Learning to Hash for Personalized Image Authentication

Zhiyong Su, Liang Yao, Jialin Mei, Lang Zhou, and Weiqing Li

***Abstract*—This paper takes a fresh look at the image authentication problem and proposes an alternative framework for personalized authentication based on the hash learning technology. Conventional image authentication methods tend to provide a general authentication framework for all images with the fixed quantization strategy and fixed control parameters determined based on given limited images and attacks. However, they may suffer from more or less misjudgments in practice, not to mention performance degradation when encountering out-of-sample images. Instead of proposing a new feature extraction algorithm, a novel personalized authentication framework which incorporates the distance metric learning technology and supervised quantization strategy to the process of image authentication is proposed in this paper. The tamper detection task is reformulated as a new supervised manipulation classification problem. For each input image, various content-preserving and content-changing samples are generated automatically firstly. Then, feature representations of all samples can be obtained by existing feature extraction methods. After that, a weighted large margin for manipulation classification (WLMMC) scheme is proposed to learn an effective feature mapping space to improve the classification performance between content-changing samples and content-preserving samples. During the quantization stage, a novel supervised personalized quantization strategy (SPQ), which is motivated by the observation that different attacks have different degrees of influence on feature components, is proposed to learn more compact yet discriminative binary codes for each input image. Effectiveness of the proposed framework is qualitatively and quantitatively demonstrated on a variety of images. Extensive experiments show that the proposed framework can significantly improve the authentication performance over the state-of-the-art techniques while achieve more compact hash codes flexibly as required.**

***Index Terms*—Metric learning, supervised quantization, image authentication, image hashing, LMNN.**

I. Introduction

IMAGE hashing is a technique for deriving a content-based compact representation from the input image, which has been widely investigated and proposed as a primitive method to solve problems of image content authentication in recent years [1]–[11]. The main concept behind image authentication is to extract image characteristics of the human perception and use them during the authentication process. In general, the framework of traditional hashing based image authentication techniques can be decomposed into two main procedures, namely, feature extraction and feature quantization. In the feature extraction stage, image features, which are in accordance with the perceptual characteristic of the human visual system, are extracted from the image perceptual content. The feature quantization stage concerns about quantization, compaction and binarization of feature vectors. Image authentication is performed via comparing the hash value of an original image with the hash value of a doubted image. As well known, image hashing is expected to be robust against a wide range of content-preserving operations, while the discriminative capability to content-changing attacks or other different images should be also provided at the same time. Therefore, current trends in hashing based authentication research are headed toward developing a good feature extraction method to achieve robustness, sensitivity and discriminability [2].

Although extensive research efforts have been invested for image authentication and many image hashing algorithms and their variations have been proposed, the state-of-the-art performance is still far from satisfactory.

(1) Existing methods are designed to be a general framework with fixed control parameters, which is expected to apply to all images, especially out-of-sample images. To achieve satisfactory overall performance, various control parameters should be carefully selected through extensive experiments carried out on a large number of images collected from custom datasets or public image datasets, such as the UCID image database [12] and ImageNet [13]. However, in theory, those fixed parameters are not the most appropriate for every test image, not to mention out-of-sample images. This inference has also been confirmed by inevitable misjudgments in the experiments of existing hashing based authentication algorithms [4]–[6], [8], [10].

(2) It is sometimes difficult to distinguish between some content-changing samples and content-preserving samples in the original feature space directly, although a variety of well designed feature extraction methods, including hand-crafted methods and deep learning based methods, have been developed in prior works [8], [10], [14]. Therefore, it can be seen from experiments that the overlapping regions between the distance distribution of content-preserving images and content-changing images always exist for every hashing algorithm [10]. Threshold values, which are always introduced

Manuscript received April 00, 0000; revised August 00, 0000. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1004904, in part by the Fundamental Research Funds for the Central Universities under Grant 30918012203 and in part by the National Natural Science Foundation of China under Grant 61300160. (*Corresponding author: Zhiyong Su*.)

Z. Su, L. Yao and J. Mei are with the School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: su@njust.edu.cn; 1317018550@qq.com; 1904329513@qq.com).

L. Zhou is with the College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China (e-mail: yzzhoulang@126.com).

W. Li is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: li_weiqing@njust.edu.cn).

to judge whether the received image is tampered or not, should be carefully considered to reach a desirable balance among robustness, sensitivity, and discriminability through lots of experiments carried out on as many images as possible. However, due to the limited performance of existing feature representations in the original feature space, current methods still suffer from misjudgments according to their reported experimental results [4]–[6], [8], [10], [14]. And, it is definitely still a challenge to develop a satisfactory feature extraction method which can meet a wide range of authentication requirements.

(3) Existing quantization techniques for authenticating images overlook an important fact that different attacks have various degrees of influence on each dimension of feature vectors even obtained by same feature extraction methods [8], [10], [14], [15]. They quantize feature vectors of different image with the same length through a fixed quantization strategy without considering the discrimination distribution of each dimension. However, it is observed that the discriminative ability of each dimension varies from image to image under same or different attacks. Therefore, fixed quantization strategies also make contributions to misjudgments in practice.

A. Motivation

To explore a possible solution to above problems, contrary to previous works which concentrate on designing a general authentication system with fixed quantization strategy and fixed control parameters for all images, this paper argues that it is a feasible choice to adopt a personalized authentication framework for different images. The motivation behind this paper covers two different aspects: distance metric learning for manipulation classification as well as supervised personalized quantization.

(1) Firstly, this paper suggests formulating the tamper detection task as a new supervised manipulation classification problem, and introduces the distance metric learning (DML) technology to enhance the manipulation classification performance of existing feature extraction methods.

Recently, the emergence of DML has opened the door to a new family of methods for many potential scenarios, such as image classification and image retrieval. DML refers to learn a desired distance metric from given training samples, measured by which the samples from the same class are as close as possible, while the samples from different classes are as far as possible [**?**]. DML algorithms can be categorized as unsupervised, semi-supervised or supervised, according to the availability of supervision information during the distance metric learning process [16]–[18]. Various literature has demonstrated, either theoretically or empirically, that learning a good distance metric can significantly improve the performance of classification, clustering and retrieval tasks in recent years. And, it has been used in various applications such as computer vision, information retrieval and bioinformatics [16], [17], [19]–[22].

However, to the best of our knowledge, few literatures considered potential applications of DML in the field of image authentication. A distance metric learning algorithm for a fingerprinting system was proposed to identify a query content by finding the fingerprint in the database that measures the shortest distance to the query fingerprint [23]. For a given training set consisting of original and distorted fingerprints, a distance metric equivalent to the l_p norm of the difference between two linearly projected fingerprints was learned by minimizing the false-positive rate for a given false-negative rate.

Therefore, based on existing feature extraction methods, this paper aims to learn a more discriminative metric space and powerful feature representations from all samples attacked by different manipulations of each original image through the DML technology. And, the problems of misjudgments as well as out-of-sample images are expected to be well alleviated by the proposed scheme.

(2) Secondly, this paper dedicates to take advantage of the discriminative ability disparities among dimensions to design a personalized quantization strategy for each image to get more compact hash codes flexibly as required while achieve competitive authentication performance.

Currently, the majority of existing hashing based image authentication methods adopt threshold-based single-bit quantization (SBQ) to binarize each dimension into 0 or 1 [8], [15]. The threshold of a certain projected dimension is usually set as the mean value or the median value of the projected values of this dimension. The hierarchical quantization (HQ) is the first proposed non-SBQ quantization method [24]. Rather than using one bit, HQ employs three thresholds to divide each dimension into four regions and allocates two bits to encode each region. The double-bit quantization (DBQ) is another quantization strategy [25]. It quantizes each projected dimension into double bits with adaptively learned thresholds that divide the real-valued axis into three regions. A variable bit quantization (VBQ) method was proposed for locality sensitive hashing, in which bits are allocated across hyperplanes [26]. Recently, an unsupervised quantization strategy called between-cluster distance-based quantization (BCDQ) was presented to learn binary image fingerprints [14]. BCDQ clusters the samples of each dimension instead of a fixed threshold to generate binary fingerprint codes, and this clustering can preserve more neighborhood structures.

However, in the image authentication scenario, it is observed that the same or different attack has different influence on each dimension of different images. The number of bits allocated to quantize each dimension should depend on the discriminative ability of the data within that dimension. Therefore, this paper investigates a new personalized quantization strategy for each image in which the discrimination information is considered for each dimension.

B. Contributions

The main contributions of this paper are summarized as follows.

