Hindawi
Advances in Multimedia
Volume 2018, Article ID 4235268, 11 pages
https://doi.org/10.1155/2018/4235268

Research Article

Image Hashing for Tamper Detection with Multiview Embedding and Perceptual Saliency

Ling Du, Zhen Chen, and Yongzhen Ke

School of Computer Science and Software Engineering, Tianjin Key Laboratory of Optoelectronic Detection Technology and System, Tianjin Polytechnic University, Tianjin 300387, China

Correspondence should be addressed to Ling Du; duling@tjpu.edu.cn

Received 29 March 2018; Accepted 3 October 2018; Published 19 November 2018

Academic Editor: Kjell Brunnström

Copyright © 2018 Ling Du et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Perceptual hashing technique for tamper detection has been intensively investigated owing to the speed and memory efficiency. Recent researches have shown that leveraging supervised information could lead to learn a high-quality hashing code. However, most existing methods generate hashing code by treating each region equally while ignoring the different perceptual saliency relating to the semantic information. We argue that the integrity for salient objects is more critical and important to be verified, since the semantic content is highly connected to them. In this paper, we propose a Multi-View Semi-supervised Hashing algorithm with Perceptual Saliency (MV-SHPS), which explores supervised information and multiple features into hashing learning simultaneously. Our method calculates the image hashing distance by taking into account the perceptual saliency rather than directly considering the distance value between total images. Extensive experiments on benchmark datasets have validated the effectiveness of our proposed method.

1. Introduction

With the widespread use of low cost and even free editing software, people can easily create a tampered image. Compared to forensic images, fake images could undergo kinds of manipulations, such as color changing, salient object changing, and copy-move forgery. Generally, there are two main problems in image forensics: one is tamper detection and the other one is tamper localization. Recently, more researchers pay attention to image tamper detection, which aims to discriminate whether a given image is pristine or fake. Image hashing based tamper detection approaches have been extensively studied recently for their great efficiency. It supports image content forensics by representing the semantic content in a compact signature, which should be robust against a wide range of content preserving attacks but sensitive to malicious manipulations.

For image hashing generation, the state-of-art hashing methods could be mainly divided into two categories: data independent hashing and data dependent hashing. In conventional image hashing methods, image hash generation is a robust feature compression process without any learning stage. It includes (1) invariant feature transform based methods, such as Wavelet transform [1], Radon transform [2], Fourier-Mellin transform [3], DCT transform [4], and QFT transform [5], which aim to extract robust features from transform domains; (2) local feature points based methods, such as SIFT [6] and end-stopped wavelet [7], which take advantages of the invariant local feature under some content preserving image processing attacks; (3) dimension reduction based methods, such as singular value decomposition (SVD) [8], nonnegative matrix factorization (NMF) [9], and Fast Johnson-Lindenstrauss transform (FJLT) [10], which embed the low level features of the high dimensional space into lower dimension; (4) statistics features based methods, such as the robust image hashing with ring partition and invariant vector distance [11]. Moreover, Wang et al. [12] propose a perceptual image hashing method by combining image block based features and key-point-based features. Yan et al. [13] use a multiscale image hashing method based on the adaptive local feature.

Since the hashing generation is independent of the data distribution, data independent hashing methods may not consider the characters of data distribution into hashing

Figure 1: Image preprocessing results for each step. (a) Original image, (b) Gaussian low-pass filtering, (c) LAB conversion, and (d) IntWT transform.

generation. Currently, more researchers begin to focus on the data dependent methods with learning for image tamper detection. Lv et al. [14] propose a semi-supervised spectral embedding method for image hashing. Efficient learning is incorporated into image hash generation by taking advantages of virtual prior attacked hash space (VPAHS). However, this algorithm only focuses on the postprocessing of image hashing. They assume the availability of real-valued image hashes and concentrate on the topic of compressing them into a short binary image hash. Currently, deep learning begins to be widely used in image forensics. Chen et al. [15] and Qian et al. [16] propose a median filtering detection and steganalysis based on convolutional neural networks (CNNs). Bayar et al. [17] propose a universal forensic approach to performing manipulation detection using deep learning. A new form of convolutional layer that is specifically designed to suppress an image's content and adaptively learn manipulation detection features is developed. Bondi et al. [18] propose a tampering detection and localization algorithm through clustering of camera-Based CNN features. The CNN is exploited to extract characteristic camera model features from image patches. Forgery patches are detected by the descriptors learned by CNN. Likewise, Yarlagadda et al. [19] propose a satellite image forgery detection and localization method using a generative adversarial network (GAN), which is also used for feature representation learning of pristine satellite images. More recently, video forgery detection [20] and camera model identification with CNNs [21, 22] are proposed. However, most of the algorithms only emphasise the feature learning by using of deep network.

Considering the abovementioned methods, there are two aspects which are not taken into full consideration. Firstly, most of the methods describe image content with single feature. Currently, most of features are only robust against for one or several types of attacks. It may not be feasible to extract one absolute robust feature which can satisfy the needs of users. Lv et al. [14] propose an image hashing algorithm based on semi-supervised spectral embedding. Two real-valued intermediate hashing methods are adopted for learning. Likewise, Yan et al. [5] proposed a quaternion-based image hashing for tampering localization. Four types of feature maps are selected for quaternion image formation. Secondly, current hashing methods usually acquire hashing detection results by treating each local region equally. Importantly, we argue that the integrity for salient objects, such as object adding, deleting, and semantic modifying, are more critical and important to be verified, since the semantic content of the image is highly connected to them. Zhang et al. [23] extract local texture features from salient regions to represent contents, which are combined with global features for computing final hash sequence. However, the saliency weights for selected regions are not taken into account for hashing metric distance. Therefore, how to efficiently combine different image features to enhance the overall performance and how to efficiently design image hashing approach based on perceptual saliency is a topic of great importance but less studied in current research.

