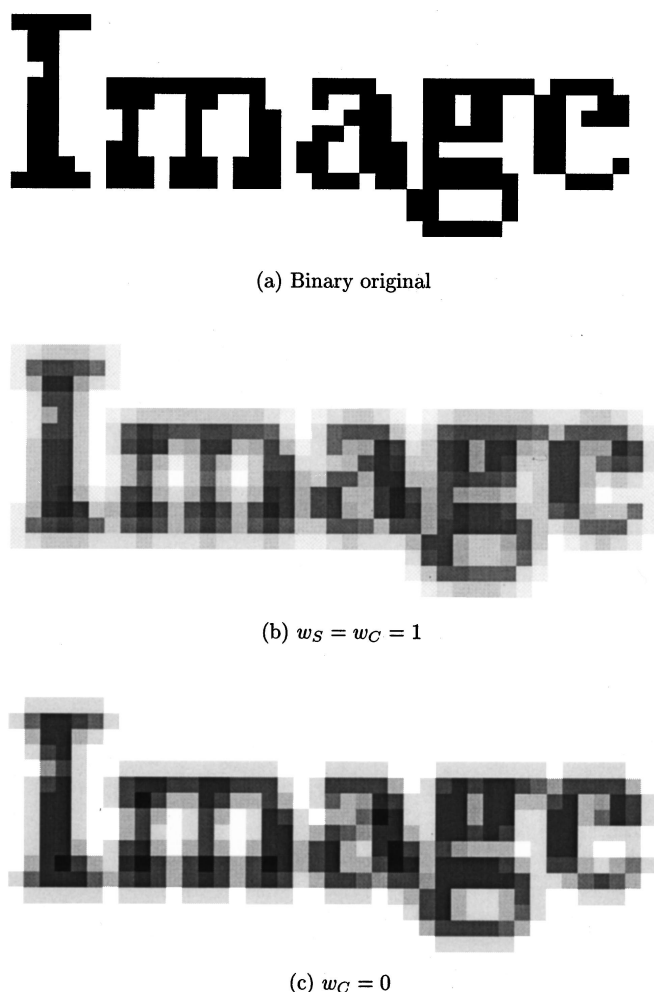


Fig. 9 Smeared images with binary constraint.

where $\lfloor x \rfloor$ is the integer floor rounding function defined by the largest integer that is less than or equal to x . Although the computational cost for this constrained approach is increased compared to the direct convolution approach in Sec. 4.1, the resulting constrained images produce superior OCR accuracy.

Figure 9 shows an original binary image and two images obtained by convolving it using this constraint. The image in Fig. 9(b) used equal values for both the side and corner weight coefficients and produced an image with a light gray border around the original black pixels. The image in Fig. 9(c) resulted from $w_S = 1$ and $w_C = 0$, where the computed grayscale values were only dependent on the side neighbors and not the corner neighbors. The gray border created in this image was not continuous as in Fig. 9(b) which can be observed in the top of the "I" character.

5 Experimental Results

A number of experiments were conducted to measure the success of our proposed text restoration system. It 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 and produce blurry results. Cubic spline expansion²¹ approximates the given discrete low-resolution pixels by 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, thus producing ringing effects. Resulting images from cubic spline expansion are shown in Figs. 12(c) and 13(b). In all observed cases cubic spline interpolations produced results superior to linear interpolations, therefore linear expansions were generally not performed.

Optical character recognition accuracy was used to quantitatively compare the results of our resolution expansion system with other methods. Experiments were conducted to measure the results for both grayscale and binary document images as well as video frames. In order to evaluate our system, document images from the University of Washington English Document Image Database were used. Twelve document images from this database are shown in Fig. 10. The choice of this database is based on the fact that it is available in the public domain and is widely used in the document research community. To demonstrate the capability of our system to expand low-resolution video images, a number of frames containing text were expanded as well. Figure 11 displays eight of these video frames. Caere's OMNIPAGE PRO 9.0 OCR package, the world's best-selling desktop OCR software, was used throughout these experiments to recognize the characters from the images. The resulting OCR text files were compared to the ground truth provided by the University of Washington database using the OCR ACCURACY REPORT VERSION 5.3 software developed at UNLV-ISRI. The following two sections describe the experiments in detail, then are followed by an analysis of the system's performance.

