

# Super-resolution of Text Images Using Edge-Directed Tangent Field

Jyotirmoy Banerjee and C.V. Jawahar  
Centre for Visual Information Technology  
IIIT, Hyderabad 500032 INDIA

{jyotirmoy@research. and jawahar@}iiit.ac.in

## Abstract

*This paper presents an edge-directed super-resolution algorithm for gray level document images without using any training set. This technique creates an image with smooth regions in both the foreground and the background, while allowing sharp discontinuities across and smoothness along the edges. Our method preserves sharp corners in text images by using the local edge direction, which is computed first by evaluating the gradient field and then taking its tangent. Super-resolution of document images is characterized by bimodality, smoothness along the edges as well as subsampling consistency. These characteristics are enforced in a Markov Random Field (MRF) framework by defining an appropriate energy function. In our case, subsampling of super-resolution image will return the original low-resolution one, proving the correctness of the method. The super-resolution image, is generated by iteratively reducing this energy function. Experimental results on a wide variety of input images, demonstrate the effectiveness of our method for document image super-resolution.*

## 1. Introduction

With the advent of modern electronic gadgets like PDAs, cellular phones, and digital cameras, the scope of document imaging has widened. Document analysis systems are becoming increasingly visible in everyday life. For instance, one may be interested in systems that process, store, understand document images obtained by cellular phones [8]. Processing challenges in this class of documents are considerably different from the conventional scanned document images. Many of this new class of documents are characterized by low resolution and poor quality required for immediate recognition. Super resolution provides an algorithmic solution to the resolution enhancement problem by exploiting the image-specific *a priori* information.

Super-resolution of low resolution document images is becoming an important pre-requisite for design and devel-

opment of robust document analysis systems. Large scale camera based book scanners employed in digital libraries require resolution enhancement to obtain high OCR accuracies. It is also true with the text embedded in natural scenes, which could be used for indexing. Digital video compression algorithms can benefit from successful text resolution expansion techniques. Videos are often indexed and retrieved based on text information. The text observed in broadcast videos is often low in resolution. Without enhancement, a simple binarization will completely remove many strokes. In these conditions, it is virtually impossible to do character recognition as most of the OCRs are designed to work at high resolutions (say  $\geq 300$  dpi). The task of resolution enhancement is typically to increase spatial resolution, while maintaining the difference between text and background. It can further assist the cause of recognition in low-resolution text images.

This paper focuses on the issue of increasing the resolution of a single document image. There has been a substantial amount of previous work in super-resolution for general imagery [5, 6, 14]. However, document images are a distinct class of images widely differing from natural images. The problem of document super-resolution is a special case of image super-resolution because (a) document images are pseudo binary in nature and (b) the regularity of the patterns used in this “visual” language distinguishes the document images from natural scenes. Further, due to our excessive familiarity in the case of document images we have fair amount of *a priori* knowledge about the high resolution image. This increases the expectations on the document super resolution algorithms. A successful document super resolution algorithm needs to use the text-specific *a priori* information. Edges are geometric regular spatial patterns, and are among the most noticeable features in document images. The visual quality near edge areas adversely affect our perception of distortion.

In this work, we propose an algorithm for super resolution of textual document images using an edge directed tangent field. This scheme is ideally suited for the textual content where the smoothness will have to be enforced

along the edges instead of across the edges. We demonstrate the applicability of the approach on documents obtained from book scanners, cell-phone cameras and broadcast videos. We demonstrate the qualitative and quantitative improvement of this method over traditional resolution enhancement schemes.

## 2. Super-resolution in Document Images

Simple approaches to image enhancement are popular in literature. Gaussian and Wiener filters (and a host of other linear filters) have been used for smoothing the blockiness created by the low resolution imaging. Median filters (and similar nonlinear filters) tend to fare better, producing less blurry images. Interpolation methods such as cubic-spline interpolation tend to be the most common image resolution enhancement approach. There are two primary difficulties with interpolation methods for resolution enhancement. First, smoothing in interpolation is indiscriminate. It occurs in places with gradual change, as well as across the sharp edges, producing blurring. Second, these approaches are inconsistent. Subsampling the super-resolution image will not return the original low-resolution one, which implies that the high resolution image is not the “true” high resolution image, which one is interested in estimating. Hence we need a model which not only maintains consistency but also tries to ensure that smoothing does not occur in region boundaries.

