Hindawi
Security and Communication Networks
Volume 2017, Article ID 1790268, 13 pages
https://doi.org/10.1155/2017/1790268

Research Article

Vague Sets Security Measure for Steganographic System Based on High-Order Markov Model

Chun-Juan Ouyang,[1,2] Ming Leng,[1,2] Jie-Wu Xia,[1,2] and Huan Liu[1,2]

[1]*Key Laboratory of Watershed Ecology and Geographical Environment Monitoring of NASG, Jinggangshan University, Ji'an 343009, China*
[2]*School of Electronics and Information Engineering, Jinggangshan University, Ji'an 343009, China*

Correspondence should be addressed to Chun-Juan Ouyang; oycj001@163.com

Received 26 April 2017; Accepted 12 June 2017; Published 6 August 2017

Academic Editor: Yushu Zhang

Security measure is of great importance in both steganography and steganalysis. Considering that statistical feature perturbations caused by steganography in an image are always nondeterministic and that an image is considered nonstationary, in this paper, the steganography is regarded as a fuzzy process. Here a steganographic security measure is proposed. This security measure evaluates the similarity between two vague sets of cover images and stego images in terms of *n-order* Markov chain to capture the interpixel correlation. The new security measure has proven to have the properties of boundedness, commutativity, and unity. Furthermore, the security measures of zero order, first order, second order, third order, and so forth are obtained by adjusting the order value of *n-order* Markov chain. Experimental results indicate that the larger n is, the better the measuring ability of the proposed security measure will be. The proposed security measure is more sensitive than other security measures defined under a deterministic distribution model, when the embedding is low. It is expected to provide a helpful guidance for designing secure steganographic algorithms or reliable steganalytic methods.

1. Introduction

Security of the steganographic system is the fundamental issue in the field of the information hiding. Image steganography is the technique of hiding information in digital image and trying to conceal the existence of the secret information. The image with and the image without hidden information are called stego image and cover image, respectively [1]. Steganography and steganalysis are in a hide-and-seek game [2]. They try to defeat each other and also develop with each other. In recent years, steganalysis researches have made much head-way [3, 4], and many attempts have been made to build up secure steganographic algorithms [5–8]. Up until now, there is no standard security measure for steganographic system. The security of the steganography always depends on the encryption of the steganography, which contradicts Kerckhoffs' principle [9]. Hence, it is very necessary to study the security measure which can provide guidance for designing the high-secure steganography and steganalytic algorithms with high performance.

Now, the study of the security measure becomes one of the hotspots in the steganography research field. Researchers have put forward their views from different viewing angels. From the point of view of information theory, Cachin [10] proposed a security measure in terms of the relative entropy between the probability mass functions (PMF) of the cover images and the stego images. Sullivan et al. [11] employed the divergence distance of the empirical matrices to define the security measure. They modeled the sequence of image pixels as first-order Markov chain which could capture one adjacent pixel dependency. Furthermore, Zhang et al. [12] models the images pixels as n-order Markov chain to provide the security measure. Based on game theory, Liu et al. [13] presented that the counterwork relationship is modeled between steganography side and attack side. In [14], Schöttle and Böhme studied adaptive steganography while taking the

knowledge of the steganalyst into account. Liu and Tang [15] also provided the security for the adaptive steganography. In [16], Chandramouli et al. proposed an alternative security measure based on steganalyzer's ROC (Receiver Operating Characteristic) performance. From the point of feature space, Pevný and Fridrich [17] provided the MMD (Maximum Mean Discrepancy) by employing a high-dimensional feature space set as the covers models.

The security measures mentioned above all assume that accurate statistical estimations can be obtained from the finite data samples. However, an image is a nonstationary process; its local statistical correlation will change when image is changed slightly. So the statistical features change is nondeterministic after steganography processing. Meanwhile, for a steganographic system, the warden is lack of the knowledge of the cover distribution. Thus, the distribution estimates of the cover and stego image are not stable. So the security measures defined under the deterministic statistical model are hard to apply due to the lack of the accurate distribution.

To address this problem, we regard the steganography as a fuzzy and indeterministic process. The goal of this paper is to provide a practical security measure in terms of the vague sets similarity measure between cover images and the stego images. Particularly, the sequence of image pixels is modeled as an *n-order* Markov chain to capture the interpixel correlations. The main contributions of this work are as follows:

(1) We derive a security measure for a steganographic system which is different from the deterministic ones. The existing security measures are defined by evaluating the difference between cover images and stego images. In contrast, the new security measure is defined by evaluating the similarity between cover images and their stego version.

(2) The *n-order* security measure based on vague sets similarity measure is proven to have the properties of the boundedness, commutativity, and unity. The properties guarantee the security measure is indeed a real distance which indeed satisfies the symmetry and triangle inequality. The boundedness guarantees the new benchmark can measure the steganographic security.

(3) Simulation results verify the effectiveness of the new security measure by benchmarking several popular steganographic schemes. When embedding rate is low, the new security measure is more sensitive to reveal the statistical features change than other security measures. Thus, the proposed security measure can provide a better guidance for the design of steganography and steganalysis.

The rest of the paper is organized as follows. Section 2 gives a review of the two security measures with the deterministic statistical distribution model and introduces the *n-order* Markov chain model. The *n-order* secure measure based on vague sets similarity measure is presented in detail in Section 3. Experimental results are provided in Section 4 to demonstrate the effectiveness and the superiority of the proposed security measure. We draw our conclusions in Section 5.

2. Steganographic Security and Cover Model

2.1. Security Measure Based on Kullback-Leibler (K-L) Divergence. Suppose C is the set of all the covers, and it is an assumption that the selections of the covers and stegos from the set C can be described by the random variables c and s on C with the probability mass functions (PMF) P_c and P_s, respectively. Cachin [10] quantified the security of a steganographic system in terms of the Kullback-Leibler (K-L) divergence (sometimes called relative entropy); that is,

$$D\left(P_c \parallel P_s\right) = \sum_{x \in X} P_c\left(x\right) \log \frac{P_c\left(x\right)}{P_s\left(x\right)}, \tag{1}$$

where X is the set of possible pixel values. A steganographic system is called perfectly secure if (1) is zero or ε-secure if $0 \le D(P_c \parallel P_s) \le \varepsilon$ is satisfied. The K-L divergence provides a simple yet convenient method for measuring the difference between cover images and stego images.

In fact, we have little information about the PMF involved due to the large dimensionality of the set C. So the security measure is usually defined with simplified cover models, such as independent and identically distributed (i.i.d.) ones. The security measure of K-L divergence calculates the difference from the view of the first-order statistical features (such as one-dimensional histogram feature).

2.2. Security Measure Based on Divergence Distance. To account for the dependence of the pixels, Sullivan et al. [11] employed the first-order Markov chain model to capture the interpixel correlation. The divergence distance was used to quantify the statistical feature perturbations introduced by a steganography between the two empirical matrices of cover images and stego images. Suppose C and S are two random sequences of the cover image pixels and the stego image pixels, respectively, obtained by a given scanning method. Let M^c and M^s be the empirical matrixes of C and S, respectively. The divergence distance is given by

$$D\left(M^c, M^s\right) = \sum_{i,j \in R} M_{ij}^c \log\left(\frac{M_{ij}^c}{\sum_j M_{ij}^c} \frac{\sum_j M_{ij}^s}{M_{ij}^s}\right), \tag{2}$$

where $M_{ij}^c / \sum_j M_{ij}^c$ and $M_{ij}^s / \sum_j M_{ij}^s$ are the transition probabilities of cover images and stego images, respectively. The transition probability is commonly calculated by the ratio of the total number to the pixel changes from value i to value j over the total number of possible pixel changes (e.g., for an 8-bit image, the total possible pixel changes number is 256×256). The constant R is the range of all possible pixel values. Thus, the divergence distance provides the difference between cover images and their stego version from the view of the second-order statistical features (such as two-dimensional histogram feature and difference histogram feature).

