Robust perceptual image hashing using SIFT and SVD

Kh. Motilal Singh[1,]*, Arambam Neelima[1], T. Tuithung[1] and Kh. Manglem Singh[2]

[1]National Institute of Technology Nagaland, Dimapur, Nagaland 797 103, India
[2]National Institute of Technology Manipur, Imphal 795 004, India

With the advancement in technology, digital data such as image, video, etc. can be easily manipulated. Image hashing is a method that can be used for authentication and identification of digital images. In this communication, a robust image hashing technique is proposed using scale invariant feature transform (SIFT), singular value decomposition (SVD) and Zernike moment. Zernike moment is used to restore the image against the rotation attack. Potential points are generated from the image by using SIFT. Block processing of equal size is performed on the input image. Key points within the same block are used to generate hashing values. The experimental outcome shows that the proposed technique can withstand different types of attacks. Receiver operating characteristic curve comparison specifies that our method outperforms other existing methods under consideration.

Keywords: Perceptual hashing function, robustness, SIFT, SVD, Zernike moment.

WITH the rapid growth in science and technology, thousands of digital media such as text, image, audio and video have been released on the internet. This enables easy creation of illegal copies of digital media and availability through different online services. A hash function is an algorithm or method that outputs data of any size to a data of invariable size. The values returned by a hash function are termed as hash values or simply hashes or message digest. The conventional cryptographic hash[1] such as secure hash algorithm (SHA), cannot satisfy the requirements for authentication and identification of multimedia content as the cryptographic hash is sensitive to every single bit of input. Perceptual hashing is a technique used for digital media content. It works by generating the message digest from the invariant features of digital media. It does not change much even if the original data has undergone changes due to image processing operations. Many researchers have been working to improve media hashing[2] and provide a number of media hashing techniques. In this communication, we focus on image hashing. An image hash should meet the following basic two properties: (i) Perceptual robustness: Image hash must tolerate conscious user manipulation on the image such as JPEG compression, scaling, filtering, etc. (ii) Uniqueness: Different message digest should correspond to different images. This implies that the hash value should be distinct and lower than the original image.

Different types of perceptual hashing functions have been developed in recent years, including methods based on histogram[3] where the authors calculated the histogram bins and utilized the hierarchical histogram. In refs 4 and 5, invariant features like scale invariant feature transform (SIFT) and colour features are used. The transform domains such as discrete cosine transform (DCT)[6,7], discrete wavelet transform (DWT)[8] and discrete Fourier transform (DFT)[9] were also introduced. Lastly, matrix factorization such as singular value decomposition[10] and non-negative matrix factorization (NMF)[11] were also used for perceptual hashing.

Key points detected by SIFT are robust against image rotation, scaling and flipping[12]. The scale space of an image $f(x, y)$ given by SIFT is defined as

$$L(x, y, \alpha) = G(x, y, \alpha) * f(x, y), \tag{1}$$

where $*$ is the convolution operation, $G(x, y, \alpha) = (1/2\pi\alpha^2)\exp(-(x^2 + y^2)/2\alpha^2)$ is a Gaussian function with a scale of α.

To find stable key-point locations in the scale space, the difference of Gaussian (DoG) is defined as the difference between two consecutive scale images, separated by a constant multiplicative factor of k and is given below

$$\begin{aligned} D(x, y, \alpha) &= (G(x, y, k\alpha) - G(x, y, \alpha)) * f(x, y) \\ &= L(x, y, k\alpha) - L(x, y, \alpha). \end{aligned} \tag{2}$$

SIFT algorithm uses Gaussian difference, a LôG approximation. It can be achieved from the distinction between two successive images of different scale that are acquired after Gaussian blur. This helps to locate the key points in the image. From the following equation, LôG calculation is achieved

$$D(x, y, \alpha) = L(x, y, k\alpha) - L(x, y, \alpha), \tag{3}$$

where the blurred images with the blurred quantities of $k\alpha$ and α are $L(x, y, k\alpha)$ and $L(x, y, \alpha)$ respectively.

LôG's maxima and minima are generated by comparing a pixel with neighbouring scales along with its adjacent 3 × 3 block size. These values are the key points of potential. Some prospective key points are noise sensitive, including low contrast pixels and edge pixels. By using the Hessian matrix[10], these key points are separated.

