SIFT Keypoint Removal via Directed Graph Construction for Color Images

Yuanman Li, *Student Member, IEEE,* Jiantao Zhou, *Member, IEEE,* and An Cheng

***Abstract*—As one of the most successful feature extraction algorithms, *Scale Invariant Feature Transform* (SIFT) has been widely employed in many applications. Recently, the security of SIFT against malicious attack has been attracting increasing attention, and several techniques have been devised to remove SIFT keypoints intentionally. However, most of the existing methods still suffer from the following three problems: 1) the *Keypoint Removal Rate* (KRR) achieved by many techniques is unsatisfactory when removing keypoints within multiple octaves; 2) noticeable artifacts are introduced in the processed image, especially in those highly textured regions; 3) the color information is totally neglected, precluding the widespread adoption of those methods. To tackle these challenges, in this work, we propose a novel SIFT keypoint removal framework. By modeling the *Difference of Gaussian* (DoG) space as a directed weighted graph, we derive a set of strict inequality constraints to remove a SIFT keypoint along a pre-constructed acyclic path. To minimize the incurred distortion, the path is strategically designed over the directed graph. Furthermore, we propose a simple yet effective optimization framework for recovering the color information of the keypoint-removed image. Extensive experiments are provided to show the superior performance of our proposed scheme over the state-of-the-art techniques, in both the scenarios of removing keypoints in a single octave and in multiple octaves.**

***Index Terms*—SIFT, keypoint removal, directed graph, convex optimization**

I. Introduction

SIFT has been proven to be a powerful instrument in many pattern recognition and multimedia security systems [1]–[9]. Due to its excellent robustness against noise, illumination, partial occlusion and geometric transformation, SIFT feature has been widely used in *Content Based Image Retrieval* (CBIR) [3], [10], [11], 3D scene modeling [12], [13] and face recognition [14], [15]. Besides, SIFT also plays a vital role for providing evidences for decision-makers in multimedia security systems, e.g., copy-move forgery detection systems [6]–[9], [16], [17]. Unfortunately, some recent studies demonstrated that the SIFT-based systems may not be trustworthy without precluding the existence of counter-forensics, which is capable of removing or modifying SIFT keypoints [18]–[26]. Such counter-forensics brings a big threat to the reliability and security of those systems built upon SIFT. Apparently, if the extracted SIFT features are not reliable, the decisions made by the associated systems could be groundless. A telling example is that, in the copy-move forgery detection scenario, one malicious attacker may inhibit the SIFT keypoints in the cloned regions intentionally, for the purpose of malfunctioning the copy-move forgery detection systems built upon SIFT [23].

Hsu *et al.* [18] pioneered the study on security issues of SIFT. They proposed to hide the SIFT keypoint by duplicating another local extremum in the detection region, thus violating the keypoint generation condition. This technique was later proved not enough to be a threat against SIFT based systems, because new SIFT keypoints would be generated around the original ones, which can still be matched with a high probability [27]. This is the so-called *New Keypoint Generation* (NKG) problem. In [19], Do *et al.* suggested two methods. The first one called *Removal with Minimum local Distortion* (RMD) tried to erase a SIFT keypoint by lowering its contrast value below the predefined contrast threshold [19]. It was shown that RMD leads to severely affected visual quality of the resulting image. In addition, the NKG problem is still unsolved. The second one called *Global Smoothing and Local Smoothing* (GSLS) attempted to suppress the SIFT keypoints through globally and locally smoothing the image [19]. Nevertheless, GSLS tends to over-smooth the processed image, especially in those highly textured regions. Instead of removing SIFT keypoints, the technique designed in [20] aimed to modify the SIFT descriptors by changing the orientations of the associated SIFT keypoints, making the manipulated descriptors difficult to be matched. By studying different removal strategies, Amerini *et al.* [23] suggested a SIFT keypoint removal technique named *CLassification-Based Attack* (CLBA), which first classified the SIFT keypoints according to the grayscale histogram of the local patch centered at each keypoint, and then applied a different removal strategy for each class. However, the achieved gain was limited.

To improve the removal performance of [18], Lu and Hsu [21] designed an optimization framework, which partially solved the NKG problem. However, the distortion of the resulting image is still large. The reason for its degraded performance is two fold: 1) the constraints incorporated into their framework are too strict, which seriously narrows the solution space, causing large distortion; 2) the constraints in their framework are based on equality, namely, to create two equal extrema. Those equality constraints satisfied at preceding removal procedures are fragile to be broken when removing the subsequent ones, especially when multiple keypoints are highly clustered. As a result, the removed keypoints could reappear with a high probability. In [25], [26], we proposed a

This work was supported in part by the Macau Science and Technology Development Fund under grants FDCT/046/2014/A1, FDCT/022/2017/A1, in part by the Research Committee at the University of Macau under grants MYRG2015-00056-FST, MYRG2016-00137-FST, and in part by the National Science Foundation of China under Grant 61402547.

Yuanman Li and Jiantao Zhou are with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. e-mails: yb57410@umac.mo, jtzhou@umac.mo

An Cheng is with Meitu, Inc. e-mail: ca@meitu.com

(Corresponding author: Jiantao Zhou, email: jtzhou@umac.mo)

SIFT keypoint removal strategy called *Removal via Convex Relaxation* (RCR), where the removal problem was more formally formulated into an optimization framework, and the NKG problem was alleviated by suppressing new keypoints within a local cuboid of the scale space. Nevertheless, as will be discussed in Section II-B, the constraints designed in [25], [26] are *not* strict inequality based; in many cases, the optimal solution is achieved when the inequality constraints are satisfied with equality, making RCR suffer from the same problem of [21] to some extent. Besides, noticing that the optimal solution is often achieved at the boundary of the solution space, it is hard to maintain the constraints satisfied at preceding removal procedures against the inevitable interferences among different removals, especially when the keypoints are tightly clustered. As will be clear shortly, though RCR [25], [26] outperforms the other existing methods when removing SIFT keypoints in a single octave, its *Keypoint Removal Rate* (KRR) performance in the case of removing keypoints within multiple octaves is still limited and can be significantly improved.

Due to their unsatisfactory performance when removing SIFT keypoints within multiple octaves, the existing SIFT keypoint removal techniques are not sufficient to be a practical threat, as most of the existing systems extract SIFT features within multiple octaves. In this work, we propose a novel optimization-based SIFT keypoint removal algorithm, capable of achieving superior performance in both the scenarios of removing keypoints in a single octave and in multiple octaves. Specifically, we first strategically model the *Difference of Gaussian* (DoG) space as a directed weighted graph, and then derive a set of novel strict inequality constraints to remove a SIFT keypoint along a pre-constructed acyclic path. These constraints can also guarantee that no new extrema are generated in a local cuboid in the scale space, significantly alleviating the NKG problem. In addition, to suppress the incurred distortion, we formulate the problem of constructing the acyclic path as finding the shortest path over the pre-constructed graph. Furthermore, an effective optimization model is established to restore the color information of the keypoint-removed image. Extensive experiments are provided to show the superior performance of our proposed method over the state-of-the-art techniques. As an independent component, our proposed removal framework can be readily integrated with the existing SIFT keypoint injection algorithms, such as FMD [19] and TPKI [26]. We also demonstrate that the integrated removal and injection attack strategy can defeat the powerful keypoint removal/injection detector with respect to multiple octaves [28]. The main contributions of our work can be summarized as follows:

- We model the DoG space as a directed graph, based on which the appropriately selected acyclic paths are used to construct the constraints for removing SIFT keypoints.
- We design novel constraints based on strict inequalities, which are further tightened to prevent the optimal solution from being achieved at the boundary of the solution space. This makes the resulting removal algorithm much more robust against the interferences among different removals. As a result, our method achieves much better performance even in the scenario of removing keypoints in multiple octaves.
- We suggest an optimization framework to restore the color information of the keypoint-removed image. As far as we know, this is the first work to specifically address the problem of recovering the color information.
- We demonstrate that the proposed method is capable of defeating the SIFT-based detection systems, even when the keypoints are extracted from multiple octaves.

The rest of this paper is organized as follows. Section II briefly introduces SIFT and our previous work RCR [25], [26]. Section III and Section IV describe the proposed SIFT keypoint removal framework. The color restoration technique for the cleaned image is detailed in Section V. Section VI reports extensive experimental results to validate the performance of our proposed scheme. A case study of defeating a SIFT-based copy-move forgery detection system is given in Section VII and we finally conclude in Section VIII.

II. A BRIEF REVIEW OF SIFT AND RCR

In this section, we briefly review the SIFT algorithm, and our previous work, i.e., RCR [25], [26].

A. Introduction to SIFT

In a nutshell, SIFT algorithm can be roughly summarized into four phases: i) candidate keypoints identification via extrema detection in the scale space; ii) keypoints refinement according to a contrast threshold and an edge threshold; iii) dominant orientation assignment of the survived keypoints; and iv) feature descriptor generation.

According to the original SIFT paper [1], the input color image is converted to the grayscale version (denoted as $\mathbf{I}$) prior to the feature extraction. Fig. 1 shows the construction of the scale space. The initial image $\mathbf{I}_v$ for a specific octave v is obtained through resizing the grayscale image $\mathbf{I}$ by a factor ρ ($\rho = 2^v$). Mathematically,

$$\mathbf{I}_v(x,y) = \mathbf{I}(\lceil \rho \cdot x \rceil, \lceil \rho \cdot y \rceil); x \in \big[1, \frac{M}{\rho}\big] \cap \mathcal{Z}, y \in \big[1, \frac{N}{\rho}\big] \cap \mathcal{Z}.$$

Here, M and N are respectively the numbers of rows and columns of $\mathbf{I}$.

In the phase i), a set of successive Gaussian-blurred images $L_{\mathbf{I}}(x,y,\sigma)$ are produced by convolving $\mathbf{I}_v$ with a variable-scale Gaussian kernel $G(x,y,\sigma)$. Namely,

$$L_{\mathbf{I}}(x,y,\sigma) = \mathbf{I}_v(x,y) \otimes G(x,y,\sigma), \tag{1}$$

where $\otimes$ is the convolution operator and

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{2}$$

According to [1], the DoG image of scale s (s is an integer between 0 and 4) is computed from the difference of two adjacent scales as shown in Fig. 1. Formally,

$$D_{\mathbf{I}}(x,y,\sigma_s) = L_{\mathbf{I}}(x,y,\sigma_{s+1}) - L_{\mathbf{I}}(x,y,\sigma_s). \tag{3}$$

where σ_s denotes the standard deviation of the scale s. The candidate keypoints are then taken as the local extrema of

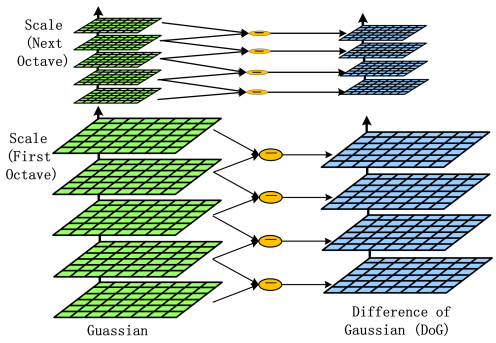


Fig. 1: The schematization of the scale space.

the DoG domain. For simplicity, let $\mathbf{x} = (x, y, \sigma_s)$ be a generic point in the scale space. It will be selected as a candidate keypoint if its DoG value $D_{\mathbf{I}}(\mathbf{x})$ is a local extremum (minimum or maximum) among all the points within the $3 \times 3 \times 3$ cube centered at itself.

To guarantee the stability and robustness of the final keypoints, in the phase ii), all the candidate keypoints are further refined according to a contrast threshold and an edge threshold. In the phase iii), a dominant orientation is calculated and assigned to each survived keypoint, to achieve the rotation invariance. In the phase iv), the surrounding information in the scale space of each keypoint is encoded into a 128-dimensional descriptor. For more details, please refer to [1].

