IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 30, 2019, accepted April 11, 2019, date of current version May 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912585

LOSGSR: Learned Opinion Score Guided Shoeprint Retrieval

YANJUN WU[1], XINNIAN WANG[1], NAMUSISI LINDA NANKABIRWA[1], AND TAO ZHANG[2]
[1]School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
[2]School of Physics and Electronic Technology, Liaoning Normal University, Dalian 116029, China

Corresponding author: Xinnian Wang (wxn@dlmu.edu.cn)

This work was supported by the Ministry of Public Security of the China Foundation of the Basic Special Project under Grant 2017GABJC09.

ABSTRACT As a kind of forensic evidence, a shoeprint conveys many important human characteristics, and it plays a vital role in forensic investigations. Millions of shoeprints are acquired from crime scenes, and it is a challenging task to retrieve the most similar shoeprints for a query shoeprint. Most shoeprint retrieval methods sort shoeprint images by feature similarities with respect to the query shoeprint; however, the results are not always what the investigator expects because the retrieval algorithm cannot determine what the investigator prefers based on only the content of the query shoeprint. This paper proposes a method to guide the shoeprint retrieval process to approximate what the user wants by applying learned opinion scores. Additionally, this paper improves shoeprint retrieval effectiveness by implementing the following four perspectives: 1) using the opinion scores of multiple examples to guide the results to meet the forensic experts' expectations; 2) proposing a learning-based method to refine opinion scores, which corrects the labeled opinion scores of multiple examples and their neighbors; 3) using a manifold ranking method to propagate the opinion scores to other dataset shoeprints; and 4) introducing a coefficient matrix to prevent the tendency of the ranking scores to become low values. The experiments show that the cumulative match scores of the proposed method are more than 96.6% in the top 2% of the dataset composed of 10,096 crime scene shoeprints.

INDEX TERMS *k* nearest neighbors, opinion scores, shoeprint image retrieval, manifold ranking, multiple examples.

I. INTRODUCTION

As a kind of forensic evidence, shoeprints cannot only serve as a clue to a case but also serve as an exhibit for bringing charges for a crime. In past decades, investigators took photos of shoeprints when a case occurred and manually compared them with those captured from other crime scenes to reveal some clues about that particular case. However, manual comparisons are challenging when there are a large number of shoeprints for comparison. Therefore, proposals for automatic shoeprint retrieval methods are necessary to perform comparisons in a more efficient manner.

The goal of shoeprint retrieval algorithms is to obtain a ranking list of shoeprints in a database according to their similarities to the query shoeprint. In the past few years, many shoeprint image retrieval methods have been proposed, and they have performed well in terms of their datasets [3], [7], [9], [11], [21], [27]. However, most of these methods only take high quality shoeprints into account, and the performance becomes comparatively low when applied to poor quality crime scene shoeprint retrieval. For example, Figure 1 shows four shoeprints that were left by the same shoe, and they represent good quality characteristics for crime scene shoeprints in practice. However, most of the existing algorithms have poor performance. We believe this poor performance is due to the following two reasons. The first reason is that few current shoeprint retrieval algorithms consider the well-known gap between semantic concepts and low-level features, and the other reason is that most of the methods neglect forensic experts' opinions.

Bouridane *et al.* [2] propose a semi-supervised learning retrieval method, which introduces experts' opinions into the retrieval method, and the method has a good performance under the condition that shoeprint examples

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

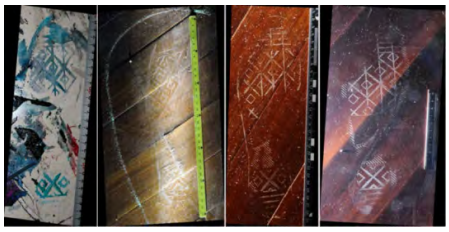

FIGURE 1. Shoeprints with the same patterns left in different crime scenes by the same culprit.

are correctly labeled. Here, the shoeprint examples denote shoeprints captured from the same scene where the query shoeprint is acquired. However, in practice, different experts may label the same example with different opinion scores depending on their perspectives, which leads to different ranking results. Therefore, we propose a learned opinion score guided shoeprint retrieval method (LOSGSR) to guide the ranking results using the learned opinion scores. Our motivations are (i) to use a learning-based method to refine opinion scores, which corrects the labeled opinion scores of multiple examples and their neighbors and (ii) to use the refined opinion scores of shoeprint examples to guide shoeprint retrieval. The experiments show that the cumulative match score of our method is improved by 3.07% and 7.79% above those of Bouridane *et al.* [2] and Zhou *et al.* [22] on average respectively, and the cumulative match score in the top 2% of the dataset is more than 96.6%. To the best of our knowledge, the performance outperformed those of other methods.

The main contributions of this paper are as follows:

1) We use the opinion scores of multiple shoeprint examples to guide the ranking scores to meet the forensic experts' expectations.

2) We propose a learning-based method to refine opinion scores, which corrects the labeled opinion scores of multiple examples and their neighbors.

3) We present a manifold ranking method which can narrow the well-known gap between low-level features and semantic concepts. The differences from the previous manifold ranking methods include the following: (i) we take into account the experts' opinion scores assigned for multiple shoeprint examples, and the results are therefore in accordance with the experts' opinions; (ii) the feature similarities between the query and dataset shoeprint images are used as a constraint criterion; (iii) the refined opinion scores of the neighbors are taken into consideration; and (iv) a coefficient matrix is used to prevent the tendency of the computed ranking scores from becoming low values.

4) Our method can work well for real crime scene shoeprint retrieval. The cumulative match score is more than 96.6% in the top 2% of the database composed of 10096 real crime scene shoeprint images.

The remainder of the paper is organized as follows. Section II provides a brief review of related work. Section III details the problem and presents the proposed method. Section IV provides the experimental results followed by the conclusion and future work.

II. RELATED WORK

In this section, we briefly review some existing shoeprint retrieval methods. According to the type of features used for retrieval existing methods can be broadly classified into three categories: 1) holistic appearance feature-based methods, 2) regional appearance feature-based methods, and 3) interest point feature-based methods.

Methods of the first class usually view a shoeprint as a whole. Bouridane *et al.* [2] provide a fractal-based feature extraction and matching method. However, this method is sensitive to variations in rotations and translations. The moments invariant features are used for automatic shoeprint retrieval, such as Hu's moments [3] and Zernike moments [4], [5], and they report good results of their methods; however, the shoeprints that they used are clear and complete; partial data is not considered. Patil and Kulkarni [6] provide a retrieval method that uses the Radon transform and the Gabor feature to align and represent a shoeprint, and the method shows good performance for partial shoeprints generated from full shoeprints in a rigorous manner. Chazal *et al.* [7] and Gueham *et al.* [8], [9] use the Fourier transform to analyze the frequency spectra of shoeprint images, but the methods are sensitive to partial shoeprints. Cervelli *et al.* [10], [12] perform the Fourier transform on the cropped shoeprint images However, this method is sensitive to geometry transformations. Alizadeh and Kose [13] implement a sparse representation method to retrieve shoeprints. They show good performance on shoeprint retrieval. However, features used in this method are sensitive to variations in rotations and translations. Richetelli *et al.* [14] compare different methods on a scene-like shoeprint database. The authors have tested phase-only correlation (POC) method, Fourier-Mellin transformation and the scale-invariant feature transform (SIFT) method on shoeprint image retrieval. Results on a scene-like shoeprint database show that the POC method has better performance than the Fourier-Mellin transformation and the SIFT method; however, their performances may considerably drop when applied to degraded crime scene shoeprints. Kong *et al.* [15] use the pre-trained convolutional neural network to extract multi-channel deep features, and match these features with their proposed multi-channel normalized cross-correlation method. Their results show good performance. However, this algorithm requires a large amount of computation to compute feature similarities.

In the second class of methods, shoeprints are always divided into different semantic regions. Tang *et al.* [16], [17] represented a shoeprint with an attributed relational graph (ARG) whose nodes are the fundamental shapes in shoes such as lines, circles, ellipses, etc. The authors show that their method is robust to distortions and partial shoeprints. However, it is a challenge work to deal with

shoeprints with random extrusions, intrusions, breaks and complex patterns that cannot be represented by such fundamental shapes Pavlou and Allinson [18], [19] use the maximally stable extremal regions (MSER) feature detector to detect region features. However, shoeprints with noises and distortions may lead to an incorrect detection. Kortylewski *et al.* [21] describe a periodic pattern-based shoeprint retrieval algorithm. The method detects periodic patterns of the shoeprint first and then evaluates the similarity through comparing the Fourier features of the periodic patterns. The algorithm can perform well on shoeprints with periodic patterns. However, how to address degraded shoeprint images remains a problem. Wang *et al.* [1] divides a shoeprint into top regions and bottom regions, and then wavelet-Fourier-Mellin transform based features of both regions are extracted for retrieval. The method performs well for its invariant features and matching score estimation method, but it does not attempt to narrow the gap between the semantic concepts and the low-level features. Kortylewski and Vetter [23] learn a compositional active basis model to each reference shoeprint, which is used to evaluate against other query images at testing time.

Methods of the third class always extract interest points first, and then the rotation, scaling and translation (RST) invariant features are used for performing shoeprint retrieval. Pavlou and Allinson [19] utilizes the MSER features and SIFT descriptors to describe the shoeprint images for shoeprint retrieval. Nibouche *et al.* [24] use the SIFT descriptor to represent shoeprint images, and the random sample consensus (RANSAC) method is used to estimate the matching performance. Crookes *et al.* [25] and Su *et al.* [26] also perform shoeprint retrieval with a Harris detector and the SIFT descriptors. Almaadeed *et al.* [28] use multiscale Harris and Hessian detectors to extract interest points, and the SIFT descriptor is employed to describe the shoeprint images Wang *et al.* [29] utilize the SIFT descriptors to describe shoeprints, they report good performance with clear shoeprints, but the performance may dramatically diminish with degraded crime scene shoeprints Most of the interest point-based methods have good performance with clear shoeprints, but the performance may considerably diminish with degraded crime scene shoeprints. A possible reason for this outcome may be that the real crime scene shoeprints are highly degraded and randomly incomplete.

Bouridane *et al.* [2] considers both feature similarity and manually labeled opinion scores of multiple examples, and the results correlate well with the forensic experts' opinions, but they neglect the effect of the mislabeled opinion scores and the contribution of the opinion scores of the neighborhood.

III. METHODS

A. PROBLEM STATEMENT

Let $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N\} \subset \mathbb{R}_+^m$ denote N shoeprint images collected from different crime scenes. $\mathbf{q}_1$ represents the query shoeprint image, and $\{\mathbf{q}_2, \cdots, \mathbf{q}_n\} \subset \mathbb{R}_+^m (n \geq 2)$ denotes multiple shoeprint examples acquired from the same crime scene as the query shoeprint. $o_i (i = 2, \cdots, n)$ denotes the rating score of $\mathbf{q}_i$ assigned by the forensic expert interactively, and this is according to its similarity to the query shoeprint $\mathbf{q}_1$

Let $\mathbf{U} = \mathbf{Q} \cup \mathbf{D} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, u_n, \mathbf{u}_{n+1}, \ldots, \mathbf{u}_{n+N}\} \subset \mathbb{R}_+^m \mathbf{Q} = \mathbf{q}_1 \cup \{\mathbf{q}_2, \cdots, \mathbf{q}_n\} \subset \mathbb{R}_+^m$, and $\mathbf{O} = o_1 \cup \{o_2, \cdots, o_n\} \subset \mathbb{R}_+^m$ Given $\mathbf{D}$, $\mathbf{Q}$ and $\mathbf{O}$, the goal of this paper is to find a function $f : \mathbf{U} \to \mathbb{R}_+$ that assigns to each shoeprint $\mathbf{u}_i \in \mathbf{U}$ a ranking score $\mathbf{f}_i \in \mathbb{R}_+, 0 \leq \mathbf{f}_i \leq 1$, and $\mathbf{f}_i$ should be as close as possible to the opinion score assigned by the forensic expert. Let $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_K]$, and $K = n + N = |\mathbf{U}|$

B. MOTIVATIONS AND FORMULATIONS

Most of the shoeprint retrieval methods sort shoeprint images by feature similarities in contrast to the query shoeprint Additionally, sometimes the ranking lists are not truly what the investigator expects because the algorithms do not take into consideration what the investigator requires. Therefore our motivation for this paper is to design and train an algorithm with features that describe what the investigator would prefer by assigning opinion scores to shoeprint examples.

