JPEG Steganography with Estimated Side-information

Weixiang Li, Kejiang Chen, Weiming Zhang, Hang Zhou, Yaofei Wang, Nenghai Yu

Abstract—Previous studies have exhibited that incorporating side-information, e.g., a high-quality “precover” image, can significantly improve steganographic security for JPEG images. This motivates us to estimate the side-information for traditional steganographic scenario in which only a JPEG image is available. It is expected to achieve high-level security by utilizing the estimated side-information similar to side-informed steganography, even though the estimated side-information is not perfectly precise. In this paper, a general framework of side-information estimated (SIE) JPEG steganography is proposed, under which the core problems are how to better estimate the precover and modulate the distortion function correspondingly. To address the two problems, we test several denoising filters and a deblocking filter to obtain the estimated precover, and we introduce two implementation models for modulating the costs. We finally recommend the combination of the deblocking filter and the modulation model using the polarity of the estimated rounding error. The experimental results show that the proposed method dramatically improves the existing additive distortions for images of an arbitrary quality factor and outperforms state-of-the-art methods based on estimating side-information when resisting modern steganalysis.

Index Terms—Steganography, JPEG images, minimal distortion, side-information, denoising, deblocking.

I. INTRODUCTION

Modern steganography is a science and art of covert communication that slightly modifies a digital cover object to transmit a covert message without drawing suspicions from steganalysis [1]. Since JPEG images are the widely adopted format for image storage and transmission, steganography on JPEG images has become a research hotspot over the past few years. Based on the minimal distortion model [2], various content-adaptive distortion functions [3]–[6] are designed for JPEG steganography by preferably exploiting image texture complexity to strengthen the steganographic security. Meanwhile, microscale steganography and the cost spreading rule [7], the JPEG controversial-pixel-prior rule [8] and the block-boundary-continuity principle [9] are extended from spatial image steganography to help improve the performance of the above additive distortion functions.

Denote $x$ as the cover element and $I$ as the range of the embedding operation at $x$. In the context of digital image steganography, ternary embedding ($\pm 1$ embedding with $I = \{x - 1, x, x + 1\}$) is more commonly used than binary embedding (with $I = \{x, x\}$ where $x$ is $x$ flipping its Least Significant Bit (LSB)), since it can achieve a smaller embedding impact. In $\pm 1$ embedding, the costs of changing the quantized JPEG coefficient by $+1$ and $-1$ are equivalent [3]–[8]. It is widely recognized that incorporating side-information at the sender can significantly improve steganographic security in practice, where the costs of $+1$ and $-1$ are not the same but modulated by some additional information. For JPEG steganography, the side-information may be in the form of an uncompressed image (called the precover [10]) or concretely as the non-rounded DCT coefficient, which partially compensates for the lack of knowledge of the cover model when it is highly non-stationary. Numerous heuristic cost-modulated schemes were introduced in [3], [7], [11]–[14]. Side-informed (SI) steganography [13] allowed a ternary embedding operation rather than a binary approach and computed the costs from the uncompressed cover, both of which appeared to improve the empirical security of existing distortion functions by a rather large margin. Under the condition of the sender lacking access to a precover, [14] used a set of multiple JPEG images of the same scene to modulate the costs, achieving a high-level steganographic security. By utilizing the additional information that is unavailable to the Warden, the hope is that the embedding will disturb the statistical properties of the cover source less.

In the traditional and common steganographic scenario (non-SI steganography [3]–[9]), the sender has no access to a precover but experiences only one JPEG image for message embedding. The high-level security of SI steganography benefits from the premise that the sender has the precover [13] or multiple images of the same scene [14]. Motivated by SI steganography, relatively high security of non-SI steganography can be expectedly achieved if we can estimate the precover from the JPEG cover as precisely as possible. Recently, [15], [16] attempted to estimate the side-information with the average and the Wiener filter, respectively, for improving the distortion functions. However, the filters for estimating the precover seem to be primitive, and the method of minimizing the distance of steganalytic feature space leads to excessive time consumption for message embedding. Therefore, a universal and efficient framework of estimating side-information for JPEG steganography needs to be established.