One of the earliest attempts to do super-resolution of document images was by Li and Doermann [9]. They used the method of projection onto convex sets (POCS), to deblur scene text in video sequences. This was particularly suitable for their application since overlaid text, usually have pure translation between frames. A pure translational model is a common assumption due to its simplicity and ease of implementation. In a parallel work, Capel and Zisserman [1] used a projective transform motion model for super-resolution of text specifically for image sequences in which the point-to-point image transformation was of enough complexity to demand such consideration. Both these methods successfully demonstrate the use of super-resolution to improve the document images. In a recent work, Teager filter (a quadratic unsharp masking filter) was adopted by Mancas-Thillou and Mirmehdi [11] for the extraction of high frequencies thus enhancing character edges. Most of these prior models did not reflect any text image property. This has been identified as a promising direction to derive super resolution algorithms specially suited for document images. Donaldson and Myers proposed a text specific prior model, which modeled the bimodality and the local smoothness with step discontinuity [4]. They use the Gibbs prior with a Huber gradient penalty function as their smoothness function. This piecewise smoothness prior is good at reduc-

ing false speckle details in the results, but it undermines the importance of enhancing edges. Dalley *et al.* [3] employed a training-based method, in a Bayesian framework. A database is built that indicates which high-resolution patch should be output given an input low-resolution patch. Park *et al.* [13] developed an alternative approach, an edge-based super-resolution technique that attempts to locate the edges to subpixel accuracy in a sequence of images taken from training examples, and then fuses the conglomerated edge information into the super-resolved image using a MRF formulation.

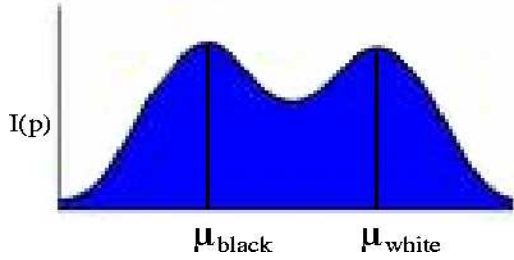
A variety of methods have been proposed to improve the contrast within a single text image. They include methods based on multi-resolution pyramid with fuzzy edge detectors [15], and a mixed approach using topological features and contour beautification [12]. There has been only limited work in the area of single frame non-training based super-resolution. Thouin and Chang [16] used nonlinear optimization on a gray scale input image to minimize a Bimodal Smoothness Average (BSA) score. Though this method works well, the processed direction of the smoothness constraint, which is a differential equation based method, is defined by gradient magnitude of the image, where the random attribute of image is not considered. Therefore it fails to preserve some edge and texture, specially the corner edges of text image.

In general, most of the previous approaches treated document image super-resolution very similar to that of super-resolution of natural images. This resulted in adverse characteristics near the character edges and corners. Textual content in document images are primarily binary and the smoothness will have to be preserved *along* the edges and *not across* the edges. We demonstrate that such an edge preserving resolution enhancement technique is ideally suited for document images.

## 3. Text Specific Prior Estimation

Though existing methods super-resolve documents, they have less emphasis on enhancing the edges. Significant amount of degradation takes place at the edges in the resolution expansion methods. Preserving character edges is most vital in document images. However, edges in low-resolution document images appear as spatially blurred edges due to degradation, sensor noise and focal blur. When edges are blur, it is difficult to explicitly locate the edges and its digital directions. This makes the super-resolution with focus on explicit enhancement of edges in document images difficult.

To avoid the difficulties with explicit edge enhancement approach, implicit edge-directed super-resolution method is proposed in this paper. The presented Markov Random Field (MRF) based edge-directed super-resolution method,



**Figure 1. Bimodal Distribution:**  $I(p)$  is the gray level at pixel  $p$ .  $\mu_{black}$  and  $\mu_{white}$  are foreground and background peaks, respectively.

is an implicit edge-directed restoration. It generates the edge-directed information on the fly, making the method independent of training set.

Ideally, any algorithm to perform document image super-resolution, should have the following characteristics:

- it should be able to successfully handle the bimodal distribution so typical of a document image.
- it should preserve and enhance the edges and corners.
- expanded images are constrained such that the sub-sampling the super-resolution image should return the original low-resolution one.