B. A brief review of RCR

In this subsection, we give a brief review of our previous work, i.e., RCR [26]. For a specific octave v, let $\mathbf{k}_o = (x_o, y_o, \sigma_{s_o})$ denote a SIFT keypoint in the scale space, and $\mathcal{S}_o$ be the 27 points within the $3 \times 3 \times 3$ cube centered at $\mathbf{k}_o$. We can write

$$\mathcal{S}_o = \Big\{(x, y, \sigma_s)\Big| |x - x_o| \le 1, |y - y_o| \le 1, |s - s_o| \le 1, \\ x, y, s \in \mathcal{Z}\Big\}. \tag{4}$$

As discussed in Section II-A, the keypoint $\mathbf{k}_o$ is either the minimum or maximum in $\mathcal{S}_o$ of the image $\mathbf{I}$. We have

$$D_{\mathbf{I}}(\mathbf{k}_o) > D_{\mathbf{I}}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_o \setminus \{\mathbf{k}_o\}, \tag{5}$$

or

$$D_{\mathbf{I}}(\mathbf{k}_o) < D_{\mathbf{I}}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_o \setminus \{\mathbf{k}_o\}. \tag{6}$$

An effective way to remove $\mathbf{k}_o$ is to make it not an extremum in $\mathcal{S}_o$, i.e., violating conditions (5) and (6) simultaneously. In RCR, the problem of removing $\mathbf{k}_o$ is constructed into the following generic constrained optimization framework:

$$\min_{\hat{\mathbf{f}}_o} ||\mathbf{f}_o - \hat{\mathbf{f}}_o||_2^2, \tag{7}$$
$$\text{s.t.} \quad (C.1): \mathbf{k}_o \text{ is not an extremum in } \mathcal{S}_o \text{ of } \hat{\mathbf{I}},$$
$$(C.2): \text{no new keypoints generated around } \mathbf{k}_o,$$

where

- $\mathbf{f}_o \in \mathcal{R}^{P \times P}$ (P is a user-defined parameter) is a local patch centered at (x_o, y_o) from the image $\mathbf{I}$. Equivalently, we can write $\mathbf{f}_o = \mathbf{E}_o \circ \mathbf{I}$, where $\mathbf{E}_o$ serves as the patch extracting matrix;
- $\hat{\mathbf{f}}_o \in \mathcal{R}^{P \times P}$ is our targeted local patch;
- $\hat{\mathbf{I}}$ is the keypoint-removed image, which is unknown.

The objective function aims to minimize the distortion between the estimated patch and the original one. The constrains $(C.1)$ and $(C.2)$ guarantee that, in the resulting image $\hat{\mathbf{I}}$ accommodating $\hat{\mathbf{f}}_o$, $\mathbf{k}_o$ is not a keypoint, and no new keypoints are generated in its surroundings. Clearly, the resulting image $\hat{\mathbf{I}}$ can be directly obtained from $\mathbf{I}$ by replacing $\mathbf{f}_o$ with $\hat{\mathbf{f}}_o$.

Given a point $\mathbf{q}$ in the scale space of $\mathbf{I}$, the associated local minimum and maximum in its $3 \times 3 \times 3$ cube can be respectively defined as

$$\mathbf{x}_{\min}^q = \arg\min_{\mathbf{x} \in \mathcal{S}_q} D_{\mathbf{I}}(\mathbf{x}), \tag{8}$$

$$\mathbf{x}_{\max}^q = \arg\max_{\mathbf{x} \in \mathcal{S}_q} D_{\mathbf{I}}(\mathbf{x}), \tag{9}$$

where $\mathcal{S}_q$ can be similarly constructed as (4). [1]

As $\mathbf{k}_o$ is a keypoint, we can simply assume that it is the maximum in $\mathcal{S}_o$. In RCR, to remove $\mathbf{k}_o$, the constraint $(C.1)$ is designed to make the following inequality hold

$$D_{\hat{\mathbf{I}}}(\mathbf{k}_o) \le D_{\hat{\mathbf{I}}}(\mathbf{x}'^q_{\max}), \tag{10}$$

where

$$\mathbf{x}'^q_{\max} = \arg\max_{\mathbf{x} \in \mathcal{S}_o \setminus \{\mathbf{k}_o\}} D_{\mathbf{I}}(\mathbf{x}). \tag{11}$$

Apparently, $\mathbf{k}_o$ will no longer be a local maximum of $\hat{\mathbf{I}}$ when the condition (10) is satisfied. For simplicity, let $\mathbf{p} = \mathbf{x}'^q_{\max}$. To alleviate the NKG problem, RCR also prevents $\mathbf{p}$ from becoming a new extremum, and the constraint $(C.2)$ designed for $\mathbf{p}$ is

$$D_{\hat{\mathbf{I}}}(\mathbf{x}^p_{\min}) \le D_{\hat{\mathbf{I}}}(\mathbf{p}) \le D_{\hat{\mathbf{I}}}(\mathbf{x}^p_{\max}). \tag{12}$$

Denote $\mathcal{S}_p$ as the $3 \times 3 \times 3$ cube centered at $\mathbf{p}$. Obviously, we have $\mathbf{k}_o \in \mathcal{S}_p$, and $\mathcal{S}_p$ is overlapped with $\mathcal{S}_o$. It can be readily calculated that the overlapping percentage between $\mathcal{S}_p$ and $\mathcal{S}_o$ is between $\frac{8}{27}$ to $\frac{2}{3}$, depending on the location of $\mathbf{p}$. Since $\mathbf{k}_o$ is the maximum in $\mathcal{S}_o$ of $\mathbf{I}$, it is of high probability that $\mathbf{k}_o$ is also the maximum in $\mathcal{S}_p$ of $\mathbf{I}$, due to the high percentage of overlapping. When $\mathbf{k}_o$ is the maximum of both $\mathcal{S}_o$ and $\mathcal{S}_p$, we have $\mathbf{x}^p_{\max} = \mathbf{k}_o$. In this case, the inequality constraint on the right hand of (12) and the inequality constraint (10) are reduced to an equality constraint, namely, $D_{\hat{\mathbf{I}}}(\mathbf{k}_o) = D_{\hat{\mathbf{I}}}(\mathbf{p})$. Note that this is not a rare case. To illustrate this, we randomly select 10000 keypoints, and experimentally find that for more than half of the keypoints, the above inequality constraints are satisfied with equality. As previously discussed in Section I, it is difficult to maintain such equality constraints, which can be easily broken when removing the subsequent SIFT keypoints. As a result, those removed keypoints could reappear with a high probability. As will be shown in the experimental stage, such drawback of RCR leads to unsatisfactory performance when removing keypoints within multiple octaves.

[1] In the following, we replace the superscript/subscript q of $\mathbf{x}^q_{\min}$, $\mathbf{x}^q_{\max}$ and $\mathcal{S}_q$ with other notations (e.g., p), to denote the local minimum, local maximum and the $3 \times 3 \times 3$ cube of other points in the scale space (e.g., the point $\mathbf{p}$).

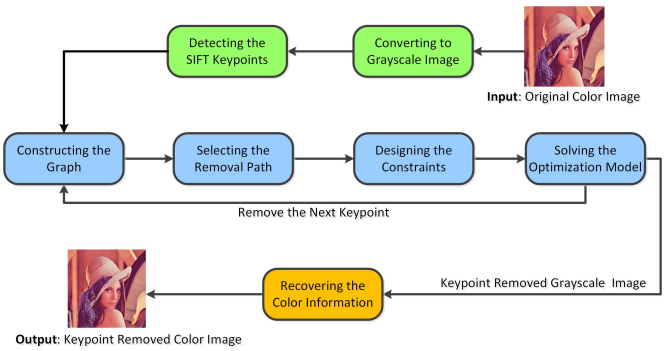


Fig. 2: The framework of our proposed scheme.

III. SIFT KEYPOINT REMOVAL VIA DIRECTED GRAPH CONSTRUCTION

In this section, we present our proposed SIFT keypoint removal algorithm. The whole framework is shown in Fig. 2. According to the SIFT algorithm [1], the input color image is converted to the grayscale version $\mathbf{I}$ prior to the feature extraction. To prepare for the subsequent SIFT keypoint removal, we first identify the keypoints in $\mathbf{I}$ by using the SIFT algorithm. We then construct the DoG space as a directed weighted graph, and compute the removal path in a minimum distortion sense over the constructed graph. The details regarding the graph construction and path computation are deferred to Section IV. Based on the pre-computed removal path, we design a series of strict inequality constraints to remove a SIFT keypoint, under a convex optimization framework. The details are given in the following Section III-A-C. With the keypoint removed grayscale image, we finally suggest an optimization-based restoration approach to recover the color information. The details can be found in Section V. Clearly, a sophisticated removal method should satisfy the following two requirements: i) significantly remove the original keypoints, and suppress new keypoint generation around the removed ones (i.e., alleviate the NKG problem); ii) minimize the incurred distortion.

In this work, we embark from the generic constrained optimization framework (7) to remove the keypoint $\mathbf{k}_o$. The key difference from [26] lies in the completely new ways of determining the constraints $(C.1)$ and $(C.2)$. Note that in [18], [21], [25], [26], the associated equality constraints make the optimal solutions have to be obtained at the boundary of the solution space. Consequently, those constraints satisfied at preceding removal procedures are hard to be maintained against the small disturbances among different removals. Furthermore, in practice, the resulting solutions will be saved as integers, where the rounding operations may also break those constraints. These two factors lead to the degraded performance of [18], [21], [25], [26], especially when removing keypoints within multiple octaves, where a large number of keypoints are highly clustered. However, different from [18], [21], [25], [26], the constraints employed in our framework are strict inequalities, which are further tightened to avoid the optimal solution to be achieved at the boundary of the original solution space. This makes our proposed scheme more robust against the disturbances introduced by different removals. As a result, compared with [18], [21], [25], [26], our proposed method can achieve much better KRR performance and higher quality of the resulting image, especially when removing the keypoints within multiple octaves. Noticing that most of the existing systems extract SIFT keypoints within multiple octaves, our proposed method is more meaningful in practice.

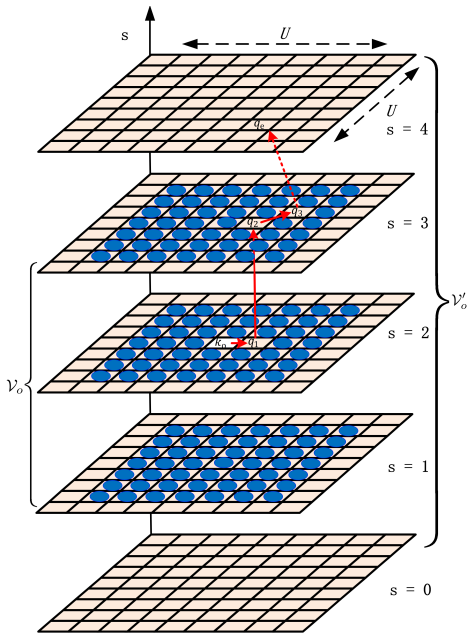


Fig. 3: The schematization of $\mathcal{V}_o$ (constructed by the blue dots) and $\mathcal{V}'_o$ (constructed by all the cells). Here, the parameter $U = 7$. The red arrows represent an example of the acyclic *path of inequality* in $\mathcal{V}'_o$.

A. Acyclic path of inequality in a keypoint free cuboid

To be consistent with [1], we assume that there are 5 scales within each octave, indexed by $s = 0$ to $s = 4$, respectively. Note that SIFT keypoints can only be generated within the inner 3 scales, i.e., $s = 1, 2, 3$ [1]. For a given keypoint $\mathbf{k}_o$ in the octave v, we construct two local cuboids $\mathcal{V}_o$ and $\mathcal{V}'_o$ in the scale space, which are formally defined as

$$\begin{aligned}\mathcal{V}_o = &\Big\{(x, y, \sigma_s)\Big||x - x_o| \le \frac{U-1}{2}, |y - y_o| \le \frac{U-1}{2}, \\ &1 \le s \le 3, x, y, s \in \mathcal{Z}\Big\} \backslash \Big\{\mathbf{k}_o\Big\},\end{aligned} \tag{13}$$

and

$$\begin{aligned}\mathcal{V}'_o = &\Big\{(x, y, \sigma_s)\Big||x - x_o| \le \frac{U+1}{2}, |y - y_o| \le \frac{U+1}{2}, \\ &0 \le s \le 4, x, y, s \in \mathcal{Z}\Big\},\end{aligned} \tag{14}$$

where U is a parameter determining the sizes of $\mathcal{V}_o$ and $\mathcal{V}'_o$. Apparently, we have $\mathcal{V}_o \subset \mathcal{V}'_o$. The schematization of $\mathcal{V}_o$ and $\mathcal{V}'_o$ is shown in Fig. 3.

