특집논문-02-07-2-10

Ordinal Measure of DCT Coefficients for Image Correspondence and Its Application to Copy Detection

Changick Kim*

Abstract

This paper proposes a novel method to detect unauthorized copies of digital images. This copy detection scheme can be used as either an alternative approach or a complementary approach to watermarking. A test image is reduced to 8×8 sub-image by intensity averaging, and the AC coefficients of its discrete cosine transform (DCT) are used to compute distance from those generated from the query image, of which a user wants to find copies. Copies may be processed to avoid copy detection or enhance image quality. We show ordinal measure of DCT coefficients, which is based on relative ordering of AC magnitude values and using distance metrics between two rank permutations, are robust to various modifications of the original image. The optimal threshold selection scheme using the maximum a posteriori (MAP) criterion is also addressed.

Keywords - image copy detection, copyright protection, ordinal measure.

I. Introduction

The success of the Internet and cost-effective digital storage device has made it possible to replicate, transmit, and distribute digital content in an effortless way. Thus, the protection of Intellectual Property Right (IPR) has become a crucial legal issue. Detecting copies of digital media (images, audio and video) is a basic requirement for IPR protection (or copyright protection). The applications of copy detection include usage tracking and copyright violation enforcement[1]

There are two approaches to protect copyright on digital image: watermarking and content-based copy detection. Watermarking embeds information into the image prior to distribution. Thus, all copies of the marked content contain the watermark, which can be extracted, to prove ownership. In the latter case, it is content-based, which means it does not require additional information but image itself. Generally, image contains enough unique information that can be used for detecting copies, especially illegally distributed copies. For instance, if an owner of an image suspects his/her image is being illegally distributed on the Internet, he/she can raise a query to a copy detection system. This copy detector has an image database, where images are collected from millions of web pages. In practice, instead of having database of images, the system can maintain a database of signatures (feature vectors) and URLs of the images. This is desirable for avoiding copyright issues as well as for saving memory space, thus the stored images need to be deleted after their feature vectors are extracted. To this end, some automated image searching tool, which crawls the web periodically, profiles newly discovered images, and then indexes them based on their features can be used[2].

The content-based copy detection can also be a complementary approach to watermarking. After the copy detector provides a creator or a distributor with a suspect list, the actual owner of the media can use

* Epson Palo Alto Laboratory Epson Research and Development Inc. CA, USA

watermark or other authentication techniques to prove ownership.

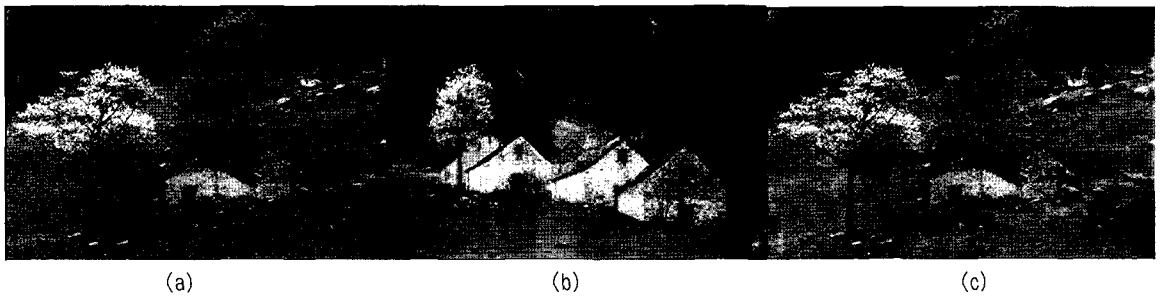
Content-based copy detection schemes extract signature from the original image[1][2][4]. The same signature, extracted from the test image, is compared to the original image signature to determine if the test image is a copy of the original image. Absolutely, the key advantage of the content-based copy detection over watermarking is the fact that the signature extraction is not required to be conducted before the image is distributed. A challenge is that copies may not same as original image. A third party may generate various modifications to avoid copy detection or enhance image quality.

This paper focuses on image copy detection, and the rest of the paper is organized as follows. In Section 2, a brief review of existing image matching techniques and their appropriateness for image copy detection is addressed. Section 3 provides a detailed presentation of the proposed copy detection system. Experimental results and conclusion are included in Section 4 and 5, respectively.

Ⅱ. Background

Color histogram-based methods such as histogram intersection[5] have been popularly used in content-based image retrieval systems. However, they are not suitable for copy detection system since the color histogram does not preserve information about the spatial distribution of colors[1]. A method that can consider the pixel locations is the partition based approach. In this method, image is divided into sub-images. Hsu et al [6] chose a set of colors that described all the image colors, then the image was partitioned into sub-images. The color information of each partition is obtained by a local color histogram. The similarity of two images is measured by comparing their local color histogram, and by considering the similarity of all the sub-images. The problems of this method are high computational cost and search time.

While the image copy detection problem can be viewed as one aspect of image retrieval, there is a fundamental difference between content-based image retrieval and image copy detection. Image copy detector searches for all copies of a query image, whereas a content-based image retrieval system searches for similar images, usually in terms of color[11][12][13]. For instance, Fig. 1 shows three images, where the left one is an original (or query) image and the middle one is an image similar to the left one in terms of color, and the right one hue changed from the original one. In case of color-based image retrieval system, the middle image must be considered more relevant than the right one, while the right image should be selected as a copy of the original image by the image copy detector. In other words, a good image copy detector should detect copies tolerating some extent of modifications. The modifications include changes in brightness and saturation, shift in hue, and spatial distortions including rotating, flipping, and so on. Chang et al. [2] proposed the use of wavelet-based replicated image detection on the web. They tested their



(a) (b) (c)

Fig. 1. Example of difference between color-based retrieval and content-based copy detection. For the query image in (a)a color-based retrieval system would retrieve the image in (b)first, whereas a copy detector should recognize the image in (c)as a copy of the original one.

