Appl. Math. Inf. Sci. 6 No. 2S pp. 643S-650S (2012)

Applied Mathematics & Information Sciences
An International Journal

Perceptual Hashing for Color Images Using Invariant Moments

Zhenjun Tang, Yumin Dai and Xianquan Zhang
Department of Computer Science, Guangxi Normal University, Guilin 541004, P.R. China
Corresponding author: Zhenjun Tang, Email: tangzj230@163.com, zjtang@gxnu.edu.cn

Received July 1, 2011; Revised Sep. 2, 2011; Accepted Sep. 13, 2011
Published online: February 1, 2012

Abstract: Image hashing is a new technology in multimedia security. It maps visually identical images to the same or similar short strings called image hashes, and finds applications in image retrieval, image authentication, digital watermarking, image indexing, and image copy detection. This paper presents a perceptual hashing for color images. The input image in RGB color space is firstly converted into a normalized image by interpolation and filtering. Color space conversions from RGB to YCbCr and HSI are then performed. Next, invariant moments of each component of the above two color spaces are calculated. The image hash is finally obtained by concatenating the invariant moments of these components. Similarity between image hashes is evaluated by L2 norm. Experiments show that the proposed hashing is robust against normal digital processing, such as JPEG compression, watermark embedding, gamma correction, Gaussian low-pass filtering, adjustments of brightness and contrast, image scaling, and image rotation. Receiver operating characteristics (ROC) comparisons between the proposed hashing and singular value decompositions (SVD) based hashing, also called SVD-SVD hashing, presented by Kozat et al. at the 11th International Conference on Image Processing (ICIP'04) are conducted, and the results indicate that the proposed hashing shows better performances in robustness and discriminative capability than the SVD-SVD hashing.

Keywords: Image Hashing, Color Space, Invariant Moments, Singular Value Decomposition (SVD)

1 Introduction

Image hashing is a new technology in multimedia security, and attracts many researchers' attentions in the past decade. It maps visually identical images to the same or similar short strings called image hashes, and finds applications in image retrieval, image authentication, digital watermarking, image indexing, and image copy detection. Conventional cryptographic hash functions such as MD5 and SHA-1 can extract a short string from the input data, but they are very sensitive to bit-level change and therefore not suitable for image hashing. Since digital images often undergo normal digital processing, e.g. JPEG compression and image enhancement, image hash should represent visual appearance of the image no matter what its digital representation is. Generally, an ideal image hash function should have two basic properties. (1) Perceptual robustness: Visually identical images have the same or very similar hashes. In other words, image hash should be robust against normal digital processing, such as image compression and filtering operations. (2) Discriminative capability: Different images have different image hashes. It means that distance between hashes of two different images should be large enough.

Various image hashing algorithms have been reported in literature. Some useful techniques for image hashing include discrete wavelet transform (DWT) [1, 2, 3], discrete cosine transform (DCT)

[4, 5], Radon transform [6, 7], Fourier-Mellin transform [8], singular value decomposition (SVD) [9], and non-negative matrix factorization (NMF) [10, 11].

In 2000, Venkatesan et al. [1] exploited DWT to extract image hash by using statistics of wavelet coefficients. This method is resilient to compression and small geometric distortion. However, it is not robust enough against some normal processing, e.g. contrast adjustment and gamma correction. Monga and Evans [2] used the end-stopped wavelet transform to detect visually significant points for constructing robust hashes. This scheme is robust against JPEG compression, scaling and small-angle rotation. In [3], Ahmed et al. presented a hash scheme for image authentication by using DWT and SHA-1. It can be applied to tamper detection, but fragile to brightness adjustment and contrast adjustment. Observing that DCT coefficients can indicate the visual content of images, Fridrich and Goljan [4] exploited it to construct perceptual hash for digital watermarking. The hash is also sensitive to image rotation. In another study [5], Li and Chang designed a hashing method based on invariant relation between DCT coefficients at the same position in separate blocks. Their method can distinguish JPEG compression from malicious attacks.