Let $\mathbf{k}_o$ be a SIFT keypoint (either a local minimum or maximum) of $\mathbf{I}$. If we successfully remove all the SIFT keypoints out of $\mathcal{V}_o$, we will obtain an acyclic path of inequality

associated with $\mathbf{k}_o$ with respect to $\hat{\mathbf{I}}$. The derivation is given below [2].

We first consider the case that the keypoint $\mathbf{k}_o$ is a local maximum of $\mathbf{I}$. After removing $\mathbf{k}_o$, then $\mathbf{k}_o$ is no longer a local maximum of $\hat{\mathbf{I}}$. Namely, there exists a point $\mathbf{q}_1 \in \mathcal{S}_o \backslash \{\mathbf{k}_o\}$, such that

$$D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1). \tag{15}$$

Because all the keypoints have been removed in $\mathcal{V}_o$, then $\mathbf{q}_1$ is not a local maximum neither. There exists a point $\mathbf{q}_2 \in \mathcal{S}_{q_1} \backslash \{\mathbf{q}_1\}$, such that

$$D_{\hat{\mathbf{I}}}(\mathbf{q}_1) < D_{\hat{\mathbf{I}}}(\mathbf{q}_2). \tag{16}$$

By continuing such process, we eventually obtain an acyclic path of inequality denoted by $\mathcal{P}(\mathbf{k}_o, <)$, where all the points on the path belong to $\mathcal{V}_o'$. Specifically, we define

$$\mathcal{P}(\mathbf{k}_o, <): \ D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1) < \cdots < D_{\hat{\mathbf{I}}}(\mathbf{q}_e). \tag{17}$$

Here, the ending point $\mathbf{q}_e$ should satisfy one of the following two conditions; otherwise, such process could be further continued.

i) $\mathbf{q}_e \in \mathcal{V}_o$ is a local maximum but not a SIFT keypoint of $\hat{\mathbf{I}}$. This means that $\mathbf{q}_e$ is rejected to be a final SIFT keypoint in the refinement stage [1]. This process stops since there does not exist a point $\mathbf{q} \in \mathcal{S}_{q_e} \backslash \{\mathbf{q}_e\}$ such that $D_{\hat{\mathbf{I}}}(\mathbf{q}_e) < D_{\hat{\mathbf{I}}}(\mathbf{q})$;

ii) $\mathbf{q}_e \in \mathcal{V}_o' \backslash \mathcal{V}_o$. This process stops as $\mathbf{q}_e$ is not a point in $\mathcal{V}_o$.

In the sequel, we call $\mathbf{q}_1, \cdots, \mathbf{q}_{e-1}$ the *nonterminal* points, and $\mathbf{q}_e$ the *terminal* point of $\mathcal{P}(\mathbf{k}_o, <)$. A schematization of the acyclic path is also shown in Fig. 3 as the red arrows direct.

What might be interesting is that as long as $\mathcal{P}(\mathbf{k}_o, <)$ is known, we can employ (17) as the constraints to remove the keypoint $\mathbf{k}_o$ out of $\mathcal{V}_o$ along the path $\mathcal{P}(\mathbf{k}_o, <)$, ensuring that no local extrema (hence keypoints) are generated on the nonterminal points of the path.

Symmetrically, if the keypoint $\mathbf{k}_o$ is a local minimum of $\mathbf{I}$, and all the keypoints within $\mathcal{V}_o$ are removed, we can also obtain another acyclic path of inequality $\mathcal{P}(\mathbf{k}_o, >)$ associated with $\mathbf{k}_o$:

$$\mathcal{P}(\mathbf{k}_o, >): \ D_{\hat{\mathbf{I}}}(\mathbf{k}_o) > D_{\hat{\mathbf{I}}}(\mathbf{q}_1) > \cdots > D_{\hat{\mathbf{I}}}(\mathbf{q}_e). \tag{18}$$

Here, $\mathbf{q}_e$ is a local minimum (but not a SIFT keypoint) in $\mathcal{V}_o$ of $\hat{\mathbf{I}}$ or $\mathbf{q}_e \in \mathcal{V}_o' \backslash \mathcal{V}_o$. Similarly, we can use (18) as the constraints to remove $\mathbf{k}_o$ (local minimum) from $\mathcal{V}_o$ along the path $\mathcal{P}(\mathbf{k}_o, >)$, provided that $\mathcal{P}(\mathbf{k}_o, >)$ is known.

Unfortunately, in practice, neither $\mathcal{P}(\mathbf{k}_o, <)$ nor $\mathcal{P}(\mathbf{k}_o, >)$ can be directly derived from $\hat{\mathbf{I}}$, as $\hat{\mathbf{I}}$ is unknown. One of our major contributions lies in determining the paths $\mathcal{P}(\mathbf{k}_o, <)$ (when $\mathbf{k}_o$ is a local maximum) and $\mathcal{P}(\mathbf{k}_o, >)$ (when $\mathbf{k}_o$ is a local minimum) from the available $\mathbf{I}$. In the next two subsections, we assume that both $\mathcal{P}(\mathbf{k}_o, <)$ and $\mathcal{P}(\mathbf{k}_o, >)$ are known, and explicitly discuss the problem of determining the conditions $(C.1)$ and $(C.2)$ under the optimization framework given in (7). The discussion of constructing the paths $\mathcal{P}(\mathbf{k}_o, <)$ and $\mathcal{P}(\mathbf{k}_o, >)$ is deferred to Section IV.

[2]As discussed previously, though making two points with the same DoG value can remove one keypoint, the associated equality constraints are fragile to be broken against the small disturbances among different removals. All the constraints incorporated in our framework are strict inequalities, and hence, the associated points have different DoG values.

B. Determination of the new condition $(C.1)$

The purpose of imposing the constraint $(C.1)$ is to ensure that $\mathbf{k}_o$ is not a keypoint in the new image $\hat{\mathbf{I}}$. An effective way to this end is to make the inequalities given in (5) and (6) invalid simultaneously.

As $\mathbf{k}_o$ is a SIFT keypoint of $\mathbf{I}$, it can be either a local maximum or a local minimum. We first assume that $\mathbf{k}_o$ is a local maximum. As the path $\mathcal{P}(\mathbf{k}_o, <)$ is assumed to be known, we can readily obtain

$$D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1). \tag{19}$$

This constraint makes $\mathbf{k}_o$ no longer a local maximum of the processed image $\hat{\mathbf{I}}$. On the other hand, since $D_{\mathbf{I}}(\mathbf{k}_o) > D_{\mathbf{I}}(\mathbf{q}_1)$, we can also infer from (19) that the local order of $\mathbf{k}_o$ and $\mathbf{q}_1$ in the scale space will be reversed with respect to $\hat{\mathbf{I}}$, which inevitably causes distortion of $\hat{\mathbf{I}}$. As will be discussed in Section IV, how to design the associated path is crucial for minimizing the incurred distortion.

Since $\mathbf{k}_o$ is the maximum in $\mathcal{S}_o$ of $\mathbf{I}$, we also have $D_{\mathbf{I}}(\mathbf{x}_{\min}^{k_o}) < D_{\mathbf{I}}(\mathbf{k}_o)$. Here, $\mathbf{x}_{\min}^{k_o}$ is the minimum element in $\mathcal{S}_o$, which can be similarly defined as (8). To prevent $\mathbf{k}_o$ from becoming a local minimum, we force

$$D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^{k_o}) < D_{\hat{\mathbf{I}}}(\mathbf{k}_o). \tag{20}$$

A byproduct of (20) is that the relative order of $\mathbf{x}_{\min}^{k_o}$ and $\mathbf{k}_o$ in the scale space of $\hat{\mathbf{I}}$ is preserved, preventing the local structure to be changed too much when removing $\mathbf{k}_o$. The order-preserving technique was also shown to be a powerful regularization strategy in recent studies [29].

Eventually, the condition $(C.1)$ can be written as

$$D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^{k_o}) < D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1). \tag{21}$$

Apparently, when the condition (21) holds, $\mathbf{k}_o$ will not be identified as a SIFT keypoint as it is no longer an extremum in $\mathcal{S}_o$ with respect to $\hat{\mathbf{I}}$. Note that since $\mathbf{x}_{\min}^{k_o}$, $\mathbf{k}_o$ and $\mathbf{q}_1$ are all fixed, and the DoG function $D_{\hat{\mathbf{I}}}(\mathbf{x})$ is linear with respect to $\hat{\mathbf{I}}$, the above condition $(C.1)$ is also linear (and hence convex).

Symmetrically, if $\mathbf{k}_o$ is a local minimum of $\mathbf{I}$, the condition $(C.1)$ can be similarly designed as

$$D_{\hat{\mathbf{I}}}(\mathbf{q}_1) < D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^{k_o}). \tag{22}$$

C. Determination of the new condition $(C.2)$

In this subsection, we discuss how to determine the condition $(C.2)$ to suppress the new keypoints generation when removing $\mathbf{k}_o$. As explained in Section I, the new keypoints generated around the original ones can still be matched with a high probability, causing the NKG problem. Do *et. al* [19] showed that the correct matches can be effectively destroyed when the keypoints are shifted in the scale space with large offsets. In light of this knowledge, the condition $(C.2)$ is designed to ensure that no new local extrema (hence keypoints)

不要这个线

不用对齐

$$\begin{aligned}
\mathcal{L}_1 &= \mathcal{V}_o \bigcap \mathcal{N} \bigcap \mathcal{Q}_o^c, && \text{(not extrema \& not points on the path)} \\
\mathcal{L}_2 &= \mathcal{V}_o \bigcap \{\mathcal{Q}_o \backslash \mathbf{q}_e\} = \{\mathbf{q}_1, \cdots, \mathbf{q}_{e-1}\}, && \text{(nonterminal points on the path)} \\
\mathcal{L}_3 &= \mathcal{V}_o \bigcap \mathcal{N}^c, && \text{(extrema)}
\end{aligned} \tag{23}$$

where $\mathcal{Q}_o^c$ and $\mathcal{N}^c$ denote the complementary sets of $\mathcal{Q}_o$ and $\mathcal{N}$, respectively.

are generated in a local cuboid $\mathcal{V}_o$ in the scale space, where $\mathcal{V}_o$ is defined in (13).

We first consider the case that $\mathbf{k}_o$ is a local maximum of $\mathbf{I}$. Let $\mathcal{N}$ record all the points that are not extrema of $\mathbf{I}$ (i.e., non-extremum points). Assume that the acyclic path $\mathcal{P}(\mathbf{k}_o, <)$ is known, and let $\mathcal{Q}_o = \{\mathbf{q}_1, ..., \mathbf{q}_e\}$ comprise all the points on the path. Due to the different characteristics of the points in $\mathcal{V}_o$, we classify them into three classes $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ as shown in (23). Specifically, $\mathcal{L}_1$ contains the points which are neither the extrema in $\mathcal{V}_o$ nor the points on the path; $\mathcal{L}_2$ contains the nonterminal points on the path, and $\mathcal{L}_3$ contains the local extrema in $\mathcal{V}_o$. From the discussion in Section III-A, we know that $\mathbf{q}_e$ is a local extremum in $\mathcal{V}_o$ (i.e., $\mathbf{q}_e \in \mathcal{L}_3$) or $\mathbf{q}_e \in \mathcal{V}_o' \backslash \mathcal{V}_o$. Clearly, we have $\mathcal{V}_o = \mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3$.

Let $\mathbf{q}$ be a generic point in $\mathcal{V}_o$. We now discuss how to design the constraint imposed on $\mathbf{q}$ to suppress new keypoint generation. Considering the different properties of the points in different classes, three cases are addressed as follows.

Case 1: $\mathbf{q} \in \mathcal{L}_1$, i.e., $\mathbf{q}$ is neither a point on the acyclic path, nor a local extremum of $\mathbf{I}$. By resorting to the order-preserving technique discussed in Section III-B, the associated constraint can be designed as

$$D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^q) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^q), \ \mathbf{q} \in \mathcal{L}_1. \tag{24}$$

As $\mathbf{I}$ is available, both $\mathbf{x}_{\min}^q$ and $\mathbf{x}_{\max}^q$ are fixed. Evidently, constraint (24) can ensure that $\mathbf{q}$ is not an extremum (hence a keypoint) of $\hat{\mathbf{I}}$.