Also, we would like methods which are reasonably fast, and try to imbibe all the above properties in our formulation.

The general framework for the problems can be defined as follows. Let  $\mathcal{P}$  be the set of pixels in an image and  $\mathcal{L}$  be a set of labels for e.g. in gray level images, there are 256 labels. The labels correspond to quantities that we want to estimate at each pixel. A labeling  $f$  assigns a label  $f_p \in \mathcal{L}$  to each pixel  $p \in \mathcal{P}$ . The quality of a labeling is given by the energy function  $E(f)$ , and is defined in terms of its clique system. A neighborhood structure  $\mathcal{N}_p$ , which contains neighboring pixels of site  $p$  ( $p$  is not included), is first defined. Then a clique is defined on the neighborhood structure  $\mathcal{N}_p$ . A set of pixel sites  $c$  in  $\mathcal{N}_p$  is a clique if all pairs of sites in  $c$  are neighbors. Lastly, a function  $V_c$  called potential function defines the interactions of pixel sites in clique  $c$ . Spatial constraints is imposed through the formulation of potential function  $V_c$ . The potential function is related to the energy function as  $E(f) = \sum_{c \in \mathcal{C}} V_c(f)$ .

### 3.1. Bimodality Constraint

Images of text are usually smooth in both the foreground and background regions with sharp transitions only at the edges. Thus, text images typically have bimodal distributions, as shown in Figure 1, with large black and white peaks. The peak occurs at  $\mu_{white}$ , the background (white)

values, since the majority of pixels on a text page is background. There is a second peak at  $\mu_{black}$ , representing the black letters. 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 amount of these intermediate gray levels is related to the amount of blur in the document image. The textual content is almost always rendered with a high contrast, otherwise, the content provider risks the probability of the viewer not noticing the content. In order to obtain a clear view, we wish to obtain a sharp bimodal image, pushing the intermediate gray level towards their nearest peaks. To incorporate this property we define the energy function as

$$B_p(f_p) = (f_p - \mu_{white})^2 (f_p - \mu_{black})^2 \quad (1)$$

where  $B_p(f_p)$  is the cost of assigning label  $f_p$  to pixel  $p$ , effecting the distribution, and is referred to as the bimodal cost. The expression measures how far the assigned label  $f_p$  is from the assigned bimodal peaks. Minimizing this expression will assign labels  $f_p$  to pixel  $p$ , values that are close to either of the peaks, making the peaks in the distribution increasingly sharper. It is interesting to notice that this energy component is capable of regulating the distribution of the document image, thus the MRF operating at a global level. As we shall see later that we try to minimize this energy component, resulting in a sharp bimodal image.

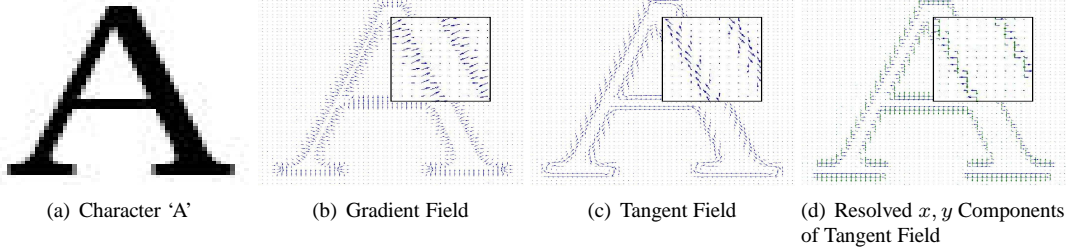
Let us now, introduce the energy term for clique with single site, which is defined as

$$D_p(f_p) = (I(p) - f_p)^2 + B_p(f_p) \quad (2)$$

where  $D_p(f_p)$  is the cost of assigning label  $f_p$  to pixel  $p$ , which is referred to as the data cost.  $I(p)$  is the gray level at pixel  $p$ .