In this paper, we present a Multi-View Semi-supervised Hashing with Perceptual Saliency (MV-SHPS) algorithm. The contributions are as follows:

(1) We effectively exploit simultaneously the supervised information and multiple features into the hashing learning.

(2) Instead of learning metric distance on global image, we explore the local hashing distance by considering the perceptual saliency effect among different regions.

(3) An extensive set of experiments on image datasets demonstrates that the proposed method outperforms several state-of-the-art perceptual image hashing techniques.

2. Proposed Method

2.1. Preprocessing. To alleviate effects of commonly used digital signal processing manipulations, the preprocessing is needed, as shown in Figure 1. All the input images are first converted to a standard $N \times N$ image by bilinear interpolation. The purpose of resizing is to resist possible resizing operations and ensure that those images with different sizes have the fixed hash length. And then, Gaussian low-pass filtering is applied to the standard image (Figure 1(b)), which can reduce the influence of minor modifications, such as noise contamination or filtering. As the CIE LAB color space is more perceptually uniform than other color space and the L component closely matches human perception of lightness. The RGB color image is firstly converted into the corresponding XYZ color space, and the XYZ color space is then converted into the corresponding LAB color space by the following [24, 25]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (1)$$

$$L = 116 f\left(\frac{Y}{Y_w}\right) - 16, \tag{2}$$

$$A = 500\left[f\left(\frac{X}{X_w}\right) - f\left(\frac{Y}{Y_w}\right)\right], \tag{3}$$

$$B = 200\left[f\left(\frac{Y}{Y_w}\right) - f\left(\frac{Z}{Z_w}\right)\right], \tag{4}$$

where R, G, and B are the red, green, and blue component of a pixel, X, Y, and Z are the CIE XYZ tristimulus values (1), and L, A, and B ((2), (4), and (3)) are color lightness, chromaticity, and coordinates, respectively. Xw=0.950456, Yw=1.0, and Zw=1.088754 are the CIE XYZ tristimulus values of the reference white point, and f(t) is calculated by the following rule:

$$f(t) = \begin{cases} t^{1/3}, & t > 0.008856 \\ 7.787t + \dfrac{16}{116}, & Otherwise, \end{cases} \tag{5}$$

and the L component is then taken for image representation (Figure 1(c)). Integer Wavelet Transform (IntWT) is an approximation of original image and is more robust against signal processing attacks. Therefore, we finally apply one-level IntWT to the L component and take the low frequency subband (LL) as the semantic perceptual image (Figure 1(d)), from which multiple types of feature are extracted for hash generation.

2.2. Hashing Learning. Suppose there are n images in the given whole set, represented as $\chi = \{\mathbf{x}_i\}, i = 1, 2, \ldots, n$, where $\mathbf{x}_i \in \mathbf{R}^D$ represents feature vector. For each image, we extract their V types of features. The task of multiview perceptual image hashing is to learn hash functions by simultaneously utilizing the feature matrices $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(V)}$, with $\mathbf{X}^{(v)} = [\mathbf{x}_1^{(v)}, \mathbf{x}_2^{(v)}, \ldots, \mathbf{x}_n^{(v)}]$ corresponding to the *v-th* type of feature matrix. Let $\mathbf{X} = \{\mathbf{X}_1 : \mathbf{X}_2 : \cdots : \mathbf{X}_n\}$ denote the combined matrix for multiview feature, where $\mathbf{X} \in \mathbf{R}^{D\times n}$, $D = \sum_{v=1}^{V} d_v$, and d_v is the dimension of *v-th* type feature. The goal of our algorithm is to learn hash functions that map $\mathbf{X} \in \mathbf{R}^{D\times n}$ to a compact representation $\mathbf{B}^{K\times n}$ in a low-dimensional Hamming space, where K is the digits length.

In the set χ, there are l labeled images, $l \ll n$, which are associated with at least one of the two categorizes $\mathscr{M}$ and $\mathscr{C}$. Specifically, a pair $(\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{M}$ is denoted as perceptually similar pair when $(\mathbf{x}_i, \mathbf{x}_j)$ are the images that have been under content-preserved un-malicious distortions and attacks. $(\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{C}$ is denoted as perceptually dissimilar pair when two samples are the original image and the one that is suffered from malicious manipulations or perceptually significant attacks such as object insertion and removal. Let us denote the feature matrix formed by the corresponding l columns of $\mathbf{X}$ as $\mathbf{X}_l \in \mathbf{R}^{D\times l}$. Note that the feature matrices are normalized to zero-centered.

We define the perceptual confidence measurement for each image example. The matrix $\mathbf{S} \in \mathbf{R}^{l\times l}$ incorporating the pairwise labeled information from $\mathbf{X}_l$, S_{ij} is the pairwise relationship for $(\mathbf{x}_i, \mathbf{x}_j)$, which is defined as

$$\mathbf{S}_{ij} = \begin{cases} 1 & (\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{M} \\ -1 & (\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{C} \\ 0 & otherwise. \end{cases} \tag{6}$$

Suppose we want to learn K hash functions that leading to a K-digit representation $\mathbf{B}$ of $\mathbf{X}$. For each digit $k = 1, 2, \ldots, K$, its hash function is defined as

$$h_k(\mathbf{x}_i) = \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i, \tag{7}$$

where $\mathbf{w}_k \in \mathbf{R}^D$ is the coefficient vector. Let $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k] \in \mathbf{R}^{D\times K}$ and the representation $\mathbf{B}$ of the feature matrix $\mathbf{X}$ for image set χ is

$$\mathbf{B} = \mathbf{W}^{\mathrm{T}} \mathbf{X}. \tag{8}$$

Our goal is to learn a $\mathbf{W}$ that is simultaneously maximizing the empirical accuracy on the labeled image and variance of hash bits over all images. The empirical accuracy on the labeled image is defined as

$$J_1(\mathbf{W}) = \sum_k \left\{ \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{M}} \mathbf{S}_{ij} h_k(\mathbf{x}_i) h_k(\mathbf{x}_j) + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{C}} \mathbf{S}_{ij} h_k(\mathbf{x}_i) h_k(\mathbf{x}_j) \right\}. \tag{9}$$

