Localization of Embedding Regions for Watermarking in Medical Images

[1]RandheerBagi, [2]Tanima Dutta, [3]Hari Prabhat Gupta
[123]Dept. of CSE
[123]IIT (BHU) Varanasi, India

Abstract: --The localization of regions which are not sensitive to the human visual system in an image is one of the most important parts of watermarking in multimedia objects. Traditionally, the localization of such regions is performed manually using segmentation or background subtraction. In this paper, we proposed an automatic localization method for finding embedding regions which are insensitive to the human visual system. Localization of the embedding region is done by using the convolutional neural network. The experimental results show that the accuracy of the proposed technique is above 95%. This proves the efficiency of the proposed method. We perform the training by using block selection method on Lenna (256×256) image by setting the threshold on the flatten level of pixels.

Introduction

Data hiding techniques like digital watermarking and steganography are promising technologies for multimedia information protection and rights management. The multimedia objects are audio, video, images, and texts. Digital representation of multimedia objects brings many advantages when compared to analog representations, such as lossless recording and copying, convenient distribution over networks, easy editing and modification, and durable, cheaper, easily searchable archival. Unfortunately, the transmission of such multimedia objects over the communication network suffers threats of different intruders and attackers which include widespread copyright violation, illegal copying and distribution, problematic authentication, and easy forging. Piracy of digital multimedia objects is already a common phenomenon on the Internet. Today, digital photographs or videos cannot be used in the chain of custody as evidence in the court because of the nonexistence of a reliable mechanism for authenticating digital images or tamper detection. Data hiding in digital multimedia objects provides a means for overcoming those problems. Therefore, digital watermarking and watermarking techniques, gain importance for authenticity of the such multimedia objects. It is also important to identify the Region of Interest (ROI) for hiding data, such that the hidden information may not degrade the quality of the image significantly as well as the image can be authenticated using that hidden information whenever required.

In classical methods, ROI in images are localized using different image processing techniques, such as image segmentation. Such methods are broadly categorized as integer transforms based, histogram bin shifting based, data compression, prediction of pixel values, and modification of frequency domain characteristics. The objective of such techniques is to maximize the embedding capacity in an image in such a way that the degradation of the image due to embedding is imperceptible to the Human Visual System (HVS). However, exact localization of ROI in the images is a challenging task. Also, techniques for embedding varies when the type of the image changes, e.g., medical images, military images, and grayscale images. The manual extraction of ROI from an image involves domain expertise in the identification of ROI. Improper localization of the region of interest may affect the watermarking accuracy in a very drastic way.

In this paper, we focus on automatically predicting ROI which is insensitive to HVS based on the existing images present in the database. We are not dependent on any dedicated image processing technique. To the best of our knowledge, none of the frameworks focus on the automatic extraction of ROI based on the existing images in the database and without using dedicated image processing techniques. We build a model that automatically localizes the ROI in an image. The localized ROI represents the embedding regions where we can hide some data within the image. The embedding region is selected in such a way that after embedding the hidden information into the localized region, the degradation of image is not perceptible by HVS. This model minimizes the error in localizing ROI with less computation and time complexity.

The contribution of this paper is summarized as follows:

1) In our proposed method, we highly enhance the embedding capacity of the image, so that more regions are available in ROI for watermarking. Therefore, more amount of hidden information can be embedded. The in- crease in embedding capacity, however, does not impact the quality of the image.

2) The increase in embedding capacity also increases the security in the watermarking process. Reason for better security is that, when more ROI is available higher combination of watermarking regions can be obtained.

3) Extensive simulations are conducted to evaluate the accuracy of the proposed technique. The testing is performed on datasets [1], [2]. Experimental results indicate that the number of ROIs is high even in noisy images without degrading the quality of the image.

The rest of the paper is organized as follows. The Section II and III describe the related work and the Preliminaries, respectively. In Section IV, the proposed work is presented. Section V validates the proposed work using the several experimental results. Finally, we conclude our work in Section VI.

Related Work

With the advancements in the field of digital multimedia processing during the last decade, digital data hiding techniques such as watermarking and steganography have gained wide popularity. Several algorithms have been proposed in the field of hiding information and its retrieval from the multimedia objects, especially in images. Image watermarking techniques can be categorized into the spatial and frequency domain embedding [3], [4] and hiding into the encrypted images [5], [6].