5.1 Experiment 1—University of Washington Grayscale Documents

The first experiment involved enhancing a total of 74 grayscale document images and video frames. Sixty-four of the images were from the University of Washington English Document Image Database and the remaining images were video frames. The document images were initially scanned at 100 dpi resolution using 8-bit grayscale quantization and then expanded by a factor of 3 in both the horizontal and vertical directions to produce 300 dpi results. A document example comparing our system with linear interpolation and cubic spline expansion is shown in Fig. 12. The original word "applications" depicted in Fig. 12(a) is severely blocky. Linear interpolation by a factor of 3 was used to produce the image in Fig. 12(b) which is blurry and lacks contrast. A cubic spline interpolation resulted in an image with better contrast but was still somewhat blurred as shown in Fig. 12(c). The BSA-restored image shown in Fig. 12(d) had the best contrast and produced superior character separation as evident between the characters in "app." The video images were obtained from broadcast television and captured at a resolution of 320×240 pixels for each frame. Image restoration results for a section of a sample video frame are shown in Fig. 13. The original grayscale image is displayed in Fig. 13(a) where blockiness

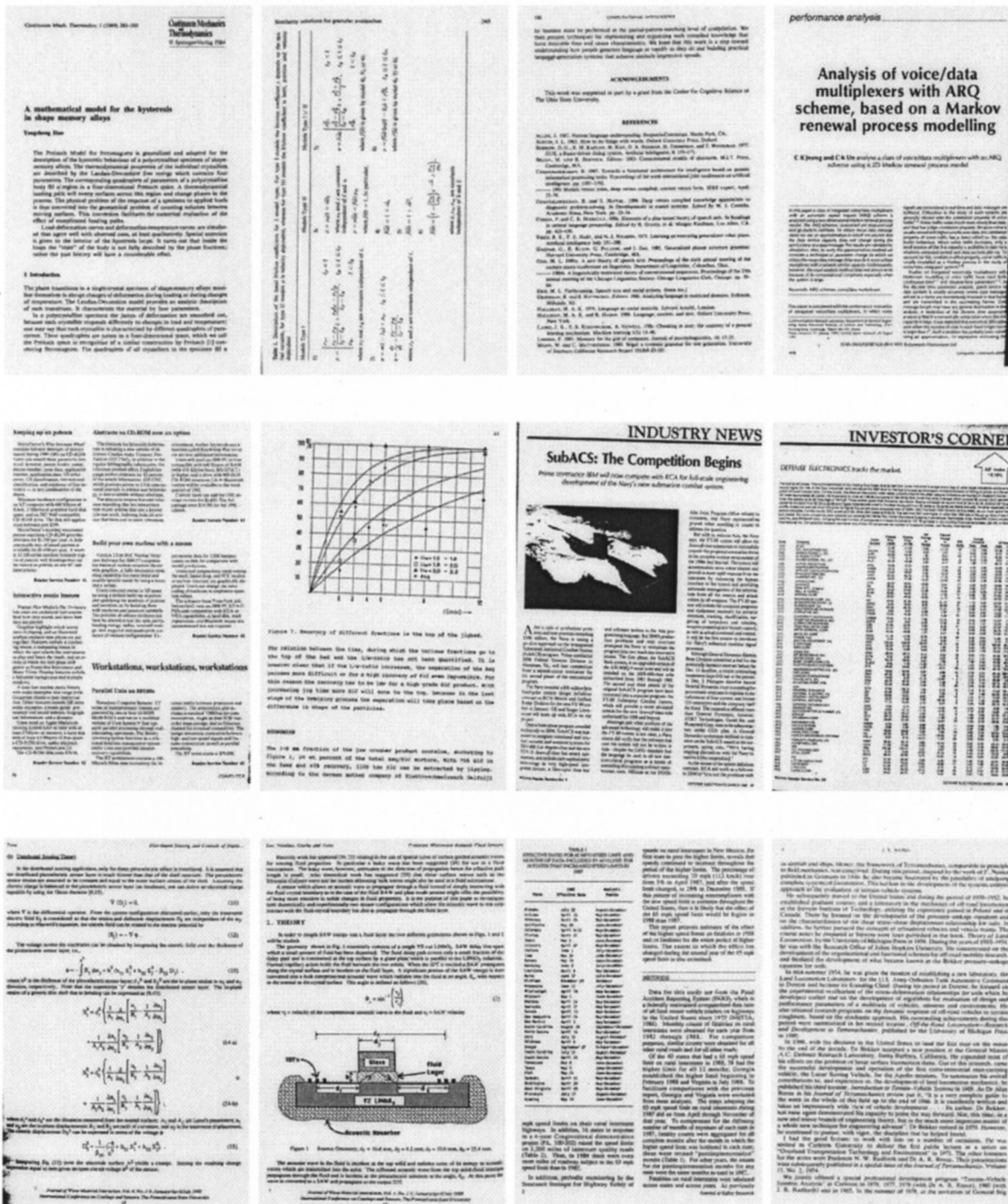


Fig. 10 Experimental document images.

and touching characters were abundant. A cubic spline expansion by a factor of 4 resulted in an image shown in Fig. 13(b), which was much improved although somewhat blurry. The image obtained from BSA restoration in Fig. 13(c) was visually superior with a high contrast and uniform background.

To quantitatively compare our system with other expansion methods, resulting images were compared with cubic spline expansion to measure OCR performance. Shown in Fig. 14(a) is the resulting 300 dpi image obtained using

cubic spline expansion on a paragraph from one of the University of Washington document images. The text file created by OCR is shown in Fig. 14(b) where mistakes have been highlighted. For this example, the OCR results for the cubic spline image had seven areas where mistakes were made. The same image paragraph was processed by our system resulting in the image in Fig. 15(a) which has better contrast than its corresponding cubic spline image in Fig. 14(a). OCR results are improved as well; only four mistake areas are highlighted in Fig. 15(b). The OCR character ac-