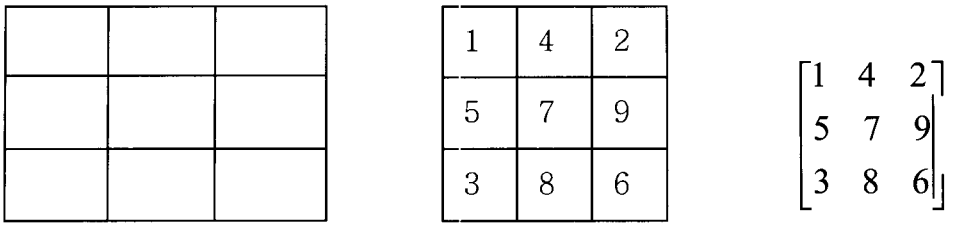


Fig. 2. (a)an image is divided into mn blocks (33 in this example), (b)average values of blocks, and (c) rank matrix of (b)

scheme by using a set of copies, which were modified from their originals by operations like sharpening, softening and despeckling. Their results indicated that out of ten modified images inserted to an image database of 30,000, they were able to correctly detect eight copies. However, it is not clear how the algorithm would do with some big distortions like histogram equalization, or high contrast enhancement, and so on. Also, this wavelet-based method would fail to deal with flipping or rotating. Naphade et al. [7] proposed a variation of color histogram intersection for matching video clips. However, this technique did not address the variations that commonly exist between different copies of the same material.

A fundamental method for matching images is a correlation-based method, which is based on the sum of pixel differences. Let I_1 and I_2 represent intensities in two images. There exist N tuples, $(I_1^1, I_2^1), \ldots, (I_1^n, I_2^n), \ldots, (I_1^N, I_2^N)$, N denoting the number of pixels in an image. The quantity $\frac{\Sigma_{i=1}^{N} \mid I_1^i - I_2^i \mid}{N}$ measures the distance between (I_1, I_2). I_1^ncan also be an average intensity of the n-th block, when N denotes the number of blocks. However, these measures are not robust in that a single outlying pixel (or block) can distort them arbitrarily. Furthermore, it is not suitable in the presence of nonlinear intensity variation at corresponding pixels.

To avoid this substantial problem, Bhat *et al.* proposed to use ordinal measures for stereo image matching[3][10]. As an example of using ordinal measure, an image is partitioned into m×n equal-sized blocks, which makes the system independent of input image sizes, and m×n sub-image is calculated by taking average value of each block (see Fig. 2-(b)). This array is converted to a rank matrix as shown in Fig. 2-(c). Suppose that the intensity values in Fig. 2-(b) are changed in the copied image so its sub-image has values: {{30, 60, 40}, {70, 90, 110}, {50, 100, 80}}. Nevertheless, its rank matrix is identical to that shown in Fig. 2-(c) and thus perfect matching with original image can be achieved.

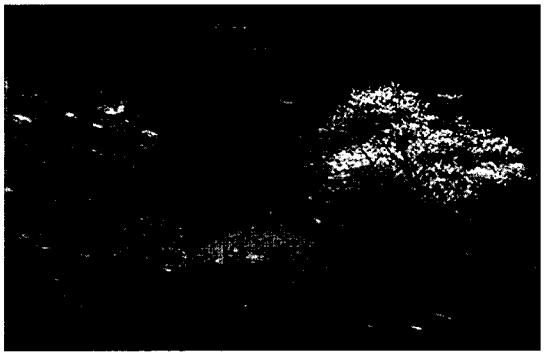


Fig. 3. Horizontally flipped image to avoid copy detecti

However, what if a tamperer changes image contents in more irregular way? Suppose a landscape image in Fig. 1-(a) is horizontally flipped as shown in Fig. 3. This modified image is not expected to be detected even by using ordinal measure of intensity values while the context of the image is well maintained. This can be easily overcome by using magnitudes of DCT coefficients. Fig. 4 shows an example of DCT coefficients of horizontally flipped matrices. The magnitude of each element is same in (c) and (d). Thus, perfect matching using the magnitudes of DCT

coefficients can be conducted in this case.

$$\begin{bmatrix} 1 & 4 & 2 \\ 5 & 7 & 9 \\ 3 & 8 & 6 \end{bmatrix} \quad \begin{bmatrix} 2 & 4 & 1 \\ 9 & 7 & 5 \\ 6 & 8 & 3 \end{bmatrix}$$

(a) (b)

$$\begin{bmatrix} 15.0 & -3.3 & -2.8 \\ -4.1 & 1.0 & 0.6 \\ -4.2 & 1.2 & -2.0 \end{bmatrix} \quad \begin{bmatrix} 15.0 & 3.3 & -2.8 \\ -4.1 & -1.0 & 0.6 \\ -4.2 & -1.2 & -2.0 \end{bmatrix}$$

(c) (d)

Fig. 4. (a) 33 matrix, (b) horizontally flipped matrix, (c) DCT of (a), and (d) DCT of (b)

A similar scheme using DCT to capture the spatial distribution of color has been approved as a MPEG-7 color descriptor, called color layout descriptor (CLD)[8]. In this paper, we take advantages of both DCT magnitudes and ordinal measure. In the following Section, we propose a robust and efficient copy detection system, which uses ordinal measure of DCT coefficients.