(1) First of all, a novel personalized framework, which incorporates the distance metric learning technology and supervised personalized quantization strategy, is proposed to authenticate each image with individualized parameters. The

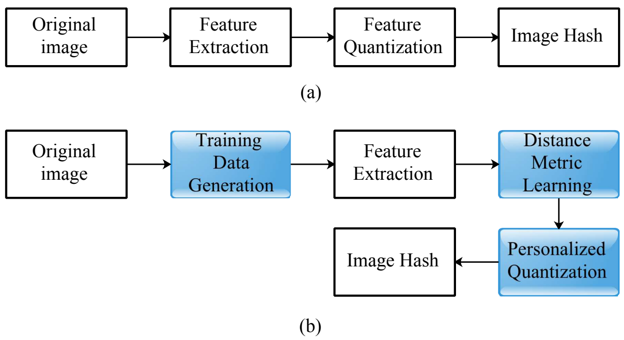


Fig. 1. Comparison of the proposed personalized authentication framework with the traditional framework. (a) Flowchart of the traditional image hashing. (b) Flowchart of the proposed personalized authentication framework. The key differences are highlighted in blue boxes.

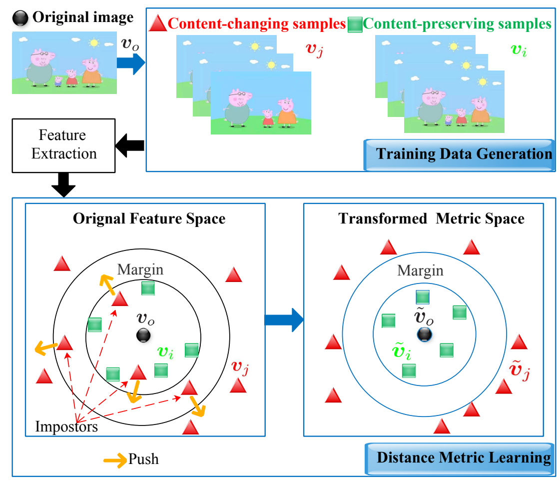


Fig. 2. Flowchart of the distance metric learning for manipulation classification. The samples in the training sets that invade the margin are called impostors.

personalized framework is designed to couple with feature extraction methods of existing works.

(2) Second, a novel distance metric learning algorithm, called weighted large margin for manipulation classification (WLMMC), is proposed to learn an effective feature mapping space for each original image from its training samples, in which its content-preserving samples are expected to be mapped close to it and its content-changing samples are mapped farther apart. The ultimate goal of WLMMC is to improve the classification accuracy of content-preserving and content-changing manipulations by learning a linear transformation matrix.

(3) Third, a novel supervised personalized quantization strategy (SPQ), which fully exploits the discriminative ability of each dimension, is proposed to learn more compact binary codes for each image. Furthermore, it can allocate variable hash bits for each individual image as required, while achieve competitive authentication performance as possible.

There is no question that, for the image hashing scheme, feature extraction and representation methods are very critical to the final authentication performance. However, it is worth to clarify that, the proposed personalized framework aims to enhance the authentication performance of existing feature extraction methods through the WLMMC and SPQ schemes as opposed to replacing the methods themselves. The source codes will be publicly available at: https://zhiyongsu.github.io.

The remainder of this paper is organized as follows. Section II overviews the framework of the proposed scheme. The proposed WLMMC scheme is described in Section III, which is followed by the supervised personalized quantization in Section IV. Section V presents the performance analysis and experimental results. Finally, the conclusions and future works are given in Section VI.

II. Overview of the Framework

The proposed framework for personalized authentication is illustrated in Fig.1(b). Firstly, for each input original image, two training sets consisting of attacked samples generated through content-changing and content-preserving operations are constructed automatically. Secondly, elaborated feature representations of all attacked samples can be obtained by existing feature extraction methods. Thirdly, a novel distance metric learning algorithm, termed weighted large margin for manipulation classification (WLMMC), is proposed to learn an effective mapping space from the two training sets in which content-changing and content-preserving samples are mapped farther apart. Finally, a novel supervised personalized quantization strategy (SPQ) is proposed to learn compact binary codes based on the statistical discrimination distribution of all dimensions.

Compared with traditional hashing based image authentication algorithms as illustrated in Fig.1(a), the differences and advantages of the proposed personalized authentication framework are mainly embodied in the distance metric learning and quantization procedures. The proposed framework dedicates to learn personalized and compact hash codes through learning a distinct metric matrix and determining a personalized bit allocation strategy from the constructed training sets for each image. It should be pointed out, however, that this paper does not focus on the feature extraction procedure.

III. Distance Metric Learning for Manipulation Classification

The tamper detection task is reformulated as a new supervised manipulation classification problem in this paper. And, a novel distance metric learning algorithm named weighted large margin for manipulation classification (WLMMC) , as illustrated in Fig.2, is introduced in this section, which is inspired by recent works on distance metric learning for large margin nearest neighbor classification (LMNN) [22], [27]–[29].

A. Terminology and Intuition

Given an original image I_o with its feature vector $\boldsymbol{v}_o$, let $S^p = \{(\boldsymbol{v}_i\}_{i=1}^n$ and $S^c = \{(\boldsymbol{v}_j)\}_{j=1}^m$ denote the training set of n labeled content-preserving samples and the training set of m labeled content-changing samples of I_o respectively, where $\boldsymbol{v}_o, \boldsymbol{v}_i, \boldsymbol{v}_j \in \mathbb{R}^d$. The proposed WLMMC seeks to learn a linear transformation $\boldsymbol{L}:\mathbb{R}^d \rightarrow \mathbb{R}^d$, which is used to compute squared Mahalanobis distances as

$$\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)=\parallel \boldsymbol{L}(\boldsymbol{v}_i-\boldsymbol{v}_o)\parallel_2^2=(\boldsymbol{v}_i-\boldsymbol{v}_o)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{v}_i-\boldsymbol{v}_o) \quad (1)$$

$$\mathcal{D}^c(\boldsymbol{v}_j,\boldsymbol{v}_o)=\parallel \boldsymbol{L}(\boldsymbol{v}_j-\boldsymbol{v}_o)\parallel_2^2=(\boldsymbol{v}_j-\boldsymbol{v}_o)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{v}_j-\boldsymbol{v}_o) \quad (2)$$

where $\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)$ denotes the squared Mahalanobis distances between the original image $\boldsymbol{v}_o$ and its content-preserving samples $\boldsymbol{v}_i$, $\mathcal{D}^c(\boldsymbol{v}_j,\boldsymbol{v}_o)$ denotes the squared Mahalanobis distances between the original image $\boldsymbol{v}_o$ and its content-changing samples $\boldsymbol{v}_j$, $\boldsymbol{M}=\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}$ is a symmetric positive definite matrix ($\boldsymbol{M}\geq 0$).

The intuition of WLMMC is that, for an original image I_o, its manipulated versions with different labels should be widely separated so that it's easy to distinguish content-preserving samples from its content-changing samples. The proposed WLMMC aims at learning a Mahalanobis distance metric that keeps content-preserving samples closer to its original image than all content-changing samples by the distance Θ, i.e. large margin in the new metric space. And, the distance Θ, which depends on the adopted feature descriptor, is used to set the scale for the linear transformation.

$$\mathcal{D}^c(\boldsymbol{v}_j,\boldsymbol{v}_o)-\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)\geq\Theta \quad (3)$$

More precisely, the distance $\mathcal{D}(\boldsymbol{v}_r,\boldsymbol{v}_o)$ between the received image I_r and its original image I_o should be greater than $\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)$ if I_r is inauthentic. Otherwise, $\mathcal{D}(\boldsymbol{v}_r,\boldsymbol{v}_o)$ is expected to be no more than $\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)$. Fig.2 shows the idealized scenario where manipulation classification errors in the original feature space are corrected by learning an appropriate metric space.

B. Loss Function

Ideally, content-preserving samples should be closer to its original image than all content-changing samples in the original feature space. However, due to the limited performance of existing feature extraction methods, some content-changing samples and content-preserving samples are always neighboring in the original feature space, which makes it difficult to distinguish them. And, this can also be demonstrated by the inevitable misjudgments of existing works [4]–[6], [8], [10]. Therefore, how to widen the distance between neighboring samples with different labels is the key to reducing misjudgments.