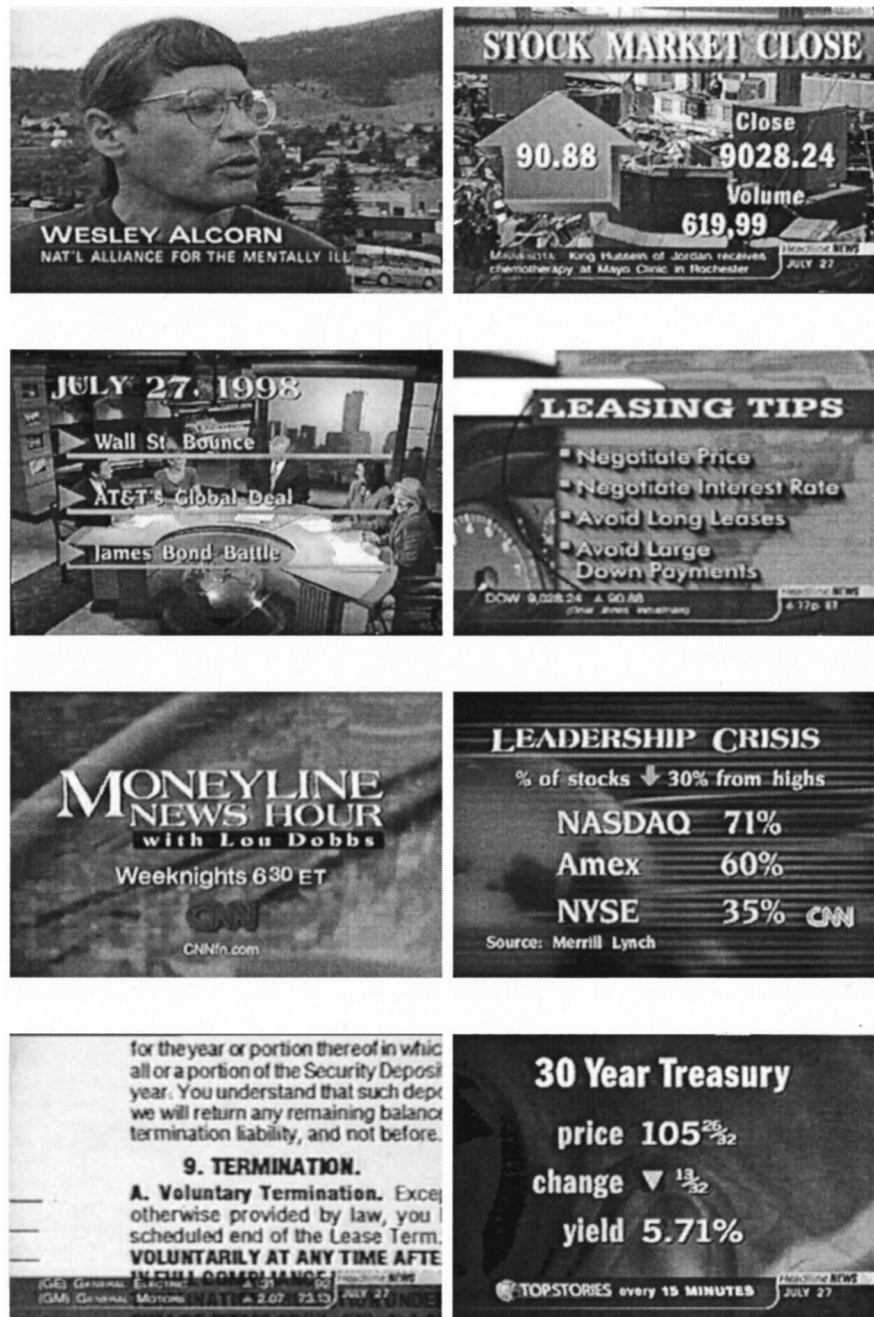


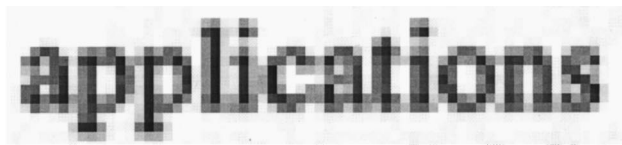
Fig. 11 Experimental video frames.

curacy for all 74 grayscale images is plotted in Fig. 16. There were over 200 000 characters in this experiment. The cubic spline expansion had an overall character accuracy of 88.0%. Images restored using the BSA-based system had an overall character accuracy of 90.9%, which is a 24% reduction in character errors. For these images, optical character recognition shows the superiority of our restoration system compared with the results produced by the cubic spline expansion.

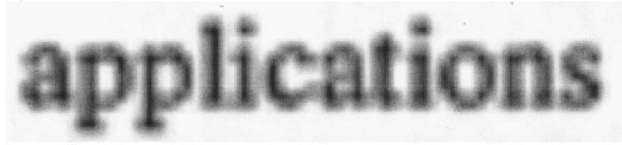
5.2 Experiment 2—University of Washington Binary Documents

To measure the success of the system in restoring binary document images, a second experiment was conducted us-

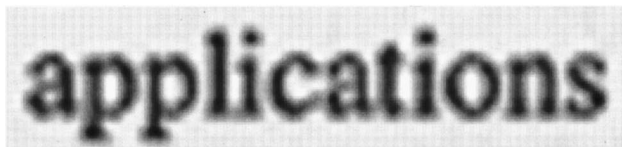
ing a total of 48 document pages from the University of Washington database. These images were scanned at 100 dpi using binary quantization and then expanded by a factor of 3 using the binary BSA restoration technique described in Sec. 4 to create 300 dpi images. Throughout these experiments, the constrained binary to grayscale conversion described in Sec. 4.2 was used because it consistently produced approximately 10% more accurate OCR results than the unconstrained conversion discussed in Sec. 4.1. The binary enhancement process for the sample word "representative" from this experiment is shown in Fig. 17 to visually illustrate the improvements. The original binary image in Fig. 17(a) was convolved with a spatial mask of



