 Copyright 2000, IS&T

Content-based Document Enhancement and Resizing

Mohamed N. Ahmed, Brian E. Cooper, and Shaun T. Love
Lexmark International, Inc.
Lexington, KY, USA

Abstract

Recent advances in information and communications technologies have increased the need for automated reading and processing of documents. Most of today's documents contain not only text and background, but also graphics, tables, and images. Common image enhancement and interpolation methods apply an interpolation or enhancement function indiscriminately to the whole image. The resulting document image suffers from objectionable moiré patterns, edge blurring and aliasing. Therefore, scanned documents must often be segmented before other document processing techniques, such as filtering, resizing, and compression can be applied. In this paper, we present a new system to segment and label document images into text, halftone images, and background using feature extraction and unsupervised clustering. Once the segmentation is performed, a specific enhancement or interpolation kernel can be applied to each document component. Each pixel is assigned a feature pattern consisting of a scaled family of differential geometrical invariant features and texture features extracted from the co-occurrence matrix. The invariant feature pattern is then assigned to a specific region using a two-stage neural network system. The first stage is a self-organizing principal components analysis (SOPCA) network that is used to project the feature vector onto its leading principal axes found by using principal components analysis (PCA). The next step is to cluster the output of the SOPCA network into different regions. This is accomplished using a self-organizing feature-map (SOFM) network. In this paper, we demonstrate the power of the SOPCA-SOFM approach to segment document images into text, halftone, and background. The proposed filtering and interpolation method results in a noticeable improvement in the enhanced image.

1. Introduction

Digital copying, in which a digital image is obtained from a scanning device and then printed, involves a variety of inherent factors that compromise image quality. Ordered halftone patterns in the original document interact with the periodic sampling of the scanner, producing objectionable moiré patterns. These are exacerbated when the copy is reprinted with an ordered halftone pattern. In addition, limited scan resolution blurs edges, degrading the appearance of detail such as text. Fine detail also suffers from flare, caused by the reflection and scattering of light from the scanner's illumination source. Flare blends together nearby colors, blurring the high-frequency content of the document. Typical examples of scanned documents are shown in Figure 1.

To suppress moiré, a filter may be constructed that is customized to the frequencies of interest. However, both the detection of the input halftone frequencies and the frequency-domain filtering itself can require significant computational effort. Although crude, a simple, small low-pass filter can correct the majority of moiré artifacts. Unfortunately, low-pass filtering affects detail as well, blurring it even further.

Sharpening improves the appearance of text and fine detail, countering the effects of limited scan resolution and flare. Edges become clear and distinct. Of course, other artifacts such as noise and moiré become sharper as well. The solution is simple in concept: determine the content of regions within the scanned image and then apply the appropriate filter to each region. Sharpening should be performed on fine detail, while moiré suppression should be applied to certain periodic artifacts.

In previous work [1], document image quality was improved significantly using this approach. After segmenting the image, a particular enhancement algorithm was selected according to the classification of each pixel. However, the enhanced image had the same resolution as the original image.

Image resizing is also used extensively in picture processing to magnify or reduce images. Standard approaches fit the original data with a continuous model and then resample this two dimensional function on a new sampling grid. In the case of nearest neighbor interpolation, the underlying image model is a polynomial spline of order zero (piecewise constant). This is extremely simple to implement but tends to produce images with a "blocky" appearance, in which the large, individual blocks of expanded pixels become objectionally apparent. A smoother reconstruction can be obtained with bilinear interpolation. Bicubic spline interpolation yields even better results, providing smooth transitions that are not as blurry [2]. These interpolation methods, however, apply an interpolation function indiscriminately to the whole image. As shown in Figure 2, the resulting document image in general suffers from objectionable edge blurring and aliasing.

From the above discussion, we can conclude that for an image enhancement system to work properly, a preprocessing step should include a segmentation of the document into text, halftone and background. If this step is successfully completed, the application of an appropriate filter/interpolation should be straightforward.

