Enhancement of Document Images from Cameras

M. J. Taylor and C. R. Dance

Xerox Research Centre Europe, 61 Regent Street, Cambridge CB2 1AB, UK.

ABSTRACT

As digital cameras become cheaper and more powerful, driven by the consumer digital photography market, we anticipate significant value in extending their utility as a general office peripheral by adding a paper scanning capability. The main technical challenges in realizing this new scanning interface are insufficient resolution, blur and lighting variations. We have developed an efficient technique for the recovery of text from digital camera images, which simultaneously treats these three problems, unlike other local thresholding algorithms which do not cope with blur and resolution enhancement.

The technique first performs deblurring by deconvolution, and then resolution enhancement by linear interpolation. We compare the performance of a threshold derived from the local mean and variance of *all* pixel values within a neighborhood[1] with a threshold derived from the local mean of just those pixels with high gradient.[2] We assess performance using OCR error scores.

Keywords: Scanning, digital cameras, thresholding, image restoration, OCR

1. INTRODUCTION

Devices for scanning paper are common in many work environments. Common examples range from the office photocopier to the fax machine. These two particular devices are similar in that the result of the scanning operation is traditionally printed on paper. This has the disadvantage that if a further copy of a document is required, then the original needs to be scanned again. The increasing importance of the internet as a communication medium has recently led to a shift in focus within the document imaging industry. The typical destination for document images is no longer assumed to be hard copy. Increasingly an electronic version is required for storage in a document database that can be searched on-line. In turn, this has led to a rise in the importance of scanners. Today's scanners typically come in one of three main forms: namely the flat-bed scanner, the sheet-feed (or keyboard) scanner and the hand-held scanner.

All three are rather cumbersome to use. The flat-bed and sheet-feed scanners require that the document be moved from its place of reading for scanning. The flat-bed is worst in this respect as it takes up rather a lot of desk space and is therefore often situated away from the reader's desk. The sheet-feed scanner usually resides on the reader's desk but only takes single sheets of a limited range of paper sizes and therefore cannot handle newspapers, magazines, books or bound reports. Those documents it can scan typically need staples removing, and a time-consuming manual feed. The hand-held scanner usually requires several swipes of the scanning head to capture a desired region.

2. SCANNING PAPER WITH DIGITAL CAMERAS

Over-the-desk scanning with digital cameras[3] has many advantages over traditional scanning techniques. Perhaps the most obvious is that of convenience. Figure 1a shows a user at a desk reading a document. A camera is pointed at the area of the desk that is usually used for reading. This camera can be thought of as a portal between the physical desk and the electronic desk-top. Live video of the desk is displayed in a window on the electronic desk-top. This arrangement affords a very convenient scanning interface. When the reader finds something he would like to scan, he simply selects that portion from the view of the physical desk (figure 1b) and drags it, automatically invoking OCR when appropriate, into a word processor or database, where the information may be stored, indexed and shared with others.

Note that reading is *minimally interrupted*. Documents need not be moved from their normal place of reading as is the case with flat-bed and sheet-feed scanners. Hand-held scanners also do not require the reader to move the

E-mail: mjt@cre.canon.co.uk, dance@xrce.xerox.com

a)  b) 

Figure 1. Face-up scanning: *a) A live view of the document that the user is reading is displayed on the workstation monitor. b) When the reader wishes to copy something, it is copied and pasted into other applications in the normal way. If text is copied, it is usually converted to an editable form (via* OCR*) before pasting into a word processor.*

document, but instead it is necessary to pass the scanning head carefully over the page in strips, which is hardly more convenient.

Of course, this new type of scanning interface is not helpful in applications where many *unbound* pages need to be scanned sequentially. In these situations, one would still prefer to use the familiar flat-bed scanner with a sheet feeder, operating at as many sheets per minute as possible. However, when small amounts (say upto a couple of pages) need to be scanned, the overhead in getting the document from the place of reading to the scanner and back again becomes more significant. This effect is greatest when only small parts from a page are required. Here are four examples:

- **Libraries**: In library work, people regularly write out reasonably long passages from books verbatim[4] because the alternative usually involves going to a photocopier which can take longer.
- **Translation**: Kanji script is difficult to enter into a computer via a keyboard, especially for non-native speakers. Overhead scanning could be used, in conjunction with an on-line dictionary, as an invaluable reading aid.[5]
- **Fax**: Much office scanning is currently performed by low resolution fax machines (200 dpi) for communication, and many fax messages consist of only one or two sheets. Overhead scanning can be used as a very convenient alternative to the traditional sheet-feed fax front-ends. A couple of pages may be faxed (or sent by e-mail) with the minimum of disruption.
- **Hypertext**: when the user selects a printed URL in the live window, the new scanning interface can automatically invoke a web browser

Overhead scanning thus enhances the value of printed information. By making scanning more convenient, we are affecting the role of paper in traditional work processes. We are making printing less of a one-way process by concentrating on the mechanics of getting small amounts of information off paper as easily as possible. We call this type of scanning *casual* scanning. This is not common in today's office because existing scanner interfaces are too cumbersome. However, with the advent of high quality digital cameras, casual scanning has the potential to become an increasingly frequent activity.

The resolution of today's consumer digital cameras has been largely determined by the resolution of current computer displays and the legacy of the analogue TV standards. As a consequence, images are typically as small as 640×480. If such a device is used to capture a full 8" $\times$ 11" page then a resolution of approximately 60 *pixels per inch* (ppi) is obtained. If it is assumed for the moment that a camera pixel is equivalent to a contact scanning *dot*, then it is clear that a typical consumer digital camera is only capable of scanning a sheet of paper at about 60 dpi. This is insufficient for most scanning tasks. For example, resolutions in the range of 200-300 dpi are required for reasonable OCR.

