Watermark by Learning Non Saliency

Dao Nam Anh1, Pham Quang Huy1 and Luong Chi Mai2

1 Electric Power University, Hanoi, Vietnam
2 Institute of Information Tehnology, Hanoi, Vietnam

**Abstract** – This paper studies the capability of non-salient image regions and pattern selection for advancing the security robustness of region based watermark. However, considerable unpredictability in the assigning image region for encryption exists, since saliency map itself is changed specifically by particular encryption algorithm and saliency detection techniques. The change may lead to misguided selection of image region for decryption. A new solution for watermarking in non-salient region is described and evaluated to derive improvement of secrecy, integrity and availability of images. This is achieved by using machine learning approach, which favors the advantageous selection of encryption region based on the variation between the salient maps before and after the encryption process. We validate the method by region selection evaluation and the correctness of decrypted messages. Experimental results show that a range of saliency models are adequate for the proposed watermark solution.

1. Introduction

It is big concern facing digital technology today is how to protect copyrights from cyber infringements. Whilst multimedia applications facilitate the copying, modification and distribution, ideally they should be capable for protection of copyrights and security of content. Digital watermarking [1], [2] is a process in which digital information is embedded in a host signal such as audio, image or video. The process's aim is tracing authenticity protection with especially adhering confidentiality and data integrity. However, the robustness relies on appropriateness of method and type of data. Four data types have been studied covering text, sound, image and video. In the case of image especially significant to keep track of large number of image types and kind of embedded information which can be text or image. Since visual attention [3] of human perception is addressed by saliency models [4], [5], [6], [7] saliency map can be used for watermarking. The purpose of this letter is to adapt a few saliency model to use the non-salient region for invisible watermarking with machine learning to advance the reliability of watermark. The method, termed here the watermark by learning non saliency (WLNS) method is described and demonstrated. The key contributions of this paper are: (a) A new invisible watermarking method with non salient regions to increase imperceptibility is defined. To significantly reduce change of non-salient region we apply a machine learning method. (b) Experiments for a benchmark image data result in a significant robustness for a few saliency models while leading to a slight loss of correctness. (c) The concept of non-salient watermarking can be checked with different saliency models and encryption methods.

1. Study Area

The concept of image saliency was first conceived by [3] allowing estimation of perception for visual attention. In watermark, the concept is used for the analysis of location for information encryption. Returning to the embedding a visible watermark for copyright notice, the method of [8] in Fig. 1 searches focus region for each video frame expecting that the region catches user’s attention the most. Hence, regions far from focus region are selected to embed a visible watermark in order to avoid overlay focus region but showing copyright notice.



**Fig 1.**  Map of watermarking saliency based methods in relation to imperceptibility, frequency patterns, image region and machine learning.

As an opposite of the visible mark, suitable regions obtained from saliency map, in principle, are used for invisible watermarking. Thus, visually salient and non-salient regions are reserved for embedding lower and higher strength watermarks accordingly in [9]. In fact, besides the difference related to the spatial contrast, the salient feature for watermarking can be calculated from luminance adaptation and just noticeable distortion (JND) which refers to the maximum distortion threshold that the human visual system cannot perceive [10], [11]. A robust and imperceptible blocks can be selected from such salient feature for watermarking in discrete cosine transform (DCT) domain. There is a large literature on the watermarking derived from saliency analysis which includes frequency patterns [12], [13], [14] calculate features of the reference image or host image as well as embedding values associated with frequency patterns. For example, a host image is decomposed into sub-bands by wavelet transform in a quality aware image approach to achieve the features [15], [16], [17] and by the least significant bits (LSB) [18].

In case of video watermarking, global motion compensated wavelet based visual attention is estimated for embedding information in accordance with the visual attentiveness [19]. Hiding information in image region is an effective solution of watermarking especially in relation to availability of saliency models. There has been a reported particular watermarking map by combination of non-saliency and heterogeneity- brightness [20] for locating the best places to embed watermark. Here, the DCT middle frequency factors of the places are used for hiding data [21], [22], [23], [24], [25]. Saliency based watermark methods addressed above and their relation to imperceptibility, frequency patterns, region base and machine learning are reported in Fig.1 which illustrates the relations by a map of domains. According to the domain map, our method is region based providing an imperceptible watermarking solution with machine learning approach.