Lefebvre et al. [6] first used the Radon transform to obtain image features resilient to rotation and scaling. It is also robust against basic image processing operations. Wu et al. [7] designed a hashing algorithm combining Radon transform, DWT and Fourier transform. This algorithm is robust again print-scan attack. In [8], Swaminathan et al. used Fourier-Mellin coefficients to generate image hashes. This hash function is robust against several content-preserving modifications such as moderate geometric transforms and filtering. In another work, Kozat et al. [9] viewed images and attacks as a sequence of linear operators, and proposed to calculate hashes using SVDs. The SVD-SVD hashing is robust against geometric attacks, e.g. rotation, at the cost of significantly increasing misclassification. Tang et al. [10] observed the invariant relation existing in the coefficient matrix of NMF and used it to construct robust hashes. In [11], Tang et al. proposed a lexicographical framework for robust image hashing and gave an implementation using NMF and DCT. The algorithms of [10, 11] are both robust against JPEG compression, moderate noise contamination, Gaussian low-pass filtering and watermark embedding, but fragile to rotation.

Most of the above algorithms are sensitive to image rotation. Several methods, such as the SVD-SVD hashing [9], can tolerate big angle rotation, but their classification performances is hurt. Moreover, the above mentioned schemes just consider gray images. For color images, they use luminance components for representation. As some color features such as hue and saturation are discarded, their discriminative capabilities, e.g. color differentiating, are limited. In this work, we propose a perceptual hashing for color images by using invariant moments. The proposed algorithm is resistant to image rotation and reaches a desirable classification between perceptual robustness and discriminative capability. The rest of the paper is organized as follows. Section 2 describes the proposed algorithm, and Section 3 gives the experiments. Conclusions are drawn in Section 4.

2 Proposed Image Hashing

As shown in Figure 1, the proposed image hashing is composed of three steps. We convert the input image into a normalized image by interpolation and filtering. Next, we perform color space conversion on the normalized image. Finally, we extract invariant moments from each component of these color spaces, and use them to form image hash. In the following subsections, we will give brief reviews of color space conversion and invariant moments and then describe detailed steps of the proposed hashing.

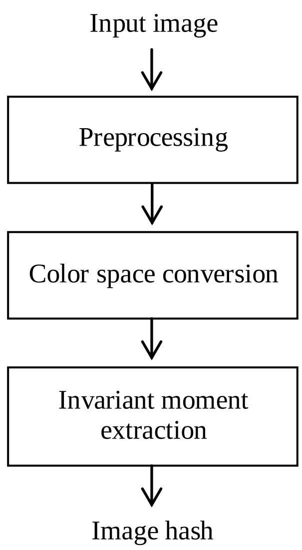


Figure 1: Block diagram of the proposed hashing

2.1 Color space conversion

Generally, a RGB color image can be represented by its hue, saturation and luminance, where the hue represents color appearance, the saturation also called chroma indicates purity or amount of white contained in the color, and the luminance also called intensity is an indicator of brightness. Clearly, taking luminance component for representation cannot indicate all characteristics of a color image. In this work, hue, saturation and luminance components are all exploited for hash generation. To do so, we firstly convert a RGB color image into HSI color space. Let R, G, and B be the red, green and blue component of a pixel, where the ranges of R, G, and B are [0, 1]. Suppose that H, S, and I are the hue, saturation, and intensity of the pixel color. Thus, conversion from RGB color space to HSI color space can be done by the following equations:

$$H = \begin{cases} \theta & , \text{ If } B \le G \\ 360 - \theta & , \text{ otherwise} \end{cases} \tag{1}$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \tag{2}$$

$$I = \frac{1}{3}(R+G+B) \tag{3}$$

where min(R, G, B) is the minimum value among R, G and B, and θ is defined as follows:

$$\theta = \cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\} \tag{4}$$

Since YCbCr color space is used in JPEG images, we also convert RGB color images into YCbCr color space for feature extraction. This is to make our extracted features resistant to JPEG compression. Suppose that Y, C_b, and C_r are the luminance, blue-difference chroma and red-difference chroma, respectively. Thus, we can obtain their values as follows.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{5}$$

