An Efficient Re-Indexing Algorithm for Color-Mapped Images

Sebastiano Battiato, Giovanni Gallo, Gaetano Impoco, and Filippo Stanco, *Member, IEEE*

***Abstract*—The efficiency of lossless compression algorithms for fixed-palette images (indexed images) may change if a different indexing scheme is adopted. Many lossless compression algorithms adopt a differential-predictive approach. Hence, if the spatial distribution of the indexes over the image is smooth, greater compression ratios may be obtained. Because of this, finding an indexing scheme that realizes such a smooth distribution is a relevant issue. Obtaining an optimal re-indexing scheme is suspected to be a hard problem and only approximate solutions have been provided in literature. In this paper, we restate the re-indexing problem as a graph optimization problem: an optimal re-indexing corresponds to the heaviest Hamiltonian path in a weighted graph. It follows that any algorithm which finds a good approximate solution to this graph-theoretical problem also provides a good re-indexing. We propose a simple and easy-to-implement approximation algorithm to find such a path. The proposed technique compares favorably with most of the algorithms proposed in literature, both in terms of computational complexity and of compression ratio.**

***Index Terms*—Hamiltonian path, indexed images, NP completeness, re-indexing, traveling salesman problem (TSP).**

I. Introduction

In THIS paper, we address a relevant issue regarding the compression of indexed images. Indexed images encode colors using a fixed lookup table (*palette*). Each entry in the table is generally a triplet of RGB values. For each pixel in the image, only the index of the corresponding color needs to be stored. The efficiency of a lossless compression algorithm for indexed images may greatly depend on the assignment of indexes in the relative lookup table. Hence, lossless compression can be optimized by choosing a different palette ordering. In particular, a palette which assigns consecutive indexes to colors sharing many adjacent pixels in the image will provide better compression ratios using adaptive compression algorithms.

Unfortunately, given an image I with M colors, the number of possible color indexings is $M!$, and only heuristic methods to pick an optimal ordering out of the many possible ones are known. Experimental data suggest that the problem is likely to be intrinsically difficult (NP hard). If our claim was true, the only strategy to avoid the exponential amount of computational resources required by an exhaustive approach is to adopt some suitable heuristic.

We propose a graph theoretical setting to achieve a content-dependent smooth index distribution in the re-indexed image. A re-indexing is obtained in correspondence with a suboptimal solution of an instance of the *traveling salesman problem* (TSP) in a weighted graph.

We tested our algorithm with different kind of images (e.g., continuous-tone, graphics), with different resolution and number of colors. Experimental results show the effectiveness of our approach.

This paper is structured as follows. Section II introduces the re-indexing problem, while Section III reviews some related work. Our technique is described in Section IV and some heuristics to (approximately) solve the TSP problem are discussed. Experimental results are shown in Section V. Conclusions are drawn in Section VI.

II. Color Re-Indexing Problem

The re-indexing problem can be stated as follows. Let I be an image of $m \times n$ pixels and M be the number of distinct colors. I can be represented as $I(x,y) = P(I'(x,y))$, where $P = \{S_1, S_2, \ldots, S_M\}$ is the set of all the colors in I, and I' is a $m \times n$ matrix of indexes in $\{1, 2, \ldots, M\}$. An image represented in such a fashion is called *indexed image* and P is its *palette*. Typical values for M are 16, 64, or 256.

Most of the compression engines proceed by scanning in some sequential order the indexes in I'. Once an ordered scan has been performed, the pixels encountered may be named $p_1, \ldots, p_{m \times n}$. If a differential approach to coding and compression is adopted the information needed to reconstruct the original image is

1) the color of pixel p_1;
2) a table providing the correspondence between colors $S_1, S_2, \ldots, S_M$ with index $i_1, i_2, \ldots, i_M$;
3) the sequence of differences

$$d_h = (\text{index of color in pixel } h+1) - (\text{index of color in pixel } h).$$

Let $D(I')$ be the set of all differences d_j with $j = 1, 2, \ldots, (n \times m) - 1$. Information theory tells us that any lossless scheme to encode the set of differences $D(I')$ requires a number of bits per pixel (bpp) greater or equal to the zero-order entropy of the statistical distribution of $D(I')$.