The objective function for empirical accuracy can be represented as

$$J_1(\mathbf{W}) = \frac{1}{2} \mathbf{tr}\left\{ \left(\mathbf{W}^{\mathrm{T}} \mathbf{X}_l\right) \mathbf{S} \left(\mathbf{W}^{\mathrm{T}} \mathbf{X}_l\right)^{\mathrm{T}} \right\}. \tag{10}$$

Then, the empirical accuracy $J_1(\mathbf{W})$ is presented as

$$J_1(\mathbf{W}) = \frac{1}{2} \mathbf{tr}\left\{ \mathbf{W}^{\mathrm{T}} \mathbf{X}_l \mathbf{S} \mathbf{X}_l^{\mathrm{T}} \mathbf{W} \right\}. \tag{11}$$

Moreover, to maximize the information provided by each bit, the variance of hash bits over all data $\mathbf{X}$ is also measured and taken as a regularization term:

$$R(\mathbf{W}) = \sum_k \mathrm{var}\left[h_k(\mathbf{X})\right] = \sum_k \mathrm{var}\left[\mathbf{w}_k^{\mathrm{T}} \mathbf{X}\right]. \tag{12}$$

Maximizing the above function with respect to $\mathbf{W}$ is still hard due to its nondifferentiability. As the maximum variance of a hash function is lower bounded by the scaled variance of the projected data, the information theoretic regularization is represented as

$$J_2(\mathbf{W}) = \frac{1}{2} \mathbf{tr}\left\{ \left(\mathbf{W}^{\mathrm{T}} \mathbf{X}\right) \left(\mathbf{W}^{\mathrm{T}} \mathbf{X}\right)^{\mathrm{T}} \right\}. \tag{13}$$

Finally, the overall semi-supervised objective function combines the relaxed empirical fitness term from (11) and

the regularization term from (13). We get the following optimization problem [26]:

$$\max_{\mathbf{W}} \quad J(\mathbf{W}) \qquad \text{s.t.} \quad \mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{I}, \tag{14}$$

with

$$J(\mathbf{W}) = J_1(\mathbf{W}) + \eta J_2(\mathbf{W}) = \frac{1}{2}\mathrm{tr}\left\{\mathbf{W}^{\mathrm{T}}\mathbf{M}\mathbf{W}\right\}, \tag{15}$$

where $\mathbf{M} = \mathbf{X}_l\mathbf{S}\mathbf{X}_l^{\mathrm{T}} + \eta\mathbf{X}\mathbf{X}^{\mathrm{T}}$, η is a tradeoff parameter, and the constraint $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{I}$ makes the projection directions orthogonal. Learning the optimal projections $\mathbf{W}$ can be solved by eigenvalue decomposition on matrix $\mathbf{M}$.

2.3. Perceptual Saliency. Image forgeries are often created by combining several images, including object adding, deleting, replacing, etc., which are highly relevant to the human perception. In other words, these object modifications usually affect the perceptual saliency of the corresponding image. In this paper, we call this variation on saliency map between trust and test image as perceptual saliency and consider it as a hint for tamper. Therefore, in our proposed method, we explore the computing of image hashing by considering the perceptual saliency effect rather than hashing acquiring from total image directly. According to [27], we take the structured matrix decomposition (SMD) model that treats the (salient) foreground/background separation as a problem of low-rank and structured-sparse matrix decomposition, to compute the saliency map of a given image.

Given the feature matrix of an input image, it can be decomposed as a low-rank matrix $\mathbf{L}$ corresponding to the nonsalient background and a sparse matrix $\mathbf{V}$ corresponding to the salient foreground objects. The structured matrix decomposition model can be formulated as

$$\min_{\mathbf{L},\mathbf{V}} \boldsymbol{\Psi}(L) + \alpha\boldsymbol{\Omega}(\mathbf{V}) + \beta\boldsymbol{\Theta}, (\mathbf{L}, \mathbf{V}), \tag{16}$$

where $\boldsymbol{\Psi}(L)$ is a low-rank constraint to allow identification of the intrinsic feature subspace of the redundant background patches, $\boldsymbol{\Omega}(\mathbf{V})$ is structured-sparsity Regularization, $\boldsymbol{\Theta}(\mathbf{L}, \mathbf{V})$ is Laplacian regularization, and α and β are positive tradeoff parameters.

2.4. Tamper Detection. For tamper detection, a forensic hash should be calculated from a trusted image and sent to a destination after encoding. Divide the original image into overlapping and pseudo randomly selected rectangular regions R_r, $r = 1, 2, \ldots, Num$. For each region, we extract V type of feature matrix $\mathbf{X}_a$ and obtain the corresponding hashing code $\mathbf{B}_a$. Likewise, the same procedures are employed to the test image to calculate hashing code $\mathbf{B}_t$ with respect to feature matrix $\mathbf{X}_t$.

$$\mathbf{B}_a = \mathbf{W}^{\mathrm{T}}\mathbf{X}_a, \qquad \mathbf{B}_t = \mathbf{W}^{\mathrm{T}}\mathbf{X}_t, \tag{17}$$

