Distortion Function for Emoji Image Steganography
Lina Shi, Zichi Wang, Zhenxing Qian, Nannan Huang, Pauline Puteaux, Xinpeng Zhang

To cite this version:

HAL Id: hal-03160124
https://hal.archives-ouvertes.fr/hal-03160124
Submitted on 5 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Distortion Function for Emoji Image Steganography

Lina Shi¹, Zichi Wang¹, Zhenxing Qian¹, *, Nannan Huang¹, Pauline Puteaux² and Xinpeng Zhang¹

Abstract: Nowadays, emoji image is widely used in social networks. To achieve covert communication in emoji images, this paper proposes a distortion function for emoji images steganography. The profile of image content, the intra- and inter-frame correlation are taken into account in the proposed distortion function to fit the unique properties of emoji image. The three parts are combined together to measure the risks of detection due to the modification on the cover data. With the popular syndrome trellis coding (STC), the distortion of stego emoji image is minimized using the proposed distortion function. As a result, less detectable artifacts could be found in the stego images. Experimental results show that the proposed distortion function performs much higher undetectability than current state-of-the-art distortion function HILL which is designed for natural image.

Keywords: Steganography, emoji image, distortion function.

1 Introduction

Data hiding embeds additional data into digital media without causing seriously distortion to guarantee the usability of cover object [Dong, Zhang and Liu (2018); Qian, Xu, Luo et al. (2018); Qian and Zhang (2016)]. The purposes of data hiding are covert communication and copyright protection usually [Qian, Zhang and Wang (2014); Qian, Zhou, Zhang et al. (2016); Wang, Qian, Zhang et al. (2018)]. To achieve covert communication, steganography aims to embed secret data into digital media without drawing suspicion by slightly modifying cover data [Li and Zhang (2019); Wang, Zhang and Yin (2018); Wang, Yin and Zhang (2019)]. Early works by Fridrich et al. [Fridrich and Soukal (2006); Zhang and Wang (2006); Zhang, Zhang and Wang (2008)] increasing the undetectability of steganography by decreasing the quantity of modifications on cover data. However, the undetectability is not satisfactory since the security performance of steganography is also related to the modified locations. The currently popular approach by Fridrich et al. [Fridrich and Filler (2007)] is to minimize the additive distortion between the cover and the stego objec, which is achieved by syndrome trellis coding (STC) [Filler, Judas and Fridrich (2011)]. In this framework, a user-defined distortion

¹ Shanghai Institute for Advanced Communication and Data Science, School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China.
² Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier, Centre National de la Recherche Scientifique, University of Montpellier, Montpellier, 34095, France.
* Corresponding Author: Zhenxing Qian. Email: zxqian@shu.edu.cn.
function is used to assign embedding costs for all cover elements to quantify the effects of modification. There are many distortion functions designed for spatial images, such as Holub et al. [Holub and Fridrich (2012); Holub, Fridrich and Denemark (2014); Li, Wang, Huang et al. (2014); Sedighi, Cogranne and Fridrich (2016)] and JPEG images, such as Guo et al. [Guo, Ni and Shi (2014); Guo, Ni, Su et al. (2015); Wang, Zhang and Yin (2016); Wei, Yin, Wang et al. (2018); Du, Yin and Zhang (2018)].

For other kinds of images, new distortion functions should be proposed to fit their unique properties. In the age of big data currently, many kinds of digital images [Guan, Zhang, Wu et al. (2019); Wu, Dong, Ota et al. (2018)] are emerged. Specially, emoji image is widely used in social networks, e.g., Twitter, Facebook, and instant messaging systems, e.g., Skype, WeChat, to express emotion vividly. Different with natural image, as shown in Fig. 1, the emoji image is constituted by several curves with legible profile. The correlation between pixels is different from natural image which can be modeled as Markov chain. To save storage space, the usual format of emoji image is palette (a typical example: graphics interchange format). For the vitality of expression, most of the emoji images are motional. That means there are more than one frame contained in each emoji image. In this case, the correlation between the frames should also be considered for steganography. Furthermore, this inter-frame correlation is different from the correlation in natural images which are motionless.