However, it is likely that the resolution of digital cameras will continue to improve, and it is not inconceivable that 200 dpi cameras with 2200×1600 (3.5 MPixel) arrays will become common in the relatively near future. Such cameras will largely replace chemical film for large market sectors. Our aim is to investigate the emerging image processing and user-interface issues that will in turn enable the *scan-as-you-read* interface between paper and electronic media.

3. DOCUMENT IMAGE DEGRADATION

Convenience in the sense of temporal efficiency is, of course, not the only relevant measure of a scanning system. The other principal measure is quality, usually measured in dots per inch (dpi). We would like to maintain the convenience that overhead scanning affords, and address the resulting problem of poor image quality. Our problem is to restore the original printed page from a low-resolution camera image, making the best use of the available camera pixels. To understand how to restore these images, it is necessary to understand the nature of the degradation caused by the channel between the paper source and the digitised image shown in figure 2.

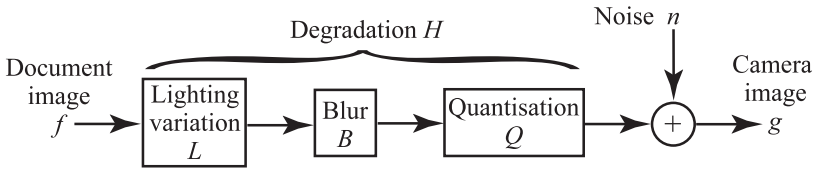

Figure 2. Degradation channel

3.1. Lighting variation

This effect is caused by uneven illumination and also the shape and orientation of the page relative to the light source (figure 3b).

3.2. Blur

There are two main sources of blur.[6] The first is lens imperfections or *aberrations*. The second is displacement between the sensor array and the focal plane: the camera is said to be out-of-focus if the image formed by the lens does not coincide with the sensor array or film. The sensitivity of a camera to this effect is often referred to as *depth of field*.

The depth of field is improved with a smaller aperture. In other words, the PSF radius reduces with the aperture radius, until a diffraction limit is reached. However, reducing the size of the aperture also reduces the amount of light reaching the imaging sensors. This means that longer exposure times are required in order to record the same SNR for given lighting conditions. Focal length also affects the depth of field. As focal length increases, or equivalently, angular field of view decreases, the depth of field increases.

It is clear that lens aberrations and a small depth of field could combine to give a spatially varying PSF. However, for the rest of this paper, we will assume a spatially invariant cylindrical PSF, parameterised by radius. This model is based on geometric optics, and is common in the engineering literature[7] (figure 3c).

3.3. Quantisation

Light is integrated over small rectangles into pixels whose area determines the level of quantisation. If the total field of view of the camera is fixed, say to be the size of a normal 8" $\times$ 11" page, then the severity of the quantisation of the page image is inversely proportional to the number of pixels in the array, which is usually referred to as the *resolution* of the digital camera (figure 3d).

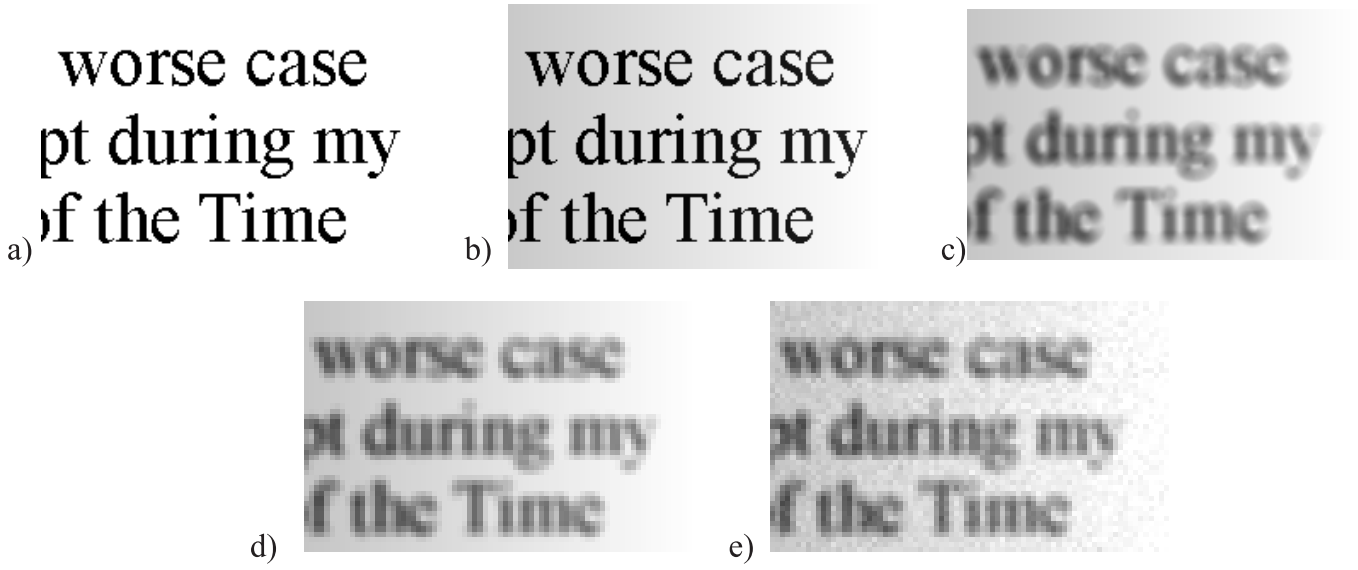

Figure 3. The effect of the different degradations *a) Original printed image, b) ligthing variation, c) blur, d) quantisation, e) noise.*

3.4. Noise

The sensor noise in an image has many sources.[8] The arrival of photons is usually best modeled by a Poisson process parameterised by a mean arrival rate λ that corresponds to the intensity of the incident light. The resulting variance of the number of arrivals in a sensor integration period of T seconds is also λT: the variance of the recorded charge in the integration period T increases with the mean intensity λ.