Considering the perceptual saliency, the metric distance between two hashing code is calculated by

$$Dist_r = \left(1 + abs\left(\lambda_{ra} - \lambda_{rt}\right)\right)\left\|\frac{\mathbf{B}_a - \mathbf{B}_t}{2\sqrt{\|\mathbf{B}_a\|\,\|\mathbf{B}_t\|}}\right\|, \tag{18}$$

where Num is the number of random selected regions and λ_{ra} and λ_{rt} are the salience weights for each region of original and tampered images. We finally find the distance that leads to the highest difference value and call it $Dist$, which is obtained by

$$Dist = \arg\max_{r=1,2,\ldots,Num} Dist_r. \tag{19}$$

For tamper detection, a forensic hash should be calculated from a trusted image and sent to destination after encoding. Finally, the threshold τ is defined to judge whether the test image $\mathbf{I}$ is a similar image or a tampered image.

$$\mathbf{I} = \begin{cases} is\ pristine & Dist \le \tau \\ is\ fake & Dist > \tau. \end{cases} \tag{20}$$

The metric distance threshold parameter for tamper detection in our method is set as 0.16. Here, we test the probability distribution of the detection results with varying thresholds on our newly created database and finally determine it based on empirical value.

3. Experiments

3.1. Experiment Setting

3.1.1. Dataset. In our experiments, we employ four real-world datasets for evaluation. Our training dataset is generated in the basis of Kodak (http://r0k.us/graphics/kodak/). We adopt 18 thousands unique images in the training set generated from Kodak as our training data and randomly sampled 5K images as labeled subset. It includes similar images with different type of content preserving attacks and tampered images with particular logo insert. For each image, the ground-truth similar images are derived from index label; i.e., images from the same index are deemed to be similar. For test, three other real-world datasets are CASIA v1.0 [26], Realistic Tampering Dataset(RTD) [28, 29], and our newly created database. CASIA consists of 800 original images and 921 tampered images, which are in JPEG format, with a size of 256 × 384, and belong to various categories according to their content scene, animal, architecture, character, plant, article, nature, texture, etc. Realistic Tampering Dataset (RTD) consists of 220 original images and 220 tampered images, which are in TIFF format, with a size of 1920 × 1080. Our newly created database includes 280 original color images and 280 tampered images with the size about 600 × 800. The tampered images are generated by changing colors of the scene elements, inserting or deleting different objects into the source images and substituting image background.

3.1.2. Metric and Parameter Setting. For algorithm parameters, we set the dimension of the test images $N = 256$, hashing learning parameter $\eta = 0.25$, and metric distance parameter $\tau = 0.16$ for all datasets. For each image, we extract

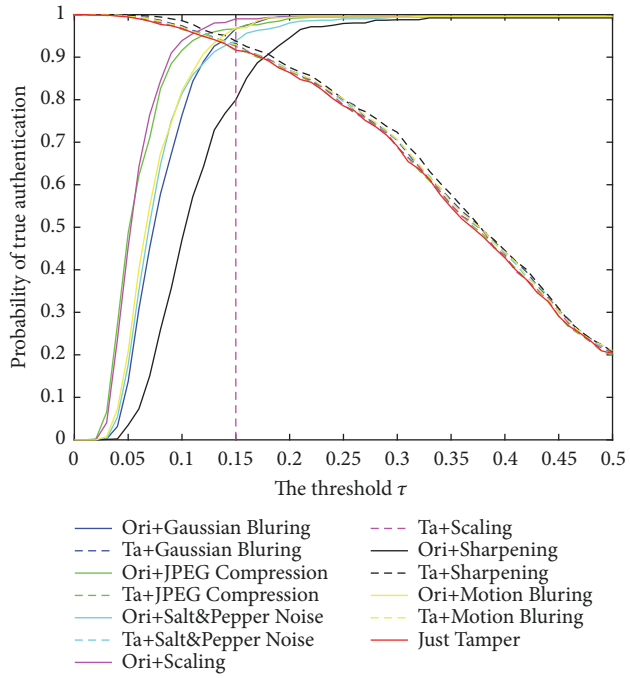


Figure 2: Image authentication performances with varying thresholds.

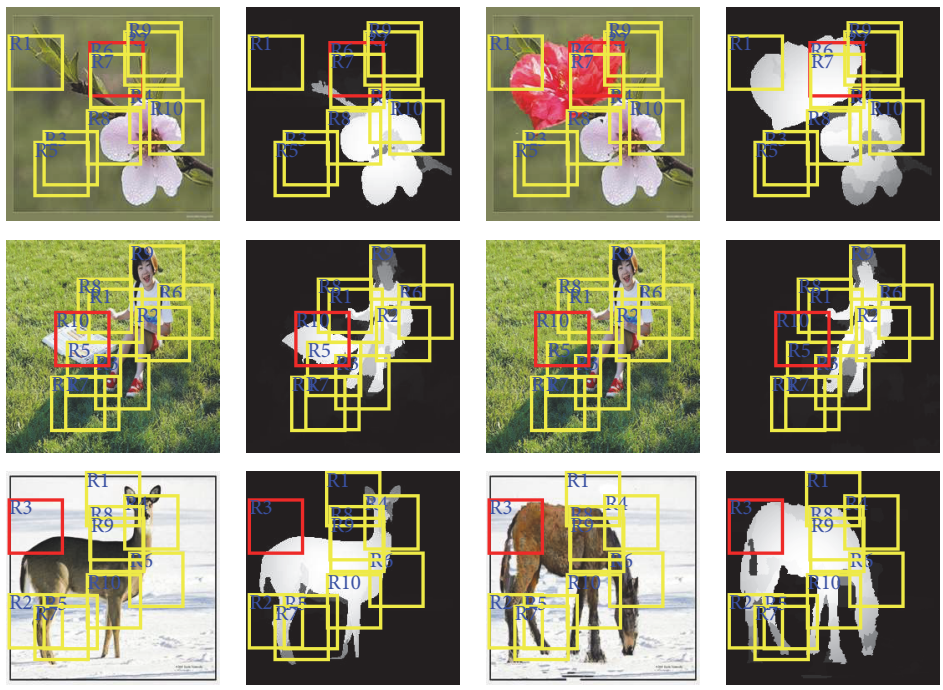
three type of features (view number $V = 3$): wavelet [1], SVD [8], and statistical [11] features as multiview observation. In this paper, we use the probability of true authentication, which means the ratio of similar/tampered images judged as similar/tampered images to the total number of corresponding type of images, as the evaluation metric for all the algorithms.