Case 2: $\mathbf{q} \in \mathcal{L}_2$, i.e., $\mathbf{q}$ is a nonterminal point on the pre-constructed path $\mathcal{P}(\mathbf{k}_o, <)$. Since $\mathcal{P}(\mathbf{k}_o, <)$ is assumed to be known, thus $\mathcal{Q}_o = \{\mathbf{q}_1, ..., \mathbf{q}_e\}$ is fixed. For simplicity, we set $\mathbf{q}_0 \triangleq \mathbf{k}_o$, and denote $\mathbf{q}_{i-1}$ and $\mathbf{q}_{i+1}$ respectively as the points before (*pre-point*) and after (*post-point*) $\mathbf{q}_i$ on the path, where $i = 1, ..., e-1$. The constraint for $\mathbf{q}$ can be directly defined as

$$D_{\hat{\mathbf{I}}}(\mathbf{q}_{i-1}) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{q}_{i+1}), \ \mathbf{q} = \mathbf{q}_1, ..., \mathbf{q}_{e-1}. \tag{25}$$

Obviously, (25) can guarantee that $\mathbf{q}$ is not a keypoint of $\hat{\mathbf{I}}$.

Case 3: $\mathbf{q} \in \mathcal{L}_3$, i.e., $\mathbf{q}$ is a local extremum in $\mathcal{V}_o$ of $\mathbf{I}$. Since the design goal of $(C.2)$ is to solve the NKG problem, we do not impose any constraints on the existing local extrema in $\mathcal{V}_o$ of $\mathbf{I}$. This is reasonable, because though $\mathbf{q}$ is a local extremum of $\mathbf{I}$, it does not necessarily mean that it would be survived as a final SIFT keypoint. Obviously, removing $\mathbf{q}$ in this case would cause unnecessary distortion. In addition, if unfortunately $\mathbf{q}$ is a SIFT keypoint in $\mathcal{V}_o$, it will be handled later because our removal method operates in a keypoint by keypoint manner.

Taking an overall consideration of Cases 1-3, the condition $(C.2)$ can be written as

$$\begin{aligned}
&D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^q) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^q), \quad \forall \mathbf{q} \in \mathcal{L}_1, \\
&D_{\hat{\mathbf{I}}}(\mathbf{q}_{i-1}) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{q}_{i+1}), \quad \mathbf{q} = \mathbf{q}_1, ..., \mathbf{q}_{e-1}.
\end{aligned} \tag{26}$$

Incorporating all the conditions $(C.1)$ and $(C.2)$, we finally arrive at the convex optimization problem for removing the SIFT keypoint $\mathbf{k}_o$ (local maximum)

$$\begin{aligned}
&\min_{\hat{\mathbf{f}}_o} ||\mathbf{f}_o - \hat{\mathbf{f}}_o||_2^2, \\
&\text{s.t. } D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^{k_o}) < D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1), \\
&\quad D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^q) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^q), \ \forall \mathbf{q} \in \mathcal{L}_1, \\
&\quad D_{\hat{\mathbf{I}}}(\mathbf{q}_{i-1}) < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{q}_{i+1}), \ \mathbf{q} = \mathbf{q}_1, ..., \mathbf{q}_{e-1}.
\end{aligned} \tag{27}$$

From the viewpoint of optimization theory, the optimal solution is usually obtained at the boundary of the solution space [30]. This brings two problems: 1) in practice, the resulting solution $\hat{\mathbf{f}}_o$ will be saved as integers, where the rounding operations may break the constraints defined in (27); 2) since the removal process operates in a keypoint by keypoint manner, the constraints satisfied for the preceding points could be easily broken by the disturbances introduced by the subsequent removals, especially when the keypoints are tightly clustered. These two problems can also be regarded as a part of reasons for the degraded performance of [18], [21], [25], [26]. In [18] and [21], the strategy of removing $\mathbf{k}_o$ is to create two equal extrema. Clearly, the associated equality constraints are very difficult to be maintained.

To achieve the robustness against small disturbances, we propose to adopt slightly tighter constraints, precluding the optimal solution from being obtained at the boundary of the original solution space. Specifically, the new framework with better robustness can be written as

$$\begin{aligned}
&\min_{\hat{\mathbf{f}}_o} ||\mathbf{f}_o - \hat{\mathbf{f}}_o||_2^2, \\
&\text{s.t. } D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^{k_o}) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{q}_1) - \epsilon, \\
&\quad D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^q) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^q) - \epsilon, \forall \mathbf{q} \in \mathcal{L}_1, \\
&\quad D_{\hat{\mathbf{I}}}(\mathbf{q}_{i-1}) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{q}_{i+1}) - \epsilon, \mathbf{q} = \mathbf{q}_1, ..., \mathbf{q}_{e-1},
\end{aligned} \tag{28}$$

where ϵ is a small positive constant. To balance the incurred distortion and the keypoint removal rate, we empirically set $\epsilon = 0.01$.

We should also emphasize that such design by introducing ϵ cannot be directly applied for RCR [26]. From the discussion of Section II-B, in many cases, the constraints of RCR are based on equality implicitly; such extension obviously makes the resulting optimization problem infeasible.

Symmetrically, if $\mathbf{k}_o$ is a local minimum of $\mathbf{I}$, we can similarly define the problem of removing the SIFT keypoint

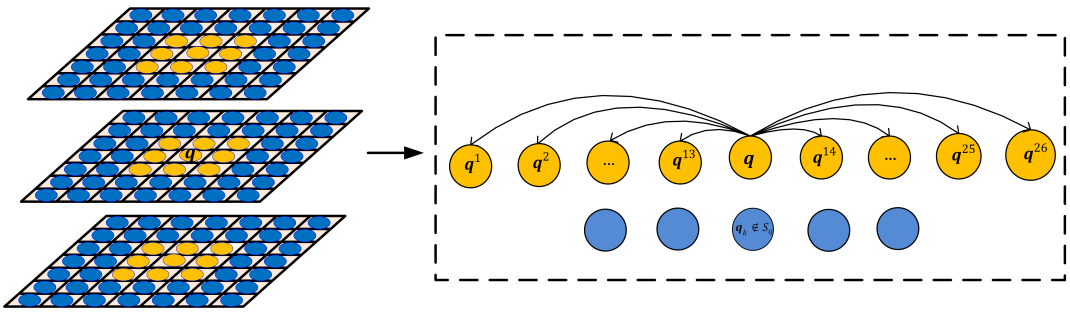


Fig. 4: The schematization of the graph $\mathcal{G}_o$. Here, we only show the subgraph of $\mathcal{G}_o$ associated with $\mathbf{q} \in \mathcal{V}_o$. $\mathbf{q}^1 \sim \mathbf{q}^{26}$ are the 26 points in $\mathcal{S}_q \backslash \{\mathbf{q}\}$. As we can see, $\mathbf{q}$ is only connected with $\mathbf{q}^1 \sim \mathbf{q}^{26}$.

$\mathbf{k}_o$ (local minimum) as

$$\begin{aligned} &\min_{\hat{\mathbf{f}}_o} ||\mathbf{f}_o - \hat{\mathbf{f}}_o||_2^2, \\ \text{s.t. } & D_{\hat{\mathbf{I}}}(\mathbf{q}_1) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{k}_o) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^{k_o}) - \epsilon, \\ & D_{\hat{\mathbf{I}}}(\mathbf{x}_{\min}^{q}) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{x}_{\max}^{q}) - \epsilon, \forall \mathbf{q} \in \mathcal{L}_1, \\ & D_{\hat{\mathbf{I}}}(\mathbf{q}_{i+1}) + \epsilon < D_{\hat{\mathbf{I}}}(\mathbf{q}) < D_{\hat{\mathbf{I}}}(\mathbf{q}_{i-1}) - \epsilon, \mathbf{q} = \mathbf{q}_1, ..., \mathbf{q}_{e-1}. \end{aligned} \tag{29}$$

An intuitional explanation of the constraints defined in (28) and (29) is that we push $\mathbf{k}_o$ along a pre-constructed acyclic path to a proper terminal point $\mathbf{q}_e$, ensuring that no new extrema (hence keypoints) are generated around. Apparently, how to construct the associated path is critical to minimize the incurred distortion, which will be discussed next.

IV. Construction of the Acyclic Paths

In this section, we present how to construct the acyclic paths $\mathcal{P}(\mathbf{k}_o, <)$ (17) and $\mathcal{P}(\mathbf{k}_o, >)$ (18) in a minimum distortion sense. Note that identifying $\mathbf{k}_o$ as a local extremum only relies on the relationship of the 27 points in $\mathcal{S}_o$. In this work, we model the scale space as a directed weighted graph, where each vertex (point) is only connected with its surrounding 26 vertices. Specifically, for a given cuboid $\mathcal{V}_o'$ as defined in (14) and illustrated in Fig. 3, we denote the associated graph as $\mathcal{G}_o$. Formally, we can write

$$\mathcal{G}_o = (\mathcal{V}_o', \mathcal{E}, \mathbf{W}), \tag{30}$$

where $\mathcal{E}$ is a set containing all the edges of the graph. Mathematically,

$$\mathcal{E} = \left\{ < \mathbf{q}_m, \mathbf{q}_n > \,\middle|\, \text{if } \mathbf{q}_m \in \mathcal{V}_o \wedge \mathbf{q}_n \in \mathcal{S}_{q_m} \setminus \{\mathbf{q}_m\} \right\}, \tag{31}$$

where $\mathcal{S}_{q_m}$ is the $3 \times 3 \times 3$ cube centered at $\mathbf{q}_m$, $\mathbf{W} \in \mathbb{R}^{T \times T}$ is the weighted adjacency matrix, and $T = (U+2) \times (U+2) \times 5$. To give a better understanding, Fig. 4 depicts the subgraph of $\mathcal{G}_o$ associated with $\mathbf{q} \in \mathcal{V}_o$.

Recall that our strategy of removing $\mathbf{k}_o$ is equivalent to push it out of $\mathcal{V}_o$ along an acyclic path, and ensure that no new local extrema are generated around. This implies that some local orders among the points in the scale space will be reversed after removing $\mathbf{k}_o$, which inevitably causes distortion of the resulting image. As will be clear soon, the adjacency matrix $\mathbf{W}$ plays a vital role for the path selection and distortion suppression. Define

$$d_{m,n} = D_{\mathbf{I}}(\mathbf{q}_m) - D_{\mathbf{I}}(\mathbf{q}_n), \tag{32}$$

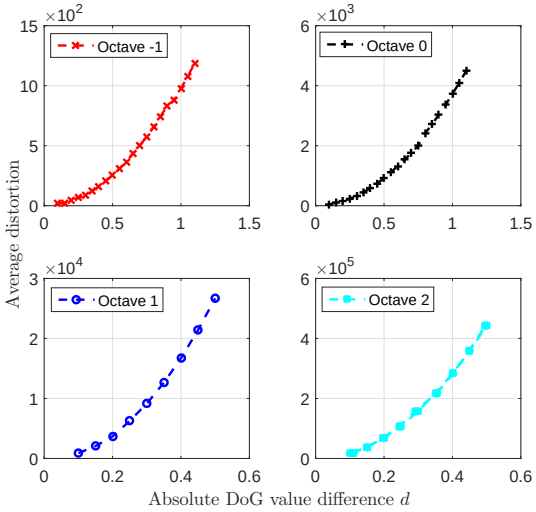


Fig. 5: The relationship between the average distortion and the absolute DoG value difference d, when reversing the order of point pairs in the scale space.

where $\mathbf{q}_m$ and $\mathbf{q}_n$ are two points satisfying $\mathbf{q}_n \in \mathcal{S}_{q_m}$.

***Assumption** 1*: Let $\mathcal{Q}_d$ record the point pairs in the scale space with the same absolute DoG value difference d, where $d \geq 0$. Mathematically,

$$\mathcal{Q}_d = \left\{ (\mathbf{q}_m, \mathbf{q}_n) \,\middle|\, |d_{m,n}| = d \wedge \mathbf{q}_n \in \mathcal{S}_{q_m} \right\}. \tag{33}$$

Let also $H_{m,n}$ be the incurred distortion of the resulting image $\hat{\mathbf{I}}$, when reversing the order of $(\mathbf{q}_m, \mathbf{q}_n) \in \mathcal{Q}_d$, i.e., making $D_{\hat{\mathbf{I}}}(\mathbf{q}_m) \leq D_{\hat{\mathbf{I}}}(\mathbf{q}_n)$ if $d_{m,n} > 0$ or making $D_{\hat{\mathbf{I}}}(\mathbf{q}_m) > D_{\hat{\mathbf{I}}}(\mathbf{q}_n)$ if $d_{m,n} \leq 0$. Then the average distortion over all the point pairs in $\mathcal{Q}_d$, namely,

$$H_{\text{ave}} = \frac{1}{|\mathcal{Q}_d|} \sum_{(\mathbf{q}_m, \mathbf{q}_n) \in \mathcal{Q}_d} H_{m,n}, \tag{34}$$

is positively correlated to d.

