

Automatic Segmentation and Descreening of Scanned Color Documents

A. Ravishankar Rao and Gerhard Thompson
IBM Research, Yorktown Heights, New York

Abstract

Printing books-on-demand is a new technology that is revolutionizing the book printing and publishing industry. One of the biggest bottlenecks in this process is the conversion of existing books into digital form. This typically involves digitization of original books through scanning, which is a slow and labor-intensive process. Careful attention must be paid to maintain the quality of the reproduced books and in particular of the images they contain. Halftoned image areas in the original books cause the most reproduction problems, as there is the potential that moiré patterns may form when these image areas are re-screened.

In order to avoid these moiré patterns, it is necessary to detect the image areas of the document and remove the screen pattern present in those areas. In the past, we have presented techniques to perform these operations in the case of grayscale images. In this paper, we extend these techniques to handle color images. We present efficient and robust techniques to segment a color document into halftone image areas, detect the presence and frequency of screen patterns in halftone areas and suppress the detected screens. Halftoned image areas are segmented by using a measure of image activity; image activity is low in text areas and high in halftoned areas. We use 2D Fourier spectral analysis to identify the screen frequencies present. The screens are then suppressed by low-pass filtering.

Our technique speeds up the conversion process of books to digital form, and overcomes quality problems in the reproduction of halftoned images.

Introduction

Printing books-on-demand is a new technology that is revolutionizing the book printing and publishing industry. Instead of printing thousands of copies of a book and storing and distributing each copy, publishers can create a digital version of the book and print exactly the desired number of copies. This solution eliminates inventory costs and the need to forecast expected demand for a book. Furthermore, a book can never go out of print as a digital version persists indefinitely.

A significant number of books exist for which there is no available digital version. Such books must be scanned and converted to a digital representation. The digital representation can vary from being a facsimile copy of the original book (i.e. each page is represented as an image) to being an OCRed and tagged version of the original book. Facsimile digital representations are easier to create, but are limited in their ability to provide searchable content.

Regardless of the final digital representation used, an important objective of the conversion process is to maintain the quality of the reproduced books, particularly in the image areas. Typically, printed text does not pose reproduction problems. However, halftoned image areas are challenging to reproduce, as there is potential for the formation of moiré patterns during scanning and post-processing steps such as image scaling, re-sampling and printing.

A common technique to process scanned halftoned images is to apply an operation called descreening, which involves blurring the image with a low-pass filter. This effectively blurs the screen so it is not visible. The drawback of applying this operation is that image features themselves may get blurred. There is a tradeoff between the removal of the screen and the blurring of image features. Hence it is important to choose the size of the blur filter to be the minimum required to blur the screen. Since screens of varying frequency occur in printed matter, it is necessary to dynamically estimate the screen frequency, which in turn determines the size of the blur filter.

The reproduction requirements for text areas are different from halftoned areas. Text areas should appear sharp, with boundaries of characters properly delineated. Thus it is undesirable to blur the text areas in the same way the halftone areas need to be blurred. Due to these differing requirements, each scanned page must be segmented into text and halftone image areas. The descreening operation is applied exclusively to the halftoned image areas. The remainder of this paper covers these operations in more detail.

Previous work by Rao et al.[1] and Jaimes et al.[2] described techniques to perform segmentation and descreening operations on bi-level and grayscale images. This paper extends those techniques to handle color images. Cheung and Ulichney,[3] Eschbach,[4] and Ng[5] present techniques for moiré reduction.

Page Segmentation

We have adapted the grayscale page segmentation technique described in Jaimes et al.[2] to handle color images. Since the color image is composed of three channels, it is advantageous to perform the segmentation in a single channel. Different choices are possible in the creation of this single channel such as L* or hue. We offer another possibility, which is to use the red channel. By empirical observation we have noted that the red channel typically carries the relevant information about halftone screens that can be used to distinguish halftoned regions from non-halftoned regions, i.e. text regions. The red channel contains the highest contrast information which makes it the easiest channel to use with our segmentation technique. The L* channel, on the other hand, combines information from all the overlapping halftone color screens and has less clearly defined whites and blacks.

Figure 1 depicts an overview of the processing stages used in our algorithm.