The recorded pixel value f is not generally a linear function of the charge q in the sensor bin. Typically the response is described by a model $f = \alpha q^{\beta}$ where α and β are device dependent constants.[9] This camera response clearly affects the noise variance over the intensity range of the sensor.

Despite these observations, we will assume in this paper that the noise is a zero mean additive Gaussian process that is independent of pixel value and time: the noise process is assumed spatially and temporally invariant (figure 3e).

4. DOCUMENT IMAGE RESTORATION

Restoration is difficult without some assumptions about both the original image and the degradation process, and the techniques that do exist are usually computationally expensive and improvements are often limited. We therefore need to make some reasonable simplifying assumptions that are well grounded in the application that we have thus far described.

4.1. Assumptions about the channel

It is difficult to restore images without prior knowledge of the blur and noise characteristics of the channel. This process known as blind deconvolution[10] would involve simultaneous estimation of the original image together with the blur and noise parameters. There are many proposed techniques for doing this[11–13] but they are all unnecessarily computationally intensive for this application. In section 3.2 it was shown how the PSF was dependent upon the aperture of the camera. The aperture need only vary if the amount of incident light on the document changes significantly. However, light levels in offices tend to change slowly. In section 3.4 we argued that noise levels are mainly dependent upon camera characteristics, and also do not vary significantly over time.

We therefore claim that it is not necessary to infer the blur and noise characteristics of the degradation process each time that the user performs a scan. Here, blur and noise parameters are best estimated by a background process using techniques such as those described in.[7,14–16] This means that the restoration process can assume that these parameters are known, and so it will be more efficient. Note however that lighting variation across the page (section 3.1) *is* expected to change from page to page, and as such needs to be treated at each scan.

4.2. Document images

In section 2 we described applications of overhead scanning that involve OCR which requires in the region of 200-300 dpi. We also noted that much scanning in the office is performed by fax machines, which are typically only 200 dpi. This observation suggests that much useful casual scanning *can* be performed with lower resolutions than the 600 dpi that is standard with today's contact scanners.

The widespread use of the fax machine suggests that much document information can be successfully communicated in binary form (black and white). This is not to say that photographs, graphics and text in full colour are not important, but rather that much useful information can be inferred from binary representations of the document. For example, if it can be assumed that the image contains text, then for many OCR applications it is unimportant what colour the text is as long as the text can be reliably recognised.

With this in mind, the first important assumption that we shall make is that the document image is *binary* – that is that there are only two colours in the original printed image. These are typically the foreground and a background colour, although in fact we do not make any assumptions about polarity. Images which contain more than one foregrond/background pair are a assumed to have been segmented into bimodal regions.

Figure 4 shows how a binary edge can be reconstructed to sub-pixel accuracy from its low-resolution grey-scale image. If the image is binarised at the same resolution then much information will be lost. We need a process for generating a higher resolution binary image from the grey-scale original *. Our task is to do this in two dimensions in the presence of lighting variation, blur and noise. In essence, we are attempting to trade the unwanted grey-scale resolution for increased spatial resolution, making optimal use of the degradation parameters. Computational efficiency is significant in our application because the scanning interface is designed to be a labour-saving tool: the longer the user has to wait between copying and pasting OCRd text, the less satisfying will be the result.

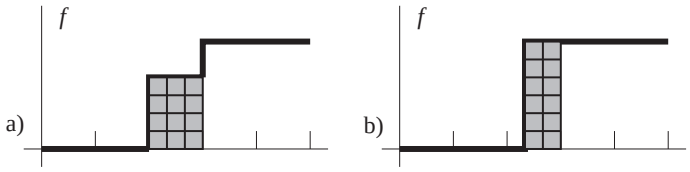

Figure 4. Super-resolution reconstruction. *a) The quantised low-resolution grey-level input, and b) the super-resolution binary reconstruction.*

4.3. OCR as a metric for binary image restoration

We choose to test our approach using OCR performance for three main reasons. The first reason, as discussed in section 2, is that many of the tasks that would most benefit from the convenience of camera scanning are likely to be based around copying text into computers. The second reason is that the bimodality assumption made in section 4.2 is more likely to be satisfied by text than line-art where the more complex use of colour is common. The third reason is that OCR performance is a pertinent task-based metric[18] with which we can compare the performance of camera scanning with that of flat-bed scanning; the equivalent for line-art is not so obvious.

4.4. Binary image restoration by thresholding

Many adaptive thresholding algorithms[18,19] only attempt to classify pixels in the presence of lighting variation.[1,2] They often assume that the image has come from source where the blur is at a much smaller scale than the objects, and is thus irrelevant. We note that the optimal level of a local threshold in our application is very much affected by blur (figure 3). An exception is to be found in the work of Mitchell *et al.*[20] who address the application of image coding of frozen frames in video-conferencing. They assume that the original image is binary. They perform high-frequency boosting to compensate for blur. They quantise the video image to three levels to enable efficient compression ratios, while maintaining legibility.

Most thresholding algorithms assume also that resolution is in abundance. This is usually a good assumption for document applications[18] where the image source is typically a flat-bed scanner. Other applications using camera

*One can think of quantised binary images as having been crudely anti-aliased.[17]

images include segmentation for automatic visual inspection of object silhouettes.[2] Here the object size is usually very large compared to the size of a pixel and resolution is not a concern.