If indexes $i_1, i_2, \ldots, i_M$ are ordered so as to produce an almost uniform distribution of values d_h, the entropy value will be large. Conversely, a zero-peaked distribution in $D(I')$ gives a

Manuscript received May 6, 2002; revised April 12, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Amri Said.

S. Battiato and G. Gallo are with the Dipartimento di Matematica e Informatica, University of Catania, 95125 Catania, Italy (e-mail: battiato@dmi.unict.it; gallo@dmi.unict.it).

G. Impoco is with the VCG, ISTI-CNR Pisa, 56100 Pisa, Italy (e-mail: impoco@di.unipi.it).

F. Stanco is with the Dipartimento di Elettrotecnica, Elettronica e Informatica, University of Trieste, 34127 Trieste, Italy (e-mail: fstanco@units.it).

Digital Object Identifier 10.1109/TIP.2004.836183

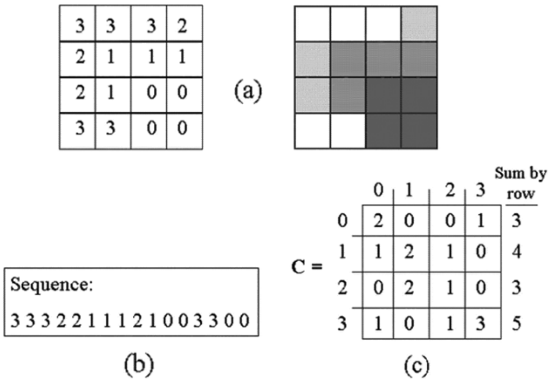

Fig. 1. (a) Toy example of an image with four colors. (b) Sequence of color symbols scanning the image in row order. (c) Co-occurrence matrix C when scanning scheme (b) is adopted.

lower entropy value. Hence, finding an optimal indexing scheme is a crucial step for differential lossless compression of indexed images.

III. Related Work

The existing solutions to the re-indexing problem may be classified into two groups, according to the particular model/strategy adopted. The first group of solutions performs the re-indexing of color indexes according to perceptive similarity between different colors. In [7], [8], and [12], consecutive symbols are assigned to visually similar colors. Although several perceptual similarity measures can be adopted, the most widely used is luminance. In [2], the idea of color ordering has been further refined introducing the zero-order entropy as a quantitative estimate of the "goodness" of a candidate re-indexing. Indexes are ordered accordingly by means of a simulated annealing algorithm for entropy minimization. Memon investigated the problem of ordering the palette with respect to the compression ratio obtained with a suite of different compression algorithms [6]. The bottleneck of this group of solutions is the relative inefficiency of running a simulated annealing algorithm to optimize the palette re-indexing. To overcome this problem a *pairwise merge* (PM) heuristic has been proposed in [6]. However, it is still too burdensome to yield fast timings.

The second group of re-indexing algorithms is guided by both information theory and local adaptive considerations. Some techniques specifically devoted to work on a bit plane basis are presented in [3] and [4], while [11] introduces a color correlation sorting algorithm. The most widely known representative in this class is the solution proposed by Zeng et alii [13], [14]. This technique is based on a greedy algorithm to maximize a suitable potential function. The potential function has been heuristically selected in such a way that large values correspond to more peaked distributions of the set $D(I')$. Co-occurrence statistics of the image indexes are collected in a matrix C. Each entry $C(i,j)$ reports the number of times the pair of neighboring color symbols (S_i, S_j) has been observed over the pixel sequence $p_1, \ldots, p_{n\times m}$. The co-occurrence matrix could be re-arranged in order to profile the particular predictive scheme adopted by the compression engine. Fig. 1(a) shows a toy example of an image with four colors. Fig. 1(c) shows the relative matrix C when a raster row-by-row scanning scheme is adopted [Fig. 1(b)].