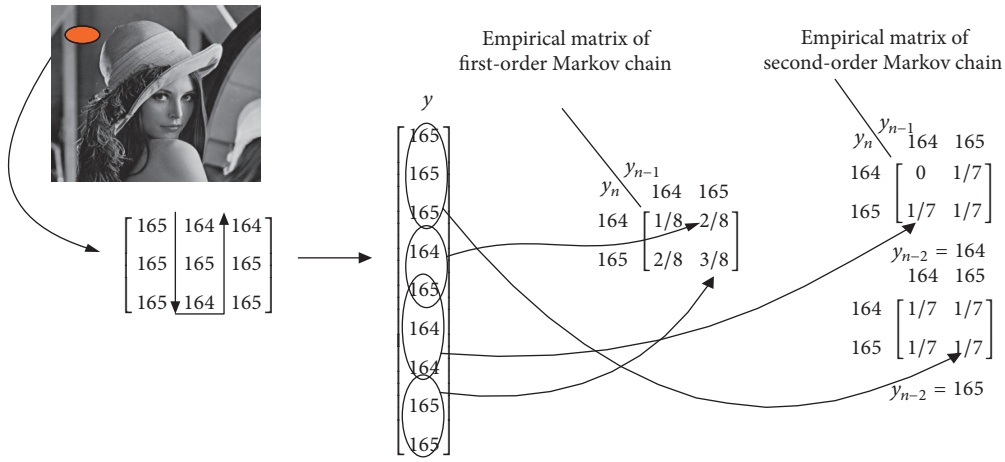


Figure 1: The generating process of the empirical matrixes of first-order and second-order Markov chain.

The two security measures mentioned above are defined based on the Shannon information theory under the assumption that the image data statistical distribution is deterministic. Most of the security measures proposed later are also defined under the same assumption. However, the image data shows the sceneries in the aspects of gray, texture, shape, and so forth. There are many a kind of indeterministic factors (such as noise) in a steganography process. Therefore, the security measures with the deterministic statistical distribution model cannot measure the security accurately.

2.3. n-Order Markov Chain Model. The weakness of the above two security measure lies in the fact that the image model such as i.i.d and first-order Markov are too simple to capture interpixel dependency. Therefore, here we model the sequence of image pixels as an n-order Markov chain. The *n-order* Markov chain is a random sequence indexing the image pixels scanned by a given mode. For instance, when $n = 2$, the second-order Markov chain accounting for two adjacent pixels' correlation meets the following condition:

$$P\left(Y_m \mid Y_{m-1}, Y_{m-2}, \ldots, Y_1\right) = P\left(Y_m \mid Y_{m-1}, Y_{m-2}\right). \quad (3)$$

There are at least two reasons for us to select *n-order* Markov chain model. First, the model is flexible. When $n = 0$, it turns out to be the i.i.d model, in which the image pixels are assumed to be unrelated. When $n = 1$, the first-order Markov chain can capture only one adjacent pixel dependence. Furthermore, the n-order Markov chain can capture more interpixel relationships among the pixels when $n \geq 2$. Second, compared with the Markov random field model [9], the Markov chain model, though simple, is able to calculate the statistical estimation of the image samples. For *n-order* Markov chain, it is easy to calculate the realistic statistical estimates using the empirical matrixes. In the following, we construct the empirical matrixes of the first-order and second-order Markov chain.

Let $\{Y_n,\ n = 1, 2, \ldots, L\}$ be an *n-order* Markov chain on the finite set ω, where Y_n is the n-indexed set of pixels obtained by a row, column, zigzag, or Hilbert scanning method. ω is the possible gray scale values. When $n = 1$, the first-order Markov chain source is defined by the transition matrixes $T_{i_1,i_2} = P(Y_n = i_1 \mid Y_{n-1} = i_2)$ and marginal probabilities $p_{i_1} = P(Y_n = i_1)$. For a realization, $y = (y_1, y_2, \ldots, y_L)^T$. Let η_{i_1,i_2} be the number of transitions from values i_1 to i_2 in y. The empirical matrixes are $M_1(y) = \eta_{i_1,i_2}(y)/(L-1)$. That is, the i_1, i_2 element represent the proportion of spatially adjacent pixel pairs with the grayscale value of i_1 followed by i_2. Thus the empirical matrixes provide an estimation of the transition matrixes and marginal probabilities. The empirical matrixes are similar to the concurrence matrixes of the image. It can be recognized as a matrix form of the two-dimensional normalized histogram for estimating the joint probability mass function (PMF) of a source image. Similarly, when $n = 2$, we can get the empirical matrixes of the second-order Markov chain, denoted by $M_2(y) = \eta_{i_1,i_2,i_3}(y)/(L-1)$. $\eta_{i_1,i_2,i_3}(y)$ is the number of transitions from values i_1 to i_3 via i_2 in y. For an 8-bit image, the size of the empirical matrixes $M_2(y)$ is 256×256×256. The element of the empirical matrixes represents the proportion of spatially adjacent pixel group with a grayscale value of i_1 followed by i_2 and i_3. A simple example of generating the empirical matrixes of first-order and second-order Markov chain is shown in Figure 1.

In Figure 1, the small block is derived from the standard image "Lena." Its size is 3 × 3, including pixels 164 and 165. The example image pixels are scanned vertically. The size of the empirical matrixes of first-order Markov chain in Figure 1 is 2 × 2. The element represents the proportion of spatially adjacent pixel pairs with (164, 164), (164, 165), (165, 164), and (165, 165). The right-hand side of Figure 1 demonstrates the procedure of the empirical matrixes of second-order Markov chain. Its size is 2 × 2 × 2, in which the element represents the proportion of spatially adjacent pixel groups with (164, 164, 164), (164, 165, 164), (165, 164, 164), (165, 165, 164), and so forth.

Since the cover sources are strongly correlated, the probabilities of two adjacency samples are equal or nearly equal. As a result, in the empirical matrixes, the masses are more