4.5. Other approaches to binary image restoration

Geman and Geman[21] use MAP estimation to restore images using Markov random field models. These techniques have been used for binary image restoration with limited success.[22] The binary models are difficult to form and the resulting optimisation problems are extremely computationally expensive to solve, usually involving a simulated annealing process.

Wang and Pavlidis[23] assume that reliable knowledge about the degradation is difficult to use. Hence they do not attempt to restore to the original binary image space but instead attempt to extract topographical features for OCR directly from the degraded grey-scale image.

5. OUR APPROACH

We adopt the philosophy that any knowledge of the degradation of the binary image can and should be used in any process that tries to ascertain the content of the original: valuable prior information such as blur and noise and background and foreground colours should be used to increase the performance of the restoration algorithms.

The method we adopt for restoring binary document images has the following three steps: deblurring, interpolation and adaptive thresholding.

5.1. Deblurring

Deblurring is a huge field[6,24] that is principally concerned with the problem of deconvolution. Initially we have decided to use Tikhonov-Miller regularisation.[25] The solutions to the ill-posed inverse problem of deblurring are regularised with a smoothness constraint. If the degraded image g and the deblurred image $\hat{f}$ are written as lexicographically ordered vectors, then this constraint can be written as a block-circulant matrix S which is designed to behave like a high-pass filter, thus penalising unwanted high frequencies that would otherwise not affect the cost J because the blur B behaves like a low-pass filter:

$$J\left(\hat{f}\right) = \left\| g - B\hat{f} \right\|^2 + \lambda \left\| S\hat{f} \right\|^2 . \tag{1}$$

The least squares solution is:

$$\hat{f}(\lambda) = \left(B^{\mathrm{T}}B + \lambda S^{\mathrm{T}}S\right)^{-1} B^{\mathrm{T}}g. \tag{2}$$

This can be computed efficiently in the frequency domain via the DFT by approximating the block-circulant matrices by Toeplitz matrices. Note that $\hat{f}$ depends upon the regularisation parameter λ. Intuitively, the deblurred image will depend upon the relative weighting between the measurement error and the smoothness term in J. We use a standard technique (appendix A) for determining a suitable value.[26]

A smoothness regularisation operator is clearly not ideal for images in which much pertinent information is non-smooth. For example, images of text contain many edges which would tend to be smoothed in the regularised solution. This observation has yielded two possible extensions to this basic technique. The first concerns the study of regularising functionals which penalise non-smoothness in a convex but non-quadratic manner.[27,28] This has a clear computational disadvantage as the regularisation operator is no longer linear.

A second approach is to make λ spatially variant.[29] The idea here is to make λ adapt to the local properties of the image, reducing regularisation near edges, and increasing it in homogeneous regions of background and foreground colour. This has the effect of under-regularising in the vicinity of edges, but this is reasonable since it is known from psychophysical experiments that the human visual system is rather insensitive to noise around sharp intensity gradients.[30]

We achieve a similar effect in an efficient way. In a post-processing step, we use a statistical test on the local variance to decide whether the original camera image (i.e. before deblurring) is near an edge or not. We do this by calculating the sample variance s in the local M-pixel neighbourhood. The statistic $(M-1)s/\sigma^2$ is distributed as χ^2_{M-1}, where σ^2 is the variance of the additive noise. If it is less than a certain confidence limit, then the null hypothesis that the local neighbourhood is a homogeneous region is accepted, and we replace the deblurred pixel with the local average in the original camera image. This has the effect of suppressing the noise in the background (figure 5) where it is most noticeable.[30]

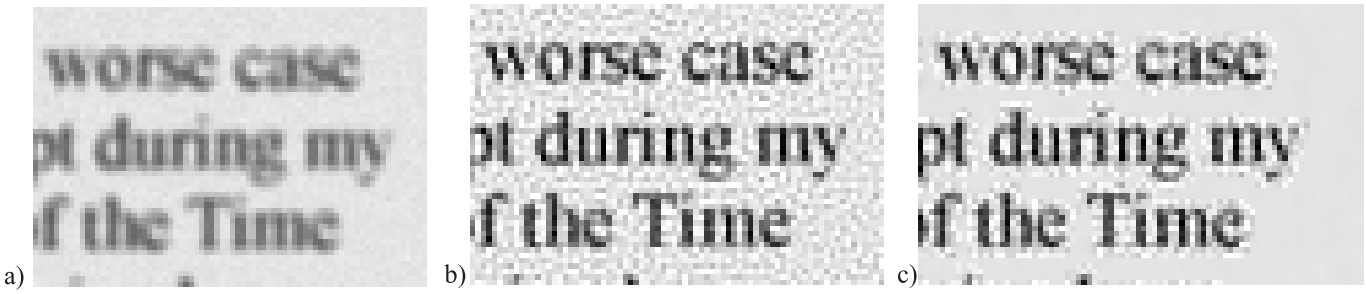

Figure 5. Regularisation and background noise suppression. *a) The original image from figure 3a subsampled and blurred with a cylindrical point spread function with radius $r = 1.5$ pixels and additive Gaussian noise with standard deviation of $\sigma = 3$ grey-levels. b) The result of regularisation. c) After background noise suppression*

5.2. Interpolation

We next perform bilinear interpolation. This is achieved by a separable operation where extra pixels are first fitted on a straight line between original pixels along a row, and then similarly for the pixels down each column. Arbitrary increases in the size of the image are possible, but we only attempt a factor of three in both dimensions.

5.3. Adaptive thresholding

The main task of this stage of the processing is to select a threshold for each pixel that separates local foreground from local background. This threshold surface typically has large scale structure due to lighting gradients, and small scale structure due to the blur still remaining after regularisation.

