PAPER

DCT Sign-Based Similarity Measure for JPEG Image Retrieval

Fitri ARNIA[†a)], **Ikue IIZUKA**[†*], ***Student Members***, **Masaaki FUJIYOSHI**[††], ***and*** **Hitoshi KIYA**[††], ***Members***

SUMMARY We propose a method to retrieve similar and duplicate images from a JPEG (Joint Photographic Image Group) image database. Similarity level is decided based on the DCT (Discrete Cosine Transform) coefficients signs. The method is simple and fast because it uses the DCT coefficients signs as features, which can be obtained directly after partial decoding of JPEG bitstream. The method is robust to JPEG compression, in which similarity level of duplicate images, i.e., images that are compressed from the same original images with different compression ratios, is not disguised due to JPEG compression. Simulation results showed the superiority of the method compared to previous methods in terms of computational complexity and robustness to JPEG compression.

key words: *image similarity, DCT coefficients signs, image retrieval, JPEG, JPEG 2000*

1. Introduction

A large volume of digital images are available in the compressed domain recently. Therefore, it is more advantageous if image retrieval can be accomplished in the compressed domain, in which full image decoding is not required. In addition, different applications demand different image availability. For example, certain applications may require to retrieve similar images including their compressed versions, others may need to find all compressed versions of an image. Retrieval methods should address these conditions to ensure that all desired images can be retrieved.

Several image retrieval methods [1]–[9] have been proposed previously. The method described in [1]–[3] is aimed at retrieving JPEG images that are compressed from the same original image with different compression ratios, which we referred to as duplicate images. The method uses the DCT coefficients signs as image features. The DCT coefficients signs are not changeable because of scalar quantization employed in JPEG compression. Furthermore, different JPEG compression ratios result in a different number of zero coefficients due to different quantization step sizes, but maintaining the signs of non-zero coefficients. Utilizing these properties, the method is robust to JPEG compression, in which duplicate images that are generated from the same original image and have equal or different (JPEG) compression ratios can be retrieved. Moreover, the method is simple, because there is no additional processing required to extract image features. The features can be obtained directly after partial decoding of bitstream. The limitation of the method described in [1]–[3] is that it is not able to retrieve similar images.

Other methods [5]–[9] are mainly designed to acquire similar JPEG and JPEG 2000 (JP2) images. Among them, Chang et al. [5] presents the most similar approach to our proposed method. However, Chang's method utilizes both magnitudes and signs of AC and DC coefficients. The method processes those coefficients in several steps to obtain image features. Since JPEG compression alters the magnitude of AC and DC coefficients, it is presumed that the method may not be robust to JPEG compression. Furthermore, because the image features are obtained through a sequence of operations, longer processing time is needed.

In this paper, we propose a novel method for retrieving images based on the DCT coefficients signs that solves limitations of the previous methods [1]–[3], [5]. The proposed method is able to retrieve both similar images and duplicate images that are generated from the same original images with various compression ratios, in one retrieval task. The latter is referred to as identification. Being able to accomplish these, the method is said to be robust to JPEG compression. Robustness to JPEG compression can be achieved due to the method using the DCT coefficients signs as features, as in [1]–[3]. Scalar quantization applied in the JPEG compression changes the number of zero coefficients, but does not change the DCT coefficients signs. Furthermore, the method is simple because the DCT coefficients signs can be obtained without any additional calculation. They can be directly obtained by only entropy decoding the bitstream or following the DCT calculations.

The remaining of this paper is organized as follows. The first part of Sect. 2 describes decoding of DCT coefficients from JPEG bitstream. It will be followed by a review on previous work that uses the DCT coefficients signs to retrieve duplicate JPEG images. The proposed method is discussed in Sect. 3. Finally, results of simulations and conclusions are presented in Sects. 4 and 5, respectively.

Manuscript received September 14, 2006.
Manuscript revised February 25, 2007.
Final manuscript received May 15, 2007.

[†]The authors are with the Graduate School of Engineering, Tokyo Metropolitan University, Hino-shi, 191–0065 Japan.
[††]The authors are with the Faculty of System Design, Tokyo Metropolitan University, Hino-shi, 191–0065 Japan.
[*]Presently, with Hitachi, Ltd., Japan.
a) E-mail: arnia@ieee.org

DOI: 10.1093/ietfec/e90–a.9.1976

2. Background

2.1 Decoding of DCT Coefficients Signs from JPEG Bitstream

Procedure of JPEG decoding will be explained by firstly reviewing the encoding process, since the decoding procedure are simply the reverse version of encoding process.

The JPEG encoding procedure is shown in Fig. 1 and can be summarized as follows [10]: The original image is tiled into 8×8 non-overlapped blocks, and a two-dimensional Forward Discrete Cosine Transform (FDCT) function is applied to each block, producing 1 DC and 63 AC coefficients. Each coefficient is then quantized. Afterwards, according to the zig-zag scan, Differential Pulse Code Modulation (DPCM) is applied on DC coefficient, and AC coefficients are run-length coded (RLC). Finally, all the coefficients are entropy (Huffman or arithmetic) encoded. The output of the entropy encoder and some additional information form the JPEG bitstream.

A more detailed representation of a JPEG bitstream is shown in Fig. 2. It is a structure of an interchange format of a gray scale image. The Start of Image (SOI) marker is the first marker in the JPEG stream. The next two entries are the tables that are required to decode the image, namely: quantization and Huffman tables. The Huffman code contains the compressed image data bitstream. Eventually, the End of Image (EOI) marker follows the last byte of the compressed data.

The compressed image data bitstream can be regarded as a collection of DCT coefficients, including signs and magnitudes. Each DCT coefficient sign is obtained by the following procedures. The decoder determines run/size of a given segment of a bitstream by searching the Huffman tables. Knowing the run/size, the next step is to obtain the coefficient exact value (sign and amplitude), which is defined as an additional bit. Positive and negative signs are represented as two's complement-related number [11]. For example, if bit "0" precedes a segment of bits, it represents a negative sign, and vice versa. Therefore, positive and negative signs can be extracted by only Huffman (run/size) decoding, followed by signs evaluation.

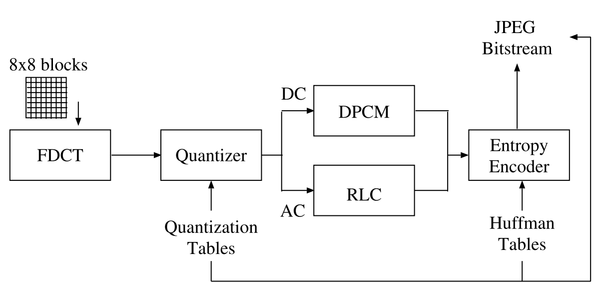


Fig. 1 Processing steps for JPEG DCT-based encoder.