### 3.2. Smoothness Along Edges

A sharp edge in an image corresponds to relatively large intensity gradients concentrated along the edge, while a relatively smooth area is composed of a more scattered set of weaker or almost no gradients. 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 document image has character images with sharp curves along the boundaries as shown in Figure 2a. The relation between high- and low-resolution image is essentially dependent on the smoothness of the edge direction. Character edges generally consist of piece wise smooth curves. The join of two curves are the corners of the characters. The enhancement approach needs to discriminate between smooth curve and the corners in the text image. Therefore, while restoring these character image the smoothness along the character edges have to be



**Figure 2. Tangent Field: (a) Character ‘A’ (b) The gradient field (c) The tangent field and (d) The resolved x and y components of the tangent field (c).**

enforced in the formulation, on the other hand maintaining sharp discontinuities across the edges. To find the edge direction we first compute the gradient of image as shown in Figure 2b. Then we take vectors tangential to this gradient field. This tangent field consists of vectors pointing along the boundaries of character images as shown in Figure 2c. Let the tangent vector at pixel location  $p$  of the degraded low-resolution image  $I$  be  $T^p$ . The tangent field is further resolved into its  $x$ -axis and  $y$ -axis components as shown in Figure 2d, which are denoted as  $T_x^p$  and  $T_y^p$ , respectively. This is done because the edges are four-connected image grid graph. These potentials are used in assigning labels  $f_p$  and  $f_q$  to two neighboring pixels. We define the energy function with a quadratic cost function for the clique with two sites is given by

$$V(f_p, f_q) = \begin{cases} \min(sT_x(f_p - f_q)^2, d) & \text{if } (p, q) \text{ are} \\ & \text{along x-axis} \\ \min(sT_y(f_p - f_q)^2, d) & \text{if } (p, q) \text{ are} \\ & \text{along y-axis} \end{cases} \quad (3)$$

where  $s$  is the rate of increase in the cost, and  $d$  controls when the cost stops increasing.  $V(f_p, f_q)$  is the cost of assigning labels  $f_p$  and  $f_q$  to two neighboring pixels, and is normally referred to as the smoothness cost. The truncated quadratic cost changes smoothly from being almost quadratic near the origin to a constant value as the cost increases.

To understand why this approach is effective, notice that a character edge has either sharp corners or smooth curves. These geometric spacial constraints can be described by local tangent field. Our proposed MRF model-based method is an implicit edge-directed approach. In this formulation, the edge direction of an edge pixel are indicated by the continuity strength in that direction. Instead of labeling each direction as either edge or non-edge direction, we measure the continuity strength in each direction with the strength of the tangent field. These values are derived from the intensity variations, i.e., the gradient. The relative continuity strengths of the directions are used as edge direction

information to formulate the geometric regularity spatial constraint, which can be summarized as smoothness along edge directions and sharpness across edge directions. Areas where the gradient is zero or negligible, the smoothness cost function is very low and does not have much influence. In these places the bimodal cost is the major deciding factor, thus rendering a highly smooth surface in those regions.

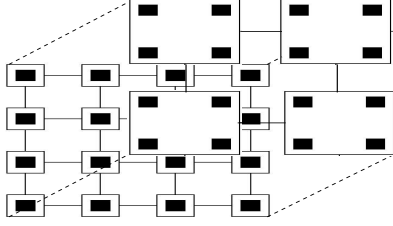
### 3.3. Subsampling Consistency

The subsampling consistency should be preserved between the low-resolution and its corresponding high-resolution image, which means that when you subsample a high-resolution image generated by the method, it should recover the original input image. In this section we describe a dualscale technique to circumvent this problem. The basic idea is to impose the criteria that the expanded images are constrained such 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. We use hierarchy to impose this constraint on the successive finer level.

In establishing the coarse-to-fine relation, we use the notion of dualscale image grid, as shown in Figure 3. The lower level corresponds to the original labeling problem we want to solve. The higher level consists of blocks of  $2^m \times 2^m$  pixel locations grouped together, where  $m$  is the magnification factor, and the resulting blocks are connected in a grid structure. The lower level and the higher level in Figure 3 correspond to the high- and low-resolution images respectively, and a block in the higher level corresponds to a pixel in the low-resolution image. The subsampling consistency can then be conditioned as

$$S_p(f_p, b) = (I(b) - f_p)^2 \quad (4)$$

where  $p$  is a pixel at the lower level and block  $b$  is the corresponding pixel at the higher level.  $I(b)$  is the gray level at pixel  $b$ .  $S_p(f_p, b)$  is the cost of assigning label  $f_p$  to pixel  $p$  based on block  $b$  that measures its distance from the corresponding block at the higher level. It is referred to as the subsampling cost.