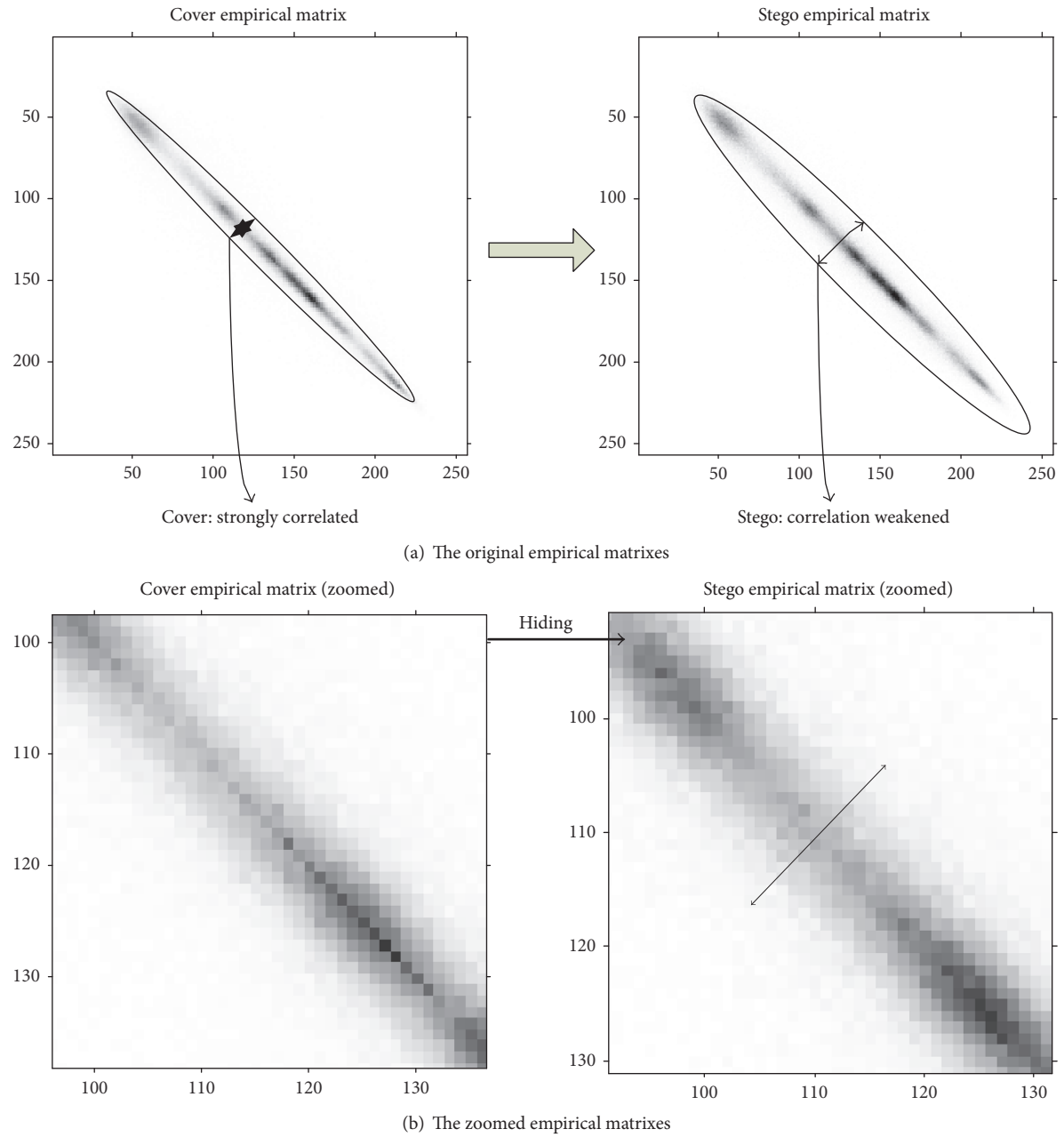


Figure 2: Empirical matrixes of a cover image and its stego image.

concentrated near the main diagonal in a correlated source. In [18], Harmsen and Pearlman considered that information hiding can be viewed as adding the additive noise to the cover image. The secret information (additive noise) is uncorrelated after hiding, and its empirical matrixes spread evenly over the main diagonal. Thus we see that hiding weakens the dependencies among the cover samples, which is illustrated in Figure 2(a). Figure 2(b) is part of the zoomed empirical matrixes. According to the above analysis, the steganography tends to spread the density of the pixels pairs away from the main diagonal of the empirical matrixes. This property may shed some light on designing of the security measure for a steganographic system. Thus, in Section 3, we will propose an *n-order* security measure in terms of the vague sets similarity measure by modeling the sequence of images pixels as an *n-order* Markov chain.

3. Security Measure Based on Vague Sets Similarity Measure

The vague sets similarity measure [19, 20] describes the matching degree of two vague sets. In a practical steganographic system, there are many indeterministic factors introduced by steganography. In this work, we regard the responding probability distribution sets of the cover samples and the stego samples as two discrete vague sets. Then a new security measure is proposed below in terms of vague sets similarity

measure to measure the similarity between cover images and stego images.

3.1. Vague Sets. Roughly speaking, a fuzzy set is a class with fuzzy boundaries. The fuzzy set A is a class of objects X along with a grade of membership function $\mu_A(x)$, $x \in X$. It assigns a single value to each object. This single value combines the evidence for x ∈ X and the evidence against $x \in X$. And it is only a measure of the pros/cons evidence. However, in many practical applications we often require pros and cons evidence simultaneously. Gau and Buehrer [21] advanced the concept of vague sets. The vague sets theory adopts a true membership function t_A and a false membership function f_A to record the lower bounds on μ_A. These lower bounds are used to create a subinterval on [0, 1], namely, $[t_A(x_i), 1-f_A(x_i)]$, to generalize $\mu_A(x_i)$ of fuzzy sets, where $t_A(x_i) \le \mu_A(x_i) \le 1 - f_A(x_i)$. Vague sets expand the value of the membership function to a subinterval of [0, 1] instead of a single value; thus it has stronger ability to reveal the indeterminacy than the fuzzy set theory. The related definitions of vague sets are as follows.

Definition 1 (vague sets). Let X be the universe of discourse, $X = \{x_1, x_2, \ldots, x_n\}$. $V(x)$ denotes all the vague sets of X, $\forall A \in V(x)$. The vague set A is characterized by a true membership function t_A and a false membership function f_A:

$$\begin{aligned} t_A &: X \longrightarrow [0,1], \\ f_A &: X \longrightarrow [0,1], \end{aligned} \tag{4}$$

where $t_A(x_i)$ is the lower bound on the grade of membership of x_i derived from the evidence for x_i. $f_A(x_i)$ is a lower bound on the negation of x_i derived from the evidence against x_i, satisfying $t_A(x_i) + f_A(x_i) \le 1$. The grade of membership of x_i is bounded to a subinterval $[t_A(x_i), 1-f_A(x_i)]$ of [0, 1]. When X is discrete, a vague set A can be written as

$$A = \sum_{i=1}^{n} \frac{\left[t_A(x_i), 1 - f_A(x_i)\right]}{x_i}, \quad x_i \in X. \tag{5}$$

Definition 2. Let X be the universe of discourse, $X = \{x_1, x_2, \ldots, x_n\}$. A and B are two vague sets of X. The entropy of the vague set A, $E(A)$, is defined as

$$E(A) = -\frac{1}{n \ln 2} \sum_{i=1}^{n} \left[t_A(x_i) \ln t_A(x_i) + f_A(x_i) \ln f_A(x_i)\right]. \tag{6}$$

Definition 3. Let X be the universe of discourse, $X = \{x_1, x_2, \ldots, x_n\}$. A and B are two vague sets of X. The partial entropy of vague set A against vague set B, $E_B(A)$, is defined as

$$E_B(A) = -\sum_{i=1}^{n} \left[t_B(x_i) \ln t_A(x_i) + f_B(x_i) \ln f_A(x_i)\right]. \tag{7}$$

3.2. The n-Order Security Measure Based on Vague Sets Similarity Measure. As discussed in Section 2.3, the n-order Markov chain model can capture sufficient inherent correlations. Additionally, the changes in image statistical features, introduced by steganography, are indeterministic. Therefore, in the new security measure, we model the sequence of the image pixels as an *n-order* Markov chain. Simultaneously, the empirical matrixes of the *n-order* Markov chain of cover images and stego images are regarded as two vague sets. Then the n-order security measure based on the vague sets similarity measure is defined as follows.