Ⅲ. Content-Ontent-Based Image Copy Detection

In this Section, we propose to use ordinal measures of AC coefficients for distance measure, with mentioning the robustness and discriminability issues of the proposed measure. We also discuss the selection scheme of the optimal threshold based on maximum a posteriori (MAP) criterion.

1. Ordinal Measures Of AC Coefficients

We begin our discussion with the problem definition of image copy detection. Let us define an image database as $T=\{C,R\}$, where $C=\{Q,C_1,\ldots,C_m\}$ denotes a set of query (original) image Q and its copies with/without some modifications, and $R=\{R_1,\ldots,R_n\}$ remaining images in a database. The detector (classifier) should have high discriminability between two classes (we call them *'class C'* for set C and *'class R'* for set R, respectively, throughout this paper), whereas having robustness to various image modifications. Let T_i be a test image from a database, then the copy detection is expressed in terms of hypothesis testing terms

$$H_o: T_i \in C; H_1: T_i \in R \tag{1}$$

where H_o is the null hypothesis and H_1 is the alternate hypothesis. The null hypothesis states that the test image is correlated to the query image, and the alternate states otherwise.

Our goal is to detect as many copies of query image as possible, minimizing false detections. We use ordinal measure of DCT coefficients as our features to represent images. In detail, the magnitudes of AC coefficients of 8×8 sub-image are ranked in the descending (or ascending) order. This is simply a permutation of integers and called a rank matrix. More specifically, if S_N denotes the set of all permutations of integers $[1,2,\ldots,N]$, then any rank matrix is an element of this set. To measure correlation between two rank matrices, r_i and r_j, derived from image I_i and I_j, respectively, it is required to define a distance metric $d(r_i,r_j)$. The distance between two images is expressed by L_1 norm of Minkowski metric between their rank matrices, r_i and r_j:

$$d(r_i,r_j) = \|r_i - r_j\| = \Sigma_{l=1}^{N} |r_{il} - r_{jl}|, \ \forall (r_i,r_j) \in S_N, \tag{2}$$

Where N is a size of the rank matrix. This L_1 norm metric is known more robust to outliers than L_2 norm [14]. Furthermore, L_1 norm is computationally more efficient.

In the proposed scheme, images in arbitrary formats are converted to YUV format, and only Y component is used since color does not play an important role in copy detection whereas it becomes a crucial feature in the

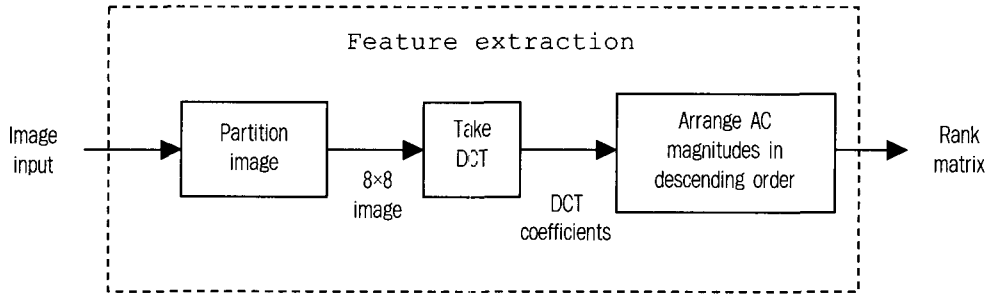


Fig. 5. A block diagram for feature extraction

image retrieval system. The procedure of the proposed matching algorithm is summarized as follows (also see Fig. 5 for feature extraction):

Algorithm : Ordinal measure of AC magnitudes of 8×8 sub-image

1. An input image is divided into 64 equal-sized blocks (or 8×8 subimage) and their average intensities are derived.
2. The derived average intensities are transformed into a series of coefficients by performing 8×8 2-dimensional (2-D) DCT.
3. For ordinal measure of AC coefficients, 1×63 rank matrix is generated, which contains the ranks of sixty three AC magnitudes.
4. Let the rank matrix of the original (query) image Q, be $q=[q_1, q_2, \cdots, q_N]$ and that of a test image T, $t=[t_1, t_2, \cdots, t_N]$, where N=63, at this moment. Then the ordinal measure between two images $D(Q, T)$ becomes L_1 norm between two rank matrices, i.e., $D(Q, T)=d(q, t)=\Sigma_{i=1}^{N}|q_1-t_i|$.

There are two issues concerning performance of the copy detector; its robustness and discriminability. In the following sub-sections, we address these issues.

2. Robustness Issue: Comparison With Different Measures

The robustness of the copy detector determines the amount of data inconsistency that can be tolerated by the system before mismatches begin to occur. To show the robustness of the proposed measure to various modifications, we compare our measure with other three measures, which are all grid based methods, wherein the image is divided into 64 equal-sized blocks (or 8×8 subimage) and their average intensities are derived.

Measure I. This is a fundamental correlation-based method. L_1 norm is calculated between two 8×8 sub-images.

Measure II. This is an ordinal measure of sub-images. L_1 norm is calculated between two rank matrices generated from 8×8 sub-images.

Measure III. L_1 norm between two sets of DCT coefficients. Firstly, AC coefficients of 8×8 sub-image are calculated by performing 2-D DCT. L_1 norm is calculated between magnitudes of the AC coefficients from query image and those from test image.

Measure IV (proposed algorithm). L_1 norm using the ordinal measure of AC magnitudes as introduced in sub-section 3.1.

For the tests introduced above, we inserted 13 images

Table 1. Copy detection results. Only the proposed method (Measure IV) detected all of modified copies.