Using the statistical information provided by the co-occurrence matrix, Zeng's algorithm computes a new order for the color symbols by iteratively adding new colors to an ordered list. A potential function is maximized.

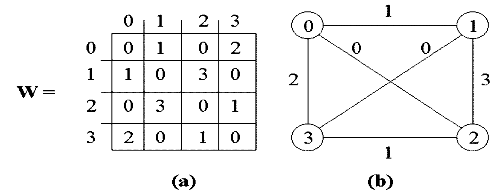

Fig. 2. (a) Matrix W relative to the example of Fig. 1. (b) Graph induced by matrix W.

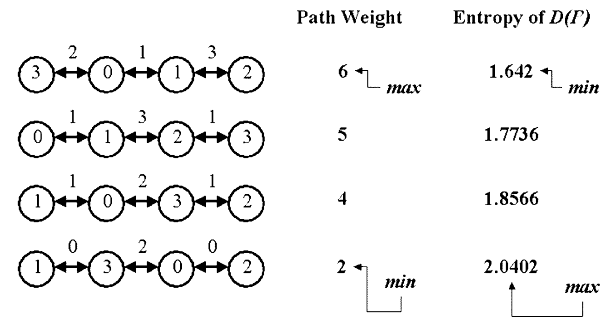

Fig. 3. Weight, set D(I'), and entropy value for some Hamiltonian paths over the graph in Fig. 2.

IV. TSP-Based Re-Indexing Scheme

Our solution derives a suitable indexing scheme solving an optimization problem for a Hamiltonian path in a weighted graph. Given an indexed image I represented by the pair (I', P), with M different colors and co-occurrence matrix C computed as above, consider a complete nondirected weighted graph $G = (V, E, w)$ where

$$V = \{S_1, S_2, \ldots, S_M\}$$
$$E = V \times V$$

$w : E \longrightarrow N$ defined as follows:

$$\begin{cases} w(i,j) = w(j,i) = C(i,j) + C(j,i), & \text{if } i \neq j \\ w(i,j) = 0, & \text{if } i = j \end{cases}$$

with $i, j = 1, \ldots, M$.

Let us call $W = \{w_{ij}\}$ an $M \times M$ matrix such that $w_{ij} = w(i,j)$. The analysis of some examples and informal heuristic considerations suggest that there is a strong negative correlation between the weight of a Hamiltonian path over the graph and the entropy of set $D(I')$, when the color symbols are assigned according to their order in the path. In other words, the greater the weight of an Hamiltonian path, the smaller is the entropy of $D(I')$ (see Figs. 2 and 3).

We have not been able to provide a formal proof of the idea sketched above but we believe that the following can be stated.

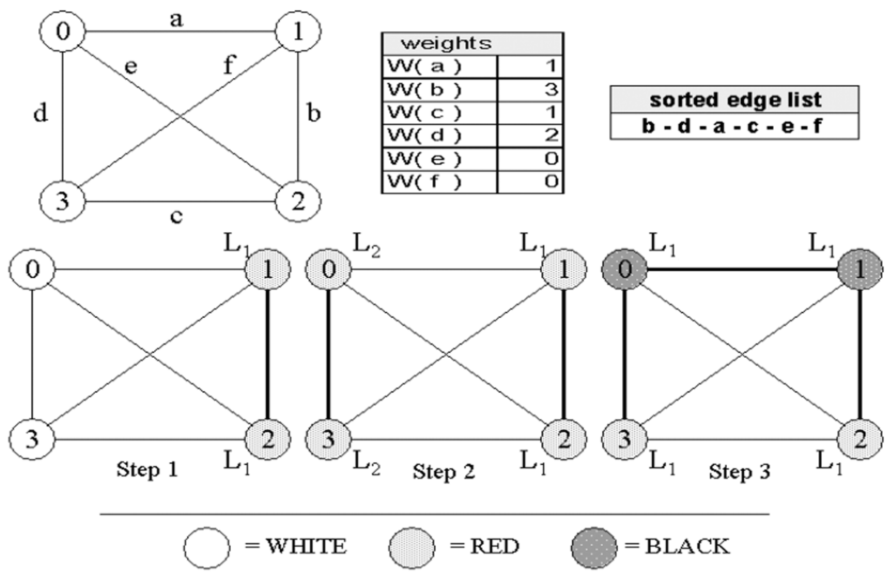

Fig. 4. Successive step of the construction of the Hamiltonian heavy path according to the algorithm described in the text.