**Figure 3. Dualscale structure: Each node in lower level(Super-resolved Image) corresponds to a block of four nodes in the higher level(Low-resolution Image). In this case the magnification factor  $m = 2$ .**

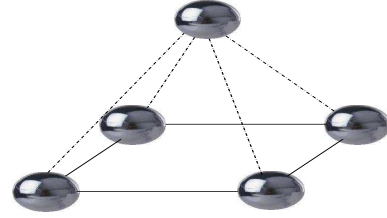
#### 4. MRF Formulation to Document Super-resolution

We model the spatial relationships in images using a Markov network, which has many well-known uses in image processing [2]. This means that the probability distribution of a node on the intermediate-resolved image is conditionally independent of all but the neighborhood of the node. Figure 4a shows the neighborhood of a node of the MRF and Figure 4b shows cliques in the neighborhood system. In Figure 4a, circles represent network nodes, and the lines indicate statistical dependencies between nodes. In Figure 4b, we define two kinds of cliques  $c_1 \in C_1$  and  $c_2 \in C_2$ . Therefore, for each node on intermediate-resolved image, there are six cliques related to it, one  $c_1$  clique and five  $c_2$  cliques. The clique  $c_1$  represents the dependency between the intermediate-resolved image and the bimodality of the restored image. The clique attains higher energy value as the pixel moves away from the bimodal peak. Lowering the energy, facilitates in deriving a sharp bimodal image. The clique  $c_2$  represents the dependency between two neighboring nodes. Clique  $c_2$  performs two distinct tasks. First, the selective smoothing using an tangent field is performed to improve the local smoothness of each region of text region. Second, it ensures that the high resolution image does not drift far from the corresponding low resolution image and is done by establishing a relation between the low- and high-resolution image.

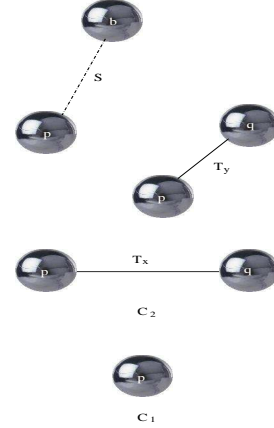
The quality of a labeling in general restoration problem is given by an energy function,

$$E(f) = \sum_{(p,q) \in \mathcal{N}} (V(f_p, f_q) + S_p(f_p, b)) + \sum_{p \in \mathcal{P}} D_p(f_p) \quad (5)$$

where  $\mathcal{N}$  are the edges in the five-connected image grid graph shown in Figure 4a.  $p$  and  $q$  are nodes belonging to the same level and node  $b$  belongs to the immediate higher level. Finding a labeling with minimum energy corresponds



(a) Neighborhood of a node



(b) Cliques in the neighborhood system

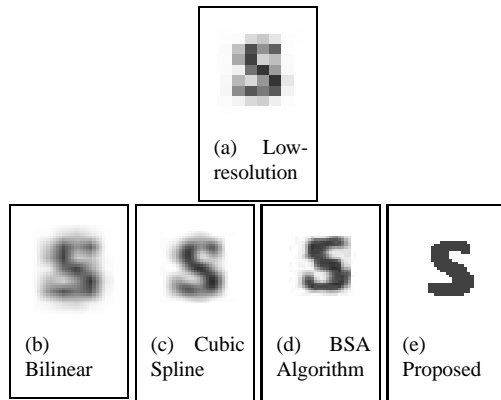
**Figure 4. Clique system in proposed MRF**

to the Maximum A Posteriori (MAP) estimation problem for an appropriately defined MRF.

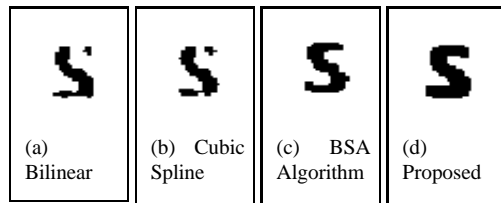
##### 4.1. Algorithm Details

Proposed super-resolution process embeds the MRF super-resolution framework (Figure 4a) through iteration. The bicubic-interpolated image of an observed low-resolution image is given as an initial intermediate resolved image. We predict missing image details in the interpolated image to create the super-resolution output. And the intermediate-resolved image is improved by the MRF framework. The edge weights are calculated both from neighbors from same level and the immediate higher level. The quality of the final super-resolved result varies with the number of iterations. The edge weights of the same level are extracted from the tangent field is given in Equation. 3. The edge that connects to the higher level in the dualscale structure (Figure 3) which passes the coarse-to-fine information, estimated using Equation 4.