PANK	Measure I	Measure II	Measure III	Measure IV
1	A	A	A	A
2	D	D	D	D
3	F	F	F	F
4	N	B	N	N
5	K	J	J	J
6	L	N	H	B
7	I	M	G	H
8	FD	E	E	G
9	E	K	I	M
10	FD	L	L	E
11	FD	I	K	I
12	FD	C	FD	L
13	FD	FD	FD	C
14	FD	FD	FD	K

FD: False Detection

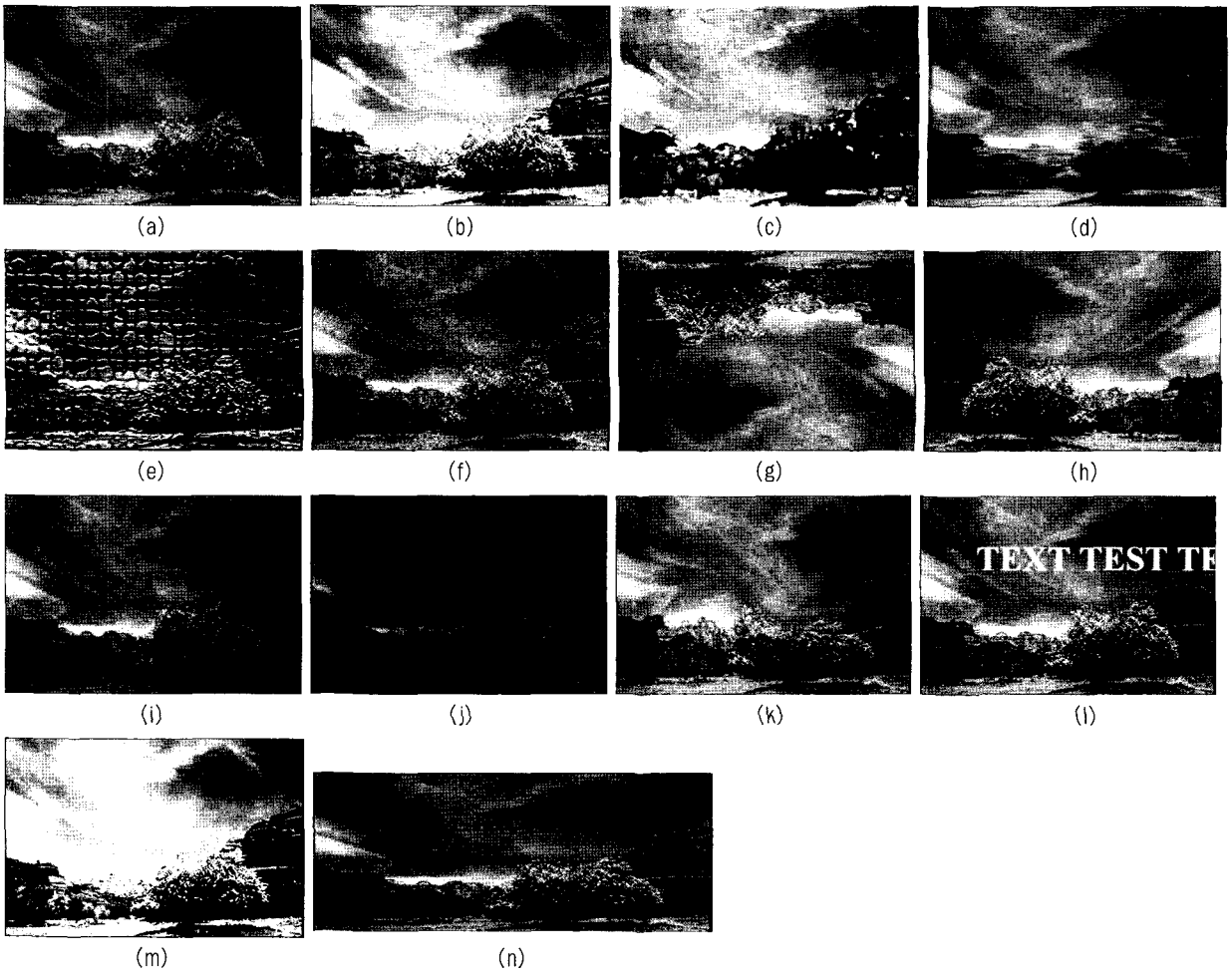


Fig. 6 Original and modified images. (a)original image (374251), (b)histogram equalization (c)water coloring, (d)motion blurring, (e)mosaic tiling, (f)Gaussian noising, (g)180rotating, (h)horizontal flipping, (i)hue change, (j)carkening, (k)twirling, (l)inserting text, (m)contrast enhancement, and (n)resizing (500200).

modified from the same image into the database of 40,000 images. This is similar to the method proposed in [2] but modifications are more irregular than those reported in[2]. Fig. 6 shows an original image and its 13 modifications. Table 1 shows top 14 ranks from each copy detection tests, where the alphabets denote images shown in Fig. 6. As shown in the table, distance measures of ordinal representation (Measure II and IV) were more robust than those of non-ordinal representation (Measure I and III). However, the results from Measure II show that the ordinal measure of pixel values could not tolerate some modifications like rotating and flipping as one would expect, whereas the ordinal measure of AC magnitude values of DCT coefficients (Measure IV) was able to detect all of thirteen copies.

3. Discriminability Issue: Finding The Best Rank Matrix Size N

The experimental results, provided in the previous sub-section, show the proposed measure is robust to various distortions. However, this robustness could turn into a liability when comparing images that are not correlated. This is due to the ordinal measures and using

DCT coefficients. Thus, the second issue, discriminability (or discriminatory power) of the system also becomes important in that it is concerned with its ability to reject irrelevant images such that false detections do not occur.