![Figure 1: Several emoji images](image)

Existing distortion functions [Holub and Fridrich (2012); Holub, Fridrich and Denemark (2014); Li, Wang, Huang et al. (2014); Sedighi, Cogranne and Fridrich (2016); Guo, Ni and Shi (2014); Guo, Ni, Su et al. (2015); Wang, Zhang and Yin (2016); Wei, Yin, Wang et al. (2018)] are designed for natural image, which aim to restrain embedding changes into texture and complex regions to conceal the modification trace [Wang, Yin and Zhang (2018)]. Although these distortion functions perform well in natural image, they are not suitable for emoji image since the profile has not been used enough. Therefore, it is necessary to develop customized distortion function for emoji image. To the best of our knowledge, there is no distortion function designed for emoji image.

To fill up this gap, we propose a distortion function for steganography in emoji image. Different with existing distortion functions designed for natural image, the proposed distortion function combines the profile of image content, the intra- and inter-frame correlation together to measure the risks of detection due to the modification on the cover data. In this way, the unique properties of emoji image are considered. When secret data is embedded with syndrome trellis coding, the obtained stego emoji exposes less detectable artifacts.
2 Structure

The structure of the proposed method is shown in Fig. 2. To fit the properties of emoji image, the profile of image content, the intra- and inter-frame correlation are employed to form the profile, texture, and variation cost respectively. Then the three parts are combined together to measure the risks of detection due to the modification on the cover data.

**Figure 2:** Structure of the proposed distortion function

2.1 Emoji image

The usual format of emoji image is palette. As shown in Fig. 3, each palette image is composed of a color palette and a color index matrix. The color palette is a list of entries of representative colors in the image, and the elements in the color index matrix are pointers to those palette entries that specify the red-green-blue (RGB) colors [Tzeng, Yang and Tsai (2004)].

![Color palette and color index matrix](image)

**Figure 3:** Demonstration of the palette format

Since there is more than one frame (color index matrix) contained in each emoji image, each emoji image is composed of a color palette and several color index matrices corresponding to the frames. In other words, an emoji image is composed of several palette images with only one-color palette. The color palette is shared with all color index matrices.

2.2 Distortion function design

According to the properties of the emoji images, a new distortion function is designed. For an emoji image with \( k \) colors and \( l \) color index matrices, denote the \( i \)-th index in color palette as \( c_i, \ i \in \{0, \ldots, k\} \), the \( j \)-th color index matrix with size \( M \times N \) as \( X_j = \{x_{(u,v)}\} \in \mathbb{R}^{M \times N} \).
\{c_i\}^{M \times N}, j \in \{0, \ldots , l\}. The proposed distortion function assigns a embedding cost for each \( x_j(u,v) \). The details are as follows.

Denote the RGB color values corresponding to \( c_i \) as \( R_i, G_i, \) and \( B_i \) respectively. To minimize the color value distortion cause by the modifications made on \( x_j(u,v) \) during steganography, the value-similar \( x_j(u,v) \) should corresponding to similar \( (R_i, G_i, B_i) \) values.

To achieve this, all the \( X_j \) are modified using Algorithm 1.

**Algorithm 1 Color index matrix adjustment**

**Input:** The \( j \)-th color index matrix \( X_j \), color palette \( c_i \).

**Output:** Adjusted \( X_j \).

1. Find all the minimal \( x_j(a_0,b_0) \) in \( X_j \), \( a_0 \in \{1, \ldots , M\}, b_0 \in \{1, \ldots , N\} \), then set all \( x_j(a_0,b_0) \) as 0;
2. Calculate \( \theta_i = R_i^2 + G_i^2 + B_i^2 \) for all the \( M \times N \) \( x_j(u,v) \).
3. Set \( w=1 \);
4. Find all \( x_j(a_w,b_w) \) with the most similar \( \theta_i \) with \( x_j(a_{w-1},b_{w-1}) \), then set all \( x_j(a_w,b_w) \) as \( w \);
5. Repeat step (4) for all the cases of \( w > 1 \) until all the \( x_j(u,v) \) in \( X_j \) are adjusted.