Conjecture: A simple Hamiltonian path in G of maximum weight gives an optimal re-indexing scheme.

The proof of the above conjecture is not a trivial task because it requires a reduction from a graph-theoretical problem to the optimization of a nonlinear function. However, the experimental results reported in Section V provide some evidence of the truth of our conjecture.

Nonetheless, the previous claim suggests a criterion to properly arrange the distribution of local relative differences. The idea is to select the re-indexing associated with the heaviest Hamiltonian path in G (i.e., a solution for the TSP leads to an optimal re-indexing). Unfortunately, since TSP is an NP-complete problem, no deterministic polynomial solution is known. Hence, only approximate solutions can be provided in realistic times. We have investigated several heuristics in order to find a good approximate solution to the TSP problem. However, the majority of the existing approaches are too burdensome and time consuming for our application. That is why we have devised a simple and fast heuristic in order to find good solution. We start sorting the edge set E in decreasing weight order and marking as WHITE all vertices in V. The algorithm continues incrementally adding to the path nonvisited edges with maximal weight. This procedure is detailed in the following pseudo-code.

Until there are unprocessed edges in E, *do* the following.

1.1) Extract the edge $e_{i,j}$ of maximal weight that has not yet been processed.

1.2) The following three cases may arise:

- 1.2.1) both vertices i, j are WHITE. In this case, both of them are set to RED and they get the same new label;
- 1.2.2) one of the two vertices is WHITE and the other one is RED. In this case, the WHITE vertex becomes RED and gets the same label that has been assigned to the RED vertex; the RED vertex, in turn, becomes BLACK;
- 1.2.3) both vertices are RED. In this case, both vertices have been already visited by the algorithm. Two cases may further happen:
  - a) both vertices are extrema of a chain of vertices with the same label. In this case, do nothing and skip edge $e_{i,j}$;
  - b) the vertices are extrema of two distinct chains of vertices. In this case, they are both set to BLACK and the relative labels are unified;
- 1.2.4) one vertex is BLACK. The edge is skipped.

The new indexing is obtained according to the relative position of the various color symbols in the path, scanning the path in one of the two possible directions. Fig. 4 shows the weight matrix W together with the Hamiltonian path built in a run of the proposed algorithm.

A simple variation of this algorithm is to select the local best node instead of the best edge. A single-edge chain is built by selecting the heaviest edge incident to the vertex with the highest energy (i.e., the vertex with the highest sum of all co-occurrences between the selected node and the others). The remaining edges are added iteratively by choosing the heaviest edge incident to one of the endpoints of the chain; care must be taken to prevent the chain from autointersecting. This step is repeated until a complete Hamiltonian path is found. This heuristic is as fast as the previous one, but its performance never exceeds that of our previous approach. For this reason, it has not been included in the results.

We have tried also other, more sophisticated heuristics to find an approximate solution of the TSP problem. Among the many techniques proposed in literature, we have also considered genetic algorithms [9], but we were unable to obtain satisfactory results. Moreover, the extra complexity added to the algorithm is not matched by an improvement in terms of compression ratio. This negative result should indeed be expected: there is no known polynomial time algorithm able to find a solution for

TABLE I
ENTROPY VALUES FOR IMAGES WITH DIFFERENT NUMBER OF COLORS

	Zeng	Memon	Approx TSP	Brute force TSP
8 colors	0.8160	0.8053	0.8051	0.8051
12 colors	0.9496	0.9428	0.9291	0.9237

TABLE II
ENTROPY, BPP, AND TIME VALUES

	Luminance	Random	Zeng	Memon	Approx TSP
Entropy	2.21	2.46	1.97	1.86	1.95
bpp	5.56	6.56	5.14	4.48	5.06
Time	15	10	1255	18738	433