Suppose C and S are n-order Markov chain sequence of cover images and stego images, respectively, and then scan them by a given mode (such as horizontal, vertical, zigzag, and Hilbert mode). MC and MS represent the corresponding empirical matrixes. $m_{i_1,i_2,\ldots,i_{n+1}}$, the element of empirical matrixes, denotes the joint probability distribution from pixels i_1 to i_{n+1} via the states of $i_2, i_3, \ldots$ and i_n. The $i_1, i_2, \ldots, i_{n+1}$ is the image pixel value, $i \in [0, 255]$. G denotes the set of all possible values of $m_{i_1,i_2,\ldots,i_{n+1}}$. Let $M_{i_1,i_2,\ldots,i_{n+1}}$ be the universe of discourse composed of $m_{i_1,i_2,\ldots,i_{n+1}}$. Then MC and MS are two vague sets on $M_{i_1,i_2,\ldots,i_{n+1}}$. That is,

$$\begin{aligned} MC &= \frac{\sum_{i=0}^{255}\left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right), 1 - f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right]}{m_{i_1,i_2\cdots i_{n+1}}}, \\ &\qquad m_{i_1,i_2\cdots i_{n+1}} \in M_{i_1,i_2\cdots i_{n+1}}, \\ MS &= \frac{\sum_{i=0}^{255}\left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right), 1 - f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right]}{m_{i_1,i_2\cdots i_{n+1}}}, \\ &\qquad m_{i_1,i_2\cdots i_{n+1}} \in M_{i_1,i_2\cdots i_{n+1}}. \end{aligned} \tag{8}$$

Definition 4. Let $M_{i_1,i_2,\ldots,i_{n+1}}$ be the universe of discourse. MC and MS are two vague sets of $M_{i_1,i_2,\ldots,i_{n+1}}$. The similarity measure $T_n(MC, MS)$ between the vague sets MC and MS is defined as the *n-order* secure measure for a steganographic system; that is,

$$T_n(MC, MS) = \frac{m \ln 2 \left(E(MC) + E(MS)\right)}{E_{MC}(MS) + E_{MS}(MC)}, \tag{9}$$

where $E(MC)$ and $E(MS)$ denote the entropy of the vague set MC and MS, respectively; $E_{MC}(MS)$ stands for the partial entropy of vague set MS against vague set MC; $E_{MS}(MC)$ is the partial entropy of vague set MC against vague set MS. $E(MC)$ and $E_{MC}(MS)$ can be written as

$$\begin{aligned} E(MC) &= -\frac{1}{m \ln 2} \\ &\quad \cdot \sum_{i_1,i_2\cdots i_{n+1} \in G} \left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \ln t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right. \\ &\quad \left. + f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \ln f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right], \\ E_{MS}(MC) &= -\sum_{i_1,i_2\cdots i_{n+1} \in G} \left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \ln t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right. \\ &\quad \left. + f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \ln f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right]. \end{aligned} \tag{10}$$

Similarly, $E(MS)$ and $E_{MC}(MS)$ can be written as

$$\begin{aligned} E(MS) &= -\frac{1}{m \ln 2} \\ &\cdot \sum_{i_1, i_2 \cdots i_{n+1} \in G} \left[t_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \ln t_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \right. \\ &\left. + f_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \ln f_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \right], \\ E_{MC}(MS) &= -\sum_{i_1, i_2 \cdots i_{n+1} \in G} \left[t_{MC}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \ln t_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \right. \\ &\left. + f_{MC}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \ln f_{MS}\left(m_{i_1, i_2 \cdots i_{n+1}}\right) \right]. \end{aligned} \tag{11}$$

Moreover, a steganographic system is called perfectly secure if $T_n(MC, MS) = 1$ or ε-secure if $T_n(MC, MS) = \varepsilon$, $\varepsilon \in (0, 1)$. $T_n(MC, MS) = 0$.

Theorem 5. *Let $T_n(MC, MS)$ be the n-order secure measure of a steganographic system based on vague set similarity measure. Then $T_n(MC, MS)$ satisfies the following.*

(1) Boundedness is

$$0 \le T_n(MC, MS) \le 1. \tag{12}$$

(2) Commutativity is

$$T_n(MC, MS) = T_n(MS, MC). \tag{13}$$

(3) Unity is

$$\begin{aligned} T_n(MC, MS) = 1 &\Longleftrightarrow \\ MC &= MS. \end{aligned} \tag{14}$$

$T_n(MC, MS)$ provides a security measure for a steganographic system in terms of the similarity between cover images and stego images. $T_n(MC, MS)$ is limited in a finite interval of $[0, 1]$, where 1 denotes "perfectly secure," while 0 denotes "definitely unsecure." However, other security measures under the deterministic statistical model calculate the difference between cover images and stego images. The values range in an infinite interval $[0, \infty)$. The property of the boundedness guarantees the proposed security measure can measure a steganographic algorithm quantitatively. Hence, it has stronger ability to reveal the statistical changes of the cover images. When $n = 0$, the image pixels distribution is said to be i.i.d., and $T_0(MC, MS)$ is called the zero-order security measure. When $n = 1$, the sequence of image pixels is considered to be a first-order Markov chain, and $T_1(MC, MS)$ is defined as the first-order security measure. Thus, a different order security measure can be obtained by adjusting the value of n.

4. Experimental Results and Discussion

In this section, we report experimental results that demonstrate the capability of the new security measure. First of all in Section 4.1 the image databases used for the experiment are described. Afterwards, in Section 4.2, we benchmark several different steganographic methods with *n-order* security measure based on vague sets, with particular attention to the effectiveness of low embedding rate. Finally, we compare the proposed security measure with previously used benchmarks designed under the deterministic statistical model.

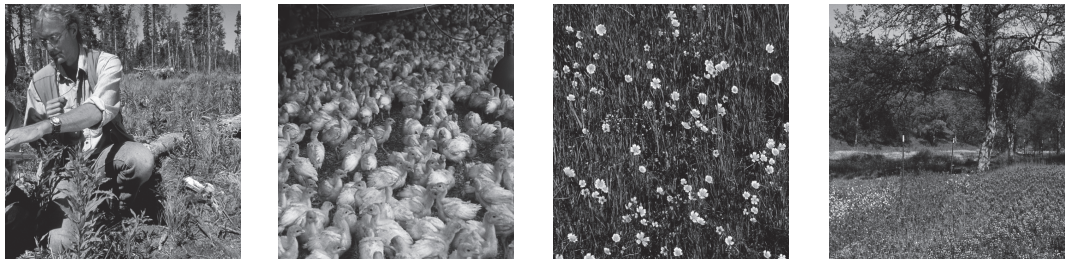
4.1. Image Database. For the experimental validation we used two image databases. The first one is BOWS2 [22] image database including 10000 grayscale images with fixed size 512×512. The other one is NRCS Photo Gallery [23]. We selected 1500 images from NRCS Photo Gallery. All images were converted into grayscale and central cropped to a size of 512×512 for experimental purposes. The images in our experiments show a wide range of scenarios including house, manmade objects, and animal. Some images are shown in Figure 3.