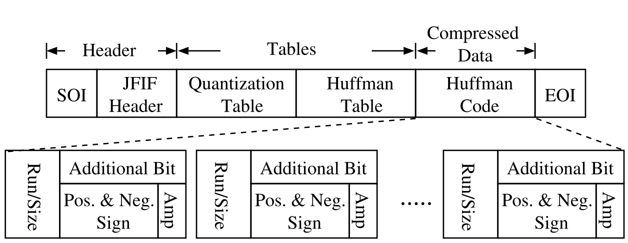


Fig. 2 JPEG bitstream.

2.2 Related Works

A. Retrieval Method Based on the DCT Coefficients Signs

Previous work that uses DCT coefficients signs as image features is aimed at retrieving duplicate JPEG images that are generated from the same original image with different compression ratios [1]–[3]. The desired JPEG images are obtained by comparing DCT coefficients signs of query image and images in the database. Scalar quantization employed in JPEG compression does not change the DCT coefficients signs. Furthermore, different JPEG compression ratios results in different number of zero coefficients due to different quantization step sizes, but maintaining the signs of non-zero coefficients. Due to these, the method that uses the DCT coefficients signs as image features is robust to JPEG compression. In addition, the method is simple, because there is no additional processing required to extract image features. The features can be obtained directly after Huffman decoding. Despite several advantages of [1]–[3], it is not able to retrieve similar images.

B. Retrieval Method Based on the DCT Coefficients

Chang's method [5] is considered as another similar approach to the proposed method. To obtain image features, the method applies Huffman decoding to DC and AC coefficients, the same as the proposed and the related [1]–[3] methods. The Huffman decoding is then followed by run-length decoding of AC coefficients, which is not required by the proposed method. In particular, Chang's method needs to re-obtain the DCT coefficients in the quantized domain. The method extracts DC and AC features separately. The DC features of an image are obtained by following steps. The difference values of the DC coefficients of subsequent 8×8-DCT blocks are taken. If a difference value is greater than or equal to zero, then it will be replaced by "1." Otherwise, it will be replaced by "0." The AC features can be extracted in a similar way to that of DC coefficients, yet coefficient value differences are taken within each 8×8-DCT blocks. Then these 1s and 0s indices are treated as a binary value and converted to decimal number. Finally, for each image Chang's method produces a histogram using the converted number of each block.

To obtain similarity level of images, two distance are calculated, namely DC and AC distances. These calculations involved a series of Exclusive-OR (XOR), subtraction and addition operations. In addition to the extensive operations of feature extractions, Chang's method uses DC and AC coefficients magnitudes, which can be altered due to JPEG compression. As a consequence, the method may not be robust to JPEG compression.

3. Proposed Methods

3.1 Model of Retrieval from JPEG Image Database

Model of image retrieval system from a JPEG database is shown in Fig. 3. Depending on the query images, querying can be classified into three kinds, namely querying with

1. Images with identical compression format (in this case JPEG) and identical size.
2. Images with different compression format and may be different size (we select JP2 image).
3. Others, which include raw image format.

For querying with JPEG images, it is only required to partially decode the JPEG bitstream. As explained in Sect. 2.1, the features, that is the DCT coefficients signs $\mathbf{S}_Q$ can be obtained by merely Huffman decoding the bitstream, followed by signs evaluation. For querying with different compression format, JP2 images is taken into consideration due to its spreading use and increasing popularity. The JP2 uses the wavelet transform instead of the DCT. In JP2 format, several thumbnail image can be obtained using its scalability feature. Querying with JP2 image is started by preprocessing the image. The processes are illustrated in the dashed box in Fig. 3. For querying with the same image sizes, the query image is first decoded, then the 8×8 DCT is applied to the decoded image to extract the DCT coefficients signs $\mathbf{S}_Q$. For querying with a JP2 thumbnail image, interpolation is conducted prior to applying the DCT. If the query is a raw image, the DCT coefficients signs $\mathbf{S}_Q$ can be extracted following calculation of the DCT coefficients.

In all kinds of querying, the DCT coefficients signs of JPEG image from database $\mathbf{D}$ are extracted in the same way as the JPEG query image. Similarity between $\mathbf{S}_Q$ and $\mathbf{S}_D$ is calculated finally.

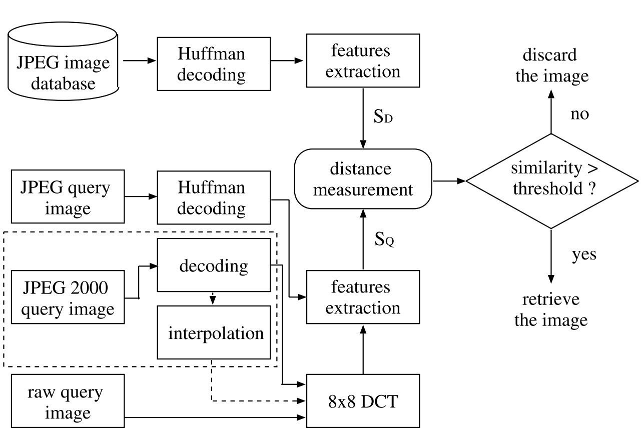


Fig. 3 Model of image retrieval from a JPEG image database.

3.2 Notations and Terminologies

Unless otherwise stated, several notations and terminologies that are used in the following sections are listed here.

- Q represents the query image and D represents the database's image.
- $\mathbf{S}_Q$ and $\mathbf{S}_D$ represent the DCT coefficients signs obtained from query image Q and database image D respectively.
- M represents the number of constituent 8×8 block of an image. $0 \leq m \leq M-1$.
- N represents position of DCT coefficient according to the zig-zag scanning in each 8×8 block. $0 \leq N \leq 63$. Each coefficients is obtained according to zig-zag scanning.
- $\zeta_{Q,D}(m)$ is a counter of non-zero DCT coefficients signs of image Q and D in the mth block.
- $\zeta_Q(m)$ is a counter of non-zero DCT coefficients signs of image Q in the mth block.
- $\chi_{Q,D}(m)$ is a counter of the same DCT coefficients signs of image Q and D in the mth block.