1. Proposed method

Our method requires saliency detection and sub-region selection for securing hidden message in watermarking process. We take a color image and a message m we wish to embed into a region of the image. The saliency detection method *S* has its parameters that allow us to estimate the salient map *Is* by (1) and its feature *s* by (2) with respect to define a sub region we wish to hide a message:

|  |  |
| --- | --- |
|  | (**1**) |
|  | (**2**) |

A predefined set of sub regions *R* is used to allow us easily selecting a sub region *u* for embedding the message. We note the task by formula *Ie*=*E* (*I*, *R*, *S*, *m*), where *Ie* is watermarked image. Hence, we can see the saliency detection model with its parameters *S* and predefined set of sub regions *R* like private keys which are transferred from sender (Alice) to receiver (Bob) by a secure channel, while watermarked image *Ie* is sent by unsecured channel (Fig. 2). As the saliency detection model with its parameters *S* and a predefined set of sub regions *R* are securely known, the receiver Bob applies saliency detection for the watermarked image by the saliency model *S* to get a saliency map *Ie* and its feature *e* which may differ from saliency feature *s*. By using the predefined set of sub regions *R* in analyzing the saliency map to get saliency feature *e*, a sub region *r* is calculated and message *m* is decrypted from the sub region:

|  |  |
| --- | --- |
|  | (**3**) |
|  | (**4**) |



**Fig 2.**  Watermarking scheme for image *I* in communication between sender (Alice) and receiver (Bob) using saliency features with a secure channel for exchanging private keys *R*, *S*.



**Fig 3.**  Saliency based watermarking scheme, *s* - saliency feature, *u* - selected sub region *e-*saliency feature of watermarked image, and *r*- selected sub region of watermarked image.

Fig. 3 presents the concept for hiding message by saliency based watermark which consists of saliency feature of carrier image *s*, selected sub region *u* of the image, saliency feature of watermarked image *e*, and selected sub region *r* of watermarked image. Obviously, the method is validated when sub regions of watermarked image are identified accurately to get back the encrypted message.

3.1 Saliency Features

As shown in the Fig. 2, a set of sub regions *R* is predefined for embedding message into one of the sub regions. The set is designed secretly for encryption and decryption. A simple example of the set consists of four possible locations of rectangular, which are located along the border of image and covers 1/4 of image area, see Fig. 4. Such a set of sub regions is designed with expectation that salient regions are usually allocated inside of image and non-salient regions are often seen along the borders.



**Fig 4.**  Variable locations of sub region of image covering ¼ area of image

Given a salient map of image *Is*, in which salient level is in a range of [0, 1], salient mask for each pixel {0, 1} can be generated by a simple comparison with a threshold *λ*, by default *λ*=.5. Note that a morphological operation like *imerode* is efficient to apply to the mask for removing small objects and presenting salient regions by large areas. Once the salient mask has been found, the smallest distance from the salient objects to image frame borders can be estimated to show a non-salient area along image frame borders in a form of rectangular. Our saliency feature of image is indicated by the ratio of area of rectangular *vi* per area of whole image by (5). As such, sub regions addressed above with example in Fig.4 can have their indexes sorted by the order of the areas of rectangular *vi* by (6).

|  |  |
| --- | --- |
|  | (**5**) |
|  | (**6**) |

The indexes of sub regions *u* show levels of suitability for invisible hiding information as they are corresponding to non-salient regions along image borders. Thus the predefined sub regions with sorted indexes are formulated by. The message *m* now is hidden into the image in a sub region resulting watermarked image *Ie* by (7). For embedding text message into image there are many methods available which can be implemented to a sub region. Here, we choose LSB [18] for text embedding to check robustness by comparing encrypted message with its recovered version after changing of saliency feature affected by encryption task in the level of bits. In the receiver’s side, the watermarked image *Ie* is analyzed to get its saliency map and associated saliency feature *e*. Similarly to (5) the feature *e* is measured by the areas of rectangular *vj* by (8) and subregions of *Ie* can have its indexes sorted by the order of areas of rectangular *vj* by (9).

|  |  |
| --- | --- |
|  | (**7**) |
|  | (**8**) |
|  | (**9**) |

We use the provided descriptions for estimation of features {*s*, *u*, *e*, *r*} which characterize the watermarking process of encryption and decryption.