The first part of the proposed distortion function aims to take use of the intra-frame correlation. This purpose is similar with the distortion functions for natural images. Current state-of-the-art distortion function is HILL, which constituted by a high-pass filter and two low-pass filters to make the modifications clustered. We employ the approach of HILL to assign the texture cost \( \rho_j^H(u,v) \) for each \( x_j(u,v) \). The details are as follows.

Let \( F_h \) be a high-pass filter, the residuals \( R_j \) of \( X_j \) are calculated using Eq. (1) firstly.

\[
R_j = X_j \otimes F_h
\]

where
\[
F_h = \begin{bmatrix}
-1 & 2 & -1 \\
2 & -4 & 2 \\
-1 & 2 & -1
\end{bmatrix}
\]

The nonexistent pixel which is out of the image boundary would be obtained by pixel symmetric padding. For example, \( x_j(u+1,v) \) is obtained by copying \( x_j(u-1,v) \) when it is out of the block boundary, and vice versa. Then two low-pass filters \( F_{l1} \) and \( F_{l2} \) are employed to obtained the cost matrix \( \rho_j^{H}\{\rho_j^{H}(u,v)\}^{M \times N} \) in HILL [Li, Wang, Huang et al. (2014)] using Eq. (3).

\[
\rho_j^{H} = \frac{1}{|R_j| \otimes F_{l1} \otimes F_{l2}} \quad (3)
\]

where \( |\cdot| \) represents the absolute value operation, and \( F_{l1} \) and \( F_{l2} \) are average filters sized 3×3 and 15×15 respectively.
Denote the color value distortion caused by modifying \( x_j(u,v) \) into \( x_j(u,v) + 1 \) as \( \delta_j(u,v) \), as shown in Eq. (4). Where \( R_j(u,v) \), \( G_j(u,v) \), \( B_j(u,v) \), and \( R_j^+(u,v) \), \( G_j^+(u,v) \), \( B_j^+(u,v) \) are the corresponding RGB color values of \( x_j(u,v) \) and \( x_j(u,v) + 1 \) respectively.

\[
\delta_j(u,v) = [R_j(u,v) - R_j^+(u,v)]^2 + [G_j(u,v) - G_j^+(u,v)]^2 + [B_j(u,v) - B_j^+(u,v)]^2
\]

Thus, the texture cost \( \rho_j^T(u,v) \) for each \( x_j(u,v) \) is defined in Eq. (5). Where the cost \( \rho_j^H(u,v) \) is calculate using Eq. (3), which is the same with the cost in HILL.

\[
\rho_j^T(u,v) = \delta_j(u,v) \cdot \rho_j^H(u,v)
\]

To extract the content profile of frame \( X_j \), it is transformed into grayscale image firstly by replacing \( x_j(u,v) \) with the corresponding real luminance value. Then the obtained grayscale image is further transformed into binary image \( Y_j = \{y_j(u,v)\}^{M \times N} \). Thus, the value of the elements in \( Y_j \) are 0 or 1. The content profile of \( X_j \) is the locations corresponding to “0” in \( Y_j \).