In this paper, we focus on the traditional non-SI steganographic scenario and propose a general framework of side-information estimated (SIE) JPEG steganography, with its
several implementation methods for improving the existing distortion functions. Under the SIE framework, the core problems for strengthening the steganographic security include estimating the precover as precisely as possible and modulating the distortion according to the estimated non-rounded coefficient. Based on the simple SIE model that only considers the polarity of the rounding error, we examine the performance of various denoising and deblocking filters on estimating the precover. The experimental results show that the proposed SIE model equipped with the selected deblocking filter can significantly improve the empirical security of existing JPEG additive distortion functions. The proposed method also outperforms state-of-the-art methods using estimated side-information.

The rest of this paper is organized as follows. In Section II, we briefly review side-informed steganography and introduce its degradation model. The general framework of estimating side-information for JPEG steganography is proposed with several implementations in Section III. The experimental results and comparisons are presented in Section IV, and the paper is concluded in Section V.

II. JPEG STEGANOGRAPHY WITH PRECOVER

With the helpful additional information of precover [13] or the same scene-based multiple JPEG images [14] that are used to modulate the costs, side-informed (SI)-based JPEG steganography significantly improved the steganographic security of the existing additive distortion functions. In [13], a $\pm 1$ embedding version of SI-based steganography was studied using the non-rounded DCT coefficients after compressing the precover. Denote $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, $\mathbf{u} = (u_1, u_2, \cdots, u_n)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ as the cover (quantized DCT coefficients for embedding), the non-rounded coefficients of the precover, and the stego, respectively. The rounding error $e_i = u_i - x_i$ ($|e_i| \leq 0.5$, $1 \leq i \leq n$) is used to adjust the original cost $\rho_i^{(A)}$ of changing $x_i$ by $\pm 1$ with

$$
\begin{cases} 
\rho_i^{(SI)+} = (1 - 2|e_i|)\rho_i^{(A)} & \text{if } y_i = x_i + \text{sign}(e_i) \\
\rho_i^{(SI)-} = \rho_i^{(A)} & \text{if } y_i = x_i - \text{sign}(e_i)
\end{cases}
$$

(1)

where $\rho_i^{(A)}$ can be defined by any distortion function $\mathcal{A} [3]–[8]$. It is recommended in [13] to compute the costs from the precover $\mathbf{u}$ instead of the cover $\mathbf{x}$. Intuitively, the modulated costs (1) not only reflect the local image complexity but also account for the distortion w.r.t. the precover. With the near-optimal STCs [2], the actual embedding will approach the minimal average distortion $E_{\mathcal{A}}(D) = \sum_{i=1}^{n} (\pi_i^{(SI)+} + \pi_i^{(SI)-})$ by modifying $x_i$ by $\pm \text{sign}(e_i)$ with probability

$$
\pi_i^{(\pm)} = \frac{\exp(-\lambda\rho_i^{(SI)\pm})}{1 + \exp(-\lambda\rho_i^{(SI)+}) + \exp(-\lambda\rho_i^{(SI)-})}
$$

(2)

with $\lambda (\lambda > 0)$ determined by the message length of $m$ bits

$$
m = -\sum_{i=1}^{n} \left( \pi_i^{(+)\log_2 \pi_i^{(+)} + \pi_i^{(-)\log_2 \pi_i^{(-)}} + (1 - \pi_i^{(+) - \pi_i^{(-)}})\log_2(1 - \pi_i^{(+) - \pi_i^{(-)}}) \right).
$$

(3)

We mark the specific SI-based method as SI-$\mathcal{A}^{(*)}$ where $* \in \{u, x\}$ represents the cost computation from the precover $\mathbf{u}$ or the cover $\mathbf{x}$.

Here, we introduce a degradation model of the SI steganography that only focuses on the polarity of the rounding error. In the SI-polarity (SIp) model, the cost modulation neglects the amplitude of the rounding error by adjusting each cost with the same parameter $\alpha (0 \leq \alpha \leq 1)$, that is,

$$
\begin{cases} 
\rho_i^{(SIp)+} = \alpha \cdot \rho_i^{(A)} & \text{if } y_i = x_i + \text{sign}(e_i) \\
\rho_i^{(SIp)-} = \rho_i^{(A)} & \text{if } y_i = x_i - \text{sign}(e_i)
\end{cases}
$$

(4)

Correspondingly, we mark the specific SIp-based method as SIp-$\mathcal{A}^{(*)}$. The SI and SIp models introduced here inspire the similar models of the proposed framework of estimating side-information for JPEG steganography as described below.