One of popular technique for reversible watermarking is implemented using the Difference Expansion (DE) method [4]. This DE method is further generalized in [7] where a pixel pair is replaced by arbitrary block size. Other enhancements of DE are described in [8] that include the prediction error expansion, which considers the correlation between larger neighborhood pixels. In [9], adaptive embedding is discussed. It uses DE with sorting technique (sorting pixel pairs according to local variance) to improve the original DE.

In frequency domain, histogram shifting is one the popular technique [3], [10] used for watermarking. In [11], a lossless watermarking technique uses the pixel difference histogram shifting. It supports multi-layer embedding which improves the hiding capacity. Similarly, in [12], reference pixels and adaptive block selection method are used for watermarking where the reference pixel is considered as a median pixel value of the block. In [13], the tamper localization is performed in medical images. The medical image is divided into the regions, Region of Interest (ROI) and Region of Non-Interest (RONI) by 8×8 non-overlapping blocks. Then the tampered region is localized by using block tamper localization vector. Similarly, in [14], the localization of tamper is performed by constructing blocks of 8×8 using discrete cosine transform where data is hidden into the middle frequency region of the coefficients.

Visual localization and segmentation are done by feedback Convolutional Neural Network (CNN) by capturing high-level semantic [15]. These semantic concepts are transformed into the image space and generate energy gaps. In [16], a scalable detection algorithm was proposed which uses high-capacity CNN for object segmentation and localization. To improve the performance it uses region proposals with CNN at training time. In [17], CNN activates the semantically meaningful regions. These regions will be provided as the input for the CNN to train the model. This specific classifier predicts the class labels with reduce computational complexity. CNN is also used in object localization in remote sensing Images [18]. It uses unsupervised score-based bounding box regression algorithm for accurate object localization process and non- maximum suppression for minimizing overlapped regions.

To the best of our knowledge, none of the method exists in the literature that uses a deep network model to predict the embedding regions which are insensitive to the human visual system but important of the image. The main focus of this paper is to localize maximum number of regions that can be used for hiding the data in the spatial domain of the image without degrading the quality of the image.

Preliminaries

This section covers a brief description of the working principle of CNN. Some of the important steps in CNN are Conv2D, Relu - Activation Function, Maxpooling and Fully- connected layer.

A. Convolutional Neural Network (CNN)

In machine learning, CNN is a class of deep feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNN uses a variation of multilayer perceptron designed to require minimal pre-processing. They are also known as shift invariant or space invariant artificial neural networks, based on their shared weights architecture and translation invariance characteristics.

B. Working of CNN

Convolutional layer is the first layer in CNN. It takes an input image as an array of pixel values. In the convolution a filter (an array of number, i.e., weights or parameters) is convolve over the input image array. The depth of the filter has to be same as the depth of input image array. It multiplies the values in the filter with the original pixel values of the image. These multiplications are all summed up, so now you have a single number. This number is just representative of when the filter is at the top left of the image. Now we repeat this process for every location on the input volume. The next step is moving the filter to the right by 1 unit, then right again by 1 and so on. Every unique location on the input volume produces a number. After sliding the filter overall the locations, we will find out an activation map or feature map. Here each of these filters can be thought of as a feature identifier.

Proposed Work

In this section discuss the automatic localization of ROI from an image using non-overlapping block selection method by setting the threshold on the flatten value level. To perform training and testing, we use the convolutional neural network (ConvNet). This section compromises training and testing phase which are involved in the localization of the region of interest (ROI).

A. Training Phase

In this phase, firstly we consider a standard image (Lenna) $I_{p\times q}$ as an input in the pre-processing phase. We divide the image $I_{p\times q}$ into square optimal blocks (b_1, $b_2 \dots b_n$) in the non-overlapping fashion. Non-overlapping block is define as the not covering the image regions which are selected earlier to perform some operation. With the variation in number of stride (1, 2, 3, 4, and 5) and kernel (3, 4, 5, 7) which are non-overlapping we select the optimal block size $I_{s\times s}$. Now we have an image $I_{p\times q}$ which is divided into small images blocks i.e., optimal block of size $I_{s\times s}$. So the total numbers of optimal square block we obtain are:

$$n = \frac{I_{p\times q}}{I_{s\times s}}. \qquad (1)$$

Let the number of pixels in a block b_i is z where z = s× s. The optimal block S_{opt} shows the maximum number of flatten value (FV) for watermarking. We compute a value α for each block b_i, which is define as the ratio of median and standard deviation for every block b_i i.e.,

$$\alpha_i = \frac{median\ [b_i]}{standard\ deviation\ [b_i]}. \qquad (2)$$

After computing the α_i of each block we subtract the each pixel z of the block b_i with α_i obtain from Eq. 2 of corresponding block,

$$D_i[j,k] = \alpha_i - \ b_i[j,k]. \qquad (3)$$

After calculating difference D_i[j, k], we count the number of flatten value (FV) of each block b_i. The flatten values are those value where D_i[j, k] == 0 for each b_i. It is define as:

$$f(u) = \begin{cases} ||FV|| = 1, & if\ D_i[i,j] == 0 \\ ||FV|| = 0, & otherwise. \end{cases} \qquad (4)$$

Here, ||.|| represent the cardinality of flatten value. Now for each block b_i we set a common threshold λ on flatten value (i.e., FV ≥ λ) based on domain knowledge. Here, for simplicity we set the λ as an average of the flatten value count from each b_i,

$$\lambda = \frac{1}{b_i} \sum_{D_i[j,k]=0} ||FV||. \qquad (5)$$

Now, after performing block division and computing difference D_i[j, k], two different set of images are created $ROI[I_{s\times s}]$ and $RONI[I_{s\times s}]$ based on the threshold value λ selected as in Eq. 5,

$$f(v) = \begin{cases} ROI[I_{s\times s}], & if\ FV\ \geq \lambda \\ RONI[I_{s\times s}], & otherwise. \end{cases} \qquad (6)$$

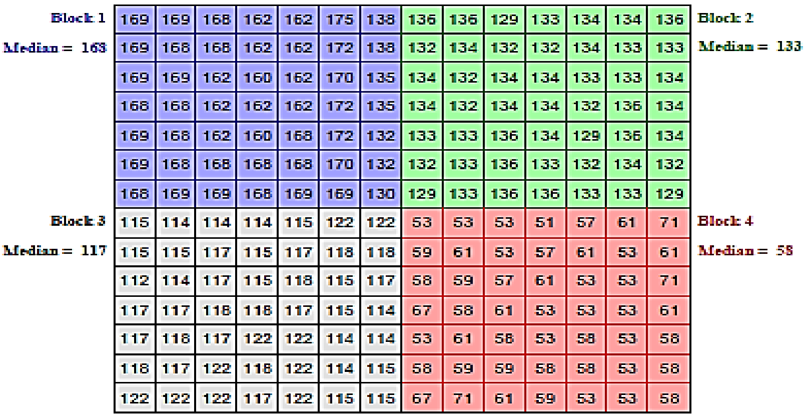
Let X is a set of $ROI[I_{s\times s}]$, and $RONI[I_{s\times s}]$i.e.,

$$X = ROI[I_{s\times s}] \cup RONI[I_{s\times s}]. \qquad (7)$$

Now we pass X to the ConvNet for training the model. At the very first step, the features are extracted from the images of set X using ConvNet. Let the extracted features of the images are: $X_f = [x_1, x_2 \dots x_n]$, where x_1 is the feature vector. Then these extracted features are pass-through the ConvNet with different activation functions like relu, tanh, leaky relu, elu and 6 hidden layers. A ConvNet model H is trained using every input of X and its corresponding feature vectors in X_f. The training is performed by back propagation process with learning rate 0.001 and dropout 0.8.