The above assumption indicates that more severe change of the DoG magnitude generally leads to larger distortion of the resulting image. This assumption was also adopted in [21], where the two points with the closest DoG magnitudes were set to be equal to minimize the incurred distortion. Note that ***Assumption** 1* describes the statistical behavior of the relationship between the distortion caused by reversing

$$\mathbf{W}_{m,n} = \begin{cases} \max(0, h \times (D_{\mathbf{I}}(\mathbf{q}_m) - D_{\mathbf{I}}(\mathbf{q}_n))), & \text{if } <\mathbf{q}_m, \mathbf{q}_n> \in \mathcal{E} \\ \infty, & \text{otherwise} \end{cases} \tag{35}$$

where

$$h = \begin{cases} 1, & \text{if } \mathbf{k}_o \text{ is a local maximum} \\ -1, & \text{if } \mathbf{k}_o \text{ is a local minimum} \end{cases}$$

the order of $(\mathbf{q}_m, \mathbf{q}_n)$ and $|d_{m,n}|$, in an average sense. To further validate ***Assumption 1*** experimentally, we randomly select 10000 point pairs $(\mathbf{q}_m, \mathbf{q}_n)$'s for each octave from 50 images. We then reverse the order of each point pair $(\mathbf{q}_m, \mathbf{q}_n)$ in the scale space, and compute the incurred distortion in ℓ_2 sense, as well as the associated $|d_{m,n}|$. More specifically, in our experiment, if $D_{\mathbf{I}}(\mathbf{q}_m) > D_{\mathbf{I}}(\mathbf{q}_n)$, we reverse their order by making $D_{\hat{\mathbf{I}}}(\mathbf{q}_m) < D_{\hat{\mathbf{I}}}(\mathbf{q}_n) - \epsilon$; otherwise, we make $D_{\hat{\mathbf{I}}}(\mathbf{q}_m) > D_{\hat{\mathbf{I}}}(\mathbf{q}_n) + \epsilon$, where ϵ is a small positive constant set to be 0.01. We quantize each $|d_{m,n}|$ by using a uniform scalar quantizer with step size 0.05. Then we collect all the point pairs $(\mathbf{q}_m, \mathbf{q}_n)$'s whose $|d_{m,n}|$'s fall into the same quantization bin, and calculate the average distortion over all these point pairs. Fig. 5 depicts the curves of the relationship between the average distortion and d for octaves -1, 0, 1 and 2. It can be easily seen that the average distortion induced by the order reversing increases as the DoG magnitudes difference d becomes larger.

In light of ***Assumption 1***, we should avoid to reverse the local order of points $\mathbf{q}_m$ and $\mathbf{q}_n$ in the scale space when $|d_{m,n}|$ is large. This motivates us to design a penalty coefficient $\mathbf{W}_{m,n}$ to reflect the distortion caused by reversing the order of $\mathbf{q}_m$ and $\mathbf{q}_n$ in the scale space.

When $\mathbf{k}_o$ is a local maximum of $\mathbf{I}$, we need to construct the acyclic path $\mathcal{P}(\mathbf{k}_o, <)$. According to (17), if the path goes through $\mathbf{q}_m$ and $\mathbf{q}_n$ sequentially, then we have constraint $D_{\hat{\mathbf{I}}}(\mathbf{q}_m) < D_{\hat{\mathbf{I}}}(\mathbf{q}_n)$. Apparently, the order of $\mathbf{q}_m$ and $\mathbf{q}_n$ will be reversed if $D_{\mathbf{I}}(\mathbf{q}_m) \geq D_{\mathbf{I}}(\mathbf{q}_n)$, i.e., $d_{m,n} \geq 0$. In this case we set the penalty coefficient $\mathbf{W}_{m,n}$ positively correlate to $d_{m,n}$, to reflect the associated distortion. Otherwise if $d_{m,n} < 0$, the order of $\mathbf{q}_m$ and $\mathbf{q}_n$ will not be changed according to (17), then we simply set $\mathbf{W}_{m,n} = 0$.

Symmetrically, when constructing the path $\mathcal{P}(\mathbf{k}_o, >)$, we allow $\mathbf{W}_{m,n}$ to positively correlate to $-d_{m,n}$ if $d_{m,n} \leq 0$, and set $\mathbf{W}_{m,n} = 0$ if $d_{m,n} > 0$. To take an overall consideration, we design the weighted adjacency matrix $\mathbf{W}$ as (35), where the parameter h is to ensure that the penalty coefficient $\mathbf{W}_{m,n}$ is a positive value when reversing the order of $\mathbf{q}_m$ and $\mathbf{q}_n$.

According to ***Assumption 1***, we propose to construct the acyclic path by finding the shortest path from $\mathbf{k}_o$ to a terminal point $\mathbf{q}_e$ over the graph $\mathcal{G}_o$. As discussed in Section III-A, $\mathbf{q}_e$ is a point satisfying one of the two conditions: 1) $\mathbf{q}_e \in \mathcal{V}_o' \backslash \mathcal{V}_o$; 2) $\mathbf{q}_e$ is a local extremum in $\mathcal{V}_o$. Note that the points satisfying condition 1) or condition 2) are *not* unique in $\mathcal{V}_o'$. In this work, we choose $\mathbf{q}_e$ as the point with the shortest distance to $\mathbf{k}_o$, and satisfying condition 1) or condition 2). To this end, we adopt the well-known *Dijkstra's Algorithm* to find the shortest path [31]. The algorithm is terminated when it first reaches any point (i.e., $\mathbf{q}_e$) satisfying condition 1) or condition 2), and it finally returns the path. Upon having the path, the optimization problem (28) and (29) can be readily solved by plugging $\mathbf{q}_1, ..., \mathbf{q}_e$ in.

Straightforwardly, we can sequentially apply our proposed method on all SIFT keypoints detected in $\mathbf{I}$. Note that after one round of removal, some new keypoints could be generated due to the interferences introduced by different removals. Furthermore, our algorithm can only guarantee that no keypoints exist in a local cuboid in the scale space; while some new keypoints generated outside the cuboid could still be matched with the original keypoints. As a result, same with all the existing removal methods, our proposed removal algorithm also works in an iterative manner. It terminates when the KRR is desired or the maximum iteration threshold T is reached ($T = 50$ in our experiment).

Remark: When developing our optimization-based SIFT keypoint removal algorithm, we make several assumptions such as ***Assumption 1***. Though these assumptions are reasonable, they may not perfectly hold in practice. In addition, we can observe from Fig. 2 that our removal algorithm is split into several subproblems and works in an iterative manner. As a result, the solution obtained from our developed framework is only suboptimal in general, and we do not preclude the existence of better removal algorithms. In fact, finding the removal performance limit is still an open problem, which seems to be very challenging. Hence, the performance loss with respect to the performance limit is unknown. Certainly, this is an interesting problem that needs to be investigated in the future. Despite the suboptimality in general, we can see that our proposed method achieves much better performance than all the other existing algorithms. This, in turn, indicates that the constraints in our optimization framework are well designed.

V. Recover the color information of the cleaned Grayscale image

As mentioned previously, SIFT algorithm [1] extracts features only on grayscale image. One color image will be first converted to the grayscale format prior to SIFT keypoint extraction, causing the color information of the anti-forensically modified image $\hat{\mathbf{I}}$ totally lost. This is a very critical problem to preclude the widespread adoption of the proposed algorithm, since most of images are with color information nowadays. To the best of our knowledge, there is no related published work to handle this issue so far. In this section, we present a simple yet effective algorithm to recover the color information from the anti-forensically modified grayscale image $\hat{\mathbf{I}}$.

For simplicity, assume that $\hat{\mathbf{I}}$ is vectorized into a column vector, and the original color image $\mathbf{I}_{rgb}$ consists of three

TABLE I: THE NUMBER OF KEYPOINTS

Image	#KP in Octave 0	#KP in Multiple Octaves
Baboon	415	2971
Barbara	205	1446
Bridge	73	909
F16	448	1834
Goldhill	386	1532
Lena	203	1069
Peppers	188	677
Sailboat	165	944
Average	260	1423

channels, respectively denoted by $\mathbf{I}_r, \mathbf{I}_g$ and $\mathbf{I}_b$. Let $\mathbf{y}_i = [y_i^r, y_i^g, y_i^b]^T \in \mathcal{R}^{3\times 1}$ be the ith pixel of $\mathbf{I}_{rgb}$. We can write $\mathbf{I}_{rgb} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_J]^T$, where $J = M \times N$. Our goal is to generate a new color image $\hat{\mathbf{I}}_{rgb} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_J]^T$ close to $\mathbf{I}_{rgb}$, satisfying

$$\hat{\mathbf{I}}_{rgb}\mathbf{w} = \hat{\mathbf{I}}, \tag{36}$$

where $\mathbf{w}$ is the mapping operator, and is commonly set as $\mathbf{w} = [0.2989, 0.5870, 0.1140]^T$. Our problem can then be formulated as

$$\begin{aligned} \min_{\hat{\mathbf{y}}_1,\hat{\mathbf{y}}_2,...,\hat{\mathbf{y}}_J} \quad & f(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_J) = \frac{1}{2}\|\mathbf{I}_{rgb} - \hat{\mathbf{I}}_{rgb}\|_F^2 \\ & = \frac{1}{2}\sum_{j=1}^{J} \|\mathbf{y}_j - \hat{\mathbf{y}}_j\|_2^2, \\ \text{s.t.} \quad & \mathbf{w}^T\hat{\mathbf{y}}_j = \hat{\mathbf{I}}_j, \quad j = 1, ..., J, \end{aligned} \tag{37}$$

where $\hat{\mathbf{I}}_j$ represents the jth pixel of $\hat{\mathbf{I}}$. The objective function of (37) is designed to minimize the ℓ_2 distortion between the resulting image and the original one. It is also possible to use some other distortion metrics, such as SSIM [32]. Nevertheless, for the sake of simplicity and convexity, we use ℓ_2 distortion metric in this work. Such choice also makes the above optimization problem convex and separable; namely solving each pixel of $\hat{\mathbf{I}}_{rgb}$ is independent. The problem (37) can be solved by using the method of Lagrange multipliers, which leads to the following closed-form solution

$$\hat{\mathbf{I}}_{rgb} = \mathbf{I}_{rgb} - \frac{1}{\mathbf{w}^T\mathbf{w}}(\mathbf{I}_{rgb}\mathbf{w} - \hat{\mathbf{I}})\mathbf{w}^T. \tag{38}$$

As will be shown in the experiments, the proposed color restoration technique is effective in recovering the color information of the keypoint-removed images.

In the following sections, we refer our proposed method as SIFT keypoint ***Removal*** via ***Directed Graph*** Construction (RDG). Note that as an independent component, our proposed removal method RDG can be readily incorporated with the existing SIFT keypoint injection algorithms, such as FMD [19] and TPKI [26].

VI. Experimental Results

Now we investigate the effectiveness of our proposed SIFT keypoint removal method RDG. Two data sets are employed in our experiment. The first one was adopted in [21], which contains 8 standard images (`Baboon`, `Barbara`, `Bridge`, `F16`, `Goldhill`, `Lena`, `Peppers` and `Sailboat`) of size 512×512. In the sequel, we refer it as `Dataset8`. The second one is `UCID-v2` corpus [33], which contains 1338 images of size 512×384 with various characteristics.