In order to improve the manipulation classification performance, a WLMMC method that minimizes the following objective function is proposed. The loss function consists of two terms, one which aims to penalize small distances between the original image and its content-changing samples by employing the geometry information of the samples in the feature space, and another which is a regularizer on distance metric $\boldsymbol{M}$.

$$\min_{\boldsymbol{M}\geq 0}\sum_{\boldsymbol{v}_i\in S^p} w_i[\Theta-(\mathcal{D}^c(\boldsymbol{v}_{j_k},\boldsymbol{v}_o)-\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o))]_+ +\lambda\|\boldsymbol{M}\|_F^2 \quad (4)$$

where $[z]_+ = max(z,0)$ denotes the standard hinge loss which monitors the inequality in Eq.(3), $\|\boldsymbol{M}\|_F^2$ is the Frobenius norm of metric $\boldsymbol{M}$ and stands for the regularizer on the expected output, λ is a non-negative coefficient balancing the two involved terms, and the coefficient w_i is introduced to adaptively adjust the penalty weight of the hinge loss caused by the invading triples defined in Eq.(3). More precisely, the closer the distance between $\boldsymbol{v}_{j_k}$ and $\boldsymbol{v}_o$, the bigger the weight w_i. Specifically, w_i is defined as:

$$w_i=\frac{\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)-\mathcal{D}^c(\boldsymbol{v}_{j_k},\boldsymbol{v}_o)+\Theta}{\sum\limits_{\boldsymbol{v}_i\in S^p}(\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)-\mathcal{D}^c(\boldsymbol{v}_{j_k},\boldsymbol{v}_o)+\Theta)} \quad (5)$$

where

$$j_k=\operatorname*{argmin}_{\boldsymbol{v}_j\in S^c}\mathcal{D}^c(\boldsymbol{v}_j,\boldsymbol{v}_o) \quad (6)$$

and

$$\sum_{i=1}^{n} w_i = 1 \quad (7)$$

That is to say, for the original image $\boldsymbol{v}_o$ and its content-preserving samples $\boldsymbol{v}_i$, if the nearest content-changing sample $\boldsymbol{v}_{j_k}$ satisfies the inequality in Eq.(3), all other content-changing samples would follow such inequality relation too.

C. Convex Optimization

To improve the computational efficiency, the optimization of Eq. (4) can be reformulated as an instance of semi-definite programming (SDP). A SDP problem is a linear program that incorporates an additional constraint on a symmetric matrix whose elements are linear in the unknown variables. According to [27], [29], by introducing slack variables ξ_i to simplify the hinge loss in Eq.(4), the resulting SDP is given by:

$$\begin{aligned}&\textbf{Minimize}\sum_{\boldsymbol{v}_i\in S^p}\xi_i+\lambda\|\boldsymbol{M}\|_F^2 \quad \textbf{subject to}:\\ &\mathcal{D}^c(\boldsymbol{v}_{j_k},\boldsymbol{v}_o)-\mathcal{D}^p(\boldsymbol{v}_i,\boldsymbol{v}_o)\geq\Theta-\xi_i,\\ &\forall\boldsymbol{v}_i\in S^p,\forall\boldsymbol{v}_j\in S^c,\xi_i\geq 0,\\ &\boldsymbol{M}\succeq 0.\end{aligned} \quad (8)$$

where $\boldsymbol{M}\succeq 0$ indicates that matrix $\boldsymbol{M}$ is required to be positive semi-definite, and the linear transformation matrix $\boldsymbol{L}$ can be calculated by matrix decomposition of $\boldsymbol{M}$.

The proposed WLMMC method can be solved based on the sub-gradient descent method [29]. Let $\boldsymbol{v}_{j,o}=(\boldsymbol{v}_j-\boldsymbol{v}_o)(\boldsymbol{v}_j-\boldsymbol{v}_o)^{\mathrm{T}}$, $\boldsymbol{v}_{i,o}=(\boldsymbol{v}_i-\boldsymbol{v}_o)(\boldsymbol{v}_i-\boldsymbol{v}_o)^{\mathrm{T}}$, the squared WLMMC distance corresponding to $\boldsymbol{M}_t$ generated in the t-th iteration can be defined as:

$$\mathcal{D}_t^c(\boldsymbol{v}_j,\boldsymbol{v}_o)=Tr(\boldsymbol{M}_t\boldsymbol{v}_{j,o}) \quad (9)$$

$$\mathcal{D}_t^p(\boldsymbol{v}_i,\boldsymbol{v}_o)=Tr(\boldsymbol{M}_t\boldsymbol{v}_{i,o}) \quad (10)$$

where $Tr(\boldsymbol{X})$ is the trace of matrix $\boldsymbol{X}$. Therefore, the objective function of Eq. (4) can be rewritten as follows:

$$\Gamma(\boldsymbol{M}_t)=\sum_{\boldsymbol{v}_i\in S^p} w_i^t[\Theta-(Tr(\boldsymbol{M}_t\boldsymbol{v}_{j_k,o})-Tr(\boldsymbol{M}_t\boldsymbol{v}_{i,o}))]_+ +\lambda\|\boldsymbol{M}_t\|_F^2 \quad (11)$$

where w_i^t is the penalty weight at iteration t. The gradient of $\Gamma(\boldsymbol{M}_t)$ respect to $\boldsymbol{M}$ at iteration t is described as follow:

$$\frac{\partial\Gamma(\boldsymbol{M}_t)}{\partial\boldsymbol{M}_t} = \sum_{\boldsymbol{v}_i\in S^p}[\xi_{ij_k}]_+(\boldsymbol{v}_{i,o}-\boldsymbol{v}_{j_k,o}) + 2\lambda\boldsymbol{M}_t$$
$$\xi_{ij_k} = w_i^t(\Theta - (Tr(\boldsymbol{M}_t\boldsymbol{v}_{j_k,o}) - Tr(\boldsymbol{M}_t\boldsymbol{v}_{i,o}))) \tag{12}$$

where $[\xi_{ij_k}]_+ = 1$ if $\xi_{ij_k} > 0$, else $[\xi_{ij_k}]_+ = 0$.

The details of the gradient method are presented in Algorithm 1.

Algorithm 1: Weighted Large Margin for Manipulation Classification (WLMMC)

Input: Data sets $S^p = \{(\boldsymbol{v}_i, y_i)\}_{i=1}^n$, $S^c = \{(\boldsymbol{v}_j, y_j)\}_{j=1}^m$, λ, Θ, step parameter ζ.
Output: $\boldsymbol{L}$, $\boldsymbol{M}$.
Initialization: $\boldsymbol{M}_0$;
repeat
  Compute the gradient $\frac{\partial\Gamma(\boldsymbol{M}_t)}{\partial\boldsymbol{M}_t}$ by Eq. (12) based on $\boldsymbol{M}_t$;
  Calculate $\tilde{\boldsymbol{M}}_{t+1} = \boldsymbol{M}_t + \zeta\frac{\partial\Gamma(\boldsymbol{M}_t)}{\partial\boldsymbol{M}_t}$;
  Do eigenvalue decomposition on $\tilde{\boldsymbol{M}}_{t+1}$ to obtain $\boldsymbol{U}$ and Σ_+, where $\boldsymbol{U}$ is a orthogonal unit matrix which makes $\Sigma = \boldsymbol{U}^{\mathrm{T}}\tilde{\boldsymbol{M}}_{t+1}\boldsymbol{U}$ diagonal, and $\Sigma_+ = abs(\Sigma)$;
  $\boldsymbol{M}_{t+1} = \boldsymbol{U}\Sigma_+\boldsymbol{U}^{\mathrm{T}}$, $\boldsymbol{L}_{t+1} = (\boldsymbol{U}sqrt(\Sigma_+))^{\mathrm{T}}$;
until *Convergence*;

IV. Supervised Personalized Quantization

A novel supervised personalized quantization strategy (SPQ) is proposed to learn binary codes for the original image in this section, as illustrated in Fig.3. The proposed SPQ algorithm is motivated by the observation that different attacks have various degrees of influence on feature components of different images even with the same feature representation method.

Given $\tilde{\boldsymbol{v}}_o$, $\tilde{\boldsymbol{v}}_i$, $\tilde{\boldsymbol{v}}_j \in \mathbb{R}^d$ be the feature vectors of $\boldsymbol{v}_o$, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ in the transformed metric space, respectively, the goal is to learn compact, yet discriminative binary hash codes that encode the the original image I_o into L-length hash codes $H \in \{0,1\}^L$, which show sensitivity to content-changing attacks and robustness against content-preserving operations. The proposed SPQ procedure can be divided into two phases: supervised dimension selection and bit allocation as well as supervised quantization.