Several approaches for document segmentation have been proposed [3]-[8]. These techniques can be broadly classified as bottom-up or top-down. Bottom-up methods start from the pixel level and merge regions together into larger and larger components. Top-down techniques apply a priori knowledge about the page to hypothesize and split the page into blocks which are subsequently identified and further subdivided. Top-down approaches work well with pre-specified layouts such as technical papers. However, the performance of these techniques degrades significantly when different components are touching or overlapping. Among bottom-up approaches, texture-based schemes have attracted much attention [6]-[8]. These methods treat different components of a document image as different textures. The scanned document images are convolved with a set of masks to generate feature vectors. Each feature vector is then classified into different classes using a pre-trained classifier. One problem associated with these approaches is the mask size for extracting local features. If the mask size is too small, it is difficult to detect large scale textures such as large fonts. On the contrary, if a large mask is chosen, the computational complexity will increase dramatically [8]. In this paper, we propose to combine texture features with a multiscale representation, presented in the next section, to exploit both local features and image structure.

2. System Description

The presented method consists of the following steps: feature extraction, feature selection, and clustering and labeling.

2.1. Feature Extraction

As shown in Figure 3, the first step in our approach is to extract a feature pattern for each pixel in a gray level image. A feature vector $\mathbf{X}$ is a set of measurements $\{x_1, x_2, \cdot\cdot, x_d\}$ which condenses the description of relevant properties of the image into a small, Euclidean feature space of d dimension. The number of needed features depends on the complexity of the image. The components of the feature vector may include gray values, gray values through different filters, texture measures, Markov random field features, fractal dimension measures, and gradient magnitudes and directions [7]. In this paper, we will construct our feature vector from measurements obtained from gray level distribution (first order gray level parameters), texture measures estimated from the gray level co-occurrence matrix (second order gray level parameters), and multiscale features. The reason behind the selection of these types of features is their ease of implementation and their strong discriminating power. These feature patterns will then undergo a process of feature selection to extract the most discriminating features for image segmentation.

2.1.1. Multiscale features

Multiscale image representations are a powerful tool for analyzing many image features. An image is decomposed into a set of descriptions, each making explicit image features at a specific scale. The scale space, a one-parameter family of blurred replicas of the input image, is based on the diffusion equation and was proposed by Witkin [9] and Koenderink [11] as the image representation for multiscale analysis. A linked set of image replicas has been called a *stack* of 2D images.

The stack $H = \{I_1, I_2, ...I_N\}$ consists of a set of replicas I_j of the original image I and its derivatives, blurred by a Gaussian kernel of increasing width (σ)

$$I_j = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\alpha,\beta)\ \times G_\sigma(x-\alpha, y-\beta)\, d\alpha\, d\beta. \tag{1}$$

2.1.2. First order gray level parameters

In this category, the features for a pixel are derived from the gray level distribution in a window, $\mathcal{W}$, of dimension $w \times w$ centered around each pixel. These features describe the first order gray level distribution without considering the spatial interdependence. Two features were selected:

1. The mean gray level, μ,

$$\mu = \frac{1}{w^2} \sum_{(x,y)\in\mathcal{W}} I(x,y) \qquad (2)$$

where $I(x,y)$ is the gray level at location (x,y).

2. The variance σ^2 of the gray level variation

$$\sigma^2 = \frac{1}{w^2-1} \sum_{(x,y)\in\mathcal{W}} (I(x,y)-\mu)^2. \qquad (3)$$

2.1.3. Texture measures

Texture is a property of the spatial distribution of gray levels or the overall pattern of gray level changes in an image. Texture properties can be derived using several approaches such as a gray-level co-occurrence matrix, first-order gradient distribution, edge co-occurrence matrix, or run-length matrix [13].

Several texture-based approaches for document segmentation have been presented [6]-[8]. These methods treat different components of a document image as different textures. The scanned document images are convolved with a set of masks to generate feature vectors. Each feature vector is then classified into different classes using a pre-trained classifier. One problem associated with these approaches is the mask size for extracting local features. If the mask size is too small, it is difficult to detect large scale textures such as large fonts. On the contrary, if a large mask is chosen, the computation complexity will increase dramatically [8]. In this paper, we propose to combine texture features with a multiscale representation, presented in the next subsection, to exploit both local features and image structure.