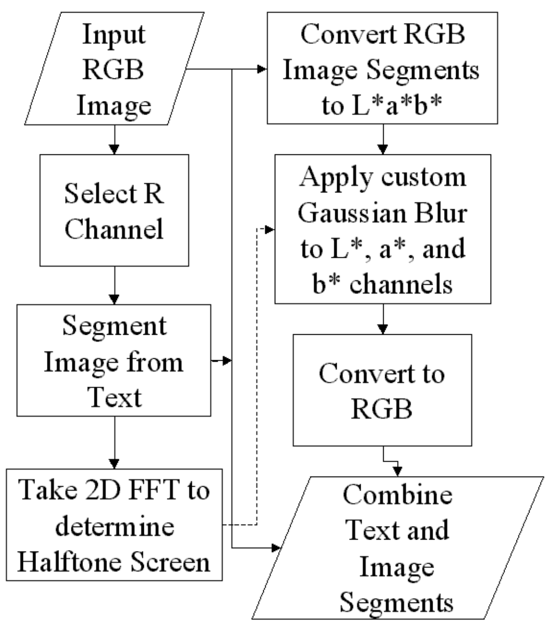

Figure 1. A flowchart of the algorithm.

Halftoned regions are distinguished from non-halftone regions based on a measure of image spatial frequency. The intensity variations in halftoned areas have higher spatial frequency than text areas. Thus a measure of spatial frequency can be used to discriminate text vs. halftoned areas.

Figure 2 top shows a scanned test image with text and halftone areas. Though a 2D FFT can be used to measure spatial frequency over the entire image, it is a computationally costly operation. We instead use a measure of the number of times the image intensity function crosses the mean image intensity in a single dimension. We term this measure the alpha-crossing, where alpha is the mean image intensity over a local region, say a 25 by 25 window. This is computationally simpler and yields appropriate results, as described in [2].

The red channel is analyzed to determine regions having high frequency based on the alpha-crossing density. Regions having high frequency are associated with halftone image areas. Rectangular bounding boxes are created around the halftoned image areas as shown in Figure 2 (middle). Further processing is confined to these halftoned image areas.

Determination of Screen Frequency

Since the screen frequency of scanned halftones may not be known in advance, it is best for this frequency to be determined dynamically. This way the descreening filter size is adapted to the precise screen frequency present. We use the 2D FFT to determine the dominant screen frequency in a color image. Since the 2D FFT is computationally expensive, we apply it to small regions within the halftoned image. The considerations here are that an adequate area of the halftoned image be sampled while keeping the computational costs down. Hence we calculate the 2D FFT over non-overlapping 128x128 windows along the diagonal of the image areas. The small size of the window keeps computational costs down, while the use of multiple windows ensures an adequate area is used for estimation.

Let $\boldsymbol{S}$ denote the 2D FFT of the image $\boldsymbol{s}$. The power spectrum, $\boldsymbol{P}$, is computed as,

$$P = S\,S^*$$

The power spectrum shows the distribution of spectral energy at various spatial frequencies. A typical plot is shown in Figure 3.

Observe that a halftoned area will show significant spectral peaks located at the screen frequencies. Additional peaks are located at harmonics of these screen frequencies. For the purpose of estimating the screen frequency, we locate the position of the highest peak. We also compute a measure of confidence that the power spectrum does possess strong peaks as opposed to a flat distribution where all frequencies have equal energy. A simple measure of non-uniformity in the power spectrum is the ratio of the highest peak to the next-highest peak. For highly peaked distributions this ratio will be large, say greater than 10. For uniform distributions, this ratio will tend to 1.

The measure of confidence is used to weight the different spectral peaks found in the 128x128 windows. The result is a distance from the origin to the dominant frequency. It is sufficient to eliminate this dominant frequency. This is done by using a Gaussian filter which is chosen to reduce this spectral peak to some pre-determined value, say 1%. We use Gaussian filters because the combination of screens in color printing gives rise to patterns with circular symmetry. Gaussian filters are efficiently implementable as a convolution of two separable 1D kernels.

