Block Mean Value Based Image Perceptual Hashing for Content Identification

Abstract. Image perceptual hashing has been proposed to identify or authenticate image contents in a robust way against distortions caused by compression, noise, common signal processing and geometrical modifications, while still holding a good discriminability to tell the same content from different ones in sense of human perception. We propose and compare in this paper four normalized block mean value based image perceptual hashing algorithms which demonstrate higher performances than other existing algorithms in robustness-and-discriminalibility and simplicity for implementation. Overlapped blocking and complexity-reduced rotation operations are employed to enhance the robustness to geometrical distortions, especially to small to middle degree rotations. Given fixed original image contents, identification ratio is used to evaluate the algorithms' content identification. Given fixed content classification, receiver operating curves is obtained to evaluate the proposed algorithms' content robustness and discriminability based on inter-and intra-content hashing distances.

1 Introduction

In the process of storing, managing and distributing of digitized information, the reliable identification of content is always required. Cryptographic hashing techniques digest input data into short binary strings and have been used for identification of binary data like executables. However, most cryptographic hashes are very sensitive to the changes in every bit of the data, in another word, it meet difficulties when coming to multimedia content which has different representation versions, such as format conversion, compression or various enhancing modifications. Thus, other efficient technologies are needed for multimedia content in application of identification and authentication.

Meanwhile, due to the increase of counterfeiting of commercial products, new technologies have been invented to protect physical goods. In this area, technologies are developed that are related to perceptual hashing functions of digital content: physical features are extracted and digitally processed to result in a digital identifier.

In contrast to cryptographic hash functions which are restricted to applications where content is not processed at all, perceptual hash functions are designed to overcome this drawback: Only those manipulations that change the content apparently could affect the calculated perceptual hash function. Moreover, technologies for the identification of multimedia content are also included in the upcoming and ongoing standardization activities such as MPEG-7 or MPEG-21, or as the so called 'Persistent Association Technologies' (PAT)[1].

Considering about its application scenarios, as stated in [2], perceptual hashing function should satisfy the following requirements:

- Ability of discrimination
- Robustness against attacks or modifications
- Dimension of perceptual hash
- Complexity and performance of the hash calculation
- Complexity and performance of the hash retrieval
- Algorithm Security

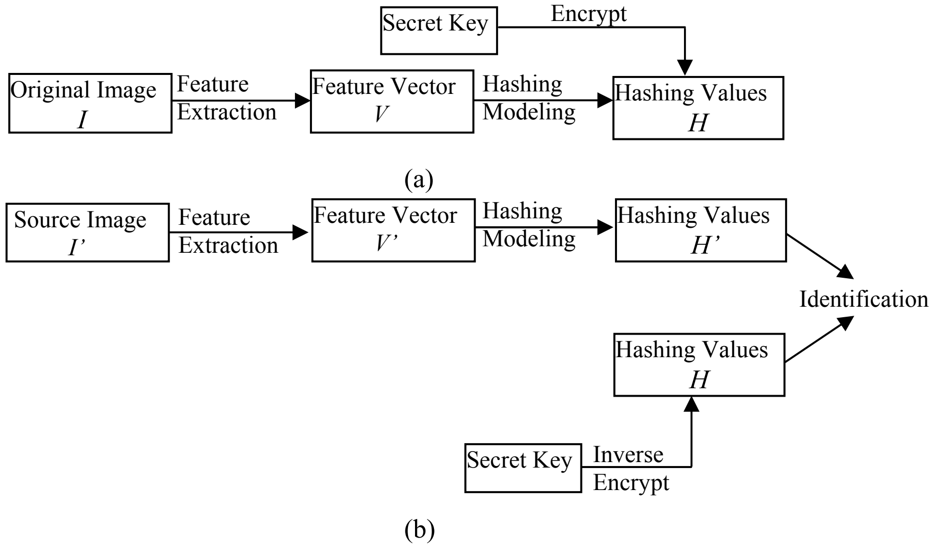

Fig. 1. Diagram of content identification. (a) is the process of obtaining hashing values; (b) is the process of identification.