3.2. Watermarking Validation

To help clarify the effect of the watermarking method it may be helpful to view the process from Bayesian view [26] and easy to appreciate aspect of sub region selection basing on saliency features by (10). As *r* and *u* are sub region indexes, they has domain {0, 1, 2, 3}. Ideally, the sub region *r* should be the same the sub region *u*, *r*=*u* for ensuring encrypted message in the sub region can be detected and decrypted. The event *r*=*u* is presented through saliency feature*s* by (11).

|  |  |
| --- | --- |
|  | (**10**) |
|  | **(11)** |

To benefit from causality of features by (10) we expand (11) into (12) representing consequence of watermarking tasks. As near the starting event (*s*) where the location of sub region *u* is determined, we present (13) with potential.

|  |  |
| --- | --- |
|  | (**12**) |
|  | **(13)** |

This is repeated for saliency feature *e* of modified image by embedding message with potential by (14). For instance, the accumulation allows us to construct a final view of dependent events by (15).

|  |  |
| --- | --- |
|  | (**14**) |
|  | (**15**) |

Given initial image with saliency feature *s*, probability of having the same location of sub region *r*=*u* before and after watermarking is associated with feature *e*. A similar expression holds for *r*≠*u* by (16). Here we take the ratio of *p*(*r*=*u* |*s*) and *p*(*r*≠*u* |*s*) to know if degree of robustness for a watermarking case with saliency feature *s* by (17).

|  |  |
| --- | --- |
|  | (**16**) |
|  | (**17**) |

The way of creating saliency map are variable by different methods of saliency detection, and methods for encryption - decryption of message in images are different too. This follows the change of saliency map and saliency feature before and after watermarking. However, the high expectation of having *r*=*u* for all by (17) is impractical for different carrier images *I* and variable text messages *m* in general. If exists at least one sub region *r* which brings *r*=*u* after watermarking we already have proper location of decrypted sub region for message decryption although saliency feature *e* is different with *s* by (18). By bounding to the condition of existing a sub region *i* instead to all sub regions (18), we can see the expectation of having *r*=*u* in expression (19), representing dependency on the feature *e* of the watermarked image.

|  |  |
| --- | --- |
|  | (**18**) |
|  | (**19**) |

Consider the problem in classification approach for feature space {*s*}. Write {0, 1} for class domain meaning *r*≠*u* or *r*=*u*. As mentioned in the Fig. 3, to define which sub region *r* for embedding message given *s*, the saliency feature *e* has its impact like a hidden state. It is necessary to apply machine learning techniques for learning the influence of the hidden state (Fig. 5). Feature *s* is exported for each images *I* from a database by a set of saliency detection models to allow watermarking encryption and decryption. Message is embedded to each sub region *u* to create specific watermarked image *Ie* for which feature *e* is extracted in decryption task. The sub region *r* is defined for decryption while robustness is estimated by comparing *u* with *r*.



**Fig. 5.** Watermarking basing saliency detection schema

Let *θ* present the saliency model used for extracting saliency feature, the likelihood function *L*(*θ*) measures robustness for the saliency model by addressing to statistical data exported in watermarking for images *I* from dataset *D* by (20). The maximum of the likelihood function with respect to *θ* is evaluated by maximum likelihood (ML) estimate [27] by (21). Four saliency models for implementing the watermarking methods are tested for estimating the likelihood function *L*(*θ*) and experimental results are reported in section 4.

|  |  |
| --- | --- |
|  | (**20**) |
|  | (**21**) |

The ML estimate is realized by our leaning model with support vector machine (SVM) [28]. Features are collected in feature detection by watermarking basing saliency detection process mentioned above. The SVM learning is conducted for the assembled features to construct kernel machines.

