Case Study

JOURNAL OF UNIVERSITY OF BABYLON
For Pure and Applied Sciences (JUBPAS)

Vol.32; No.2.| 2024

Using Support Vector Machine to Detect Data Hiding in Color Images

Zaid Ibrahim Rasool*

- Computer Techniques Engineering Department, Al- Mustaqbal University, 51001, Hilla, Iraq

*Corresponding author email: zaid.ibrahim@uomus.edu.iq

Accepted:	6/4/2024			Published:	30/6/2024

ABSTRACT

Background:

A steganography investigation model that uses improved dim scale List of capabilities, for identifying information inside uncompressed RGB pictures.

Materials and Methods:

in this paper, a bunch of 2000 RGB pictures was made utilizing normal tones, in BMP design with a size of 512 x 512 pixels. Clean pictures are Inserted secret picture information, utilizing two Payload plans, 2 pieces for every channel (BPC) and 4 pieces for each channel. The arrangement of chosen highlights comprises of 24 elements for every variety channel, and 72 highlights for each variety channel The picture, incorporates dim-level co-event grid highlights, entropy Elements, and factual proportions of variety. The list of capabilities components is determined as Individual channels, joined into picture vector highlights. The cycle depends on steganography examination On AI, utilizing Paired vector machine (SVM) support Execute the exercise manual in MATLAB.

Results:

The outcomes showed that the new method had overcome the benchmark with more than 80% identification accuracy

conclusion :

a new Steganalysis model was introduced by using SVM. The model can recognize the presence of stowed-away information inside RGB variety pictures, removing the highlights from datasets of perfect and messy pictures and ordering utilizing the support Vector Machine calculation. the results were compared with the DNA algorithm.

Keywords: highlight set, entropy, AI, SVM, bit per channel, identification exactness, RGB.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جـامعة بـابل للعلـوم الصـرفة والتطـبيقية مجلة جـامعة بـابل للعلـوم الصـرفة والتطـبيقية مجلة جـامعة بـابل للعلـوم الصـرفة والتطـبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

INTRODUCTION

In this paper, most common way of finding the presence of stowed-away information inside a variety of pictures is a test in different fields like network protection and data security. The need to guarantee the trustworthiness of information has become clear through the rising dependence on advanced interchanges and the far-reaching utilization of picture-sharing stages.

Steganography, which is the principal purpose of the exploration paper around here, is a strategy where private data is concealed inside cover media, for example, a variety of pictures. Steganography methods exploit the properties of the RGB variety space to implant information at all critical pieces (LSBs) of pixel values or by altering specific parts of the picture, for example, variety channels.

High-level steganography strategies plan to limit distinguishable changes via cautiously changing installing calculations to safeguard the nature of the picture containing the restricted intel. This requires the improvement of cutting-edge identification strategies equipped for distinguishing unpretentious alterations that may not be apparent to the natural eye.

Factors that influence the identification of stowed-away information inside a variety of pictures are pressure, commotion, and different types of picture handling that can darken the presence of the data. Thus, discovery calculations have been utilized that can successfully distinguish stowed away happy across assorted datasets and under various circumstances.

Factual investigation, signal handling procedures, and AI calculations have been utilized to distinguish and alleviate chances related to undercover interchanges. Resolving this issue is basic to guaranteeing the respectability, privacy, and legitimacy of computerized pictures.

8-bit dark scale Steganalysis of pictures has been the subject of broad examination [1-5], which has brought about models and capabilities that depend on dissecting the textural highlights of pictures. In [6], a list of capabilities is introduced which consolidates GLCM highlights [7], Entropy, and other factual proportions of variety. highlights are coordinated to quantify the properties of 50% of pictures' bytes. steganalysis of a variety of pictures has been applied to compacted pictures, for example, JPEG [8], and uncompressed pictures [9]. Research on highlight-based steganalysis of variety pictures has brought about sets that depend on characteristics of variety channels, and the connections between channels, like Variety Angle (CGCM) model [10, 11], and the Variety Rich Model [12].