A general scheme of the perceptual hashing technology is shown in Fig. 1[3]. The most challenging part of this scheme is to extract the feature vector which has to represent the image and yet robust to various distortions. Normally, image features that are invariant to allowed content-preserving image processing operations are identified and used to generate the hash functions. Some of the features that have been proposed in the literature include block-based histograms, image edge information, relative magnitudes of the DCT coefficients. Image edge information [4] was initially proposed to generate the hash function, but the resulting hash could not be tolerant to rotation distortion. In [5], Seo et al., proposed an affine transform resilient image fingerprinting algorithm based on the auto-correlation of the Radon transform, the log mapping and the Fourier transform was presented to deal with affine transform. In [6], Venkatesan et al. presented an image hashing technique, which divided a wavelet transformed image into tiles and extract the statistical features of tiles as hashes. In [7], Fridrich proposed a robust hashing method whose hashing vector were created by projections of DCT coefficients to key-dependent random patterns. Lu et al. proposed a robust mesh-based hashing [11] that aims at resisting more geometrical distortions by firstly, extracting robust mesh and secondly, extracting mesh-based robust hash and finally matching hash for similarity measurement. However, they still allow limited resistance to geometrical distortions. In [8], Yu et al. stated a cumulant-based image fingerprint, which calculate cumulants as an initial feature vector from an image which reduced dimensions. Continuous with its theoretical advantages, cumulants-based algorithm have an improvement in terms of robustness and discriminability.

In this paper, we introduce and compare four normalized block mean based image perceptual hashing functions for content identification. The first two methods are based on normal block mean values of original content while the third and forth methods combined simplified Radon transform, which enhanced the rotation attacks. At the same time, the second and forth methods are overlap version of the first and third methods, respectively. By overlapping block to a certain degree, the overall robustness enhanced effectively. Our proposed methods have high discriminability and low complexity in terms of calculation. We also show that our proposed schemes are robust to geometric distortions (such as rotation, scaling, translation and their combinations), filtering operations (such as median filtering), JPEG compression and various content preserving manipulations. Security was achieved by encrypting the hashing values by secret key.

The paper is organized as the following: In section 2, four normalized block mean based hashing function are detailed. In section 3, test metrics are introduced. In section 4, experiments are performed and results are illustrated. Conclusions are summarized in section 5.

2 Normalized block mean based algorithms

In this section, four normalized block mean based algorithms are introduced in details. The first method is a direct block mean based method which is the simplest case, while the second method is constructed on the overlap blocks. The third method rests on the rotation version of block mean values, and the fourth method is concerning the overlap rotated block mean values.

2.1 Method 1

The first perceptual hashing function is based on the mean values of the blocks, which is specifically stated in the following:

Step 1: Normalize the original image into a preset sizes;

Step 2: Divide the size-normalized image I into unoverlapped blocks $I_1, I_2, \ldots, I_N$, in which N is the block number equal to length of the final hash bit string;

Step 3: Encrypt the indices of the block sequence $\{I_1, I_2, \ldots, I_N\}$ using a secret key K to obtain a block sequence with a new scanning order $\{I'_1, I'_2, \ldots, I'_N\}$;

Step 4: Calculate the mean value sequence $\{M_1, M_2, \ldots, M_N\}$ from corresponding block sequence $\{I'_1, I'_2, \ldots, I'_N\}$ and obtain the median value Md of this sequence as

$$Md = \text{median}(M_i) \qquad (i=1,2,\ldots,\text{N}) \tag{1}$$

Step 5: Normalize the mean value sequence into a binary form and obtain the hash values h as:

$$h(i) = \begin{cases} 0, M_i < Md \\ 1, M_i \geq Md \end{cases} \tag{2}$$

2.2 Method 2

Overlapping strategy for the blocks [10] is generally used to enhance the robustness. Here, the second perceptual hashing function, which is also based on the mean values of the blocks, takes advantage of overlapping. The degree of overlapping is set to be half-sized a block and then we employ the Method 1 to calculate hash values.