3.3 Retrieval Algorithm

The retrieval method is accomplished according to the following steps:

1. Set $m := 0$.
2. Set $\zeta_{Q,D}(m) := 0, \chi_{Q,D}(m) := 0$ and $n := 0$.
3. If $S_Q(m,n) \neq 0$ and $S_D(m,n) \neq 0$, $\zeta_{Q,D}(m) := \zeta_{Q,D}(m)+1$.
   If $S_Q(m,n) = 0$ or $S_D(m,n) = 0$, proceed to step 5.
4. If $S_Q(m,n) = S_D(m,n), \chi_{Q,D}(m) := \chi_{Q,D}(m) + 1$.
   If $S_Q(m,n) \neq S_D(m,n), \chi_{Q,D}(m) := \chi_{Q,D}(m) - 1$.
5. Set $n := n+1$. If $n \leq 63$, proceed to step 3.
6. Normalize $\chi_{Q,D}(m)$ with respect to $\zeta_{Q,D}(m)$.

$$r_{Q,D}(m) = \frac{\chi_{Q,D}(m)}{\zeta_{Q,D}(m)}, \tag{1}$$

   here, if $\zeta_{Q,D}(m) := 0$, then set $r_{Q,D}(m) := 1, m := m+1$. If $m \leq M-1$, proceed to step 2.
7. Take average $\sigma_{Q,D}$ of $r_{Q,D}(m)$ with respect to M,

$$\sigma_{Q,D} = \frac{1}{M}\sum_{m=0}^{M-1} r_{Q,D}(m). \tag{2}$$

   $\sigma_{Q,D}$ is defined as a similarity level between images Q and D.

The above algorithm depends on the non zero DCT coefficients signs of both Q and D, and is referred to as the "NZ-QD." This method can retrieve both similar and duplicate images.

For retrieving only similar images with better performance (with less ability of identifying duplicate images), we introduce to consider only the non-zero DCT coefficients signs of Q (we refer to this method as the "NZ-Q"). This can be achieved by slightly modifying the above algorithm, as follows :

- Replacing step no.2 by:
  Set $\zeta_Q(m) := 0$, $\chi_{Q,D}(m) := 0$ and $n := 0$.
- Replacing step no.3 by:
  If $S_Q(m,n) \neq 0$, $\zeta_Q(m) := \zeta_Q(m) + 1$. If $S_Q(m,n) = 0$ or $S_D(m,n) = 0$, proceed to step 5, and
- Replacing step no.6 by:
  Normalize $\chi_{Q,D}(m)$ with respect to $\zeta_Q(m)$.

$$r_{Q,D}(m) = \frac{\chi_{Q,D}(m)}{\zeta_Q(m)}, \tag{3}$$

here, if $\zeta_Q(m) := 0$, then set $r_{Q,D}(m) := 1$, $m := m + 1$. If $m \leq M - 1$, proceed to step 2.

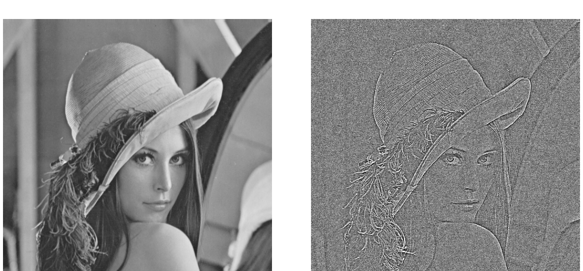
All the above-mentioned processes, for the "NZ-QD" or "NZ-Q," are applied to all images in the database **D** to obtain their similarity level. As a matter of fact, $\chi_{Q,D}(m)$ which is determined in step 4 can be calculated by following equation:

$$\chi_{Q,D} = \sum_{n=0}^{N-1} S_Q(m,n) \cdot S_D(m,n). \tag{4}$$

That is the sum of DCT coefficients signs product of query image Q and database image D. There are only two possible values of sign product values, namely "1" and "−1." If both $S_Q(m,n)$ and $S_D(m,n)$ are the same, their product will be "1," and if $S_Q(m,n)$ is different from $S_D(m,n)$, their product will be "−1." This is in accordance with step 4.

3.4 Similarity Measure Based on DCT Coefficients Signs

The DCT coefficients signs of an image contain a significant amount of information of the image. A DCT coefficients signs only image (DSOI) preserves information such as outline and edges of the original image. Figure 4 illustrates this condition. Figure 4(a) is the original image and (b) is the DSOI of image (a). We can recognize the original image by observing its DSOI. Therefore, similar images may be retrieved by considering only their DCT coefficients signs.



(a) Original image. (b) DCT sign only image (DSOI).

Fig. 4 Image of DCT coefficients signs.

From JPEG compression point of view, incorporating only DCT coefficients signs is a benefit. It is known that different JPEG compression ratios result in different number of zero-valued DCT coefficients (because of different quantization step sizes), but they do not change the signs of DCT coefficients. The proposed method utilizes these properties to obtain a simple method, yet robust to JPEG compression.

The similarity of two images is measured based on the count of equivalent DCT coefficients signs of the images, which are tiled into 8×8 non-overlapped blocks. Equation (4) in Sect. 3.3 calculates the similarity level of two 8×8-DCT blocks. Each pair of equivalent DCT coefficients signs contributes to the similarity increment and vice versa. Images similarity is obtained by averaging the similarity of all 8×8 blocks.

3.5 Computational Complexity

The number of calculations required to retrieve similar images, which is commonly referred to as computational complexity, includes features extraction calculation and similarity level calculation.

We compared computational complexity of the proposed method with Chang's method, since the latter proposed a similar approach to our method. We also compared the proposed method with pixel correlation techniques. Pixel correlation is a simple and widely-acceptable similarity measure and the same as the proposed method, it does not require feature calculation. Therefore, we considered it for comparison purpose. Correlation is given by

$$r = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \cdot \sum\left(Y_i - \overline{Y}\right)^2}}, \tag{5}$$

with $\overline{X}$ denote pixel value average of image X. Value $r = 0$ indicates no linear relationship between image X and Y, and $r = 1$ suggests a strong linear relationship between image X and Y. A summary of computational complexity of the proposed, Chang's and pixel correlation method for querying one image is presented in Table 1.

For features extraction (Table 1(a)), the proposed method does not require additional calculation. The signs of DCT coefficients can be extracted after the Huffman de-

Table 1 Computational complexity. M is total of 8×8 blocks in an image.

(a) Features extraction.

Operation	Proposed	Chang's	Pixel Corr.
±	-	$9M$	-
No. of Trans.	-	M	-

(b) Similarity level calculation.

Operation	Proposed	Chang's	Pixel Corr.
±	M	$M + 256$	$448M$
⊕	-	$M + 256$	-
÷/×	$M + 1$	-	$192M$

coding process (when querying with JPEG images) or following the DCT calculations (when querying with other formats). Pixel correlation also requires no features calculation, however a full decoding process (Huffman, DPCM and RLC decoding) and inverse IDCT, have to be conducted instead. Compared to the proposed method, pixel correlation requires a lot more computation. Calculations required to fully decode the image are not shown in the table. On the other hand, Chang's method requires as many as $9M$ additions/subtractions, and M transformation from binary to decimal number. Here, M is the total number of 8×8 blocks in an image.