1. Experimental Results

The task of identifying specific sub region for watermarking by learning non saliency feature needs clarification of robustness, since the likelihood function *L*(*θ*) by (20) is expected to be 100% for a selected saliency model. Therefore, our experiment has focused on the searching some saliency models which satisfy the expectation. Our initial set of methods for the learning includes the saliency using region covariance by [4], sparse saliency by [5], rare saliency by [6] and spectral residual saliency by [7].

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  | a.Original image I |  |
|  |  |  |  |
| b.Saliency map by [4] | Bottom sub region, *m*=”*wmgokhgksn*” | Watermarked *Ie* | New salient map  |
|  |  |  |  |
| c.Saliency map by [5] | Top sub region, *m*=“*vqiujhgxfklx*” | Watermarked *Ie* | New saliency map |
|  |  |  |  |
| d.Saliency map by [6] | Top sub region, *m*=“*dzuxzegwt*” | Watermarked *Ie* | New saliency map |
|  |  |  |  |
| e.Saliency map by [7] | Top sub region, *m*=“*qvdswweou*” | Watermarked *Ie* | New saliency map |

**Fig. 6.** Example of watermarking by various saliency models.

A saliency dataset of ten thousand images [29] is used in the learning experiments. For this dataset, saliency objects in images are diversified. A string is generated randomly for each text embedding. The image database is split randomly into two datasets: a dataset for training *DL* and a dataset for testing *DT*. For each saliency model, SVM training is conducted to achieve kernels. To select a sub-region for hiding message, the SVM checks saliency feature *s*, which extracted from saliency map of image *I*. An example is shown in Fig 6a: SVM with the saliency map by [4] advices to take the top sub-region for hiding a message *wmgokhgksn* that leads to new image *Ie* and new salient map. The map is different to its initial map. Note that the watermark in the image is imperceptive by human eyes. Having two images *I* and *Ie*, their difference can be measured by precision, recall, fmeasure [29] and reversed mean squared error (MSE) [30], sum of absolute differences (SAD) [31], structural similarity index measure (SSIM) and peak signal to noise ratio (PSNR) [2] for estimating resemblance instead of distinction. The first group measures for learning the degree of image change by watermarking derives the statistical report in Table I.

**Table 1.** Invisibility of watermark basing on non-saliency

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Saliency model | Precision\* | Recall | Fmeasure | rMSE |
| Covariance [4] | 0.8895 | 0.9947 | 0.9391 | 0.9638 |
| Highlighting [5] | **0.9266** | *0.9959* | **0.9573** | *0.9652* |
| Rare [6] | *0.9189* | **0.9965** | *0.9510* | 0.9651 |
| Spectral [7] | 0.9154 | 0.9944 | 0.9505 | **0.9655** |

\**Scores printed in bold are the best, italic scores are the second.*

Evaluating the scores of precision, recall, fmeasure and rMSE highlights the best and the second scores of invisibility mostly for highlighting sparse saliency [5] and rare saliency [6] models. The average scores from the test of image set DT are resumed in Table II.A. The performance of the learning with sparse saliency [5] and rare saliency [6] models with the best scores. This shows that the sub-regions with embedded messages by Alice for all the test images are always detected by Bob.

**Table 2.** Stability of saliency map in watermarking & Message correspondence

|  |  |
| --- | --- |
| A. Stability of saliency map in watermarking | B. Message correspondence |
| Saliency model | Precision | Recall | Fmeasure | rMSE | *jw* distance | Encryption time *(s*) | Decryption time (*s*) |
| Covariance [4] | 0.9365 | 0.9370 | 0.9367 | 0.9051 | 0.9017 | 0.2667 | 20.2770 |
| Highlighting [5] | **1.0000** | **1.0000** | **1.0000** | **1.0000** | **1.0000** | 0.1269 | 0.1705 |
| Rare [6] | **1.0000** | **1.0000** | **1.0000** | **1.0000** | **1.0000** | 0.0614 | 1.0465 |
| Spectral [7] | 0.9011 | 0.9053 | 0.9032 | 0.8545 | 0.8634 | 0.0462 | 0.1360 |

