FAST IDENTIFICATION OF JPEG 2000 IMAGES FOR DIGITAL CINEMA PROFILES

Osamu WATANABE†, Takahiro FUKUHARA , and Hitoshi KIYA‡*

†Takushoku University, Dpet. of Electronics & Computer Systems, 815–1, Tatemachi, Hachioji-shi, Tokyo, JAPAN
* Sony Corporation, CPDG PSG, 4–14–1, Asahi-cho, Atsugi-shi, Kanagawa, JAPAN
‡Tokyo Metropolitan University, Dept. of Info. and Commun. Systems Eng., 6–6 Asahigaoka, Hino-shi, Tokyo, JAPAN

ABSTRACT

A method for identifying JPEG 2000 images with different coding parameters, such as DWT-filters, code-block sizes, quantization step sizes, and resolution levels, is presented. The proposed method does not produce false-negative matches regardless of the different coding parameters. Moreover, the proposed method is fast because it uses the number of zero-bit-planes. This number can be extracted from the JPEG 2000 codestream by only parsing the header information. Experimental results showed the effectiveness of the proposed method for digital cinema applications.

Index Terms— JPEG 2000, Image identification, Image search, Digital Cinema

1. INTRODUCTION

Recently, the use of digital images and video sequences has greatly increased because of the rapid growth of the Internet and multimedia systems. It is often necessary to identify a certain image in a database that has a large amount of digital images. In connection, several international standards for searching still or moving images and retrieval systems have been developed [1–4]. In this work, "Identification" is defined as an operation for finding an image that is identical to a given original image from among compressed images. Database images are generally stored in compressed forms to reduce the amount of data.

JPEG 2000 [5] has been officially selected as the standard compression/decompression technology for digital cinema by the Digital Cinema Initiatives consortium [6]. The identification system used for digital cinema systems must be able to handle a large number of JPEG 2000-encoded frames in a sufficiently short processing time.

Several methods have been developed for identifying compressed images [7–9]. The method described in Ref. [7] is for JPEG images and uses the signs of the discrete cosine transform (DCT) coefficients of the images. One method for JPEG 2000 [8] , uses the signs of the discrete wavelet transform (DWT) coefficients. An algorithm for both JPEG 2000 and JPEG was proposed in Ref. [9]. Although these methods are for compressed images, they use transformed coefficients, which are not available without decoding. Codestream-based identification methods for JPEG 2000 images [10, 11] have also been proposed. Codestream-based means that there is no need for embedded block coding with optimized truncation (EBCOT) decoding, which is the most time consuming process in a JPEG 2000 decoder; therefore, such methods feature fast identification.

However, these codestream-based methods do not work with compressed images with different JPEG 2000 coding parameters, such as DWT filters, code-block sizes, DWT resolution levels, and quantization step sizes. This disadvantage is not trivial because the JPEG 2000 coding profiles for digital cinema defined by [12] may differ in coding parameters.

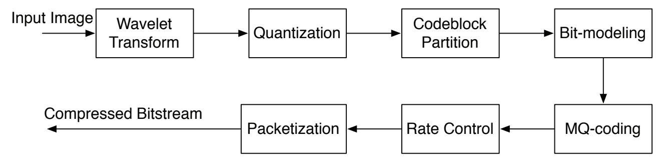

Fig. 1. Block diagram of JPEG 2000 encoder

To address this, a method for identifying JPEG 2000 images with different coding parameters is presented in this paper. The proposed method is fast because it uses the number of zero-bit-planes, which is obtained by parsing only the header part of a JPEG 2000 codestream. In addition, the proposed method does not produce false-negative matches, regardless of differences in the JPEG 2000 coding parameters. Moreover, by using the scalability of the JPEG 2000 codestream, scalable identification can be performed. Simple binary comparison of the compressed data does not have these features.

2. NUMBER OF ZERO-BIT-PLANES IN JPEG 2000

The number of zero-bit-planes is part of the header information of JPEG 2000 codestreams. This information is easily obtained by parsing the main header of JPEG 2000. In other words, extracting the number of zero-bit-planes does not require full JPEG 2000 decoding, i.e., it does not require heavy EBCOT decoding.