For similarity level calculation (Table 1 (b)), the proposed method requires as many as M additions/subtractions and and $M + 1$ divisions/multiplications. While Chang's method requires $M + 256$ additions/subtractions and XOR operations. Pixel correlation is the most calculation-consuming method. It requires about as many as $448M$ additions/subtractions and $192M$ divisions/multiplications.

As an illustration, let us consider an images with size of 512×512 pixels, therefore $M = 4096$. The computational complexity of the proposed method is 4096 additions/subtractions and 4097 multiplications/divisions. Chang's method requires as many as 41216 additions/subtractions and 4352 XOR operations. While pixel correlation, excluding the decoding calculations, requires 1835008 additions/subtractions and 786432 multiplications/divisions.

4. Simulation

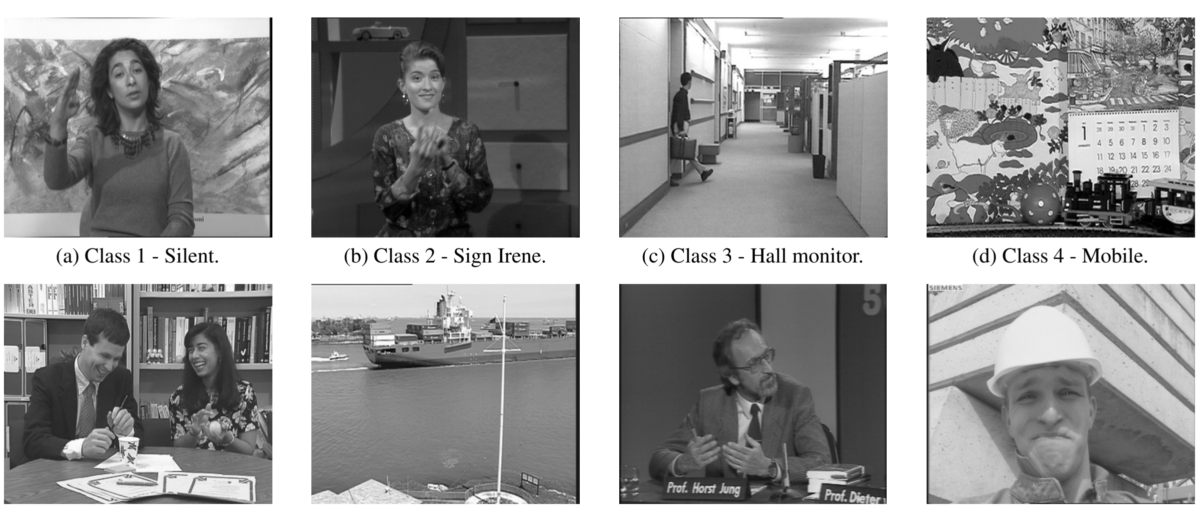
To evaluate the performance of the proposed method, we built an image database that comprised of 8 individual video sequences with different sceneries. Details of the database are summarized in Table 2, and frame samples of each sequence are shown in Fig. 5. All frames were JPEG compressed with three different QFs, 50, 200, and 350 respectively. The "mobile" is a sequence with camera panning, while "diskus" is a sequence with zooming operations. Figures 6 and 7 illustrate these operations as the frames progress through the "mobile" and "diskus" sequences, respectively. Here, "sequence" frame no. x is represented as "sequence-x," and "sequence" frame no. x with QF y will be referred to as "sequence x-y." The same notations will be used in the remaining sections. Each sequence was referred to as a class, in which similar images were grouped together. Each class consisted of 100 images (frames) and was represented by 2 queries. Totally there were 16 queries, each with compression ratio or Q-factor (QF) of 50. The query images are listed in Table 3.

Four methods were evaluated: the "NZ-QD," "NZ-Q," Chang's, and pixel correlation, respectively. The methods under evaluation use similarity measure to decide similar images, except Chang's method that uses distance measure. When using the similarity measure, the most similar image will have similarity level equal to "1." Conversely, in distance measure, the most similar image will have distance value equal to "0."

Simulations were conducted to test the performance of the proposed method considering the frame-work mentioned in the introduction part, namely: retrieving (1) similar images and (2) images that are compressed from the same original image with different compression ratios (identification

Table 2 Image database. Each QF is composed of 800 images.

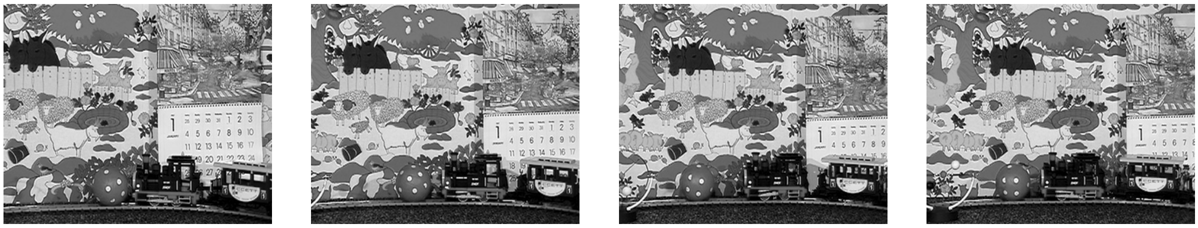
Database	gray scale, 352×288, 8 bpp
class 1	silent (frame no. 1–100)
class 2	sign irene (frame no. 101–200)
class 3	hall monitor (frame no. 201–300)
class 4	mobile(frame no. 301–400)
class 5	paris(frame no. 401–500)
class 6	container(frame no. 501–600)
class 7	diskus(frame no. 601–700)
class 8	foreman(frame no. 701–800)
Q-factor	50, 200, 350



(a) Class 1 - Silent. (b) Class 2 - Sign Irene. (c) Class 3 - Hall monitor. (d) Class 4 - Mobile.

(e) Class 5 - Paris. (f) Class 6 - Container. (g) Class 7 - Diskus. (h) Class 8 - Foreman.

Fig. 5 Frame samples from "collected sequence."



(a) “mobile-325.” (b) “mobile-350.” (c) “mobile-375.” (d) “mobile-400.”

Fig. 6 Camera panning in “mobile.”

(a) “diskus-660.” (b) “diskus-673.” (c) “diskus-687.” (d) “diskus-700.”

Fig. 7 Zooming in “diskus.”

Table 3 List of query images.