In the proposed scheme, one factor that affects discriminability is the rank matrix size. The rank matrix size determines the number of AC coefficients that will be used for ordinal measure. It is interesting to investigate the effect of AC coefficients in high frequencies, since some abrupt changes in pixel values might result in changes in high frequencies. Here, we seek to find some measure of the ease of discriminating.

Let $x = d(q, t)$ be a distance between query image and a test image in the database. We define a normalized distance, $\hat{x}$:

$$\hat{x} = x/M,\ 0 \le \hat{x} \le 1, \tag{3}$$

where M is the maximum value of $d(r_i, r_j)$, $\forall\ (r_i, r_j) \in S_N$. M is obtained when the two permutations are reverse of each other, and N is a rank matrix size.

We assume here that we know the state of nature (class label) and the decision of the system. Such information allows us to find the rank matrix size that has the highest discriminability. To this end, we consider the two probabilities:

- $P(\hat{x} > \hat{x}^* \mid \hat{x} \in w_1)$:

$$\text{False Rejection} = \frac{\text{number of copies whose normalized distances are greater than } \hat{x}^*}{\text{number of total copies}}$$

- $P(\hat{x} > \hat{x}^* \mid \hat{x} \in w_2)$:

$$\text{Correct Rejection} = \frac{\text{number of non-copies whose normalized distances are greater than } \hat{x}^*}{\text{number of total non-copies}}$$

where $\hat{x}$ is a normalized threshold value for determining whether or not a test image in a database belongs to the *class C*. If we have a large number of trials, we can determine these probabilities experimentally. We plot a point representing these rates on a two dimensional graph with $\hat{x}^*$ being changed from 0 to 1. This graph is called a receiver operating characteristic or ROC curve[9]. Fig 7 shows ROC curves for four different rank matrix sizes. The rank matrix sizes tested here are as follows:

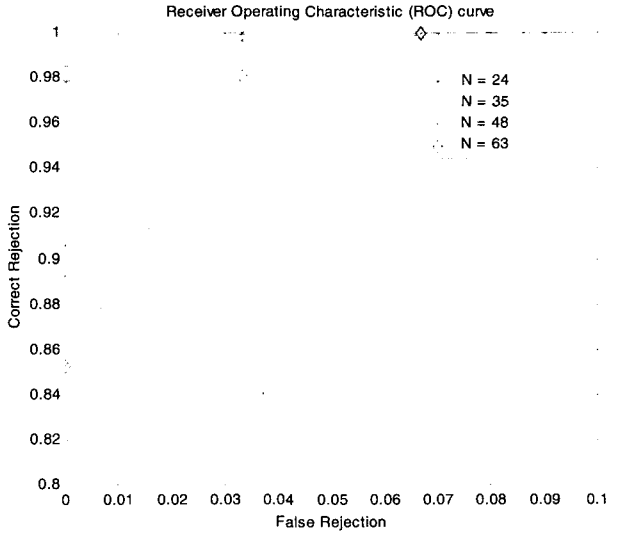


Fig. 7. ROC curve to find N that has the highest discriminability. N is rank matrix size.

- N = 63: All of AC coefficients of 8×8 DCT are used for ordinal measure.
- N = 48: Only 48 low frequency AC coefficients are taken and used for ordinal measure, (or 7×7 upper-left coefficients except DC coefficient)
- N = 35: Only 35 low frequency AC coefficients are taken and used for ordinal measure, (or 6×6 upper-left coefficients except DC coefficient)
- N = 24: Only 24 low frequency AC coefficients are taken and used for ordinal measure, (or 5×5 upper-left coefficients except DC coefficient)

To discriminate the performances of the proposed measure with respect to the rank matrix size, we increased the size of the set C to 30 from 14 by adding more various modifications, such as spherizing, radial blurring, brush strokes, and so on. The ideal curve must pass through $(0, 1) - 0$ false rejection and 1 correct rejection. As shown in Fig. 7, the discriminability is

highest when the size of rank matrix N is 35. This denotes that appropriate removal of AC coefficients in the high frequencies improves ordinal measure of AC coefficients. Reduction of the rank matrix size to 35 from 63 also save memory resource for indexing.

4. Selection Of Optimal Threshold

4.1. Decision By *Maximum A Posteriori* (MAP) Criterion

In the preceding sub-sections, we showed the robustness of the proposed measure, and a selection scheme of the rank matrix size that has the highest discriminability. Lastly, a complete detector (classifier) should employ a threshold value for determining whether a test image from a database belongs to the *class C*. Given an optimal threshold θ_x, we can write an expression for the copy detection system as follows:

$$C = \bigcup_i T_i \text{ such that } x = d(q, t_i) < \theta_x$$

where C is a set of images belonging to the *class C* and T_i is a test image from the image database. The distance $d(q, t_i)$ between the query (original) image and a test image can be modeled as a random variable x, whose value x has mean μ_1 for images belonging to the *class C*, and mean μ_2 for remaining images in a database. We assume the distributions are normal - that is, $p(x \mid w_i) \sim N(\mu_i, \sigma_1^n)$, where w_i denotes the possible class, with $i = 1$ for *class C* and $i = 2$ for *class R*. A reasonable starting point for finding the optimal threshold θ_x is to use the likelihood ratio test as follows:

Decide w_1 if $P(w_1 \mid x) > P(w_2 \mid x)$: otherwise decide w_2.