III. FRAMEWORK OF ESTIMATING SIDE-INFORMATION FOR JPEG STEGANOGRAPHY AND ITS IMPLEMENTATIONS

Under the traditional and common steganographic scenario, the sender possesses only a JPEG image without any side-information, i.e., the sender has no access to a precover. Inspired by the SI steganography, a clever sender may estimate some sufficiently accurate side-information from the JPEG image to aid the steganography, i.e., he or she may be able to modulate the costs with the help of the estimated side-information.

![Fig. 1: General framework of side-information estimated (SIE) JPEG steganography.](Image 365x340 to 510x390)

In this section, we propose a general framework of side-information estimated (SIE) JPEG steganography as depicted in Fig. 1. The SIE framework is formulated as

$$
\mathcal{S}(\mathcal{F}) \cdot \mathcal{A}^{(*)}.
$$

In Fig. 1, the JPEG image $\mathbf{x}$ is used to estimate the precover $\hat{\mathbf{u}}$ (the estimated non-rounded coefficient) with a filter $\mathcal{F}$, and the original cost $\rho$ can be defined by an arbitrary distortion function $\mathcal{A}^{(*)}$ with $* \in \{x, \hat{u}\}$ representing the cost computation of $\mathbf{x}$ or $\hat{\mathbf{u}}$. Moreover, in terms of the update strategy $\mathcal{S}$, the SIE-based cost $\rho^{(SIE)}$ is obtained by modulating $\rho$ with $\hat{\mathbf{u}}$. Obviously, under the SIE framework, the two core issues for improving the steganographic security are the following: 1) how to precisely estimate $\hat{\mathbf{u}}$, i.e., choose better $\mathcal{F}$, and 2) how to adjust the costs according to $\hat{\mathbf{u}}$, i.e., design better $\mathcal{S}$. If the estimated $\hat{\mathbf{u}}$ is sufficiently close to the precover $\mathbf{u}$, we can simply refer to the SI method (1). However, it is of practical significance to design an appropriate update strategy $\mathcal{S}$ when $\hat{\mathbf{u}}$ cannot be perfect enough.
A. Estimating the precover with various filters

The first critical problem under the SIE framework, that is, how to better estimate the precover, can be regarded as the image restoration problem. It is known that the information loss for JPEG compression takes place in the stage of quantization, leading to round-off errors in each block, which inevitably produces blocking artifacts. To suppress blocking artifacts and obtain a high-quality estimated precover, several denoising and deblocking filters are studied in this paper.

1) Denoising filters: The denoising filters are applied in the spatial domain, which is the decompression image of the JPEG image. We introduce the average filter, median filter, Gaussian filter and Wiener filter, which are abbreviated as Avg, Med, Gau and Wie, respectively, and each has window sizes of $3 \times 3$ and $5 \times 5$. The output of the average filter is simply the average value of pixels contained in the neighborhood of the filter mask. In the median filter, the current pixel is replaced with the median value among its neighboring pixels. The Gaussian filter is a non-uniform low-pass filter, which modifies the input signal by convolution with a Gaussian function. The Wiener filter is optimal in terms of the mean square error by using a pixelwise adaptive Wiener filtering method based on statistics estimated from a local neighborhood of each pixel. Specifically, the Gaussian and Wiener filtering outputs are obtained using \texttt{fspecial}('gaussian') and \texttt{wiener2} in MATLAB, respectively.

2) Deblocking filter: We also select a type of image-restoration-based deblocking filter, called SSRQC [17], to obtain a better estimated precover. Image deblocking is usually formulated as an ill-posed image inverse problem by exploiting the information in the JPEG compressed bit-stream, such as the decompressed image and the quantization matrix. SSRQC [17] was proposed for image deblocking by using a structural sparse representation (SSR) prior and a quantization constraint (QC) prior with a new split Bregman iteration-based method, which greatly improved the existing image-deblocking quality.