Niblack[1] assumes that the background/foreground polarity is known. He calculates the sample mean μ_M and standard deviation s_M in a small $M \times M$ neighbourhood, and then uses the threshold:

$$t_M(x, y) = \mu_M - k s_M \tag{3}$$

where $k \sim 0.2$ and $M \sim 15$ are found to yield good results for dark objects on a light background.[18] From (3) it can be seen that the value of k is used to adjust how much of the total print object boundary is taken as part of the given object. Negative values for k tend to make the dark objects thinner.

In homogeneous regions larger than M it is clear that (3) cannot be applied. Here the expected sample variance $E[s_M^2]$ quickly drops to σ^2, and the threshold becomes $\mu_M - 0.2\sigma$, producing noisy homogeneous regions. Negative k would make dark and light regions predominantly white. To prevent this, it is common to perform a post-processing operation where connected components are removed if the mean gradient of the grey-scale image around their boundaries is smaller than some threshold.[18]

Yanowitz and Bruckstein[2] fit a potential surface that passes through local maxima of gradient pixels, and use this to threshold the image. This surface is constructed with an iterative interpolation scheme that is somewhat computationally intensive to compute, but can potentially cope well with both small and large scale threshold surface structure.

6. EXPERIMENTS

We generate a binary image of 10 point Times New Roman text at 300 dpi. This is then degraded artificially to simulate the effect of a camera sampling at 100 ppi without lighting variation, but with different blurs and noise. We perform regularisation and interpolation, followed by adaptive thresholding. We set out to investigate whether the information pertinent to threshold estimation is best inferred from the local mean of *all* pixels after Niblack or from the edges after Yanowitz and Bruckstein. Having found the best information to use, we asses the sensitivity of the OCR error rate to blur and noise.

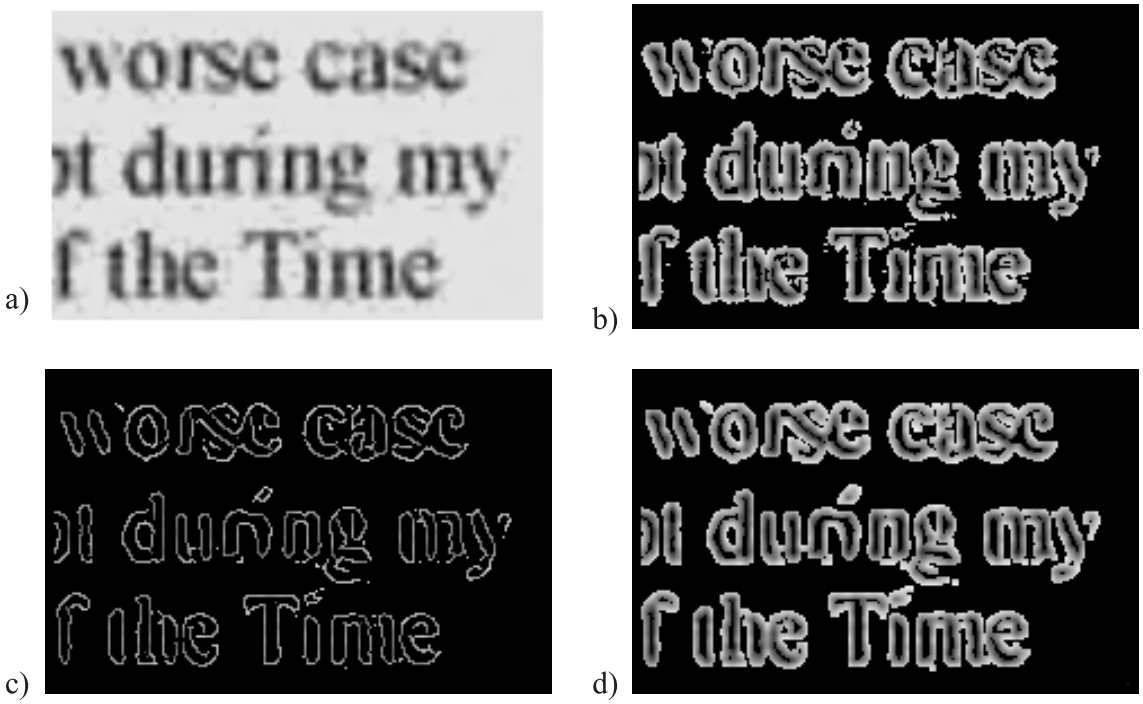

Figure 6. Estimation of local thresholds. *a) All pixels. b) Pixels with a local gradient above a threshold* $T = 100$. *c) Local maxima of gradient found using the Deriche operator. d) A dilated version of c) with dilation radius* $d = 1$.

6.1. Choice of local threshold

Figure 6 shows various alternatives for selecting pixels from which we can estimate a suitable local average. Figure 6a shows all the pixels in a region after interpolation. Figure 6b shows just those pixels with a Sobel gradient[9] that is greater than 160. Figure 6c shows edge pixels found by use of the Deriche operator,[31] and figure 6d shows these edge pixels after a dilation operation of radius $d = 1$. A local threshold for each pixel is estimated by calculating the mean of the non-zero pixels within a small neighbourhood (7×7) of the images shown in figure 6. This threshold is bounded so that the local threshold is not permitted to come within 3σ of the background or foreground levels to avoid noisy results in homogeneous regions. The thresholded images are sent to an OCR engine[†], and the resulting text file is compared with ground truth to produce an OCR error rate E.