where $P(w_1 \mid x)$ and $P(w_2 \mid x)$represent the *a posteriori* conditional probabilities that correspond to H_0 and H_1, respectively. If we apply Bayes theorem on both sides of the expression and rearrange terms as shown below

$$\frac{p(x \mid w_1)}{p(x \mid w_2)} \underset{H_1}{\overset{H_0}{\gtrless}} \frac{P(w_2)}{P(w_1)} \tag{4}$$

the left-hand ratio is known as the *likelihood ratio*, and the entire equation is often referred to as the *likelihood ratio test*. Since the test is based on choosing the image class with maximum *a posteriori* probability, the decision criterion is called the *maximum a posteriori* (MAP) *criterion*. It is also called the *minimum error criterion*, since on the average, this criterion yields the minimum number of incorrect decisions. Let K denote the ratio of priors $P(w_2)/P(w_1)$, then (4) can be written as follows:

$$\frac{p(x \mid w_1)}{p(x \mid w_2)} \underset{H_1}{\overset{H_0}{\gtrless}} K \tag{5}$$

We propose to model the class-conditional probability density functions by Gaussian distributions

$$p(x \mid w_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] \tag{6}$$

$$p(x \mid w_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \tag{7}$$

We believe that these distributions approximately model the real data. By substituting (6) and (7) into (5),

$$\exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \underset{H_1}{\overset{H_0}{\gtrless}} K\frac{\sigma_1}{\sigma_2} \tag{8}$$

By taking log on each side,

$$-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 \underset{H_1}{\overset{H_0}{\gtrless}} \ln\left(K\frac{\sigma_1}{\sigma_2}\right) \tag{9}$$

Now, we assume the both classes have same variance. By this assumption, we can statistically constrain the

class C so that too much distorted copies are regarded to have lost the identity of the original image. Since $\sigma = \sigma_1 = \sigma_2$, (9) is reduced to

$$2x(\mu_1 - \mu_2) + \mu_2^2 - \mu_1^2 \underset{H_1}{\overset{H_0}{\gtrless}} 2\sigma^2 \ln(K) \qquad (10)$$

Finally, the problem can be reduced to finding an optimal threshold θ_x for x by rearranging (10)

$$x \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{\mu_1 + \mu_2}{2} - \frac{\sigma^2 \ln(K)}{\mu_2 - \mu_1} = \theta_x \qquad (11)$$

(11) provides us with several intuitions. When $K = 1$, or the prior probabilities for both classes are same, the optimal threshold becomes the average value of two mean values, or $(\mu_1 + \mu_2)/2$. As K increases, the optimal threshold decreases since $\frac{\sigma^2 \ln(K)}{\mu_2 - \mu_1} > 0$. In practice, K would have much bigger value. To calculate the optimal threshold by (11), we used same test materials, which were generated for the experiments in sub-section 3.3 (e.g., 30 copies including original image shown in Fig. 6-(a)), into the database of 40,000 images. Then, we calculated μ_1, μ_2, and σ_2^2 whose normalized values are 0.1464, 0.5842, and 0.0043, respectively. In this case, the optimal threshold by (11) is 0.30 with K=40,000/30. Note that the threshold value is 0.37 when $K = 1$. This indicates that the optimal threshold must be placed lower than the simple average of two means from both classes as K increases. This is easily understood since a test image from a database is not likely to belong to the class of copies as the database size becomes large. Also, note that the threshold value is not dramatically reduced even though the database size gets tremendously big. For example, the calculated optimal threshold value is 0.26 when K= 1,000,000/30.

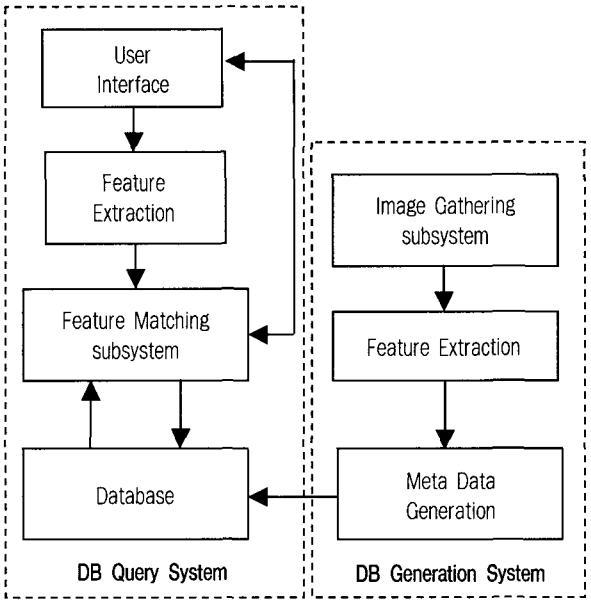


Fig. 8. A block diagram of copy detection syst

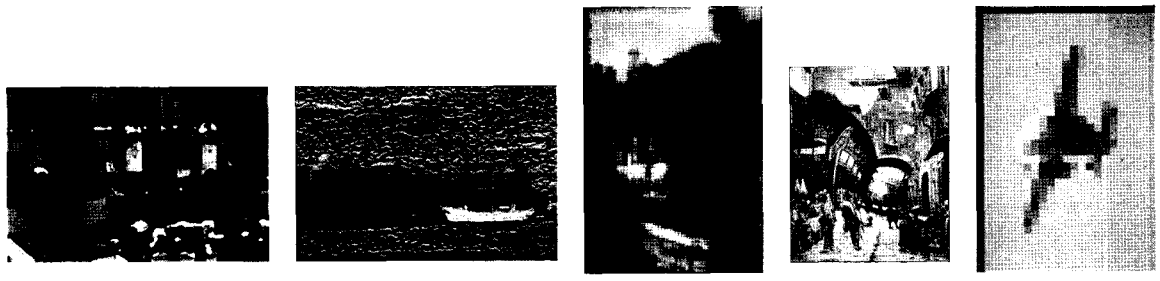
Ⅳ. Experimental Results

We discuss here the performance of the system we developed for copy detection. All images used in the experiments and shown in figures throughout this paper are part of Corel stock photo library and stored in JPEG format with arbitrary sizes. Only Y component images are extracted from color images and used in the tests. Fig. 8 shows the block diagram of the implemented system, which consists of two subsystems: the database generation subsystem and the database query subsystem. The inputs to the database generation subsystem are images collected by the image gathering subsystem, which can be operated manually or by using software like web crawler. The rank matrices are extracted from images and then appropriately indexed to be used by the database query subsystem. The database query subsystem matches a query image to the indexed elements of the database.