The purpose of the suggested model is to identify whether an RGB cover image contains a hidden message, the identification problem requires the prior development of a paired classifier using controlled learning techniques on portions of a dataset of both clean and messy photos. The teganalysis method relies on combining single-channel components to create a multi-channel include set, without considering the relationship between different channels. Using a dataset of RGB clean images and the corresponding messy images, the identification accuracy of the suggested model is evaluated using implanting rates of 4 pieces per channel and 2 pieces per channel.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 1. Single channel features

Feature name	Feature description
CC-LR	Correlation coefficient between left and right half-bytes
CoV-FB	coefficient of variation of fuul-bytes
CoV-RHB	coefficient of variation of right half-bytes
GLCM-FB	Contrast, correlation, homogeneity, energy, of full bytes
GLCM-RHB	Contrast, correlation, homogeneity, energy, of right half- bytes
GLCM-3LSB	Contrast, correlation, homogeneity, energy, of 3 LSB part of bytes
GLCM-2LSB	Contrast, correlation, homogeneity, energy, of 2 LSB part of bytes
Entropy-FB	Entropy of full bytes
Entropy-RHB	Entropy of right half bytes
Skew-FB	Skewness of full bytes
Skew-RHB	Skewness of of right half bytes
Avediff-RHB	Average deference between successive right half bytes

The selected cover picture format is uncompressed RGB-BMP, which lacks the alpha channel in all three channels. Two public datasets are used so that the suggested model can be approved numerous times. The primary approved dataset consists of one thousand clear pictures in an alpha channel Altercation design that were retrieved from the NRC image collection [13]. The second approval dataset relies on CALTECH's bird pictures dataset [14], which is in a compacted variety JPEG. Figure 1 and figure 2 provide an example of the scaled and altered NRC and CALTECH images. Downloads of the two datasets are available in [15]. To create the messy datasets, the excellent images from the datasets were placed with privileged information of varying sizes. Three images served as the restricted information: a large image for high-limit inserting with the 4bpc plan, a medium image for low-limit installing with the 2bpc plan. The picture "House.bmp" from the USC-SIPI Picture Data collection contains the privileged information for the 3-channel high payload implanting with 4bpc [16], as seen in figure 3. It was scaled with a payload that was half the size of the cover photo in order to suit the most severe implanting limit of the selected cover pictures.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Figure 1. NRC cover image sample

Figure 2. CALTECH cover image sample

Figure 3. "Large secret image" House.bmp,379 KB, 50% payload

The image "Peppers.bmp" from Gonzales dataset [17] is used for the three-channel low payload installed with 2bpc, as shown in figure 4. It was downsized to fit the largest installation limit of the selected cover photo at a payload of 25% of cover picture's size.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Figure 4. 25% payload of medium secret image peppers.bmp,189 KB.

EXPERIMENTAL WORK

The suggested model is subjected to an exploratory assessment as part of the study approach, which involves implanting, extracting highlight vectors from both excellent and messy pictures, and arranging the extricated vectors using artificial intelligence [18].

Using spatial space steganography, clean images of the two datasets were inserted with privileged information. The three-channel implanting method used consecutively insertion of privileged information into each picture's pixel, whereby the largest parts were replaced with 2 or 4 implants per direct in each pixel. for single-channel implanting, just the NRC cover images were used, and 2bpc was used to put the blue variety channel in each pixel. Five messy datasets have been produced by the implanting cycles: NRC-LSB2, NRC-LSB4, CALTECH-LSB2, CALTECH-LSB4, and NRC-2LSB-Blue.

From each variety channel of the perfect and messy pictures, the selected channel-based highlights are obtained. The elements GLCM, Entropy, Standard Deviation, Relationship Coefficient, Mean, and Skewness are extracted using MATLAB. The (CoV) is specified as below:

$$COV= \frac{\text{Standard Deviation(n)}}{\text{Mean(n)}} \qquad (1)$$

The component extraction procedure yields a highlights information record in the Succeed CSV design that includes single direct elements in discrete worksheets. The single channel worksheets are joined into a picture highlights information record which comprises 72 component components. The picture highlights record is partitioned into a preparation subset and a testing subset, as indicated by the three-overlap cross-approval strategy. The preparation record is shaped by combining an equivalent number of spotless and messy picture highlight information. Likewise, the testing (concealed) records contain highlight information of pictures that are not pieces of the preparation information.