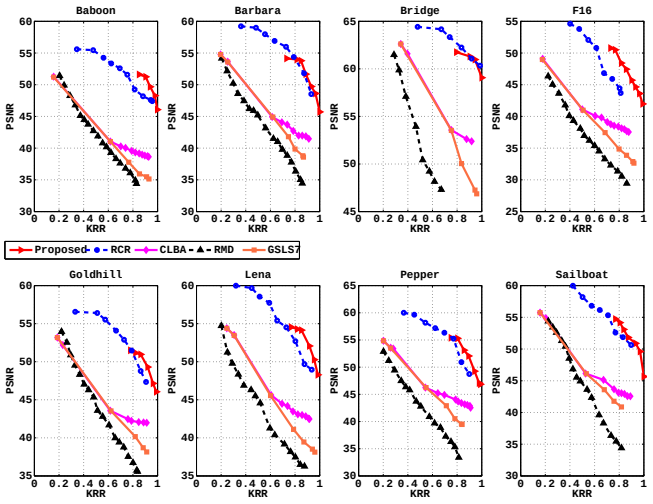


Fig. 6: Comparison with the proposed RDG, RCR [26], RMD [19], GSLS7 [19] and CLBA [23] in terms of average KRR-D performance on octave 0 for `Dataset8`.

The widely adopted SIFT-VLFEAT [34] is used in our work for SIFT keypoint extraction. To be consistent with [23], [26], [28], [35], [36], the associated peak and edge thresholds are respectively set as 4 and 10. A keypoint is regarded as removed if it is no longer a keypoint in the processed image, or if it is not matched with the original keypoints. The criterion is also adopted in all the other SIFT keypoint removal algorithms. It should be noted that in [23], [26], [28], [35], [36], only those keypoints in the first octave (octave 0) are considered. However, most of the existing systems built upon SIFT extract keypoints in multiple octaves. Therefore, in our experiment, we conduct extensive experiments for both a single octave and multiple octaves. Compared with removing keypoints in a single octave, removing keypoints in multiple octaves is much more challenging, since much more highly clustered keypoints need to be removed. Table I lists the number of keypoints in a single octave (octave 0) and in multiple octaves (octaves -1, 0, 1 and 2), where we can see that the number of keypoints detected in multiple octaves is generally more than 5 times larger than that in octave 0. As will be shown later, our proposed method RDG outperforms the state-of-the-art techniques for both a single octave and multiple octaves.

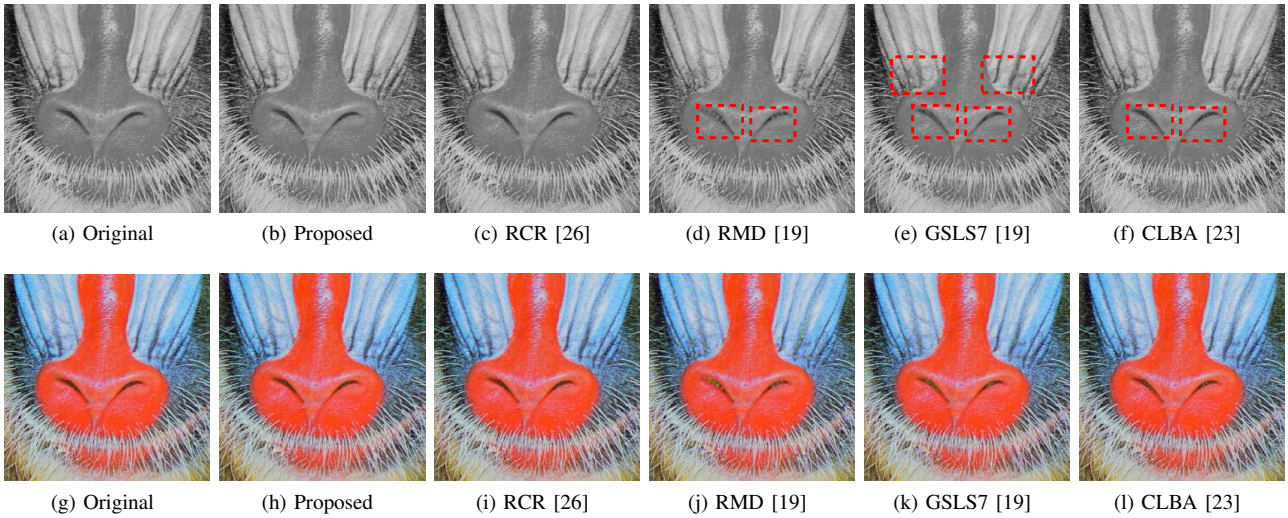
The performance of SIFT keypoint removal is evaluated in terms of *Keypoint Removal Rate-Distortion* (KRR-D) metric, where the KRR is defined as

$$KRR = 1 - \frac{\#\text{ correctly matched keypoints after removal}}{\#\text{ number of original keypoints}},$$

and the Distortion is measured in terms of the *Peak Signal-to-Noise Ratio* (PSNR) metric, between the original image $\mathbf{I}$ and the resulting image $\hat{\mathbf{I}}$.

A. Performance of SIFT keypoint removal on single octave

We first conduct the experiments for removing the SIFT keypoints in octave 0. The comparison is performed among



(a) Original (b) Proposed (c) RCR [26] (d) RMD [19] (e) GSLS7 [19] (f) CLBA [23]

(g) Original (h) Proposed (i) RCR [26] (j) RMD [19] (k) GSLS7 [19] (l) CLBA [23]

Fig. 7: Visual quality comparison with different removal methods for octave 0. (a, g) original, (b, h) Proposed RDG (KRR 99.52%, PSNR 47.37 dB, SSIM 0.9992, PSNR-C 47.12 dB, SSIM-C 0.9989), (c, i) RCR (KRR 95.66%, PSNR 47.40 dB, SSIM 0.9989, PSNR-C 46.96 dB, SSIM-C 0.9987), (d, j) RMD (KRR 82.89%, PSNR 34.40 dB, SSIM 0.9891, PSNR-C 35.55 dB, SSIM-C 0.9920), (e, k) GSLS7 (KRR 85.30%, PSNR 35.94 dB, SSIM 0.9875, PSNR-C 37.08 dB, SSIM-C 0.9912), and (f, l) CLBA (KRR 90.12%, PSNR 38.63 dB, SSIM 0.9931, PSNR-C 39.68 dB, SSIM-C 0.9949).

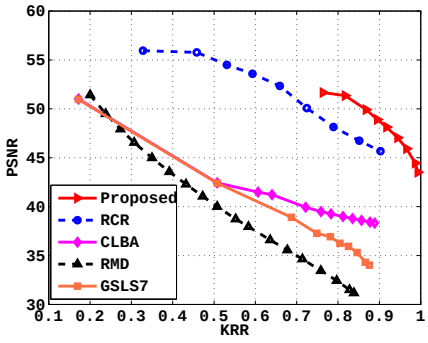


Fig. 8: Comparison with the proposed RDG, RCR [26], RMD [19], GSLS7 [19] and CLBA [23] in terms of average KRR-D performance for octave 0 over `UCID-v2`.

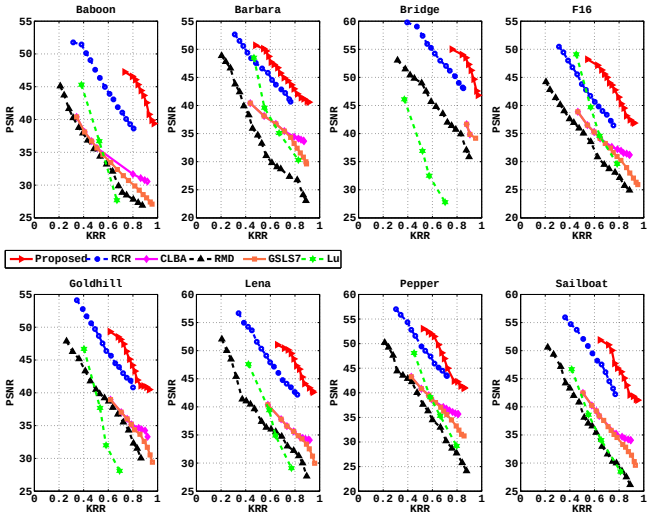
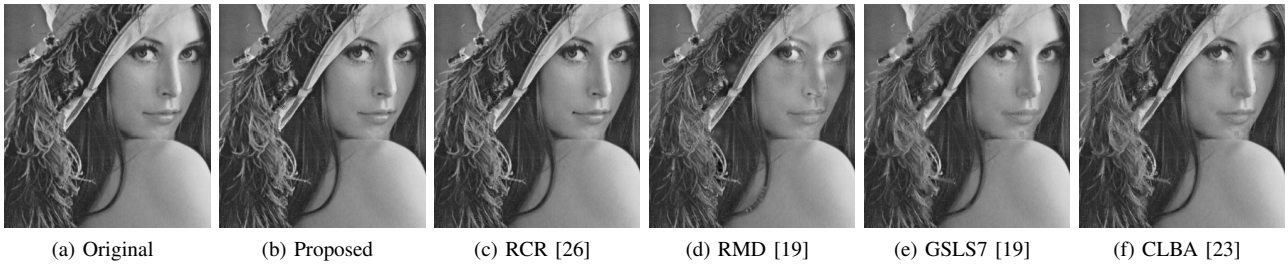


Fig. 9: Comparison with [19] and [21] in terms of KRR-D performance on multiple octaves for `Dataset8`.

our proposed RDG, GSLS7 [19], RMD [19], CLBA [23], RCR [26] over both `Dataset8` and `UCID-v2`. All the associated parameters of each method are carefully tuned to achieve the best KRR-D performance for the sake of fairness. As will be clear soon, our proposed RDG is capable of effectively reducing the keypoints, while controlling the incurred distortion to an imperceptible level.

1) Removal effectiveness on `Dataset8`: Fig. 6 shows the KRR-D performance on `Dataset8` with respect to the above five attack methods, where each point in the KRR-D plot represents the result of a certain iteration. We can observe that our proposed RDG significantly outperforms the competing methods RMD [19], GSLS7 [19] and CLBA [23] in terms of KRR-D metric. Take `Lena` as an example. The gains of PSNR over RMD, GSLS7 and CLBA are about 17 dB, 14 dB and 12 dB, respectively when KRR is 0.8. Further, it can also be noticed from Fig. 6 that RDG slightly outperforms RCR [26] in most cases. For `Lena`, the PSNR gain over RCR is about 2 dB when KRR is 0.8. Though the performance gap is not quite significant, it will be shown later, our proposed RDG outperforms RCR by a big margin in the case of removing keypoints in multiple octaves.

In addition to the KRR-D performance, we also present the visual quality comparison among these five competing methods in Fig. 7, where the first row shows the keypoint-removed grayscale images, while the second row lists the corresponding color images. To better measure the distortion, we provide both PSNR and SSIM results for different algorithms. In addition, we use "PSNR-C" and "SSIM-C" to present PSNR and SSIM metrics for color images, which are calculated by



(a) Original (b) Proposed (c) RCR [26] (d) RMD [19] (e) GSLS7 [19] (f) CLBA [23]

(g) Original (h) Proposed (i) RCR [26] (j) RMD [19] (k) GSLS7 [19] (l) CLBA [23]

Fig. 10: Visual quality comparison with different removal methods for multiple octaves. (a, g) original, (b, h) Proposed RDG (KRR 91.58%, PSNR 43.38 dB, SSIM 0.9972, PSNR-C 42.53 dB, SSIM-C 0.9952), (c, i) RCR (KRR 81.01%, PSNR 42.37 dB, SSIM 0.9956, PSNR-C 42.31 dB, SSIM-C 0.9944), (d, j) RMD (KRR 84.47%, PSNR 30.49 dB, SSIM 0.9689, PSNR-C 31.67 dB, SSIM-C 0.9751), (e, k) GSLS7 (KRR 87.56%, PSNR 33.72 dB, SSIM 0.9700, PSNR-C 34.87 dB, SSIM-C 0.9770), and (f, l) CLBA (KRR 89.99%, PSNR 34.23 dB, SSIM 0.9750, PSNR-C 35.35 dB, SSIM-C 0.9803).