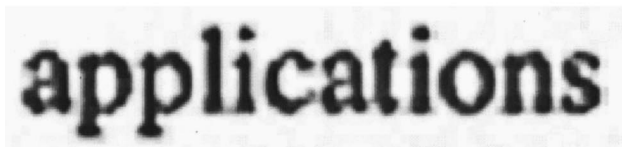
(a) Original image



(b) Linear interpolated image



(c) Cubic spline image



(d) BSA restored image

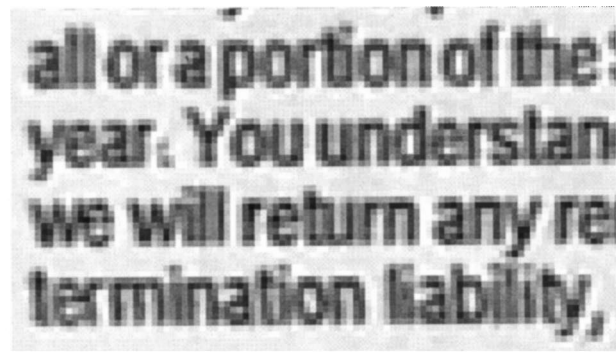
Fig. 12 Results of grayscale document restoration.

weights $w_S=1$ and $w_C=0.707$ to produce a grayscale image in Fig. 17(b). This low-resolution grayscale image was enhanced by the system to produce a high-resolution grayscale image in Fig. 17(c), which was converted to a binary image to produce a restored image in Fig. 17(d). The resulting high-resolution image was noticeably less blocky than the original.

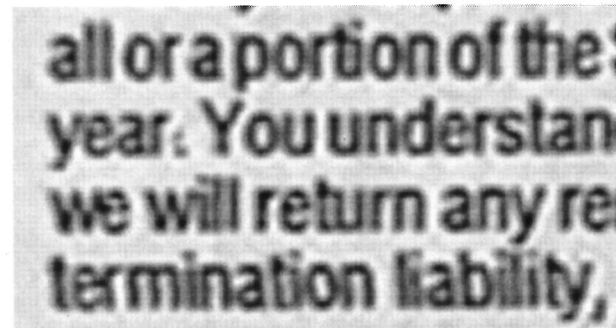
To quantitatively measure the success of the system in enhancing binary document images, OCR accuracy was compared before and after resolution expansion. The expanded images produced by our system were compared with 300 dpi images created by pixel replication. Figure 18 shows the OCR character accuracy results for both sets of 300 dpi images. There were a total of over 140 000 characters in this dataset. The overall character accuracy for the original images was 82.0% and the overall character accuracy for the restored images was 89.1%, which was a 39.5% reduction in errors—a significant improvement over pixel replication.

5.3 System Performance

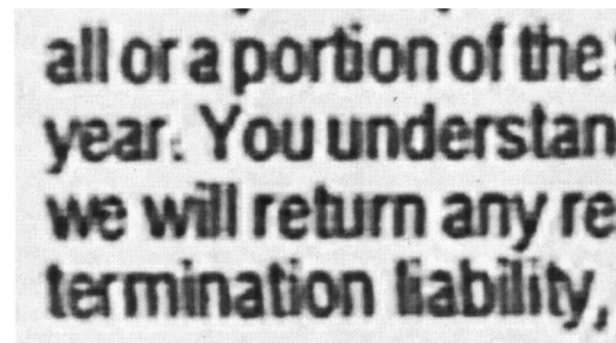
The previous experimental results demonstrate the advantages of our text restoration system compared to standard methods of interpolation such as linear interpolation and cubic spline expansion. The system was tested using images from the widely available University of Washington document database and evaluated using the commercial Caere OMNIPAGE PRO OCR software package to measure character accuracy. For the document and video images in



(a) Original image



(b) Cubic spline image



(c) BSA restored image

Fig. 13 Results of grayscale video frame restoration.

these experiments, the images produced by our restoration system resulted in a higher OCR accuracy compared to other standard methods.

As stated earlier in this paper, the goal of this resolution expansion system is to improve the OCR accuracy. Real-time image processing is not a requirement. Our system takes approximately 10 min to expand a single full-page 100 dpi document image by a factor of 3 on a 250 MHz workstation. Cubic spline interpolation is much faster than our system and takes about 10 s for a single document page. Video frames are typically smaller than low-resolution document images and are therefore processed more quickly by the system. A 320×240 video frame takes approximately 2 min for our system to expand by a factor of 3. Cubic spline expansion of a video frame only takes

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) Cubic 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) Cubic spline OCR results

Fig. 14 Example OCR results for cubic spline.

about 2 s. If the speed concern becomes an issue in the future, our approach has a potential for massive parallelization because the images can be divided into blocks of pixels that can be restored independently.