From the above introduction, the candidate filter \( \mathcal{F} \in \{ \text{Avg}(w), \text{Med}(w), \text{Gau}(w), \text{Wie}(w), \text{SSRQC} \} \) with window size \( w \in \{3 \times 3, 5 \times 5\} \) will be investigated in this paper.

B. Modulating the costs with two implementation models

To further enrich the SIE framework, we provide here two heuristic update strategies that are inspired by (1) and (4). Denote \( \hat{e}_i = \tilde{u}_i - x_i \) as the estimated rounding error. The first strategy, which we name SIEg model, utilizes \( \hat{e}_i \) to modulate the original cost \( \rho_i^{(A)} \) by \( g(\hat{e}_i) \), that is,

\[
\begin{align*}
\rho_i^{(\text{SIEg})+} &= g(\hat{e}_i) \cdot \rho_i^{(A)} \quad \text{if } y_i = x_i + \text{sign}(\hat{e}_i), \\
\rho_i^{(\text{SIEg})-} &= g(\hat{e}_i) \cdot \rho_i^{(A)} \quad \text{if } y_i = x_i - \text{sign}(\hat{e}_i).
\end{align*}
\]

(5)

Note that the value range of \( \hat{e}_i \) is erratic (not in the value range \([0, 0.5]\) as SI) and determined by the precision of \( \tilde{u} \). Therefore, we carefully design a heuristic form of \( g(\cdot) \),

\[ g(\hat{e}_i) = \begin{cases} 
1 - 2|\hat{e}_i| & \text{if } |\hat{e}_i| \leq 0.5 \\
\beta & \text{otherwise}
\end{cases} \]

(6)

where \( \beta (0 \leq \beta \leq 1) \) is to ensure that the cost corresponding to \( |\hat{e}_i| > 0.5 \) remains non-negative when using the modulation according to (1).

Since it is easier to estimate the accurate polarity of the rounding error than the amplitude of the rounding error, we introduce a compromising strategy inspired by the SIE model (4). The second strategy, called the SIE-polarity (SIEp) model, solely considers the polarity of \( \hat{e}_i \) by assigning the same parameter \( \gamma (0 \leq \gamma \leq 1) \) for modulating the cost of \( \pm \text{sign}(\hat{e}_i) \),

\[
\begin{align*}
\rho_i^{(\text{SIEp})+} &= \gamma \cdot \rho_i^{(A)} \quad \text{if } y_i = x_i + \text{sign}(\hat{e}_i), \\
\rho_i^{(\text{SIEp})-} &= \rho_i^{(A)} \quad \text{if } y_i = x_i - \text{sign}(\hat{e}_i).
\end{align*}
\]

(7)

Therefore, the cost update strategy \( S \in \{ \text{SIEg}, \text{SIEp} \} \) will be investigated in this paper.

C. Generalizing other related methods into the SIE framework

The methods in [15], [16] also attempted to utilize the estimated side-information for embedding, which can be considered as two implementation instances of the proposed SIE framework. Specifically, the method in [15] obtained the estimated precover by replacing the decompressed boundary pixels with the \( 3 \times 3 \) average filtered boundary pixels (i.e., another \( \mathcal{F} \)) and adjusted the costs by the estimated quantized DCT coefficients instead of the estimated non-rounded coefficients (i.e., another \( S \)). Similarly, the method in [16] used the \( 3 \times 3 \) Wiener filter to estimate the precover and searched the best parameter for cost modulation by minimizing the steganalytic feature distance (i.e., another \( S \)), which is time-consuming with 20-message embedding and 20-feature extraction. Comparatively, the proposed general SIE framework enables the combination of various \( S, \mathcal{F} \) and \( A \) for designing a better steganographic scheme.