4.2. Verification of the Effectiveness of the Proposed Security Measure. To evaluate the performance of the proposed method for measuring the security of the steganographic algorithms, the new security measure with different orders is used to measure the security of different steganographic algorithms with different embedding rates. First, we select some spatial-domain steganographic algorithms, including LSBM (least significant bit matching) [24], LSB ± 2, HUGO [25] (highly undetectable steganography). We use 2000 images from BOWS2 image database; all the images are grayscale with the fixed size 512×512. As discussed in Section 2.3, first-order and second-order Markov chain models have captured sufficient interpixel correlations. Additionally, considering the computation complexity, we use the zero-order, first-order, and second-order security measure based on vague sets to measure the LSBM, LSBM2, and HUGO steganographic methods with the embedding rate ranging from 0.1 bpp (bits per pixel) to 1 bpp in a step size of 0.1 bpp. The average measure results for zero-order, first-order, and second-order security measure of 2000 images with different embedding rates are depicted in Figure 4.

In Figure 4, all curves indicate that the value of security measure gradually decreases with an increase in the embedding rate for the same steganographic algorithm. It is consistent with the definition of the security measure based on vague sets. Its value is limited in an interval of $[0, 1]$, where 1 denotes "perfectly secure" for the steganographic system. Hence the value of the *n*-order security measure satisfies monotonic decreasing property; that is, the higher the security of the stego images, the larger the value of the security measure. Furthermore, as is evident in Figure 4, the values of the same order security measure are different for different stego schemes with the same embedding rate. Note that LSB ± 2 obtains the lowest value in Figure 4, implying that it is most unsecure among the three hiding methods under the same condition. On the contrary, HUGO gains the highest value. All the measure results are coincident with the theoretical analysis of the three embedding schemes.

Furthermore, in order to evaluate the measuring ability of different order security measures, we compare the security

(a) Some images of BOWS2



(b) Some images of NRCS

Figure 3: Some images of image database.

for the same steganographic algorithm using different order security measures. Figure 6 shows the average measure results of zero-order, first-order, and second-order security measure for LSBM, LSB ± 2, and HUGO, respectively. In fact, all the data in Figure 5 is derived from Figure 4. As demonstrated in Figure 5, for the same steganographic method, the values of the zero-order, first-order, and second-order security measure are different at the same embedding rate. It is demonstrated that the value of the first-order security measure is smaller than that of the zero-order measure but larger than that of the second-order measure for the same steganographic method with the same embedding rate. The experiments show that the second-order security measure provides the largest measure interval to reveal the security change of the cover images with the embedding rate ranging from 0.1 bpp to 1 bpp. So we can conclude that second-order security measure can provide more obvious statistical distributed changes caused by steganography.

To further verify the effectiveness of the proposed security measure. We used it to benchmark JPEG steganographic algorithms schemes on different database. And we focus on low payloads to see if any of the test steganographic schemes becomes distinguishable by using the vague sets security measure with finite image sample.

We selected 1500 images from NRCS Photo Gallery. All images were converted into grayscale and central cropped to a size of 512 × 512 for experimental purposes. The images were embedded with pseudorandom payloads with 5%, 10%, 15%, and 20% bpac (bits per nonzero AC coefficient). The tested stego schemes include F3, F5 without shrinkage (nsF5) [26], Model Based Steganography without deblocking (MB1) [27], and Model Based Steganography with deblocking (MB2) [28]. The cover images were single-compressed JPEGs with quality factor 70. The measure results using zero-order, first-order, and second-order security measure based on vague sets are showed in Table 1. The data in Table 1 indicates that, for the same steganography, the larger the embedding rate, the lower the value of the same security measure. It also exhibited that, for the same steganography, the higher the order of the security measure, the smaller the value of the security measure, suggesting that second-order security measure can get a value lower than the other two security measures under the same condition.

The data in Table 1 also shows, according to the same order security measure, the MB2 is the least statistically detectable, followed by MB1 and nsF5, while F3 is the most detectable. All the measure results are coincident with the theoretical security among adopted stego algorithms. In a word, the experimental results indicate that the proposed security measure is effective for measuring the security for different steganographic methods on different image database. Meanwhile, the greater the order, the stronger the measure ability of the security measure.

4.3. Comparison with Security Measure under Deterministic Statistical Model. To show the superiority of the proposed security measure $T_n(MC, MS)$, we compare it with two security measures under the deterministic statistical model. One is the Kullback-Leibler (K-L) divergence between the probability mass functions (PMF) proposed by Anderson [9], denoted by $D(P_c \parallel P_s)$. The other, denoted as $D(M_c, M_s)$, is the divergence distance between the two empirical matrices proposed by Cachin [10]. To be unbiased, the zero-order measure $T_0(MC, MS)$ is compared with $D(P_c \parallel P_s)$ when

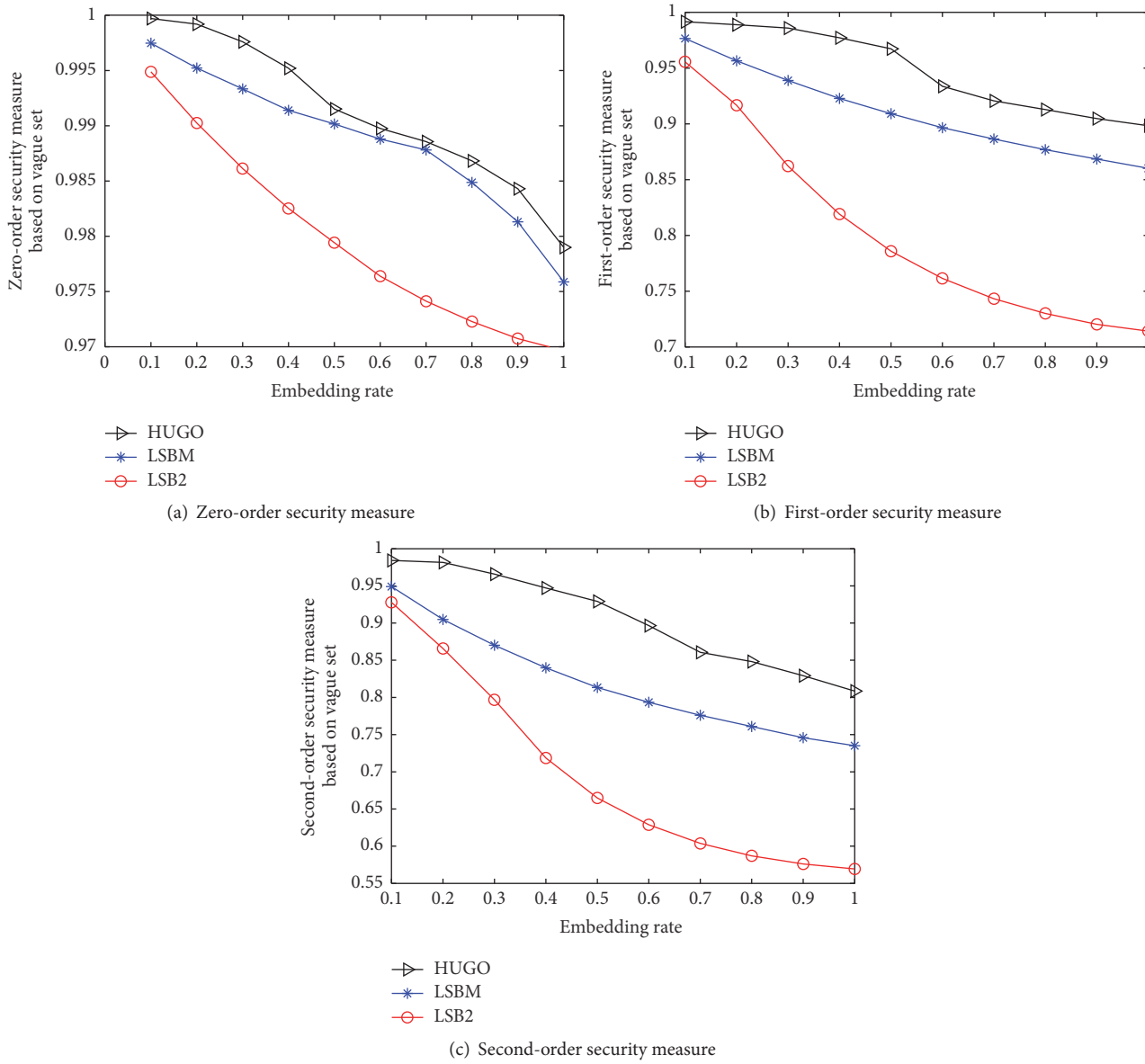


Figure 4: The same order security measure for different steganographic methods with different embedding rate.