To prove the performance of image authentication we evaluate the image authentication performance using CASIA, Realistic Tampering Dataset (RTD) and newly created database. Except for simple tampering (TP) for image content, six types of content-preserving attacks are performed to verify the robustness of our proposed method: scaling with the percentage as 1.5, JPEG compression with the quality factor as 50, sharpening with the value as 0.49, Gaussian blurring with the size of the filter as 3,. and the standard deviation of the filter as 10, motion blurring with the amount of the linear motion as 3 and the angle of the motion blurring filter as 45, and salt & pepper noise with the noise density as 0.005.

For thresholds determination, we analyze the probability distribution of the authentication results with varying threshold t. As shown in Figure 2, the results shown by the solid line and dashed line indicate the probability distribution of similar images and tampered images under six types of content-preserving manipulations, respectively. The probability distribution results of similar and tampered images approximately intersect at $t = 0.15$. Therefore, in our experiments, we set $t = 0.15$ to distinguish the similar images and forgery images. For the compared methods, we tune all the parameters to best performances.

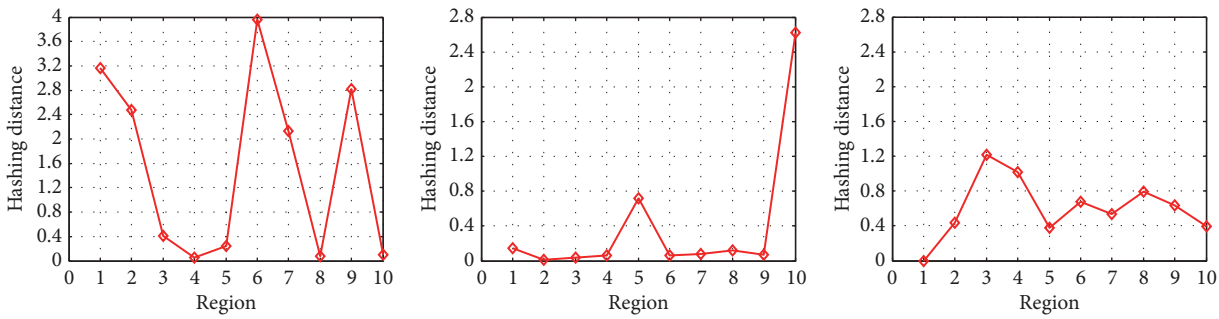
3.2. Perceptual Saliency Analysis. To evaluate the impacts of perceptual saliency for tamper detection results, Figure 3 shows some examples of metric distance for tamper detection. Here, we extract the image saliency map followed by randomly select ten regions ($Num = 10$) with size 64×64. The accuracy is based on smaller size and larger number. However, it also leads to higher hashing code length, which will decrease memory efficiency. We set the final parameter values by making a tradeoff between accuracy and efficiency. For each region, we resize it into 256×256 and compute the hashing code using (12). As shown in Figure 3, the modification for original image (columns (a) and (c)) are effectively map into the saliency map (columns (b) and (d)). For example, the semantic content for region six (R_6) of image A is modified by adding a red flower, leading to the higher perceptual difference in R_6 between two images. As shown in (13), we mark such difference and take it as weight for final hashing distance computing. Likewise, for object deleting and modifing, regions R_{10} and R_3 also reflect such tamper. Figure 4 illustrates the hashing distance for different regions with perceptual saliency corresponding to three images in Figure 3. Our perceptual saliency design for image hashing fully considers and improves the impact of local features. Figure 5 shows the probability of true authentication capability for tamper detection on newly created database with/without perceptual saliency under different threshold settings.

3.3. Comparison Results. We compare our method with the following baselines. Wavelet-based image hashing [1] develops an image hash based on an image statistics vector extracted from the various subbands in a wavelet decomposition of the image. SVD-based image hashing [8] uses spectral matrix invariants as embodied by singular value decomposition. RPIVD-based image hashing [11] incorporates ring partition and invariant vector distance to image hashing algorithm for enhancing rotation robustness and discriminative capability. Quaternion-based image hashing [5] constructs quaternion image, which combines advantages of both color and structural features, to implement the quaternion Fourier transform for image feature hashing generation. We report the the tamper detection results due to our emphasis. Table 1 shows the probability of true authentication capability of the proposed method compared to the methods proposed in [1, 5, 8, 11]. Note that the region size in our method is set as 64×64. We use the probability of true authentication capability to comprehensively evaluate the performance. For similar images, it records the ratio of similar images judged as similar images to the total number of similar images, which indicate the algorithm robustness. For tampered images, they record the ratio of tampered images judged as tampered images to the total number of tampered images, which indicate the algorithm discrimination. We conducted many experiments and calculated the corresponding results under various attacks. As is shown, the probability of true authentication capability with different content-preserving attacks on three databases is illustrated. Note that higher values indicate better performance for all metrics. In a big picture, our approach outperforms all the baselines. For the tamper



(a) (b) (c) (d)

Figure 3: Metric distance for tamper detection (row 1, Image A: object add, row 2, Image B: object delete, and row 3, Image C: object modify) based on salience map. (a) Original image, (b) Original saliency map, (c) Tampered image, and (d) Tampered saliency map. Ten regions with different salient map weights (colored rectangles) are selected for metric distance computing. Regions R_6, R_{10}, and R_3 corresponding to images A-C contribute the max distance for tamper detection.