In this work, we extracted our texture parameters from the gray level co-occurrence matrix described in Haralik [13]. The co-occurrence matrix is based on the estimation of the second-order joint conditional probability density function, $C(i,j|d,\theta)$, which is the probability of going from gray level i to gray level j, given the inter-sample spacing d and the angle θ.

We calculated a gray level co-occurrence matrix for each pixel neighborhood $\mathcal{W}$, and we derived five measures that contain information of image texture characteristics such as homogeneity, gray-level linear dependencies, and contrast [13].

3. Feature Selection

A key problem encountered in statistical pattern recognition is that of feature selection. Feature selection refers to a process whereby a data space is transformed into a feature space that, in such a way that the data set may be represented by a reduced number of "effective" features and yet retain most of the intrinsic information content of the data. Principal Components Analysis (PCA) is perhaps the oldest and best-known technique in multivariate analysis [15]. The practical value of PCA is that it provides an effective technique for dimensionality reduction. The first stage in our two-stage network is a neural network that performs principal components analysis of arbitrary size on the input vector. As shown in Figure 3, this network is a feedforward network composed of a single layer of linear neurons. The only aspect of the network that is subject to training is the set of synaptic weights $\{w_{ji}\}$ connecting source node i in the input layer to computation node in the output layer, where $i = 0,1,\ldots,n-1$, and $j = 0,1,\ldots,m-1$.

The output $y_j(k)$ of neuron j at time k, produced as the response to the set of inputs
$\mathbf{X}(k) = \{x_0(k), x_1(k), \ldots, x_{n-1}(k)\}$, is given by

$$y_j(k) = \sum_{i=0}^{n-1} w_{ji}x_i(k), \qquad j = 0,1,\ldots,m-1. \qquad (4)$$

The synaptic weights $\{w_{ji}\}$ is updated in accordance with a generalized form of Hebbian learning,

$$\Delta w_{ji}(k) = \eta[y_j(k)x_i(k) - y_j(k)\sum_{h=0}^{j} w_{hi}(k)y_h(k)], \quad i = 0,1,\ldots,n-1, \;\; j = 0,1,\ldots,m-1 \qquad (5)$$

where $\Delta w_{ji}(k)$ is the change applied to the synaptic weight $w_{ji}(k)$ at time k, and η is the learning-rate parameter.

3.1. Self-Organizing Feature-Mapping (SOFM)

In this section we will develop the SOFM algorithm to cluster our modified feature vector, $\mathbf{y}$, into different clusters corresponding to different document regions.

Consider the network in Figure 3, which depicts a one-dimensional array of neurons. The input vector, representing the set of input signals, is denoted by

$$\mathbf{y} = \begin{pmatrix} y_0 & y_1 & \cdot & \cdot & y_{m-1} \end{pmatrix}^T. \qquad (6)$$

The synaptic weight vector of neuron j is denoted by

$$\mathbf{V}_j = \begin{pmatrix} v_{j0} & v_{j1} & \cdot & \cdot & v_{j(m-1)} \end{pmatrix}^T \quad j = 1,2,\ldots,L. \qquad (7)$$

To find the best match of the input vector $\mathbf{y}$ with the synaptic weight vectors $\mathbf{V}_j$, we simply compare the inner products $\mathbf{V}_j{}^T\mathbf{y}$ for $j = 1, 2, \ldots, L$ and select the largest. In the formulation of an adaptive algorithm, it is more convenient to normalize the weight vectors $\mathbf{v}_j$ to a constant Euclidean norm [15]. In such a situation, the best matching criterion described here is equivalent to the minimum Euclidean distance between vectors. If we denote the neuron that best matches the input vector $\mathbf{y}$ by $r(\mathbf{y})$, then we may determine $r(\mathbf{y})$ by applying the condition

$$r(\mathbf{y}) = min\{||\mathbf{y} - \mathbf{V}_j|| \quad such\ \ that \quad j = 1, 2, \ldots, L\}. \tag{8}$$