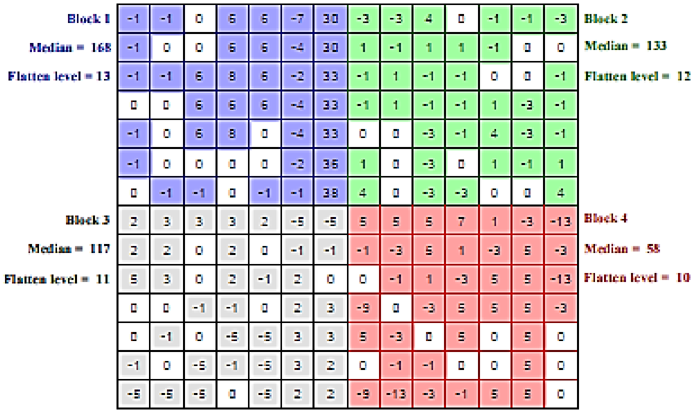
B. Testing Phase

In the testing phase, a test image Itest is first pre-processed. When we get I'test as a set of test image blocks comprising of ROI and RONI regions of Itest. After obtaining Itest it is passed through ConvNet H for extracting the features. These features are same as the once extracted during the training phase. Let the extracted features are Y1 = [y_1, $y_2 \ldots y_n$] where y1 is the feature vector. The extracted features are then passed through the ConvNet model H to obtain the predicted image Y2.



Block 1	169	169	168	162	162	175	138	136	136	129	133	134	134	136	Block 2
Median = 168	169	168	168	162	162	172	138	132	134	132	132	134	133	133	Median = 133
	169	169	162	160	162	170	135	134	132	134	134	133	133	134	
	168	168	162	162	162	172	135	134	132	134	134	132	136	134	
	169	168	162	160	168	172	132	133	133	136	134	129	136	134	
	169	168	168	168	168	170	132	132	133	136	133	132	134	132	
	168	169	169	168	169	169	130	129	133	136	136	133	133	129	
Block 3	115	114	114	114	115	122	122	53	53	53	51	57	61	71	Block 4
Median = 117	115	115	117	115	117	118	118	59	61	53	57	61	53	61	Median = 58
	112	114	117	115	118	115	117	58	59	57	61	53	53	71	
	117	117	118	118	117	115	114	67	58	61	53	53	53	61	
	117	118	117	122	122	114	114	53	61	58	53	58	53	58	
	118	117	122	118	122	114	115	58	59	59	58	58	53	58	
	122	122	122	117	122	115	115	67	71	61	59	53	53	58	

Fig. 1. A sample of 4 block for adaptive block selection.



Block 1	-1	-1	0	6	6	-7	30	-3	-3	4	0	-1	-1	-3	Block 2
Median = 168	-1	0	0	6	6	-4	30	1	-1	1	1	-1	0	0	Median = 133
Flatten level = 13	-1	-1	6	8	6	-2	33	-1	1	-1	-1	0	0	-1	Flatten level = 12
	0	0	6	6	6	-4	33	-1	1	-1	-1	1	-3	-1	
	-1	0	6	8	0	-4	33	0	0	-3	-1	4	-3	-1	
	-1	0	0	0	0	-2	36	1	0	-3	0	1	-1	1	
	0	-1	-1	0	-1	-1	38	4	0	-3	-3	0	0	4	
Block 3	2	3	3	3	2	-5	-5	5	5	5	7	1	-3	-13	Block 4
Median = 117	2	2	0	2	0	-1	-1	-1	-3	5	1	-3	5	-3	Median = 58
Flatten level = 11	5	3	0	2	-1	2	0	0	-1	1	-3	5	5	-13	Flatten level = 10
	0	0	-1	-1	0	2	3	-9	0	-3	5	5	5	-3	
	0	-1	0	-5	-5	3	3	5	-3	0	5	0	5	0	
	-1	0	-5	-1	-5	3	2	0	-1	-1	0	0	5	0	
	-5	-5	-5	0	-5	2	2	-9	-13	-3	-1	5	5	0	

Fig. 2. Median difference for 4 blocks with flatten level.

The last layer of ConvNet model H is a fully connected. It uses softmax function for the prediction of regions. Now, the accuracy of the proposed system is measured by a confusion matrix. Here, the accuracy is a measure of the ratio of correctly predicted ROI and the total predicted ROI in the final output image Itest with localized ROI

$$Accuracy = \frac{Correct\ Prediction}{Total\ Prediction} \times 100.$$

Experimental Results

In this section, we analyze the performance of our proposed method. The models were written and trained in Python. We get an optimal block size with stride and kernel as 1 and 7 x 7, respectively for training the ConvNet model H. We also measure the quality of the noisy images by computing Mean Standard Error (MSE), Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The soul idea behind working in noisy images is to find out the regions in the noisy images for watermarking which are insensitive to human eyes. In Method 1, it is done by computing$D_i[j,k]$ i.e.,

$$D_i[j,k] = med - b_i[j,k].$$

Where as in the Method 2, for each block b_i we calculate. The is the ratio of median and standard deviation of each block i.e.,

$$D_i[j,k] = \alpha_i - b_i[j,k].$$

Table I. Number of region of interest for original image and for Gaussian noise image.