(a) Image A (b) Image B (c) Image C

Figure 4: Hashing distance for different regions with perceptual saliency corresponding to three images in Figure 3.

detection, including removal, insertion, and replacement of objects, color modification, and background substitution, our method outperforms other methods, especially under various attacks. It should be noted that, for all experiments, we set our hashing length K as 64 digits, which is relative short compared with other methods.

3.4. Complexity Analysis. The complexity of the proposed image hashing algorithm that will be discussed here includes semi-supervised learning, saliency map generation and tamper detection. In the semi-supervised learning, it is actually the most time consuming step in our method because most of the time is spent on learning W. We sample a subset of the training items (e.g., containing l items). The pairwise similarity preserving considers the similarities of all pairs of items in the subset. The time complexity is $O(l^2K + l^2d)$. K is the number of hash code and d is the dimension of image feature. It is important to note that the semi-supervised learning process is an offline procedure, and the produced optimal projections W are then fixed for the whole procedure of proposed method. Practically, the training procedure has been done with a nonoptimized MATLAB code on a regular personal computer. This procedure can be preprocessed by any user on the personal computer with common configurations. As for our proposed scheme, the computational complexity mainly depends on saliency map and hashing

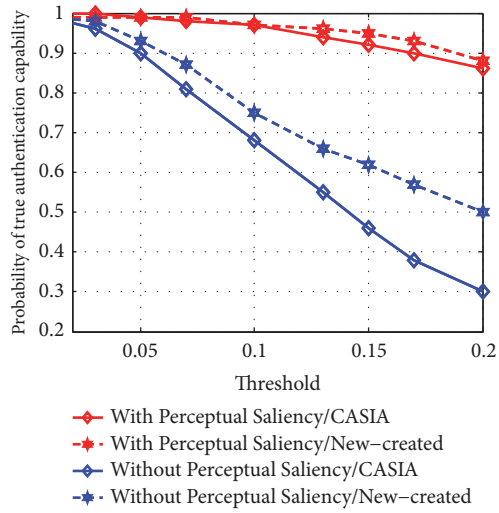


Figure 5: The probability of true authentication capability for tamper detection for CASIA/newly created database with/without perceptual saliency.

distance computations. For saliency map generation, the complexity depends on the salient detection algorithm. The current fast model is about 0.017 seconds per image. For tamper detection, our algorithm is to efficiently produce a sequence ordered by the increasing distances between the original and tamper images. The time complexity cost is $O(Num \log Num)$.

3.5. Comparison with Deep Learning Based Methods. For current forensics application, Chen et al. [15] and Qian et al. [16] propose a median filtering detection and steganalysis based on convolutional neural networks (CNNs). Likewise, Bayar et al. [17] propose image manipulation detection using deep learning. All of these methods focus on image manipulation, which are content preserving attacks. As for hashing application, the hashing code is robust against a wide range of content preserving attacks but sensitive to malicious manipulations. Bondi et al. [18] and Yarlagadda et al. [19] propose tampering detection and localization algorithms. However, these algorithms are not based on hashing operation. For image hashing based algorithm, the image semantic content is represented in a compact signature. Moreover, video forgery detection [20] and camera model identification with CNNs [21, 22] are proposed. In summary, most of current proposed algorithm are focused on image content preserving manipulations or not based on hashing representation. Our proposed method effectively exploits simultaneously the supervised information and multiple features into the hashing learning and performs tamper detection by considering the perceptual saliency effect among different regions. The comparison with deep learning based methods for image forensics is shown in Table 2.

3.6. Discussion. From the description of our MV-SHPS algorithm and the experimental results, we draw the conclusion that there are three aspects that importantly affect the perceptual hashing algorithm.

(1) Learning based image hashing: In our proposed method, we effectively exploit the supervised information into the hashing learning. The experimental results have shown that data dependent methods with learning can lead to high quality hashing. The process is trained to optimally fit data distributions and specific objective functions, which produce better hashing codes to preserve the local similarity. Therefore, how to efficiently learn hashing code based on the image data is a first topic of great importance in the future research.

(2) Image Hashing based on multiview embedding: Most of the methods describe image content with single feature. Since most features are only robust against for one or several types of attacks, it may not be feasible to extract one absolute robust feature which can satisfy the needs of users. Therefore, how to efficiently combine different image features to enhance the overall performance is a second topic of great importance but less studied in current research.

(3) Image hashing based on saliency detection: Current hashing methods usually acquire hashing detection results by treating each local region equally. Instead of learning metric distance on global image, we explore the local hashing distance by considering the perceptual saliency effect among different regions. Therefore, how to efficiently design image hashing approach based on perceptual saliency is a third topic of great importance for perceptual hashing algorithm on tamper detection.

4. Conclusion

In this paper, we proposed a novel Multi-View Semi-supervised Hashing algorithm with Perceptual Saliency (MV-SHPS). In summary, our proposed method has several desirable contributions: first, we effectively exploited simultaneously the supervised information and multiple features into the hashing learning. Second, instead of assuming only global image hashing contributes to metric distance of hash code for tamper detection, we explored the local hashing distance by considering the perceptual saliency effect among different regions. We performed extensive experiments on three image datasets compared with the state-of-the-art hashing techniques. Experimental results demonstrated that the proposed semi-supervised hashing with multiview features and perceptual saliency yields superior performance. The current work can be extended with the design of coregularized hashing for multiple features, which is expected to show even better performance.

Data Availability

All data generated or analysed during this study are included in this paper. For the datasets used in this paper, CASIA v1.0 and RTD can be downloaded from http://forensics.idealtest.org/ and http://kt.agh.edu.pl/~korus/downloads/dataset-realistic-tampering/. The new-created dataset is available from the corresponding author (duling@tjpu.edu.cn) on request.