SVD[10] is a factorizing technique for real or complex rectangular matrix. SVD of a matrix $A \in R^{M\times N}$, where R represents the real number domain, is given by

$$A = USV^T, \tag{4}$$

*For correspondence. (e-mail: khmotilal@gmail.com) where $U \in R^{M\times M}$ and $V \in R^{N\times N}$ are the unitary matrices and $S \in R^{M\times N}$ is a diagonal matrix with singular values in the diagonal denoted by s_i, such that $s_1 \geq s_2 \geq \cdots \geq s_r > s_{k+1} = 0 = 0 \cdots s_n = 0$ with order k, $k \leq N$ and T is the transpose operator.

SVD gives very stable singular values of the image and a small perturbation to the image does not change them significantly.

Zernike moment[5] of the digital image $f(x, y)$ of size $M \times N$ of order n with repetition m over the unit disk is

$$Z_{mn} = \frac{n+1}{\pi}\sum_{x=0}^{M-1}\sum_{y}^{N-1} f(x, y) \times V_{mn}^{*}$$

$$= \frac{n+1}{\pi}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x, y) \times R_{mn}\mathrm{e}^{jm\theta}, \tag{5}$$

where $V_{mn}^{*} = R_{mn}\exp(jm\theta)$ forms a complete and orthogonal set over the unit disk and

$$R_{mn} = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s},$$

is called radical Zernike polynomials, where $(x^2+y^2)^{1/2} \leq 1$ is the length of the vector ρ from the origin to the pixel at the coordinates (x, y), $\theta = \tan^{-1}(x/y)$ represents the angle between the vector ρ and the x-axis in the counterclockwise direction, $j = \sqrt{-1}$, $n - |m| =$ even, and $|m| \leq n$.

If an image is rotated by an angle of α degrees, the relation between the original and the rotated images is

$$Z'_{mn} = Z_{mn}\,\mathrm{e}^{-jm\alpha}, \tag{6}$$

with the magnitude $|Z'_{mn}| = |Z_{\mathrm{mm}}|$ and the phase shift as

$$\alpha = \frac{\arg(Z'_{mn}) - \arg(Z_{mn})}{m}, \quad m \neq 0. \tag{7}$$

It is found that magnitudes of Zernike moments have invariant properties against rotation, while the moments give change in phase shift.

The proposed hashing consists of six stages as shown in Figure 1. The stages are as follows.

Pre-processing: The pre-processing consists of resizing the image using bicubic interpolation, conversion into a grey-scale image for colour image and low-pass filtering. No conversion is required for the grey-scale image.

Zernike moment: Zernike moment of the input image is found using eq. (5). The angle of rotation is calculated from the difference between the dominant feature angles of the original input image and the attacked image using eq. (7).

SIFT feature extraction: Potential key points are extracted from the pre-processed image using SIFT invariant feature transform using eq. (3).

Block processing: The pre-processed image is divided into many blocks of equal size. A block is selected at random by using a key.

Singular value decomposition: A block selected by the key is decomposed into three matrices using eq. (4). The average singular value *arg* of all biggest singular values $B(x)$ from the selected blocks x is found.

Generate hash value: Hash bit is generated by comparing the average singular value and the largest singular value of randomly selected block as shown below

$$H_i(x) = \begin{cases} 1, \text{ if } B(x) \geq \text{avg} \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

The performance of the proposed method is analysed using many test images of different sizes shown in Figure 2. The images are resized to 512 × 512 and we extract the hash values from each image and their attacked versions using eq. (8). Similarity between the original image and attacked image of length L is measured by Hamming distance (HD)

$$\mathrm{HD} = \frac{\sum_{i=1}^{L}|H_1(i) - H_2(i)|}{L}. \tag{9}$$

The standard test images are resized and manipulated using MATLAB. Hashes from each image are calculated and are processed through the proposed method. Different attacks are impulse noise attack, Gaussian noise attack,

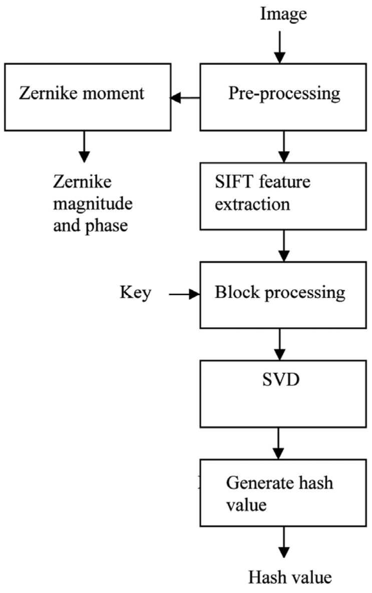

Figure 1. Block diagram of the proposed method.