2.2 Invariant moments

Invariant moments are firstly introduced by Hu [12]. They are invariant to translation, scaling and rotation, and have been widely used in image classification [13], image matching [14], character recognition, and so on. The aim of choosing invariant moments as image features is to make our hash resilient to image rotation. Let $f(x, y)$ be gray value of a pixel in a digital image sized $m \times n$, where $0 \le x \le m$ and $0 \le y \le n$. Thus, the seven invariant moments are defined as follows:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{6}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{7}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{8}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{9}$$

$$\begin{aligned} \phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \\ & [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03}) \\ & (\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{10}$$

$$\begin{aligned} \phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \end{aligned} \tag{11}$$

$$\begin{aligned} \phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \\ & [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12}) \\ & (\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{12}$$

where η_{pq} (p, q = 0,1, 2, …) are the normalized central moments defined as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{13}$$

in which γ is determined by:

$$\gamma = \frac{p+q}{2} + 1 \qquad p+q = 2,3,\cdots \tag{14}$$

and μ_{pq} are the central moments calculated by:

$$\mu_{pq} = \sum_{x=0}^{m}\sum_{y=0}^{n}(x-\bar{x})^p(y-\bar{y})^q f(x,y) \tag{15}$$

where

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}} \tag{16}$$

and M_{pq} are the (p+q)-th order moments:

$$M_{pq} = \sum_{x=0}^{m}\sum_{y=0}^{n} x^p y^q f(x,y) \tag{17}$$

2.3 Detailed steps

The detailed steps of the proposed image hashing are as follows:

(1) *Preprocessing*. The input image is firstly changed to a standard size by bi-cubic interpolation. This ensures all image hashes to have the same length, and makes the hash scaling-resistant. Next, the standard size image is filtered by a Gaussian low-pass filtering. This is to reduce high frequency components and alleviate influences

of minor image modifications, e.g. noise contamination and filtering, on the hash value.

(2) *Color space conversion*. Having obtained the normalized RGB color image, we convert it into HSI and YCbCr color spaces by using the equations presented in Subsection 2.1. Thus, six color components are available.

(3) *Invariant moment extraction*. For each color component, we calculate its invariant moments by using the equations (6)−(12). The image hash is then obtained by concatenating the invariant moments of these six components. As seven moments are extracted from each component, the hash length is 42 decimal digits.

To evaluate similarity between image hashes, L2 norm is exploited. Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be two image hashes. Thus, the L2 norm is defined as follows:

$$d(\mathbf{h}_1,\mathbf{h}_2)=\sqrt{\sum_{i=1}^{42}[h_1(i)-h_2(i)]^2} \qquad (18)$$

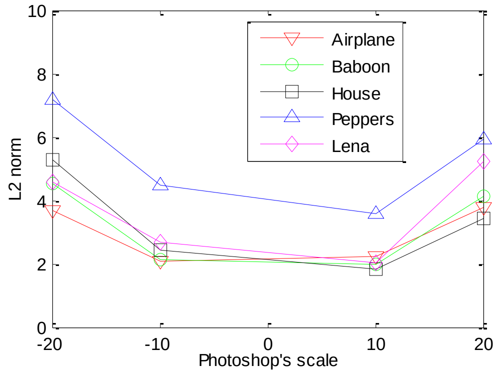
where $h_1(i)$ and $h_2(i)$ are the i-th elements of $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively. The more similar the images of the input hashes, the smaller the d value. If d is smaller than a pre-defined threshold T, the two images are considered as visually identical images. Otherwise, they are different images.

3 Experiments

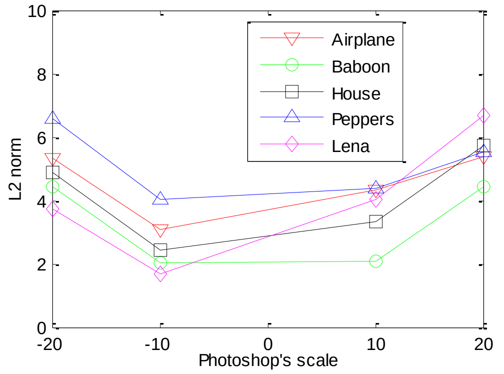
In experiments, all images are resized to 512×512, and then passed through a 3×3 Gaussian low-pass filter with a unit standard deviation. In the following subsections, we conduct experiments about perceptual robustness and discriminative capability to validate performances of the proposed algorithm. To show our advantages, comparisons with the SVD-SVD hashing [9] are also done.