Figure 1 shows a block diagram of the JPEG 2000 encoder and Fig. 2 shows an example of an image I analyzed using DWT. R^I denotes the number of DWT resolution levels, which is three in Fig. 2, and $r^I = 1, 2, ..., R^I$ is the index of the resolution level. b denotes the index of the subband, where 1, 2, 3, and 4 are for LL, HL, LH, and HH, respectively. The subbands are divided into $X^I \times Y^I$ -sized code-blocks, and $J_1(r^I) \times J_2(r^I)$ code-blocks are in a subband of resolution level r^I.

Each quantized coefficient is separated into its sign and absolute magnitude, where the absolute magnitudes are factorized as bit-planes from the most significant bit to the least significant bit. All the samples in the bit-planes are either zero or one. A special bit-plane, in which all samples are zero, frequently appears. This special bit-plane is referred to as a "zero-bit-plane." When all bit-planes are zero-bit-planes, the code-block is defined as "not included" because it does not include any data to be encoded. Note that the numbers of zero-bit-planes in images compressed with different DWT filters and quantization step sizes may be different, even if the compressed images are generated from one original image and the other coding parameters are the same.

Hereafter, $Z^I_{r^I,b}[j_1, j_2]$ represents the number of zero-bit-planes in the code-block at $[j_1, j_2]$, where $j_1 = 1, 2, ..., J_1(r^I)$ and $j_2 = 1, 2, ..., J_2(r^I)$. Let $a[j_1, j_2] \equiv a[\boldsymbol{j}]$ denote the $J_1 \times J_2$ array, where $\boldsymbol{j} \in [1, J_1] \times [1, J_2]$, of code-blocks in a subband.

Therefore, $Z^I_{r^I,b}[\boldsymbol{j}]$ stands for the array of the number of zero-bit-planes in a subband.

3. PROPOSED METHOD

The proposed method consists of three steps. The first step defines regions to be compared. In the second step, the numbers of zero-bit-planes in the compared regions are derived. The third step compares the query and database images by using the numbers of zero-bit-planes in the regions.

It is assumed that query image Q has $X^Q \times Y^Q$-sized code-blocks and is decomposed to R^Q resolution levels, whereas database image D has $X^D \times Y^D$-sized code-blocks and consists of R^D resolution levels.

3.1. Definition of Regions to Be Compared

To normalize or unify the regions to be compared, this step defines the size of compared regions and the compared depth of resolution levels as shown in Fig. 3. The detail of this step is as follows,

1. By Eqs. (1) and (2), the numbers of code-blocks in regions to be compared are derived.

$$(M^Q, M^D) = \begin{cases} (\frac{X^D}{X^Q}, 1) & X^Q \le X^D \\ (1, \frac{X^Q}{X^D}) & others \end{cases} \tag{1}$$

$$(N^Q, N^D) = \begin{cases} (\frac{Y^D}{Y^Q}, 1) & Y^Q \le Y^D \\ (1, \frac{Y^Q}{Y^D}) & others \end{cases} \tag{2}$$

where $M^Q \times N^Q$ code-blocks form a region to be compared in query image Q, and $M^D \times N^D$ code-blocks from a region to be compared in a database image.

2. The compared depth of resolution levels is given by Eq. (3).

$$L = \begin{cases} R^Q & R^Q = R^D \\ R^Q - 1 & R^Q < R^D \\ R^D - 1 & R^Q > R^D \end{cases} \tag{3}$$

where L is the compared depth. Note that the comparable depth decreases by one without decoding under the condition that $R^Q \ne R^D$. The number of compared regions, c, is given as

$$c = \begin{cases} J_1^{min}(L)J_2^{min}(L) + \sum_{b=2}^{4}\sum_{l=1}^{L-1} J_1^{min}(l)J_2^{min}(l) & R^Q = R^D \\ \sum_{b=2}^{4}\sum_{l=1}^{L-1} J_1^{min}(l)J_2^{min}(l) & others \end{cases} \tag{4}$$