A. Supervised Dimension Selection

Firstly, normalized distance histograms are created to describe the overall intensity distribution of distances between the original image and its manipulated samples in each dimension l $(1 \le l \le d)$. Let $d_{io}^l = |\tilde{\boldsymbol{v}}_i(l) - \tilde{\boldsymbol{v}}_o(l)|$ denote the distance between $\tilde{\boldsymbol{v}}_i$ and $\tilde{\boldsymbol{v}}_o$ in the l-th dimension, as well as $d_{jo}^l = |\tilde{\boldsymbol{v}}_j(l) - \tilde{\boldsymbol{v}}_o(l)|$ be the distance between $\tilde{\boldsymbol{v}}_j$ and $\tilde{\boldsymbol{v}}_o$ in the l-th dimension too. Given the content-preserving sample set $S^p(\tilde{\boldsymbol{v}}_i \in S^p)$ and content-changing sample set $S^c(\tilde{\boldsymbol{v}}_j \in S^c)$,

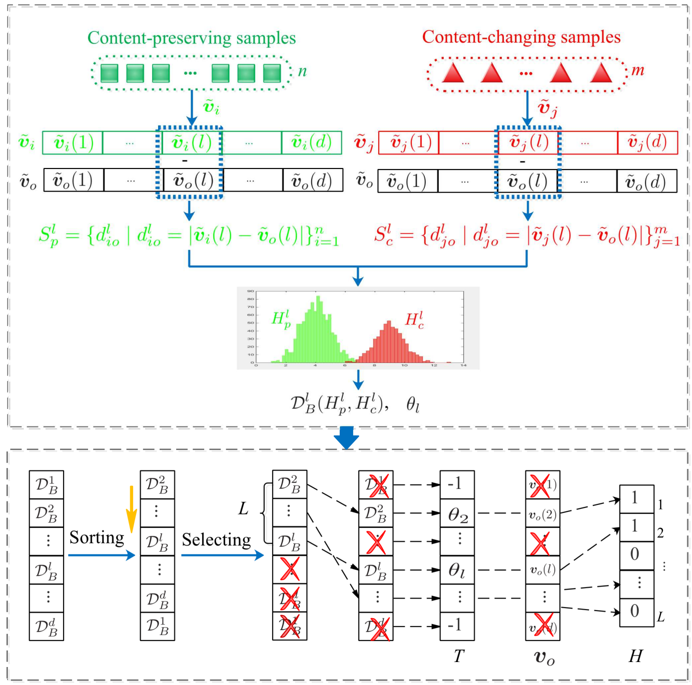


Fig. 3. Overview of the supervised personalized quantization strategy.

two distance sets $S_p^l = \{d_{io}^l\}_{i=1}^n$ and $S_c^l = \{d_{jo}^l\}_{j=1}^m$ can be calculated for each dimension l. Then, two normalized distance histograms H_p^l and H_c^l over the same domain X^l are generated, respectively, where the x-axis shows the distance values ranging from d_{min}^l to d_{max}^l, the y-axis corresponds to normalized count of distance values. And, the number of subdivisions in x-axis is set to n_b.

$$d_{min}^l = \min(\min(S_p^l), \min(S_c^l)) \tag{13}$$

$$d_{max}^l = \max(\max(S_p^l), \max(S_c^l)) \tag{14}$$

$$X^l(k) = \lfloor d_{min}^l \rfloor + k \times \frac{\lceil d_{max}^l \rceil - \lfloor d_{min}^l \rfloor}{n_b}, k \in [0, n_b] \tag{15}$$

Secondly, the Bhattacharyya distance $\mathcal{D}_B^l$ is employed to measure the histogram distance between H_c^l and H_p^l of the l-th dimension:

$$\mathcal{D}_B^l(H_p^l, H_c^l) = -\ln(BC(H_p^l, H_c^l)) \tag{16}$$

where

$$BC(H_p^l, H_c^l) = \sum_{x\in X^l}\sqrt{H_p^l(x)H_c^l(x)} \tag{17}$$

is the Bhattacharyya coefficient for discrete probability distributions, $0 \le BC \le 1$, $0 \le \mathcal{D}_B^l \le \infty$. It should be noted that, in essence, the histogram distance $\mathcal{D}_B^l$ also indicates the classification performance between content-preserving samples and content-changing samples in each dimension l. The greater the $\mathcal{D}_B^l$, the better the classification performance.

Finally, $L(1 \le L \le d)$ discriminative dimensions are selected according to the Bhattacharyya distance of each dimension via a quantization template $T \in \mathbb{R}^d$. The hash length

L is determined by the compression rate δ which indicates the percentage of dimensions that need to be reserved.

$$L = \lfloor d \times \delta \rfloor \tag{18}$$

All dimensions are first arranged in descending order according to their Bhattacharyya distances. Then, the top L dimensions are chosen as required and the quantization template $T(i)$ is set to be 1 if the i-th dimension is selected, otherwise, $T(i)$ is set to -1. One bit is allocated to each reserved dimension in this paper.

With the supervised dimension selection strategy discussed above, the final hash length L can vary from 1 to d on demand, since the proposed method is capable of allocating zero bit to undiscriminating dimensions.

B. Supervised Quantization

L selected dimensions are quantized into a L-length binary string H by thresholding. Firstly, for the l-th dimension of $\tilde{\boldsymbol{v}}_o$, a threshold θ_l is defined if this dimension is selected ($T(l) = 1$):

$$\theta_l = \operatorname*{argmin}_{\theta_l \in X^l} \left(\sum_{\tilde{\boldsymbol{v}}_i \in S^p} f(\tilde{\boldsymbol{v}}_i(l), \theta_l) + \sum_{\tilde{\boldsymbol{v}}_j \in S^c} f(\theta_l, \tilde{\boldsymbol{v}}_j(l)) \right) \tag{19}$$

where

$$f(a,b) = \begin{cases} 1, & a > b \\ 0, & a \le b \end{cases} \tag{20}$$

Then, $T(l)$ is updated by θ_l. After that, the real-value of $\tilde{\boldsymbol{v}}_o(l)$ is quantized by thresholding. More specifically, $h(l) = 1$ if $\tilde{\boldsymbol{v}}_o(l) \ge \theta_l$. Otherwise, $h(l) = -1$. Consequently, L bits are collected and form the final hash codes H.

In summary, the proposed SPQ approach is summarized in Algorithm 2.

V. Experimental Results And Performance Analysis

In this section, the performance of proposed WLMMC is evaluated firstly. Then, performance comparisons with regard to robustness, sensitivity and discrimination between traditional image hashing algorithms and the proposed personalized framework are conducted.

It must be emphasized again that the proposed personalized framework is designed to combine with feature extraction methods of existing image hashing schemes, and further enhance their authentication performance. The feature extraction method used in the experiments are the Ring Partition (RP) based descriptor [6], and the Tensor Decomposition (TD) based descriptor [10]. The total dimensions of feature vectors of RP based descriptor and TD based descriptor are 160 and 96, respectively. Parameters for WLMMC are $\Theta = 10$, $\zeta = 10^{-9}$ for the RP based descriptor; $\Theta = 0.1$, $\zeta = 10^{-6}$ for the TD based descriptor; and $\lambda = 0.1$. The source codes will be publicly available at: https://zhiyongsu.github.io.

Algorithm 2: Supervised Personalized Quantization (SPQ)

Input: Original image $\tilde{\boldsymbol{v}}_o$, content-preserving samples $\{\tilde{\boldsymbol{v}}_i\}_{i=1}^n$, content-changing samples $\{\tilde{\boldsymbol{v}}_j\}_{j=1}^m$, compression rate δ, $\tilde{\boldsymbol{v}}_o$, $\tilde{\boldsymbol{v}}_i$, $\tilde{\boldsymbol{v}}_j \in \mathbb{R}^d$.

Output: Quantization template $T \in \mathbb{R}^d$, L-length hash codes H.

for $l = 1$ *to* d **do**
  Calculate two distance sets $S_p^l = \{d_{io}^l\}_{i=1}^n$ and $S_c^l = \{d_{jo}^l\}_{j=1}^m$ in the l-th dimension of all samples, respectively, where $d_{io}^l = |\tilde{\boldsymbol{v}}_i(l) - \tilde{\boldsymbol{v}}_o(l)|$, $d_{jo}^l = |\tilde{\boldsymbol{v}}_j(l) - \tilde{\boldsymbol{v}}_o(l)|$;
  Calculate two normalized distance histograms H_p^l and H_c^l over the same domain with respect to S_p^l and S_c^l, respectively;
  Calculate the Bhattacharyya distance $\mathcal{D}_B^l$ between H_p^l and H_c^l.
end
Arrange all d dimensions in descending order according to their Bhattacharyya distances;
Select the top $L = \lfloor d \times \delta \rfloor$ dimensions on the basis of the compression rate δ;
Calculate the threshold θ_l if the l-th dimension is selected, and let $T(l) = \theta_l$, otherwise $T(l) = -1$;
Get final hash codes H of the original image $\tilde{\boldsymbol{v}}_o$ by thresholding according to the quantization template T.

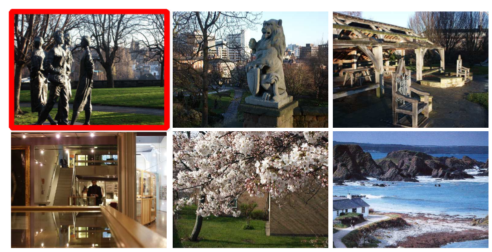


Fig. 4. Some typical images of UCID. The first image with red boxes is selected as the benchmark image in the experiment.