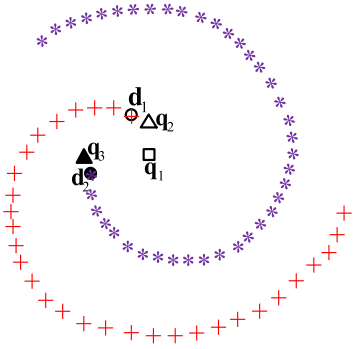

FIGURE 2. Intuitive illustration of the proposed LOSGSR method.

Figure 2 shows the basic idea of our proposed LOSGSR method. The samples from the dataset $\mathbf{D}$ are represented with '+'s and '∗'s, and symbol denotes a certain kind of shoeprint in the feature space. In addition, the samples from each kind of shoeprint have a closer distance in the semantic space, and $\mathbf{d}_1$ and $\mathbf{d}_2$ are samples from different types of shoeprints that are represented by a circle and a filled circle. $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ are shoeprint examples from a crime scene. $\mathbf{q}_1$ is the query shoeprint, which is represented by a rectangle; $\mathbf{q}_2$ and $\mathbf{q}_3$ are represented by a triangle and a filled triangle, respectively. We aim to implement ranking scores that can meet the following two criteria: (i) shoeprints closer to the query shoeprint in the semantic space should have higher ranking scores; and (ii) nearby shoeprints in the semantic space should share similar ranking scores. In the feature space, it is very difficult to discern which one of the $\mathbf{d}_1$ and $\mathbf{d}_2$ is closer to the query shoeprint $\mathbf{q}_1$. To rank the shoeprints

according to the investigator's expectations, we introduce the investigator's opinions into the shoeprint retrieval. The investigator assigns rating scores o_2 and o_3 to $\mathbf{q}_2$ and $\mathbf{q}_3$, according to their similarity in the semantic space to $\mathbf{q}_1$. Assume that o_2 is greater than o_3. Through the guidance provided from the opinion scores of $\mathbf{q}_2$ and $\mathbf{q}_3$, the ranking score of $\mathbf{d}_1$ that is computed by the retrieval algorithm should be greater than that of $\mathbf{d}_2$; otherwise, the ranking results will not be consistent with the human evaluation. Finally, through the opinion scores of $\mathbf{q}_2$, $\mathbf{q}_3$ and their neighbors, the ranking scores can be guided to meet the investigator's expectations.

Based on these results, we formulate our method using the following optimization problem.

$$\begin{aligned}\mathbf{f}^* = \arg\min_{\mathbf{f}} Q(\mathbf{f}) = &\frac{1}{2}\alpha \sum_{\mathbf{u}_i\in\mathbf{U}} \mathbf{R}_{i,i}\left(\mathbf{f}_i - \mathbf{y}_i\right)^2 \\ &+ \frac{1}{4}\beta \sum_{\mathbf{u}_i\in\mathbf{U}}\sum_{\mathbf{u}_j\in\mathbf{U}} \mathbf{W}_{ij}\left(\frac{1}{\sqrt{\mathbf{C}_{ii}}}\mathbf{f}_i - \frac{1}{\sqrt{\mathbf{C}_{jj}}}\mathbf{f}_j\right)^2 \\ &+ \frac{1}{2}\gamma \sum_{\mathbf{q}_i\in\mathbf{Q}}\left[\left(\mathbf{f}_i - \hat{\mathbf{y}}_i\right)^2 + \sum_{\mathbf{u}_l\in N_k(\mathbf{q}_i)}\left(\mathbf{f}_l - \hat{\mathbf{y}}_l\right)^2\right] \end{aligned} \quad (1)$$

where α, β, γ are the regularization parameters, and $0 < \alpha, \beta, \gamma < 1$, and $\alpha < \beta$. $N_k(\mathbf{q}_i)$ denotes the k nearest neighbors of $\mathbf{q}_i$. We denote the initial ranking score list as $\mathbf{y} = \{\mathbf{y}_i | i = 1, 2, \cdots, K\}^{\mathrm{T}}$, where $\mathbf{y}_i$ is the feature similarity value for the ith shoeprint image in the dataset. $\mathbf{y}_i$ is computed according to (9). For the ranking score list is considered to be reasonable, the target image has a relatively high probability of ranking at the top of the ranking list; thus, we re-rank the top-m images in the ranking score list and encourage the ranking scores of the rest to be low values. The newly defined initial ranking score list is $\mathbf{y} = \{\mathbf{y}_i | i = 1, 2, \cdots, K\}^{\mathrm{T}}$. If a dataset image $\mathbf{u}_i$ does not rank at the top-m of the ranking list, then its corresponding initial ranking score $\mathbf{y}_i = 0$. $\hat{\mathbf{y}}_i$ denotes the learning-based opinion score of the ith example from the crime scene, and $\hat{\mathbf{y}}_l$ denotes that of one of its neighbors.

The first term is the initial ranking score correlation term. The shoeprints closer to the query shoeprint in the feature space should have higher ranking scores. $\mathbf{R}$ is a diagonal $K \times K$ matrix, in which the ith element $\mathbf{R}_{i,i}$ represents the initial rank order index in a descending order of the feature similarity values to the query shoeprint. The process of computing $\mathbf{R}_{i,i}$ is as follows: (i) compute the feature similarity values between the query image and the dataset images according to (9); (ii) rank the dataset images in descending order depending on the feature similarity values, and let $Ind\left(\mathbf{u}_i\right)$ denote the index position of $\mathbf{u}_i$ in the ranking list; and (iii) the initial rank order index is $\mathbf{R}_{i,i} = \frac{Ind(\mathbf{u}_i)}{K}$.

The second term is the smoothness term. The shoeprint images nearby in the feature space should share similar ranking scores. $\mathbf{W}$ represents an affinity matrix, and $\mathbf{W}_{i,j} = w\left(\mathbf{u}_i, \mathbf{u}_j\right)$ is computed according to (9). $\mathbf{C}$ denotes a diagonal matrix, in which the ith element $\mathbf{C}_{i,i}$ denotes the sum of ith row of the affinity matrix $\mathbf{W}$. Formally, $\mathbf{C}_{i,i} = \sum_{j=1}^{K} \mathbf{W}_{i,j}$.

The third term is the fitting term. This term associates a cost for the degree of deviation from the learning-based opinion scores. With the help of the second term, the opinion scores can be propagated to other shoeprint images in the dataset.

Due to the convenience of solving the problem, we rearrange the third term in (1); thus, the cost function can be rewritten as follows:

$$\begin{aligned} Q(\mathbf{f}) = &\frac{1}{2}\alpha \sum_{i=1}^{K} \mathbf{R}_{i,i}\left(\mathbf{f}_i - \mathbf{y}_i\right)^2 \\ &+ \frac{1}{4}\beta \sum_{i=1}^{K}\sum_{j=1}^{K} \mathbf{W}_{ij}\left(\frac{1}{\sqrt{\mathbf{C}_{ii}}}\mathbf{f}_i - \frac{1}{\sqrt{\mathbf{C}_{jj}}}\mathbf{f}_j\right)^2 \\ &+ \frac{1}{2}\gamma \sum_{i=1}^{K} \mathbf{A}_{i,i}\left(\mathbf{f}_i - \hat{\mathbf{y}}_i\right)^2 \end{aligned} \quad (2)$$

where $\mathbf{A}_{i,i} = 1_{\mathbf{Q}\cup N_k(\mathbf{Q})}\left(\mathbf{u}_i\right)$, which indicates whether $\mathbf{u}_i$ is in the union set of $\mathbf{Q}$ and its neighborhood set $N_k(\mathbf{Q})$ or not, and $N_k(\mathbf{Q}) = \{N_k(\mathbf{q}_1), \cdots, N_k(\mathbf{q}_n)\}$.

The matrix-vector formulation of (2) is:

$$\begin{aligned} Q(\mathbf{f}) = &\frac{1}{2}\alpha\left(\mathbf{f} - \mathbf{y}\right)^{\mathrm{T}}\mathbf{R}\left(\mathbf{f} - \mathbf{y}\right) + \frac{1}{4}\beta\mathbf{f}^{\mathrm{T}}\mathbf{L}\mathbf{f} \\ &+ \frac{1}{2}\gamma\left(\mathbf{f} - \hat{\mathbf{y}}\right)^{\mathrm{T}}\mathbf{A}\left(\mathbf{f} - \hat{\mathbf{y}}\right) \end{aligned} \quad (3)$$

where $\mathbf{L}$ denotes a symmetric normalized Laplacian matrix, $\mathbf{L} = \mathbf{I} - \mathbf{C}^{-1/2}\mathbf{W}\mathbf{C}^{-1/2}$ and $\mathbf{f} = \{f_i | i = 1, 2, \cdots, K\}^{\mathrm{T}}$, and $\hat{\mathbf{y}} = \left\{\hat{y}_i | i = 1, 2, \cdots, K\right\}^{\mathrm{T}}$. If a dataset image $\mathbf{u}_i$ does not belong to the neighbor set of any examples, then its corresponding opinion score $\hat{y}_i = 0$.

The proposed LOSGSR algorithm has three phases: off-line feature extraction offline opinion score prediction model training and on-line retrieval. The flow diagram of the proposed LOSGSR is shown in Figure 3. In the off-line feature extraction phase, wavelet-Fourier-Mellin based features of every image in the database $\mathbf{D}$ are firstly extracted and pooled into the feature database; and then the affinity matrix of the database $\mathbf{D}$ is computed and prepared for retrieval. In the offline opinion score prediction model training phase, firstly, forensic experts are asked to assign rating scores to the training shoeprints depending on their similarities to the reference shoeprint; and then the feature similarities of these training shoeprints to the reference image are computed; finally, parameters of the opinion score prediction model can be acquired. In the on-line retrieval phase, firstly, the rating scores of multiple examples $\mathbf{q}_i (i = 2, \cdots, n)$ are assigned by the user, and then the labeled opinion scores of multiple examples and their neighbors can also be computed; secondly, the affinity matrix of the database $\mathbf{U}$ is constructed; and then the predicted opinion score can be acquired by using the trained prediction model, and the opinion scores can be refined based on the labeled and predicted opinion scores; finally the ranking scores are computed and the ranking list is

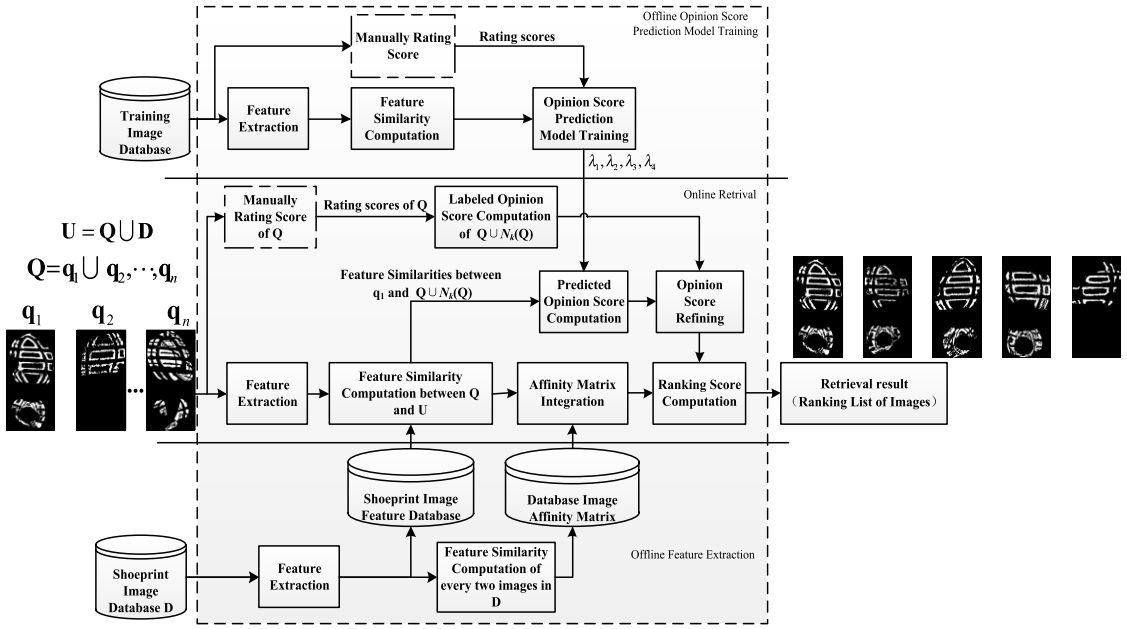

FIGURE 3. The flow diagram of the proposed LOSGSR method Two dotted boxes denote the process of manually rating scores. In the off-line opinion score prediction model training phase, it is in accordance with forensic experts' opinions. In the on-line retrieval phase, it depends on the users' opinions.

acquired based on the ranking scores. We detail the proposed LOSGSR algorithm in the following sections.