The first experiment, shown in table 1, shows how E is affected by the gradient threshold T. The best results are obtained when $T = 0$ (figure 7a) corresponding to Niblack's method with $k = 0$. In this case, $E = 2.5\%$ which is similar to that obtained by using a flat-bed scanner at 200 dpi or a fax machine. Figure 7b shows the thresholded result with $T = 160$ where it can be seen that more characters are being split. This highlights the fundamental compromise between splitting single characters and merging neighbouring characters.

Table 1. OCR *error rate,* E*, as a function of gradient threshold,* T *(*$r = 1.5$*,* $\sigma = 3.0$*).*

T	0	10	40	160
E (%)	2.5	3.1	4.6	7.6

The second experiment, shown in table 2, shows how E is affected by the radius of the dilation operation d performed on the edges. The best results are obtained when $d > 3$ (figure 7d). This level of dilation covers most of the pixels

[†] TextBridge from Xerox

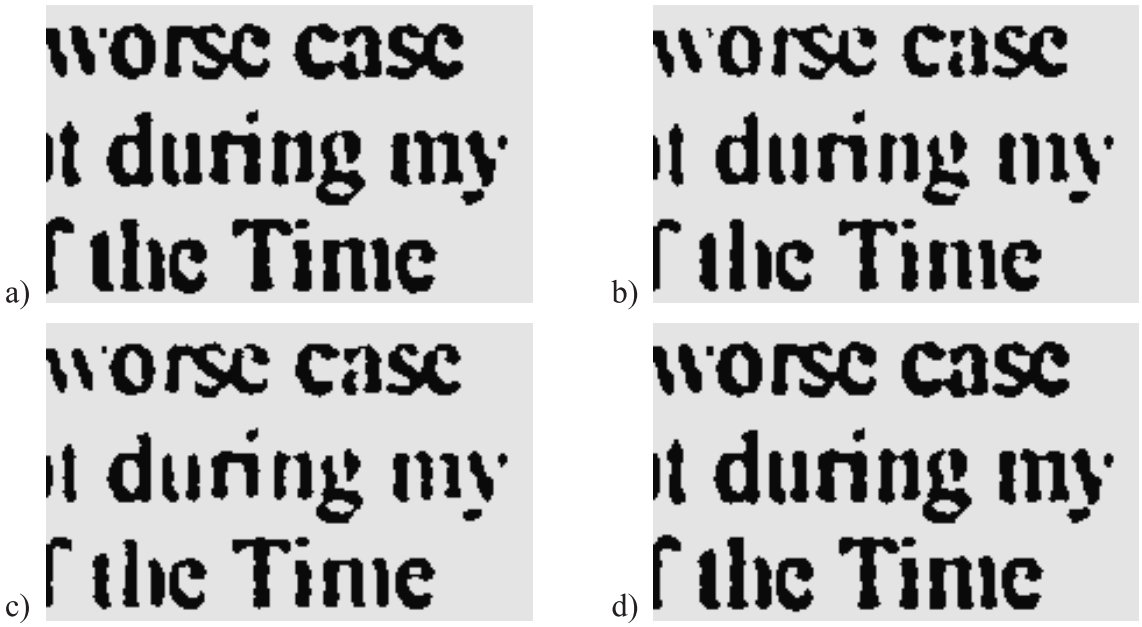

Figure 7. Thresholded results, *using a local average of: a) all pixels, b) pixels with a local gradient above a threshold $T = 160$. c) Local maxima of gradient found using the Deriche operator, d) a dilated version of c) with dilation radius $d = 1$.*

around the characters and thus closely approximates Niblack's method with $k = 0$. Figure 7c shows the thresholded result using just the edge pixels ($d = 0$). Here again it can be seen that characters are being split.

Table 2. OCR *error rate, E, as a function of the radius of dilation of Deriche high-gradient pixels, d ($r = 1.5$, $\sigma = 3.0$).*

Dilation radius (pixels)	0	2	4	6
E (%)	7.2	6.5	2.5	2.5

We conclude that using maxima of gradient does not aid the estimation of local thresholds for OCR.

6.2. Sensitivity to blur and noise

Having demonstrated that the best OCR results can be obtained from a threshold calculated from the local average of the deblurred and interpolated image, we now show how this method degrades with increasing blur and noise. Figure 8 compares the binary output with and without deblurring. Table 3 and figure 9a-b shows how E increases with the radius of the cylindrical point spread function. Table 4 and figure 9c-d shows how E increases with the standard deviation of the additive Gaussian noise.

Table 3. OCR error rate, E, as a function of blur radius, r ($\sigma = 3.0$).

r (pixels)	1.5	1.6	1.7	1.8	1.9
E (%)	2.5	3.8	9.0	11.0	14.0

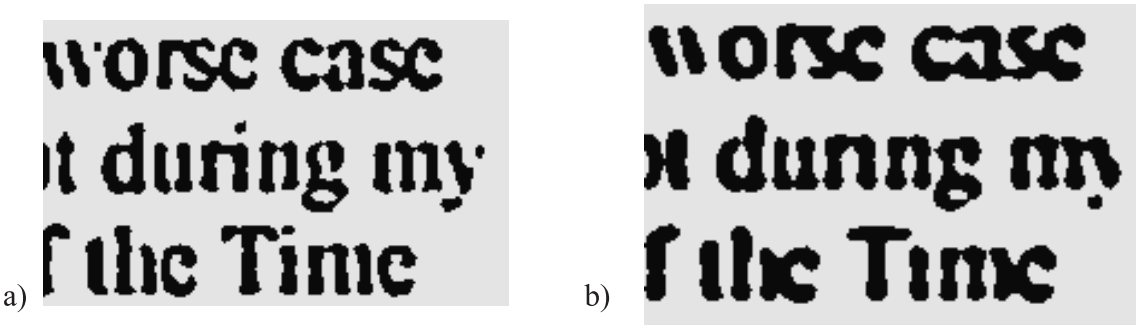

Figure 8. *Blur radius $r = 1.5$ pixels and $\sigma = 3$ grey-levels. a) Binary output with deblurring ($E = 2.5\%$). b) Binary output without deblurring ($E = 26.3\%$).*