No	Query Name	No	Query Name
1	silent30-50	9	paris430-50
2	silent70-50	10	paris470-50
3	sign130-50	11	container535-50
4	sign160-50	12	container570-50
5	hall225-50	13	diskus635-50
6	hall275-50	14	diskus670-50
7	mobile330-50	15	foreman735-50
8	mobile370-50	16	foreman770-50

of duplicate images). For those purposes, the simulations were organized into four parts, namely, simulations of:

- Global performance of “NZ-QD.”
- Similar image retrieval by querying JPEG image.
- Duplicate image retrieval by querying JPEG image.
- Similar image retrieval by querying JP2 image.

4.1 Global Performance of the “NZ-QD”

A global performance of a similarity measure can be evaluated if average similarity and variance between image classes are available. Here, we calculated the average similarity and variance based on Eqs. (6), (7), and (8) given in the following.

Let $p_{C_m,i}$ and $q_{C_n,j}$ be the i-th image of class “C_m” and the j-th image of class “C_n” respectively. Similarity between images in class “C_m” and “C_n” is defined in a way described by Eq. (6),

$$sim(p_{C_m,i}, q_{C_n,j}) \quad i = 1, \ldots, |C_m|, \quad j = 1, \ldots, |C_n|, \quad m, n = 1, \ldots, \text{number of classes} \tag{6}$$

Here, $sim(p_{C_m,i}, q_{C_n,j})$ was calculated based on the “NZ-QD” method in Eq. (2). Similarity average μ_{C_m,C_n} and variance $\sigma^2_{C_m,C_n}$ between class “C_m” and “C_n” were calculated by Eqs. (7) and (8), respectively.

$$\mu_{C_m,C_n} = \frac{1}{|C_m||C_n|} \sum_{i=1}^{|C_m|} \sum_{j=1}^{|C_n|} sim(p_{C_m,i}, q_{C_n,j}) \tag{7}$$

$$\sigma^2_{C_m,C_n} = \frac{1}{|C_m||C_n|} \sum_{i=1}^{|C_m|} \sum_{j=1}^{|C_n|} (sim(p_{C_m,i}, q_{C_n,j}) - \mu_{C_m,C_n})^2 \tag{8}$$

If the μ_{C_m,C_n} becomes larger, the images between class C_m and C_n become more similar in average. It is worth noting that when $m = n$, Eq. (7) measures the self similarity of images within the same class, referred to as the *intra-class similarity*. And when $m \neq n$, Eq. (7) measures the image similarity between classes and referred to as *inter-class similarity*.

The global performance of the “NZ-QD,” which was calculated using Eqs. (6), (7), and (8), are provided in Table 4. For each class (C_m or C_n), the averaged intra-class similarity μ_{C_m,C_m} is the largerst among the inter-class similarity μ_{C_m,C_n}. As shown, for sequences that are obtained by steady camera (e.g. “hall monitor” and “silent”), the proposed similarity measure resulted in a higher intra-class similarity μ_{C_m,C_m}. In this case, $\mu_{C_3,C_3} = 0.9102$ for “hall monitor” and $\mu_{C_1,C_1} = 0.7908$ for “silent.” On the other hand, sequence that is obtained by free-hand camera (“foreman”) has lower μ_{C_m,C_m}, that is $\mu_{C_8,C_8} = 0.3963$. Sequences that are obtained by steady camera but with panned and zoomed frames (“mobile” and “diskus”) also have lower μ_{C_m,C_m}. In this case, $\mu_{C_4,C_4} = 0.1188$ for “mobile” and $\mu_{C_7,C_7} = 0.6625$ for “diskus.” Therefore, the proposed similarity measure defines frames in “hall monitor” are more similar with each other than those in “mobile” or “diskus.”

The inter-class similarity is measured by the value of μ_{C_m,C_n}. Two classes are considered similar if the μ_{C_m,C_n}

Table 4 Global performance of the "NZ-QD." 1-st row and 1-st column are class number. QFs of the images: 50.

Class	1	2	3	4	5	6	7	8	
1	0.7908	−0.1232	0.0711	−0.0154	0.0131	0.0986	−0.1118	0.1661	μ
	0.0038	0.0003	0.0001	0.0003	0.0001	0.0002	0.0037	0.0003	σ^2
2	−0.1232	0.7602	−0.2480	−0.0190	0.1954	−0.2195	0.6109	−0.2217	μ
	0.0003	0.0027	0.0005	0.0004	0.0002	0.0001	0.0007	0.0003	σ^2
3	0.0711	−0.2480	0.9102	−0.0143	−0.0661	0.2195	−0.3241	0.2376	μ
	0.0001	0.0005	0.0016	0.0002	0.0003	0.0002	0.0016	0.0008	σ^2
4	−0.0154	−0.0190	−0.0143	0.1188	−0.0353	0.0835	−0.0654	−0.0050	μ
	0.0003	0.0004	0.0002	0.0188	0.0002	0.0004	0.0003	0.0007	σ^2
5	0.0131	0.1954	−0.0661	−0.0353	0.7625	−0.1053	0.2331	−0.0692	μ
	0.0001	0.0002	0.0003	0.0002	0.0055	0.0001	0.0008	0.0002	σ^2
6	0.0986	−0.2195	0.2195	0.0835	−0.1053	0.7397	−0.3076	0.1435	μ
	0.0002	0.0001	0.0002	0.0004	0.0001	0.0098	0.0005	0.0007	σ^2
7	−0.1118	0.6109	−0.3241	−0.0654	0.2331	−0.3076	0.6625	−0.2725	μ
	0.0037	0.0007	0.0016	0.0003	0.0008	0.0005	0.0047	0.0014	σ^2
8	0.1661	−0.2217	0.2376	−0.0050	−0.0692	0.1435	−0.2725	0.3963	μ
	0.0003	0.0003	0.0008	0.0007	0.0002	0.0007	0.0014	0.0167	σ^2

Table 5 Recall of all methods with $K = 50$. Query: 16 images, QF: 50. Total JPEG image in database 800. Maximum recall value = 0.5.

Query	NZ-QD	NZ-Q	Chang's	Pixel corr.
silent30-50	0.5	0.5	0.5	0.5
silent70-50	0.5	0.5	0.5	0.5
sign130-50	0.5	0.5	0.5	0.5
sign160-50	0.5	0.5	0.5	0.5
hall225-50	0.5	0.5	0.5	0.5
hall275-50	0.5	0.5	0.5	0.5
mobile330-50	0.26	0.5	0.4	0.5
mobile370-50	0.49	0.5	0.5	0.5
paris430-50	0.5	0.5	0.5	0.5
paris470-50	0.5	0.5	0.5	0.5
container535-50	0.5	0.5	0.5	0.5
container570-50	0.5	0.5	0.5	0.5
diskus635-50	0.5	0.5	0.5	0.38
diskus670-50	0.49	0.5	0.5	0.45
foreman735-50	0.5	0.5	0.5	0.5
foreman770-50	0.5	0.5	0.5	0.5
Av. Recall	0.484	0.5	0.494	0.489

Table 6 Precision of all methods with $K = 50$. Query: 16 images, QF: 50. Total JPEG image in database 800. Maximum precision value = 1.