For performance evaluation, we inserted five sets of copies into the database, where each set has 10 copies of original images (see Fig. 9 for original images). A subset of the inserted copies is shown in Fig. 10. It is

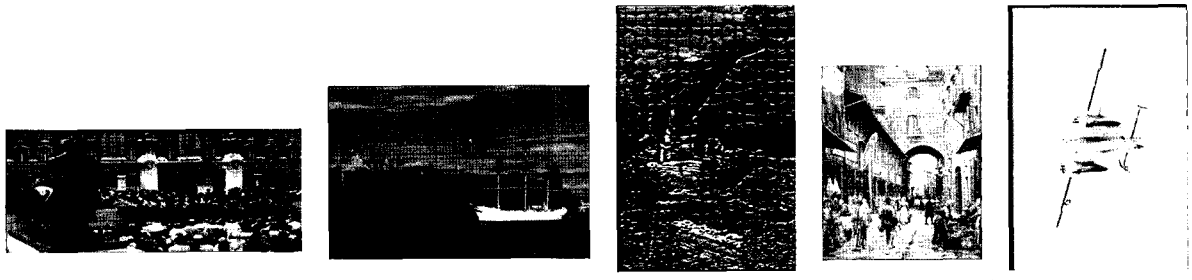
A B C D E

Fig. 9. Original (query) images of five test sets.



(a)Water coloring (b)Craquelure (c)Water paper (d)Twirl (e)Mosaic

(f)Flip horizontally (g)Ripple (h)Histogram equalize (i)Oceanripple (j)Sponge



(k)Aspect ratio changed (l) Saturation increased (m) Mosaic tiles (n)Hue increased (o)Contrast increased

Fig. 10. A subset of copies used in the test. The images in the subsetshow that a variety of modifications are applied to generate copies.

expected color-based schemes would fail to detect some images where color or contrast are altered, while correlation-based schemes would fail in flipped images and be weak in the presence of nonlinear intensity variation at corresponding pixels. Also, texture-based approaches may have difficulty in dealing with copies where edge locations are severely altered or new edge points are added (see (b), (i), and (m) in Fig. 10).

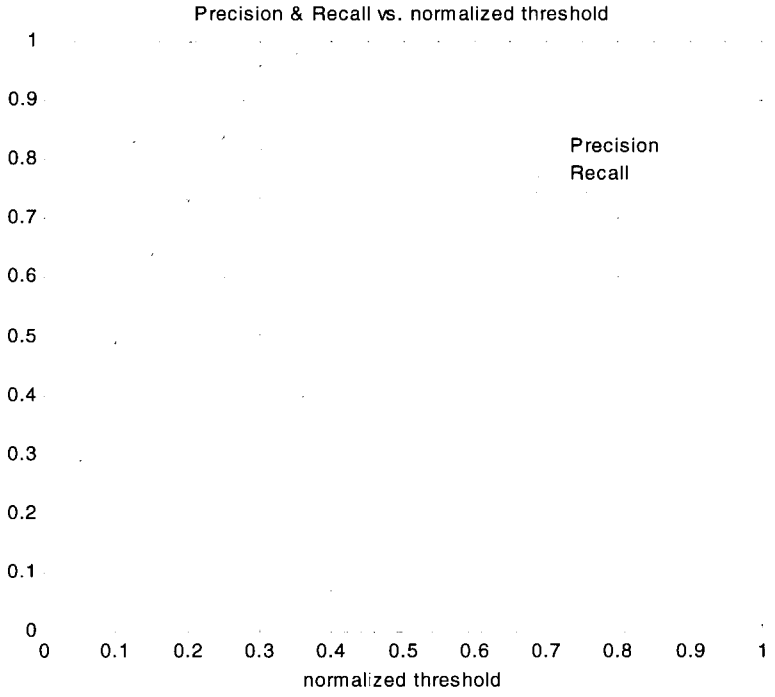


Fig. 11. Precision & Recall vs. normalized thresho

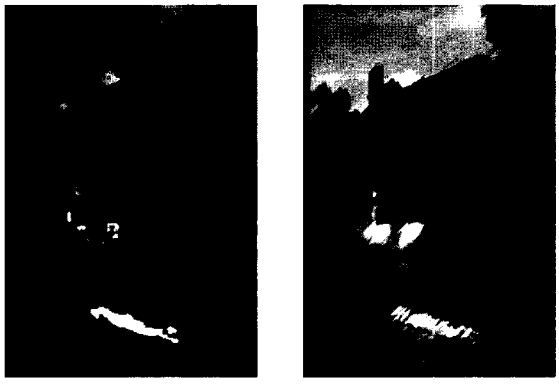
To evaluate the performance of the proposed algorithm, the precision and recall rates are defined as follows:

$$\text{Precision}(\hat{x}^*) = \frac{\text{number of copies detected whose } \hat{x} > \hat{x}^*}{\text{number of detections whowe } \hat{x} > \hat{x}^*}, \quad (13)$$

$$\text{Recall}(\hat{x}^*) = \frac{\text{number of copies detected whose } \hat{x} > \hat{x}^*}{\text{number of total copies}}, \quad (14)$$

where $\hat{x}^*$ is a normalized threshold.