Finding the exact solution can be computationally intractable, but we find good results using the approximate solution obtained by running a fast, iterative algorithm called efficient belief propagation [7]. The algorithm runs at one level of resolution and then uses the messages at that level in order to get estimates for the messages at the next finer



**Figure 5. Character ‘s’ super-resolved by a factor of 4 times**



**Figure 6. Thresholded version of the results of several methods in Figure 5.**

level, and so on, down to the original grid. Three or four iterations at each level and a maximum of five levels of grid hierarchy are sufficient. Inference algorithms based on belief propagation have been found to yield accurate results, but despite recent advances are often too slow for practical use [7]. For a full-size page (of size  $1600 \times 2600$ ) the processing time is way beyond commercial use. To make it work more efficiently for a document page, we try to divide and conquer the problem. Each character in a document is visually an independent entity, not to mention about a word or a line. Geometrically it does not depend on one another. Thus they can be dealt separately without effecting the overall document image. In our method, the low resolution full size paper is segmented to word or line level as per the feasibility. These small chunks of images are then fed into the algorithm, drastically reducing the time and space complexity.

## 5. Experimental Results

We demonstrate the performance of our algorithm on textual content in video frames as well as the document images obtained by book scanners, and cellphone cameras. We quantitatively and qualitatively demonstrate the superiority of the proposed model.

common agenda

(a) Ideal High-resolution Image

common agenda

(b) Low-resolution Image

common agenda

(c) Bilinear Interpolated Image

common agenda

(d) Cubic Spline Interpolated Image

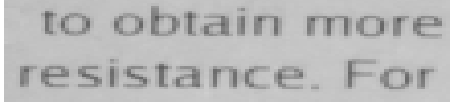
common agenda

(e) Using Proposed Method Restored Image

**Figure 7. Text super-resolved by a factor of 4 times**

To show the effectiveness of our method, we compare the results with several common methods, including bilinear interpolation, cubic-spline interpolation and BSA algorithm [16]. Figure 5 shows resulting images obtained from linear interpolation, cubic spline expansion, BSA algorithm and our method. The character ‘s’ from an image scanned at 75 dots per inch (dpi) using 8-bit gray scale quantization is shown in Figure 5a where significant blockiness is apparent. Bilinear interpolation results in a continuous curve, with a discontinuous derivative. These images naturally tend to be smooth, without sharp discontinuities, producing blurry results. Bilinear interpolation by a factor of four was used to create the image in Figure 5b, which is very blurry and lacks good contrast. Cubic-spline interpolation is an alternate popular scheme. The disadvantage of cubic splines is that they could oscillate in the neighborhood of an outlier producing a ringing effect. Figure 5c 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 Figure 5d has superior to the images obtained using other interpolation methods for this example. This method allows for sharp edges but does not discriminate between general text edge and corners. Our method presents an edge-directed super-resolution algorithm. Consequently the local edge direction are represented well by this method. Figure 5e shows, image quality is improved, strokes are reconstructed more precisely, linearity and smoothness of contours are improved, stroke





(a) Low-resolution Camera Based Image



(b) Our method

**Figure 8. Camera based results. A small portion of the text is magnified and displayed.**



(a) Low-resolution Image from television broadcast

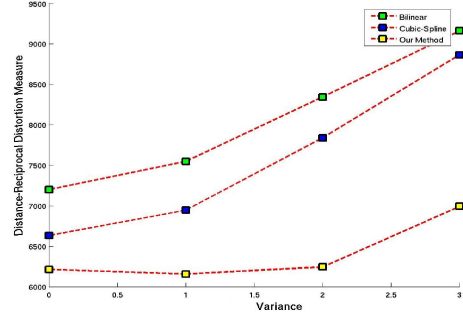


(b) Our method

**Figure 9. Result on text from television broadcast frames.**

width is more uniform, and shape features of fonts are reconstructed finely. Figure 6 shows the thresholded version of the results of several methods in Figure 5. Bilinear and cubic spline methods introduces cut in the thresholded image as shown in Figure 6a and Figure 6b, respectively. Figure 6c shows that there is still blockiness left on the smooth surface of the character ‘s’, introduced by BSA algorithm. The reason being that the algorithm breaks the whole image into  $4 \times 4$  blocks and each of these blocks are handled independently, resulting in lack of continuity across the blocks. Figure 6d shows that there is not much difference from the original image in Figure 5d even after thresholding as our method generates a sharp bimodal image. The image obtained using our method has smooth edges and is superior to the images obtained using other methods.