Query	NZ-QD	NZ-Q	Chang's	Pixel corr.
silent30-50	1	1	1	1
silent70-50	1	1	1	1
sign130-50	1	1	1	1
sign160-50	1	1	1	1
hall225-50	1	1	1	1
hall275-50	1	1	1	1
mobile330-50	0.52	1	0.8	1
mobile370-50	0.98	1	1	1
paris430-50	1	1	1	1
paris470-50	1	1	1	1
container535-50	1	1	1	1
container570-50	1	1	1	1
diskus635-50	1	1	1	0.76
diskus670-50	0.98	1	1	0.9
foreman735-50	1	1	1	1
foreman770-50	1	1	1	1
Av. Precision	0.968	1	0.988	0.979

between those classes is high. For example, in Table 4, the highest average similarity between classes is $\mu_{C_7,C_2} = 0.6109$. That is between sequence "diskus" and "sign irene."

Based on the data and description provided above, we confirmed that the proposed method is appropriate for similarity measure.

4.2 Similar Image Retrieval by Querying JPEG Image

The second simulation was aimed at retrieving similar images from a JPEG database using JPEG images as query. The metrics used to measure the similarity performance were *recall* and *precision* [12], which is abbreviated as the RP. Let n_c, n_m and n_f be the number of correct, missed and wrongly retrieved images respectively among the first K retrievals. The RP of a query image q were defined in Eq. (9) and Eq. (10), respectively.

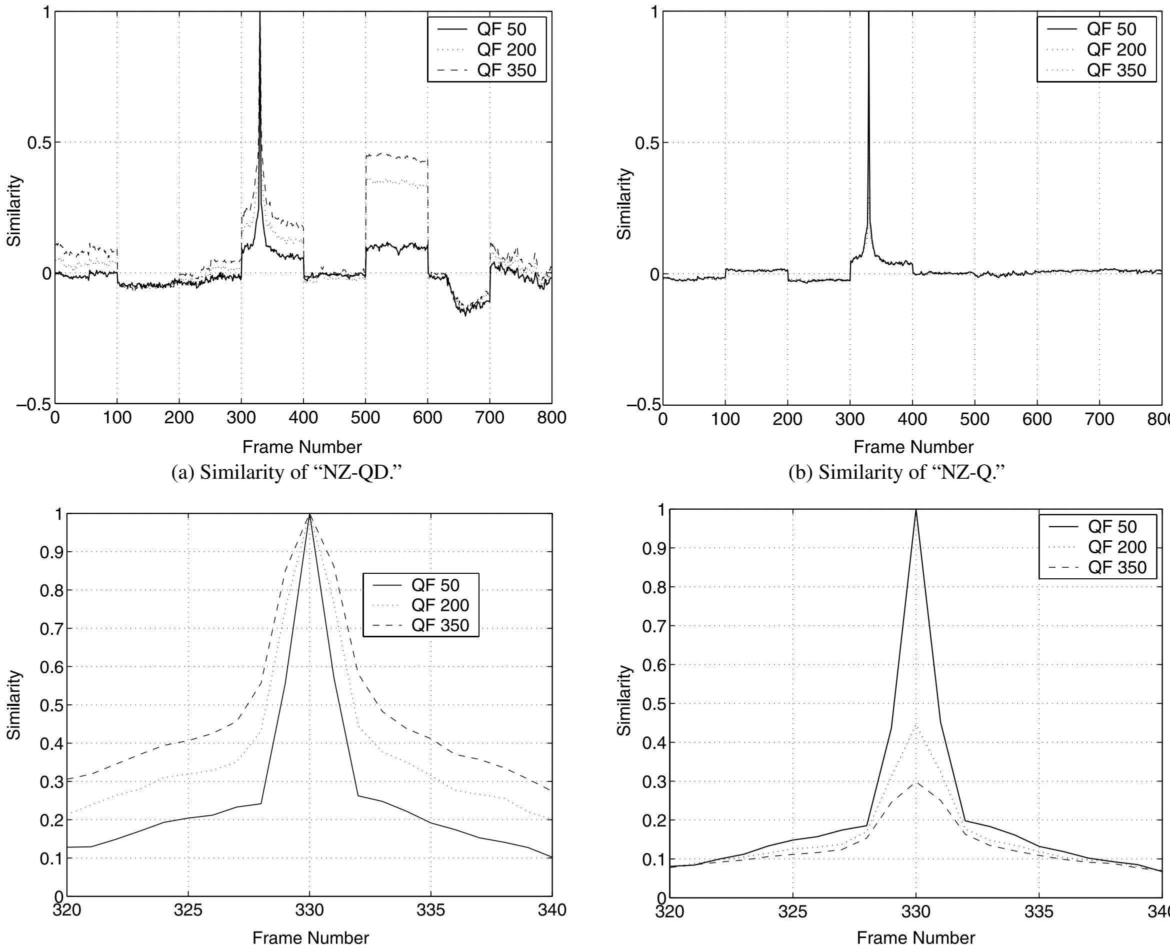
$$Recall_q = \frac{n_c}{n_c + n_m} \tag{9}$$

$$Precision_q = \frac{n_c}{n_c + n_f} \tag{10}$$

In this evaluation, 800 images with QF of 50 in the database were considered. Tables 5 and 6 show the recalls/precisions of 16 queries of all methods with $K = 50$ respectively. The averaged recall/precision over 16 queries are shown in the last row of Tables 5 and 6 respectively.

In general, the RPs of the "NZ-QD," as well as Chang's method or pixel correlation, decrease in sequences with camera panning ("mobile") and zooming ("diskus") operations. While the "NZ-Q" can achieve a perfect RPs, namely 0.5 and 1 respectively.

Figures 8(a) and 8(b) visualize the similarity of the "NZ-QD" and "NZ-Q" when they were queried with "mobile330-50." From the figures, it is confirmed that the "NZ-Q" performed better than the "NZ-QD" in retrieving "mobile330-50." As we can see in Fig. 8(a), the "NZ-QD" also retrieved the images from "container" (class 6), resulted



(a) Similarity of "NZ-QD."
(b) Similarity of "NZ-Q."
(c) Identification of "NZ-QD."
(d) Identification of "NZ-Q."