IV. Experimental Results and Analysis

In this section, we will demonstrate the performance of different SIE-based methods and the comparison with the methods in [15], [16]. The experiments are mainly conducted on BOSSBase 1.01 [18], which contains 10,000 gray-scale images of size \( 512 \times 512 \) pixels. All of the images are compressed into the JPEG domain with quality factors \( QF=50, 75 \) and 95, which are adopted as datasets for experimental comparisons. To verify the generalization of the proposed method, we perform experiments on another popular image set BOWS2 [19], which also contains 10,000 gray-scale images of size \( 512 \times 512 \) and is JPEG compressed by \( QF=75 \). We use the mainstream distortion functions UERD [4] and J-UNIWARD [3] with the optimal simulator [20] for message embedding, and the steganalyzer is trained by using state-of-the-art DCTR-8,000D [21] and GFR-17,000D [22] with the FLD ensemble [23] by default. The FLD ensemble can minimize the total classification error probability under equal priors \( P_E = \min P_{A0} \frac{1}{2}(P_{FA} + P_{MD}) \) where \( P_{FA} \) and \( P_{MD} \) are the false-alarm probability and the missed-detection probability, respectively. The ultimate security is qualified by the average error rate \( \overline{P_E} \) averaged over 10 random 5000/5000 splits of the dataset, and larger \( \overline{P_E} \) means stronger security.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2925118, IEEE Transactions on Circuits and Systems for Video Technology

1. Investigation on different $\mathcal{F}$ on estimating the precover: The goal of employing $\mathcal{F}$ on the JPEG image is to obtain an estimated precover as close as possible to the real precover. Intuitively, we can evaluate the effect of different $\mathcal{F}$ via the peak signal-to-noise ratio (PSNR) between the real precover and the estimated precover, where $\mathcal{F}$ with larger PSNR may have a better ability to estimate the precover. In addition, we measure the ratio of correct polarities of the estimated rounding errors, i.e., $R_p = (\sum_{i=1}^{n} \delta_{sign}(\hat{e}_i) = sign(e_i))/n$ where the Iverson bracket $[\cdot]$ is defined to be 1 if the logical expression $I$ is true and 0 otherwise. Because the polarity of the rounding error, which directs the DCT coefficient to be rounded to “the other side” [13], is significantly important for the steganographic security, $\mathcal{F}$ of higher $R_p$ with more correct polarities may also correspond to stronger security.

We randomly select 1,000 JPEG images from BOSSBase of QF=75 and perform different $\mathcal{F}$ on these images to observe the average PSNR and $R_p$. As shown in TABLE I, each denoising filter of the window size $3 \times 3$ has a larger PSNR and $R_p$ than that of the $5 \times 5$ window, which implies that the filter size should be sufficiently small because of the strong correlation among neighboring pixels. Among denoising filters, the $3 \times 3$ Wiener filter achieves the best performance. However, the deblocking filter $SSRCQ$ is even better than the $3 \times 3$ Wiener filter. Therefore, we believe that the deblocking filter designed for technically removing blocking artifacts and restoring the image is more suitable for estimating the precover.

Since the SIEp model is easier to be implemented, we first investigate the impact of different $\mathcal{F}$ on improving the steganographic security when combined with the SIEp model. As shown in TABLE II, the security w.r.t. $\mathcal{F}$ has a consistent trend that is similar to the PSNR and $R_p$ in TABLE I. The $3 \times 3$ Wiener filter is the best denoising filter, and it is still worse than the deblocking filter $SSRCQ$. Therefore, we select the best filter $SSRCQ$ while assigning the optimal $\gamma = 0.65$ to SIEp in (7) for the following experiments.

2) Investigation on SI-based and SIE-based models: We now study the gap between the SI-based and SIE-based methods as well as the impact of cost computation of $x$ or $u/u$.

Similar to the choice of $\gamma$ as shown in TABLE II, the optimal parameters in TABLE III, such that $\alpha = 0.4$ for Slp in (4) and $\beta = 0.65$ for SIEg in (6), are determined at 0.3bpnzac UERD on BOSSBase of QF=75 against DCTR by traversal search with a step of 0.05. According to the experimental results, the selected optimal parameters ($\alpha = 0.4$, $\beta = 0.65$, $\gamma = 0.65$) are applicable to other distortion functions, relative payloads, quality factors and image sources.