$D(P_c \parallel P_s)$ is used under the assumption that the image model is i.i.d. Similarly, the first-order measure $T_1(MC, MS)$ is compared with $D(M_c, M_s)$ since their image pixel sequences are all modeled as the first-order Markov chain. In the experiments, the same 2000 images from BOWS2 are adopted. $T_0(MC, MS)$, $D(P_c \parallel P_s)$, $T_1(MC, MS)$, and $D(M_c, M_s)$ are used to measure the security of the HUGO with the embedding rate ranging from 0.05 bpp to 1 bpp in a step size of 0.05 bpp. Figures 6(a) and 6(b) show the average measure of $T_0(MC, MS)$ and $D(P_c \parallel P_s)$ with different embedding rates, respectively. The average measure values of $T_1(MC, MS)$ and $D(M_c, M_s)$ are also illustrated in Figures 7(a) and 7(b), respectively.

Looking at Figures 6 and 7, we see that the value of security measure based on vague sets decreases as the embedding rate increases, whereas the value of security measure under the deterministic distribution model increases as the embedding rate increases. All the curves in Figures 6 and 7 indicate that both the security measure models are effective in measuring the security of the steganography. In order to show the superiority of the proposed security measure, we define $\delta = \Delta y/y$ as the sensitivity of, where Δy is the security measure variation of a given embedding rate change range, andyis the total security measure variation of the embedding rate change. Obviously, Figures 6(b) and 7(b) demonstrate that δ of security measure is very small when embedding rate is lower than 0.5 bpp. So its corresponding

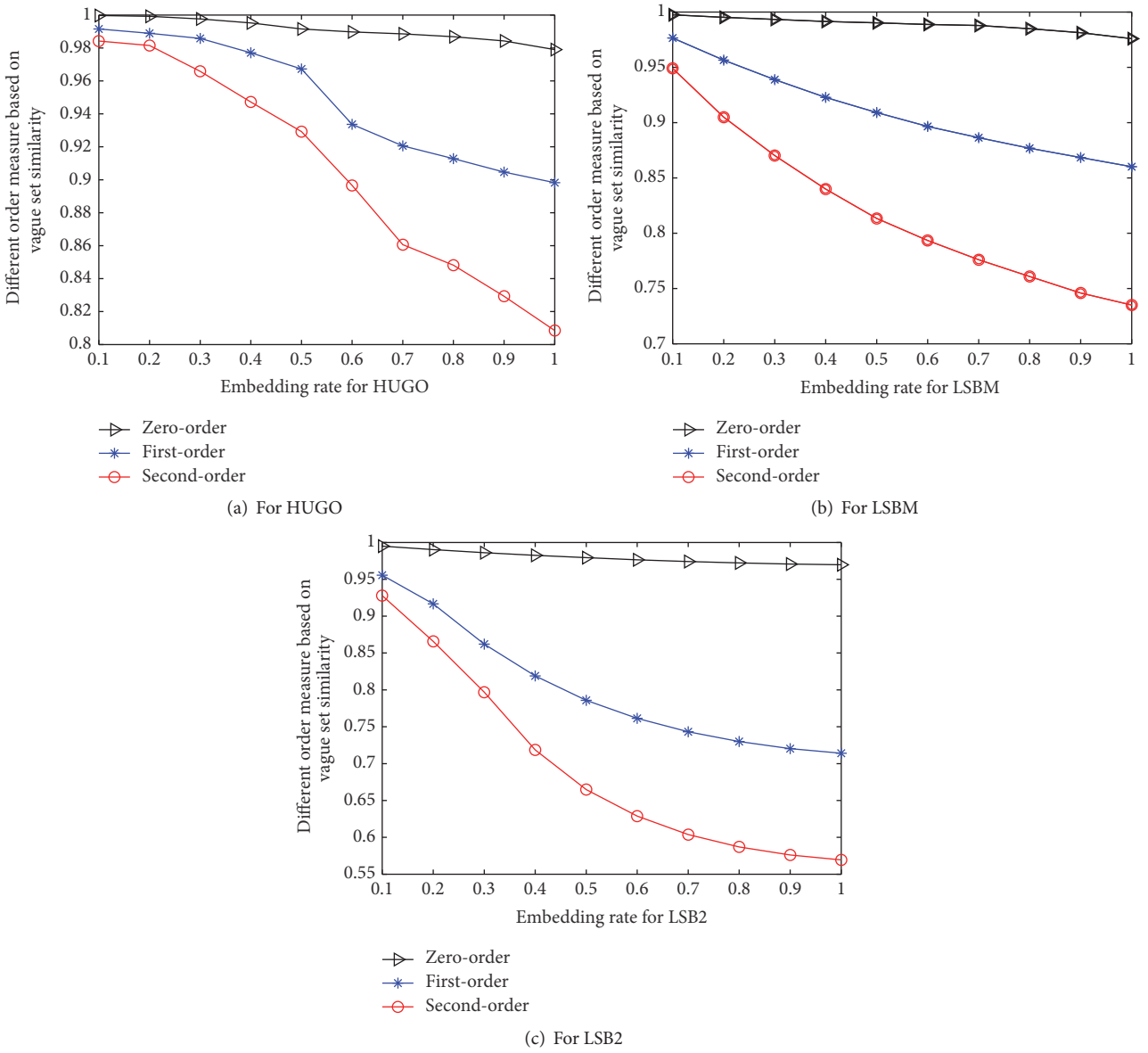
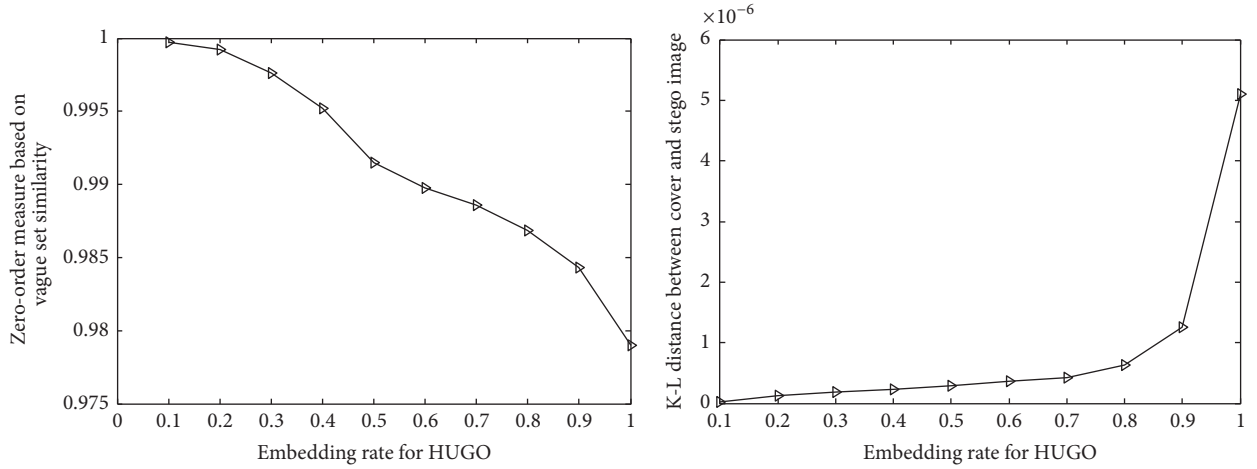