Table 1. Description of different attacks

Attack	Description	Parameter value
Impulse noise	Noise ratio	0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40
Gaussian noise	Variance	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1
Median filter	Filter size	2, 3, 4, 5, 6, 7
Scaling	Ratio	2, 3, 4, 5, 6, 7,
Rotation	Angle	–45°, –40°, –35°, –30°, –2, –5°, 5°, –10°, 15°, 20°, 25°, 30°, 35°, 40°, 45°
Gamma correction	Gamma	0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5
Sharpening	Alpha	0.1, 0.2, 0.3, 0.4, 0, 0.6, 0.7, 0.8, 0.9, 1
JPEG compression	Quality	20, 30, 40, 50, 60, 70, 80, 90

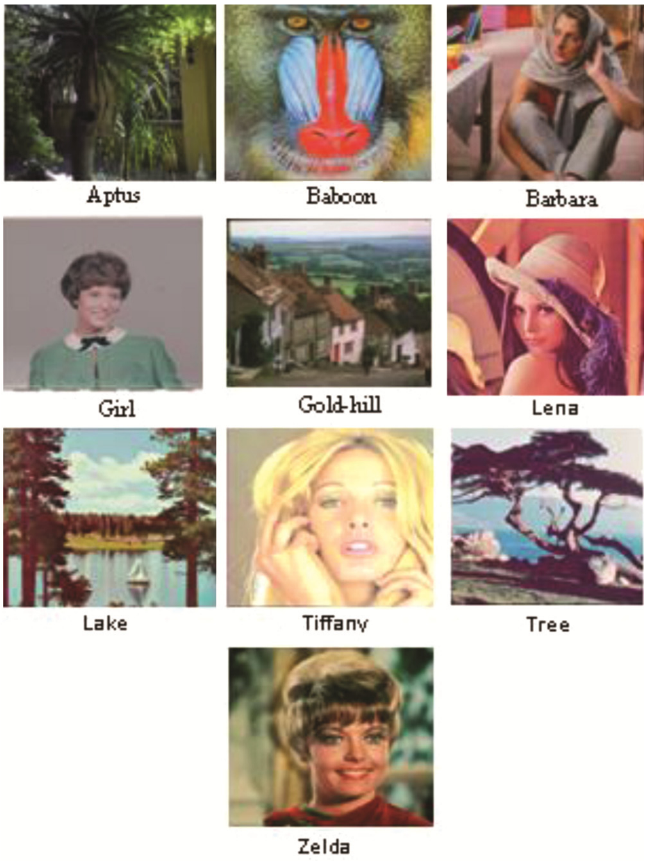

Figure 2. Standard test images.

filtering attack, scaling attack, rotation attack, gamma correction attack, sharpening attack and compression attack. The descriptions of the attacks used are given in Table 1. HDs of the original images and attacked images are shown in Figure 3 for different attacks. Experimentally it was found that for all the standard images, HDs are below the threshold $T = 0.20$. We can conclude that for any attack methods, if HDs are below the threshold $T = 0.20$, then our proposed method is considered robust against the particular attack.

Impulse noise is injected to the original images for impulse noise ratio from 5% to 40%. It degrades some pixels in the original image, but does not corrupt all pixels. The maximum HD is 0.19 for Aptus image shown in Figure 3 *a* when the noise ratio is 40%. It shows the robustness against impulse noise attack.

Gaussian noise is injected to the original images for Gaussian noise for zero mean and variance from 0.01 to 0.10. It degrades all pixels in the original image. The maximum HD is 0.19 for Aptus image, as shown in Figure 3 *b* when the variance is 0.08. It shows the robustness against Gaussian noise attack.

Median filter attack changes the original image by filtering. The filter sizes of the mask range from 2 × 2 to 7 × 7. The maximum HD is 0.0781 for Aptus image in Figure 3 *c* when the filter mask is 6 × 6. It shows the robustness against median filter attack.

Scaling attack resizes the original image by downsizing to the smaller values and then upsizing to the original values. The mask of the filter sizes ranges from 2 × 2 to 7 × 7. The maximum HD is 0.0859 which is obtained for Girl image as shown in Figure 3 *d* when the scale ratio is set to 5 × 5. It shows the robustness against scaling attack.