The security of Slp-$A(\cdot)$ is not as superior as that of SI-$A(\cdot)$, but it is still far better than the non-SI $A(\cdot)$ even though it only utilizes the direction of the rounding error. SIEp($SSRCQ$)-$A(\cdot)$ is naturally worse than SI-$A(\cdot)$ (the upper bound with the real precover), and the security of

<table>
<thead>
<tr>
<th>Filter $\mathcal{F}$</th>
<th>Avg($3\times3$)</th>
<th>Med($3\times3$)</th>
<th>Gau($3\times3$)</th>
<th>Wte($3\times3$)</th>
<th>$SSRCQ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PSNR$ (dB) $R_p$ (%)</td>
<td>33.36</td>
<td>34.25</td>
<td>34.99</td>
<td>37.12</td>
<td>38.92</td>
</tr>
<tr>
<td>55.79</td>
<td>56.02</td>
<td>56.09</td>
<td>57.59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Filter $\mathcal{F}$</th>
<th>Avg($5\times5$)</th>
<th>Med($5\times5$)</th>
<th>Gau($5\times5$)</th>
<th>Wte($5\times5$)</th>
<th>$SSRCQ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PSNR$ (dB) $R_p$ (%)</td>
<td>28.39</td>
<td>30.79</td>
<td>31.65</td>
<td>34.62</td>
<td></td>
</tr>
<tr>
<td>53.19</td>
<td>54.46</td>
<td>54.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme</th>
<th>$A(\cdot)$</th>
<th>$SI-A(\cdot)$</th>
<th>$SIp-A(\cdot)$</th>
<th>$SIEg(\cdot)$</th>
<th>$SIEp(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UERD</td>
<td>23.06</td>
<td>45.07</td>
<td>32.27</td>
<td>17.50</td>
<td>37.42</td>
</tr>
<tr>
<td>SI-UNIWARD</td>
<td>24.35</td>
<td>44.82</td>
<td>31.60</td>
<td>17.54</td>
<td>40.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>27.77</td>
<td>29.56</td>
<td>30.50</td>
<td>31.46</td>
<td>31.32</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>30.75</td>
<td>31.88</td>
<td>31.90</td>
<td>31.81</td>
<td>31.15</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>27.85</td>
<td>30.09</td>
<td>33.59</td>
<td>34.31</td>
<td>31.37</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>32.74</td>
<td>33.69</td>
<td>37.32</td>
<td>36.73</td>
<td>32.30</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>34.85</td>
<td>36.28</td>
<td>37.42</td>
<td>37.06</td>
<td>34.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>26.09</th>
<th>28.07</th>
<th>28.33</th>
<th>28.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>29.04</td>
<td>29.39</td>
<td>29.14</td>
<td>28.67</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>27.20</td>
<td>29.12</td>
<td>24.46</td>
<td>29.15</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>31.17</td>
<td>31.14</td>
<td>30.31</td>
<td>29.50</td>
</tr>
</tbody>
</table>
A. Universality verification of the proposed SIE method

SIEp(SSRQC)-A(*) is terrible, given the fact that estimating the precise amplitude of the rounding error is more difficult. What is beyond our expectation is that SIEp(SSRQC)-A(*) outperforms Slp-A(*). Since Slp and SIEp are focusing on the polarity of the rounding error, Slp should be the ideal bound that SIEp can approach with more correct polarities. Interestingly, this phenomenon implies that simply pursuing larger $R_p$ may not be the best choice for improving the SIEp method. We attempt to explore and explain this phenomenon via the distribution of the rounding errors whose polarities are correctly estimated. Suppose a DCT coefficient $x_1$ with the cost $p_1$ and the rounding error $|e_1| = 0.4$ and another DCT coefficient $x_2$ with the same cost $p_2 = p_1$ and the rounding error $|e_2| = 0.2$. According to SI (1), the modulated cost of $x_1$ is smaller than that of $x_2$ (i.e., $x_1$ is more suitable for modification), which distinguishes the modification priorities of these two coefficients with different rounding errors and thus leads to the high-level security of SI. However, in view of Slp (4), these two coefficients still experience the same modulated costs regardless of the amplitudes of their rounding errors. Since only a part (w.r.t. $R_p$) of coefficients whose polarities of the rounding errors are correctly estimated (i.e., $\text{sign}(\hat{e}_i) = \text{sign}(e_i)$) can be selected for cost modulation, the modification priorities of the coefficients with different $|e|$ will be reasonably reflected if more coefficients with larger $|e|$ are selected for cost modulation. As verified in Fig. 2, with increasing the amplitude of $|e|$, the ratio of the selected $|e|$ w.r.t. SSRQC increases. In this way, it is very likely that $x_1$ with $|e_1| = 0.4$ is selected by SIEp and $x_2$ with $|e_2| = 0.2$ is not, such that the modification priorities of them are reflected similar to the optimal SI. Overall, unlike Slp that gives the same priorities to the coefficients with different $|e|$s, SIEp will focus more on the coefficients with larger $|e|$s that are more suitable for modification, making the modification priorities of the coefficients with different $|e|$s more reasonable. And we believe that this contributes to the secure advantage of SIEp to Slp.