6 Conclusions

In this paper, we described a system designed to restore document and text images using resolution expansion. Our system was shown to be capable of enhancing grayscale documents and video images as well as binary document images. The success of the system was demonstrated by experiments using images from a standard document 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.

(a) Cubic 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) Cubic spline OCR results

Fig. 15 Example OCR results for BSA.

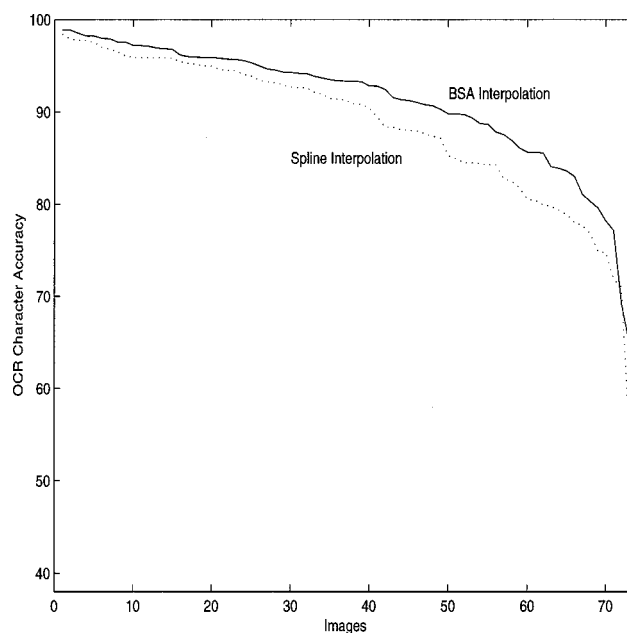


Fig. 16 OCR results for grayscale restoration.

database and a commercial OCR package. Restoration of grayscale images was performed by optimizing bimodal, smoothness, and average (BSA) scores that measure desired properties of text images. These scores were combined to form a single scoring function which produced images that were strongly bimodal and smooth, while satisfying the average constraint score. When the original image was binary, it was initially converted to a grayscale

representative

(a) Original binary image

representative

(b) Convolved grayscale image

representative

(c) Restored grayscale image

representative

(d) Restored binary image

Fig. 17 Results of binary document restoration.

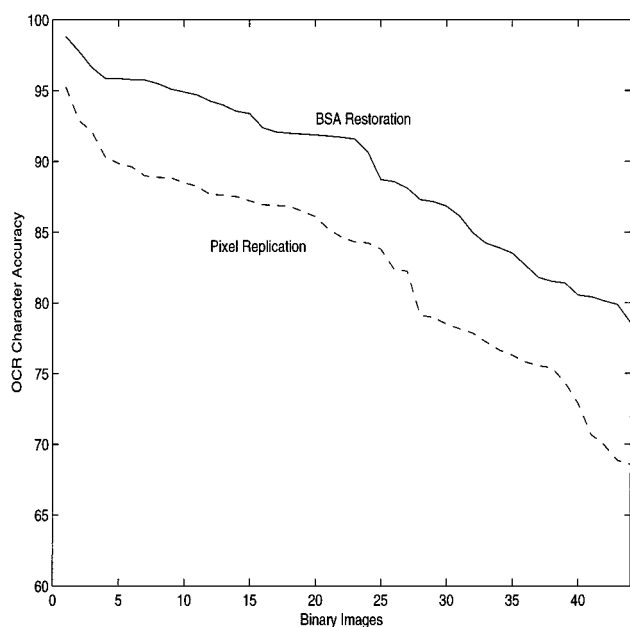


Fig. 18 OCR results for binary documents.

image using a spatial convolution mask and then processed as grayscale images. These resultant images restored using our system were shown to be superior to images expanded using existing linear interpolation and cubic spline expansion techniques.

Optical character recognition accuracy was also used to quantitatively measure the success of various resolution expansion methods. Both binary and grayscale text images restored with our system were experimentally shown to improve OCR accuracy compared to linear interpolation and cubic spline expansion. Although our system was designed to enhance text images, it can also be used to accurately expand checks, tables, charts, graphs, line drawings, and other types of documents as well because of their similarities.