Fig. 8 Similarity and identification visualizations when the "NZ-QD" and "NZ-Q" are queried with "mobile330-50."

in lower RPs (see Tables 5 and 6).

In regard with the "NZ-QD," it can bee seen in Table 4, that "mobile" sequence has a low intra-class similarity average, that is $\mu_{4,4} = 0.1188$, with a large variance, that is $\sigma^2_{4,4} = 0.0188$. In querying with "mobile," frames of "container" sequence that have the closest inter-class similarity ($\mu_{4,6} = 0.0835$) were retrieved. Furthermore, the "diskus" has intra-class similarity $\mu_{7,7} = 0.6625$, which is considerably high. However, this number is very close to $\mu_{7,2} = 0.6109$. In other words, "diskus" is considered to be very similar to "sign irene." Therefore, when the method was queried by frames from "diskus," images from "sign irene" were retrieved.

It is worth noting that although Chang's and pixel correlation methods performed closely to the proposed methods, they need a much higher computational complexity (longer processing time), as shown in Table 1. As an illustration, querying time of "mobile-970" (database image = 800) of the "NZ-QD" and Chang's method were 15.72 seconds and 1665,93 seconds, respectively, when using a machine with CPU of 3.6 GHz and RAM of 2 GB.

4.3 Duplicate Image Retrieval by Querying JPEG Image

The third simulation was to evaluate whether the method can identify duplicate images, which are generated from the same original images with different JPEG compression ratios, in addition to retrieving similar images. For this purposes, all 2400 images in the database (QFs 50, 200 and 350) were taken into account.

Metrics to measure the identification performance was the occurence number of false positive (FP) and false negative (FN). A good identification method should retrieve all targeted images without leakage, i.e., without FP and FN.

Table 7 shows the identification results of all methods. Table 7(a) is the results of the "NZ-QD" when it is queried with "mobile330-50." It can be seen that all other versions of the query image ("mobile330-200" and "mobile330-350") were perfectly retrieved with similarity value of "1," and were put at the first three highest ranks (including the query image). Figure 8(c) visualizes this identification. Note that simulations with worse QFs, namely 400, 450 and 500 also produced similarity value of "1." The

Table 7 Results of image identification of "mobile330-50" to JPEG database. Query: 16 images, QFs: 50, 200, 350. Total JPEG image in database: 2400.

(a). NZ-QD

Rank	Retrieved Image	Similarity Value	Note.
1	mobile330-50	1	identified
2	mobile330-200	1	identified
3	mobile330-350	1	identified

(b). NZ-Q

Rank	Retrieved Image	Similarity Value	Note.
1	mobile330-50	1	identified
2	*mobile331-50*	*0.4516*	*FP*
3	mobile330-200	*0.4454*	identified
7	*mobile330-350*	*0.2982*	*FN*

(c). Chang's method

Rank	Retrieved Image	Distance Value	Note.
1	mobile330-50	0	identified
2	mobile330-200	*79*	identified
3	*mobile331-50*	*107*	*FP*
6	*mobile330-360*	*133*	*FN*

(d). Pixel correlation

Rank	Retrieved Image	Similarity Value	Note.
1	mobile330-50	1	identified
2	mobile330-200	0.9665	identified
3	mobile330-350	0.9524	identified

same results were achieved for all other query images.

On the other hand, the "NZ-Q," Chang's, and pixel correlation method were not able to identify all compressed versions of the query image. In this case, false positive (FP) and false negative (FN) occurred. For the "NZ-Q," the identification result and its visualization can be seen in Table 7(b) and Fig. 8(d) respectively. In this case, the "NZ-Q" put "mobile330-200" at rank 3 instead of rank 2. Similar leakages also occurred in Chang's method. The pixel correlation could identify all compressed versions of "mobile330-50." However, even with the highest computational complexity, the similarity values of the other compressed versions were not equal to "1." In addition, for other queries, FP and FN also occurred in pixel correlation.

In the "NZ-QD," the similarity of "1" corresponds to the identification of duplicate images that are originated from the same images with different JPEG compression ratios.

4.4 Similar Image Retrieval by Querying JP2 Images

The last simulation was to evaluate the "NZ-QD" and the "NZ-Q" methods when a JPEG image database is queried with images from different format, in this case the JP2 images. The metrics used to measure the performance were also recall and precision.

The query images were compressed from the same original (uncompressed) image as images in JPEG database and then pre-processed according to steps described in Fig. 3. Decomposition and reconstruction of JP2 image were done using 9/7 wavelet filter with 5-level decomposition. There were two kinds of JP2 query images used as queries. One was image with the same size as database's image. Another one was image with smaller size (thumbnail), which obtained by decoding the compressed version with only 4-level reconstruction followed by zero order interpolation.

The recall/precision values of querying with JP2 images are shown in Tables 8 and 9 respectively. The QF of images in the database was 50. From those tables, we can see that the RPs of the proposed methods when it is queried with JP2 images is as good as it is queried with JPEG images. In this case, the "NZ-Q" performed better than the "NZ-QD."

Table 8 Recall of the proposed methods when queried with JP2 images with $K = 50$. Query: 16 images, QF: 50. Total JPEG image in database: 800. "JP2-NS" stands for JP2 normal size image and "JP2-T" stands for JP2 thumbnail image. Maximum recall value = 0.5.

Query	(a). NZ-QD		(b). NZ-Q	
	JP2-NS	JP2-T	JP2-NS	JP2-T
silent-30	0.5	0.5	0.5	0.5
silent-70	0.5	0.5	0.5	0.5
sign-130	0.5	0.5	0.5	0.5
sign-160	0.5	0.5	0.5	0.5
hall-225	0.5	0.5	0.5	0.5
hall-275	0.5	0.5	0.5	0.5
mobile-330	0.26	0.34	0.5	0.5
mobile-370	0.49	0.5	0.5	0.5
paris-430	0.5	0.5	0.5	0.5
paris-470	0.5	0.5	0.5	0.5
container-535	0.5	0.5	0.5	0.5
container-570	0.5	0.5	0.5	0.5
diskus-635	0.5	0.5	0.5	0.5
diskus-670	0.5	0.5	0.5	0.5
foreman-735	0.5	0.5	0.5	0.5
foreman-770	0.5	0.5	0.5	0.5
Av. Recall	0.484	0.490	0.5	0.5

Table 9 Precision of the proposed methods when queried with JP2 with $K = 50$. Query: 16 images, QF: 50. Total JPEG image in database: 800. "JP2-NS" stands for JP2 normal size image and "JP2-T" stands for JP2 thumbnail image. Maximum precision value = 1.

