ROBUST IMAGE HASHING BASED ON LOW-RANK AND SPARSE DECOMPOSITION

Yue nan Li, Ping Wang

School of Electronic and Information Engineering, Tianjin University, China

ABSTRACT

We propose in this paper a low-rank and sparse decomposition based image hashing algorithm, aiming to summarize the structural information and sparse salient components of digital image to compact digest. More specifically, we leverage compressive sampling and random projection to separately aggregate the low-rank approximation of input image and the spatial layout of salient components into binary hash. Owing to its capability of capturing and fusing intrinsic visual characteristics, the proposed work demonstrates high robustness and discriminability. As observed in content identification experiments, it shows much higher accuracy than state-of-the-art algorithms. Furthermore, we also analytically evaluate the security of the proposed hashing algorithm using the entropy based metric, and its performance in content identification is analyzed using the channel coding theorem.

Index Terms— Content identification, low-rank and sparse decomposition, performance analysis, robust hashing.

1. INTRODUCTION

Robust image hashing is a one-way mapping from digital image to a succinct digest, and it was proposed as an alternative to cryptographic hashing. As an effective tool for integrity verification, cryptographic hashing is highly sensitive to the variation on message. A single bit modification on source message could incur drastic changes in output hash. However, what the human visual system assimilates from an image are its perceptual characteristics, and the changes on digital representation do not necessarily affects the visual appearance of an image. Accordingly, the hashing function for image needs to tolerant content-preserving distortions, and hash values should be computed on the basis of perceptual contents. Despite of its robustness, robust hashing function also inherits the collision-free property of cryptographic hashing. It is desired that the hash strings of perceptually irrelevant images are statistically independent, and the capability of robust hashing in distinguishing between distinct images is termed as discriminability.

Robust hashing has found extensive applications in copy detection, visual searching, content tracking, authentication, etc. In particular, since it can make a succinct and informative representation of perceptual contents, robust hashing has become a competing solution for reliable and efficient content identification. Recently, some content-sharing sites, like YouTube, have resorted to robust hashing to detect unauthorized uploading of copyrighted contents. In general, most existing robust hashing algorithms have a feature extraction component followed by a quantizer that encodes features to fix-length hash. Some commonly used features include statistics [1–3], frequency-domain coefficients [4–6], matrix invariance [7, 8] and key points [9, 10]. Feature quantizer aims to reduce the redundancies in features while balancing the robustness and discriminability of resulting hash. In Mıhçak *et al.*'s pioneering work [11], features are encoded to hash using a randomized Lloyd-Max quantizer. In [12], we proposed a dithered lattice vector quantizer for image hashing, and our comparative studies corroborate that multi-dimensional quantizer can tolerant a higher degree of distortions than scalar ones. Monga *et al.* formulated feature quantization as an optimization problem and developed a clustering based quantizer [13]. Similarly, the work in [14] leverages spectral embedding to achieve the optimal balance between robustness and discriminability. In addition to algorithm design, some theoretical studies on the performance of robust hashing have been reported in the literature. Varna *et al.* leveraged the game theory to examine the strategic interaction between algorithm designer and an adversary trying to fool the content identification system [16]. In [17], the random projection based hashing is formulated as a communication model, and a side information assisted hashing algorithm is proposed. The work in [18] establishes the connection between content identification and list decoder, and the error bounds and maximum achievable rate of content identification are derived.

In this paper, we develop a robust hashing algorithm for content identification. By decomposing digital image into low-rank and sparse terms, the proposed algorithm learns visual features from both the principle structures and salient components of an image. Moreover, we also investigate its security and content identification performance from the perspective of information theory. Experimental results demon-

This work was supported by the National Natural Science Foundation of China under Grants 61202164 and 61572352, the Fund for Doctor Stations of Ministry of Education under Grant 20120032120026, Tianjin Research Program of Application Foundation and Advanced Technology under Grant No.14JCQNJC01500.

strate the superiorities of our work to competing hashing algorithms.

The rest of this paper is organized as follow. In Section 2, we describe the proposed image hashing algorithm in detail. Analytical studies are presented in Section 3. Section 4 reports the results of comparative experiments, and conclusions are drawn in the final section.