2.3 Method 3

To have higher robustness against the rotation attack, the third image perceptual hashing function is realized as a rotation version of the first perception hashing function. Steps are taken as follows:
Step 1: Normalize the original image into a preset sizes;
Step 2: Divide the size-normalized image I into unoverlapped blocks I_1, I_2, …, I_N, in which N is the block number equal to length of the final hash bit string;
Step 3: Encrypt the indices of the block sequence $\{I_1, I_2, \ldots, I_\text{N}\}$ using a secret key K to obtain a block sequence with a new scanning order $\{I'_1, I'_2, \ldots, I'_\text{N}\}$;
Step 4: Calculate the mean value sequence $\{M_1, M_2, \ldots, M_\text{N}\}$ from corresponding block sequence $\{I'_1, I'_2, \ldots, I'_\text{N}\}$ and obtain the median value Md of this sequence as in Method 1;
Step 5: Rotate by D degrees the matrix $\mathbf{M}$ formed by $\{M_1, M_2, \ldots, M_\text{N}\}$, $D = \{0, 15, 30, \ldots, 345\}$. Divide rotated matrix $\mathbf{Mi}$ ($\mathbf{i}$ = 1, 2, …, 24) into N blocks. Obtain the mean value sequence $\{M\text{i}_1, M\text{i}_2, \ldots, M\text{i}_\text{N}\}$of each block and median value Mdi of this sequence, which forms 24 groups of sequences;
Step 6: Perform the same equation (2) for the 24 groups respectively and obtain the final hash value matrix.

2.4 Method 4

The fourth perceptual hashing function is also based on the rotated mean values of the blocks (Method 3), while the difference lies in that the blocks in *step 1* are half overlapped same as in Method 2.

3 Test Metrics

To measure the performance of image hashing, we chose the following three metrics which are effectively used to characterize the robustness, discriminability and authentication ability of the perceptual hashing function: bit error rate (BER), hit rate, and receiver operation characteristics (ROC).

3.1 Bit Error Rate (BER)

In consideration of the binary character of perceptual hash, the number of mismatched bits *i* normalized by the number of bits per perceptual hashes *n* describes the distance between two perceptual hashes. It is called bit error rate (BER) and denoted as ρ :

$$\rho = \frac{i}{n} \tag{1}$$

Where $i \in \{0,1,2,\ldots n\}$ and $0 \le \rho \le 1$. The smaller the BER, the higher the possibility that the corresponding multimedia data contains perceptually identical contents is. Two kinds of distribution of ρ is considered here: the distribution of ρ results from comparison of perceptually identical contents which implies the robustness of perceptual hashing, and the distribution of ρ resulting from matching between different contents which shows the ability of discriminability of perceptual hashing. In the first case, the BER should be as small as possible while in the second one, the BER is expected to be large enough.

3.2 Hit Rate

Originally, hit rate is the chief measurement of a cache, which is the percentage of all accesses that are satisfied by the data in the cache. In our case, it describes the accuracy of detection, which is defined by the right detections over all the detections.

3.3 Receiver Operating Characteristics (ROC)

In most applications, the process of image authentication is similar to a hypothesis testing process with the following two hypothesizes [9]:

H_0: Image is not authentic;

H_1: Image is authentic;

The ROC curve demonstrates the receiver's performance by classifying the received signal into one of the hypothesis states. It is built upon false acceptance rate (FAR), which is the probability that hypothesis H_1 is accepted when the hypothesis H_0 is true and false rejection rate (FRR), which is the probability that the hypothesis H_0 is accepted when the hypothesis H_1 is true. ROC curve is a plot of the probability of detection 1-FAR for the false alarm probability FRR. The closer the curve is to the corner of small FRR value and large 1-FAR value, the higher detection ability the algorithm has.

4 Experimental Results

Our experiment is performed on a database of 2600 images, with the size of 256× 256. In this database, there are 100 original gray scale images, which consist of classic benchmark images, such as Lena, Baboon, Pepper, etc., and a variety of scenery and human activity photos from [12]. These camera photos were converted to gray scales and downsampled to 256×256. For each original image in the set, 25 versions are generated by manipulating the original image according to the proposed distortions as follows:

Compression
1 JPG Q=10
2 JPG Q=40
3 JPG Q=90
Filtering Operation
4 Gaussian 3-by-3 filtering
5 Sharpening 3-by-3 enhancement
6 Median filtering 3-by-3
Geometric Distortions
7 Scaling 0.5
8 Scaling 2.0
9 Cropping 2%
10 Cropping 10%
11 Cropping 20%
12 Random removal of 5%
13 Shearing 2%
14 Shearing 10%
15 Rotation 2 degree + scaling + cropping
16 Rotation 5 degree + scaling + cropping
17 Rotation 10 degree + scaling + cropping
18 Rotation 15 degree + scaling + cropping
19 Rotation 30 degree + scaling + cropping
20 Rotation 90 degree + scaling + cropping
Additive Noise
21 Gaussian noise 0.01
22 Salt&Pepper noise 0.02
23 Speckle noise 0.01
Luminance Distortion
24 Histogram Equalization
25 Brightness Enhancement

4.1 Bit Error Rate

Bit error rate for the first two methods are tabulated in Table 1, which consists of two parts: different content and identical content. Distriminability is demonstrated by the BER of different content, while the robustness is shown by that of the same content with different attacks. From the table, it can be observed that after overlapping operation, robustness is enhanced except for shearing, cropping and rotation. Comparing with the results from [5] and [8], the proposed two methods have much better performance except for larger degree rotations in light of the blocking processing in the proposed methods.