3.1 Perceptual robustness

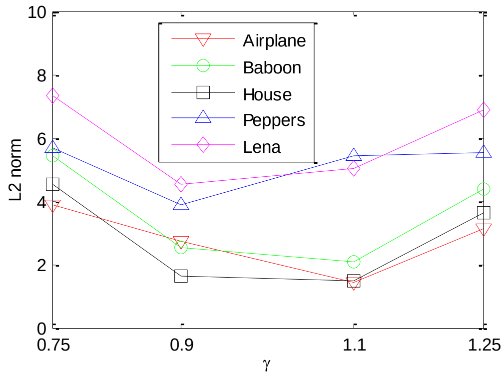
We take five standard color images sized 512×512, i.e. Airplane, Baboon, House, Peppers, and Lena, as test images, and use StirMark 4.0 [15] to perform attacks including JPEG compression, watermark embedding, scaling, and rotation. As rotation will expand the sizes of the processed images, we only take the center parts sized 361×361 from the original and the processed images for hash generation. In addition, we exploit Photoshop and MATLAB to produce attacked images, where the used operations are brightness adjustment, contrast adjustment, and gamma correction, 3×3 Gaussian low-pass filtering, respectively. Thus, each image has 60 attacked images. Detailed parameter values of different operations are listed in Table 1. Extract image hashes of the original image and its attacked images, and then calculate their similarities by the L2 norm. Figure 2 is the results and Table 2 presents the maximum, minimum and mean L2 norms of different operations and their standard deviations. We find that the maximum L2 norms of all operations are smaller than 8.0 except rotation. This means that one can take $T = 8$ as a threshold to resist the above operations except some big angle rotations. In this case, 7.00% attacked images are falsely judged as different images. When T reaches 12.5, the proposed hashing will be robust against all the above operations.