2. LOW-RANK AND SPARSE DECOMPOSITION BASED IMAGE HASHING

Image hash is desired to capture the intrinsic characteristics of digital image. The research on machine learning reveals that an original image and its perceptually similar versions span a low-dimensional subspace [19]. Thus, estimating the subspace helps hashing algorithm learn the intrinsic and stable image features. On the other hand, hash values should also represent the unique visual characteristics that can best distinguish an image from others. The layout of salient regions is one of the most distinctive visual characteristics. The formation of salient regions is the consequence of the visual attention mechanism. Confronted with overwhelming amount of information sensed by vision system, human brain first selectively processes a small fraction of most important visual stimuli. Salient regions are those win the competition for the priorities of information processing [20]; hence, they tend to be very sparse and show remarkable contrast with neighbors.

In light of the aforementioned facts, we propose to compute robust hash from both the low-rank and sparse salient components of an image. We first smooth the input image using a Gaussian low-pass filter and then resize it to $N \times N$. The proposed algorithm decomposes the pre-processed image $\boldsymbol{I} \in \mathbb{R}^{N\times N}$ as the sum of low-rank and sparse terms [21]. In (1), the low-rank matrix $\boldsymbol{L}$ captures the principle structures of $\boldsymbol{I}$, and $\boldsymbol{S}$ is a sparse matrix representing salient components:

$$\begin{aligned} \min_{\boldsymbol{L},\boldsymbol{S}} \quad & \text{Rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{S}\|_0 \\ \text{s.t.} \quad & \boldsymbol{I} = \boldsymbol{L} + \boldsymbol{S}, \end{aligned} \tag{1}$$

where $\|\cdot\|_0$ denotes the l_0 norm, and λ is a const. We solve (1) by the inexact augmented Lagrange multipliers (IALM) based algorithm proposed in [21]. Fig.1 shows the decomposition results of a testing image. Apparently, $\boldsymbol{L}$ depicts the coarse structures of $\boldsymbol{I}$, while $\boldsymbol{S}$ provides a sparse representation of salient regions. To identify the most salient parts, we binarize $\boldsymbol{S}$ by setting the top 10% elements with largest absolute values to one. However, it is worth noting that the binarized $\boldsymbol{S}$ still contains some unstable patterns that may not survive content-preserving manipulations. The unstable patterns appear as isolated noise-like spots, as can be seen from Fig.1(c). Based on this observation, we refine $\boldsymbol{S}$ using the following morphological operations:

$$\boldsymbol{S} = (\boldsymbol{S} \ominus \boldsymbol{E}) \oplus \boldsymbol{E}, \tag{2}$$

(a) (b) (c)

Fig. 1. Low-rank and sparse decomposition of an image. (a) original image, (b) low-rank component, (c) binarized sparse component.

where $\ominus$ and $\oplus$ denote the erosion and dilation operations, representatively, and $\boldsymbol{E}$ is the 2×2 square structuring element.

Hash values are generated by mapping $\boldsymbol{L}$ and $\boldsymbol{S}$ to binary bits. Despite of its low-rank, $\boldsymbol{L}$ is not free of redundancy. To meet the compactness requirement of robust hashing, the proposed algorithm utilizes random compressive sampling (RCS) [22] to encode $\boldsymbol{L}$ to a short measurement vector:

$$\boldsymbol{m} = \boldsymbol{U} \begin{pmatrix} \boldsymbol{D} & & \\ & \ddots & \\ & & \boldsymbol{D} \end{pmatrix} \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{v} \in \mathbb{R}^{N^2\times 1}$ is the vectorization of $\boldsymbol{L} \in \mathbb{R}^{N\times N}$, with its elements randomly permuted, $\boldsymbol{D} \in \mathbb{R}^{W\times W}$ is the W-point DCT matrix, and $\boldsymbol{U} \in \{0,1\}^{K\times N^2}$ is a binary matrix that randomly samples K elements from a vector of length N^2, ($K \ll N^2$). The permutation and sampling processes are controlled by a secret key. As will be discussed later, the RCS process not only reduces redundancies, but also endows robust hashing with the security against forgery attack. After RCS, the first K hash bits $\{b_1, \cdots, b_K\}$ are generated by comparing the elements of $\boldsymbol{m} \in \mathbb{R}^K$ with zero:

$$b_i = \begin{cases} 1 & \text{if} \quad m_i > 0, \\ 0 & \text{otherwise.} \end{cases} \quad i = 1, \cdots, K. \tag{4}$$