Table 1. Bit Error Rate of the first two method for four example images. The first four rows in the table are the original images and the next 25 rows are 25 modifications of the original images.

	Method1				Method2			
Images and Modifications	Lena	Baboon	Barbara	Airplane	Lena	Baboon	Barbara	Airplane
Lena	0.000	0.500	0.550	0.461	0.000	0.476	0.500	0.386
Baboon	0.496	0.000	0.531	0.457	0.477	0.000	0.476	0.410
Barbara	0.551	0.531	0.000	0.480	0.500	0.476	0.000	0.433
Airplane	0.461	0.457	0.480	0.000	0.387	0.410	0.433	0.000
JPG Q=10	0.004	0.008	0.008	0.004	0.004	0.012	0.008	0.004
JPG Q=40	0.004	0.000	0.004	0.000	0.004	0.000	0.000	0.004
JPG Q=90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gaussian 3-by-3 filtering	0.000	0.000	0.004	0.000	0.004	0.000	0.000	0.000
Sharpening 3-by-3	0.012	0.027	0.023	0.004	0.008	0.015	0.031	0.004
Median filtering 3-by-3	0.000	0.004	0.000	0.000	0.000	0.008	0.004	0.008
Scaling 0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scaling 2.0	0.004	0.000	0.004	0.000	0.012	0.000	0.000	0.004
Cropping 2%	0.012	0.043	0.074	0.027	0.035	0.027	0.035	0.012
Cropping 10%	0.246	0.219	0.234	0.121	0.145	0.125	0.136	0.066
Cropping 20%	0.328	0.340	0.363	0.219	0.273	0.257	0.257	0.125
Rand. removal 5%	0.012	0.020	0.027	0.016	0.035	0.031	0.027	0.031
Shearing 2%	0.016	0.059	0.070	0.020	0.039	0.015	0.039	0.012
Shearing 10%	0.211	0.141	0.218	0.070	0.156	0.113	0.105	0.039
Rotation 2 degree + scaling + cropping	0.070	0.098	0.066	0.051	0.102	0.035	0.058	0.054
Rotation 5 degree + scaling + cropping	0.219	0.168	0.089	0.125	0.160	0.136	0.136	0.093
Rotation 10 degree + scaling + cropping	0.313	0.273	0.226	0.215	0.227	0.230	0.222	0.160
Rotation 15 degree + scaling + cropping	0.375	0.352	0.328	0.246	0.301	0.305	0.304	0.195

Rotation 30 degree + scaling + cropping	**0.426**	**0.484**	**0.468**	**0.395**	**0.383**	**0.387**	**0.382**	**0.308**
Rotation 90 degree + scaling + cropping	**0.480**	**0.500**	**0.585**	**0.492**	**0.426**	**0.430**	**0.468**	**0.414**
Gaussian noise 0.01	**0.000**	**0.004**	**0.004**	**0.004**	**0.012**	**0.000**	**0.008**	**0.008**
S&P noise 0.02	**0.004**	**0.000**	**0.004**	**0.000**	**0.004**	**0.000**	**0.008**	**0.004**
Speckle noise 0.01	**0.000**	**0.004**	**0.004**	**0.004**	**0.000**	**0.012**	**0.004**	**0.000**
Histogram Equalization	**0.008**	**0.008**	**0.012**	**0.188**	**0.070**	**0.008**	**0.016**	**0.137**
Brightness Enhancement	**0.012**	**0.012**	**0.023**	**0.008**	**0.035**	**0.016**	**0.031**	**0.004**

4.2 Hit Rate

Hit Rate is calculated and illustrated in Figure 2. It can be seen that most of the results in Method 2 are higher than Method 1, while most of the values of Method 4 are higher than Method 3. In the rotation modifications, Method 3 and 4 have higher robustness.