C. LEARNING-BASED OPINION SCORE COMPUTATION METHOD

There are mainly two methods to obtain the opinion scores. One method requires the forensic experts to manually label the examples, and the other method requires learning. In practice, the former method may involve different experts labeling the same example with different opinion scores according to their opinions; hence this method leads to different ranking results. Therefore, we proposed a learning-based method to correct the labeled opinion scores. The learning-based opinion scores $\hat{\mathbf{y}}$ can be formally defined as follows:

$$\hat{\mathbf{y}}(\mathbf{u}_i) = (1 - a_{\mathrm{p}})\, \mathbf{y}_{\mathrm{s}}(\mathbf{u}_i) + a_{\mathrm{p}} \mathbf{y}_{\mathrm{p}}(\mathbf{u}_i) \tag{4}$$

where $\mathbf{y}_{\mathrm{s}}$ and $\mathbf{y}_{\mathrm{p}}$ denote the labeled opinion scores and the predicted opinion scores, respectively, and a_{p} denotes the weighted value.

The labeled opinion score $\mathbf{y}_{\mathrm{s}}$ is defined as follows:

$$\mathbf{y}_{\mathrm{s}}(\mathbf{u}_i) = \begin{cases} o_i & \mathbf{u}_i \in \mathbf{Q} \\ w(\mathbf{q}_j, \mathbf{u}_i)\, o_j & \mathbf{u}_i \in N_k(\mathbf{q}_j) \end{cases} \tag{5}$$

where $w(\mathbf{q}_j, \mathbf{u}_i)$ denotes the similarity value between $\mathbf{q}_j$ and $\mathbf{u}_i$. This equation means that if $\mathbf{u}_i$ is manually labeled, its labeled opinion score is equal to the rating score o_i; otherwise, its labeled opinion score is a weighted version of the training score of the closest example to it.

The predicted opinion score $\mathbf{y}_{\mathrm{p}}$ is defined as

$$\mathbf{y}_{\mathrm{p}}(\mathbf{u}_i) = h(w(\mathbf{q}_1, \mathbf{u}_i), \lambda), \quad \mathbf{u}_i \in \mathbf{Q} \cup N_k(\mathbf{Q}) \tag{6}$$

where $h(w(\mathbf{q}_1, \mathbf{u}_j), \lambda)$ denotes the pre-learned logistic function, λ denotes the model parameter vector, and $w(\mathbf{q}_j, \mathbf{u}_i)$ denotes the similarity value between the shoeprint $\mathbf{q}_j$ and $\mathbf{u}_i$.

1) LABELED OPINION SCORES

For an image $\mathbf{q}_i \in \mathbf{Q}$, its labeled opinion score equals the rating score that is labeled by the experts. The procedures are as follows:

Step 1: The labeled opinion score of the query shoeprint image $\mathbf{y}_{\mathrm{s}}(\mathbf{q}_1) = o_1$;

Step 2: Five levels are used to evaluate the similarity between the query shoeprint and other examples These levels are the extremely high, high, medium, low and extremely low levels, in decreasing order. The ratings between the extremely low and extremely high levels are mapped to numbers between 0.2 and 1, and the numbers are defined as the rating score shown in Table 1.

TABLE 1. Shoeprint similarity rating scale.

Similarity level	Extremely Low	Low	Medium	High	Extremely High
Rating score	0.2	0.4	0.6	0.8	1

Step 3: An expert is then asked to judge the similarity level of each example $\mathbf{q}_i$ and to assign a score by using the provided shoeprint similarity rating scale shown in Table 1.

For example, six shoeprints collected at a crime scene are shown in Figure 4, and let $o_1, o_2, o_3, o_4, o_5, o_6$ denote the rating scores of the six images, which are interactively labeled by the investigator according to the rating scheme

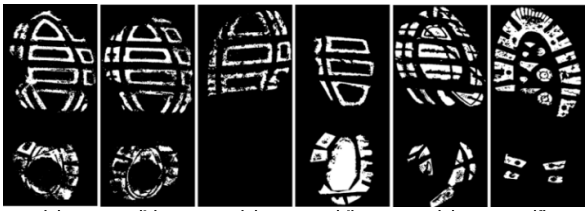

FIGURE 4. Rating scores of the query shoeprint and the other examples. (a) $o_1 = 1$. (b) $o_2 = 1$. (c) $o_3 = 0.8$. (d) $o_4 = 0.6$. (e) $o_5 = 0.4$. (f) $o_6 = 0.2$.

described in Table 1. We assume that the first image is the query shoeprint; thus, we set o_1 equal to 1, and o_2 should have the same opinion score because the second shoeprint has the same pattern as the query shoeprint. Similarly, o_3, o_4, o_5, o_6, corresponding to their similarity level, can be labeled 0.8, 0.6, 0.4 and 0.2, respectively.

For an image $\mathbf{u}_i \in N_k\left(\mathbf{q}_j\right)$, its labeled opinion score is the weighted version of the rating score of $\mathbf{q}_j$, and it is defined as $\mathbf{y}_{\mathrm{s}}\left(\mathbf{u}_i\right) = w\left(\mathbf{q}_j, \mathbf{u}_i\right) o_j$. If $\mathbf{u}_i$ is in the intersection of the neighbor sets of different multiple examples, $\mathbf{q}_j$ represents the one closest to it. That is, $\mathbf{q}_j = \arg\max_{\mathbf{q}_{j^*}} \left(w\left(\mathbf{q}_{j^*}, \mathbf{u}_i\right)\right)$.

2) PREDICTED OPINION SCORES COMPUTATION

Inspired by the methods of objective image quality assessment [30], we approximate the predicted opinion score $\mathbf{y}_{\mathrm{p}}(i)$ by using a four-parameter logistic function, which is defined as follows:

$$\mathbf{y}_{\mathrm{p}}(i) = h\left(w\left(\mathbf{q}_1, \mathbf{u}_i\right), \lambda\right) = \frac{\lambda_1 - \lambda_2}{1 + \exp\left(w\left(\mathbf{q}_1, \mathbf{u}_i\right) - \lambda_3 / |\lambda_4|\right)} + \lambda_2 \tag{7}$$

Here, $\boldsymbol{\lambda} = \left[\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4\right]$, and the λ_i's are the parameters that parameterize the function, mapping the similarity value to the opinion score.

Given a training set $\mathbf{T} = \left\{\left(x^{(i)}, y^{(i)}\right), i = 1, 2, \cdots, M\right\}$, the optimal values of parameters λ can be acquired by using the least square method:

$$\lambda^* = \arg\min_{\lambda} \sum_{i=1}^{M} \left\{\frac{\lambda_1 - v\lambda_2}{1 + v\exp\left(x^{(i)} - \lambda_3 / |\lambda_4|\right)} + \lambda_2 - \mathbf{y}^{(i)}\right\}^2 \tag{8}$$

where $x^{(i)}$ represents the feature similarity between the ith image and its reference image, and $\mathbf{y}^{(i)}$ represents the rating scores.

The training set $\mathbf{T}$ is constructed as follows:

Step 1: The training shoeprint images are randomly divided into L groups. Each group $\mathbf{G}_l$ $(l = 1, 2, \ldots, L)$ consists of one reference shoeprint image and n_l different shoeprint images.

Step 2: Each group of shoeprint images is shown to different forensic experts, and the forensic experts will score other images in this group, depending on their similarities to the reference image, while basing their score on the rating scheme described in Table 1.

Step 3: The mean score of different experts for each image is recorded as $y^{(i)}$, and its feature similarity to the reference image is recorded as $x^{(i)}$.

Step 4: The training set

$$\mathbf{T} = \left\{\left(x^{(i)}, y^{(i)}\right), i = 1, 2, \cdots, \sum_{l=1}^{L} n_l\right\}.$$

D. AFFINITY MATRIX COMPUTATION METHOD

Each of the elements in the affinity matrix $\mathbf{W}$ denotes the similarity of every two images in the feature space, and we use the holistic and regional appearance-based features proposed in [2] to compute the similarity value. Here, we give a brief overview about this method.

The matching score between two shoeprint images is defined as a weighted sum of both holistic and regional similarity values, and it is formally defined as:

$$\mathbf{W}(i, j) = a_{\mathrm{r}} S_{\mathrm{r}}\left(\mathbf{u}_i, \mathbf{u}_j\right) + a_{\mathrm{h}} S_{\mathrm{h}}\left(\mathbf{u}_i, \mathbf{u}_j\right) \tag{9}$$

where a_{r} and a_{h} denote the weighted parameters and subject to $a_{\mathrm{r}} + a_{\mathrm{h}} = 1$. $S_{\mathrm{r}}\left(\mathbf{u}_i, \mathbf{u}_j\right)$ denotes the regional appearance feature similarity value, and $S_{\mathrm{h}}\left(\mathbf{u}_i, \mathbf{u}_j\right)$ denotes the holistic appearance feature similarity value.

For a shoeprint image $\mathbf{u}_i$, its feature extraction process has four steps as follows.

Step 1: Separate the shoeprints from different backgrounds and then normalize the shoeprints.

Step 1.1: Binarized shoeprint image extraction. Firstly, the collected crime scene shoeprints are split into a grid of cells; secondly, a thresholding method (e.g. Otsu's method) is used to extract sub shoeprints in each cell; finally, morphological operations are applied to smooth edges and eliminate small holes.

Step 1.2: Resolution normalization. The picture of the print is taken with a forensic scale near to the print in a forensic scene, and it is rescaled to a predefined dpi.

Step 1.3: Orientation normalization. The shoeprint contour model (SPCM) described in [1] is applied to normalize the shoeprint.

Step 2: The binarized and normalized shoeprint image $\mathbf{u}_i$ is divided into the top region and the bottom region, and they are denoted as $\mathbf{S}_{\mathrm{top}}(i)$ and $\mathbf{S}_{\mathrm{bottom}}(i)$, respectively.

Step 3: The whole shoeprint $\mathbf{u}_i$ and its two corresponding regions $\mathbf{S}_{\mathrm{top}}(i)$ and $\mathbf{S}_{\mathrm{bottom}}(i)$ are decomposed at a specified number of levels by using the Haar wavelet. At each level, we can acquire one approximation and three details. The coefficients can be represented as follows:

$$\begin{aligned}
&\mathbf{FW}\left(\mathbf{u}_i\right) = \left\{\mathbf{F}_{\mathrm{W}}^{(\mathbf{u}_i)}(l, h, v) \,|0 \le l \le L, h, v = 0, 1\right\} \\
&\mathbf{FW}\left(\mathbf{S}_{\mathrm{top}}(i)\right) = \left\{\mathbf{F}_{\mathrm{W}}^{(\mathbf{S}_{\mathrm{top}}(i))}(l, h, v) \,|0 \le l \le L, h, v = 0, 1\right\} \\
&\mathbf{FW}\left(\mathbf{S}_{\mathrm{bottom}}(i)\right) = \left\{\mathbf{F}_{\mathrm{W}}^{(\mathbf{S}_{\mathrm{bottom}}(i))}(l, h, v) \,|0 \le l \le L, h, v = 0, 1\right\}
\end{aligned} \tag{10}$$

where L is the maximum level. To avoid merging the useful neighbor patterns, L should be able to meet the criterion: $2^{L-1} \leq D_{\min}$, where $D_{\min}$ represents the minimum distance between two neighbor patterns, which can be specified interactively.