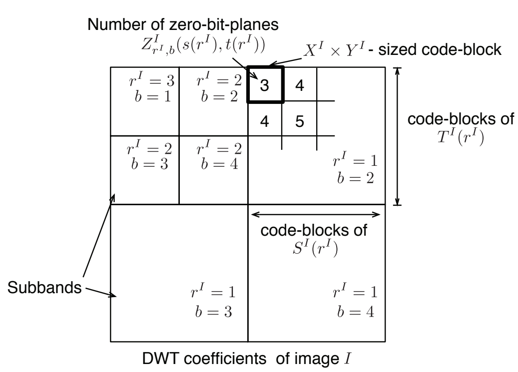

Fig. 2. Definition of subband, code-block, and resolution level, where number of DWT decomposition $R^I = 3$

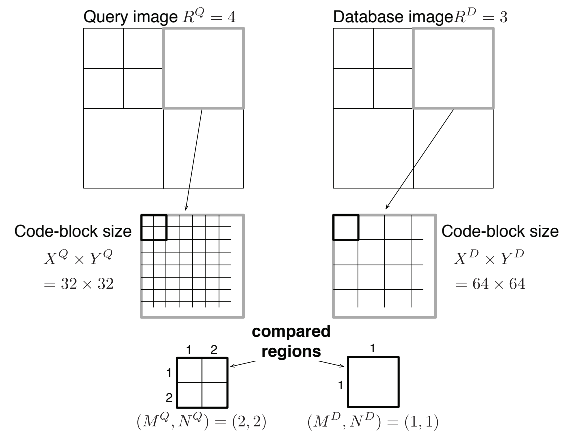

Fig. 3. Regions to be compared: the horizontal and vertical size of the regions are set to the least common multiple of M_Q and M_D, N_Q and N_D.

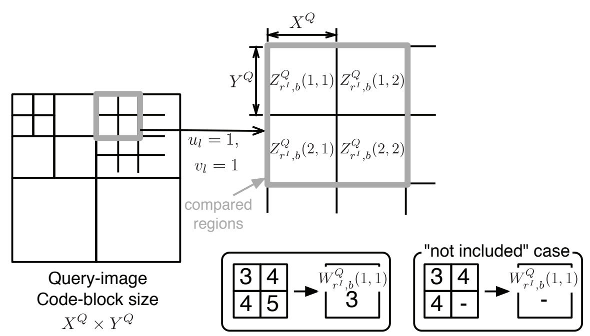

Fig. 4. Number of zero-bit-planes in regions to be compared: If multiple number of zero-bit-planes are obtained in the region, only the minimum value is used in subsequent processes. If there is "Not included", the minimum value is replaced by "Not included."

where $J_1^{min}(l) = \min_{I\in\{Q,D\}} J_1^I(l)$ and $J_2^{min}(l) = \min_{I\in\{Q,D\}} J_2^I(l)$.

3.2. Derivation of Numbers of Zero-Bit-Planes

This step derives the number of zero-bit-planes in all compared regions, i.e., in each compared region in subband b at the l-th resolution level, where $b = 1, 2, 3,$ and 4, and $l = 1, 2, \ldots, L$. The algorithm is as follows,

Algorithm 3.1: DERIVATENZBP(a, a)

for $l \leftarrow 1$ **to** L
  if $R^Q \ne R^D$
    then $b_{start} \leftarrow 2$
    else $b_{start} \leftarrow 1$
  for $b \leftarrow b_{start}$ **to** 4
    for each $\boldsymbol{j} \in [1, J_1^{min}] \times [1, J_2^{min}]$
      for $m^I = 1, 2, \ldots, M^I$
        for $n^I = 1, 2, \ldots, N^I$
          $W^I_{l,b}[\boldsymbol{j}] = \min_{m^I,n^I} Z^I_{l,b}[(j_1 - 1)M^I + m^I,$
          $(j_2 - 1)N^I + n^I]$

Note that the values $(j_1 - 1)M^I + m^I$ and $(j_2 - 1)N^I + n^I$ should

3	3	3	4	4	3	3	4
3	3	3	3	3	4	4	4
3	4	3	4	5	4	3	4
4	3	3	4	4	4	5	5

(a) 9x7 DWT filter

4	4	4	4	4	4	4	5
4	3	4	3	4	4	4	5
3	4	4	4	5	5	4	5
4	4	4	5	4	5	5	6

(b) 5x3 DWT filter

Fig. 5. Example of numbers of zero-bit-planes produced by different types of DWT filters: The values are obtained with exactly the same an input image and coding-parameters except DWT filter.