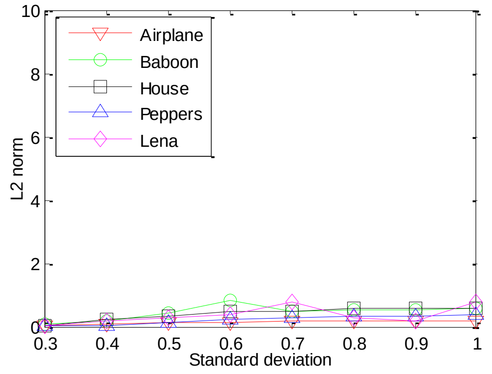
(a) Brightness adjustment



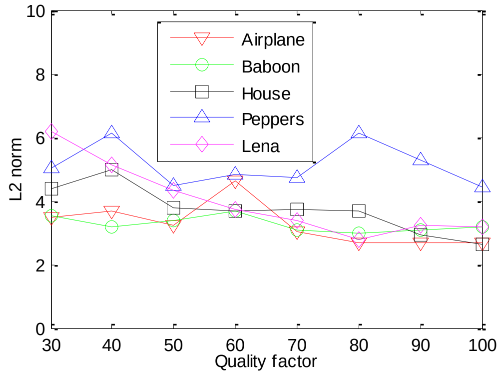
(b) Contrast adjustment



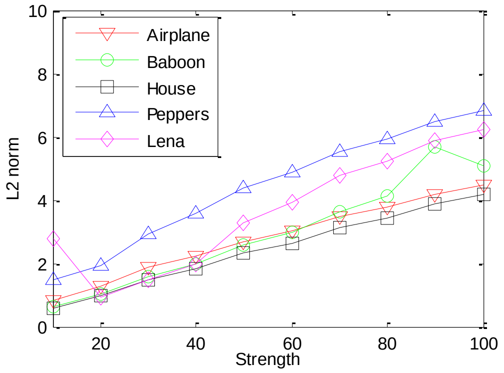
(c) Gamma correction



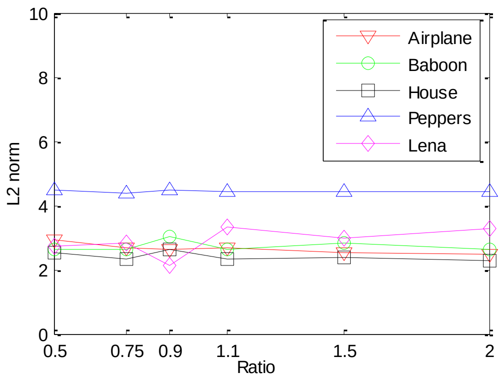
(d) 3×3 Gaussian low-pass filtering



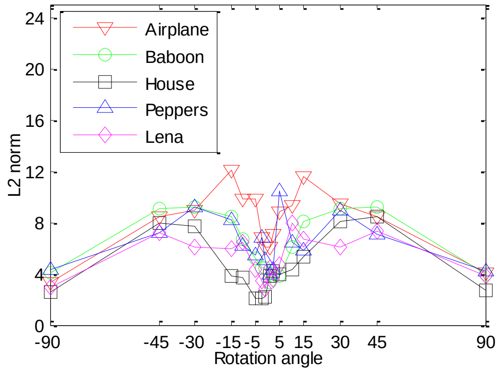
(e) JPEG compression



(f) Watermark embedding



(g) Scaling



(h) Rotation

Figure 2: Robustness performances under different operations

Table 1: The used operations and their parameter values

Operation	Description	Parameter value
Brightness adjustment	Photoshop's scale	10, 20, −10, −20
Contrast adjustment	Photoshop's scale	10, 20, −10, −20
Gamma correction	γ	0.75, 0.9, 1.1, 1.25
3x3 Gaussian low-pass filtering	Standard deviation	0.3, 0.4, ..., 1.0
JPEG compression	Quality factor	30, 40, ..., 100
Watermark embedding	Strength	10, 20, ..., 100
Scaling	Ratio	0.5, 0.75, 0.9, 1.1, 1.5, 2.0
Rotation	Angle in degree	1,2,5,10,15,30 ,45,90,−1, −2, −5, −10, −15, −30, −45, −90

Table 2: Maximum, minimum and mean L2 norms of different operations and their standard deviations

Operation	Max.	Min.	Mean	Standard deviation
Brightness adjustment	7.19	1.82	3.66	1.51
Contrast adjustment	6.65	1.69	4.20	1.45
Gamma correction	7.32	1.44	4.05	1.73
3×3 Gaussian low-pass filtering	0.85	0.01	0.32	0.22
JPEG compression	6.18	2.61	3.87	1.00
Watermark embedding	6.85	0.58	3.23	1.69
Scaling	4.47	2.13	3.02	0.76
Rotation	12.05	2.03	6.10	2.38

3.2 Perceptual robustness

We download 67 images from Internet and take 33 images captured by digital cameras to form a database with 100 different images, where image sizes range from 256×256 to 2048×1536. Calculate the L2 norm between each pair of hashes, and then obtain 4950 results. Figure 3 is the distribution of these 4950 results, where the abscissa is the L2 norm and the ordinate represents the frequency of L2 norm. The maximum, minimum, mean distances and the standard deviation are 51.36, 6.46, 22.46 and 7.37, respectively. We find that 0.10% and 5.82% different images are falsely considered as similar images when T = 8 and T = 12.5, respectively.

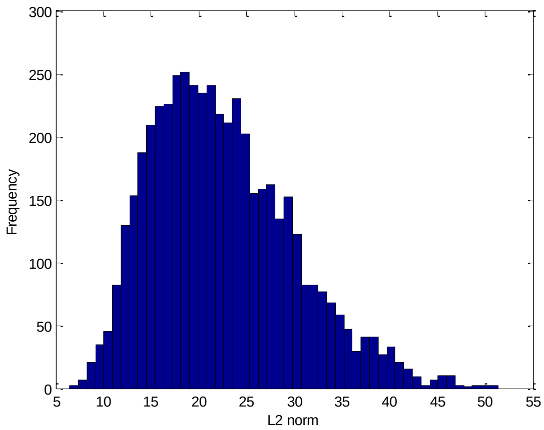


Figure 3: Distribution of L2 norms between hashes of different images

3.3 Performance comparisons

We compare the proposed hashing with the SVD-SVD hashing [9] to show advantages. To make fair comparisons, we exploit the same images to validate perceptual robustness and discriminative capability of the SVD-SVD hashing. As the SVD-SVD hashing only considers gray images, the luminance component of color images in YCbCr color space is extracted for hash generation [9]. The used parameter values of the SVD-SVD hashing are: the first number of overlapping rectangles is 100, rectangle size is 64×64, the second number of overlapping rectangles is 20 and the rectangle size is 40×40. The L2 norm used in [9] is also taken as metric here.