Due to its widespread use in example and image processing applications, the Support Vector Machine (SVM) classifier is selected in this step for the paired characterization task of picture handling [19]. The paired SVM calculation is executed in MATLAB by making use of the "svmtrain" and "svm order" functions. The "svm train" capability was produced using the preparation subset of highlights information vectors of a collection of 1000 perfect and 1000 messy photographs. The "svm characterize" capability was used to break down the unlabeled testing subset of the elements information vectors of 500 ideal and 500 messy pictures in light of the preparation result, where the terms "clean" and "messy" are assigned to every unlabeled test picture. A three-overlap cross-approval technique was used to repeat the testing and preparation procedure several times.

The accompanying measurements have been utilized in assessing the location execution of the proposed model: True Negative (TN): it refers to the proportion of true negative identifications to the quantity of clean pictures, True Positive (TP): it refers to the proportion of true positive location to the quantity of messy pictures, False Negative (FN): it refers to the proportion of false negative location of the quantity of messy pictures, False Positive (FP): it refers to the proportion of false positive location of the quantity of clean pictures.

Location Exactness: The proportion of accurately identified spotless and messy pictures to the all-out number of clean and messy pictures, as beneath:

$$\text{Detection Accuracy} = \frac{(TN+TP)}{(TN+TP+FN+FP)} \quad (2)$$

The experimental work is carried out in the stages shown in activity diagram Figure 5

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية
ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

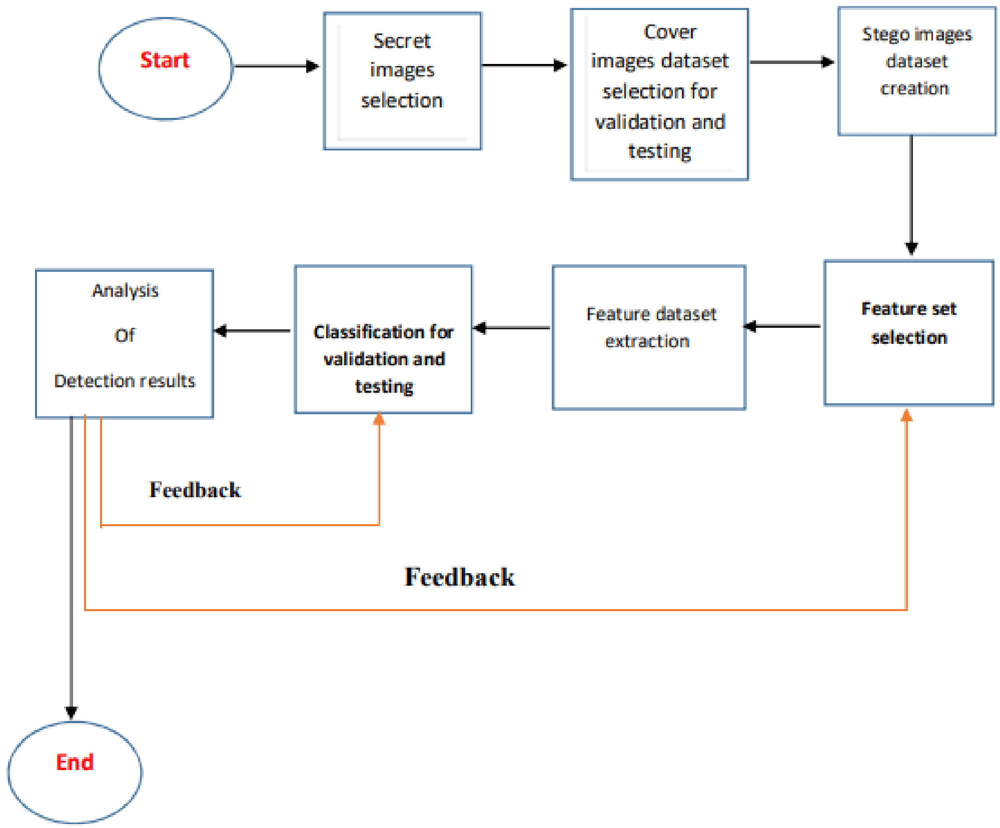

Figure 5 Stages of the experimental work

RESULTS AND DISCUSSION

Using a three-overlap cross-approval, the bunch order of the two data sets' perfect and messy pictures resulted in error rates of FN, TN, TP, and FP. Table 2 presents the identification accuracy effects as well as the error rates of the NRC dataset steganalysis using 4-bpc as well as 2-bpc installation. Likewise, the fact that the false positive (FP) and false negative (FN) values are identical shows that the suggested model has a reasonable placement for the two errors.