To describe the layout of salient regions, we compute the distances from each non-zero element in $\boldsymbol{S}$ to image center, based on which a 16-bin distance histogram $\boldsymbol{h} \in \mathbb{R}^{16\times 1}$ is constructed. The histogram is then projected onto a matrix $\boldsymbol{P} \in \mathbb{R}^{Q\times 16}$ ($Q < 16$) whose elements are randomly drawn from the normal distribution $\mathcal{N}(0,1)$. As in (4), the vector $\boldsymbol{Ph} \in \mathbb{R}^{Q\times 1}$ is binarized to Q bits via thresholding. Finally, the $(K+Q)$ binary bits generated by $\boldsymbol{L}$ and $\boldsymbol{S}$ are concatenated to form the hash string of input image. To summary, Fig.2 presents the flowchart of the hash computation process.

3. INFORMATION-THEORY BASED PERFORMANCE ANALYSIS

3.1. Randomness analysis

The security of hashing algorithm against forgery attack is determined by its amount of randomness. In this subsec-

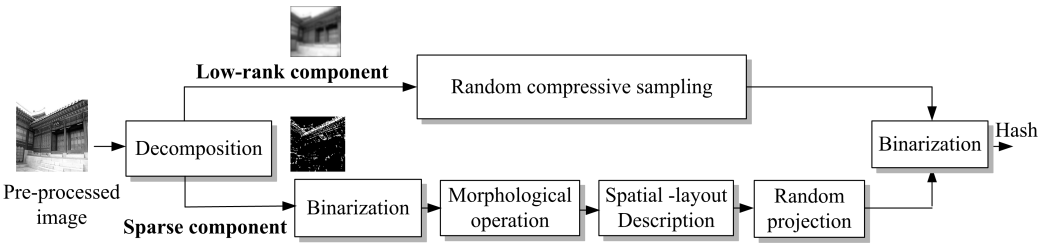

Fig. 2. Flowchart of the proposed algorithm.

tion, we analyze the security of our algorithm by quantifying its randomness using the entropy based metric proposed in [4]. Randomness of the proposed algorithm is introduced by the following processes: permutating the elements of $\boldsymbol{L} \in \mathbb{R}^{N\times N}$, sampling K DCT coefficients from N^2 coefficients, and projecting the distance histogram onto Q random directions. Given $\boldsymbol{L}$, there are $N^2!$ permutations and $C_{N^2}^K$ combinations. Note that permutation and sub-sampling are independently conducted, so the probability that a specific measurement vector $\boldsymbol{m}$ being observed is $(N^2!C_{N^2}^K)^{-1}$, and the entropy of the RCS process is:

$$H(\boldsymbol{m}|\boldsymbol{L}) = \log_2(N^2!) + \log_2(C_{N^2}^K). \tag{5}$$

In this work, we chose $N = 128$ and $K = 80$. Even though an adversary can assess to $\boldsymbol{L}$, the probability of correctly estimating $\boldsymbol{m}$ without the knowledge of the exact key is less than 10^{-200}. We now analyze the entropy of the random projection of distance histogram. Let us consider the j-th projection value $(1 \le j \le Q)$, $\boldsymbol{p}_j\boldsymbol{h}$, where $\boldsymbol{p}_j$ is the j-th row of $\boldsymbol{P}$. Since the elements of $\boldsymbol{p}_j$ follow the standard normal distribution, it is easy to verify that $\boldsymbol{p}_j\boldsymbol{h} \sim \mathcal{N}(0, \|\boldsymbol{h}\|_2^2)$ and its conditional entropy is:

$$H(\boldsymbol{p}_j\boldsymbol{h}|\boldsymbol{h}) = \frac{1}{2}\log_2(2\pi e\|\boldsymbol{h}\|_2^2). \tag{6}$$