As receiver operating characteristics (ROC) graph [16] is a useful tool for visualizing classification performances, we use it for comparing classification performances between the robustness and the discriminative capability. So we calculate true positive rate (TP rate) and false positive rate (FP rate) of the respective algorithms, which are defined as follows.

$$\text{TP rate} = \frac{n_1}{N_1} \tag{19}$$

$$\text{FP rate} = \frac{n_2}{N_2} \tag{20}$$

where n_1 is the number of the pairs of visually identical images considered as similar images, N_1 is the total pairs of visually identical images, n_2 is the number of the pairs of different images considered as similar images, and N_2 is the total pairs of different images. Actually, TP rate and FP rate indicate the robustness and the discriminative capability, respectively. For two algorithms with the same FP rate, the method with big TP rate outperforms the one with small value. Similarly, if they have the same TP rate, the hashing with small FP rate is better than that with big value. We choose thresholds for the proposed hashing and SVD-SVD hashing respectively, and calculate their TP rates and FP rates. We repeat this process for different thresholds and obtain the ROC graph as shown in Figure 4, where the ordinate is the TP rate and the abscissa is the FP rate. The used thresholds for the proposed hashing are: 1, 4, 8, 12, 15, 18, 22, 27, 30 and 34, and those for the SVD-SVD hashing are: 0.1, 0.2, 0.3, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8 and 0.9. From Figure 4, we observe that the ROC curve of the proposed hashing is above that of the SVD-SVD hashing. This means that the proposed hashing has better performances than the SVD-SVD hashing in robustness and discriminative capability. For examples, when FP rate is near 0, TP rate of the proposed hashing is 0.93 while that of the SVD-SVD hashing is about 0.11. Similarly, when TP rate reaches 1.0, optimal FP rate of the proposed hashing is about 0.058 and that of the SVD-SVD hashing is approximately 0.95.

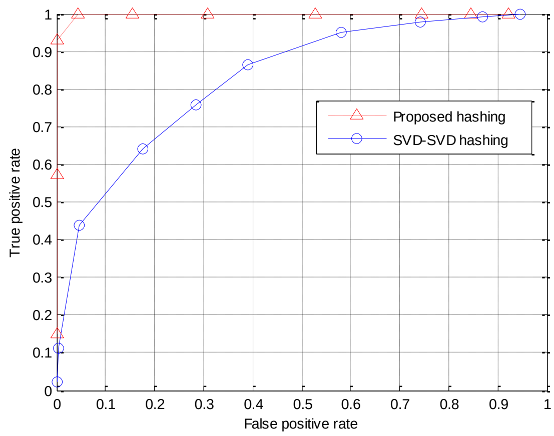


Figure 4: ROC comparisons between the proposed hashing and the SVD-SVD hashing

4 Conclusions

In this work, we have proposed a perceptual hashing for color images. A key component of the proposed algorithm is the use of invariant moments. Since the moments are insensitive to rotation, robustness against image rotation is achieved. As the hue, saturation and luminance of the image are all considered, discriminative capability of the proposed hashing is strengthened. Experimental results show that the proposed hashing is robust against normal digital processing, such as JPEG compression, watermark embedding, gamma correction, Gaussian low-pass filtering, adjustments of brightness and contrast, image scaling, and image rotation. ROC comparisons between the proposed hashing and the SVD-SVD hashing indicate that our algorithm has better performances in robustness and discriminative capability.

Acknowledgements

This work was partially supported by the Natural Science Foundation of China (61165009, 60963008), the Guangxi Natural Science Foundation (2011GXNSFD018026, 0832104), the Project of the Education Administration of Guangxi (200911MS55), the Scientific Research and Technological Development Program of Guangxi (10123005–8), and the Scientific Research Foundation of Guangxi Normal University for Doctor Programs. The authors would like to thank the anonymous referees for their valuable comments and suggestions.

References

[1] R. Venkatesan, S.-M. Koon, M. H. Jakubowski and P. Moulin, Robust image hashing. In: Proceedings of IEEE International Conference on Image Processing (ICIP'00). (2000), 664–666.