The original images are rotated in the counter-clockwise and clockwise direction for angles from –45° to 45°. The invariant properties of Zernike moment against rotation make the proposed method robust against rotation attack. The maximum HD is 0.0625 for Aptus image as shown in Figure 3 *e* when the image is rotated to 45°. It shows the robustness against rotation attack.

Gamma correction attack is applied to the original image for different values of gamma. Gamma ranges from 0.6 to 1.5 in our work, where gamma is defined in MATLAB. This attack can darken as well as brighten the image. The maximum HD is 0.1172 for Aptus image as shown in Figure 3 *f* when the gamma value is set to 1.5. It shows the robustness against gamma correction attack.

Sharpening attack enhances the edges, corners and lines textures present in the images. It is applied to the original images for Alpha from 0.1 to 1.0, where Alpha is defined in MATLAB. The maximum HD is 0.1953 for girl image as shown in Figure 3 *g* when the value of Alpha is set to 0.2. It shows the robustness against sharpening attack.

JPEG compression attack is applied to the original image for different compression quality value ranges from 40 to 90 in our work. This attack compresses the quality

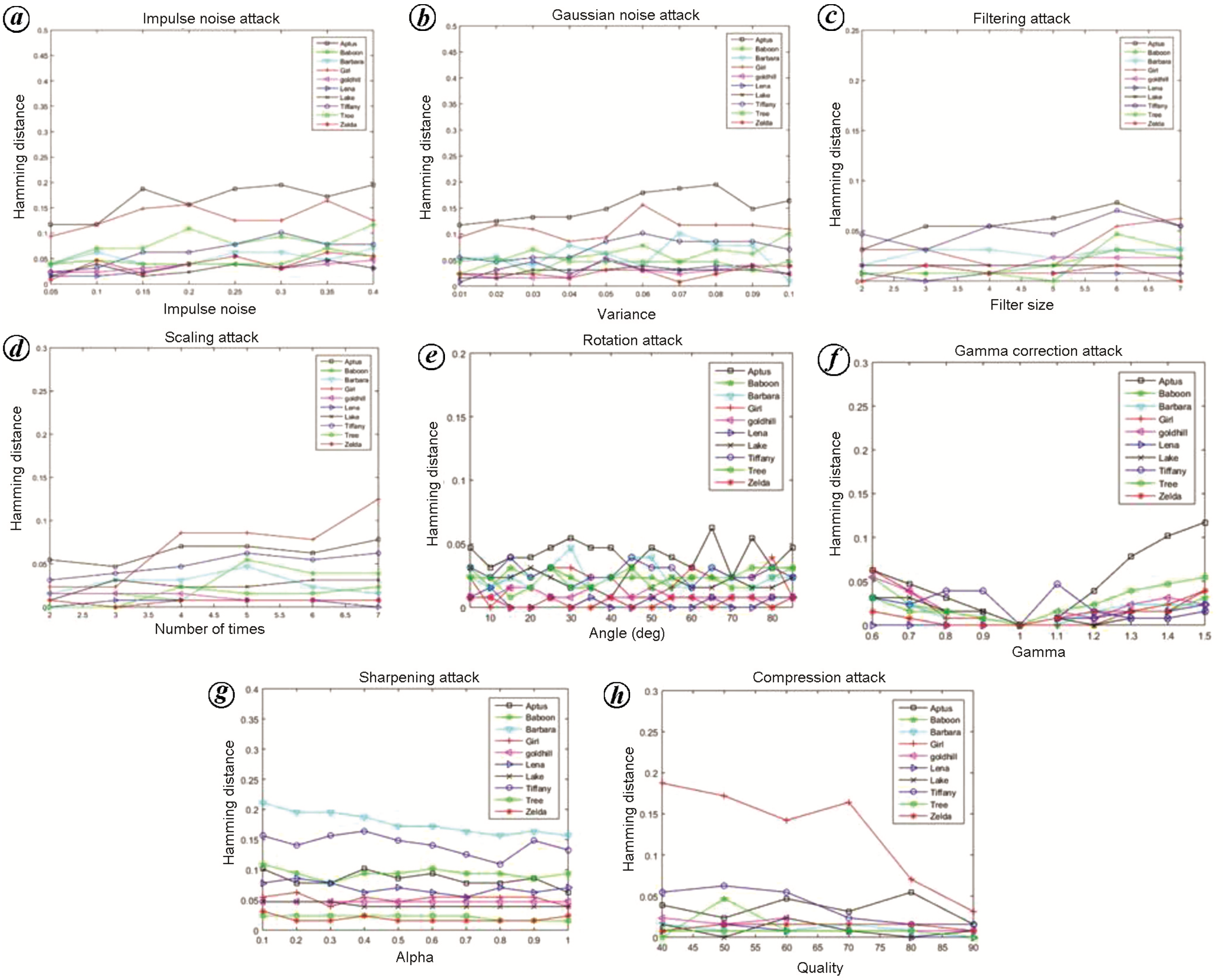

Figure 3. Hamming distances based on 10 standard images and their manipulated images: ***a***, impulse noise; ***b***, Gaussian noise; ***c***, median filtering; ***d***, scaling; ***e***, rotation; ***f***, gamma-correction; ***g***, sharpening; ***h***, JPEG compression.

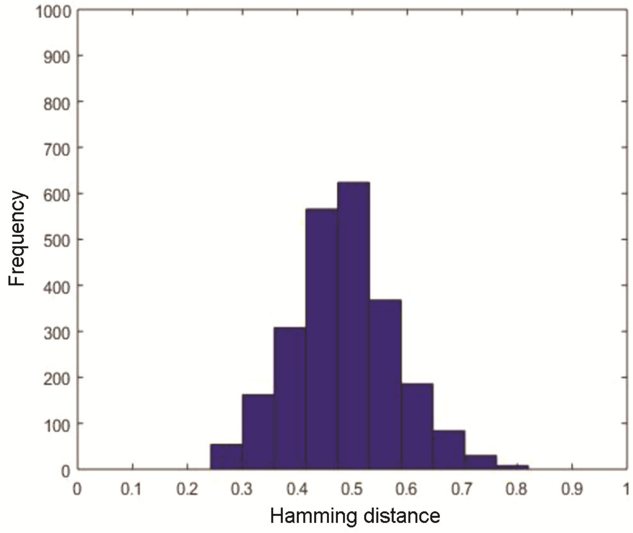

Figure 4. Hamming distance between different image hashes.

of the image. The maximum HD is 0.1875 for girl image as shown in Figure 3 *h* when the quality is set at 40. It shows the robustness against JPEG compression attack.