The projection value, $\boldsymbol{p}_j\boldsymbol{h}$, is then binarized to hash bit, b_{K+j}, by comparing it with zero. As discussed above, $\boldsymbol{p}_j\boldsymbol{h}$ follows zero mean normal distribution, so the resulting binary bit is uniformly distributed in $\{0, 1\}$: $\Pr(b_{K+j} = 0|\boldsymbol{h}) = \Pr(b_{K+j} = 1|\boldsymbol{h}) = \frac{1}{2}$. Recall that projection directions are independently generated, the conditional entropy of the Q hash bits generated by $\boldsymbol{h}$ is

$$H(b_{K+1}, \cdots, b_{K+Q}|\boldsymbol{h}) = \sum_{j=1}^{Q} H(b_{K+j}|\boldsymbol{h}) = 2Q. \tag{7}$$

As (7) shows, the random projection and binarization schemes can maximize the randomness of output hash.

3.2. Content identification performance analysis

In the following, we analyze the content identification performance of the proposed algorithm from the information theory perspective. Similar to [17], we first model robust hashing based content identification as the communication over noisy channel. Consider a database of W reference images, the indices $\{1, \cdots, W\}$ form the message set in communication. Without loss of generality, we consider n-length hash: $\boldsymbol{b}^{(n)} = \{b_1, \cdots, b_n\} \in \{0, 1\}^n$. Each image in this database is represented by a hash string, and this process can be modeled as an encoder: $\mathscr{E}(\cdot) : \{1, \cdots, M\} \to \{0, 1\}^n$. More specifically, the hash string of the i-th reference image serves as the codeword of the symbol i. Given a query image, content identification system compares its hash string $\boldsymbol{y}^{(n)}$ with those in database to determine whether it is perceptually similar to any reference image. Denote the hash string of query image by $\boldsymbol{y}^{(n)}$, then the content identification process is equivalent to the channel decoding stage: $\mathscr{D}(\cdot) : \{0, 1\}^n \to \{1, \cdots, M, \emptyset\}$[1]. Accordingly, the error of content identification can be measured by

$$P_e = \frac{1}{M}\sum_{i=1}^{M} \Pr(\mathscr{D}(\boldsymbol{y}^{(n)}) \neq i|\boldsymbol{b}^{(n)} = \mathscr{E}(i)). \tag{8}$$

Since the hash string computed by the proposed algorithm is bit-wise independent, the channel transition probability obeys:

$$p(\boldsymbol{y}^{(n)}|\boldsymbol{b}^{(n)}) = \prod_{i=1}^{n} p(y_i|b_i), \tag{9}$$

Assume that content-preserving distortions flip each hash bit with equal probability, i.e. $p(y_i|b_i) = p$, the analytical model can be viewed as a binary symmetric channel whose capacity is:

$$C = \max I(b_i; y_i) = 1 - p. \tag{10}$$

The capacity is achieved when hash bits are independently and uniformly distributed in $\{0, 1\}^n$. According to the channel coding theorem and the Fano's inequality [23], we can derive the following performance bound:

$$(1 - P_e)\log_2 M \le 1 + nC. \tag{11}$$

By estimating p and substituting (10) into (11), we can get the relationship between error rate P_e, hash length N and database size M.

[1] $\emptyset$ corresponds to the case that the query is not perceptually similar to any reference image.

4. EXPERIMENTAL RESULTS

The testing database in content identification experiments was composed of 2,000 reference images and 134,000 distorted images generated by a variety of content-preserving manipulations (as listed Table 1). The parameter settings of the proposed algorithm are as follows: $N = 128$, $W = 32$ $K = 80$, $Q = 10$ and $\lambda = 1/\sqrt{128}$. The proposed work was compared with four representative image hashing algorithms: the sparse coding based hashing (SC) [15], the Gabor filtering and dithered lattice vector quantization based hashing (GF-DLVQ) [12], the nonnegative matrix factorization (NMF) based hashing [7], and the ring-partition and NMF based hashing (Ring-NMF) [8].