The weight vectors of the winning neuron $r(\mathbf{y})$, as well as all the neurons that lie in a neighborhood $\Omega_{r(\mathbf{y})}(k)$, are pulled into the direction of the input pattern. This gives the Kohonen learning rule

$$\mathbf{V}_{r(\mathbf{y})}^{k+1} = \mathbf{V}_{r(\mathbf{y})}^{k} + \eta^k(\mathbf{y} - \mathbf{V}_{r(\mathbf{y})}^{k}) \tag{9}$$

and

$$\mathbf{V}_j^{k+1} = \mathbf{V}_j^k + (\eta^k)^2(\mathbf{y} - \mathbf{V}_j^k) \qquad \forall j \in \Omega_{r(\mathbf{y})}(k) \tag{10}$$

where $0 < \eta^k < 1$ is the learning rate at time k.

4. Interpolation

Bicubic interpolation [2] provides a reasonably simple but effective method for enlarging many images. Transitions between expanded pixels remain smooth, and edge content is preserved better than with bilinear interpolation. Nevertheless, detail in document images remains overly blurry, even with bicubic interpolation. Rather than trying increasingly complex interpolation algorithms, we propose integrating the interpolation along with image enhancement, after segmentation.

Consider the simple example shown in Figure 4, in which two classes are present. The left column of pixels (shown as black dots) represent a text region, while the other two columns (shown as gray dots) represent a background region. (Beneath the classified pixels are the corresponding numeric values of the pixels.) The left figure uses conventional interpolation, in which all neighboring pixels contribute to the interpolated pixel value, 81. Note, however, that the interpolated pixel falls within the background region. More reliable results may be obtained when the interpolated pixel is derived only from neighboring pixels within the background class. This approach is presented in the right figure. In this case, the interpolated pixel has a value of 123. By restricting the contributing pixels to the same class, blurring between boundaries is considerably reduced.

5. Results

We tested our algorithm with several images scanned at 300 dpi. Thirty different features were extracted for every pixel. These features included first order statistics, texture measures at different orientations, and multiscale features. The images were segmented into text, images, and background. Results are shown in Figures 5-7.

Figure 5 illustrates the segmentation of a typical document into three classes: text (black), image (gray), and background (white). Such a classification determines both the appropriate enhancement algorithm as well as the neighborhood of pixels from which interpolated pixels will be estimated. Another document image is shown in Figure 6. At the bottom is an enlarged portion of this document, obtained with bicubic interpolation applied to the entire image. Compare this to the top of Figure 7. Here, content-based enhancement is performed, followed by global bicubic interpolation. Although the background has been eliminated, the text remains just as blurry. In contrast, the bottom of Figure 7 demonstrates the value of performing interpolation, as well as enhancement, according to the classified content of the image. The enlarged text is sharper and considerably improved.

6. Conclusions

In this paper, we presented a new technique to segment document images. With this technique, each pixel is assigned a feature pattern consisting of a scaled family of differential geometrical invariant features. The invariant feature pattern is then assigned to a specific region using a two-stage neural network system. The first stage is a self-organizing principal components analysis (SOPCA) network that is used to project the feature vector onto its leading principal axes found by using principal components analysis. This step provides an effective basis for feature extraction. The second stage consists of a self-organizing feature map (SOFM) which will automatically cluster the input vector into different regions. Finally, a connected component labeling algorithm is applied to ensure region connectivity. The eight most effective features that the SOPCA selected included the variance, contrast at 0 and 90 degrees, correlation at $0, 90$, and 270 degrees, the first and second order multiscale features. These features have the most discriminating

 Copyright 2000, IS&T

power in the segmentation process. Once the segmentation step is performed, specific filters and interpolation functions can be applied to each document component.

References

[1] M. Ahmed, B. E. Cooper, and S. Love, "Document Image Segmentation Using a Two-stage Neural Network," *Proc. of SPIE* , Vol. 3962, pp. 25-33, January 2000.