We demonstrate the effectiveness by creating low-resolution images from high-resolution originals, expanding the low-resolution imagery, and then measuring the distance to the originals. To achieve this an anti-aliasing process is performed by blurring (low pass filter) the image followed by block averaging (subsampling). For an image  $I$ , of  $r$  rows and  $c$  columns and a low pass filter with impulse response  $G$ , the resulting image  $i$  subsampled at each



**Figure 10. Text super-resolved by a factor of 4 times**

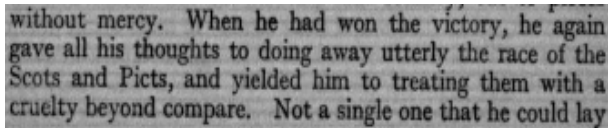
$\Delta$  pixels would be represented by

$$i(x, y) = \sum_{j=1}^r \sum_{i=1}^c I(i, j)G(x - i, y - j),$$

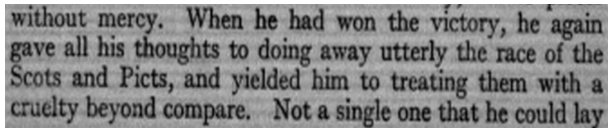
where  $x = 1, \dots, c/\Delta$  and  $y = 1, \dots, r/\Delta$ . Restored images are then compared with the original to determine the success of restoration numerically. For binary document images, the PSNR does not match well with subjective assessment, since it is a point-based measurement, and mutual relations between pixels are not taken into account. Hence, we use the Distance-Reciprocal Distortion Measure (DRDM) that measures the visual distortion in digital binary document images and matches well to the subjective evaluation by human visual perception [10]. The DRDM was used to compare the various methods of image resolution expansion. We initially take a  $70 \times 380$  size image at 300dpi as shown in Figure 7a. The low resolution image is generated by the process of anti-aliasing, where a Gaussian low pass filter of standard variance  $\sigma = 1$  and block averaged with  $\Delta = 4$  was used shown in Figure 7b. The bilinear interpolation produces a severely blurred image shown in Figure 7c, reducing the DRDM to 7549.7. The cubic-spline gives better result in Figure 7d with DRDM reduced to 6945.3. Our method produced the best image shown in Figure 7e by reducing the DRDM to 6156.4. The sharp decline in DRDM score justifies our claim. A comparative study of the reduction in DRDM for the various image expansion techniques is plotted in Figure 10. We observe that higher blur factor leads to greater error during restoration.

Experiments with Camera-Based images is conducted by capturing document images using a Cannon hand held camera. Result on camera-based image is displayed in Figure 8 Text in a video broadcast frames are rendered in very low-resolution. Result obtained by super-resolving these images is shown in Figure 9.

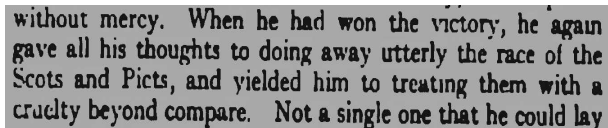
We examined effectiveness of the proposed method for improving OCR accuracy. A set of 20 page from a book

A low-resolution scan of a page of text. The text is blurry and difficult to read. The visible text is: "without mercy. When he had won the victory, he again gave all his thoughts to doing away utterly the race of the Scots and Picts, and yielded him to treating them with a cruelty beyond compare. Not a single one that he could lay".

(a) Low-resolution Image

A spline interpolated version of the low-resolution image. The text is sharper but still has some artifacts, particularly in the spacing and alignment of characters. The visible text is: "without mercy. When he had won the victory, he again gave all his thoughts to doing away utterly the race of the Scots and Picts, and yielded him to treating them with a cruelty beyond compare. Not a single one that he could lay".