Table 2. Three-fold cross validation of RGB features results.

Metric	Average of three folds(%)	
	2-bpc	4-bpc
FN	0.07%	0.00%
FP	0.07%	0.00%
TN	83.88%	87.00%
TP	83.88%	87.00%
Acuracy	83.88%	87.00%

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

The error rates and detection accuracy outcomes employing 2bpc and 4-bpc embedding are displayed in Table 3. There is less noise in the results than in the NRC photos since they follow similar patterns.

Table 3. Results for Three-fold cross Validation for the Caltech dataset of RGB features.

Metric	Average of three folds(%)	
	2-bpc	4-bpc
FN	0.43%	0.00%
FP	1.22%	0.06%
TN	89.67%	90.00%
TP	90.50%	91.00%
Acuracy	90.10%	90.97%

The outcomes showed that the new method had overcome the benchmark with more than 80% identification accuracy .

CONCLUSION

Steganalysis model was introduced in the work to recognize the presence of stowed-away information inside RGB variety pictures, utilizing factual textural highlights that depend on properties of gray-scale 8-digit pictures, the chosen highlights were removed from datasets of perfect and messy pictures and ordered utilizing the support Vector Machine calculation.

Conflict of interests.

There are no conflicts to declare.

References

[1] A. Cheddad, J. Condell, K. Curran, and P. McKevitt, "Digital image steganography: survey and analysis of current methods," Signal Processing, vol. 90, pp. 727-752, 2010.

[2] K. L. Chiew, "Steganalysis of binary images," PhD thesis, Macquarie University, 2011.

[3] A. Aljarf, S. Amin, J. Filippas, and J. Shuttelworth, "Develop a detection system for grey and colour stego images," International Journal of Modeling and Optimization, vol. 3, no. 5, p. 458, 2013.

[4] A. Ker, "Steganalysis of LSB matching in grayscale images," IEEE Signal Processing Letters, vol. 12, no. 6, pp. 441-444, 2005.

[5] D. Lerch-Hostalot and D. Megıas, "LSB matching steganalysis based on patterns of pixel differences and random embedding," Computers & Security, Volume 32, Pages 192- 206. February 2013.

[6] M. M. Al-Jarrah, Z. Al-Taei, and A. Abuarqoub, "Steganalysis Using LSB-Focused Statistical Features," in Proceedings of the International Conference on Future Networks and Distributed Systems (ICFNDS '17), Vol Part F130522, Article number 3109814,2017.
[7] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.
[8] V. Holub, and J. Fredrich, "Low-complexity Features for JPEG steganalysis using undecimated DCT," IEEE Transactions on Information Forensics and Security, vol. 10, Issue: 2, Feb. 2015.
[9] H. Abdulrahman, M. Chaumont, P. Montesinos, B. Magnier, "Color images steganalysis using RGB channel geometric transformation measures,"Security and Communication Networks, vol. 9,issue 15, pp. 2945-2956, 2015.
[10] R. Gong, and H. Wang, "Steganalysis for GIF images based on colors-gradient co-occurrence matrix." Optics Communications, (vol. 285), no. 24, pp. 4961-4965, 2012.
[11] A. Aljarf, S. Amin, J. Filippas, and J. Shuttelworth, "The Development of an Images Detection System Based on Extracting the Colour Gradient Co-occurrence Matrix Features," in 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 2016, pp. 260-267. doi: 10.1109/DeSE.2016.
[12] M. Goljan, J. Fridrich, and R. Cogranne, "Rich model for steganalysis of color images," IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 185-190). IEEE, 2014.
[13] NRCS Photo Gallery, natural resource, and conservation related photos, https://photogallery.sc.egov.usda.gov/res/sites/photo gallery, viewed on 1/2/2017.
[14] H. Olguin-Garcia, O. U. Juarez-Sandoval, M. Nakano-Miyatake, and H. Perez-Meana, "Color Image Steganalysis Method for LSB Matching," 2015.
[15] H. "Color image steganalysis method for LSB matching," in Computer Engineering and Applied Computing (World Comp), pp. 27-30, (2015).
[16] Patel, C. B., Wandra, K. H., & Shah, S. (2016). Pixel swapping and parity based image
[17] steganography algorithm. In Electrical, Electronics and Computer Science (SCEECS).