Table 1. CONTENT-PRESERVING MANIPULATIONS

Manipulation	Strength
JPEG Compression	Quality factor $\in [1, 95]$
Circular Filtering	Radius $\in [1, 3]$
Median Filtering	Filter size $\in [2, 20]$
Gaussian Noise	Zero mean, variance $\in [0.1, 1]$
Speckle Noise	Zero mean, variance $\in [0.01, 0.3]$
Rotation+Cropping	$\theta \in [1, 10]$
Histogram Equalization	Number of gray levels $\in [8, 224]$
Gamma Correction	$\gamma \in [0.55, 1.45]$

The proposed work, SC and GF-DLVQ are all binary hashing algorithms, so the normalized hamming distance was adopted to measure the similarity between hash strings. For NMF and Ring-NMF that output real-valued hash, we use the l_2 norm and correlation coefficient based distance metrics, as suggested in [7] and [8], respectively. Fig.3 shows the receiver operating characteristic (ROC) curves of these hashing algorithms with the false rejection rate (FRR) in logarithmic scale, and their F_1 scores are listed in Table 2 to make a quantitative comparison on accuracy. The proposed algorithm and SC have the highest F_1 score. However, as can be seen from Fig.3, the proposed algorithm outperforms SC in the region with low false rejection rate (FRR). Hence, it is more suitable for the applications that have stringent requirement on FRR, such as digital right management. It is obvious from Fig.3 and Table 2 that the proposed algorithm demonstrates notable superiority over NMF and Ring-NMF. As mentioned in Section 2, a part of hash bits are generated from the low-rank representation of input image. Similarly, NMF and Ring-NMF aggregates the low-rank and non-negative matrix factors of input image into hash values. However, it should be noted that the our work further applies random compressive sampling (RCS) on the low-rank term $\boldsymbol{L}$. RCS can preserve the most stable components of $\boldsymbol{L}$ while reducing its redundances; hence, the measurement vector generated by RCS is more robust than the matrix factors decomposed by NMF. Moreover, our algorithm also combines the salient information of input image, making output hash more discriminative. Aside from its high accuracy, the proposed work also outperforms other algorithms in terms of compactness (the length of output hash is 90 bits, as shown in Table 3).

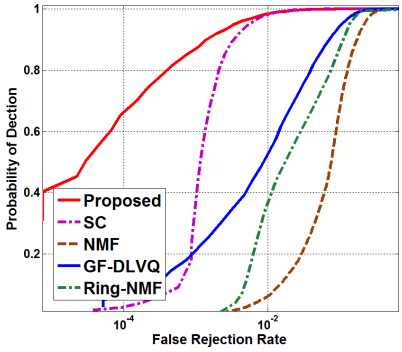

Fig. 3. ROC curves.

Table 2. COMPARISON ON F_1 SCORE

Proposed	SC	GF-DLVQ	NMF	Ring-NMF
0.9888	0.9888	0.9221	0.8652	0.9065

Table 3. COMPARISON ON HASH LENGTH

Proposed	SC	GF-DLVQ	NMF	Ring-NMF
90 bits	90 bits	120 bits	64 real numbers	64 real numbers

In Section 3.2, we have analyzed the performance of binary hashing algorithm in content identification and shown that the performance bound depends on the probability of bit flipping. We estimated this probability from the hash strings of 134,000 pairs of original and distorted images, the result of which shows $p = 0.0549$. Accordingly, the channel capacity is $C = 0.9451$.

5. CONCLUSIONS

We have proposed a low-rank and sparse decomposition based image hashing algorithm. The high robustness, discriminability and compactness of the proposed algorithm can be attributed to the following reasons: 1) the low-rank and sparsity constrains enable hashing algorithm to simultaneously capture the structural information and salient components of input image; 2) the compressive sampling process can provide a compact, stable and secure representation of the low-rank term; 3) incorporating the layout of salient regions in hash results in higher discriminability.

6. REFERENCES

[1] R. Venkatesan, S. M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in *Proc. IEEE Conf. Image Processing*, Sep. 2000, vol.3, pp.664–666.