A. Dataset

To evaluate the performance of the proposed framework, 100 color images of various sizes from the UCID [30] image database are selected and tested in experiments. Some typical images are shown in Fig.4. And, for the space limitation, the first image is selected as the benchmark image in this paper. The normalized image size is 512×512.

For each original image, two kinds of datasets, named training dataset and test dataset, are automatically constructed, respectively. The training dataset is employed to learn the distance metric matrix $\boldsymbol{L}$ and the quantization template T for each original image. The test dataset is used for the performance evaluation. Each kind of datasets consists of two subsets : content-preserving samples and content-changing samples. The adopted content-changing and content-preserving operations as well as their parameter values are given in Table I and Table II, respectively. Note that the training dataset and test dataset are generated exactly in the same way but with totally different parameter values. In order to enhance

TABLE I
TYPICAL CONTENT-PRESERVING MANIPULATIONS AND PARAMETER VALUES

Operations	Description	Parameters	Number of Training Samples	Number of Test Samples
Rotation	Rotation angle	$1^0 \sim 360^0$	18	20
Scaling	Ratio	$0.1 \sim 0.99$, $1 \sim 3$	15	10
Gaussian Noise	Variance	$0.01 \sim 0.15$	15	10
Pepper and Salt Noise	Density	$0.01 \sim 0.05$	5	5
Speckle Noise	Density	$0.01 \sim 0.2$	20	24
Random Noise	Density	$0.01 \sim 0.12$	12	10
Gaussian Filter	Size,Sigma	Size:$3 \sim 12$, Sigma:0.5	8	9
Unsharp Filter	Alpha	$0.1 \sim 0.5$	5	7
Motion Filter	Len,Theta	Len:$1 \sim 5$, Theta:0	5	5
JPEG Compression	Quality Factor	$40 \sim 100$	5	11
Total			108	112

TABLE II
TYPICAL CONTENT-CHANGING MANIPULATIONS AND PARAMETER VALUES

Operations	Description	Parameters	Number of Training Samples	Number of Test Samples
Adding Elements			4	8
Copy-Move Attack			4	8
Cropping	Ratio	$5\% \sim 60\%$	28	20
Mosaic Attack			24	20
Total			60	56

the classification performance, more content-changing and content-preserving operations are encouraged to be involved and employed to enrich training samples in practice. Thus, for each original image, 112 content-preserving samples and 56 content-changing samples are produced in the test dataset, respectively.

It should be noticed that it is impossible to cover all kinds of image operations, especially content-changing manipulations, in the training and learning procedure, as the means and approaches for image content tampering are various and diversified. However, this paper focus on improving the authentication performance of existing algorithms based on their own feature extraction methods. Therefore, like existing image authentication algorithms [4]–[6], [8], [10], a limited number of content-changing and content-preserving operations discussed above are employed to generate training samples in this paper. And, the superiority of the personalized framework will be proved through experiments under the same conditions, i. e., test images and image operations.

B. Performance Criteria

- The normalized hamming distance (NHD) is taken to measure similarity of two image hashes in this paper.

$$d_H(H_1, H_2) = \frac{1}{L}\sum_{l=1}^{L}|h_1(l)_-h_2(l)| \tag{21}$$

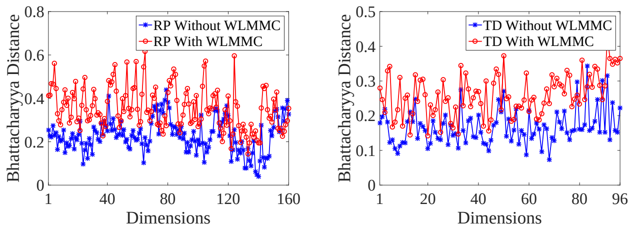
  where $h_1(l)$ and $h_2(l)$ are the l-th elements of H_1 and H_2, respectively.
- The normalized distance histogram is adopted to describe the overall intensity distribution of distances (e.g., Euclidean distance, and NHD), as introduced in Section IV-A.
- The Bhattacharyya distance is then employed to evaluate the classification performance through measuring the distance between two normalized distance histograms. The greater the Bhattacharyya distance, the better the classification performance.

Note that performance evaluations of image hashing schemes are always conducted under different thresholds [1]–[8], [10]. Given a specified distance metric (e.g., NHD), two input images are judged as visually identical images if their distance is smaller than a given threshold. Otherwise, they are different images or one is a tampered version of the other. Therefore, experiments will focus on illustrating how the proposed framework facilitates the choice of thresholds and further alleviates the misjudgment as well as enhances the authentication performance.

C. WLMMC Performance Analysis

The WLMMC is designed to enhance the manipulation classification performance of existing feature extraction methods in the learned metric space. The RP based descriptor and TD based descriptor are employed as the feature extraction methods in this section. For each descriptor, the distance metric matrix $\boldsymbol{L}$ of each original image is firstly learned through the WLMMC scheme based on the training dataset. Then, for each original image, RP based descriptors (RP without WLMMC) and TD based descriptors (TD without WLMMC) of all its attacked samples in the test dataset are extracted in their original feature space. After that, the distance metric matrix, which is learned from the training dataset, is employed to generate the corresponding transformed feature descriptors, named RP with WLMMC and TD with WLMMC, in the metric space for each sample, respectively. The overall classification performance of WLMMC is evaluated from two aspects of dimensions and samples.

1) Classification Performance on Dimensions: To evaluate the classification performance on each dimension l, for each original image, the Euclidean distance d_{io}^l between $\tilde{\boldsymbol{v}}_i$ and $\tilde{\boldsymbol{v}}_o$ in the l-th dimension of all its content-preserving samples, as well as the Euclidean distance d_{jo}^l between $\tilde{\boldsymbol{v}}_j$ and $\tilde{\boldsymbol{v}}_o$



(a) (b)

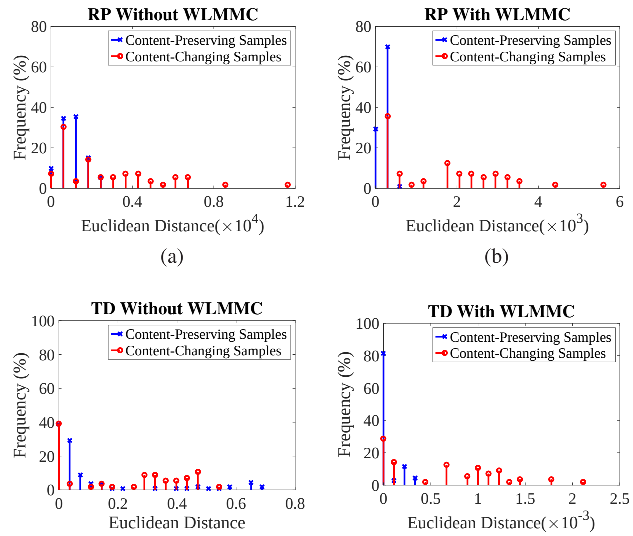
Fig. 5. Comparisons of Bhattacharyya distances of each dimension of the benchmark image under different feature representations with and without the WLMMC algorithm. The greater the Bhattacharyya distance, the better the classification performance between content-changing samples and content-preserving samples on each dimension. (a) RP based descriptor. (b) TD based descriptor.

in the l-th dimension of all its content-changing samples are calculated in the metric space, respectively. Then, two distance histograms H_p^l and H_c^l are created to represent the distance distribution of $\{d_{io}^l\}_{i=1}^n$ and $\{d_{jo}^l\}_{j=1}^m$, respectively. Finally, the Bhattacharyya distance is employed to measure the distance between H_c^l and H_p^l of each dimension l in the transformed metric space. For comparison, the Bhattacharyya distance of each dimension l in the original feature space is also calculated according to the above method.

For the space limitation, Fig.5 shows the comparisons of Bhattacharyya distances of each dimension of the benchmark image under different feature representations. The x-axis is the index of feature dimensions, and the y-axis is the Bhattacharyya distance. It can be seen from the results that Bhattacharyya distances of almost all of the dimensions become larger because of the introducing of the WLMMC scheme. As discussed in Section IV-A, the greater the Bhattacharyya distance, the better the classification performance between content-changing samples and content-preserving samples. Therefore, the proposed WLMMC algorithm can significantly improve the manipulation classification performance of most of the dimensions in the learned metric space.