Table 1: Probability of true authentication comparisons among five algorithms under malicious attacks and various content-preserving manipulations. TP and Ori in the table indicate tampered images and original images, respectively. NCD and RTD indicate the newly created dataset and the Realistic Tampering dataset, respectively.

Manipulations	Dataset & Method														
	CASIA					NCD					RTD				
	[1]	[8]	[11]	[5]	Our	[1]	[8]	[11]	[5]	Our	[1]	[8]	[11]	[5]	Our
Ori+Gaussian Bluring	0.97	0.97	0.95	0.99	0.98	0.99	0.99	0.94	0.99	0.97	0.99	0.99	0.71	0.99	0.95
Ori+JPEG Compression	0.98	0.98	0.99	0.94	0.98	0.99	0.99	0.98	0.95	0.98	0.99	0.99	0.95	0.86	0.89
Ori+Motion Bluring	0.97	0.98	0.96	0.98	0.97	0.99	0.99	0.95	0.99	0.99	0.99	0.99	0.96	0.99	0.97
Ori+Salt&Pepper Noise	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.98	0.99	0.96	0.99	0.95
Ori+ Image Scaling	0.63	0.76	0.71	0.69	0.80	0.67	0.81	0.63	0.88	0.83	0.69	0.88	0.71	0.90	0.85
Ori+ Image Sharpening	0.83	0.89	0.97	0.99	0.96	0.88	0.91	0.95	0.99	0.95	0.80	0.95	0.94	0.99	0.81
TP	0.83	0.90	0.84	0.90	0.92	0.91	0.91	0.89	0.93	0.95	0.84	0.79	0.81	0.73	0.85
TP+Gaussian Bluring	0.84	0.90	0.87	0.90	0.93	0.91	0.92	0.91	0.93	0.95	0.85	0.80	0.87	0.73	0.86
TP+JPEG Compression	0.83	0.90	0.85	0.91	0.92	0.91	0.92	0.88	0.93	0.95	0.85	0.80	0.84	0.78	0.86
TP+Motion Bluring	0.84	0.90	0.86	0.90	0.93	0.91	0.93	0.91	0.93	0.95	0.84	0.79	0.83	0.73	0.85
TP+Salt&Pepper Noise	0.83	0.90	0.86	0.90	0.93	0.91	0.92	0.91	0.93	0.95	0.85	0.80	0.82	0.73	0.87
TP+ Image Scaling	0.88	0.93	0.92	0.93	0.94	0.95	0.94	0.95	0.94	0.95	0.88	0.83	0.88	0.77	0.87
TP+ Image Sharpening	0.85	0.90	0.87	0.90	0.93	0.91	0.91	0.92	0.93	0.95	0.86	0.81	0.83	0.73	0.86

Table 2: Comparison with deep learning based methods for image forensics.

Method	Main technique	Application	Hashing
Chen [15]	Filter layer that output median filtering residual	Median Filtering detection	No
Qian [16]	Customized CNN model	Steganalysis	No
Bayar [17]	New convolutional layer to learn manipulation detection features	Image manipulation (i.e. median filtering, gaussian blurring, additive white Gaussian noise, eesampling) detection	No
Bondi [18]	Clustering of camera-based CNN features	Tampering Detection and Localization	No
Yarlagadda [19]	GAN and one-class classifier	Satellite image forgery detection and Localization	No
D'Avino [20]	Autoencoder with recurrent neural networks	Video forgery detection	No
Tuama [21]	A layer of preprocessing is added to the CNN model	Camera model identification	No
Bondi [22]	data-driven algorithm based on convolutional neural networks	Camera model identification	No
Proposed	Multiview feature, perceptual saliency, semi-supervised hashing	Image tamper detection	Yes

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was supported by National Natural Science Foundation of China (Grant no. 61602344), the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (Grant no: 2017KJ091), and Natural Science Foundation of Tianjin (Grant no. 17JCQNJC00100).

References

[1] R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, pp. 664–666, September 2000.

[2] F. Lefebvre, J. Czyz, and B. Macq, "A robust soft hash algorithm for digital image signature," in *Proceedings of the International Conference on Image Processing*, pp. 495–498, Barcelona, Spain.

[3] A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure image hashing," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 215–230, 2006.

[4] C. Kim, "Content-based image copy detection," *Signal Processing: Image Communication*, vol. 18, no. 3, pp. 169–184, 2003.

[5] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Quaternion-Based Image Hashing for Adaptive Tampering Localization," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2664–2677, 2016.

[6] X. Lv and Z. Jane Wang, "Perceptual image hashing based on shape contexts and local feature points," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1081–1093, 2012.

[7] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: Performance evaluation and tradeoffs," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3452–3465, 2006.

[8] S. S. Kozat, R. Venkatesan, and M. K. Mihçak, "Robust perceptual image hashing via matrix invariants," in *Proceedings of the International Conference on Image Processing, ICIP '04*, pp. 3443–3446, Singapore, October 2004.

[9] V. Monga and M. K. Mihçak, "Robust and secure image hashing via non-negative matrix factorizations," *IEEE Transactions on Information Forensics & Security*, vol. 2, no. 3, pp. 376–390, 2007.

[10] X. Lv and Z. J. Wang, "An extended image hashing concept: Content-based fingerprinting using FJLT," *EURASIP Journal on Information Security*, vol. 2009, pp. 1–17, 2009.

[11] Z. Tang, X. Zhang, X. Li, and S. Zhang, "Robust image hashing with ring partition and invariant vector distance," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 200–214, 2016.

[12] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, and J. Xue, "A visual model-based perceptual image hash for content authentication," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1336–1349, 2015.

[13] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Multi-scale image hashing using adaptive local feature extraction for robust tampering detection," *Signal Processing*, vol. 121, pp. 1–16, 2016.

[14] X. Lv and Z. J. Wang, "Compressed binary image hashes based on semisupervised spectral embedding," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1838–1849, 2013.