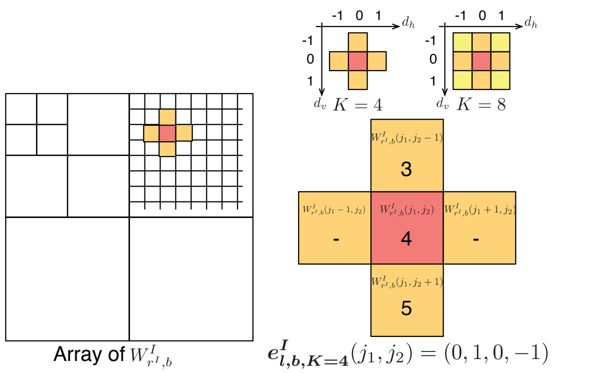

Fig. 6. Identification procedure: the number of zero-bit-planes in the focal and its k-th neighbouring regions are compared and the relation $e^I l, b, K$ is produced.

satisfy

$$1 \leq (u_l - 1)M^I + m^I \leq S^I(l), 1 \leq (v_l - 1)N^I + n^I \leq T^I(l). \quad (5)$$

Also note that if at least one $Z^I_{l,b}$ in the compared regions is "not included," then $W^I_{l,b}$ is also defined as "not included," as shown in Fig. 4.

3.3. Identification based on Numbers of Zero-bit-planes in Compared Regions

Identification is performed using the derived numbers of zero-bit-planes, $W^I_{l,b}[\boldsymbol{j}]$.

Figure 5 shows an example of numbers of zero-bit-planes produced with different DWT filters. Note that this example was obtained as the result of JPEG 2000 encoding with the same original image and the same coding parameters except the DWT filter. The numbers of zero-bit-planes clearly differ between Figs. 5(a) and 5(b).

The identification procedure starts with retrieval of the relations among the focal region and K (4 or 8) of its neighboring regions to compare images, as shown in Fig. 6. First, define the neighboring regions by the coordinate difference. That is, d_h and d_v indicate the horizontal and vertical distances between the focal region and the k-th neighboring region, respectively, c.f., Fig. 6.

The relation between the focal and the k-th neighboring region, $e^I_{l,b,k}$, is

$e^I_{l,b,k}(j_1, j_2) =$

$$\begin{cases} 1 & W^I_{l,b}(j_1, j_2) - W^I_{l,b}(j_1 + d_h, j_2 + d_v) > 0 \\ 0 & W^I_{l,b}(j_1, j_2) - W^I_{l,b}(j_1 + d_h, j_2 + d_v) = 0 \\ -1 & W^I_{l,b}(j_1, j_2) - W^I_{l,b}(j_1 + d_h, j_2 + d_v) < 0. \end{cases} \quad (6)$$

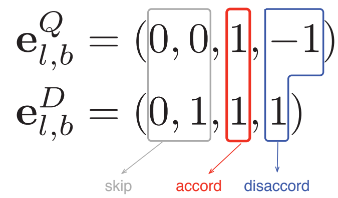

Fig. 7. Comparing relations between images: zero value of $\boldsymbol{e}$ is treated as "don't care."

Note that if either $W^I_{l,b}(j_1, j_2)$ or $W^I_{l,b}(j_1 + d_h, j_2 + d_v)$ is "not included," and if the neighboring coordinate exceeds the valid range, then $\boldsymbol{e}^I_{l,b,k} = 0$.

Now the relation values $e^I_{l,b,k}$ are ready to be compared. The decision is made by the comparing operation shown in Fig. 7. That is,

$$Decision[\boldsymbol{j}] = \begin{cases} Positive & \boldsymbol{e}^Q_{l,b,k} = \boldsymbol{e}^D_{l,b,k} \\ Negative & \boldsymbol{e}^Q_{l,b,k} \neq \boldsymbol{e}^D_{l,b,k} \end{cases} \quad (7)$$

If the values $Decision[\boldsymbol{j}]$ are "Positive" for all possible $\boldsymbol{j}$, the query and database images are considered to be identical.