[2] M. Unser, A. Aldroubi, and M. Eden, "Enlargement or Reduction of Digital Images with Minimum Loss of Information," *IEEE Trans. on Image Processing* , Vol. 4, No. 3, pp. 247-258, March 1995.

[3] Y. Y. Tang, C. Y. Suen, and M. Cheriet, "Document Analysis and Understanding: A Brief Survey," *Proc. Int. Conf. on Document Analysis and Recognition*, pp. 17-31, 1991.

[4] H. Cheng and C. Bouman, "Trainable Context Model for Multiscale Segmentation," *Proc. Int. Conf. on Image Processing (ICIP)*, October 1998.

[5] L. O'Gorman, "The Document Spectrum for Page Layout Analysis," *IEEE Trans. on Pattern Matching and Machine Intelligence (PAMI)*, Vol. 15, No. 11, November 1993.

[6] K. Etemad, D. Doermann, and R. Chellapa, "Page segmentation using decision integration and wavelet packets," *Proc. Int. Conf. of Pattern Recognition*, Vol. 2, pp. 345-349, 1994.

[7] A.K. Jain and Y. Zhong, "Page segmentation using texture analysis," *Pattern Recognition*, Vol. 29, No. 5, pp. 343-770, 1996.

[8] H. Cheng, C. Bouman, and J. Allebach, "Multiscale Document Segmentation," *Proc. of IS&T 50th Annual Conference*, pp. 417-425, May 1997.

[9] A. P. Witkin, "Scale space filtering," *Proc. International Joint Conference on Artificial Intelligence*, pp. 1019-1023, 1983.

[10] S. Haring, M. Viergever, and J. Kok "A multiscale approach to image segmentation using Kohonen networks," *Proc. IPMI*, pp. 212-224, Berlin, 1993.

[11] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, Vol. 50, pp. 363-370, 1984.

[12] L. Florack and B. M. Romeny, "Scale and the differential structure of images," *Image and Vision Computing*, Vol. 10, pp. 376-388, 1992.

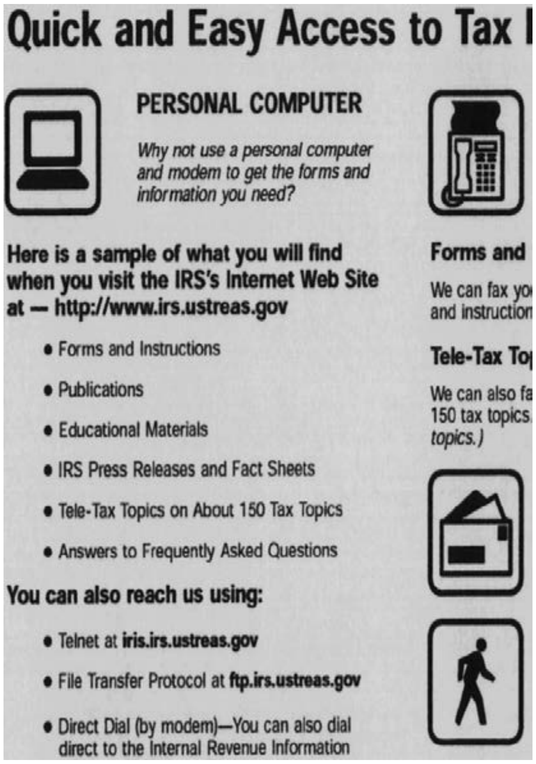
[13] R. M. Haralik and L. G. Shapiro, "Computer and Robot Vision," Addison-Wesley Publishing Company, 1992.

[14] S. Dellepiane et al., "Model generation and model matching of real images by fuzzy approach," *Pattern Recognition*, Vol. 25, No. 2, pp. 115-137, 1992.

[15] S. Haykin, "Neural Networks: A Comprehensive Foundation," Macmillan College Publishing Company, 1994.

[16] T. Kohonen, "Self Organization and Associative Memory," Springer-Verlag, 1984.