The locations corresponding to “1” in \( Y_j \) belong to the background of \( X_j \), which are so smooth that any modifications would be discovered. Therefore, these locations are not suitable for steganography. Accordingly, the profile cost \( \rho_j^P(u,v) \) can be obtained using Eq. (6).

\[
\rho_j^P(u,v) = \begin{cases} 1 & y_j(u,v) = 0 \\ \infty & y_j(u,v) = 1 \end{cases}
\]

In order to utilize the inter-frame correlation among different \( X_j \), we consider the color difference between \( X_j \) and \( X_{j-1} \). Both \( X_j \) and \( X_{j-1} \) are transformed into grayscale image to calculate the color difference. To avoid the subscript of \( X_j \) overflowing, the embedding tasks are not done on the first frame \( X_1 \). That means the first frame is kept unchanged during data embedding. Denote the \((u,v)\)th pixels of the grayscale \( X_j \) and \( X_{j-1} \) as \( p_j(u,v) \) and \( p_{j-1}(u,v) \) respectively, the color difference \( D_j = \{d_j(u,v)\} \) between \( X_j \) and \( X_{j-1} \) is,

\[
d_j(u,v) = |p_j(u,v) - p_{j-1}(u,v)|
\]

Then the variation cost \( \rho_j^V = \{\rho_j^V(u,v)\}^{M \times N} \) for each \( x_j(u,v) \) is defined in Eq. (8) to decrease the modifications on the frame which is similar with the previous frame.

\[
\rho_j^V(u,v) = \begin{cases} 1.3 & d_j(u,v) < 15 \\ 1 & d_j(u,v) \geq 15 \end{cases}
\]

where the values “1.3” and “15” are empirically determined by experiments.

Finally, the three parts (texture cost \( \rho_j^T(u,v) \), profile cost \( \rho_j^P(u,v) \), variation cost \( \rho_j^V(u,v) \) ) are combined together by multiplication. The final embedding cost \( \rho_j(u,v) \) assigned for \( x_j(u,v) \) is defined in Eq. (9).

\[
\rho_j(u,v) = \rho_j^T(u,v) \cdot \rho_j^P(u,v) \cdot \rho_j^V(u,v)
\]
analyze the quality of stego image. Finally, we provide the results of undetectability compared with HILL.

3.1 Experimental setup

To build the cover image set, we collected 560 emoji images that is widely used in social networks. These images are in palette format and each image contains 256 colors and several frames. There are 2557 frames in total of the 560 images. We have uploaded all the 560 images on https://pan.baidu.com/s/1nOsn_eoI8vLpqqo8ue8nOQ.

We compare the proposed distortion function with the popular distortion function HILL which performs the state-of-the-art undetectability. Since HILL is designed for spatial image, each frame is transformed into grayscale image firstly when embedding with HILL. In other words, for HILL embedding, there are 2557 grayscale images are used as covers. The capacity of secret data embedded in each frame is set as 600 bits, 700 bits, 800 bits, 900 bits, 1000 bits, and 1100 bits respectively. All embedding tasks are done by the embedding simulator [Pevný, Filler and Bas (2010)] since it is widely used to simulate the optimal embedding. For steganalysis, the feature sets SPAM proposed by Pevný et al. [Pevný, Bas and Fridrich (2010)] and SRMQ1 proposed by Fridrich et al. [Fridrich and Kodovsky (2012)] are employed in our experiments. The ensemble classifier proposed by Kodovsky e al. [Kodovsky, Fridrich and Holub (2012)] is used to measure the property of feature sets. In detail, half of the cover and stego feature sets are used as the training set while the remaining half are used as testing set. The criterion to evaluate the performance of feature sets is the minimal total error $P_E$ under equal priors achieved on the testing set in Kodovsky et al. [Kodovsky, Fridrich and Holub (2012)]:

$$P_E = \min_{\theta} \left( \frac{P_{FA} + P_{MD}}{2} \right)$$

where $P_{FA}$ is the false alarm rate and $P_{MD}$ is the missed detection rate. The performance is evaluated using the average value of $P_E$ over ten random tests.