[2] V. Monga and B. L. Evans, Perceptual image hashing via feature points: performance evaluation and trade-offs. IEEE Transactions on Image Processing. (2006), 15(11), 3453–3466.

[3] F. Ahmed, M.Y. Siyal and V.U. Abbas, A secure and robust hash-based scheme for image authentication. Signal Processing. (2010), 90(5), 1456–1470.

[4] J. Fridrich and M. Goljan, Robust hash functions for digital watermarking. In: Proceedings of IEEE International Conference on Information Technology: Coding and Computing. (2000), 178–183.

[5] C. Y. Lin and S. F. Chang, A robust image authentication system distinguishing JPEG compression from malicious manipulation. IEEE Transactions on Circuits System and Video Technology. (2001), 11(2), 153−168.

[6] F. Lefebvre, B. Macq and J.-D. Legat, RASH: Radon soft hash algorithm. In: Proceedings of European Signal Processing Conference. (2002), 299–302.

[7] D. Wu, X. Zhou and X. Niu, A novel image hash algorithm resistant to print–scan. Signal Processing. (2009), 89(12), 2415–2424.

[8] A. Swaminathan, Y. Mao and M. Wu, Robust and secure image hashing. IEEE Transactions on Information Forensics and Security. (2006), 1(2), 215–230.

[9] S. S. Kozat, K. Mihcak and R. Venkatesan, Robust perceptual image hashing via matrix invariants. In: Proceedings of IEEE International Conference on Image Processing (ICIP'04). (2004), 3443–3446.

[10] Z. Tang, S. Wang, X. Zhang, W. Wei and S. Su, Robust image hashing for tamper detection using non-negative matrix factorization. Journal of Ubiquitous Convergence and Technology. (2008), 2(1), 18–26.

[11] Z. Tang, S. Wang, X. Zhang, W. Wei and Y. Zhao, Lexicographical framework for image hashing with implementation based on DCT and NMF. Multimedia Tools and Applications. (2011), 52(2-3), 325–345.

[12] M. K. Hu, Visual pattern recognition by moment invariants. IRE Transaction on Information Theory. (1962), IT-8,179–187.

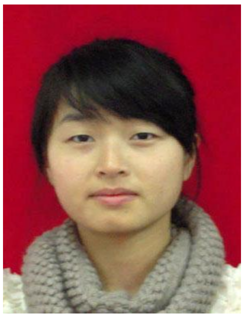
[13] T. C. HSIA, A note on invariant moments in image processing. IEEE Transactions on Systems, Man, and Cybernetics. (1981), SMC-11(12), 831–834.

[14] A. Goshtasby, Template matching in rotated images. IEEE Transactions on Pattern Analysis and Machine Intelligence. (1985), PAMI-7 (3), 338–344.

[15] F. A. P. Petitcolas, Watermarking schemes evaluation. IEEE Signal Processing Magazine. (2000), 17(5), 58–64.

[16] T. Fawcett, An introduction to ROC analysis. Pattern Recognition Letters. (2006), 27(8), 861–874.

Zhenjun Tang received the B.S. degree and the M.Eng. degree from Guangxi Normal University, Guilin, P.R. China in 2003 and 2006, respectively, and the PhD degree from Shanghai University, Shanghai, P.R. China, in 2010. He is now an associate professor with the Department of Computer Science, Guangxi Normal University. His research interests include image processing and multimedia security. He has contributed more than 50 papers. He served as a session chair of 2010 *International Conference on Computer and Computational Intelligence* (ICCCI 2010). He is also a reviewer of *IET Image Processing* and *Multimedia Tools and Applications*.



Yumin Dai received the B.Eng. degree from Xi'an International University, Xi'an, P.R. China in 2010. She is now a graduate student with the Department of Computer Science, Guangxi Normal University, Guilin, P.R. China. Her research interests include image processing and multimedia security.

Xianquan Zhang received the M.Eng. degree from Chongqing University, Chongqing, P.R. China. He is currently a professor with the Department of Computer Science, Guangxi Normal University, Guilin, P.R. China. He has contributed more than 40 papers. His research interests include image processing and computer graphics.