We collected 50 different images of different size ranging from 256 × 256 to 1024 × 1024. We used eq. (9) to calculate HDs between pairs and then produced 2450 outcomes. The lowest and highest values of the HDs are 0.24 and 0.82 respectively.

The mean and standard deviation of the similarity values are 0.484 and 0.094 respectively. If $S = 0.20$ is the threshold then various images are falsely treated as images of the same kind. If S is increased between 0.2 and 0.25, the similarity is 0.02% of false outcomes. If S is raised further to 0.30, then 0.04% of the various pictures are wrongly categorized as comparable pictures. HDs of various pictures are above 0.20, which means that the suggested technique can determine the difference. The HDs distribution of different images are shown in Figure 4.

We compared the proposed method with three notable existing methods which are based on hierarchical histogram[3], colour features[4] and dominant DCT[6]. For standard comparisons, all the colour images are resized to 512×512 and the same manipulating methods are applied. For the hierarchical histogram, 16 sets of HDs are used. For colour features and dominant DCT, 64×64 block size is taken. It is found that these methods are robust against various geometric attacks.

We use a receiver operating characteristics (ROC) graph[10] to visualize classification performance. The *abscissa* is the false positive rate (FβR) and the *ordinate* is the true positive rate (TβR). The robustness and discriminative capability are defined by the following equations

$$\text{F}\beta\text{R} = \frac{n_2}{N_2}, \tag{10}$$

$$\text{T}\beta\text{R} = \frac{n_2}{N_1}, \tag{11}$$

where n_2 is the number of pair of total different images considered as similar images and N_2 is the total pairs of different images. n_1 is the number of pair of visually identical images considered as similar images and N_1 is the total pairs of visually identical images.