Figure 5: The different order security measures for the same steganography with different embedding rate.

security measure is not sensitive to the statistical distribution change. Hence, the new security measure can reveal more obvious statistical change than the security measures under deterministic statistical distribution model when embedding rate is low.

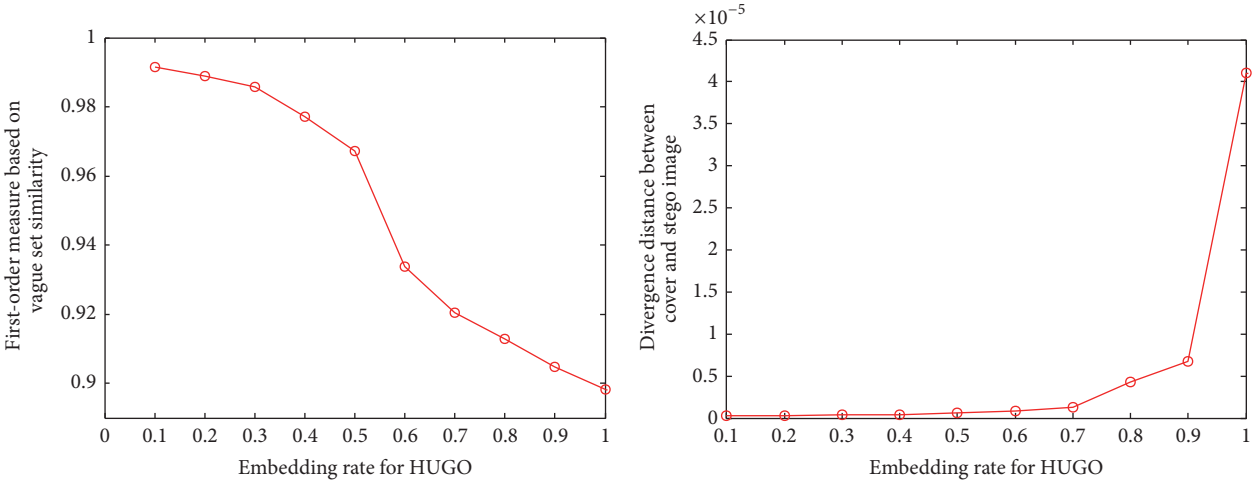
5. Conclusions

Vague sets similarity measure is a simple yet effective tool for measuring the similarity between two vague sets. In this work, a novel security measure for a steganographic system in terms of the vague sets similarity measure is proposed to measure the similarity between cover images and stego images. Particularly, in the new security measure, the sequence of image pixels is modeled as an *n-order* Markov chain to capture sufficient interpixel dependencies. The proposed security measure is proven to have such properties as boundedness, commutativity, and unity. Various order security measures can be obtained by adjusting the value of n. Experimental results confirm the effectiveness of the proposed security measure for evaluating different steganographic algorithms. Meanwhile, the security measure with a higher order always has a better measure ability. Additionally, when the embedding rate is low, the *n-order* security measure based on vague sets is more sensitive than other security measures under the deterministic distribution model. Considering the computational complexity and steganalytic ability, two issues should be tackled in our further research. One is how to use the *n-order* security measure to design reliable steganalytic methods by extracting the statistical feature from the empirical matrixes. The other is



(a) $T_0(MC, MS)$ (b) $D(P_c \parallel P_s)$

Figure 6: $T_0(MC, MS)$ and $D(P_c \parallel P_s)$ for HUGO with different embedding rates.



(a) $T_1(MC, MS)$ (b) $D(M_c, M_s)$

Figure 7: $T_1(MC, MS)$ and $D(M_c, M_s)$ for HUGO with different embedding rates.

how to use the new security measure to design highly secure steganographic algorithms.

Appendix

Proof of Theorem 5. (1)

$$
\begin{aligned}
&E_{MC}(MS) + E_{MS}(MC) - m\ln 2\,(E(MC) + E(MS)) \\
&\quad = -\sum_{i_1,i_2\cdots i_{n+1}\in G}\Big[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \\
&\quad + f_{MC}\left(m_{ij}\right)\ln f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\Big] \\
&\quad - \sum_{i_1,i_2\cdots i_{n+1}\in G}\Big[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \\
&\quad + f_{MS}\left(m_{ij}\right)\ln f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\Big] \\
&\quad + \sum_{i_1,i_2\cdots i_{n+1}\in G}\Big[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \\
&\quad + f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\Big] \\
&\quad + \sum_{i_1,i_2\cdots i_{n+1}\in G}\Big[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right) \\
&\quad + f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\Big] \\
&\quad = \sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right. \\
&\quad \left. + f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right]
\end{aligned}
$$

Table 1: Different order vague sets security measure for different steganography methods.

Steganography method	Embedding rate (bpac)	Zero-order	First-order	Second-order
F3	5%	0.9768	0.9662	0.9569
	10%	0.9755	0.9647	0.9569
	15%	0.9714	0.9608	0.9498
	20%	0.9683	0.9584	0.9477
nsF5	5%	0.9865	0.9736	0.9593
	10%	0.9847	0.9711	0.9542
	15%	0.9840	0.9687	0.9531
	20%	0.9818	0.9656	0.9515
MB1	5%	0.9994	0.9879	0.9785
	10%	0.9991	0.9866	0.9699
	15%	0.9965	0.9849	0.9673
	20%	0.9959	0.9837	0.9656
MB2	5%	0.9999	0.9868	0.9687
	10%	0.9996	0.9842	0.9624
	15%	0.9987	0.9922	0.9617
	20%	0.9982	0.9818	0.9609

$$+\sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}+f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right]. \tag{A.1}$$

Since the inequality satisfies $\ln x \geq (1-1/x)$, we have

$$\begin{aligned}
&E_{MC}(MS)+E_{MS}(MC)-m\ln 2\left(E(MC)+E(MS)\right)\\
&\geq\sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\cdot\left(1-\frac{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right)+f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\cdot\left(1-\frac{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right)\right]\\
&+\sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\cdot\left(1-\frac{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right)+f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\cdot\left(1-\frac{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right)\right]\\
&\geq\sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)-t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)+f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)-f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right]\\
&+\sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)-t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)+f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)-f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\right]=0.
\end{aligned} \tag{A.2}$$

Hence $E_{MC}(MS)+E_{MS}(MC)-m\ln 2(E(MC)+E(MS))\geq 0$, such that $E_{MC}(MS)+E_{MS}(MC)\geq m\ln 2(E(MC)+E(MS))$.