(a) Original (b) Proposed (c) RCR [26] (d) RMD [19] (e) GSLS7 [19] (f) CLBA [23]

(g) Original (h) Proposed (i) RCR [26] (j) RMD [19] (k) GSLS7 [19] (l) CLBA [23]

Fig. 11: Visual quality comparison with different removal methods for multiple octaves. (a, g) original, (b, h) Proposed RDG (KRR 93.60%, PSNR 40.14 dB, SSIM 0.9958, PSNR-C 41.09 dB, SSIM-C 0.9967), (c, i) RCR (KRR 80.58%, PSNR 38.63 dB, SSIM 0.9936, PSNR-C 39.45 dB, SSIM-C 0.9949), (d, j) RMD (KRR 88.89%, PSNR 26.80 dB, SSIM 0.9533, PSNR-C 28.06 dB, SSIM-C 0.9676), (e, k) GSLS7 (KRR 91.11%, PSNR 28.06 dB, SSIM 0.9233, PSNR-C 29.32 dB, SSIM-C 0.9494), and (f, l) CLBA (KRR 91.15%, PSNR 30.51 dB, SSIM 0.9557, PSNR-C 31.53 dB, SSIM-C 0.9704).

taking the average results for the three channels [37]. Since all the existing removal techniques do not address the color loss issue, for RCR, RMD, GSLS7 and CLBA, the color information is recovered by our proposed color restoration method (see Section V). We can see that our proposed RDG achieves more visually pleasing results than RMD, GSLS7 and CLBA. Readers are invited to examine the highlighted regions enclosed in the red rectangles in Fig. 7. RMD and CLBA tend to introduce severe noise around the highly textured regions, while GSLS7 is prone to over-smoothing the fine details. We also find that RDG and RCR both achieve high quality of the resulting images.

2) Removal effectiveness on `UCID-v2`*:* Fig. 8 reports the average KRR-D performance of RDG, RCR [26], RMD [19],

GSLS7 [19] and CLBA [23] for the dataset UCID-v2. Our proposed RDG still outperforms RMD, GSLS7 and CLBA by a big margin. For example, the PSNR gains of RDG over RMD, GSLS7 and CLBA are about 20 dB, 15 dB and 13 dB, respectively when KRR is 0.8, which is quite remarkable. In the same case, the PSNR gain of RDG over RCR is about 5 dB. Moreover for RCR, we experimentally find that the KRR performance can hardly be further enhanced when it reaches around 0.9. However, our proposed RDG can achieve KRR very close to 1.0, while still maintaining high quality of the resulting images.

每个首字母大写

B. Performance of SIFT keypoint removal on multiple octaves

To further show the effectiveness of RDG, we conduct the experiments of removing the SIFT keypoints in octaves -1, 0, 1 and 2 simultaneously. We do not consider the higher octaves since they contain a negligible number of keypoints. As shown in Table I, this generally results in more than 5 times larger number of keypoints than that in octave 0 only. Note that removing SIFT keypoints in multiple octaves simultaneously is much more challenging than that in a single octave, as the support regions of the spatially neighbored keypoints are highly overlapped. It follows that the constraints satisfied at preceding removals are more likely to be broken when removing the following keypoints, making those removed keypoints reappear. Furthermore, the distortion becomes larger when more keypoints are removed.

1) Removal effectiveness on Dataset8: Fig. 9 reports the results over Dataset8 for six competitors, including our proposed RDG, RCR [26], RMD [19], GSLS7 [19], CLBA [23] and Lu's method [21]. As the source code of [21] is not available, the results represented by the green dotted lines in Fig. 9 are extracted from Table 1 of [21]. Compared with RMD, GSLS7, CLBA and Lu's method, our proposed RDG achieves quite remarkable gains in terms of KRR-D performance. Take Lena as an example. When KRR = 0.8, the PSNR gains over RMD, GSLS7, CLBA and Lu's method are about 16 dB, 13 dB, 13 dB and 19 dB, respectively. Still, we can observe that RDG outperforms RCR in terms of KRR-D metric. In addition, the KRR achieved by RCR cannot be further increased when it reaches around 0.8. This is because those removed keypoints tend to reappear for RCR. We emphasize that both the KRR performance and the PSNR performance are important when attacking a SIFT-based system. Thanks to the robust constraints designed in our optimization framework, the KRR achieved by the proposed RDG is generally above 0.95 in the case of removing keypoints in multiple octaves, and the resulting images are still of very high quality.

The visual quality comparison among RDG, RCR, RMD, GSLS7 and CLBA for removing keypoints in multiple octaves is given in Figs. 10-11. We observe that RDG achieves much better visual quality of the resulting images than the other competitors. RMD, GSLS7 and CLBA are prone to over-smoothing the fine details. Interested readers can take a close look at the hair of Lena in Fig. 10(d)-(f). For RCR, it tends to introduce artifacts around the highly textured areas when removing keypoints in multiple octaves as shown in Fig. 10(c).

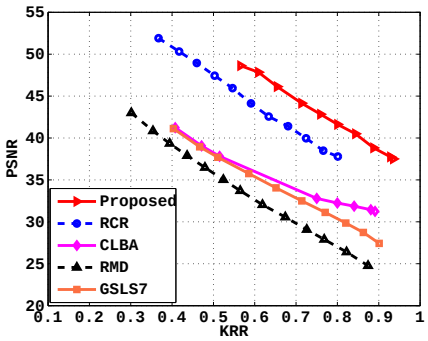


Fig. 12: Comparison with RMD method [19] in terms of average KRR-D performance for multiple octaves over UCID-v2.

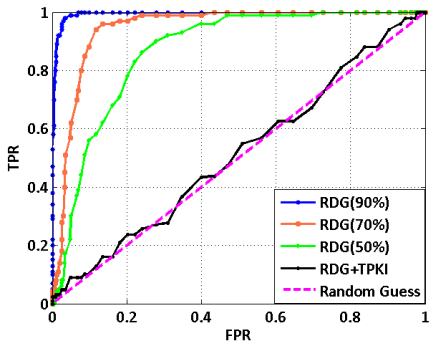
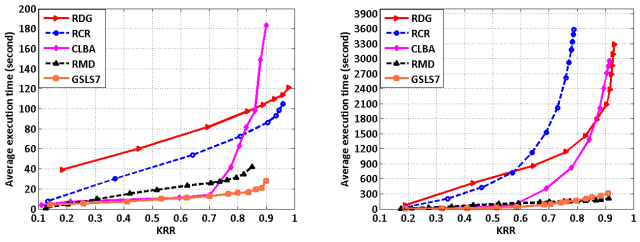


Fig. 13: ROC curves for the KCR detector with respect to multiple octaves.

2) Removal effectiveness on UCID-v2: To further investigate the KRR-D performance of different methods, we conduct the experiments to remove keypoints in multiple octaves over UCID-v2. Fig. 12 depicts the average KRR-D performance of RCR, RMD, GSLS7, CLBA and our proposed RDG. One can observe that our proposed RDG outperforms the other competitors by a big margin. For example, when KRR is 0.8, the PSNR gains over RMD, GSLS7 and CLBA are 15 dB, 12 dB and 10 dB, respectively. The PSNR gain over RCR is about 4 dB. In addition, another inferiority of RCR is that the maximally achievable KRR is only around 0.8 when removing keypoints in multiple octaves. We will show later that such limitation on KRR makes RCR insufficient for defeating some SIFT-based systems.

C. Performance against the KCR detector [28]

It was demonstrated in [28] that the SIFT keypoints lie in proximity of corners, and this property can be used to detect malicious manipulations of SIFT keypoints. Though the removal algorithms can hide the SIFT keypoints, the corners of the images do not change much after removal. The *Keypoint-to-Corner Ratio* (KCR) [28] detector was shown to be very effective in revealing the inconsistencies like the absence or the anomalous distribution of the keypoints with respect to the resulting images. To conceal the footprints of SIFT keypoint removal, some injection algorithms were proposed. Among them, the *Three-Phase SIFT Keypoint Injection* (TPKI) [26] method was illustrated to be very effective in injecting



(a) Octave 0 (b) Multiple octaves

Fig. 14: The average execution time on Dataset8. (a) The execution time comparison for octave 0; (b) the execution time comparison for multiple octaves.

fake keypoints and fooling the KCR detector. Nevertheless, it should be noted that in [26], [28], all the experiments are conducted for the keypoints in the first octave only. In this subsection, we further show the performance of RDG against the KCR detector for multiple octaves. The following types of attacks are considered.

- RDG(90%\70%\50%): remove $90\%\backslash 70\%\backslash 50\%$ of the keypoints within multiple octaves using RDG.
- RDG+TPKI: first remove the keypoints using RDG to the greatest extent, and then use TPKI algorithm to inject fake keypoints within multiple octaves.

We evaluate the performance over dataset UCID-v2, and use the attacked images as positive instances, while the original images as negative instances. In Fig. 13, we draw the ROC curves for the above attacks against the KCR detector, where the threshold of the KCR detector varies from -5 to 1 with step size being 0.02. We can observe that the KCR detector is still effective in detecting a standalone removal attack for multiple octaves. In addition, as the KRR increases, it becomes easier for KCR to detect the forged images. However, after being incorporated with the injection algorithm TPKI, the combined attack RDG+TPKI can still effectively pass the KCR detector. This is validated by the fact that the ROC curve of RDG+TPKI is very close to that of the random guess. The experimental results indicate that analyzing the distribution of the keypoints only is not sufficient for effectively revealing the footprints of the combined attack. It is necessary as well as important to investigate more advanced removal/injection detection algorithms to validate the input data, so as to achieve high reliability. One possible way is to analyze the different characteristics of the genuine SIFT keypoints and the injected ones, such that the forensic detector only needs to focus on the genuine keypoints.

D. Computational complexity

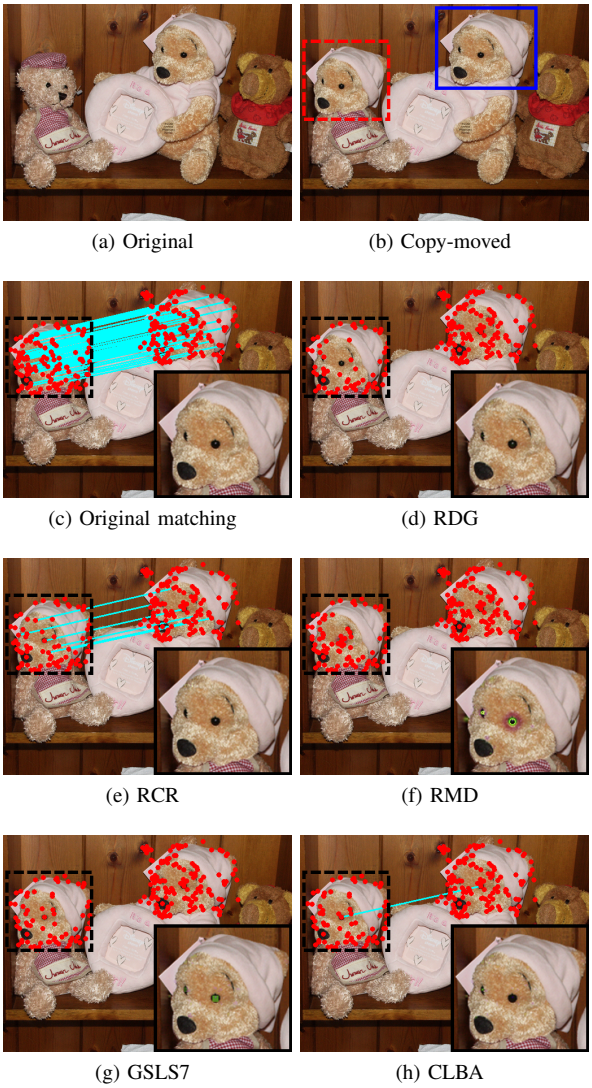
In this subsection, we give a short discussion on the computational complexity of the different removal algorithms. Fig. 14 shows the average running time for removing keypoints within octave 0 and within multiple octaves. All the results are evaluated on a desktop with Core-i7 and 8-GB RAM. We can observe that GSLS7 and RMD are much faster than the other algorithms, especially when removing keypoints within multiple octaves. This is because the main operations of GSLS7 and RMD are smoothing and evaluating closed-form solutions, which are fast. CLBA adopts GSLS7 for the first 10 iterations, and achieves low complexity at the beginning iterations. However, the complexity of CLBA increases quickly after a certain time, because the *collage* attack applied for additional iterations involves searching over a large database. For RCR and RDG, the computational complexity is higher than that of GSLS7 and RMD, since they need to solve a set of constrained optimization problems. Furthermore, for the case of multiple octaves, we notice that the complexity of RCR is higher than that of RDG when KRR is above 0.6. This is due to the phenomenon that RCR needs more iterations to achieve the same KRR when removing keypoints in multiple octaves. We should also emphasize that the KRR-D performance is generally much more crucial than the computational complexity in practice, as SIFT keypoint removal is typically conducted offline.