Statistic report with Jaro-Winkler distance is in Table II.B where *jw* scores of the watermarking learning with highlighting sparse saliency [5] and rare saliency [6] methods are the best. As such, our watermark method by learning non saliency can be employed robustly with both highlighting sparse saliency [5] and rare saliency [6] detection models. During development relationship between saliency features and selection of a sub region for hiding messages the SVM among machines learning techniques were applied. The effective robustness of sparse saliency [5] and rare saliency [6] were collected from test on separated dataset. Saliency is sensitive with image change by scaling, cropping and rotating. Hence the saliency based watermark resilience is delicate with the change of watermarked host image. The text-based watermark is more fragile than image based watermark as it requires correctness in the bits level for decryption. As a result, it can be affected by malicious attacks like image re-sizing or rotation. In order to avoid the problem, text can be transferred a small image which is then embedded into a sub-region of a host image. The ability to design a set of sub regions for embedding information is mostly large. The set displayed in Fig. 4 is the simplest having rectangles allocated along image borders. An alternative version of the set includes resizing rectangular to smaller size and moving the rectangles to the center of image by a secret small distance (Fig. 7). The pre-defined set of sub regions for hiding information is used as a private key in watermarking.

|  |
| --- |
| a. b. c. |

**Fig. 7.** Examples of arrangment of sub regions with different alignment: a. left, b. right, c. cente

Because the hiding task by a particular set of sub regions makes saliency map changed specifically by the set, this suggests to conduct learning by SVM for each set of sub regions and check robustness of the algorithm for the configuration of sub regions by (20) before using the set for watermarking.

1. Conclusions

A non saliency based watermarking method for hiding message in carrier image is described. This approach searches non-salient region for increasing invisibility for embedded message. In this work, SVM learning is special to select correctly a non salient sub region for encryption and decryption, providing solution for the problem caused by the change of saliency map in the embedding task. Four saliency detection models were tested with the methods with a saliency benchmark of ten thousands images and two of them are robust with the LSB message embedding. They are sparse saliency [5] and rare saliency [6]. The requirement of exactness of identification of sub region for decryption in LSB means that our saliency based watermarking method can be applied for hiding an image in a carrier image, which needs further study in future. The arrangement of sub regions used in our method is a variable subject and a part of private key. Hence, there may be designed specifically for each watermark case by the sender. Finally, the saliency based watermarking method can be implemented for other saliency models with robustness which is completely testable and certifiable by experiments for a large saliency benchmark database.