Figure 1: *Examples of different scanned documents at 300 dpi.*


Quick and Easy Access to Tax I

PERSONAL COMPUTER

Why not use a personal computer and modem to get the forms and information you need?

Here is a sample of what you will find when you visit the IRS's Internet Web Site at — http://www.irs.ustreas.gov

- Forms and Instructions
- Publications
- Educational Materials
- IRS Press Releases and Fact Sheets
- Tele-Tax Topics on About 150 Tax Topics
- Answers to Frequently Asked Questions

You can also reach us using:

- Telnet at **iris.irs.ustreas.gov**
- File Transfer Protocol at **ftp.irs.ustreas.gov**
- Direct Dial (by modem)—You can also dial direct to the Internal Revenue Information

Forms and

We can fax yo and instruction

Tele-Tax To

We can also fa 150 tax topics *topics.)*

Figure 2: *Result of applying a smoothing function to remove noise, followed by bicubic interpolation of the entire image. The resulting document image suffers from edge blurring and aliasing.*

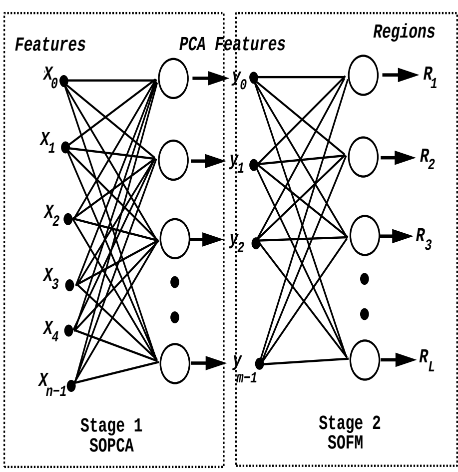


Figure 3: *Two-stage network (SOPCA-SOFM) for the unsupervised clustering of multiscale feature vectors.*

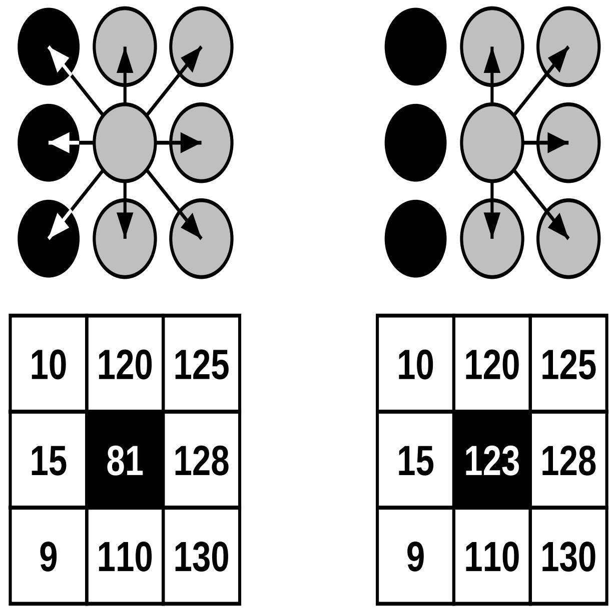


Figure 4: *Conventional interpolation (left) computes the interpolated central pixel using all pixels within a local window, causing blurring. Content-based interpolation (right) estimates the interpolated pixel using only those pixels within the same class, providing a more distinct and accurate result.*

Figure 5: *Original and segmented document into text (black), image (gray), and background (white)*