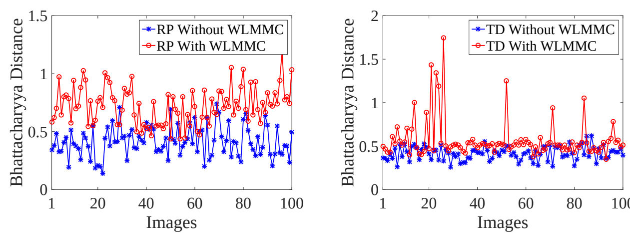
2) Classification Performance on Samples: To validate the classification performance on samples, for each original image $\tilde{\boldsymbol{v}}_o$, the Euclidean distances $D_e^p(\tilde{\boldsymbol{v}}_i, \tilde{\boldsymbol{v}}_o)$ between $\tilde{\boldsymbol{v}}_i$ and $\tilde{\boldsymbol{v}}_o$, as well as the Euclidean distance $D_e^c(\tilde{\boldsymbol{v}}_j, \tilde{\boldsymbol{v}}_o)$ between $\tilde{\boldsymbol{v}}_j$ and $\tilde{\boldsymbol{v}}_o$ are computed firstly in the transformed metric space, respectively. Then, two distance histograms H_p and H_c are created to represent the distance distribution of $\{D_e^p(\tilde{\boldsymbol{v}}_i, \tilde{\boldsymbol{v}}_o)\}_{i=1}^n$ and $\{D_e^c(\tilde{\boldsymbol{v}}_j, \tilde{\boldsymbol{v}}_o)\}_{j=1}^m$, respectively. Finally, the Bhattacharyya distance between H_p and H_c is derived. For performance comparison, the Bhattacharyya distance of each original image in the original feature space is also calculated without the using of WLMMC algorithm.

Fig.6 presents the distribution of Euclidean distance intensities of all content-changing samples and content-preserving samples of the benchmark image with and without the WLMMC algorithm respectively, where the x-axis is the Euclidean distance, and the y-axis is the frequency. As revealed in many previous works [7], [8], [10], the overlapping regions between the distance distribution of content-preserving samples and content-changing samples (or different images)



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

Fig. 6. Comparisons of Euclidean distances distribution of all content-changing samples and content-preserving samples of the benchmark image based on different feature extraction methods. The number of samples in the overlapping regions decreases greatly because of the involving of WLMMC. (a) RP without WLMMC. (b) RP with WLMMC. (c) TD without WLMMC. (d) TD with WLMMC.



(a) (b)

Fig. 7. Comparisons of Bhattacharyya distances of all test images. The Bhattacharyya distance of each test image reflects the discriminabiltity between its content-changing samples and content-preserving samples. (a) RP based descriptor. (b) TD based descriptor.

always exist for every feature extraction (or hashing) method. Furthermore, the samples in the overlapping interval will be inevitably misclassified. However, it is observed that the number of samples in the overlapping regions decreases greatly because of the involving of WLMMC, as shown in Fig.6.

Fig.7 shows the comparisons of Bhattacharyya distances of all test images with and without the WLMMC algorithm, where the x-axis is the index of the test images, and the y-axis is the Bhattacharyya distance. For each test image, its Bhattacharyya distance reflects the discriminabiltity between its content-changing samples and content-preserving samples. It can be seen that, Bhattacharyya distances of almost all of the test images increase in different extent, which indicates that the proposed WLMMC algorithm can significantly improve the classification performance between content-preserving samples and content-changing samples for each original image in the feature space.

D. Performance Comparisons

In essence, the authentication performance of image hashing schemes mainly depends on its feature extraction and representation method. The superiority of the proposed framework mainly embodies in improving the classification performance of existing image hashing algorithms on the basis of their own feature extraction methods. For the space limitation, the proposed personalized framework is compared with the state-of-the-art TD hashing algorithm to show advantages [10]. The parameter settings of the compared hashing algorithm are the same with those reported in the original paper. WS-TD hashing in the experiments refers to the proposed framework which couples with the TD based descriptor [10]. The distance metric matrix $\boldsymbol{L}$ and quantization template T of each original image is firstly learned through the WS-TD hashing based on the training dataset. Then, performance comparisons are carried out on the test dataset.

1) Robustness and Sensitivity: $100 \times 112 = 11200$ pairs of content-preserving samples are used for robustness validation, and $100 \times 56 = 5600$ pairs of content-changing samples are used for sensitivity evaluation. Image hashes of each pair of content-preserving samples are extracted by TD hashing and WS-TD hashing, respectively. The WS-TD hashing algorithm adopts four different compression rates $(\delta = 1.0, 0.75, 0.5, 0.25)$ to generate hash codes, respectively. Specifically, for each original image $\tilde{\boldsymbol{v}}_o$, its binary hash codes H, distance metric matrix $\boldsymbol{L}$, and quantization template T are firstly generated through the proposed WLMMC and SPQ schemes based on the training dataset. Then, the binary hash codes of all its attacked samples in the test dataset are calculated with the same distance metric matrix $\boldsymbol{L}$ and quantization template T. After that, the NHD $D_h^p(\tilde{\boldsymbol{v}}_i, \tilde{\boldsymbol{v}}_o)$ between the test image $\tilde{\boldsymbol{v}}_o$ and its content-preserving samples $\tilde{\boldsymbol{v}}_i$, as well as the NHD $D_h^c(\tilde{\boldsymbol{v}}_j, \tilde{\boldsymbol{v}}_o)$ between the test image $\tilde{\boldsymbol{v}}_o$ and its content-changing samples$\tilde{\boldsymbol{v}}_j$ are calculated, respectively. Two distance histograms H_p and H_c are created to represent the distribution of the distance intensities of $\{D_h^p(\tilde{\boldsymbol{v}}_i, \tilde{\boldsymbol{v}}_o)\}_{i=1}^n$ and $\{D_h^c(\tilde{\boldsymbol{v}}_j, \tilde{\boldsymbol{v}}_o)\}_{j=1}^m$, respectively. Finally, the Bhattacharyya distance between H_p and H_c is calculated.

Fig.8 shows the NHDs distribution of all content-changing samples and content-preserving samples of the benchmark image with three different compression rates. Fig.9 also presents the NHDs distribution of all content-changing samples and content-preserving samples of all test images. Ideally, the NHDs between the test image and its content-preserving samples should be close to 0, while the NHDs between the test image and its content-changing samples should be close to 1. Moreover, the smaller the number of samples in the overlapping regions between the distance distribution of content-preserving samples and content-changing samples, the better the classification performance. It is observed that the overlapping regions between the NHDs distribution of content-preserving images and content-changing images always exist for all hashing algorithms. However, the number of samples in the overlapping interval of WS-TD hashing is smaller than those in the overlapping intervals of TD hashing. The most important thing is that the number of content-changing samples whose NHDs are close to 0, as well as the number of content-preserving samples whose NHDs are close to 1, significantly decrease as shown in Fig.9 (b)-(d). Meanwhile, the proposed framework can yield better classification performance under higher compression rates.

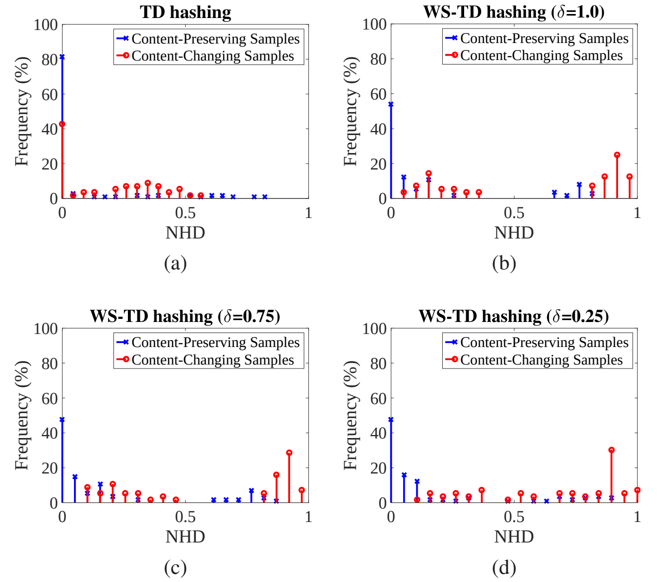


Fig. 8. Comparisons of NHDs distribution of all content-changing samples and content-preserving samples of the benchmark image with different compression rates. (a) TD hashing. (b) WS-TD hashing ($\delta = 1.0$). (c) WS-TD hashing ($\delta = 0.75$). (d) WS-TD hashing ($\delta = 0.25$).