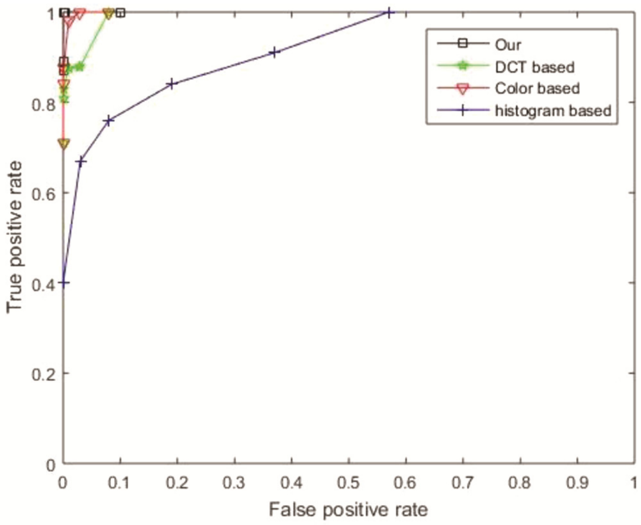

Figure 5. Receiver operating characteristics comparison among different methods.

Table 2. Threshold used

Method	Threshold
Present study	0.15, 0.2, 0.25, 0.3, 0.35, 0.4
Ref. 6	5.5, 6.5, 8.5, 10.5, 12.5, 14.5
Ref. 4	5.5, 7.5, 10.5, 15.5, 20.5, 25.5
Ref. 3	0.2, 0.25, 0.3, 0.35, 0.4, 0.45

For the methods with same FβR, the method with higher TβR is better than the lower TβR. Similarly, for the method with the same TβR, the method with the small FβR is better. We choose six threshold values and calculated their respective TβR and FβR. The ROC chart taken from the above data is illustrated in Figure 5. The ROC curve of the suggested technique is considered above other techniques. It demonstrates that the suggested technique is superior than others. The threshold used for the ROC diagram is shown in Table 2.

Over the past few decades, researchers have developed many image hashing techniques, which can resist different attacks, but their discriminative capabilities are not desirable. A robust image hashing using SIFT, SVD and Zernike moment was proposed. SIFT provides stable keypoints, SVD gives robustness against different attacks and Zernike moment gives very stable robustness against the rotation attack. Zernike moments are used to restore the image to the original position against rotation attack. The proposed method is analysed through various image processing attacks. It is experimentally found that the performance of the proposed method is better than the existing methods under consideration.

1. Schneier, B., *Applied Cryptography, Protocols, Algorithms and Source Code in C*, John Wiley, Chichester, 1996, 2nd edn.
2. Weng, I. and Preneel, B., From image hashing to video hashing, In 16th International Multimedia Modeling Conference, 2010, **5916**, 662–668.
3. Choi, Y. S. and Park, J. Y., Image hash generation method using hierarchical histogram. *Multimedia Tools Appl.*, 2012, **61**(1), 181–194.
4. Tang, Z. Zhang, X. Dai, X. Yang, J. and Wu, T., Robust image hash function using local color features. *AEU-Int. J. Electron. Commun.*, 2013, **67**, 717–722.
5. Quyang, J., Liu, Y. and Shu, H., Robust hashing for image authentication using SIFT feature and quaternion Zernike moments. *Multimed Tools Appl.*, 2017, **72**(2), 2609–2626.
6. Tang, Z., Yang, F., Huang, L. and Zhang, X., Robust image hashing with dominant DCT coefficients. *Optik*, 2014, **125**(18), 5102–5107.
7. Lin, C.-Y. and Chang, S.-F., A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**(2), 151–169.
8. Tang, Z., Zhang, X., Dai, Y. and Lan, W., Perceptual image hashing using local entropies and DWT. *Imaging Sci. J.*, 2013, **61**, 241–251.
9. Qin, Q., Chang, C. C. and Tsou, P. L., Robust image hashing using non-uniform sampling in discrete Fourier domain. *Digital Signal Proc.*, 2013, **23**(2), 578–585.
10. Neelima, A. and Manglem, Kh., Perceptual hash function based on scale-invarient feature transform and singular value decomposition. *Comput. J.*, 2015 (Online); doi:10.1093/comjnl/bxv079
11. Monga, V. and Mihcak, M. K., Robust and secure image hashing via non-negative matrix factorization. *IEEE Trans. Inform. Forens. Secur.*, 2007, **2**(3), 376–390.
12. Lowe, D. G., Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vision*, 2004, **60**, 91–110.

Received 17 December 2016; revised accepted 20 June 2019

doi: 10.18520/cs/v117/i8/1340-1344