[18] Paul, R. T. (2011). Review of robust video watermarking techniques. IJCA Special Issue on Computational Science.
[19] Prakash, G. S. (2006). Measures for Classification and Detection in Steganalysis(Doctoral dissertation, Indian Institute of Science, Bangalore).
[20] Schaathun, H. G. (2012). Machine learning in image steganalysis, Wiley and Sons,Sujatha, P. Purushothaman, S. & Rajeswari, R. (2012). Detecting the presence of hidden information using back propagation neural network classifier. International Journal of Computer Science and Information Security.
[21] Swanson, M.D., Kobayashi, M., Tewfik, A.H. (1998). Multimedia data-embedding and watermarking technologies, Proc of the IEEE, (vol. 86), no. 6, pp. 1064-1087.
[22] Thiyagarajan, P. Aghila, G. & Venkatesan, V. P. (2012). Steganalysis using color model conversion.
[23] Westfeld, A. & Pfitzmann, A. (2000). Attacks on Steganographic Systems.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

[24] Wiles, J. & Rogers, R. (2007). Techno security's guide to managing risks for it managers, auditors, and investigators. Security & Networking.

[25] Q. Liu, X. Kang, W. Zhao, and W. Xu, "Color image steganalysis using support vector machines," in IEEE International Conference on Multimedia and Expo (ICME), 2010.

[26] P. S. Hiremath and M. S. Sutaone, "Feature based detection of data hiding in color images using SVM," Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 3, no. 9, pp. 978–982, 2013.

[27] D. S. Saini, A. K. Khandelwal, and M. R. Kaim, "Steganalysis of color images using SVM classifier," Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 2, no. 7, pp. 215–218, 2012.

[28] A. Z. Połom and W. Koszela, "Data hiding detection using support vector machines and decision trees," in 10th Int. Conf. Comput. Recognition Syst. CORES, 2017, pp. 387–396.

[29] A. Cheddad, J. Condell, K. Curran, and P. McKevitt, "Steganalysis of color images using content-based features and support vector machines," IEEE Trans. Inf. Forensics Secur., vol. 5, no. 4, pp. 820–829, 2010.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

الخلاصة

خلفية:

نموذج تحقيق إخفاء المعلومات الذي يستخدم قائمة إمكانيات ذات مقياس خافت محسنة لتحديد المعلومات داخل صور RGB غير المضغوطة .

المواد والأساليب:

في هذا البحث تم إنتاج مجموعة من 2000 صورة RGB باستخدام الألوان العادية، وبتصميم BMP بحجم 512×512 بكسل. يتم إدراج الصور النظيفة كمعلومات صورة سرية، باستخدام خطتي حمولة، قطعتين لكل قناة (BPC) و4 أجزاء لكل قناة. يتكون ترتيب الإبرازات المختارة من 24 عنصرًا لكل قناة متنوعة، و72 تمييزًا لكل قناة متنوعة. تتضمن الصورة ميزات شبكة الأحداث المشتركة ذات المستوى الخافت وعناصر الإنتروبيا والنسب الفعلية للتنوع. يتم تحديد قائمة مكونات القدرات كقنوات فردية، مدمجة في تمييزات متجهات الصورة. تعتمد الدورة على فحص إخفاء المعلومات على الذكاء الاصطناعي، باستخدام دعم آلة المتجهات المقترنة (SVM) تنفيذ دليل التمرين في MATLAB.

نتائج:

وأظهرت النتائج أن الطريقة الجديدة قد تغلبت على المعيار بدقة تزيد عن 80% في تحديد الهوية

خاتمة:

تم تقديم نموذج جديد لتحليل الإخفاء باستخدام .SVM يمكن للنموذج التعرف على وجود معلومات مخزنة داخل صور متنوعةRGB ، وإزالة النقاط البارزة من مجموعات البيانات للصور المثالية والفوضوية والترتيب باستخدام حساب Vector Machine الداعم. وتمت مقارنة النتائج مع خوارزمية الحمض النووي.

الكلمات المفتاحية: مجموعة التمييز ، الإنتروبيا، الذكاء الاصطناعي، SVM، البت/لقناة، دقة التعريف، .RGB