[15] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, "Median Filtering Forensics Based on Convolutional Neural Networks," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1849–1853, 2015.

[16] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Media Watermarking, Security, and Forensics*, 2015.

[17] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proceedings of the 4th ACM Information Hiding and Multimedia Security Workshop, IH and MMSec '16*, pp. 5–10, Spain, June 2016.

[18] L. Bondi, S. Lameri, D. Guera, P. Bestagini, E. J. Delp, and S. Tubaro, "Tampering Detection and Localization Through Clustering of Camera-Based CNN Features," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW '17*, pp. 1855–1864, USA, July 2017.

[19] S. K. Yarlagadda, D. Güera, P. Bestagini, F. Maggie Zhu, S. Tubaro, and E. J. Delp, "Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier," *Journal of Electronic Imaging*, vol. 2018, no. 7, pp. 214-1–214-9, 2018.

[20] D. D'Avino, D. Cozzolino, G. Poggi, and L. Verdoliva, "Autoencoder with recurrent neural networks for video forgery detection," in *Proceedings of the Media Watermarking, Security, and Forensics 2017, MWSF 2017*, pp. 92–99, USA, February 2017.

[21] A. Tuama, F. Comby, and M. Chaumont, "Camera model identification with the use of deep convolutional neural networks," in *Proceedings of the 8th IEEE International Workshop on Information Forensics and Security, WIFS 2016*, UAE, December 2016.

[22] L. Bondi, L. Baroffio, D. Guera, P. Bestagini, E. J. Delp, and S. Tubaro, "First Steps Toward Camera Model Identification with Convolutional Neural Networks," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 259–263, 2017.

[23] Y. Zhao, S. Wang, X. Zhang, and H. Yao, "Robust hashing for image authentication using zernike moments and local features," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 55–63, 2013.

[24] W. Burger and M. J. Burge, "Principles of digital image processing - core algorithms," in *Undergraduate Topics in Computer Science*, Undergraduate Topics in Computer Science, pp. 1–24, Springer London, London, 2009.

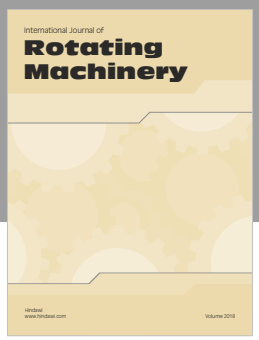
[25] R. C. González, R. E. Woods, and R. C. González, *Digital image processing*, 3rd edition, 2008.

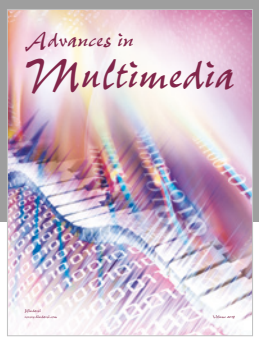
[26] J. Wang, S. Kumar, and S.-F. Chang, "Semi-supervised hashing for large-scale search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2393–2406, 2012.

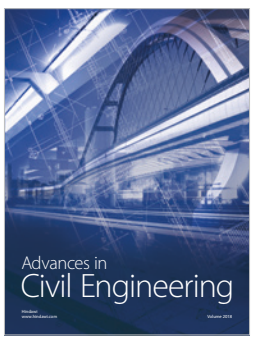
[27] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 818–832, 2017.

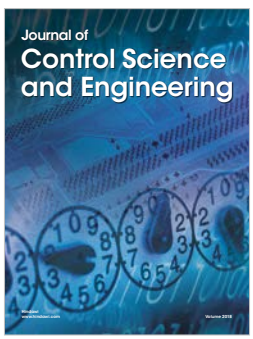
[28] P. Korus and J. Huang, "Multi-Scale Analysis Strategies in PRNU-Based Tampering Localization," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 809–824, 2017.

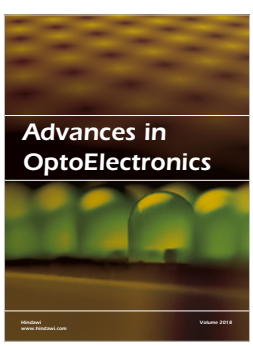
[29] P. Korus and J. Huang, "Evaluation of random field models in multi-modal unsupervised tampering localization," in *Proceedings of the 8th IEEE International Workshop on Information Forensics and Security, WIFS 2016*, pp. 1–6, UAE, December 2016.

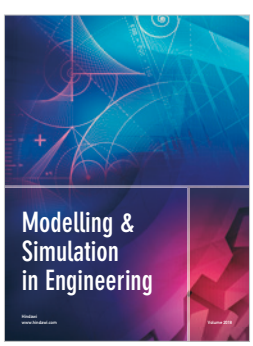


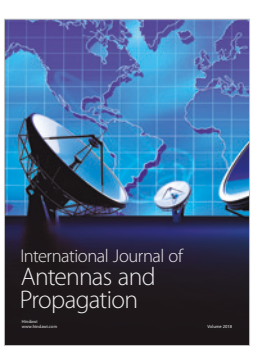


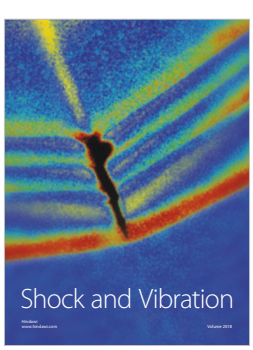


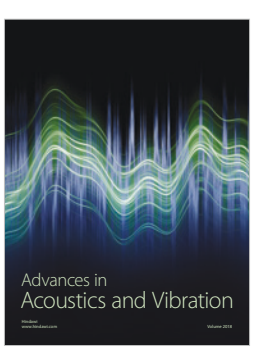












Hindawi

Submit your manuscripts at
www.hindawi.com