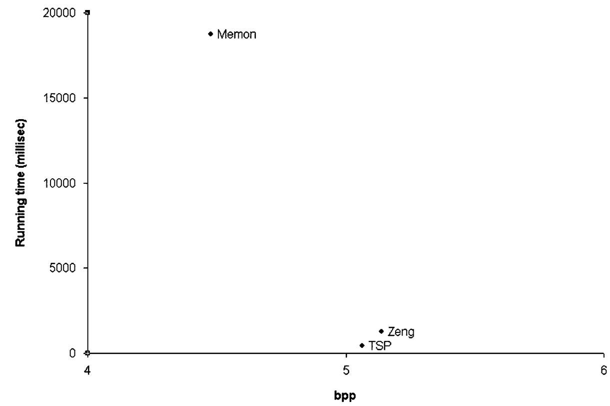

Fig. 5. Bits-per-pixel performance versus running time of three re-indexing schemes.

a general instance of the TSP problem that guarantees a prescribed error bound to the approximation [15].

V. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed method, we have performed several tests over different sets of indexed images, using different TSP heuristics. Comparisons have also been carried out between our re-indexing algorithm and the methods described in [6] (*pairwise merge*) and [14], which in our opinion are the two most effective algorithms in literature.

First of all, we gathered some experimental support to our conjecture that optimal re-indexing for an image is related to an optimal path in the corresponding TSP problem instance (Section II). We computed the optimal TSP paths for the graphs corresponding to indexed images by brute force. In order to keep the computation feasible, we have restricted our attention only to images with eight and 12 colors. Table I shows some comparisons that give some evidence of the truth of our conjecture.

Some experiments have been performed over two sets of indexed images: photographs (i.e., continuous tone) and synthetic images. As a benchmark measure, we have adopted the entropy of the sets $D(I')$ of test images, the corresponding compression ratios obtained with the JPEG2000-lossless algorithm [1], [5], [10], and the relative timing performances. Table II shows that the minimal entropy values are found in correspondence of Memon's technique at the expenses of a high computational burden. Our TSP-based approach provides performances close to Zeng's, which gives the lowest entropies and runs faster than other approaches. The slight gain in compression ratio of our algorithm with respect to Zeng method probably depends on the difference between the two greedy strategies adopted. Namely, rather than growing a single chain by choosing the best fitting color symbol as Zeng does, we generate suborderings that are eventually merged.

The results for bpp follow the same pattern observed for entropy. A simultaneous comparison of running time and bpp performance obtained with different re-indexing schemes is shown in Fig. 5. The proposed technique is a realistic alternative to the other re-indexing schemes.

Not surprisingly, experiments show that both bpp rates and timings are higher for real life photos than for synthetic images. A more interesting observation is that when working with few colors the performance of Zeng's, Memon's, and our approach is substantially the same. Performances differ as the number of colors increases. This is probably related to the degree of optimality of the TSP solution found by our approach.

As for the asymptotic complexity of different re-indexing schemes, let us observe that, while it can be proved that both Memon's and Zeng's algorithms have asymptotic time complexity of $O(M^3)$, our TSP approach requires $O(M^2 \log M)$ steps. Hence, our algorithm can be a good choice when coding images with large palettes.

VI. CONCLUSIONS AND FUTURE WORK

We have proposed an efficient re-indexing scheme for palette-based images. The re-indexing problem is translated into an optimization problem over a weighted graph and solved in an approximate fashion. The new indexing found by the proposed methodology exhibits smooth relative transitions in the indexed image, making it more suitable for lossless compression. The performances of our algorithm compare favorably with the performances of other approaches.

Future research aims at finding a better approximate algorithm for solving the related TSP problem. We believe that if a better approximation of the optimal solution can be found efficiently, this, in turn, would produce a stronger entropy decrease.