[2] Y. Zhao, S. Wang, X. Zhang, and H. Yao, "Robust hashing for image authentication using Zernike moments and local features," *IEEE Trans. Inf. Forens. Sec.*, vol.8, no.1, pp.55–63, Jan. 2013.

[3] Y. S. Choi and J. H. Park, "Image hash generation method using hierarchical histogram," *Multimed. Tools Appl.*, vol.61, no.1, pp.181–194, Nov. 2012.

[4] A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure image hashing," *IEEE Trans. Inf. Forens. Sec.*, vol.1, no.2, pp.215–230, Jun. 2006.

[5] M. M. Esmaeili, M. Fatourechi, and R. K. Ward, "A robust and fast video copy detection system using content-based fingerprinting," *IEEE Trans. Inf. Forens. Sec.*, vol.6, no.1, pp.213–226, Mar. 2011.

[6] Y. Li, P. Wang, and Y. Su,"Robust image hashing based on selective quaternion invariance," *IEEE Signal Process. Lett.*, vol.22, no.12, pp.2396–2400, Dec. 2015.

[7] V. Monga and M. K. Mıhçak, "Robust and secure image hashing via non-negative matrix factorizations," *IEEE Trans. Inf. Forens. Sec.*, vol.2, no.3, pp.376–390, Sep. 2007.

[8] Z. J. Tang, X. Q. Zhang, and S. C. Zhang, "Robust perceptual image hashing based on ring partition and NMF," *IEEE Trans. Knowl. Data Eng.*, vol.26, no.3, pp.711–724, Mar. 2014.

[9] X. Lv and Z. J. Wang, "Perceptual image hashing based on shape contexts and local feature points," *IEEE Trans. Inf. Forens. Sec.*, vol.7, no.3, pp.1081–1093, June 2012.

[10] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: performance evaluation and tradeoffs," *IEEE Trans. Image Process.*, vol.15, no.11, pp.3453–3466, Nov. 2006.

[11] M. K. Mıhçak and R. Venkatesan, "A perceptual audio hashing algorithm: a tool for robust audio identification and information hiding," in *Proc. Workshop Information Hiding*, Pittsburgh, PA, Apr. 2001, vol.2137, pp.51–65.

[12] Y. Li, Z. Lu, C. Zhu, and X. Niu, "Robust image hashing based on random Gabor filtering and dithered lattice vector quantization," *IEEE Trans. Image Process.*, vol.21, no.4, pp.1963–1980, Apr. 2012.

[13] V. Monga, A. Banerjee, and B. L. Evans, "A clustering based approach to perceptual image hashing," *IEEE Trans. Inf. Forens. Sec.*, vol.1, no.1, pp.68–79, Mar. 2006.

[14] X. Lv and Z. J. Wang, "Compressed binary image hashes based on semisupervised spectral embedding," *IEEE Trans. Inf. Forens. Sec.*, vol.8, no.11, pp.1838–1849, Nov. 2013.

[15] Y. Li, "Robust content fingerprinting algorithm based on sparse coding," *IEEE Signal Process. Lett.*, vol.22, no.9, pp.1254–1258, Sep. 2015.

[16] A. L. Varna and M. Wu, "Modeling and analysis of content identification," in *Proc. IEEE Conf. Multimedia and Expo*, June. 2009, pp.1528–1531.

[17] S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Conception and limits of robust perceptual hashing: toward side information assisted hash functions," in *Proc. SPIE Media Forens. Sec. XI*, Jan. 2009.

[18] P. Moulin, "Statistical modeling and analysis of content identification," in *Proc. Info. Theory Appl. Workshop* , Jan. 2010, pp.1–5.

[19] E. J. Candes, X. Li, and Y. Ma, "Robust principal component analysis?" *J. ACM*, vol.58, no.3, May 2011.

[20] D. M. Beck and S. Kastner, "Top-down and bottom-up mechanisms in biasing competition in the human brain," *Vis. Res.*, vol.49, no.10, pp.1154–1165, May 2009.

[21] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," UIUC Tech. Rep. UILU-ENG-09-2215, Nov. 2009.

[22] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol.60, no.1, Jan. 2012, 139–154.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ:Wiley, 2006.