Acknowledgments

References

1. A.Z.Tirkel, G.A. Rankin, R.M. Van Schyndel, W.J.Ho, N.R.A.Mee, C.F.Osborne (1993). Electronic Water Mark. DICTA 93, Macquarie University. pp.666-673.
2. Satapathy, Suresh Chandra, et al. (2016), Computer Communication, Networking and Internet Security. Proceedings of IC3T 5.
3. L. Itti and C. Koch (2001). Computational Modelling of Visual Attention. Nature Reviews Neuroscience, 2(3):194–203.
4. Erkut Erdem, Aykut Erdem (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. Journal of Vision, 13(4):11, 1-20.
5. Xiaodi Hou, Jonathan Harel, Christof Koch (2012) Image Signature: Highlighting Sparse Salient Regions. IEEE Trans. Pattern Anal. Mach. Intell. 34(1): 194-201.
6. N. Riche, M. Mancas, B. Gosselin, T. Dutoit (2012). RARE: a New Bottom-Up Saliency Model, Proceedings of the IEEE International Conference on Image Processing.
7. Xiaodi Hou, Liqing Zhang (2007). Saliency Detection: A Spectral Residual Approach, IEEE Conference on Computer Vision and Pattern Recognition.
8. Chen-Hsiu Huang, Ja-Ling Wu (2004). Attacking Visible Watermarking Schemes, IEEE Transactions on Multimedia 6(1):16 - 30.
9. Deepayan Bhowmik, Matthew Oakes, Charith Abhayaratne (2011), Visual Attention-Based Image Watermarking, International Symposium on Circuits and Systems.
10. Y. Niu, M. Kyan, L. Ma, A. Beghdadi, S.Krishnan (2011). A visual saliency modulated just noticeable distortion profile for image watermarking, 19th ESP Conf.
11. Y. Niu, M. Kyan, L. Ma, A. Beghdadi, S. Krishnan (2013). Visual saliency's modulatory effect on just noticeable distortion profile and its application in image watermarking, Signal Processing Image Communication 28(8).
12. Ingemar J. Cox (2008). Digital watermarking and steganography. Morgan Kaufmann, Burlington, MA, USA.
13. M.K. Mihcak, I.Kozintsev, K.Ramchandran, P.Moulin (1999). Low-complexity imagedenoising based on statistical modeling of wavelet coefficients. IEEE Signal ProcessingLetters, 6(12):300–303.
14. Andrew B. Watson (1993). DCT quantization matrices optimized for individual images. Human Vision, Visual Processing, and Digital Display IV, SPIE-1913:202–216.
15. Lihua Tian, Nanning Zheng, Jianru Xue, Ce Li, Xiaofeng Wang (2011). An integrated visual saliency-based watermarking approach for synchronous image authentication and copyright protection. Image Commun. 26, pp. 427-437.
16. M. Sung, Li X., I. Lee, (2016). Visual perception based robust watermarking with integral imaging, Optik – Inter. J. for Light and Electron Optics, Vol 127, Iss 24, 11828-11839.
17. C. Wang, T. Zhang, W. Wan, X. Han, M. Xu (2017). A Novel STDM Watermarking Using Visual Saliency-Based JND Model, Information (Switzerland) 8(3):103.
18. A. Basu, S. Talukdar, N. Sengupta, A. Kar, S. Lal, Chakraborty, S. K. Sarkar (2015). On the Implementation of a Saliency Based Digital Watermarking, Information Systems Design and Intelligent Applications, AISC 339. Springer India, pp.447-456.
19. M. Oakes, D. Bhowmik, C. Abhayaratne (2016). Global Motion Compensated Visual Attention based Video Watermarking, SPIE, 25(6), pp.061-624.
20. A. Gawish, C. Scharfenberger, H. Bi, A. Wong, P. Fieguth, D. Clausi, (2016). Robust Non-saliency Guided Watermarking, 13th Conf on Computer and Robot Vision.
21. Z. Zhang, L. Wu, Y. Yan (2017). An improved reversible image watermarking algorithm based on difference expansion,International Journal of Distributed Sensor Networks.
22. S.-K. Ji, W.-Hyung, K.H.-Ul, J.S.-Min, M.H.-Kyu Lee (2015). Robust Imperceptible Video Watermarking for MPEG Compression and DA-AD Conversion Using Visual Masking, IWDW 2015: Digital-Forensics and Watermarking, pp 285-298.
23. C. Wang, T. Zhang, W. Wan, X. Han, M. Xu (2017). A novel STDM watermarking using visual saliency-based JND model, Information (Switzerland) 8(3):103.
24. L. Tian, N. Zheng, J. Xue, C. Li (2013). Authentication and copyright protection watermarking scheme for H.264 based on visual saliency and secret sharing, Multimed Tools Appl.
25. Frank Y. Shih, Xin Zhong (2016).Intelligent Watermarking for High-Capacity Low-Distortion Data Embedding, Int. J. Patt. Recogn. Artif. Intell. 30, 1654003.
26. Barber, D. (2012). Bayesian Reasoning and Machine Learning, Cambridge Univ. Press.
27. Milton, J. S., Arnold, J. C. (1995). Introduction to Probability and Statistics, Principles and Applications for Engineering and the Computing Sciences. McGraw-Hill.
28. Vapnik, V. N. (2014). Invited Speaker. IPMU Information Processing and Management.
29. Ali Borji, Ming-Ming Cheng, Huaizu Jiang, Jia Li, (2015). Salient Object Detection: A Benchmark, IEEE TIP, vol. 24, number 12, pp 5706-5722.
30. Powers, David M. W, (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation (PDF). J. of Machine Learn. Tech. 2 (1): 37–63.
31. Winkler, W. E. (2006). Overview of Record Linkage and Current Research Directions (PDF). Research Report Series.