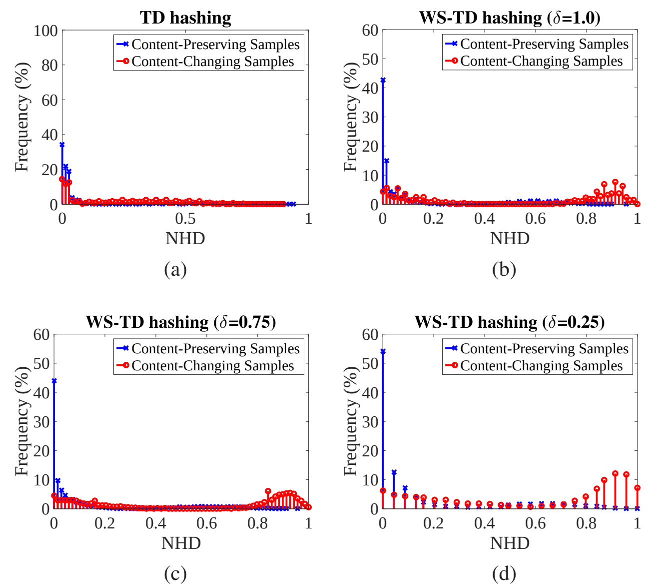


Fig. 9. Comparisons of NHDs distribution of content-changing samples and content-preserving samples of all test images with different compression rates. (a) TD hashing. (b) WS-TD hashing ($\delta = 1.0$). (c) WS-TD hashing ($\delta = 0.75$). (d) WS-TD hashing ($\delta = 0.25$)

Fig.10 gives comparisons of Bhattacharyya distances of NHDs of all test images under different compression rates δ, where the x-axis is the index of test images, and the y-axis is Bhattacharyya distances. For each test image, its Bhattacharyya distance measures the distance between its

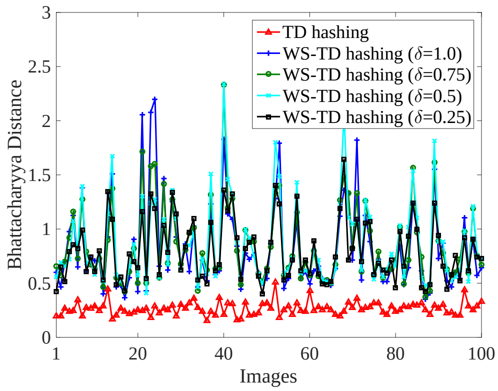


Fig. 10. Comparisons of Bhattacharyya distances of NHDs of all test images under different compression rates.

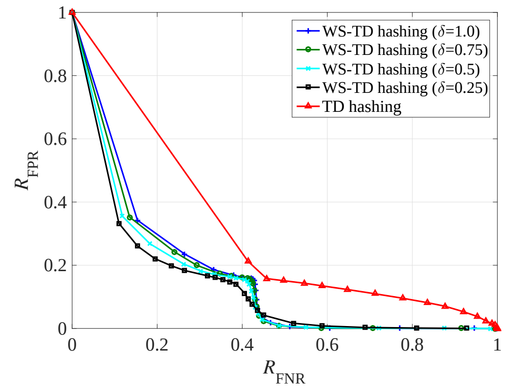


Fig. 11. ROC curve comparisons among different hashing schemes.

content-changing samples' NHDs histogram and its content-preserving samples' NHDs histogram. And, it also reflects the discriminabiltity between content-changing samples and content-preserving samples. It is observed that the Bhattacharyya distances of all test images increase in the hash space, which demonstrates that the proposed WLMMC and SPQ algorithms can significantly improve the classification performance between content-preserving and content-changing samples. Furthermore, it can be seen from the results that the proposed SPQ algorithm can get higher compression rates while yield a comparable classification performance.

The receiver operating characteristics (ROC) graph is also employed to make visual classification comparisons with respect to robustness and sensitivity. The false negative rate (FN) R_{FNR} and false positive rate (FP) R_{FPR} are defined as follows [4]:

$$R_{\text{FNR}} = \frac{N_{\text{tampAuthentic}}}{N_{\text{tampered}}} \tag{22}$$

$$R_{\text{FPR}} = \frac{N_{\text{authenticTamp}}}{N_{\text{identical}}} \tag{23}$$

where $N_{\text{tampAuthentic}}$ is the number of tampered images detected as authentic, N_{tampered} is the total number of tampered images, $N_{\text{authenticTamp}}$ is the number of authentic images detected as tampered, and $N_{\text{identical}}$ is the total number of pairs of visually identical images.

Fig.11 presents the ROC curve comparisons between the state-of-the-art TD hashing and WS-TD hashing algorithms with different compression rates. It is observed that the ROC curves of the WS-TD hashing schemes are all much closer to the left-bottom corner than the TD hashing. This means that the WS-TD hashing is superior to the compared TD hashing in classification between robustness and sensitivity.

2) Discrimination: $100 \times (100 - 1)/2 = 4950$ pairs of different images are used for discrimination testing. For each original image, also named the reference image, its hash codes, the distance metric matrix $\boldsymbol{L}$, and four quantization templates T are firstly generated through the WS-TD hashing under four different compression rates $(\delta = 1.0, 0.75, 0.5, 0.25)$ based on the training dataset. Then, the hash codes of all of the other 99 test images are generated through the same $\boldsymbol{L}$ and T of the reference image. Finlay, the NHDs between each reference image and the other 99 test images are calculated. Consequently, $100 \times (100-1)/2 = 4950$ NHDs are generated for each compression rate.

TABLE III

STATISTICS OF NHDS OF TD HASHING AND WS-TD HASHING UNDER DIFFERENT COMPRESSION RATES

Algorithm	Min	Max	Mean	Standard deviation
TD hashing	0.0417	0.8958	0.4596	0.1590
WS-TD hashing ($\delta = 1.00$)	0.6458	1	0.9110	0.0463
WS-TD hashing ($\delta = 0.75$)	0.5972	1	0.9175	0.0482
WS-TD hashing ($\delta = 0.50$)	0.4792	1	0.9191	0.0548
WS-TD hashing ($\delta = 0.25$)	0.4167	1	0.9137	0.0706

Fig.12 gives the comparisons of NHDs of all pairs of different images under different compression rates, where the x-axis is the index of test images, and the y-axis is the NHD. Table III also shows the statics of NHDs under different compression rates. The results of TD hashing are consistent with the conclusions reported in the original paper [10]. Note that discrimination performance evaluations of image hashing schemes are always conducted based on given thresholds [1]–[8], [10]. Given a specified distance metric (e.g., NHD), two input images are judged as visually identical images if their distance is smaller than a given threshold. Otherwise, they are different images. As shown in the Fig.12 and Table III, for each test image, the WS-TD hashing achieves better NHDs distribution than the TD hashing, even though the compression rate δ is set to 0.25. Meanwhile, the discriminability of the WS-TD hashing keeps nearly the same while the compression rate decreases dramatically. Therefore, the proposed framework can hold better discriminable capability while achieve more compact hash codes.

The ROC graph is also exploited to make visual classification comparisons with respect to robustness and discrimination. The true positive rate (TPR) P_{TPR} and false positive rate (FPR) P_{FPR} are first defined:

$$P_{\text{TPR}} = \frac{N_{\text{similar}}}{N_{\text{identical}}} \tag{24}$$

$$P_{\text{FPR}} = \frac{N_{\text{distinct}}}{N_{\text{different}}} \tag{25}$$

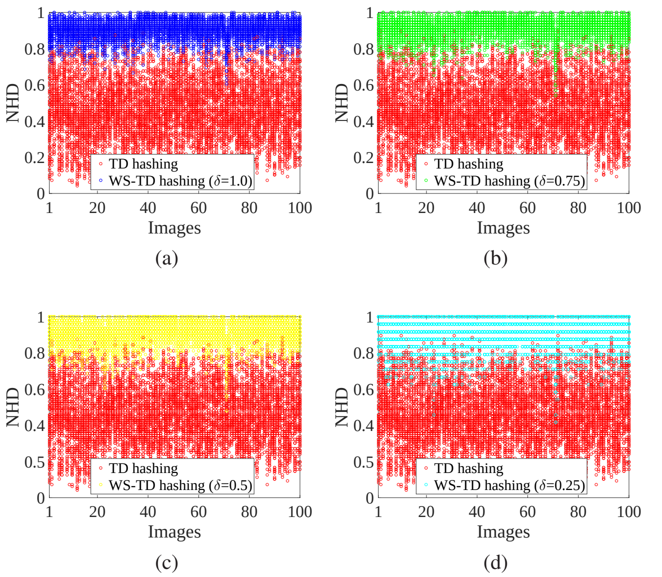


Fig. 12. Comparisons of NHDs of between each original image and the other 99 images under different compression rates. (a) $\delta = 1.0$. (b) $\delta = 0.75$. (c) $\delta = 0.5$. (d) $\delta = 0.25$.