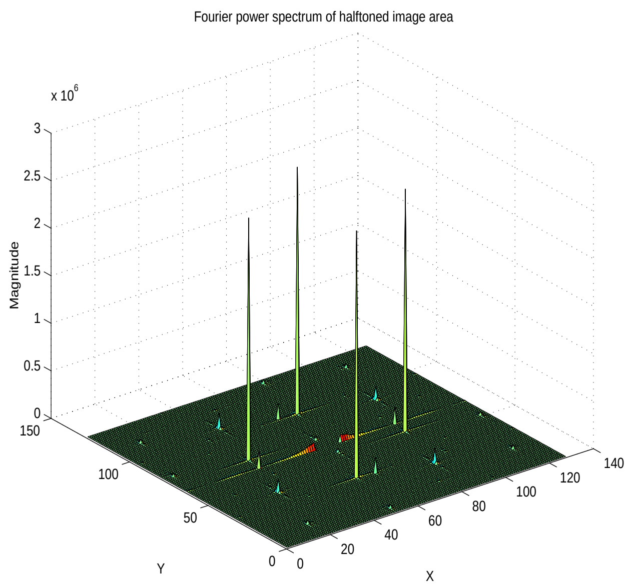

Figure 3: The Fourier power spectrum (magnitude) of the halftoned area.

If halftoned image areas are present after segmentation, the image is converted to an L*a*b* representation. Each of the L, a, b channels are blurred with a Gaussian filter, as described above, and then recombined to create the final descreened RGB image.

Finally, the halftoned image is replaced with its descreened version, while the text areas stay the same. The resulting page can now be processed without the fear of moiré patterns occurring.

Results

Figure 2, bottom, illustrates the steps of our process with a scanned color image. Due to the effective descreening of the color halftone, no visible moiré patterns are present in the final rehalftoned output.

We experimented with several images containing mixed halftone and text content, and obtained excellent results with our algorithm.

Conclusion

In this paper we presented efficient and robust techniques to segment a color document into halftone image areas and text areas. Halftoned image areas are segmented from text areas by using a measure of image activity in the R channel. We implemented an algorithm using 2D Fourier spectral analysis to detect the frequency of screen patterns in the halftoned image areas. The screens are then suppressed by low-pass filtering, where the filter size is a function of the detected screen frequency.

Our technique speeds up the conversion process of books to digital form, and overcomes quality problems in the reproduction of halftoned images.

References

1. A. R. Rao, F. Mintzer, G. Thompson, "Descreening printer-ready images," IS&T NIP14 Conference, International conference on digital printing technologies, Oct. 1998, pp. 285-289.
2. A. Jaimes, A. R. Rao, F. Mintzer, G. Thompson, ``Segmentation and automatic descreening of scanned documents," SPIE Vol. 3648, Color Imaging: Device Independent Color, Color Hardcopy and Graphic Arts IV, Jan 1999, pp. 517-528.
3. Shiufun Cheung, Robert A. Ulichney, "Low-Memory Low-Complexity Inverse Dithering," SPIE Vol. 3648, Jan 1999, pp. 529-539.
4. Reiner Eschbach, "Digital Copying of Medium Frequency Halftones," SPIE Vol. 3300, Jan 1998, pp. 398-404.
5. Yee S. Ng, "Rule-based image segmentation method", IS&T PICS Conference, March 2000, pp. 269-274.

Biography

A. Ravishankar Rao is a research staff member at the IBM T.J. Watson Research Center. His research interests include image processing, computer vision and data mining. He chaired the SPIE Conference on Machine Vision and Applications from 1996-98. He is an Associate Editor of the journals Pattern Recognition and Machine Vision and Applications. He received a Ph.D. degree in computer engineering from the University of Michigan, Ann Arbor, in 1989. He is a Senior Member of the IEEE and a member of SPIE and IS&T. His work has resulted in three patents, over twenty-five publications, and a book entitled "A Taxonomy for Texture Description and Identification." (email: ravirao@us.ibm.com)

Gerhard Thompson joined IBM in 1968 and worked on a variety of projects related to LSI chips including performance evaluation, software and hardware modeling, and circuit and logic design.

Since 1982 he has been with Image Applications at the IBM T.J. Watson Research Center. Here he developed image processing algorithms in the fields of binary and gray-scale rotation, scaling and digital halftoning for both monochrome and color printers. He hold numerous patents in these areas.