3.2 Image quality demonstrations

The demonstrations of the proposed method are shown in Fig. 4. Where Fig. 4(a) is a cover emoji image composed of several frames. After each frame is embedded with 600, 800, and 1000 bits respectively, the obtained stego images are shown in Figs. 4(b), 4(c) and 4(d) correspondingly.

It is clear that the stego images are close to the cover image, which means the visual quality of the stego images is satisfactory regardless of the capacity. Thus, the usability of emoji images is reserved after embedded adequate secret data using the proposed method.
Figure 4: Demonstrations of (a) cover and corresponding stego emoji images using the proposed method with capacity (b) 600 bits, (c) 800 bits, (d) 1000 bits

3.3 Undetectability comparison

Fig. 5 shows the undetectability comparisons of HILL and the proposed method against SPAM and SRMQ1 tested on all the 2557 frames, and Tab. 1 depicts all numerical values.

![Graph showing undetectability comparisons](image_url)

Figure 5: Comparisons of the proposed method with HILL against (a) SPAM, (b) SRMQ1

<table>
<thead>
<tr>
<th>Steganography Algorithm</th>
<th>Feature</th>
<th>Capacity (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>600</td>
<td>700</td>
</tr>
<tr>
<td>Proposed</td>
<td>SPAM</td>
<td>0.2696</td>
</tr>
<tr>
<td></td>
<td>SRMQ1</td>
<td>0.1533</td>
</tr>
<tr>
<td>HILL</td>
<td>SPAM</td>
<td>0.1199</td>
</tr>
<tr>
<td></td>
<td>SRMQ1</td>
<td>0.0624</td>
</tr>
</tbody>
</table>

It is clear that the security performance of the proposed method is much better than HILL for all cases, regardless of the steganalytic tools and capacity. Specifically, the $P_E$ values...
of the proposed method are more than two times of HILL. For the cases of large capacity, e.g., 900, 1000, 1100 bits, the $P_E$ values of the proposed method are nearly three times of HILL. The large improvement on undetectability is because that the proposed distortion function is designed by following the unique properties of emoji image, while HILL not. In addition, inter-frame correlation is the most unique property of motion image, which is mentioned in Usui et al. [Usui, Takano and Yamamoto (2017)]. The modifications of steganography should avoid destroying the correlation as far as possible. The correlation can be reflected in the difference between image frames. For this reason, we also give the undetectability comparisons on difference image in Fig. 6 to further demonstrate the superiority of the proposed method. Tab. 2 lists the corresponding numerical values.

![Figure 6: Comparisons on difference image against (a) SPAM, (b) SRMQ1](image)

**Table 2: Testing errors tested on difference image against SPAM and SRMQ1**

<table>
<thead>
<tr>
<th>Steganography Algorithm</th>
<th>Feature</th>
<th>Capacity (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>Proposed</td>
<td>SPAM</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>SRMQ1</td>
<td></td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>HILL</td>
<td>SPAM</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>SRMQ1</td>
<td></td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Since the first frame is kept unchanged during data embedding, as mentioned in Subsection 2.2, the difference images are obtained by calculating the differences between the first frame and the other frames respectively in each emoji cover and stego image. That means the first frame is kept unchanged during embedding. As shown in Fig. 6, the undetectability
tested on the difference images of the proposed method is still better than HILL.

4 Conclusion

A distortion function for emoji image steganography is proposed in this paper. To fit the properties of emoji image, the profile of image content, the intra- and inter-frame correlation are considered in the proposed distortion function. The three parts are combined together by multiplication to resist steganalysis. Experimental results proved the effectiveness of the proposed distortion function. For further study, it is significant to develop the theoretical optimal embedding for emoji image by uniting the steganographic methods for palette image.

Acknowledgement: This work was supported by the Natural Science Foundation of China (U1736213, 61572308), the Natural Science Foundation of Shanghai (18ZR1427500), the Shanghai Dawn Scholar Plan (14SG36), and the Shanghai Excellent Academic Leader Plan (16XD1401200).

References


351-358.