REFERENCES

[1] *Lossless and Near-Lossless Coding of Continuous Tone Still Images (JPEG-LS)*, July 1997. FCD 1449, ISO/IEC JTC1/SC29 WG1 (JPEG/JBIG).

[2] A. C. Hadenfelt and K. Sayood, "Compression of color-mapped images," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 442–453, May 1994.

[3] J. Fojtk and H. Vclav, "Invisible modification of the palette color image enhancing lossless compression," in *Proc. Joint IAPR Workshops*, vol. 1451, Sydney, Australia, 1998, pp. 1029–1036.

[4] M. J. Gormish, "Compression of palletized images by color," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, 1995.

[5] *Information Technology—JPEG 2000 Image Coding System—Part I*, 2000. ISO/IEC 15 444-1.

[6] N. Memon and A. Venkateswaran, "On ordering color maps for lossless predictive coding," *IEEE Trans. Image Processing*, vol. 5, pp. 1522–1527, Nov. 1996.

[7] L. M. Po and W. T. Tan, "Block address predictive color quantization image compression," *Electron. Lett.*, vol. 30, no. 2, pp. 120–121, Jan. 1994.

[8] A. Spira and D. Malah, "Improved lossless compression of color-mapped images by an approximate solution of the traveling salesman problem," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. III, Salt Lake City, UT, May 2001, pp. 1797–1800.
[9] A. Thomas. (2001) Solving the travelling salesman problem using a genetic algorithm. [Online]. Available: http://www.generation5.org/aisolutions/ga_tsp.shtml
[10] D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals*, ser. The Kluwer International Series in Engineering and Computer Science. Norwell, MA: Kluwer, 2002.
[11] P. Waldemar and T. A. Ramstad, "Subband coding of color images with limited palette size," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. V, Adelaide, Australia, Apr. 1994, pp. 353–356.
[12] A. Zaccarin and B. Liu, "A novel approach for coding color quantized images," *IEEE Trans. Image Processing*, vol. 2, pp. 442–453, Aug. 1993.
[13] W. Zeng and S. Lei, "Option of JPEG2000 for coding palettized images–a proposal," JTC1/SC29/WG1 N1453, Maui, HI, Dec. 1999.
[14] W. Zeng, J. Li, and S. Lei, "An efficient color re-indexing scheme for palette-based compression," in *Proc. IEEE Int. Conf. Image Processing*, to be published.
[15] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 1990.

Sebastiano Battiato received the Ph.D. degree in applied mathematics and computer science from Catania University, Catania, Italy, in 1999.

From 1999 to 2003, he was with STMicroelectronics, Catania, in the Advanced System Technology (AST) Catania Lab c/o the Imaging Group. He is currently a Researcher and Teaching Assistant in the Department of Mathematics and Informatics, University of Catania. His current research interests are in the areas of digital image processing, pattern recognition, and computer vision. He acts as a reviewer for several leading international conferences and journals, and he is the author of several papers and international patents.

Giovanni Gallo received the Doctorate degree in mathematics from Catania University, Catania, Italy, and the Ph.D. degree in computer science from New York University, New York, in 1992.

He has taught formal methods for computer science and computer human interaction at Catania University since 1992. His research interests include computer algebra, image processing, image database, medical images processing, and scientific visualization.

Gaetano Impoco received the degree in computer science from the University of Catania, Catania, Italy, in 2001. He is currently pursuing the Ph.D. degree in computer science at the University of Pisa, Pisa, Italy.

He is part of the Visual Computing Lab (ISTI-CNR), Pisa, which is involved in topics in the area of 3-D graphics and visualization. His current research interests concern automatic range map registration, sensor planning, and applications of 3-D models to cultural heritage.

Filippo Stanco (M'04) received the M.Sc. (*summa cum laude*) and Ph.D. degrees in computer science from the University of Catania, Catania, Italy, in 1999 and 2003, respectively.

He is currently a Research Assistant at the University of Trieste, Italy. His research interests include digital restoration, artifacts removal, filtering, super resolution, interpolation, and texture analysis. He is a Reviewer for several leading international conferences and journals.