Acknowledgments

P.D.T. 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. Caere Corporation, OMNIPAGE PRO 9.0 software package (1998).
2. Xerox Corporation, TEXTBRIDGE PRO 9.0 software package (1999).
3. H. Kaufman and A. M. Tekalp, "Survey of estimation techniques in image restoration," *IEEE Control Syst. Mag.* **11**(1), 16–24 (1991).
4. M. Y. Jaisimha, E. A. Riskin, R. Ladner, and S. Werner, "Model-based restoration of document images for OCR," *Proc. SPIE* **2660**, 297–308 (1996).
5. M.-Y. Yoon, S.-W. Lee, and J.-S. Kim, "Faxed image restoration using Kalman filtering," in *Proc. International Conf. Document Analysis Recognition*, pp. 677–681 (1995).
6. P. Stubberud, J. Kanai, and V. Kalluri, "Adaptive image restoration of text images that contain touching or broken characters," in *Proc. International Conf. Document Analysis Recognition*, pp. 778–781 (1995).
7. A. P. Whichello and H. Yan, "Linking broken character borders with variable sized masks to improve recognition," *Pattern Recogn.* **29**(8), 1429–1435 (1996).

8. J. Liang, R. M. Haralick, and I. T. Phillips, "Document image restoration using binary morphological filters," *Proc. SPIE* **2660**, 274–285 (1996).
9. Y.-C. Shin, R. Sridhar, V. Demjanenko, P. W. Palumbo, and J. J. Hull, "Contrast enhancement of mail piece images," *Proc. SPIE* **1661**, 27–37 (1992).
10. G. Ramponi and P. Fontanot, "Enhancing document images with a quadratic filter," *Signal Process.* **33**, 23–34 (1993).
11. L. Koskinen, H. Huttunen, and J. T. Astola, "Text enhancement method based on soft morphological filters," *Proc. SPIE* **2181**, 243–253 (1994).
12. R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.* **29**(6), 1153–1160 (1981).
13. T. C. Chen and R. J. P. de Figueiredo, "Image decimation and interpolation techniques based on frequency domain analysis," *IEEE Trans. Commun.* **32**(4), 479–484 (1984).
14. N. B. Karayiannis and A. N. Venetsanopoulos, "Image interpolation based on variational principles," *Signal Process.* **25**, 259–288 (1991).
15. A. D. Kulkarni and K. Sivaraman, "Interpolation of digital imagery using hyperspace approximation," *Signal Process.* **7**, 65–73 (1987).
16. R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.* **3**(3), 233–242 (1994).
17. P. D. Thouin and C.-I. Chang, "A method for restoration of low-resolution text images," in *Proc. Symp. Document Image Understanding*, pp. 143–148 (1999).
18. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA (1992).
19. J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: general algorithm," *Mon. Not. R. Astron. Soc.* **211**, 111–124 (1984).
20. H. R. Kang, *Color Technology for Electronic Imaging Devices*, SPIE, Bellingham, WA (1997).
21. H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.* **26**(6), 508–517 (1978).



Paul D. Thouin received his BS degree in electrical engineering from the University of Michigan, Ann Arbor, in 1987. In 1993, he obtained his MSEE degree from George Washington University in Washington D.C. He received his PhD degree in electrical engineering from the University of Maryland Baltimore County in 2000. Dr. 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 in the Research and Development Group. His research interests include image enhancement, statistical modeling, document analysis, pattern recognition, and multiframe video processing. Dr. Thouin is a member of SPIE and Phi Kappa Phi.



Chein-I Chang received his BS degree from Soochow University, Taiwan, in 1973, his MS degree from National Tsing Hua University, Hsinchu, Taiwan, in 1975, and his MA degree from the State University of New York at Stony Brook, in 1977, all in mathematics. He also received his MS and MSEE degrees from the University of Illinois at Urbana-Champaign, in 1982, and his PhD degree in electrical engineering from the University of Maryland, College Park, in 1987. Dr. Chang has been with the University of Maryland Baltimore County (UMBC) since 1987, as a visiting assistant professor from January 1987 to August 1987. He was a visiting research specialist at the National Cheng Kung University, Tainan, from 1994 to 1995. He has a patent on automatic pattern recognition and several pending patents on image processing techniques for hyperspectral imaging and detection of microcalcifications. He is currently on the editorial board of *Journal of High Speed Networks*. 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.