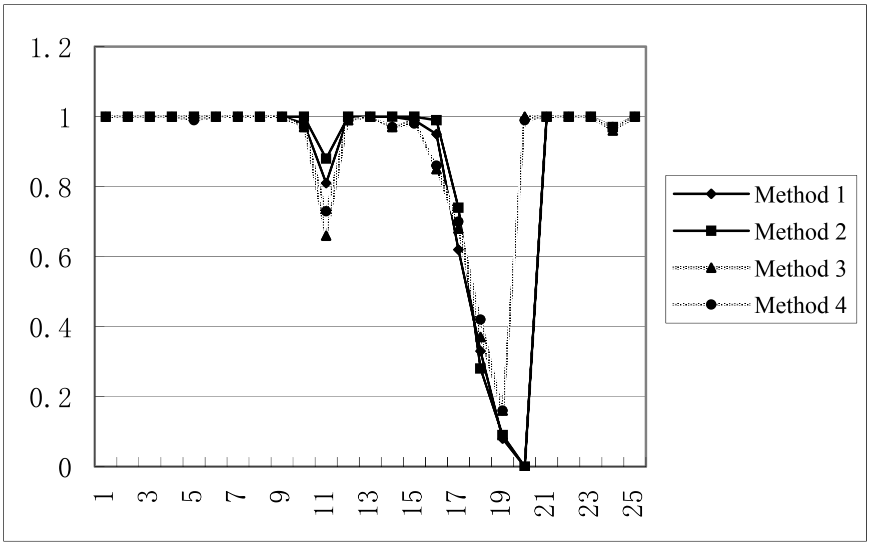

Fig. 2. Hit Rate illustration. *X*-axis is the corresponding modifications and *Y*-axis is the Hit Rate values.

4.3 ROC

ROC curves were plotted for the four methods and some of them are listed below. Each figure demonstrates the discriminability of the proposed four methods against a specified image modification. Taking the remaining 25 versions (including the original one) relevant to the input image as with same content, and other original images as different contents, we can obtain two Gaussian-shape histograms of same contents and different contents. Through threshold adjusting, we can obtain a FAR-FRR relationship and then get the ROC curve for one specified modification. In most modification cases in our experiments, the proposed four methods show desirable discrimination between different contents and the ROC curves are closer to the left upper corner, which bears higher 1-FAR values and lower FRR values. Here we only present several cases whose performances are not so satisfactory. It can be concluded that, the third and fourth methods with rotation operations are better than the former two methods especially for severe modifications such as sharpening, large degree rotation+scaling+cropping, and histogram equalization.

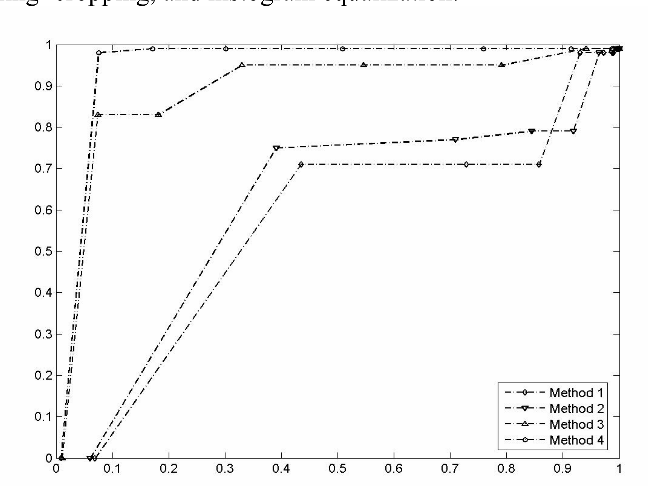

Fig. 3. ROC curve of four methods for the 5th modification: Sharpening 3-by-3 enhancement. *X*-axis is the FRR values, and the *Y*-axis is 1-FAR values.