Query	(a). NZ-QD		(b). NZ-Q	
	JP2-NS	JP2-T	JP2-NS	JP2-T
silent-30	1	1	1	1
silent-70	1	1	1	1
sign-130	1	1	1	1
sign-160	1	1	1	1
hall-225	1	1	1	1
hall-275	1	1	1	1
mobile-330	0.52	0.68	1	1
mobile-370	0.98	1	1	1
paris-430	1	1	1	1
paris-470	1	1	1	1
container-535	1	1	1	1
container-570	1	1	1	1
diskus-635	1	1	1	1
diskus-670	1	1	1	1
foreman-735	1	1	1	1
foreman-770	1	1	1	1
Av. Precision	0.969	0.980	1	1

5. Conclusions

A method to retrieve similar images from JPEG (Joint Photographic Image Group) database is proposed. Similarity of the images is determined based on the DCT (Discrete Cosine Transform) coefficients signs, which serve as features. The proposed method is simple because the features can be obtained without additional calculation. Furthermore, it is also robust to JPEG compression, because only unalterable data with respect to JPEG compression, namely non zero coefficients signs are considered. In addition to retrieving similar images, the proposed method is also able to identify duplicate images that are compressed from the same original image with different compression ratios. In terms of computational complexity and robustness to JPEG compression, the performances of the proposed method are more excellent than those of the previous ones.

References

[1] F. Arnia, I. Iizuka, M. Fujiyoshi, and H. Kiya, “Fast image identification methods for JPEG images with different compression ratios,” IEICE Trans. Fundamentals, vol.E89-A, no.6, pp.1585–1593, June 2006.

[2] F. Arnia, I. Iizuka, M. Fujiyoshi, and H. Kiya, “Fast and robust identification methods for JPEG images with various compression ratios,” Proc. 2006 IEEE ICASSP, vol.2, pp.397–400, Toulouse, France, May 2006.

[3] F. Arnia, I. Iizuka, Y. Nakano, A. Takagi, and H. Kiya, “Compressed domain based image authentication for JPEG coded images,” Proc. 2005 ITC-CSCC, vol.4, pp.1465–1466, Jeju, Korea, July 2005.

[4] I. Iizuka, F. Arnia, M. Fujiyoshi, and H. Kiya, “Consideration on DCT sign only correlation for image identification,” Proc. 2006 ITC-CSCC, vol.2, pp.481–484, Chiang Mai, Thailand, July 2006.

[5] C.-C. Chang, J.-C. Chuang, and Y.-S. Hu, “Retrieving digital images from JPEG compressed image database,” J. Image and Vision Computing, vol.22, pp.471–484, 2004.

[6] M. Shneier and M. Abdel-Mottaleb, “Exploiting the JPEG compression scheme for image retrieval,” IEEE Trans. Pattern Anal. Mach. Intell., vol.18, no.8, pp.849–853, Aug. 1996.

[7] J. Jiang, A. Armstrong, and G.C. Feng, “Web-based image indexing and retrieval in JPEG compressed domain,” Multimedia Syst., vol.9, pp.424–432, 2004.

[8] A.R. McIntyre and M.I. Heywood, “Exploring content-based image indexing techniques in compressed domain,” Proc. IEEE Canadian Conf. Elect. & Comp. Eng., vol.2, pp.957–962, May 2002.

[9] M.K. Mandal and C. Liu, “Efficient image indexing techniques in the JPEG2000 domain,” J. Electronic Imaging, vol.13, no.1, pp.182–187, Jan. 2004.

[10] Information technology — Digital compression and coding of continuous-tone still images — Requirement and guidelines, ISO/IEC International Standard 10918-1, 1992.

[11] G.K. Wallace, “The JPEG still picture compression standard,” IEEE Trans. Consum. Electron., vol.38, no.1, pp.xviii–xxxiv, Feb. 1992.

[12] K.J. Qiu, J. Jiang, G. Xiao, and S.Y. Irianto, “DCT-domain image retrieval via block-edge-patterns,” ICIAR 2006, Lecture Notes in Computer Science (LNCS), vol.4141, pp.673–684, 2006.

Fitri Arnia received her B.E. degree in electrical engineering from the University of North Sumatera, Medan, Indonesia, in 1997 and M.Eng.Sc. degree in electrical engineering from the University of New South Wales, Sydney, Australia, in 2004. She is currently a candidate for Ph.D. degree at Tokyo Metropolitan University, Tokyo, Japan, and on study leave from the Electrical Engineering Department, Syiah Kuala University, Banda Aceh, Indonesia. Her research interests are signal and image processing. She is a student Member of IEEE and ACM.

Ikue Iizuka received her B.E. and M.E. degrees in electrical engineering from Tokyo Metropolitan University, Tokyo, Japan, in 2005 and 2007, respectively. In 2007, she joined Hitachi, Ltd., Japan. Her research interest is image processing.

Masaaki Fujiyoshi received his B.Arts, M.Eng., and Ph.D. degrees from Saitama University, Japan, in 1995, 1997, and 2001, respectively. In 2001, he joined Tokyo Metropolitan University, Japan, where he is currently an Assistant Professor of Information and Communication Systems Engineering, Faculty of System Design. His research interests include image processing, secure communications, and spread spectrum communications. Dr. Fujiyoshi served as an Associate Editor for the Special Section on Selected Papers from the 19th Workshop on Circuits and Systems in Karuizawa of the IEICE Trans. Fundamentals. He has also served as an Associate Editor for the J. IEICE from 2005 to 2007. He has been a Member of the Smart Infomedia Systems Committee of the IEICE from 2005. He is also a Member of the IEEE and the ITE (Institute of Image Information and Television Engineers, Japan). He received the Young Engineer Award from the IEICE in 2001.

Hitoshi Kiya received his B.E. and M.E. degrees from Nagaoka University of Technology, Japan, in 1980 and 1982, respectively, and his D.E. degree from Tokyo Metropolitan University, Japan, in 1987. In 1982, he joined Tokyo Metropolitan University, where he is currently a Professor of Information and Communication Systems Engineering, Faculty of System Design. He was a Visiting Fellow at the University of Sydney, Australia, from Oct. 1995 to March 1996. His research interests include digital signal processing, multirate systems, adaptive filtering, image processing, and security for multimedia. Prof. Kiya served as an Associate Editor for the IEICE Trans. Fundamentals (Japanese Edition) and the IEEE Trans. Signal Processing from 1998 to 2002 and from 1998 to 2000, respectively. He is the chair of the Signal Processing Committee of the IEICE, and the former chair of the Media Engineering Committee of the ITE (Institute of Image Information and Television Engineers, Japan). He is also a Member of the ITE and the IIEEJ (Institute of Image Electronics Engineers of Japan), and a Senior Member of the IEEE.