Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp321-325)

Improving Zeng's Technique for Color Re-indexing of Palette-Based Images

LUCA MATOLA and BRUNO CARPENTIERI
Dipartimento di Informatica ed Applicazioni "R. M. Capocelli"
Università di Salerno
84081 Baronissi (SA)
ITALY

Abstract: - Palette re-indexing is a well-known and effective approach for improving the compression of color-indexed images. In this paper, we modify the re-indexing technique proposed by Zeng *et al.* in [1] to maximize the compression performance of a palette-based imaging system. The proposed scheme pays attention to the statistics of the input image in optimizing the assignment of index values to colors in a one-step look-ahead greedy way. Experimental results show the effectiveness of the proposed algorithm.

Key-Words: - Palette-based images, color-indexed image compression, lossless image compression.

1 Introduction

Many modern applications of lossless and lossy image compression deal with color images.
For images that consist of only a small number of colors, it is also a common practice that each color is mapped to an index and the indexes of the pixels are then compressed, usually losslessly. Since the size of the index space is much smaller than that of the color intensity space, this color palette approach usually can achieve better compression efficiency for images with a limited number of colors [1].
Palletized images, popular on the Internet, have a single component, representing an array of indices to a palette table, rather than multiple components as in the original color space representation. Indices are just labels for the colors they represent, and their numeric value may bear no relation to the color components. Furthermore, palletized images often contain combinations of synthetic, graphic, and text bitmap data, and might contain a sparse color set. Therefore these color-indexed images are represented by a matrix of indexes (the index image) and a color-map or palette. The indexes in the matrix point to positions in the color-map and, therefore, establish the colors of the corresponding pixels. For a particular image, the mapping between index values and colors (typically, RGB triplets) is not unique: it can be arbitrarily permuted, as long as the corresponding index image is changed accordingly [2].
Therefore, the efficiency of a lossless compression algorithm for indexed image may greatly depend on the assignment of indexes in the relative color-map [3].
The computational complexity of the problem of finding an optimal (from the compression point of view) mapping between colors and indices can be high. In fact, the number of possible configurations for a table of M colors corresponds to the number of permutations of M objects, which equals M! (and the problem above can be shown to be NP-hard).
Clearly, exhaustive search for an optimal solution is impractical for most of the interesting cases, which motivated several suboptimal, lower complexity proposals for re-indexing color images [4].
The re-indexing scheme proposed by Zeng *et al.* in [1], is actually one of the best proposals, both in terms of computational efficiency and relative performance.
Zeng uses a greedy strategy to obtain a re-indexing scheme that is such that he resulting image tends to be smoother than the initial color image, thus it is more amenable to most lossless compression algorithms [4].
Our approach aims to improve the lossless compressibility of the final re-indexed image, working on the statistical information accumulated in the early steps of Zeng's algorithm.
This paper is organized as follows. In Section 2, we give a description of the re-indexing problem and a review of Zeng's algorithm. Section 3 reports our proposal on the improvement of Zeng's method, and, finally, in Section 4 we present experimental results that compares the compression obtainable with Zeng's algorithm and the compression obtainable with our proposal.

Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp321-325)

2 The Re-Indexing problem

The re-indexing problem can be formulated as follows. Let $\boldsymbol{I}$ be an image of m x n pixels and M be the number of distinct colors.

$\boldsymbol{I}$ can be represented as: $\boldsymbol{I(x,y) = P(I'(x,y)}$, where $\boldsymbol{P = \{S_1,S_2, \dots , S_M\}}$ is the set of all the colors in $\boldsymbol{I}$, and $\boldsymbol{I'}$ is a m x n matrix of indexes in $\boldsymbol{= \{1, 2, \dots , M \}}$.

The image $\boldsymbol{I'}$ is called *indexed image* and P is its *palette*. Typical values for M are in the range [3,256] . Most of the compression engines proceed by scanning in some sequential order the indexes in $\boldsymbol{I'}$. After the previous scan operation, the pixels may be numbered as $\boldsymbol{p_1, p_2, \dots , p_{mxn}}$.

At this point, in [3], the information needed to reconstruct the original image is:

1) the color of pixel p_1 ;

2) a table providing the correspondence between colors $S_1, S_2,\dots, S_M$ with index $i_1, i_2, \dots, i_M$;

3) the sequence of the differences:
   d_h = (index of color in pixel $h+1$) – (index of color in pixel h).

Let $\boldsymbol{D(I')}$ be the set of all differences with d_j con j = 1,2,…, $m \times n$ -1 . From information theory we know that to encode the set of differences $\boldsymbol{D(I')}$ any lossless compression scheme requires a number of bits per pixel (bpp) greater or equal to the zero-order entropy of the statistical distribution of $\boldsymbol{D(I')}$. At this point, an index ordering takes a significant importance in the compression's process. Indeed , if indexes $\boldsymbol{i_1, i_2, \dots , i_m}$ are ordered in such a way to produce an almost uniform distribution of values d_h , the entropy value will be large. Conversely, a zero-peaked distribution in $\boldsymbol{D(I')}$ gives a lower entropy value.