VII. Application to Copy-Move Anti-Forensics

Copy-move forgery is a popular manipulation of creating doctored images, where one or several parts of the image are pasted elsewhere in the same image intentionally, to duplicate or hide some objects of interest. Such process usually accompanies some appropriate geometric transformations [38], [39]. In this section, we further demonstrate the superiority of our proposed SIFT keypoint removal technique RDG in defeating the SIFT-based copy-move forgery detection system [8], which is shown to be quite effective in robustly exposing the cloned regions via SIFT keypoint matching. It should be noted that in addition to the SIFT-based techniques, there also exist some other image copy-move forgery detectors, such as block-based approaches [40], [41]; the discussion of these schemes is beyond the scope of this paper though. Different from [23], [25], [35], [36], where the SIFT keypoints are extracted only in a single octave (i.e., octave 0), we conduct the experiments for removing keypoints in multiple octaves, since most of the existing SIFT-based copy-move forgery detectors extract SIFT keypoints within multiple octaves, such as [8], [16].

One example is shown in Fig. 15, where the forged image is created by copying a bear head to cover another one. We can see that before keypoint removal, the cloned region can be readily detected by the copy-move forgery detector [8], where the SIFT keypoints are extracted in multiple octaves. However, as demonstrated in Fig. 15(d), upon applying our proposed RDG, the number of matched pairs drops to 0, and the distortion of the resulting image is still controlled to an imperceptible level. This shows the effectiveness of our proposed algorithm RDG. In addition, For comparison purposes, the results for RCR [26], RMD [19], GSLS7 [19] and CLBA [23] are given in Fig. 15(e)-(h). A detailed comparison is reported in Table II.

Furthermore, we also evaluate the performance of our proposed scheme over a copy-move forgery database MICC-F220 [8], which consists of 110 tampered images, and 110 genuine images. We use the detector proposed by Amerini *et al.* [8] to test each image for the copy-move



(a) Original (b) Copy-moved

(c) Original matching (d) RDG

(e) RCR (f) RMD

(g) GSLS7 (h) CLBA

Fig. 15: An example of countering copy-move forgery detection with different attack methods.

evidences. Before SIFT keypoint removal, the detector [8] is able to identify 99.09% of the tampered images, while only misclassifying 9.09% genuine images. Then, we use the proposed RDG to attack the 110 tampered images. After SIFT keypoint removal, the detector [8] cannot identify any tampered images, which indicates that RDG achieves 100% success rate for defeating the copy-move forgery detector [8].

The above results suggest the necessity and importance of devising more sophisticated forensic detectors for SIFT keypoint manipulations. As mentioned previously, investigating the different characteristics of genuine keypoints and fake keypoints may lead to a viable solution to this end.

VIII. CONCLUSIONS

In this work, we have presented a counter-forensic technique against the SIFT based systems, and demonstrated that the SIFT keypoints can be significantly removed with small distortion of the resulting images. By modeling the DoG space as a directed weighted graph, we have derived a set of inequality constraints to remove a SIFT keypoint along a pre-constructed acyclic path. To minimize the incurred distortion, the path has been strategically designed over the pre-constructed graph. In addition, we have proposed a simple yet effective method to recover the color information of the cleaned image via an optimization framework. Extensive experiments have demonstrated that our proposed removal method RDG can achieve much better KRR-D performance over the state-of-the-art techniques, under both the scenarios of removing keypoints in a single octave and in multiple octaves.

An implication of our results is that we should not fully trust the decisions made by the SIFT-based systems (even when they extract SIFT keypoints within multiple octaves), since the SIFT keypoints of the input images could be maliciously tampered, and the forged images can still successfully pass the keypoint removal/injection detector. Therefore, investigating more advanced detection algorithms is very important to validate the input data, so as to achieve high reliability. It should also be noted that this task can be very challenging when the removal/injection operations are only conducted in small regions, e.g., in the case of copy-move forgery.

TABLE II: Keypoints and SIFT matches when copy-move forgery is carried out with different attack methods

	#Keypoints	#Matches	Local PSNR	Local SSIM
No attack	150	73	∞	1
RMD	96	0	28.65	0.9685
GSLS7	82	0	33.68	0.9844
CLBA	87	2	38.91	0.9917
RCR	90	8	43.40	0.9969
RDG	78	0	49.06	0.9990

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Comp. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. on Computer Vision*, vol. 2. IEEE, 1999, pp. 1150–1157.

[3] H. Lejsek, F. Asmundsson, B. Jonsson, and L. Amsaleg, "NV-tree: An efficient disk-based index for approximate search in very large high-dimensional collections," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 31, no. 5, pp. 869–883, 2009.

[4] Z. Qin, J. Yan, K. Ren, C. W. Chen, and C. Wang, "Towards efficient privacy-preserving image feature extraction in cloud computing," in *Proc. ACM Int. Conf. on Multimedia*. ACM, 2014, pp. 497–506.

[5] C. Strecha, A. Bronstein, M. Bronstein, and P. Fua, "LDAHash: Improved matching with smaller descriptors," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 34, no. 1, pp. 66–78, 2012.

[6] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. on Inf. Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.

[7] H. Huang, W. Guo, and Y. Zhang, "Detection of copy-move forgery in digital images using SIFT algorithm," in *Proc. IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, vol. 2. IEEE, 2008, pp. 272–276.

[8] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy–move attack detection and transformation recovery," *IEEE Trans. on Inf. Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, 2011.

[9] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, L. D. Tongo, and G. Serra, "Copy-move forgery detection and localization by means of robust clustering with j-linkage," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 659–669, 2013.

[10] K. Grauman and T. Darrell, "Efficient image matching with distributions of local invariant features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2005, pp. 627–634.

[11] Y. Ke, R. Sukthankar, and L. Huston, "Efficient near-duplicate detection and sub-image retrieval," in *Proc. ACM Int. Conf. on Multimedia*, vol. 4, no. 1. ACM, 2004, pp. 869–876.
[12] I. Skrypnyk and D. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *Proc. IEEE and ACM Int. Symp. Mixed and Augmented Reality*. IEEE, 2004, pp. 110–119.
[13] I. Gordon and D. Lowe, "What and where: 3D object recognition with accurate pose," in *Toward Category-Level Object Recognition*. Springer, 2006, vol. 4170, pp. 67–82.
[14] Y. Han, J. Yin, and J. Li, "Human face feature extraction and recognition base on SIFT," in *Proc. IEEE Int. Symposium on Computer Science and Computational Technology*, vol. 1. IEEE, 2008, pp. 719–722.
[15] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. L. Lu, "Person-specific SIFT features for face recognition," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 2. IEEE, 2007, pp. 593–596.
[16] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Trans. on Inf. Forensics and Security*, vol. 5, no. 4, pp. 857–867, 2010.
[17] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. on Inf. Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.
[18] C.-Y. Hsu, C.-S. Lu, and S.-C. Pei, "Secure and robust SIFT," in *Proc. ACM Int. Conf. on Multimedia*. ACM, 2009, pp. 637–640.
[19] T.-T. Do, E. Kijak, T. Furon, and L. Amsaleg, "Deluding image recognition in SIFT-based CBIR systems," in *Proc. ACM workshop on Multimedia in Forensics, Security and Intell.* ACM, 2010, pp. 7–12.
[20] T.-T. Do, E. Kijak, L. Amsaleg, and T. Furon, "Enlarging hacker's toolbox: deluding image recognition by attacking keypoint orientations," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. IEEE, 2012, pp. 1817–1820.
[21] C.-S. Lu and C.-Y. Hsu, "Constraint-optimized keypoint inhibition/insertion attack: security threat to scale-space image feature extraction," in *Proc. ACM Int. Conf. on Multimedia*. ACM, 2012, pp. 629–638.
[22] R. Caldelli, I. Amerini, L. Ballan, G. Serra, M. Barni, and A. Costanzo, "On the effectiveness of local warping against SIFT-based copy-move detection," in *Proc. IEEE Int. Symp. Commun., Control, Signal Process*. IEEE, May 2012, pp. 1–5.
[23] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo, "Counter-forensics of SIFT-based copy-move detection by means of keypoint classification," *EURASIP J. on Image and Video Proc.*, vol. 2013, no. 1, pp. 1–17, 2013.
[24] I. Amerini, F. Battisti, R. Caldelli, M. Carli, and A. Costanzo, "Exploiting perceptual quality issues in countering SIFT-based forensic methods," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. IEEE, May 2014, pp. 2664–2668.
[25] A. Cheng, Y. Li, and J. Zhou, "SIFT keypoint removal via convex relaxation," in *Proc. IEEE Int. Conf. on Multimedia and Expo*. IEEE, June 2015, pp. 1–6.
[26] Y. Li, J. Zhou, A. Cheng, X. Liu, and Y. Y. Tang, "SIFT keypoint removal and injection via convex relaxation," *IEEE Trans. on Inf. Forensics and Security*, vol. 11, no. 8, pp. 1722–1735, 2016.
[27] T.-T. Do, E. Kijak, T. Furon, and L. Amsaleg, "Understanding the security and robustness of SIFT," in *Proc. ACM Int. Conf. on Multimedia*. ACM, 2010, pp. 1195–1198.
[28] A. Costanzo, I. Amerini, R. Caldelli, and M. Barni, "Forensic analysis of SIFT keypoint removal and injection," *IEEE Trans. on Inf. Forensics and Security*, vol. 9, no. 9, pp. 1450–1464, 2014.
[29] B. Ni, P. Moulin, and S. Yan, "Order preserving sparse coding," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 37, no. 8, pp. 1615–1628, 2015.
[30] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[31] R. L. R. Thomas H. Cormen, Charles E. Leiserson and C. Stein, *Introduction to algorithms*, 3rd ed. MIT press, 2009.
[32] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.
[33] G. Schaefer and M. Stich, "UCID - An uncompressed colour image database," in *Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia*, vol. 5307, 2004, pp. 472–480.
[34] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms (2008)," 2012.
[35] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo, "SIFT keypoint removal and injection for countering matching-based image forensics," in *Proc. ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 123–130.
[36] I. Amerini, M. Barni, R. Caldelli, and A. Costanzo., "Removal and injection of keypoints for SIFT-based copy-move counter-forensics," *EURASIP J. on Image and Video Proc.*, vol. 2013, no. 1, pp. 1–12, 2013.
[37] W. Dong, G. Shi, Y. Ma, and X. Li, "Image restoration via simultaneous sparse coding: Where structured sparsity meets gaussian scale mixture," *Int. J. Comput. Vis.*, vol. 114, no. 2, pp. 217–232, 2015.
[38] A. J. Fridrich, B. D. Soukal, and A. J. Lukáš, "Detection of copy-move forgery in digital images," in *Proc. Digital Forensic Research Workshop*. Citeseer, 2003.
[39] O. M. Al-Qershi and B. E. Khoo, "Passive detection of copy-move forgery in digital images: State-of-the-art," *Forensic Science International*, vol. 231, no. 1, pp. 284–295, 2013.
[40] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient dense-field copy-move forgery detection," *IEEE Trans. on Inf. Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, 2015.
[41] S.-J. Ryu, M. Kirchner, M.-J. Lee, and H.-K. Lee, "Rotation invariant localization of duplicated image regions based on zernike moments," *IEEE Trans. on Inf. Forensics and Security*, vol. 8, no. 8, pp. 1355–1370, 2013.

Yuanman Li (S'15) received the B.S. in software engineering from Chongqing University, Chongqing, China, in 2012, and the M.S. degree in software engineering from University of Macau, Macau, China, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China. His research interests include multimedia security, pattern recognition and machine learning.

Jiantao Zhou (M'11) received the B.Eng. degree from the Department of Electronic Engineering, Dalian University of Technology in 2002, the M.Phil. degree from the Department of Radio Engineering, Southeast University in 2005, and the Ph.D. degree from the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology in 2009. He held various research positions with the University of Illinois at Urbana-Champaign, the Hong Kong University of Science and Technology, and McMaster University. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. His research interests include multimedia security and forensics, and multimedia signal processing. He was a coauthor of two papers that received the best paper award at the IEEE Pacific-Rim Conference on Multimedia (PCM) in 2007, and the best student paper award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, respectively. He holds four granted U.S. patents and two granted Chinese patents.

An Cheng received B.S degree from Zhuhai College of Jilin University, China in 2012 and the M.S degree from University of Macau in 2015. He is currently a researcher at Meitu, Inc. His research interests include computer vison and video processing.