3.4. Features

The features of the proposed method are as follows.

(A) Robustness against the difference in coding parameters: the processing step described in 3.1 makes it possible to identify images having different code-block sizes and/or DWT resolution levels. Images having different DWT filters and/or quantization step sizes can be identified with the processing step described in 3.3. These steps are easily combined. Moreover, the method does not produce false-negative matches regardless of the differences in any of the coding parameters. Conventional methods does not have this feature.

(B) Fast processing: The number of zero-bit-planes can be extracted without EBCOT decoding, and the data amount is much smaller than that of the whole codestream, so the proposed method is very fast.

4. EXPERIMENTAL RESULTS

The performance of the identification methods in terms of their precision and processing speed of the image identification were evaluated to verify the effectiveness of the proposed method.

4.1. Experimental Condition

In the experiment, the Standard Evaluation Test Material (StEM) [13] was used. The test set contained 14,964 frames (there are more original frames in StEM, but those images, with visually full black-color, were rejected from the test sequence). First, a sequence of JPEG 2000 compressed images was stored in a database. A query image was found and identified from the JPEG 2000 images in the database. The identification experiment was performed for all possible combinations of a query and a database image. The number of combinations of a query image and a database one was $14,964 \times 14,964$ in total. The encoding parameters for query images are listed in Table 1, and Table 2 lists the parameters for database images. In Table 2, "DWT53," "CB,""Res," and "Qstep" mean images encoded using different DWT filters, code-block sizes, DWT resolution levels, and quantization base step sizes from query images, respectively. "DWT53"

Table 1. Test sequence and JPEG 2000 encoding parameters for query images

Test sequence	StEM (DCI standard), 14,964 frames
Resolution	4,096 (H)× 1,740 (V)
Format	RGB(4:4:4) 12 bits/component
Rate Control	VBR (Variable Bit Rate)
Codestream	DCI Compliant (JPEG 2000 Part-1) DWT Filter: 9 × 7, 5 × 3 DWT Level: 5, 4 Base Step Size: 1/256, 1/200 Code-block Size = 32 × 32, 64 × 64, 128 × 32 No tile decomposition

Table 2. JPEG 2000 encoding parameters for database images

	DWT filter	DWT level	Base Step Size	Code-block size
DWT53	5 × 3	5	1/256	32 × 32
CB64	9 × 7	5	1/256	64 × 64
CB128	9 × 7	5	1/256	128 × 32
Res	9 × 7	4	1/256	32 × 32
Qstep	9 × 7	5	1/200	32 × 32

database consists of JPEG 2000 images encoded losslessly. This means all images in this database are encoded by using 5×3 DWT filter. For "CB" databases, two types of code-block size (64 × 64=CB64 and 128 × 32=CB128) were used. For "Res" databases, the number of DWT decompositions was set to four. For "Qstep" databases, the number of neighboring code-blocks was set to $K = 4$ and $K = 8$.

4.2. Experimental results

The false-positive rate (FPR) and true-positive rate (TPR) [14] of all trials are shown in Table 3. These rates are calculated as

$$FPR = FP/(FP + TN) \quad (8)$$

$$TPR = TP/(TP + FN) \quad (9)$$

where FP, TN, TP, and FN are the number of false-positive, true-negative, true-positive, and false-negative results of the identification, respectively. Table 3 shows that the identification based on the proposed method produced an under-1.0% FPR regardless of the difference of JPEG 2000 coding parameters and without any false-negative matches. Figure 8 shows the ROC curve [14]. In this figure, the performance of identifier diverges from the solid line. A plotted point placed above the line means that its performance is good. Note that all the proposed method's points are placed above the divergence line.

The experiment described in the previous subsection was performed on a workstation with a Xeon 2.50 GHz processor and 4 GB memory. The average processing time for a combination of a query and a database image was about 0.6 [ms/frame] excluding disk access time.

Table 3. False-positive rate (%) and true-positive rate (%): calculated using Eqs. (8) and (9)
The symbol "N/A" in the table means there were no available results because the conventional method could not be used for images having different coding parameters. A TPR of 0.0 (%) means the conventional method could not identify query frames at all.

