Segmentation-Based JPEG for Document Images

Mohamed Nooman Ahmed
Lexmark International Inc.
Lexington, Kentucky

Abstract

This paper describes a method for compressing document images using variable quantization tables based on document segmentation. This method aims at solving the problem of text and fine details degradation often encountered when compressing images, especially scanned document using the JPEG standard. The segmentation step first mark image elements as belonging to high/low frequency areas. Using this information, higher quantization is selected for low frequency areas while the quality of fine details is preserved by using lower quantization factors.

Introduction

One of the most popular and comprehensive continuous tone, still frame compression standard is the JPEG standard. It defines three different coding systems: (1) a lossy baseline coding system, which is based on the Discrete Cosine Transform (DCT); (2) an extended coding system for greater compression and progressive reconstruction applications; and (3) a lossless independent coding scheme for reversible compression. To be JPEG compliant, a product or system must include support for the baseline system.

Although there are a number of settings that can be predefined to achieve different compression ratios, there is only one parameter, called the quality factor, that is adjusted regularly in JPEG on an image-by-image basis with one setting for an active image. The quality factor is a single number in an arbitrary, relative scale. A high quality factor will provide a relatively high quality decompressed image, but will require a relatively large file. And, of course the lower the quality, the rougher the approximation of the image and the more compression with a correspondingly smaller file size, but also, the more visible defects, or artifacts, will be in the decompressed final image. In scanned documents, text generally shows significant compression artifacts at higher quality factors than pictures.

Therefore, a long sought goal in image compression has been to maintain maximum perceptible image quality while achieving maximum compression. Most of the techniques that have been developed in the past for compound images are based on proprietary (non-standard) compression techniques, so the images could only be decompressed using a specific company's product.

In this paper, we present an image compression system which can be used to apply different, appropriate quantization factors to small blocks of pictures and fine details to provide significant image compression. To achieve this, the method consists of a two-step compression scheme. First a segmentation step that divides the image into its different components. Second a variable DCT quantization based on the result of step one is applied. This allows high frequency areas to be compressed at a better quality than smoothly varying parts, thus improving the overall perceptual quality while minimizing the file size and at the same time being JPEG compliant.

JPEG Compression

In the base line system, often called sequential baseline system, the input and output data precision is limited to 8 bits, where as the quatized DCT values are restricted to 11 bits. The compression is performed in three sequential steps: DCT computation, quantization, and variable-length code assignment.

The image is first divided into 8x8 non-overlapping blocks which are processed left to right, top to bottom. After a normalization step, the DCT reduces data redundancy and transforms each fixed length image block by applying a 2D discrete cosine transform. This transform converts each fixed length image block into the frequency domain as a new frequency domain image block. The first coefficient in the block, the lowest frequency coefficient, is the DC coefficient and the other coefficients are the AC coefficients (e.g., for an 8 by 8 block, there will be one DC coefficient and 63 AC coefficients).

The DCT coefficients are then quatized using a defined quantization table and the reordered using a zig-zag pattern to form a 1-D sequence of quantized coefficients. Huffman codes are then applied to the resulting sequence.

Image Segmentation

The first step in our approach is to extract a feature pattern for each pixel in a gray level image. A feature vector **X** is a set of measurements $\{x_1,x_2,..,x_d\}$ which condenses the description of relevant properties of the image into a small, Euclidean feature space of N 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, and gradient magnitudes and directions.[5]

In this paper we will construct our feature vector from measurements obtained from gray level distribution in a window, 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,

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

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

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

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

Fuzzy-C-Means Clustering

The standard FCM objective function for partitioning $\{x_k\}^N_{k=1}$ into c clusters is given by

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{p} ||x_k - v_i||^2 \tag{3}$$

where $\{x_k\}^N_{k=1}$ are the feature vectors for each pixel, $\{v_i\}^c_{i=1}$ are the prototypes of the clusters and the array $[u_{ik}] = U$ represents a partition matrix, $U \in \mathcal{U}$, namely,

$$\mathcal{U}\{\ u_{ik} \in [0,1] \ \ | \ \sum_{i=1}^{c} u_{ik} = 1 \quad \forall k \quad and$$
$$0 < \sum_{k=1}^{N} u_{ik} < N \quad \forall i\} \tag{4}$$

The parameter p is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. The FCCM objective function is minimized when high membership values are assigned to pixels whose intensities are close to the centroid of its particular class, and low membership values are assigned when the pixel data is far from the centroid.

Block Classification and Variable Quantization

Once the segmentation step is performed, we classify the block to be either text or image. The number of text (High Frequency) pixels N_T and image pixels (low frequency) N_I are calculated. If $N_T > N_I$, then the block is classified as text, otherwise the block is classified as image. Image blocks are highly quantized while text blocks are left relatively untouched. This allows high frequency areas to be compressed at a better quality than smoothly varying parts, thus improving the overall perceptual quality while minimizing the file size.