Step 4: The Fourier-Mellin transform is performed on each band of the wavelet coefficients, and then a band-pass filter proposed in [35] is used to weaken the effect of the connections between patterns and noises, such as small holes, intrusions, extrusions and broken patterns. The filtered Fourier-Mellin domain coefficients of $\mathbf{FW}(\mathbf{u}_i)$ are used as holistic appearance features of the shoeprint $\mathbf{u}_i$, and those of $\mathbf{FW}(\mathbf{S}_{\text{top}}(i))$ and $\mathbf{FW}(\mathbf{S}_{\text{bottom}}(i))$ are used as region features. Here, we use $\mathbf{FMW}(\mathbf{u}_i)$, $\mathbf{FMW}(\mathbf{S}_{\text{top}}(i))$ and $\mathbf{FMW}(\mathbf{S}_{\text{bottom}}(i))$ to denote the holistic and region appearance feature of the shoeprint $\mathbf{u}_i$.

For two shoeprint images $\mathbf{u}_i$ and $\mathbf{u}_j$, the holistic appearance similarity $S_{\text{h}}(\mathbf{u}_i, \mathbf{u}_j)$ between them is computed by their correlation coefficient of the appearance features. The regional appearance similarity between them is a weighted sum of correlation coefficients of both $\mathbf{FMW}(\mathbf{S}_{\text{top}}(i))$ and $\mathbf{FMW}(\mathbf{S}_{\text{bottom}}(i))$. Please refer to (7)–(19) in [1] for details about how to set the weights adaptively.

E. SOLUTION

The dimensions of the matrix $\mathbf{R}$, $\mathbf{W}$, $\mathbf{L}$, $\mathbf{C}$ and $\mathbf{A}$ are $K \times K$s, and it requires high memory to simulate the formulation when there are a large number of shoeprints in the dataset. In this study, we apply the solution of iterative formulation as follows.

By differentiation of $Q(\mathbf{f})$ with respect to $\mathbf{f}$, we can obtain

$$\frac{\partial Q}{\partial \mathbf{f}} = \alpha\mathbf{R}(\mathbf{f}-\mathbf{y}) + \beta\mathbf{L}\mathbf{f} + \gamma\mathbf{A}(\mathbf{f}-\hat{\mathbf{y}}) = 0 \tag{11}$$

Equation (11) can be transformed into

$$\alpha\mathbf{R}\mathbf{f}(\mathbf{f}-\mathbf{y}) + \beta(\mathbf{I}-\mathbf{S})\mathbf{f} + \gamma\mathbf{A}(\mathbf{f}-\hat{\mathbf{y}}) = 0 \tag{12}$$

where $\mathbf{S} = \mathbf{C}^{-1/2}\mathbf{W}\mathbf{C}^{-1/2}$. Then, we transform (12) into the following form.

$$\mathbf{f} = \left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)\mathbf{f} + \frac{\gamma\mathbf{A}}{\beta\mathbf{I}+\gamma\mathbf{A}}\hat{\mathbf{y}} + \frac{\alpha\mathbf{R}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{y} \tag{13}$$

Finally, we can obtain the iterative formulas as follows:

$$\mathbf{f}^{(t)} = \left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)\mathbf{f}^{(t-1)} + \frac{\gamma\mathbf{A}\hat{\mathbf{y}}+\alpha\mathbf{R}\mathbf{y}}{\beta\mathbf{I}+\gamma\mathbf{A}} \tag{14}$$

where t denotes the number of iterations.

F. CONVERGENCE ANALYSIS

Similar to the convergence analysis performed in [31], we show the sequence $\{\mathbf{f}^{(t)}\}$ converges and compute the limit as follows.

We assume that $\mathbf{f}^{(0)} = \frac{\gamma\mathbf{A}\hat{\mathbf{y}}+\alpha\mathbf{R}\mathbf{y}}{\beta\mathbf{I}+\gamma\mathbf{A}}$, and by the iterative equation, we can obtain

$$\mathbf{f}^{(t)} = \left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)^t \frac{\gamma\mathbf{A}\hat{\mathbf{y}}+\alpha\mathbf{R}\mathbf{y}}{\beta\mathbf{I}+\gamma\mathbf{A}} + \frac{\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}} \times \sum_{i=0}^{t-1}\left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)^i (\gamma\mathbf{A}\hat{\mathbf{y}} + \alpha\mathbf{R}\mathbf{y}) \tag{15}$$

where $0 < \alpha, \beta, \gamma < 1$, $\alpha < \beta$ and the spectral radius satisfies $\rho\left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right) < 1$; we can determine the limit of $\mathbf{f}^{(t)}$ as follows:

$$\lim_{t\to\infty}\mathbf{f}^{(t)} = \lim_{t\to\infty}\left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)^t \frac{\gamma\mathbf{A}\hat{\mathbf{y}}+\alpha\mathbf{R}\mathbf{y}}{\beta\mathbf{I}+\gamma\mathbf{A}} + \frac{\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\lim_{t\to\infty}\sum_{i=0}^{t-1}\left(\frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} - \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)^i \times (\gamma\mathbf{A}\hat{\mathbf{y}} + \alpha\mathbf{R}\mathbf{y}) \tag{16}$$

Hence, we can determine the limit value $\mathbf{f}^*$ as follows:

$$\mathbf{f}^* = 0 + \frac{\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\left(\mathbf{I} - \frac{\beta\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{S} + \frac{\alpha\mathbf{I}}{\beta\mathbf{I}+\gamma\mathbf{A}}\mathbf{R}\right)^{-1} \times (\gamma\mathbf{A}\hat{\mathbf{y}}+\alpha\mathbf{R}\mathbf{y}) = (\alpha\mathbf{R}+\beta(\mathbf{I}-\mathbf{S})+\gamma\mathbf{A})^{-1} \times (\gamma\mathbf{A}\hat{\mathbf{y}} + \alpha\mathbf{R}\mathbf{y}) \tag{17}$$

where $\mathbf{f}^*$ is equivalent to $(\alpha\mathbf{R} + \beta(\mathbf{I}-\mathbf{S}) + \gamma\mathbf{A})^{-1}(\gamma\mathbf{A}\hat{\mathbf{y}} + \alpha\mathbf{R}\mathbf{y})$.

IV. EXPERIMENTS

A. EXPERIMENT CONFIGURATION

1) DATASETS

We test the proposed method on three shoeprint databases. The first one is a real crime scene shoeprint image database named MUES-SR10KS2S [1], and the others are public available shoeprint databases, i.e. FID-300 [23] and CS [14].

Probe image set and gallery image set in MUES-SR10KS2S are all crime scene shoeprint images. We refer this kind of shoeprint retrieval as scene to scene (S2S) shoeprint retrieval [20]. Gallery shoeprint images in FID-300 and CS are reference shoeprints and of high quality, and the probe shoeprint images are crime scene or scene-like images. We refer this kind of shoeprint retrieval as scene to reference (S2R) shoeprint retrieval [20].

a: MUES-SR10KS2S DATABASE

The MUES-SR10KS2S dataset consists of one gallery set, one probe set and one training set. The definitions and descriptions of the three data sets are detailed as follows.

(i) Probe set: A probe set is a collection of probe images that need to be retrieved [32], [33]. The probe set contains twelve kinds of patterns, and the total number of the probe set is 72. The probe images have been derived from different crime scenes with different image quality.

(ii) Gallery set: A gallery set is a collection of recorded shoeprint images against which a probe image is matched [32], [33]. The gallery set consists of 10096 crime scene shoeprints, which include 72 probe images, 432 synthetic versions of probe images and 9592 crime scene shoeprints.

(iii) Training set: The training set is a collection of shoeprint images used to train the opinion score prediction model. The training set contains 15 groups of crime scene shoeprints that are not contained in the gallery set. Each group has a different number of images, and the total number of shoeprint images in the training set is 100. There is one reference shoeprint in each group.

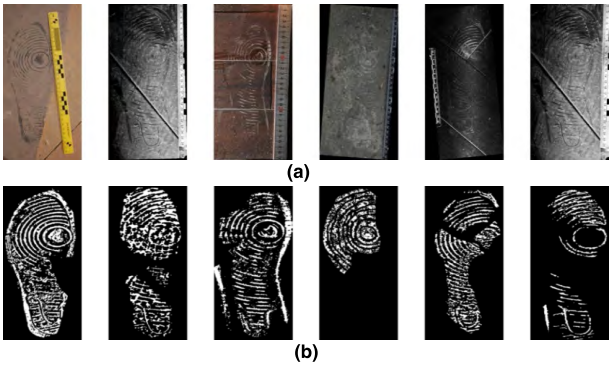

FIGURE 5. Sample shoeprints in the MUES-SR10KS2S database. (a) The probe shoeprint and its five counterparts from real crime scenes in the gallery set. (b) Corresponding binarized versions of the probe shoeprint and its five counterparts in the gallery set.

Some samples of the MUES-SR10KS2S database are shown in Figure 5. Figure 5(a) shows one probe shoeprint and its five counterparts in the gallery set, and Figure 5(b) shows their binarized and normalized versions.

b: FID-300 DATABASE

The FID-300 database contains two different kinds of shoeprint images. The first kind of shoeprint images is derived from crime scenes by forensic investigators, which are used as the query shoeprints, and the total number of the query shoeprints is 300. The second kind of shoeprint images is the reference shoeprints, which are generated by using a gelatine lifter on the outsole of the reference shoe, and then by scanning the lifters. These reference shoeprint images are used as the gallery images, and they are of very high quality. The total number of the gallery images is 1175.

Some samples of the FID-300 database are shown in Figure 6. Figure 6(a) shows one group of query shoeprints, and Figure 6(b) shows the corresponding reference shoeprints of the query shoeprints shown in Figure 6(a).

c: CS DATABASE

The CS database consists of one gallery set and four probe sets. The reference shoeprint images are used as the gallery images, and they are of very high quality. The total number of the gallery images is 100 The first kind of probe shoeprint

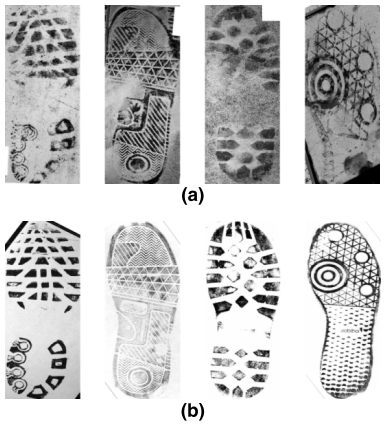

FIGURE 6. Samples of the query shoeprint and its corresponding reference shoeprints in the FID-300 database. (a) The query shoeprint from crime scenes. (b) Corresponding reference shoeprints of the query shoeprints.

images is 66 crime scene-like dust shoeprint images which are generated by using black gelatine lifters on the prints that walked by analysts after stepping in a tray of collected vacuum dust, and then by scanning the lifters. The second kind of probe shoeprint images is 53 crime scene-like blood shoeprint images which are generated by using a scanner on the crime scene-like blood impressions. The third kind of probe shoeprint images is the same 53 blood replicates enhanced by using leuco-crystal violet (LCV). The fourth kind of probe shoeprint images is 100 high quality shoeprint images. The exemplars are prepared by using an accepted Handprint method [36] and digitized by using the scanners. Samples of four kinds of query shoeprints from CS database are shown in Figure 7(c)–(f), and their corresponding reference shoeprints are shown in Figure 7(a) and Figure 7(b).

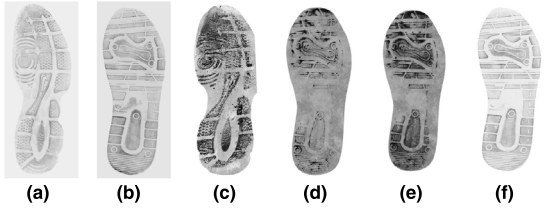

FIGURE 7. Samples of the query shoeprint and its corresponding reference shoeprints in the CS database. (a) The reference shoeprint of (c). (b) The reference shoeprint of (d), (e) and (f). (c) Dust shoeprint. (d) Blood shoeprint. (e) Blood shoeprint enhanced by LCV. (f) High quality shoeprint.