2.1 Zeng's Method

The palette re-indexing method proposed by Zeng *et al.* [1] is based on a fast and quite efficient one-step lookahead greedy approach, which aims at increasing the lossless compression efficiency of color-indexed images.

Suppose that in the initial index image, the index values 0,1, ..., M-1 represent color symbols $\boldsymbol{S_O}$, $\boldsymbol{S_1}$, ..., $\boldsymbol{S_{M-1}}$, respectively.

The algorithm starts by finding the symbol that is most frequently located adjacent to other (different) symbols, and the symbol that is most frequently found adjacent to it.

The statistics gathered from the initial index image are used to build a table that indicates the cross-counts $C(S_i, S_j)$ of two different symbols S_i and S_j .

The cross-count $C(S_i, S_j)$ is defined as the number of times that a pixel with color S_i is spatially adjacent to a pixel with color S_j, in the initial index image.

Fig.1 (a) shows as an example an image with 4 colors. Fig. 1 (c) shows the relative matrix C when a raster row-by-row scanning scheme is adopted.

The aim of this table is to assign close index value to symbols that are frequently located next to each other.

The re-indexing scheme proposed by Zeng in [1] can be summarized as:

Step 1: Calculate the cross-counts $C(S_i, S_j)$ for each pair of symbols S_i, and S_j based on the initial index image. Calculate the cumulative cross-counts $C_i = \sum_{j=0,\, j\neq i}^{M-1} C(S_i, S_j)$ for each symbol S_i .

Step 2: Find the symbol S_{Max} that has the largest cumulative cross-counts C_i. Denote it as L_0 .Put L_0 in a symbol pool ***P***, i.e., ***P***$=\{ L_0 \}$ at this point. ***P*** will consist of spatially ordered symbols. Denote the size of ***P*** as ***N***, and set ***N***= 1. A new entry can enter ***P*** *only* from either the left end or the right end. Once a symbol enters the pool ***P***, it will be indicated as assigned.

Step 3: A new unassigned symbol will be chosen and assigned to the left or the right end position of the pool ***P***. Let us first consider the left end position. The unassigned symbol S_i, that maximizes the potential function $D_i = \sum_{j=0}^{N1} w_{(N,j)} \cdot C(S_i, L_j)$ will be chosen, where $w_{(N,j)}$ are the weights controlling the impact of $C(S_i, L_j)$ on the overall potential function $\boldsymbol{D_i}$. In general, $w_{(N,j)}$ will depend on the physical distance between the current end position of the pool ***P*** and the position of L_j . The parameter ***N*** is used to indicate that the weight $w_{(N,j)}$, in general, may change after each iteration. A good choices of $w_{(N,j)}$ is $\log_2(1+1/d_{(N,j)})$ where $d_{(N,j)}$ is the physical distance between the position of L_j and the end position, as analyzed in the following. Denote this chosen symbol as S_{Lmax}. Similarly, S_{Rmax}, can be chosen for the right end position of ***P***.

Step 4: One of S_{Lmax} and S_{Rmax} that has the larger potential function value $\boldsymbol{D_i}$ is assigned to the corresponding end position in the pool ***P***, and is denoted as $\boldsymbol{L_N}$. Set N= N+1. An example of the status of ***P*** is ***P*** = $\{L_3, L_0, L_1, L_2\}$ for N=4.

Step 5: If $(N < M)$, go to Step 3.

Step 6: Assign integers 0, 1, ..., M-1 to the spatially ordered symbols in the pool ***P*** in left-to-right or right-to left order. A re-indexed index image is generated by replacing the initial index value i with the new index value assigned to S_i .

3 Our Proposal

The index image has, in general, some properties:

1) The number of colors is usually limited, much less than the number of colors of a natural image.

2) Pixels with the same index tend to cluster together.

Within a region of pixels having the same index, the compression efficiency is very high, and it is usually independent of the index value. The cost of coding depends on the transitions between regions that have different index values. In general, the larger the difference of the neighbouring index values, the more bits it takes to code the transition.

Similarly to Zeng's method, also our solution requires the transformation of the image into a linear sequence of pixel. The aim is to take advantage of the pixel similarities among neighbouring pixel positions in the linearized sequence after the raster-scan process. We named this approach **2D**.

Our proposal consists of an initial pre-processing phase, in which sequences of pixel of the same color are detected. Once these sequences are discovered, we shall use them, instead of the single pixels that Zeng's technique was using, to build the co-occurrence matrix. We are therefore considering the image as a set of subsequences, where each subsequence has a particular length and it is composed by identical pixels.