Each image was raised as a query to detect its eleven copies (including itself) from the database. The optimal threshold obtained in Section 3.4 has been used. Precision and recall rates versus normalized threshold are plotted in Fig. 11. We see that the desirable results are obtained around 0.3, which is the optimal threshold calculated in Section 3.4, since both rates are high at that position. In particular, note that having the higher value of recall rate is of importance to minimize the number of missed copies. The detail values at the optimal threshold are summarized in Table 2. As shown in the table, out of 55 inserted copies to the database of 40,000 images, 53 copies were detected at the optimal threshold, whereas only two copies were missed for the query of image C. The missed copies are shown in Fig. 12. For the query of image B, 12 false detections were included, while copies were all ranked on top 11. This



(a) (b)

Fig. 12. Only two missed copies at the optimal 'dark strokes' and (b) by 'water color'threshold. (a) is modified by

indicates it may be required to have an adaptive thresholding scheme depending on the characteristics of the query image.

Table 2. Summarization of the copy detection results at the optimal threshold $\widehat{\mu}_x$.

Query Image	X = #of copies < $\widehat{\mu}_x$	X = #of detections < $\widehat{\mu}_x$	Precision ($\widehat{\mu}_x$) = X /Y	Recall ($\widehat{\mu}_x$) = X /11	#of copies in Top 11 ranks	#of copies missed at $\widehat{\mu}_x$
A	11	12	0.92	1	11	0
B	11	23	0.48	1	11	0
C	9	9	1	0.82	10	2
D	11	11	1	1	11	0
E	11	15	0.73	1	10	0
AVERAGE			0.83	0.96	10.6	0.4

V. Conclusion

In this paper, we have shown a novel scheme to detect unauthorized copies of an image. We proposed to use ordinal measure of 35 DCT coefficients of 8×8 sub-image, and obtained promising test results that are robust to various modifications including flipping and rotating. For indexing the signature, 6bits×35/image, i.e., about 27 bytes/image are necessary. The selection scheme of the optimal threshold based on the maximum *a posteriori* (MAP) criterion was also addressed, showing that the theoretically derived threshold value made an excellent decision for detecting copies of a query image. Of course, the system can be easily modified to return a set of candidates sorted in ascending order in distance as ordinary retrieval systems do.

The detection time was about 2 seconds on Pentium III at 933 MHz, which includes feature vector extraction of the query image and distance matching with about 40,000 feature vectors in the database. To effectively deal with a huge size of database, efficient clustering and indexing scheme should be considered. We are now working on finding the suitable solutions for this issue. Extending the proposed algorithm to video copy detection is our future work.

REFERENCES

[1] A. Hampapur and R. M. Bolle, "Comparison of distance measures for video copy detection," in Proc. IEEE International Conference on Multimedia and Expo (ICME), 2001.

[2] E. Y. Chang, J. Z. Wang, C. Li, and G. Wiederhold, "RIME: A Replicated Image Detector for the World-Wide-Web," in Proc. SPIE Multimedia Storage and Archiving Systems III, Nov. 1998.

[3] D. N. Bhat and S. K. Nayar, "Ordinal Measures for Image Correspondence," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 4, April 1998, pp. 415 - 423.

[4] R. Mohan, "Video Sequence Matching," in Proc. International Conference on Audio, Speech and Signal Processing (ICASSP), 1998.

[5] M. Swain and D. Ballard, "Color Indexing," International Journal of Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.

[6] W. Hsu, T.S. Chua, and H.K.Pung, "An Integrated Color-Spatial Approach to Content-Based Image Retrieval," in Proc. ACM Multimedia, pp. 305-313, 1995.

[7] M. Naphade, M. Yeung, and B. Yeo, "A novel scheme for fast and efficient video sequence matching using compact signatures," in Proc. SPIE conference on Starage and retrieval for Media Databases, SPIE vol. 3972, pp. 564-572, Jan. 2000.

[8] B. S. Manjunath, et al., "Color and Texture Descriptors," IEEE Tr. Circuits and Systems for Video Technology, vol. 11, no. 6, June 2001, pp. 703-715.

[9] R. Duda, P. Hart, and D. Stork, Pattern Classification, Wiley-interscience, 2001.

[10] D. N. Bhat and S. K. Nayar, "Ordinal Measures for Visual Correspondence," Columbia Univ., Computer Science, Tech. Rep. CUCS-009-96, Feb. 1996.

[11] J. Huang, et al., "Image Indexing Using Color Correlograms," in Proc. CVPR, pp.762-768, 1997.

[12] M. Stricker and A. Dimai, "Color Indexing With Weak Spatial Constraints," in Proc. SPIE Storage and Retrieval for Still Image and Video Databases IV, vol. 2670, pp.29-40, 1996.

[13] Y. Deng and B. Manjunath, "An Efficient Low-Dimensional Color Indexing Scheme for Region-Based Image Retrieval," in Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP-99), pp.3017-3020, Phoenix, Arizona, March 1999.

[14] P. J. Rousseeuw and A. M. Leroy, Robust Regression and Outlier Detection, John Wiley and Sons, 1987.

저 자 소 개

김 창 익(Changick Kim)

- 1989년 : 연세대학교 전기공학과 졸업 (B.S.)
- 1991년 : 포항공과대학교 전자전기공학과 졸업 (M.S.)
- 2000년 : University of Washington (워싱톤주립대) 전기공학과 졸업 (Ph.D.)
- 현재 : 엡슨 미국 연구소 (Epson Research and Development Inc.)근무
- 주관심분야 : error resilient wireless video communication, media security/management, video analysis, and multimedia signal processing