Mr. Thompson received a B.S.E.E. degree from Drexel University in 1967, an M.S.E.E. from Princeton in 1968, and an M.S. in Computer Science from Syracuse University in 1983. (email: gerryt@us.ibm.com)

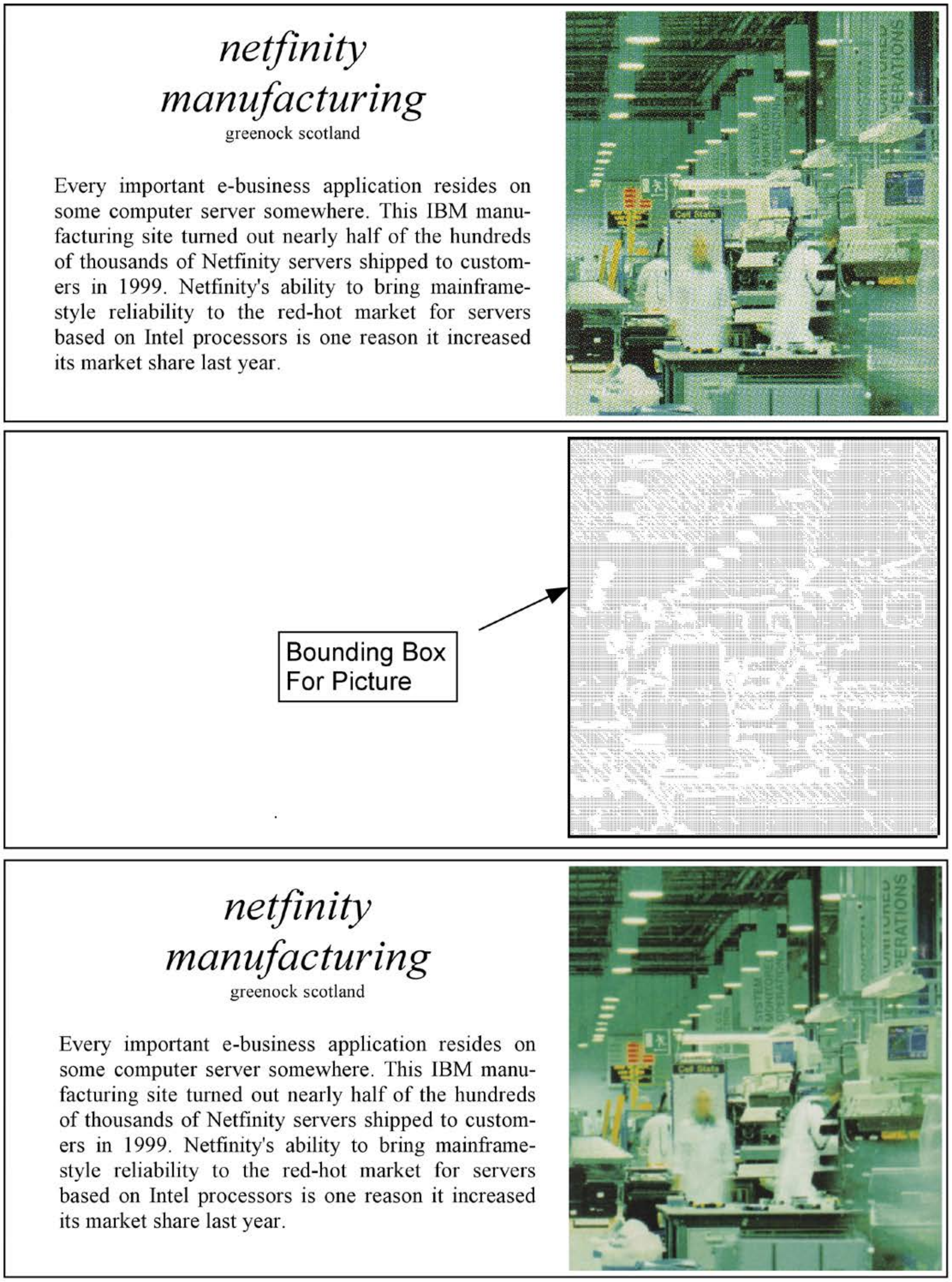

Figure 2: Original image (top), density map with bounding boxes (middle), and result of descreening (bottom).