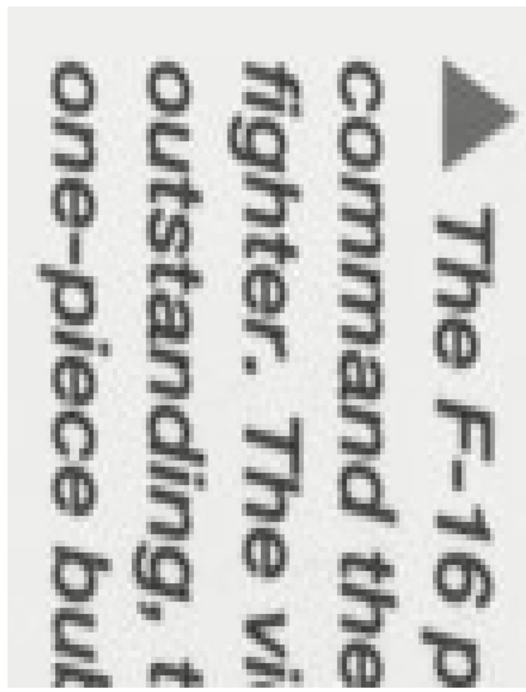

Figure 1. Original image with 8x magnification

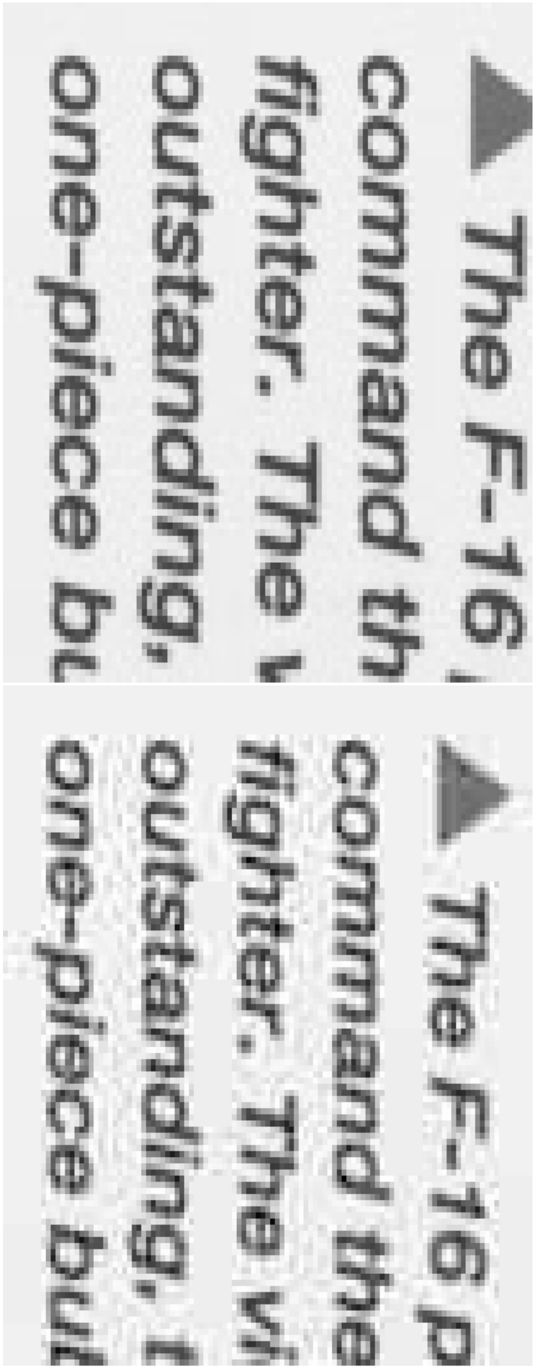

Figure 2. Top: Compressed image using Variable Quantization and, Bottom: JPEG compression with quality factor 0.4.

References

1. 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.
2. H. Cheng and C. Bouman, "Trainable Context Model for Multiscale Segmentation" Proc. Int. Conf. on Image Processing (ICIP), October, 1998.
3. 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.
4. 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.
5. A. K. Jain and Y. Zhong, "Page segmentation using texture analysis," Pattern Recognition, Vol. 29, no. 5, pp. 343-770, 1996.
6. H. Cheng, C. Bouman, and J. Allebach, "Multiscale Document Segmentation," Proc. of IS&T 50th Annual Confernece, pp. 417-425, May, 1997.
7. 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.
8. J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecing compact well-separated clusters," J. Cybern., Vol. 3, pp. 32-57, 1973.
9. J. Bezdek, "A convergence theorem for the fuzzy ISO-DATA clustering algorithms," IEEE Trans. On Pattern Ana. Machine Intell. (PAMI), 1980.
10. C. Xu, D. Pham, and J. Prince, "Finding the brain corext using fuzzy segmentation, isosurfaces, and deformable surfaces," Proc. of the XVth Int. Conf. on Informaiton Processing in Medical Imaging (IMPI 97), pp. 399-404, 1997.
11. D. L. Pham and J. L. Prince, "Adaptive Fuzzy segmentation of magnetic resonance images," IEEE Trans. Medical Imaging, Vol. 18, No. 9, pp. 737-752, 1999.