7. CONCLUSIONS

Face-up scanning with digital cameras has the potential to provide a convenient and natural way of getting paper-based information into a computer. This new type of scanning requires careful image restoration for successful OCR. To this end, we have described a three stage process based upon Tikhonov-Miller regularisation, bilinear interpolation and adaptive thresholding using a local average technique. We have demonstrated that thresholding using edges in the grey-scale image does not aid OCR performance. Assuming that the page is scanned at 100 ppi with the blur and noise parameters described above, the use of a local average thresholding method after deblurring and interpolation yields an OCR performance comparable to a 200 dpi contact scanning process for bimodal images.

Table 4. OCR *error rate, E, as a function of noise standard deviation, σ ($r = 1.5$).*

σ	3.0	4.0	5.0	6.0
E (%)	2.5	8.4	11.0	14.0

APPENDIX A. CHOOSING THE SMOOTHING PARAMETER

The estimation of $\hat{f}$ is most efficiently carried out in the frequency domain. Representing the discrete Fourier transform with bold face characters, (2) becomes:

$$\hat{\mathbf{f}} = \frac{\mathbf{H}^*}{|\mathbf{H}|^2 + \lambda |\mathbf{S}|^2}\, \mathbf{g}. \tag{4}$$

Note that (4) becomes the inverse filter $1/\mathbf{H}$ as $\lambda \to 0$. In effect, (4) defines a one parameter set of regularised solutions, parameterised by λ. Writing the expected restoration error as[26]:

$$\epsilon = E\left[\left\| f - \hat{f}(\lambda)\right\|^2\right] = \|f\|^2 + E\left[\left\|\hat{f}(\lambda)\right\|^2\right] - 2E\left[f^{\mathrm{T}}\hat{f}(\lambda)\right] \tag{5}$$

and noting that $\|f\|^2$ does not depend upon λ, we note that the minimisation of ϵ is equivalent to the minimisation of the last two terms in (5). Using Parseval's theorem:

$$E\left[\left\|\hat{f}(\lambda)\right\|^2\right] = E\left[\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \frac{|\mathbf{H}|^2\ |\mathbf{g}|^2}{(|\mathbf{H}|^2 + \lambda|\mathbf{S}|^2)^2}\right]. \tag{6}$$

and it can also be shown that[26]:

$$E\left[f^{\mathrm{T}}\hat{f}(\lambda)\right] = E\left[\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \frac{|\mathbf{H}|^2\ |\mathbf{f}|^2}{|\mathbf{H}|^2 + \lambda|\mathbf{S}|^2}\right] = E\left[\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \frac{|\mathbf{g}|^2 - \sigma^2}{|\mathbf{H}|^2 + \lambda|\mathbf{S}|^2}\right]. \tag{7}$$

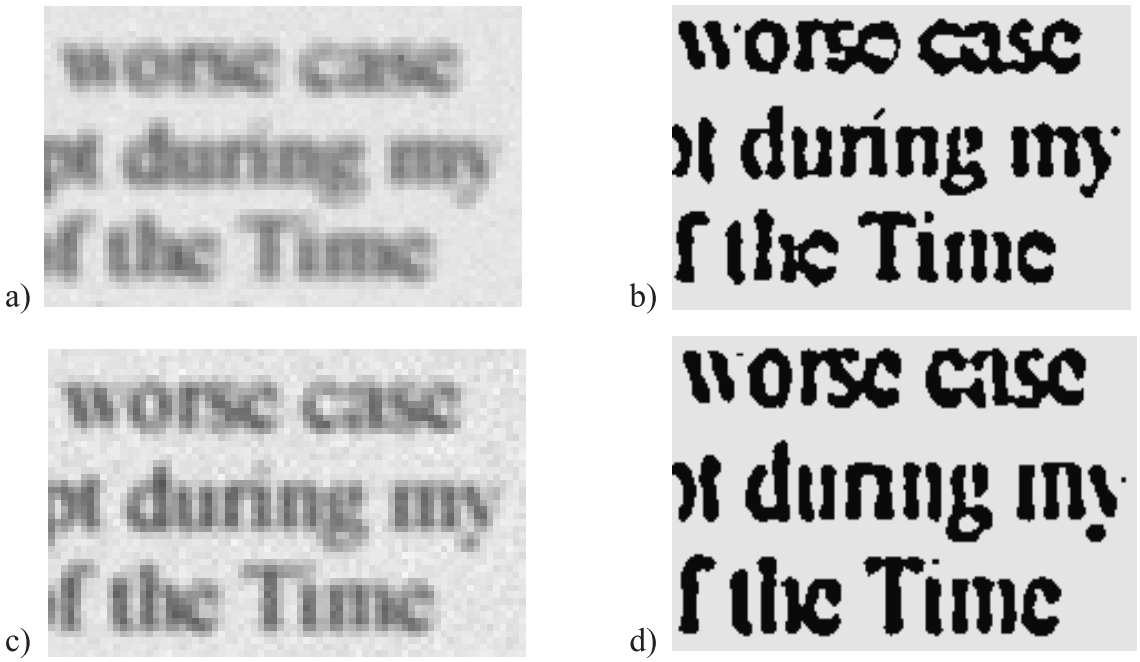

Figure 9. *a) Blur radius $r = 1.9$ pixels and $\sigma = 3$ grey-levels, and b) its restored binary image. c) Blur radius $r = 1.5$ pixels and $\sigma = 6$ grey-levels, and b) its restored binary image.*

Note that the expectations (6) and (7) are independent of the unknown f. Hence, removing the expectation operators, we can write

$$\hat{\epsilon}(\lambda) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \frac{|\mathbf{H}|^2\ |\mathbf{g}|^2}{\left(|\mathbf{H}|^2 + \lambda|\mathbf{S}|^2\right)^2} - 2\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \frac{|\mathbf{g}|^2 - \sigma^2}{|\mathbf{H}|^2 + \lambda|\mathbf{S}|^2}. \tag{8}$$

Note that this requires knowledge of σ. It can be shown that this is unimodal and so we can use a gradient descent optimisation algorithm on $\epsilon(\lambda)$ using (8) to find the best λ.