As pointed out in [13], the costs of SI-A(*) computed from the real precovers $u$ achieve better security than that from the JPEG cover $x$ because more information about the precovers source is lost due to JPEG compression. However, since $u$ does not perfectly approach the real precovers, the costs of $u$ do not show superior performance for the SIE-based methods when compared with the costs of $x$. Instead, the bold data in TABLE III demonstrate the advantage of calculating the costs of $x$ for the SIEp method. Therefore, SIEp(SSRQC)-A(*) is the recommended implementation method of the proposed SIE framework for the following experiments.

B. Universality verification of the proposed SIE method

We test the universality of the proposed SIEp(SSRQC)-A(*) via using the mainstream distortion functions UERD [4] and J-UNIWARD [3] with relative embedding payloads $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ bpnzac (bit per nonzero AC coefficient), on two image sets of different sources and quality factors in resisting the detections of state-of-the-art steganalytic features DCTR-8,000D [21] and GFR-17,000D [22]. Fig. 3 demonstrates that the proposed method can well approach the SI method and improve the traditional non-SI distortion functions by a large margin. TABLE IV verifies that the proposed method can be applied to image sets of relatively small or large quality factors and another image source.

C. Comparison with other related methods

As mentioned before, the methods in [15], [16] can be generalized as two implementation instances of the proposed SIE framework. As shown in TABLE V. SIEp($arg(3\times3)$)-UERD(x) using the same filter outperforms the method in

![Fig. 3: Performance of the proposed method SIEp(SSRQC)-A(*) using UERD/J-UNIWARD at 0.3bpnzac on other image sets of different sources and quality factors.](image)

<table>
<thead>
<tr>
<th>Set-QF</th>
<th>Scheme</th>
<th>DCTR</th>
<th>GFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS-QF50</td>
<td>UERD</td>
<td>18.19</td>
<td>13.79</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-UERD(x)</td>
<td>35.85</td>
<td>30.31</td>
</tr>
<tr>
<td></td>
<td>J-UNIWARD</td>
<td>18.51</td>
<td>13.10</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-JUNI(x)</td>
<td>37.65</td>
<td>32.78</td>
</tr>
<tr>
<td>BOSS-QF95</td>
<td>UERD</td>
<td>37.20</td>
<td>32.14</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-UERD(x)</td>
<td>44.63</td>
<td>42.84</td>
</tr>
<tr>
<td></td>
<td>J-UNIWARD</td>
<td>40.36</td>
<td>35.55</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-JUNI(x)</td>
<td>45.54</td>
<td>44.77</td>
</tr>
<tr>
<td>BOWS2-QF75</td>
<td>UERD</td>
<td>23.93</td>
<td>17.33</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-UERD(x)</td>
<td>37.23</td>
<td>33.08</td>
</tr>
<tr>
<td></td>
<td>J-UNIWARD</td>
<td>26.62</td>
<td>18.77</td>
</tr>
<tr>
<td></td>
<td>SIEp(SSRQC)-JUNI(x)</td>
<td>39.65</td>
<td>35.15</td>
</tr>
</tbody>
</table>