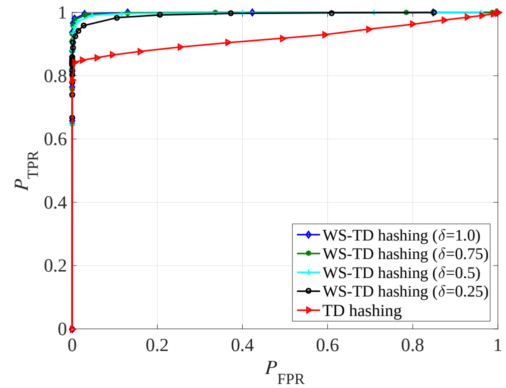


Fig. 13. Classification performance comparisons with respect to robustness and discrimination between TD hashing and WS-TD hashing with different compression rates.

where N_{similar} is the number of pairs of visually identical images that are correctly identified as similar images, N_{distinct} is the number of pairs of distinct images that are mistakenly classified as similar images, and $N_{\text{different}}$ is the total number of pairs of different images.

Fig.13 presents the ROC curve comparisons between the state-of-the-art TD hashing and WS-TD hashing algorithms with respect to robustness and discrimination. It can be seen that the ROC curves of all WS-TD hashing with different compression rates are all above those of TD hashing, and are much closer to the top-left corner. Therefore, the WS-TD hashing is superior to TD hashing in classification between robustness and discrimination.

In theory, the discrimination performance of image hashing algorithms depends mainly on the feature extraction methods. Both the WS-TD hashing and TD hashing adopt the same feature extraction method called TD based descriptor. The reason why the proposed framework gets better discrimination performance in the experiments may lie in the learned distance metric matrix $\boldsymbol{L}$ and quantization template T. On the verification stage, traditional hashing schemes (e.g. TD hashing) usually compute the hash codes of the received image with same and fixed parameters which are independent of specific images. However, to judge whether the received image is visually identical to the original image, the proposed personalized framework should calculate the hash codes of the received image based on the learned distance metric matrix $\boldsymbol{L}$ and quantization template T of the original image.

E. Limitations

The proposed framework can significantly enhance the authentication performance because of the introducing of the distance metric learning technology as well as the supervised personalized quantization strategy, on the basis of the feature extraction methods of previous works. However, it may suffer from the computational complexity and transmission overhead, compared with conventional authentication algorithms.

In terms of computational complexity, for each original image, two kinds of training sets should be generated automatically firstly. Then, the metric matrix as well as the quantization template should be learned and determined from the training samples. In terms of transmission overhead, the learned metric matrix $\boldsymbol{L}$ and quantization template T should be kept and transmitted for each image.

VI. Conclusions

This paper proposes a novel and effective personalized image authentication framework, which can make full use of feature extraction methods of existing image hashing schemes. The core technologies of the personalized framework are WLMMC and SPQ. The proposed WLMMC scheme seeks to learn an effective feature mapping space for each original image from its training samples to improve the classification performance between content-changing samples and content-preserving samples before the quantization stage. The proposed SPQ algorithm applies a supervised personalized quantization strategy for each image, taking the discriminability in each dimension into consideration. Compared with the state-of-the-art methods, the proposed personalized authentication framework can achieve better authentication performance while learn more compact binary codes. Besides, the proposed framework can also alleviate the discussed misjudgments and out-of-sample problems plagued existing methods. It is believed that the proposed personalized framework may serve as an alternative approach to the image authentication problem.

Acknowledgment

The authors would like to acknowledge the helpful comments and kindly suggestions provided by anonymous referees. Many thanks for Prof. Tang's [6], [10] and Prof. Wang's [4] source codes.

REFERENCES

[1] A. Haouzia and R. Noumeir, "Methods for image authentication: a survey," *Multimedia Tools and Applications*, vol. 39, no. 1, pp. 1–46, 2008.

[2] S. Han and C. Chu, "Content-based image authentication: current status, issues, and challenges," *International Journal of Information Security*, vol. 9, no. 1, pp. 19–32, 2010.

[3] Y. Ou and K. H. Rhee, "A survey on image hashing for image authentication," *IEICE Transactions on Information and Systems*, vol. E93D, no. 5, pp. 1020–1030, 2010.

[4] 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.

[5] X. Li, X. Sun, and Q. Liu, "Image integrity authentication scheme based on fixed point theory," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 632–645, 2015.

[6] 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.

[7] P. Korus, "Digital image integrity a survey of protection and verification techniques," *Digital Signal Processing*, vol. 71, pp. 1–26, 2017.

[8] C. Jiang and Y. Pang, "Perceptual image hashing based on a deep convolution neural network for content authentication," *Journal of Electronic Imaging*, vol. 27, no. 4, pp. 043 055:1–043 055:10, 2018.

[9] X. Liu, C. Lin, and S. Yuan, "Blind dual watermarking for color images authentication and copyright protection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1047–1055, 2018.

[10] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust image hashing with tensor decomposition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 3, pp. 549–560, 2019.

[11] F. Khelifi and A. Bouridane, "Perceptual video hashing for content identification and authentication," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 50–67, 2019.

[12] G. Schaefer and M. Stich, "Ucid: an uncompressed color image database," *Proceedings of the SPIE*, vol. 5307, pp. 472–480, 2003.

[13] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet : A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[14] X. Nie, X. Li, Y. Chai, C. Cui, X. Xi, and Y. Yin, "Robust image fingerprinting based on feature point relationship mining," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1509–1523, 2018.

[15] W. Chen, G. Ding, Z. Lin, and J. Pei, "Accelerated manhattan hashing via bit-remapping with location information," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 2441–2466, 2017.

[16] A. Bellet, A. Habrard, and M. Sebban. (2014) A survey on metric learning for feature vectors and structured data. [Online]. Available: https://arxiv.org/abs/1306.6709

[17] F. Wang and J. Sun, "Survey on distance metric learning and dimensionality reduction in data mining," *Data Mining and Knowledge Discovery*, vol. 29, no. 2, pp. 534–564, 2015.

[18] P. Yang, K. Huang, and A. Hussain, "A review on multi-task metric learning," *Big Data Analytics*, vol. 3, no. 1, p. 3, 2018.

[19] J. Wang, W. Liu, S. Kumar, and S. Chang, "Learning to hash for indexing big data - a survey," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 34–57, 2016.

[20] L. Chi and X. Zhu, "Hashing techniques: A survey and taxonomy," *ACM Computing Surveys*, vol. 50, no. 1, pp. 11:1–11:36, 2017.

[21] J. Wang, T. Zhang, J. Song, N. Sebe, and H. Shen, "A survey on learning to hash," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 769–790, 2018.

[22] J. He and D. Xu, "Large margin nearest neighbor classification with privileged information for biometric applications," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. [Online]. Available: 10.1109/TCSVT.2019.2927873

[23] D. Jang, C. D. Yoo, and T. Kalker, "Distance metric learning for content identification," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 932–944, 2010.

[24] W. Liu, J. Wang, S. Kumar, and S. Chang, "Hashing with graphs," in *Proceedings of the 28th International Conference on Machine Learning*, 2011, pp. 1–8.

[25] W. Kong and W. Li, "Double-bit quantization for hashing," in *Proceedings of the 26th Conference on Artificial Intelligence*, 2012, pp. 634–640.

[26] S. Moran, V. Lavrenko, and M. Osborne, "Variable bit quantisation for lsh," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013, pp. 753–758.

[27] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, 2005, pp. 1473–1480.

[28] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[29] K. Song, F. Nie, J. Han, and X. Li, "Parameter free large margin nearest neighbor for distance metric learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 2555–2561.

[30] G. Schaefer and M. Stich, "Ucid: an uncompressed color image database," in *Proceedings of SPIE, Storage and Retrieval Methods and Applications for Multimedia 2004*, 2004, pp. 472–480.

Zhiyong Su is currently an associate professor in the School of Automation at Nanjing University of Science and Technology, China. He received the B.S. and M.S. degrees from the School of Computer Science and Technology, Nanjing University of Science and Technology in 2004 and 2006, respectively, and received the Ph.D. from the Institute of Computing Technology, Chinese Academy of Sciences in 2009. His current interests include computer graphics, computer vision, argument reality, and machine learning.

Liang Yao received the B.S. degree from Anhui Normal University, China, in 2017. She is currently pursuing the M.S degree with the Nanjing University of Science and Technology , China. Her research interests include image processing and computer vision.

Jialin Mei received the B.S. degree from Nanjing Institute of Technology, China, in 2017. She is currently pursuing the M.S. degree with the Nanjing University of Science and Technology, China. Her research interests include image processing and computer vision.

Lang Zhou received her B.S., M.S. and Ph.D. degrees from the School of Computer Science and Technology, Nanjing University of Science and Technology in 2004, 2006 and 2010, respectively. She joined the College of Information Engineering in Nanjing University of Finance and Economics, China, in 2010. Her current interests include pattern recognition and natural language processing.

Weiqing Li is currently an associate professor in the School of Computer Science and Engineering at Nanjing University of Science and Technology, China. He received the B.S. and Ph.D. degrees from the School of Computer Sciences and Engineering, Nanjing University of Science and Technology in 1997 and 2007, respectively. His current interests include computer graphics and virtual reality.