		Proposed		Conventional [10]	
		FPR(%)	TPR(%)	FPR(%)	TPR(%)
DWT53	($K = 4$)	0.93	100	0.04	0.0
	($K = 8$)	0.84	100	0.04	0.0
CB64		0.55	100	N/A	N/A
CB128		0.60	100	N/A	N/A
Res		0.41	100	N/A	N/A
Qstep	($K = 4$)	0.79	100	0.04	0.0
	($K = 8$)	0.72	100	0.04	0.0

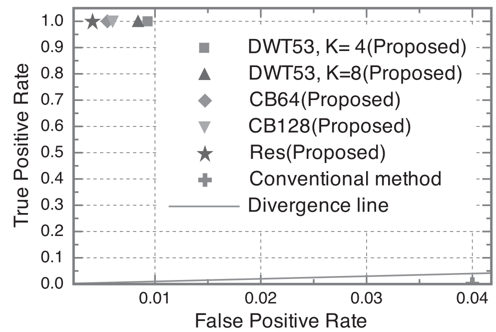

Fig. 8. ROC curve of experimental results

5. CONCLUSION

A zero-bit-plane-based identification method for JPEG 2000 images with different JPEG 2000 coding parameters has been presented in this paper. The proposed method is fast because it uses the extracted number of zero-bit-planes without requiring EBCOT decoding. Moreover, the method does not produce false-negative matches, regardless of the difference of JPEG 2000 coding parameters.

6. REFERENCES

[1] "Information technology – JPSearch – Part 1: System framework and components," International Standard ISO/IEC TR-24800-1, Dec. 2007.

[2] "Compact Descriptors for Visual Search: Applications and Use Scenarios," ISO/IEC JTC1/SC29/WG11/N11529, July 2010.

[3] "Compact Descriptors for Visual Search: Context and Objectives," ISO/IEC JTC1/SC29/WG11/N11530, July 2010.

[4] "Compact Descriptors for Visual Search: Requirements," ISO/IEC JTC1/SC29/WG11/N11531, July 2010.

[5] "Information technology — JPEG 2000 image coding system – Part 1: Core coding system," International Standard ISO/IEC IS-15444-1, Dec. 2000.

[6] "Digital Cinema System Specification V1.2," Digital Cinema Initiatives, LLC Technology Committee, Mar. 2008.

[7] F. Arnia, I. Iizuka, M. Fujiyoshi and H. Kiya, "Fast and robust indentification methods for jpeg images with various compression ratios," in *Proc. of IEEE ICASSP*. IEEE, May 2006.

[8] O. Watanabe, A. Kawana and H. Kiya, "An Identification Method for JPEG2000 Images Using the Signs of DWT Coefficients," *Technical Report of IEICE*, vol. 106, no. IE2006-217, pp. 177–181, Jan. 2007.

[9] K. O. Cheng, N. F. Law, W. C. Siu, "A fast approach for identifying similar features in retrieval of JPEG and JPEG 2000 images," in *Proc. APSIPA ASC*, Oct. 2009, number MP-P2-3.

[10] T. Fukuhara, K. Hosaka and H. Kiya, "Accurate identifying method of JPEG2000 images for digita cinema," in *Proceedings of The 14th International Multimedia Modeling Conference (MMM'08)*. Lecture Notes in Computer Science, Jan. 2008, vol. 4903.

[11] O. Watanabe, T. Iida, T. Fukuhara and H. Kiya, "Identification of JPEG 2000 images in encrypted domain for Digital Cinema," in *Proc. of IEEE ICIP*, Nov. 2009, number MA.PJ.8, pp. 2065–2068.

[12] "Information technology – JPEG 2000 image coding: Core coding system AMENDMENT 2: Extended profiled for cinema and video production and archival applications," ISO/IEC 15444-1:2004/Amd 2, 2009.

[13] Digital Cinema Initiatives, LLC Technology Committee, "StEM Access Procedures," http://www.dcimovies.com/StEM/, Sept. 2010.

[14] T. Fawcett, "Using Rule Sets to Maximize ROC Performance," in *Proc. of IEEE ICDM*. IEEE, 2001, pp. 131–138.