2) EVALUATION METRIC

The cumulative match characteristic curve (CMC) is often used to evaluate the performance of a retrieval algorithm operating in the closed-set identification task [34]. The CMC shows how often the query image appears in the top n matches, and its top n can be defined as follows:

$$CMC\ (n) = 100\frac{R_n}{|\mathbf{P}|} \tag{18}$$

TABLE 2. Comparisons with the state-of-the-art methods with the same affinity matrix.

Approaches	The cumulative match score of top percentage						
	0.1%	0.2%	0.3%	0.4%	0.5%	1%	2%
Zhou, et al. 2003 [22]	50.4%	70.4%	76.4%	78.4%	80.2%	87.7%	90.1%
Wang, et al. 2016 [20]	49.4%	74.4%	82.5%	85.1%	87.3%	92.7%	95.2%
Ours	**50.4%**	**78.6%**	**86.5%**	**90.3%**	**91.1%**	**94.6%**	**96.6%**

where $\mathbf{P}$, $|\mathbf{P}|$ and R_n denote the query images, the number of the query images and the number of images in the gallery set which match the query images in the top n ranked images, respectively.

B. PERFORMANCE EVALUATION

In this section, we conduct three kinds of experiments to evaluate the performance of the proposed method. First, we compare the proposed method with some state-of-the-art algorithms, whose inputs are affinity matrixes, so that we can evaluate the ranking performances under the condition that the affinity matrixes are the same. Second, comparisons are conducted on the public available databases so that we can compare the retrieval performance with methods whose results have been reported in the same public databases. The third experiment is conducted on the MUES-SR10KS2S dataset, and we compare the retrieval performance with the state-of-the-art methods.

1) COMPARISONS WITH STATE-OF-THE-ART METHODS WITH THE SAME AFFINITY MATRIX

We compare the proposed method with the works of Wang *et al.* [20] and Zhou *et al.* [22]. Zhou *et al.* [22] provides a manifold-based ranking algorithm, and the manifold regularization term of our proposed ranking cost function is inspired based on its idea. Wang *et al.* [20] also provides a manifold-based ranking algorithm, and it considers the manually labeled opinion scores in the ranking result.

In this subsection, to ensure a fair comparison, the affinity matrixes of all the methods are set to be the same, and they are computed according to the method in section III (D) on the dataset MUES-SR10KS2S. The cumulative match scores of the algorithms are listed in Table 2, and the CMC curves of the algorithms are shown in Figure 8. Figure 9 provides a visual illustration of the top 10 shoeprint images in the ranking lists of our algorithm. The results show that our proposed method outperforms state-of-the-art methods with the same affinity matrix on the MUES-SR10KS2S database, and the cumulative match score of our proposed methods reaches more than 96.6% of top 2% of the dataset.

The cumulative match scores show that the performance of our method is approximately 7.79% above the method of Zhou *et al.* [22] on average and is approximately 3.07% above the method of Wang *et al.* [20] on average. The reasons why our proposed method outperformed the compared methods are as follows: (i) Zhou *et al.* [22] provides a manifold-based ranking algorithm, but it loses sight of the roles for the examples collected from a crime scene with the query shoeprint; and (ii) Wang *et al.* [20] consider both the similarity rank order index and the manually labeled opinion scores, but neglect the effect of the mislabeled opinion scores and the contribution of the opinion scores of the neighborhood.

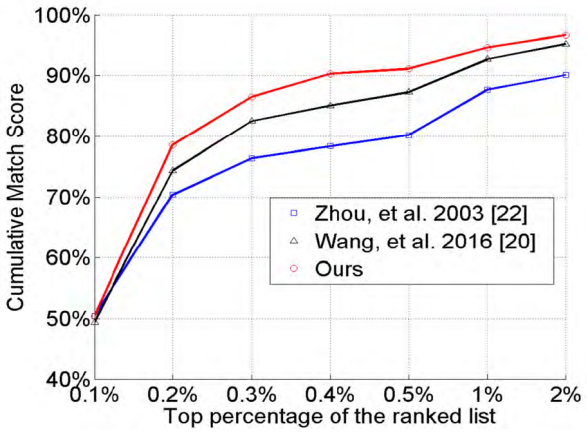

FIGURE 8. CMCs of our method and the compared methods on the MUES-SR10KS2S database (the total number of the gallery set is 10096).

2) COMPARISONS WITH STATE-OF-THE-ART METHODS ON TWO PUBLIC AVAILABLE DATABASES

To further prove the validity of our proposed method, we compared the proposed method with the works of Richetelli *et al.* [14], Wang *et al.* [20] Kortylewski and Vetter [23], Almaadeed *et al.* [28] and Kong *et al.* [15] on two public available datasets (e.g. the CS dataset [14] and the FID-300 dataset [23]). The performances of all methods are shown in Table 3 and Table 4, respectively. Among them, the results of the phase-only correlation method (POC) on the CS dataset are borrowed from Richetelli *et al.* [14], and the results of Kortylewski and Vetter [23] and Kong *et al.* [15] on the FID-300 dataset are borrowed from Figure 5 in Kong *et al.* [15], and the others are obtained by our own programs.

The results show that the performances of most of the methods fall with image quality declines on two datasets in an order: high quality > blood + LCV > blood > dust > FID-300. It indicates that the image quality can exert a great influence on the precisions of all the shoeprint retrieval methods.

The results also show that: (i) performances of interest point based methods [14] may surpass those of some fre-

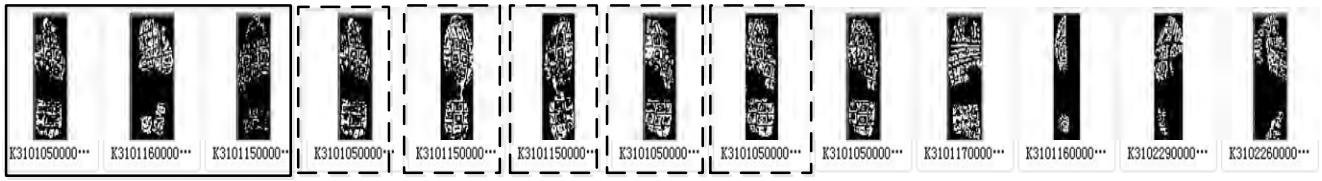

FIGURE 9. Results of our method. The top 10 in the ranking list of our method are presented. Shoeprints encircled by the solid boxes are the query shoeprint and shoeprints collected from the same crime scene, and those encircled by dotted boxes are the counterparts of the query shoeprint.

TABLE 3. Comparisons with the state-of-the-art shoeprint retrieval algorithms on the CS database.

Impression type	Method	The cumulative match score of the top percentage				
		1%	2%	3%	4%	5%
High Quality	SIFT [14]	94%	96%	98%	98%	98%
	Almaadeed et al. 2015 [28]	96%	100%	100%	100%	100%
	POC [14]	99%	99%	100%	100%	100%
	Wang, et al. 2016 [20]	99%	100%	100%	100%	100%
	Kong, et al. 2017[15]	99%	100%	100%	100%	100%
	Ours	**99%**	**100%**	**100%**	**100%**	**100%**
Dust	SIFT [14]	7.6%	15.2%	18.2%	21.2%	22.7%
	Almaadeed et al. 2015 [28]	50.0%	54.6%	63.4%	74.2%	95.5%
	POC [14]	47.0%	47.0%	51.5%	51.5%	54.5%
	Wang, et al. 2016 [20]	86.4%	93.9%	95.5%	95.5%	98.5%
	Kong, et al. 2017[15]	86.4%	89.4%	90.9%	90.9%	95.5%
	Ours	**89.4%**	**93.9%**	**95.5%**	**95.5%**	**98.5%**
Blood	SIFT [14]	3.8%	11.3%	13.2%	17.0%	18.9%
	Almaadeed et al. 2015 [28]	47.2%	52.8%	64.2%	81.1%	98.1%
	POC [14]	45.3%	50.9%	54.7%	64.2%	66.0%
	Wang, et al. 2016 [20]	86.8%	100%	100%	100%	100%
	Kong, et al. 2017[15]	92.5%	98.1%	100%	100%	100%
	Ours	**92.5%**	**100%**	**100%**	**100%**	**100%**
Blood+LCV	SIFT [14]	17.0%	22.6%	32.1%	35.9%	41.5%
	Almaadeed et al. 2015 [28]	60.4%	67.9%	73.6%	86.8%	98.1%
	POC [14]	58.5%	66.0%	71.7%	81.1%	83.0%
	Wang, et al. 2016 [20]	96.2%	100%	100%	100%	100%
	Kong, et al. 2017[15]	94.3%	100%	100%	100%	100%
	Ours	**96.2%**	**100%**	**100%**	**100%**	**100%**

TABLE 4. Comparisons with the state-of-the-art shoeprint retrieval algorithms on the FID-300 database.

Method	The cumulative match score of the top percentage				
	1%	5%	10%	15%	20%
Kortylewski, et al. 2016 [23]	22.0%	47.5%	58.0%	67.0%	71.0%
Wang, et al. 2016 [20]	68.6%	80.9%	88.7%	93.0%	96.3%
Kong, et al. 2017[15]	73.0%	82.5%	87.5%	91.0%	92.5%
SIFT [14]	8.3%	18.0%	29.0%	35.6%	37.3%
POC [14]	5.7%	14.3%	23.3%	26.3%	32.3%
Almaadeed et al. 2015 [28]	23.7%	46.0%	59.7%	69.7%	88.0%
Ours	**68.8%**	**81.3%**	**89.5%**	**93.0%**	**96.3%**

quency domain methods (e.g. POC) when there are variations on scale and orientation. For example, results on FID-300 dataset show the performances of interest point based methods surpass that of POC based method [14]; (ii) interest point based methods can achieve a similar performance to frequency-based methods on dealing with shoeprints of high quality and of the same capturing modality. For example, Table 3 shows the cumulative match scores of the top 1% of both kinds of methods are more than 94% on the high quality image subset of the CS database; and

TABLE 5. Comparisons with the state-of-the-art shoeprint retrieval algorithms on the MUES-SR10KS2S database.

Methods	Performance reported in the Literature		Performance on Our Dataset	
	Performance	Gallery set Description	The Cumulative Match Score of Top 2%	Mean Average Running Time (ms)
Kortylewski, et al. 2014 [21]	85.7%@20%	#R:1,175	38.6%	0.5696
Wang, et al. 2016 [20]	87.5%@2%	#S:10,096	95.2%	17.8238
SIFT [14]	97%@5%	#R:100	16.2%	13.3
Almaadeed et al. 2015 [28]	68.5@2.5%	#R:400	42.1%	8.7
POC [14]	100%@5%	#R:100	28.6%	0.026
Kong et al. 2017 [15]	92.5%@20%	#R:1,175	45.6%	186.0
Ours	_	_	**96.6%**	**17.903**

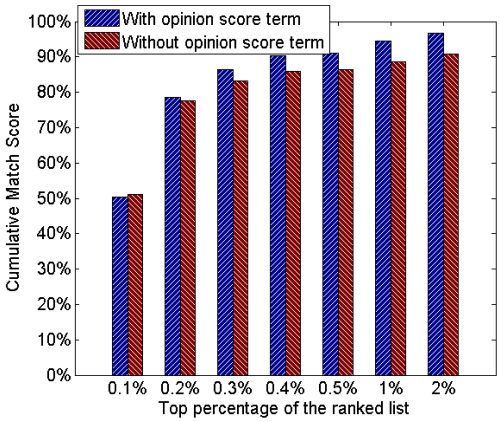

FIGURE 10. Performance of our method with and without the opinion score term.

(iii) however, performances of interest point based methods may decrease dramatically when retrieving shoeprints with low quality or of different imaging ways. For example, results on low quality image sets (e.g. Dust and Blood subset in CS dataset) show the performances of interest point based methods are much lower than those of frequency domain based methods. The possible reasons are as follows:

(i) Shoeprint images of low quality are always incomplete and interfered with fractures, complex backgrounds and other artifacts, and these factors result in many false interest points and incorrect descriptors. Furthermore, descriptors of the local corresponding patches also differ greatly because of the severe interferences. As a result, there are many wrong matching points between images. On the other hand, although shoeprints from the same shoe could differ in local appearance, their macro structural features such as orientation, periodicity of primitive pattern elements are similar; therefore, frequency domain based methods can be much more effective.