Since $t_{MC}(m_{i_1,i_2\cdots i_{n+1}})$, $t_{MS}(m_{i_1,i_2\cdots i_{n+1}})$, $f_{MC}(m_{i_1,i_2\cdots i_{n+1}})$, and $f_{MS}(m_{i_1,i_2\cdots i_{n+1}})$ are all in the range of $[0,1]$ and $0\times\ln 0=0$, $E(MC)$, $E(MS)$, $E_{MC}(MS)$, and $E_{MS}(MC)$ are all positive.

Hence $0\leq T_n(MC,MS)\leq 1$.

(2) According to the definition of the *n-order* security measure, $T_n(MC,MS)$ is described as

$$T_n(MC,MS)=\frac{m\ln 2\left(E(MC)+E(MS)\right)}{E_{MC}(MS)+E_{MS}(MC)}. \tag{A.3}$$

And it can also be described as

$$T_n(MS,MC)=\frac{m\ln 2\left(E(MS)+E(MC)\right)}{E_{MS}(MC)+E_{MC}(MS)}. \tag{A.4}$$

Hence $T_n(MC,MS)=T_n(MS,MC)$.

(3) From the proving procedure of property (1), we have

$$E_{MC}(MS) + E_{MS}(MC) - m\ln 2\,(E(MC) + E(MS)) \geq 0, \tag{A.5}$$

$$\begin{aligned} &E_{MC}(MS) + E_{MS}(MC) - m\ln 2\,(E(MC) + E(MS)) \\ &= \sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right. \\ &\left. + f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right] \\ &+ \sum_{i_1,i_2\cdots i_{n+1}\in G}\left[t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right. \\ &\left. + f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)\ln\frac{f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}{f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right)}\right]. \end{aligned} \tag{A.6}$$

If and only if

$$\begin{aligned} t_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) &= t_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right), \\ f_{MC}\left(m_{i_1,i_2\cdots i_{n+1}}\right) &= f_{MS}\left(m_{i_1,i_2\cdots i_{n+1}}\right). \end{aligned} \tag{A.7}$$

Namely, when $MC = MS$ and $MC = MS$, such that $E_{MC}(MS) + E_{MS}(MC) - m\ln 2(E(MC) + E(MS)) = 0$.

Hence $T_n(MC, MS) = 1 \Leftrightarrow MC = MS$. □

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Foundation of China (nos. 61462046, 61363014), the Science and Technology Research Projects of Jiangxi Province Education Department (nos. GJJ16079, GJJ160750), the Natural Science Foundation of Jiangxi Province (nos. 20151BAB207026, 20161BAB202050, and 20161BAB202049), Jinggangshan University Doctoral Scientific Research Foundation (nos. JZB1311, JZB15016), and Key Laboratory of Watershed Ecology and Geographical Environment Monitoring of NASG (nos. WE2015012, WE2016013).

References

[1] B. Li, J. He, J. Huang, and Y. Q. Shi, "A survey on image steganography and steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, 2011.

[2] X.-P. Zang, Z.-X. Qian, and S. Li, "Prospect of digital steganography research," *Journal of Applied Sciences-Electronics and Information Engineering*, vol. 34, no. 5, pp. 475–489, 2016.

[3] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[4] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, article A1, pp. 219–228, 2015.

[5] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[6] N. Zhou, H. Li, D. Wang, S. Pan, and Z. Zhou, "Image compression and encryption scheme based on 2D compressive sensing and fractional Mellin transform," *Optics Communications*, vol. 343, pp. 10–21, 2015.

[7] L. Guo, J. Ni, and Y. Q. Shi, "Uniform embedding for efficient JPEG steganography," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 814–825, 2014.

[8] N. Zhou, S. Pan, S. Cheng, and Z. Zhou, "Image compression-encryption scheme based on hyper-chaotic system and 2D compressive sensing," *Optics and Laser Technology*, vol. 82, pp. 121–133, 2016.

[9] R. Anderson, "Why information security is hard - An economic perspective," in *Proceedings of the 17th Annual Computer Security Applications Conference, ACSAC 2001*, pp. 358–365, usa, December 2001.

[10] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, vol. 192, no. 1, pp. 41–56, 2004.

[11] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis for Markov cover data with applications to images," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 275–287, 2006.

[12] Z. Zhang, G. J. Wang, W. Jun et al., "Steganalysis of spread spectrum image steganography based on high-order markov chain mode," *ACTA Electronica Sinica*, vol. 38, no. 11, pp. 2578–2584, 2010.

[13] G.-J. Liu, Y.-W. Dai, Y.-X. Zhao, and Z.-Q. Wang, "Modeling steganographic counterwork by game theory," *Journal of Nanjing University of Science and Technology*, vol. 32, no. 2, pp. 199–204, 2008.

[14] P. Schöttle and R. Böhme, "Game theory and adaptive steganography," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 760–773, 2016.

[15] J. Liu and G.-M. Tang, "Game research on large-payload and adaptive steganographic counterwork," *Acta Electronica Sinica*, vol. 42, no. 10, pp. 1963–1969, 2014.

[16] R. Chandramouli, M. Kharrazi, and N. Memon, "Image steganography and steganalysis: concepts and practice," in *Proceedings of the IWDW'03*, vol. 2939, pp. 35–49, 2003.

[17] T. Pevný and J. Fridrich, "Benchmarking for steganography," in *Information Hiding. 10th International Workshop*, pp. 251–267, Santa Barbara, Calif, USA, 2008.

[18] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive noise modelable information hiding," in *Proceedings of the IST/SPIE 15th Annu. Symp. Electronic Imaging Science Technology*, pp. 21–24, San Jose, Calif, USA, January 2003.

[19] F. Li and Z.-Y. Xu, "Measures of similarity between vague sets," *Journal of Software*, vol. 12, no. 6, pp. 922–927, 2001.

[20] S. Y. Quan, "The vague set similarity measure based on Meaning of Information," *Computer Engineering and Applications*, vol. 43, no. 25, pp. 87–89, 2007.

[21] W. L. Gau and D. J. Buehrer, "Vague sets," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 2, pp. 610–614, 1993.

[22] P. Bas, T. Filler, and T. Pevny, "Break our steganographic system—the ins and outs of organizing BOSS," in *Proceedings of the 13th International Workshop on Information Hiding*, pp. 59–70, Berlin, Germany, 2011.

[23] United States Department of Agriculture, Natural resources conservation service photo gallery, [DB/OL] http://photogallery.nrcs.usda.gov, 2002.

[24] T. Sharp, "An implementation of key-based digital signal steganography," in *Proceedings of the Information Hiding Workshop*, vol. 2137, pp. 13–26, 2001.

[25] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6387, pp. 161–177, 2010.

[26] J. Fridrich, D. Soukal, and M. Goljan, "Maximum likelihood estimation of length of secret message embedded using ±K steganography in spatial domain," in *Proceedings of SPIE-IS and T Electronic Imaging - Security, Steganography, and Watermarking of Multimedia Contents VII*, vol. 5681, pp. 595–606, January 2005.

[27] P. Sallee, "Model-Based Steganography," in *Digital Watermarking*, T. Kalker, Ed., vol. 2939 of *Lecture Notes in Computer Science*, pp. 154–167, Springer, Berlin, Heidelberg, 2004.

[28] P. Sallee, "Model-based methods for steganography and steganalysis," *International Journal of Image and Graphics*, vol. 5, no. 1, pp. 167–189, 2005.