(b) Spline Interpolated Image

The text restored by the proposed method. The text is very sharp and clear, with well-defined characters and proper spacing. The visible text is: "without mercy. When he had won the victory, he again gave all his thoughts to doing away utterly the race of the Scots and Picts, and yielded him to treating them with a cruelty beyond compare. Not a single one that he could lay".

(c) Our method

**Figure 11. Portion of text from the book.**

were used, where a page consist of approx 350 ~ 400 words. Each page was then scanned using 8-bit gray scale quantization at 100 dpi to create low-resolution original images using a ZEUTSCHEL OS 5000 scanner. These 100 dpi resolution pages were then expanded using various resolution expansion methods by a factor of four to create 400 dpi images which were processed by OCR. Restored images in 400 dpi were generated from input images in 100 dpi by the proposed method. The OCR accuracy, using FreeOCR Version 2.2, a freely downloadable OCR package, was compared with the results of images that were expanded using various other resolution expansion methods by a factor of four. There were a total of 28708 characters in these 20 images. Cubic spline interpolation resulted in 1558 character errors and our method had 869 character errors for an overall improvement of 44.2% for this set of images. The expansion required about  $6\frac{1}{2}$  min per page for our restoration algorithm. A sample section of restored images using cubic spline expansion and our method are shown in Figure 11. Figure 11a shows the original low-resolution image. We observe that the text is bimodal where  $\mu_{black}$  and  $\mu_{white}$  are 20 and 170, respectively. The reason for improvement in OCR-accuracy is possibly the enhancement of the edge directed tangent field. Since many OCR algorithms use directional features along contours as primary features, the contour enhancement are effective for improving OCR accuracy as well as image quality.

## 6. Conclusions

An implicit edge-directed super-resolution algorithm for document images is proposed in this paper. Edge direction information is incorporated in the formulation of the energy

function in the MRF model. The edge preserving super resolution scheme provides better results on a wide class of document images. The method is quite straightforward to implement and generate good results. Our algorithm is an instance of a general non training based approach that can be useful for document image-processing, that extracts a single high-resolution frame from a single low-resolution image, where the priors are derived from same image.

## References

- [1] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *ICPR00*, pages Vol I: 600–605, 2000.
- [2] R. Chellappa and A. Jain. *Markov Random Fields: Theory and Applications*. Academic Press, 1993.
- [3] G. Dalley, B. Freeman, and J. Marks. Single-frame text super-resolution: a bayesian approach. In *ICIP04*, pages V: 3295–3298, 2004.
- [4] K. Donaldson and G. K. Myers. Bayesian super-resolution of text in videowith a text-specific bimodal prior. *IJDAR*, 7(2-3):159–167, 2005.
- [5] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *IJIST*, 14(2):47–57, 2004.
- [6] R. Fattal. Image upsampling via imposed edge statistics. *ACM Trans. Graph.*, 26(3):95, 2007.
- [7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.
- [8] K. Kise and D. Doermann. *Second International Workshop on Camera-Based Document Analysis and Recognition*. 2007.
- [9] H. Li and D. Doermann. Superresolution-based enhancement of text in digital video. In *ICPR00*, pages Vol I: 847–850, 2000.
- [10] H. Lu, A. Kot, and Y. Shi. Distance-reciprocal distortion measure for binary document images. *SPLetters*, 11(2):228–231, February 2004.
- [11] C. Mancas-Thillou and M. Mirmehdi. Super-resolution text using the teager filter. In *First International Workshop on Camera-Based Document Analysis and Recognition*, pages 10–16, 2005.
- [12] H. Nishida. Restoring high-resolution binary images for text enhancement. In *ICIP (2)*, pages 506–509, 2005.
- [13] J. Park, Y. Kwon, and J. H. Kim. An example-based prior model for text image super-resolution. In *ICDAR05*, pages I: 374–378, 2005.
- [14] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine*, 20:21–36, 2003.
- [15] F. Sattar and D. Tay. Enhancement of document images using multiresolution and fuzzy logic techniques. *Signal Processing Letters*, 6(10):249–252, October 1999.
- [16] P. D. Thouin and C.-I. Chang. A method for restoration of low-resolution document images. *IJDAR*, 2(4):200–210, 2000.