(ii) The types of both the query shoeprint and its corresponding shoeprints are different (e.g. Dust and Blood subset in CS dataset) and this cause difference in local descriptors. There are many different types, ranging from impressions of dust or blood on hard surfaces to impressions made in soft surfaces. Different types may result in difference of local characteristics (e.g. magnitude of gradient), wrong matching points therefore occur. Frequency domain based methods can capture global not local characteristics of shoeprint patterns, and they are much more robust to imaging modality than interest point based methods for cross modal shoeprint retrieval.

Almaadeed *et al.* [28] improved the performance of the interest point based method depending on two perspectives: (i) selecting region of interest (ROI) so that specific patterns can be selected without other unwanted information; and (ii) using multiple point-of-interest detectors to detect more stable interest points. The results show that method in [28] can achieve a good performance on shoeprint retrieval and be superior to the POC based method.

The results also show that our method has a good performance on the two public databases and outperforms the compared methods except for the case that the query shoeprint is a small patch that has not periodical patterns. For example, although the performance of our method on the FID-300 dataset is approximately 3.8% above that of the works of Kong *et al.* [15] in the top 20% of the ranked list, the cumulative match score of our algorithm does not surpass that of Kong *et al.* [15] in the top 1% and 5% of the ranking list. The main reason is that some of the probe images are of small size. Kong *et al.* [15] use a template matching method to search over both translations (with a stride of 2) and rotations, and the template matching method can work well for small patch retrieval

3) COMPARISONS WITH STATE-OF-THE-ART METHODS ON THE MUES-SR10KS2S DATABASE

We also compare our method with some other shoeprint retrieval methods on the MUES-SR10KS2S database. The results of the cumulative match score of the top 2% and the mean average running time are shown in Table 5. Because some original papers do not release codes, the results in Table 5 are obtained by running our own implemented codes according to the original papers. The results of

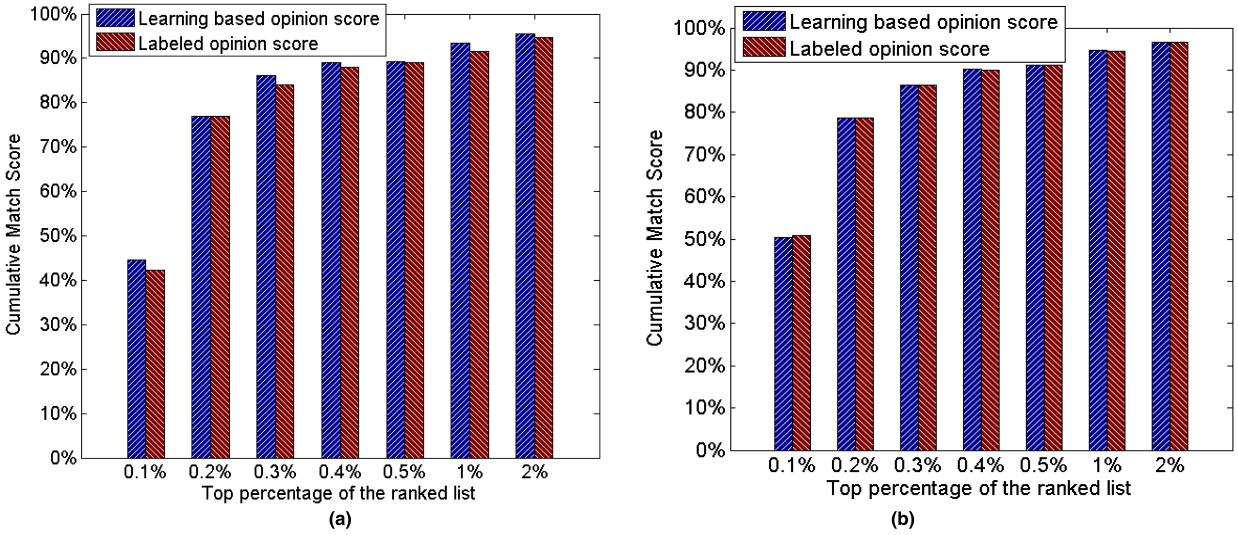

FIGURE 11. Performance of methods with two kinds of opinion scores in two scenarios. (a) Performance of incorrect labeled opinion scores. (b) Performance of correct labeled pinion scores.

Kong *et al.* [15] are obtained by running their published codes.

The results show that our method outperforms the state-of-the-art methods. The reasons are as follows: (i) the holistic and regional appearance-based features used in our method can work well on representing different modality or degraded shoeprint images; and (ii) we take into consideration forensic experts' opinions, and tell the algorithm what the investigator would prefer by assigning opinion scores to shoeprint examples.

Our hardware configuration consists of a 3.33-GHz CPU with 8-GB RAM. All of the compared methods are implemented with MATLAB codes. To compare the running times, the mean average running time is defined and applied.

$$MAT = \sum_{i=1}^{n_q} T(i)/n_q K \tag{19}$$

where i denotes the ith query shoeprint, $T(i)$ represents the running time of the ith query, n_q denotes the number of query shoeprints, and K is the number of the dataset shoeprints. Since the affinity matrix and the parameters of the prediction model used in these methods are pre-computed offline, their running times are not added.

C. ANALYSIS AND DISCUSSION

In this section, we further analyze the effect of different components of the cost function on the ranking result, which includes the roles of the opinion score, the learning-based opinion scores, the coefficient matrix A and the neighbors of multiple examples. The influence caused by the number of multiple examples is also discussed. Finally, we also verify the convergence of the proposed method.

1) EFFECTIVENESS OF THE OPINION SCORES

In the proposed method, we aim at using the opinion scores to guide shoeprint retrieval results. In (1), the third term is the opinion score correlation term that associates a cost for the degree of deviation from the opinion scores. To verify the effectiveness of the opinion scores, we conduct experiments to compute the ranking scores in two scenarios, which include the opinion score correlation term and exclude it. The performances of the experiments are shown in Figure 10. The experimental results show that opinion scores can improve the effectiveness of shoeprint retrieval.

2) EFFECTIVENESS OF THE LEARNING-BASED OPINION SCORES

To verify the effectiveness of the learning-based opinion scores for the proposed method, we conduct experiments to compute the ranking scores by using two kinds of opinion scores. The first kind of opinion score is the labeled opinion score labeled by the experts. The second kind of opinion score is the proposed learning-based opinion score computed according to (4). Comparisons between the two kinds of opinion scores are conducted in two scenarios. One scenario involves all of the labeled opinion scores assigned by skilled forensic science experts, and we refer to this kind of opinion score as the correctly labeled opinion scores. The other scenario involves the opinion scores that are incorrectly labeled opinion scores.

Figure 11 shows the results of the quantitative comparisons in the scenarios in which the examples are labeled correctly or incorrectly. The results show that the method with the learning-based opinion scores achieves a higher performance than the one using only incorrect labeled opinion scores, and they also show a similar performance to the case that using correct labeled opinion scores.

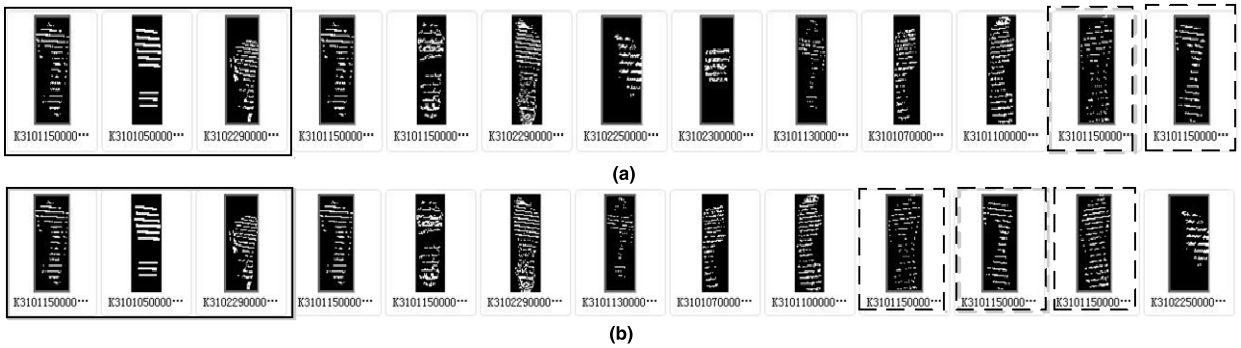

FIGURE 12. Results of the scenario in which the examples are labeled incorrectly. Shoeprints encircled by the solid boxes are the query shoeprint and shoeprint examples collected from the same crime scene, and those encircled by dotted boxes are its counterparts in the dataset. (a) Labeled opinion score. (b) Learning-based opinion score.

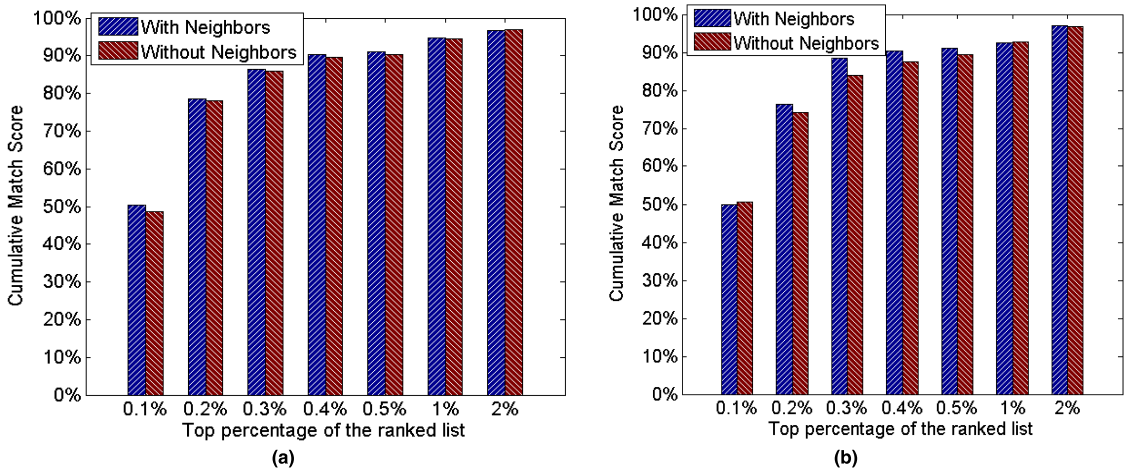

FIGURE 13. Performance of our method with and without neighbors in two scenarios. (a) Performance of our method with coefficient matrix A. (b) Performance of our method without coefficient matrix A.

Figure 12 shows a visual example of the top 10 shoe print images in the ranking lists in the scenario where examples are labeled incorrectly. The results also show that the proposed method is robust to wrong-labeled opinion scores.

3) EFFECTIVENESS OF THE NEIGHBORS OF MULTIPLE EXAMPLES AND THE COEFFICIENT MATRIX **A**

In the proposed method, we consider the contribution of the opinion scores of the neighbors, and introduce the coefficient matrix **A** to prevent the ranking scores tendency to become low values. In this section, we conduct three kinds of experiments to show how the neighbors of multiple examples and the coefficient matrix **A** affect the effectiveness of the retrieval.

The first kind of experiment is to show the effectiveness of neighbors of multiple examples. The performances of the experiments are shown in Figure 13. The experimental results show that the neighbors of multiple examples can improve the effectiveness of the shoeprint retrieval.

The second kind of experiment shows the effectiveness of the coefficient matrix **A**. The performances of the experiments are shown in Figure 14. The experimental results show that the coefficient matrix **A** can improve the effectiveness of shoeprint retrieval.

The third kind of experiment shows the effectiveness of the neighbors of multiple examples and the coefficient matrix **A**. The comparison results are shown in Figure 15. The results show that methods with **A** and neighbors significantly outperforms methods without **A** and neighbors. The results also demonstrate the effectiveness of the proposed method.

4) INFLUENCE OF EXAMPLES FROM EACH SIMILARITY LEVEL

In this section, we conduct experiments to show how the examples from different similarity levels affect the effectiveness of the shoeprint retrieval.