ACKNOWLEDGMENTS

Thanks go to A. Zappala, W. Newman, S. Taylor and A. Gee for numerous helpful discussions

REFERENCES

1. W. Niblack, *An Introduction to Digital Image Processing*, pp. 113–116. Prentice Hall, 1986.
2. S. Yanowitz and A. Bruckstein, "A new method for image segmentation," *Computer Vision, Graphics and Image Processing* **46**, pp. 82–95, April 1989.
3. P. Wellner, "Interacting with paper on the digital desk," *Communications of the ACM* **36**, pp. 87–97, July 1993.
4. K. O'Hara, F. Smith, W. Newman, and A. Sellen, "Student readers' use of library documents: Implications for digital library technologies," in *Proceedings of CHI - to appear, Los Angeles, CA.*, 1998.
5. W. Newman and P. Wellner, "A desk supporting computer-based interaction with paper documents," in *CHI '92 Human Factors in Computing Systems, Monterey CA*, pp. 373–380, May 1992.
6. J. Biemond, R. Lagendijk, and R. Mersereau, "Iterative methods for image deblurring," *Proceedings of the IEEE* **78**, pp. 856–883, May 1990.
7. A. Savakis and H. Trussell, "Blur identification by residual spectral matching," *IEEE Trans. on Image Processing* **2**, pp. 141–151, April 1993.
8. M. W. Burke, *Image Acquisition, Handbook of Machine Vision Engineering Vol. 1*, pp. 648–666. Chapman and Hall, 1996.

9. A. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.
10. D. Kundar and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine* **13**, pp. 43–64, May 1996.
11. R. Lagendijk, A. Tekalp, and J. Biemond, "Maximum likelihood image and blur identification: a unifying approach," *Optical Engineering* **29**, pp. 422–435, May 1990.
12. R. Lagendijk, J. Biemond, and D. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximisation algorithm," *IEEE Trans. on Image Processing* **38**, pp. 1108–1191, July 1990.
13. Y. You and M. Kaveh, "A regularisation approach to joint blur identification and image restoration," *IEEE Trans. on Image Processing* **5**, pp. 416–428, July 1996.
14. B. Chalmond, "PSF estimation for image deblurring," *CVGIP: Graphical Models and Image Processing* **53**, pp. 364–372, July 1991.
15. R. Fabian and D. Malah, "Robust identification of motion and out-of-focus blur parameters from blurred and noisy images," *CVGIP: Graphical Models and Image Processing* **53**, pp. 403–412, September 1991.
16. S. Reeves and R. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Trans. on Image Processing* **1**, pp. 301–311, July 1992.
17. F. Foley and A. van Dam, *Computer Graphics – Principles and Practice, Second edition*, pp. 623–647, 976–979. Addison-Wesley, 1990.
18. O. Trier and A. Jain, "Goal-directed evaluation of binarization methods," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17**, pp. 1191–1201, December 1995.
19. P. Sahoo, S. Soltani, A. Wong, and Y. Chen, "A survey of thresholding techniques," *Computer Vision, Graphics and Image Processing* **41**, pp. 233–260, 1988.
20. J. Mitchell, W. Pennebaker, D. Anastassiou, and K. Pennington, "Graphics image coding for freeze-frame videoconferencing," *IEEE Trans. on Communications* **37**, pp. 515–521, May 1989.
21. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **6**, pp. 721–741, November 1984.
22. G. Wolberg and T. Pavlidis, "Restoration of binary images using stochastic relaxation with annealing," *Pattern Recognition Letters, Elsevier* **3**, pp. 375–388, December 1985.
23. L. Wang and T. Pavlidis, "Direct gray-scale extraction of features for character recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **15**, pp. 1053–1067, October 1993.
24. M. Sezan and A. Tekalp, "Survey of recent developments in digital image restoration," *Optical Engineering* **29**, pp. 393–404, May 1990.
25. A. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*, Wiley, 1977.
26. P. Galatsanos and A. Kastaggelos, "Methods for choosing the regularisation parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **1**, pp. 322–336, July 1992.
27. R. Stevenson, B. Schmitz, and J. Delp, "Discontinuity preserving regularisation of inverse problems," *IEEE Trans. on Systems, Man and Cybernetics* **24**, pp. 455–469, March 1994.
28. R. Schultz and R. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. on Image Processing* **3**, pp. 233–241, May 1994.
29. R. Lagendijk, J. Biemond, and D. Boekee, "Regularised iterative image restoration with ringing reduction," *IEEE Trans. on Acoustics, Speech and Signal Processing* **36**, pp. 1874–1887, December 1988.
30. G. Anderson and A. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. on Systems, Man and Cybernetics* **6**, pp. 845–853, March 1994.
31. O. Faugeras, *Three-Dimensional Computer Vision*, MIT Press, Cambridge, Mass., 1993.