Now, to calculate the cross-count $C(S_i, S_j)$ for each pair of symbols we have to consider the following quantities:

1) number of times that a subsequence with symbol S_i is spatially adjacent to a subsequence with symbol S_j in the original index image.

2) number of pixels that are involved in the previous computation.

We weight 1) by multiplying it by 2) and by giving a weight to the entries in the co-occurrence matrix that is proportional to the frequency of their occurrence in the original index image.
For example, let

1 1 1 2 2 1 3 3 1 2 2 1

be the sequence derived from the raster scan visit of the original index image. Then the corresponding co-occurrence matrix is:

	1	**2**	**3**
1	x	16	3
2	12	x	0
3	3	0	x

where, for example, *C(1,2)* is obtained by multiplying the two occurrences of the transaction between the sub-sequences composed by 1 and 2 with the eight pixels involved, displayed in boldface, are

1 1 1 2 2 1 3 3 **1 2 2** 1

In the worse case, if each subsequence is composed by only one pixel, the 2D procedure becomes equivalent to Zeng's method.

4 Experimental Results

To evaluate the effectiveness of the proposed method, we have performed some preliminary experimental tests.
Comparisons have been carried out between our re-indexing algorithms and Zeng's technique [1]. We have experimented on a test set of color-indexed images having various size and number of colors.
The number of colors for the images in the test set ranges between 7 and 256.
The test set contains the following nine popular images, generally considered as a benchmark set for the research in this area.
The images are:

- "Books.ppm" (179 x 318 pixels)
- "Music.ppm" (111 x 111 pixels)
- "Winaw.ppm" (633 x 465 pixels)
- "Netscape.ppm" (612 x 100 pixels)
- "Sea_dusk.ppm" (484 x 325 pixels)
- "Benjerry.ppm" (466 x 60 pixels)
- "Gate.ppm" (564 x 108 pixels)
- "Sunset.ppm"(640 x 480 pixels)
- "Yahoo.ppm" (460 x 59 pixels) .

These images have been first reindexed using our algorithm and Zeng's original method [1], and then they have been JPEG-LS compressed.
Table 1 presents the compression results obtained by using a standard JPEG-LS codec (SPMG JPEG-LS obtained from http://spmg.ece.ubc.ca).
The results are presented in terms of bits per pixel.

Image	**Color**	**Zeng**	**2D**	**Gain 2D (%)**
Books	7	1.736	1.736	0
Music	8	1.186	**1.158**	2.4
Winaw	10	0.506	**0.494**	2.4
Netscape	27	1.83	**1.791**	2.2
Sea_disk	43	0.187	**0.177**	5.4
Benjerry	48	1.272	**1.242**	2.4
Gate	69	2.932	3.138	-6.4
Sunset	138	2.947	**2.73**	7.4
Yahoo	156	1.867	**1.745**	6.6
Average Gain				2.5%

Table 1 - Compression results, using a JPEG-LS codec. "Gain" indicates the percentage of compression of proposed method in relation to the Zeng's approach. Best results compared with Zeng's results are displayed in boldface.

5 Conclusions

Analyzing the results in Table 1, we can conclude that these preliminary experimental results show modification proposed in this note is indeed effective.
In fact, for all but two of the test images ("Books.ppm" and "Gate.ppm"), we have had significant compression (two of them with a gain of 5 % or more). For the "Books.ppm" image, the compression results are equivalent to the ones obtained by using Zeng's technique.
Future research involves a wider experimentation of our technique and a better theoretical assessment of its potentials.

References:

[1] W.Zeng, J.Li and S. Lei, "An Efficient Color Re -Indexing Scheme for Palette-Based Compression", Proceedings of the IEEE International Conference on Image Processing (ICIP-2000).

[2] A. J. Pinho and A. J. R. Neves, "A Note on Zeng's Technique for Color Re-indexing of palette-Based Images", *IEEE Signal Processing Letters*, Vol.11, No.2, February 2004.

[3] S. Battiato, G. Gallo, G. Impoco and F. Stanco, “An Efficient Re-Indexing Algorithm for Color-Mapped Images”, *IEEE Transaction on Image Processing*, Vol.13, No.11, November 2004.

[4] A. J. Pinho, A. J. R. Neves, “A Survey on Palette Reordering Methods for Improving the Compression of Color-Indexed Images”, *IEEE Transaction on Image Processing*, Vol.13, No.11, November 2004.

[5] M. Weinberger, G. Seroussi, G. Sapiro, "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS", Hewlett-Packard Laboratories Technical Report No. HPL-98-193R1, November 1998, revised October 1999.

[6] B. Carpentieri, M. Weinberger, G. Seroussi, "Lossless Compression of Continuous-Tone Images", *Proceedings of the IEEE*, Vol. 33, No. 11, November 2000.