We conduct six experiments, and we only select one example in descending similarity order to obtain a pair with the query shoeprint in each experiment. The results are listed in Table 6 and shown in Figure 16. The experimental results show that examples from each similarity level can improve the effectiveness of shoeprint retrieval. The examples that have higher similarity with the query shoeprint can obtain a better performance.

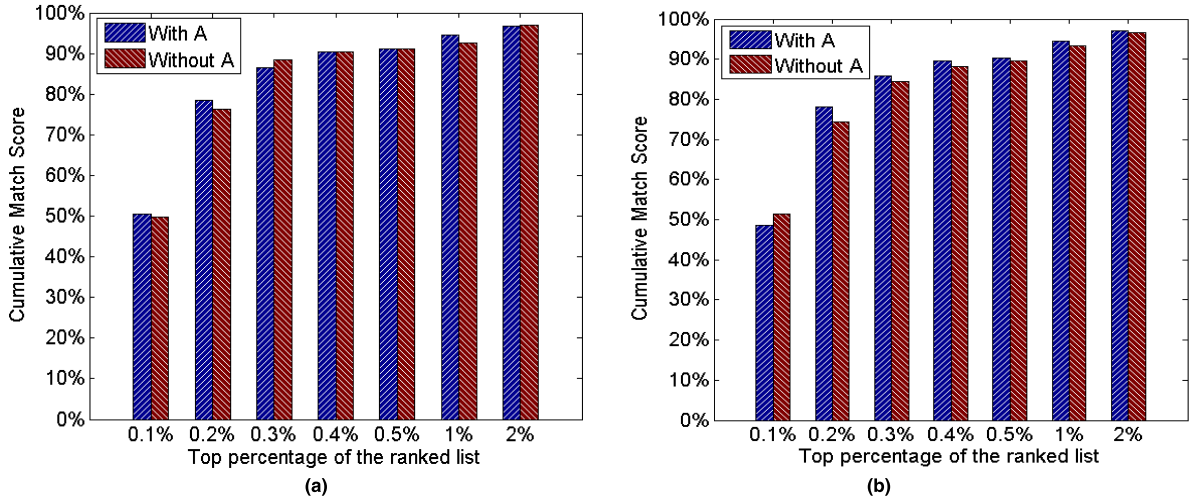

FIGURE 14. Performance of our method with and without the coefficient matrix A in two situations. (a) Performance of our method with neighbors. (b) Performance of our method without neighbors.

TABLE 6. Comparison of the performance of examples from different similarity levels.

Numbers of examples	The cumulative match score of top percentage						
	0.1%	0.2%	0.3%	0.4%	0.5%	1%	2%
Without example	50.6%	76.0%	81.7%	83.7%	85.7%	88.1%	90.9%
Only Extremely High level	53.4%	78.6%	86.5%	88.5%	90.1%	92.7%	94.2%
Only High level	51.6%	78.4%	83.9%	87.1%	88.9%	92.3%	94.0%
Only Medium level	50.8%	76.8%	84.3%	87.1%	88.3%	90.9%	94.0%
Only Low level	50.6%	76.6%	81.7%	84.3%	86.7%	90.1%	93.8%
Only Extremely Low level	50.8%	76.2%	81.9%	83.9%	85.7%	88.1%	90.9%

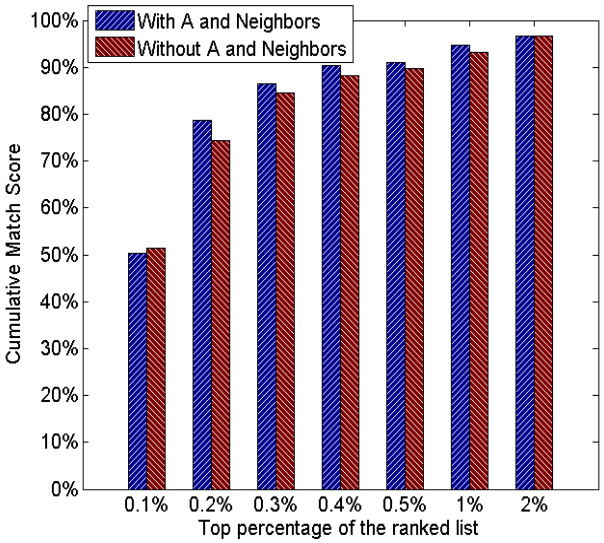

FIGURE 15. Performance of our method with and without the coefficient matrix A and neighbors.

5) INFLUENCE OF THE NUMBER OF MULTIPLE EXAMPLES

In this section, we conduct three kinds of experiments to show how the number of examples affects the effectiveness of the retrieval.

The first kind of experiment shows the performances of the different number of examples from different similarity levels. We conduct six experiments, and we append one example in descending similarity order to $\mathbf{Q}$ in each experiment. The initial element of $\mathbf{Q}$ is the query image.

The results are listed in Table 7 and shown in Figure 17. The experimental results show that increasing the number of query examples of different similarity levels can improve the effectiveness of shoeprint retrieval.

The second kind of experiment is performed to show the performances of different numbers of examples from the same similarity levels. For each similarity level, we conduct three experiments. The initial element of $\mathbf{Q}$ is the query image; then, we append one more example from the same similarity level to $\mathbf{Q}$ in each experiment. Here, we list the performance of examples from the extremely high level in Table 8. The results show that adding one example to $\mathbf{Q}$ can significantly improve shoeprint retrieval effectiveness, but that adding one more example can only offer a slight improvement. We think that the two examples have a closer distance in the semantic space, and do not complement one another. When we conduct shoeprint retrieval, the two examples may have similar effects on guiding shoeprints. To verify this result, we conduct the third kind of experiment with the

TABLE 7. Comparison of the performance of different numbers of examples.

Numbers of examples	The cumulative match score of top percentage						
	0.1%	0.2%	0.3%	0.4%	0.5%	1%	2%
Without example	50.6%	76.0%	81.7%	83.7%	85.7%	88.1%	90.9%
One example from Extremely High level	53.4%	78.6%	86.5%	88.5%	90.1%	92.7%	94.2%
One more example from High level	53.6%	78.8%	86.7%	89.9%	91.1%	93.8%	95.0%
One more example from Medium level	52.4%	78.8%	86.5%	89.9%	90.5%	94.4%	96.8%
One more example from Low level	50.4%	78.6%	86.5%	90.3%	91.1%	94.6%	96.6%
One more example from Extremely Low level	50.4%	78.6%	86.5%	90.3%	91.1%	94.6%	96.6%

TABLE 8. Comparison of the performance of the proposed method with different numbers of examples from the same similarity level.

Numbers of examples	The cumulative match score of top percentage						
	0.1%	0.2%	0.3%	0.4%	0.5%	1%	2%
Without example	50.6%	76.0%	81.7%	83.7%	85.7%	88.1%	90.9%
One-example	53.4%	78.6%	86.5%	88.5%	90.1%	92.7%	94.2%
Two-example	53.4%	78.4%	86.5%	88.9%	90.3%	92.8%	94.2%

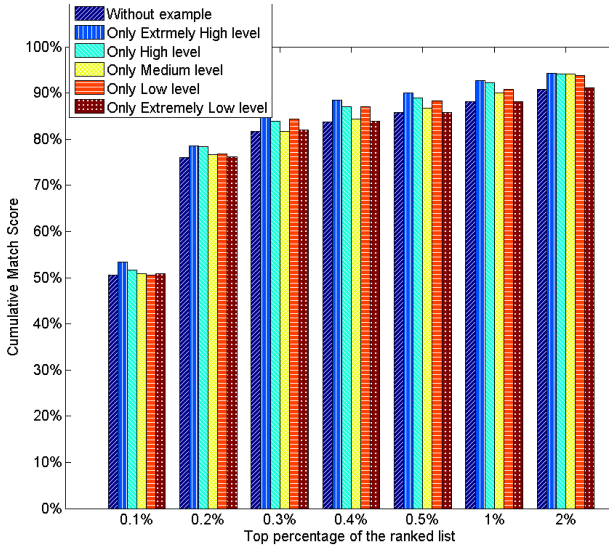

FIGURE 16. Performance of the proposed method with examples from different similarity levels.

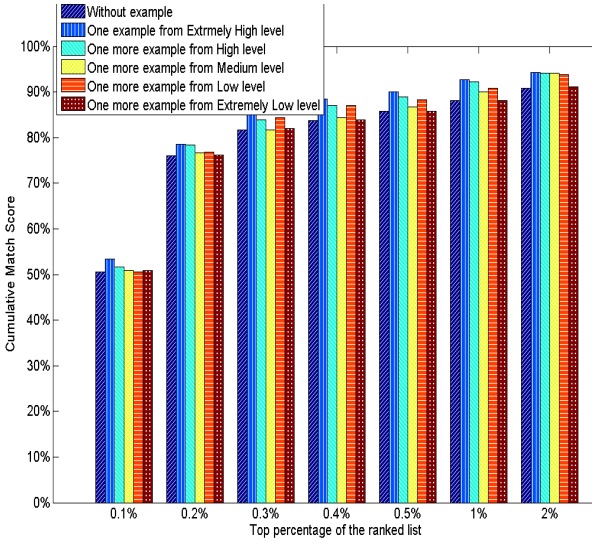

FIGURE 17. Performance of the proposed method with different numbers of examples.

two examples that complement one another from the same similarity level. The two examples are partial shoeprints generated from two full prints. The first example only contains the bottom region, and the second example only contains the top region. We list the results in Table 9. The results show that adding two examples which complement one another from the same similarity level to $\mathbf{Q}$ can significantly improve the shoeprint retrieval effectiveness.

$$MAO\,(t) = \frac{\sum_{i=1}^{n_q} Q_{(i)}\left(f^{(t)}\right)}{n_q K} \tag{20}$$

$$MAR\,(t) = \frac{\sum_{i=1}^{n_q} \left\| \left(\alpha \mathbf{R}_{(i)} + \beta \mathbf{L}_{(i)} + \gamma \mathbf{A}_{(i)}\right) f^{(t)} - \alpha \mathbf{R}_{(i)} \mathbf{y} - \gamma \mathbf{A}_{(i)} \hat{\mathbf{y}} \right\|_2}{n_q K} \tag{21}$$

TABLE 9. Comparison of the performance of the proposed method with different numbers of examples which complement one another from the same similarity level.

Numbers of examples	The cumulative match score of top percentage						
	0.1%	0.2%	0.3%	0.4%	0.5%	1%	2%
Without example	50.6%	76.0%	81.7%	83.7%	85.7%	88.1%	90.9%
One-example	50.0%	77.2%	83.1%	84.9%	86.9%	89.5%	92.5%
Two examples	51.4%	78.6%	84.7%	86.5%	88.7%	91.7%	93.7%

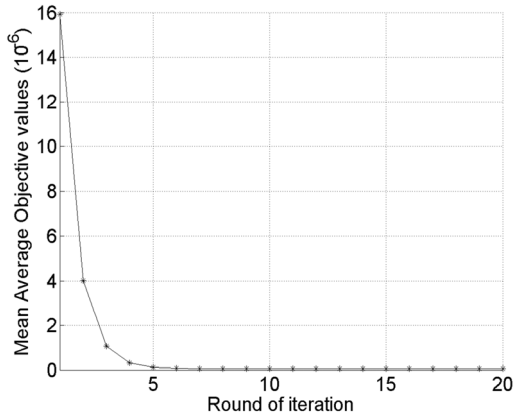

FIGURE 18. Convergence curves of the mean average objective function values of our algorithm. The figure shows that the MAO monotonically decreases until convergence within a few iterations.

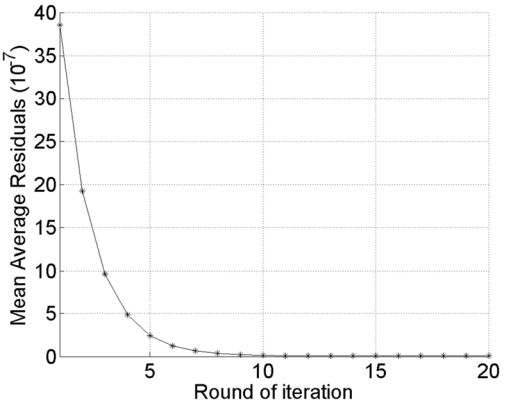

FIGURE 19. Convergence curves of the mean average residual values of our algorithm. The figure shows that the MAR monotonically decreases until convergence within a few iterations.