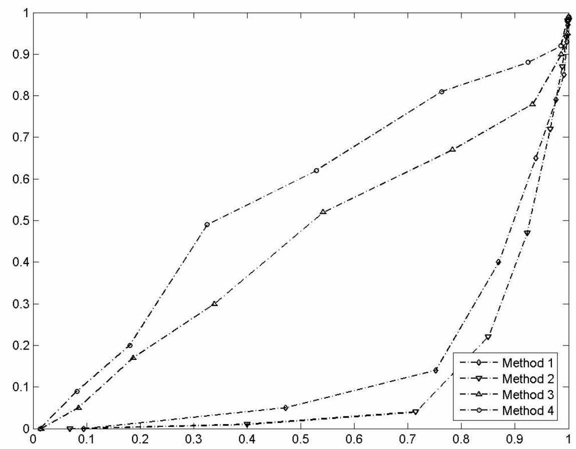

Fig. 4. ROC curve of four methods for the 16th modification: Rotation 15 degree + scaling + cropping. *X*-axis is the FRR values, and the *Y*-axis is 1-FAR values.

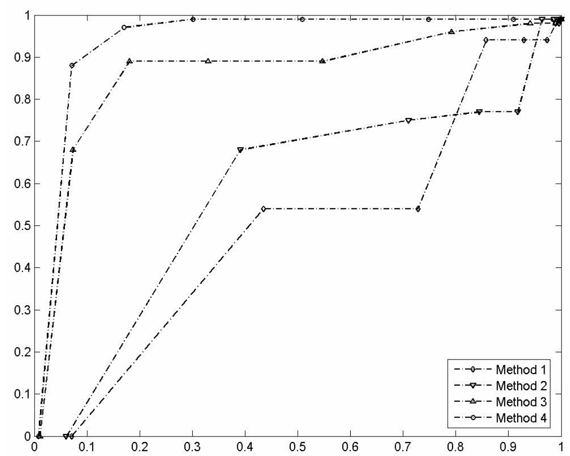

Fig. 5. ROC curve of four methods for the 16th modification: Histogram Equalization. *X*-axis is the FRR values, and the *Y*-axis is 1-FAR values.

5 Conclusions

Four normalized block mean values based hashing functions were proposed and experimental results were compared. The overall performance of the four methods is better than the previous methods in literatures in terms of robustness and discriminability. From the results it can be concluded that there is a desirable trade-off between the robustness and discriminability from our proposed methods for most common lossy compression, signal processing modifications and rotations with small degrees (<10 degree). Another advantage of the proposed methods is the low computational complexity except the latter two methods employing rotation operations. Our future work focuses on improvements of robustness to large degree geometrical distortions while keeping block-wise methods' efficiency.

Acknowledgement

This research work is supported by National Natural Science Foundation of China (Project Number: 60372052).

References

1. E. N. of Excellence. Applications, Application Requirements and Metrics. Technical report, ECRYPT, (2006).
2. E. N. of Excellence. First summary report on forensic tracking. Technical report, ECRYPT, (2005).
3. P. Cano, E. Baltle, T. Kalker, and J. Haitsma. A review of algorithms for audio fingerprinting. IEEE Workshop on Multimedia Signal Processing (2002), 169-73.
4. M. P. Queluz, Twards Robust, Content Based Techniques for Image Authentication. IEEE Signal Processing Society- Second Workshop on Multimedia Signal Processing (1998), 297-302.
5. S. J.S., J. Haitsma, T, Kalker, and C, D, Yoo, Affine transform resilient image fingerprinting, ICASSP'03, Proc. IEEE 3, (2003) 61-64.
6. R. Venkatesan, S. M. Koon, M. H. Jakubowski, and P. Moulin, Robust Image Hashing, Proc, IEEE ICIP 2000, Vancouver, CA (2000).
7. J. Fridrich and M. Golian, Robust Hash function for Digital Watermarking, IEEE Proc. International Conference on Infromation Technology: Coding and Computing 278-183, (2000) 27-29.
8. Longjing Yu, Martin Schmucker, Christoph Busch, and Shenghe Sun, Cumulant-based image fingerprints. Security, Steganography, and Watermarking of Multimedia Contents VII, Proc. Of SPIE-IS&T Electronic Imageing SPIE Vol. 5681, 68-75
9. A.Swaminathan, Y. Mao, M. Wu, Robust and Secure Image Hashing. IEEE Tran. on Information Forensics and Security (2005).
10. M. Mihcak and R. Venkatessan, A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding. Proceedings of 4th International Information Hiding Workshop, Pittsburgh, PA (2001).
11. C-Shien Lu, C-Yong Hsu, S-Wei Sun and P-Chi Chang, Robust Mesh-based Hashing for Copy Detection and Tracing of Images. IEEE (2004)
12. J.Li, Photography Image Database. http://www.start.psu.edu/jiali/index.download.html