TABLE V: Detection errors $\overline{F}_2$ (%) of the proposed SIEp(SSRQC)-A(*) using UERD/J-UNIWARD at 0.3bpnzac on other image sets of different sources and quality factors.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>DCTR</th>
<th>GFR</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>UERD</td>
<td>23.06</td>
<td>17.81</td>
<td>0.77</td>
</tr>
<tr>
<td>Method in [15] upon UERD</td>
<td>28.87</td>
<td>21.61</td>
<td>0.79</td>
</tr>
<tr>
<td>Method in [16] upon UERD</td>
<td>34.79</td>
<td>28.28</td>
<td>36.69</td>
</tr>
<tr>
<td>SIEp($arg(3\times3)$)-UERD(x)</td>
<td>30.50</td>
<td>25.35</td>
<td>0.78</td>
</tr>
<tr>
<td>SIEp($Wie(3\times3)$)-UERD(x)</td>
<td>33.92</td>
<td>27.57</td>
<td>0.80</td>
</tr>
<tr>
<td>SIEp(SSRQC)-UERD(x)</td>
<td>37.42</td>
<td>33.05</td>
<td>14.15</td>
</tr>
<tr>
<td>SIEp(SSRQC)-UERD(x) + MinDistFea</td>
<td>37.88</td>
<td>34.14</td>
<td>50.11</td>
</tr>
</tbody>
</table>
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2925118, IEEE Transactions on Circuits and Systems for Video Technology

which indicates the reasonability of directly using the filtered image as the estimated precover. The method in [16] is slightly better than SIEp(Wie(3 x 3))-UERD(x) but at the cost of heavy time consumption, which is unacceptable in real-time online applications. Instead, the execution time of SIEp(Wie(3 x 3))-UERD(x) only performing STCs once is negligible. When employing the better filter SSRQC, SIEp(SSRQC)-UERD(x) outperforms the method in [16] by 2.63% and 4.77%, respectively, in resisting DCTR and GFR. Although SIEp(SSRQC)-UERD(x) takes 14.15s for message embedding because of the high computational complexity of SSRQC, it is still much faster than the method in [16]. Obviously, SIEp(SSRQC)-UERD(x) can combine the idea of minimizing the feature distance as done in [16], but the corresponding profit is marginal. Therefore, we recommend the efficient and safe method SIEp(Wie(3 x 3))-A(x) and the safest method SIEp(SSRQC)-A(x) for real-world steganography.

V. CONCLUSION

In this paper, we proposed a general framework of estimating side-information for JPEG steganography. To solve the two critical problems under the SIE framework, we employed several denoising and deblocking filters for better estimation of the precover, and we introduced two implementation models for modulating the costs. The experimental results validated that the proposed method SIEp(SSRQC)-A(x) improved additive distortion functions by a large margin for different image sets of several quality factors, and it outperformed the state-of-the-art estimated side-information-based methods. Obviously, how to precisely estimate the side-information is critical for improving the security of the SIE-based method and thus needs further investigation. Furthermore, how to incorporate the estimated side-information as the knowledge of the selection channel [24] for steganalysis is another important and interesting issue.

REFERENCES


Weixiang Li received the B.S. degree from Xidian University, Xi’an, China, in 2016. He is currently pursuing the Ph.D. degree with the University of Science and Technology of China. His research interests include steganography and steganalysis. He was a recipient of the Best Student Paper Award of 6th ACM IH&MMSec in 2018.

Kejiang Chen received the B.S. degree in School of Communication and Information Engineering, Shanghai University, in 2015. He is currently pursuing the Ph.D. degree in Information Security in University of Science and Technology of China (USTC). His research interests include information hiding, image processing and deep learning.
Weiming Zhang received his M.S. degree and PH.D. degree in 2002 and 2005 respectively from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China. Currently, he is a professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include multimedia security, information hiding, and privacy protection.

Hang Zhou received the B.S. degree in School of Communication and Information Engineering, Shanghai University, in 2015. He is currently pursuing the Ph.D. degree in Information Security in University of Science and Technology of China (USTC). His research interests include information hiding, image processing and computer graphics.

Yaofei Wang received the B.S. degree from the School of Physical Science and Technology, Southwest Jiaotong University, in 2017. He is currently pursuing the Ph.D. degree in information security with the University of Science and Technology of China. His research interests include information hiding, image processing, and deep learning.

Nenghai Yu received his B.S. degree in 1987 from Nanjing University of Posts and Telecommunications, M.E. degree in 1992 from Tsinghua University and Ph.D. degree in 2004 from University of Science and Technology of China, where he is currently a professor. His research interests include multimedia security, multimedia information retrieval, video processing, information hiding and security, privacy and reliability in cloud computing.