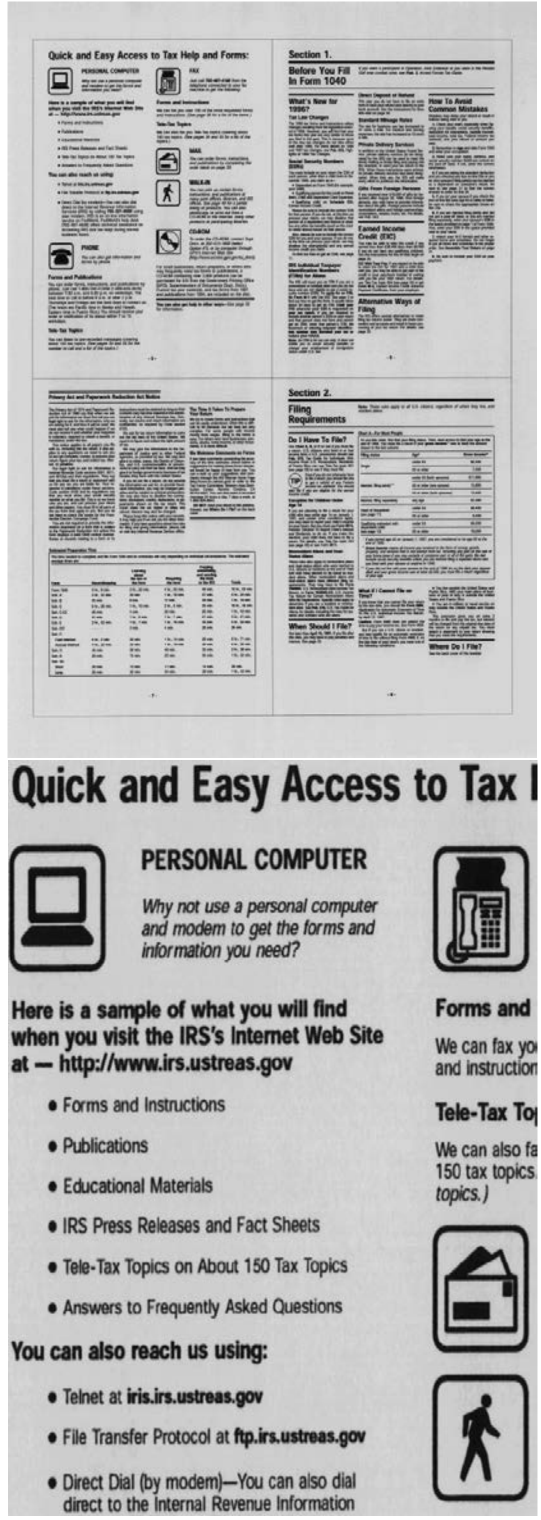


Figure 6: *Original and enlarged document using bicubic interpolation applied to the whole image*

 Copyright 2000, IS&T

Quick and Easy Access to Tax I

PERSONAL COMPUTER

Why not use a personal computer and modem to get the forms and information you need?

Here is a sample of what you will find when you visit the IRS's Internet Web Site at — http://www.irs.ustreas.gov

- Forms and Instructions
- Publications
- Educational Materials
- IRS Press Releases and Fact Sheets
- Tele-Tax Topics on About 150 Tax Topics
- Answers to Frequently Asked Questions

You can also reach us using:

- Telnet at **iris.irs.ustreas.gov**
- File Transfer Protocol at **ftp.irs.ustreas.gov**
- Direct Dial (by modem)—You can also dial direct to the Internal Revenue Information

Forms and

We can fax yo and instruction

Tele-Tax To

We can also fa 150 tax topics. *topics.)*

Quick and Easy Access to Tax I

PERSONAL COMPUTER

Why not use a personal computer and modem to get the forms and information you need?

Here is a sample of what you will find when you visit the IRS's Internet Web Site at — http://www.irs.ustreas.gov

- Forms and Instructions
- Publications
- Educational Materials
- IRS Press Releases and Fact Sheets
- Tele-Tax Topics on About 150 Tax Topics
- Answers to Frequently Asked Questions

You can also reach us using:

- Telnet at **iris.irs.ustreas.gov**
- File Transfer Protocol at **ftp.irs.ustreas.gov**
- Direct Dial (by modem)—You can also dial direct to the Internal Revenue Information

Forms and

We can fax yo and instruction

Tele-Tax To

We can also fa 150 tax topics. *topics.)*

Figure 7: *Document enlargement with content-based enhancement, with (top) bicubic interpolation performed separately, after enhancement, and (bottom) both interpolation and enhancement guided by segmentation.*