6) CONVERGENCE STUDY

In this section, we analyze the convergence behavior of the proposed method on the MUES-SR10KS2S database.

For analyzing the convergence of the proposed method, the mean average objective values (MAO) and the mean average residuals (MAR) are defined and applied. (20) and (21), as shown at the bottom of the previous page, where i represents the ith probe image, t denotes the iteration number and n_q is the number of probe images.

Figure 18 and Figure 19 show the convergence curves of the proposed method with regard to the MAO and MAR. In both figures, the abscissae represent the rounds of the iterations while the ordinates are the MAO values or MAR values for each round. The results show that the proposed method converges within 10 iterations.

V. CONCLUSION AND FUTURE WORK

We proposed a learned opinion score guided shoeprint retrieval (LOSGSR) method for shoeprint retrieval. The method improves the effectiveness of the shoeprint retrieval depending on four perspectives: (1) use the opinion scores of the multiple shoeprint examples to guide the ranking scores to meet the forensic experts' expectations; (2) propose a learning-based method to refine opinion scores, which corrects the labeled opinion scores of multiple examples and their neighbors; (3) take into account the feature similarity between the query and dataset shoeprint images; and (4) introduce a coefficient matrix to prevent the tendency of the computed ranking scores from becoming low. The experiments on the real crime scene datasets show that the cumulative match score of the proposed algorithm in the top 2% of the dataset is more than 96.6%, and it is not only significantly higher than the traditional manifold ranking method, but also higher than the state-of-the-art shoeprint retrieval algorithms.

Our future work will involve taking into account the relationship between every two shoeprint clusters.

REFERENCES

[1] X. Wang, H. Sun, Q. Yu, and C. Zhang, "Automatic shoeprint retrieval algorithm for real crime scenes," in *Proc. Asian Conf. Comput. Vis.*, Singapore, Nov. 2014, pp. 399–413. doi: 10.1007/978-3-319-16865-4_26.
[2] A. Bouridane, A. Alexander, M. Nibouche, and D. Crookes, "Application of fractals to the detection and classification of shoeprints," in *Proc. Int. Conf. Image Process.*, Vancouver, BC, Canada, Sep. 2000, pp. 474–477. doi: 10.1109/ICIP.2000.900998.
[3] G. Algarni and M. Amiane, "A novel technique for automatic shoeprint image retrieval," *Forensic Sci. Int.*, vol. 181, nos. 1–3, pp. 10–14, 2008. doi: 10.1016/j.forsciint.2008.07.004.
[4] C.-H. Wei and C.-Y. Gwo, "Alignment of core point for shoeprint analysis and retrieval," in *Proc. Int. Conf. Inf. Sci. Electron. Elect. Eng. (ISEEE)*, Sapporo, Japan, Apr. 2014, pp. 1069–1072. doi: 10.1109/InfoSEEE.2014.6947833.
[5] C.-Y. Gwo and C.-H. Wei, "Shoeprint retrieval: Core point alignment for pattern comparison," *Sci. Justice*, vol. 56, no. 5, pp. 341–350, 2016. doi: 10.1016/j.scijus.2016.06.004.
[6] P. M. Patil and J. V. Kulkarni, "Rotation and intensity invariant shoeprint matching using Gabor transform with application to forensic science," *Pattern Recognit.*, vol. 42, no. 7, pp. 1308–1317, 2009. doi: 10.1016/j.patcog.2008.11.008.
[7] P. D. Chazal, J. Flynn, and R. B. Reilly, "Automated processing of shoeprint images based on the Fourier transform for use in forensic science," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 341–350, Mar. 2005. doi: 10.1109/TPAMI.2005.48.
[8] M. Gueham, A. Bouridane, and D. Crookes, "Automatic classification of partial shoeprints using advanced correlation filters for use in forensic science," in *Proc. Int. Conf. Pattern Recognit.*, Tampa, FL, USA, Dec. 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4761058.
[9] M. Gueham, A. Bouridane, and D. Crookes, "Automatic recognition of partial shoeprints based on phase-only correlation," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, Sep./Oct. 2007, pp. IV-441–IV-444. doi: 10.1109/ICIP.2007.4380049.

[10] F. Cervelli, F. Dardi, and S. Carrato, "An automatic footwear retrieval system for shoe marks from real crime scenes," in *Proc. 6th Int. Symp. Image Signal Process. Anal.*, Salzburg, Austria, Sep. 2009, pp. 668–672. doi: 10.1109/ISPA.2009.5297667.

[11] F. Cervelli, F. Dardi, and S. Carrato, "A texture based shoe retrieval system for shoe marks of real crime scenes," in *Proc. Int. Conf. Image Anal. Process.*, Catania, Italy, 2009, pp. 384–393. doi: 10.1007/978-3-642-04146-4_42.

[12] F. Cervelli, F. Dardi, and S. Carrato, "A translational and rotational invariant descriptor for automatic footwear retrieval of real cases shoe marks," in *Proc. 18th Eur. Signal Process. Conf.*, Aalborg, Denmark, Aug. 2010, pp. 1665–1669.

[13] S. Alizadeh and C. Kose, "Automatic retrieval of shoeprint images using blocked sparse representation," *Forensic Sci. Int.*, vol. 277, pp. 103–114, Aug. 2017. doi: 10.1016/j.forsciint.2017.05.025.

[14] N. Richetelli, M. C. Lee, C. A. Lasky, M. E. Gump, and J. A. Speir, "Classification of footwear outsole patterns using Fourier transform and local interest points," *Forensic Sci. Int.*, vol. 275, pp. 102–109, Jun. 2017. doi: 10.1016/j.forsciint.2017.02.030.

[15] B. Kong, J. Supancic, D. Ramanan, and C. Fowlkes, "Cross-domain forensic shoeprint matching," in *Proc. 28th Brit. Mach. Vis. Conf. (BMVC)*, London, U.K., 2017, pp. 1–5.

[16] Y. Tang, S. N. Srihari, and H. Kasiviswanathan, "Similarity and clustering of footwear prints," in *Proc. IEEE Int. Conf. Granular Comput.*, San Jose, CA, USA, Aug. 2010, pp. 459–464. doi: 10.1109/GrC.2010.175.

[17] Y. Tang, S. N. Srihari, H. Kasiviswanathan, and J. J. Corso, "Footwear print retrieval system for real crime scene marks," in *Proc. Int. Conf. Comput. Forensics*, Tokyo, Japan, 2010, pp. 88–100. doi: 10.1007/978-3-642-19376-7_8.

[18] M. Pavlou and N. M. Allinson, "Automated encoding of footwear patterns for fast indexing," *Image Vis. Comput.*, vol. 27, no. 4, pp. 402–409, 2009. doi: 10.1016/j.imavis.2008.06.003.

[19] M. Pavlou and N. M. Allinson, "Automatic extraction and classification of footwear patterns," in *Proc. 7th Int. Conf. Intell. Data Eng. Automated Learn.*, Burgos, Spain, 2006, pp. 721–728. doi: 10.1007/11875581_87.

[20] X. Wang, C. Zhan, Y. Wu, and Y. Shu, "A manifold ranking based method using hybrid features for crime scene shoeprint retrieval," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21629–21649, 2016. doi: 10.1007/s11042-016-4029-3.

[21] A. Kortylewski, T. Albrecht, and T. Vetter, "Unsupervised footwear impression analysis and retrieval from crime scene data," in *Proc. Asian Conf. Comput. Vis.*, Singapore, 2014, pp. 644–658. doi: 10.1007/978-3-319-16628-5_46.

[22] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. Cambridge, MA, USA: MIT Press, 2003, pp. 169–176.

[23] A. Kortylewski and T. Vetter, "Probabilistic compositional active basis models for robust pattern recognition," in *Proc. 27th Brit. Mach. Vis. Conf. (BMVC)*, York, U.K., 2016, pp. 1–12.

[24] O. Nibouche, A. Bouridane, D. Crookes, M. Gueham, and M. Laadjel, "Rotation invariant matching of partial shoeprints," in *Proc. Int. Mach. Vis. Image Process. Conf.*, Dublin, Ireland, Sep. 2009, pp. 94–98. doi: 10.1109/IMVIP.2009.24.

[25] D. Crookes, A. Bouridane, H. Su, and M. Gueham, "Following the footsteps of others: Techniques for automatic shoeprint classification," in *Proc. 2nd NASA/ESA Conf. Adapt. Hardw. Syst.*, Edinburgh, U.K., Aug. 2007, pp. 67–74. doi: 10.1109/AHS.2007.56.

[26] H. Su, D. Crookes, A. Bouridane, and M. Gueham, "Local image features for shoeprint image retrieval," in *Proc. Brit. Mach. Vis. Conf.* Coventry, U.K.: Univ. Warwick, 2007, pp. 1–10.

[27] M. P. Deshmukh and P. M. Patil, "Automatic shoeprint matching system for crime scene investigation," *Int. J. Comput. Sci. Commun. Technol.*, vol. 2, no. 1, pp. 281–287, 2009.

[28] S. Almaadeed, A. Bouridane, D. Crookes, and O. Nibouche, "Partial shoeprint retrieval using multiple point-of-interest detectors and SIFT descriptors," *Integr. Comput. Aided Eng.*, vol. 22, no. 1, pp. 41–58, 2015. doi: 10.3233/ICA-140480.

[29] H. Wang, J. Fan, and Y. Li, "Research of shoeprint image matching based on SIFT algorithm," *J. Comput. Methods Sci. Eng.*, vol. 16, no. 2, pp. 349–359, 2016. doi: 10.3233/JCM-160622.

[30] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011. doi: 10.1109/TIP.2011.2131660.

[31] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems*, vol. 16, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. Cambridge, MA, USA: MIT Press, 2003, pp. 321–328. doi: 10.1007/s11222-007-9033-z.

[32] P. J. Phillips, P. Grother, and R. Micheals, "Evaluation methods in face recognition," in *Handbook Face Recognition*, K. J. Anil and Z. Li Stan, Eds. New York, NY, USA: Springer, 2005, pp. 328–348. doi: 10.1007/0-387-27257-7_15.

[33] W. Gao *et al.*, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008. doi: 10.1109/TSMCA.2007.909557.

[34] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000. doi: 10.1109/34.879790.

[35] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996. doi: 10.1109/83.506761.

[36] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, and W. Bodziak, "Quantifying randomly acquired characteristics on outsoles in terms of shape and position," *Forensic Sci. Int.*, vol. 266, pp. 399–411, Sep. 2016. doi: 10.1016/j.forsciint.2016.06.012.

YANJUN WU received the B.S. degree in electronic and information engineering from Dalian Nationalities University, Dalian, China, in 2013. He is currently pursuing the Ph.D. degree with the Information Science and Technology College, Dalian Maritime University (DMU).

His current research interests include machine learning, pattern recognition, and biometric verification.

XINNIAN WANG received the B.S. and M.S. degrees in computer science and technology from Liaoning Normal University, Dalian, China, in 2000, and the Ph.D. degree in communication and information system from Dalian Maritime University, Dalian, in 2005, where he is currently an Associate Professor and the Ph.D. Adviser of information and communication engineering.

His current research interests include forensic image processing, pattern recognition, and biometrics.

NAMUSISI LINDA NANKABIRWA received the B.S. degree in software engineering from Makerere University, Kampala, Uganda, in 2016. She is currently pursuing the master's degree with the Information Science and Technology College, Dalian Maritime University.

Her current research interests include machine learning and pattern recognition.

TAO ZHANG received the B.S. degree in electronic engineering from Liaoning Normal University, Dalian, China, in 1999, the M.S. degree in transportation information engineering and control from Dalian Jiaotong University, Dalian, China, in 2001, and the Ph.D. degree in communication and information system from Dalian Maritime University, Dalian, China, in 2015.

She is currently a Senior Engineer and Research Supervisor with Liaoning Normal University, China. Her current research interests include image processing, pattern recognition, and biometrics.