Methods	Original Image	Gaussian Noise (at mean=0, varying standard deviation)					
	ROI	0.100	0.141	0.173	0.200	0.223	0.243
M 1	554	494	494	501	477	500	513
M 2	731	721	720	723	732	727	726

The Table I represent a comparative analysis of the number of region of interest for original and for Gaussian noise using both the Methods (i.e., M 1 and M 2). It shows when we increases the noise levels for Gaussian noise, the computed ROI are nearly equal for M 1. In case of M 2, the computed ROIs are increases. The significant observation is that the M 2 for computing ROI is better than M 1 even in noisy environment.

Table Ii Comparison on Quality of Original and Noisy Images by MSE, SSIM and PSNR

Gaussian Noise with mean =0			
Noise	**MSE**	**SSIM**	**PSNR**
0.100	2.71	0.98	27.5831
0.141	2.71	0.98	27.5832
0.173	2.71	0.98	27.5831
0.200	2.72	0.98	27.5902
0.223	2.73	0.98	27.5812
0.243	2.74	0.98	27.5817

The Table II shows the quality of noisy images by measuring the parameter MSE, SSIM and PSNR for the Gaussian noise. These computed values represent the relation between different noise level and image degraded quality of noisy image. It ensures that the noise is added in such a way that its addition on the original image is not perceptible to the human eyes.

Conclusion

In this work, we proposed a machine learning based approach to localize the number of regions in the images for watermarking. The image is divided into the smaller blocks to compute a threshold, which is used to distinguish between ROI and RONI. Our approach employs ConvNet to identify the informative features from the blocks (i.e., ROI and RONI). The obtained features are used to train the model H. We evaluate the proposed approach on different images using various metrics, such as MSE, PSNR, and SSIM. The results showed that the proposed approach can find better ROIs than the existing approaches. As security is an essential aspect of the watermarking, the proposed approach also ensures the security of hidden information. Furthermore, we are working on optimizing the ROI computation such that the overall run time of the model can be reduced.

Acknowledgements

This work is supported by Science and Engineering Research Board (SERB) file number ECR/2017/002419, project entitled as A Robust Medical Image Forensics System for Smart Healthcare, and scheme Early Career Research Award.

References

[1] (2018) The Kvasir Dataset. http://datasets.simula.no/kvasir/.

[2] (2018) Hamlyn Centre Laparoscopic / Endoscopic Video Datasets. http://hamlyn.doc.ic.ac.uk/vision/.

[3] X. Li, B. Li, B. Yang, and T. Zeng, "General framework to histogram shifting-based reversible data hiding," IEEE Transactions on Image Processing, vol. 22, no. 6, pp. 2181–2191, June 2013.

[4] J. Tian, "Reversible data embedding using a difference expansion," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 8, pp. 890–896, Aug 2003.

[5] J. Raj.T and E. T. Sivadasan, "A survey paper on various reversible data hiding techniques in encrypted images," in 2015 IEEE International Advance Computing Conference (IACC), June 2015, pp. 1139–1143.

[6] X. Zhang, "Separable reversible data hiding in encrypted image," IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, pp. 826–832, April 2012.

[7] A. M. Alattar, “Reversible watermark using the difference expansion of a generalized integer transform,” IEEE Transactions on Image Processing, vol. 13, no. 8, pp. 1147–1156, Aug 2004.

[8] X. Li, J. Li, B. Li, and B. Yang, “High-fidelity reversible data hiding scheme based on pixel-value-ordering and prediction-error expansion,” Signal Processing, vol. 93, no. 1, pp. 198 – 205, 2013.

[9] L. Kamstra and H. J. A. M. Heijmans, “Reversible data embedding intoimages using wavelet techniques and sorting,” IEEE Transactions on Image Processing, vol. 14, no. 12, pp. 2082–2090, Dec 2005.

[10] M. Fallahpour and M. H. Sedaaghi, “High capacity lossless data hiding based on histogram modification,” IEICE Electronics Express, vol. 4, no. 7, pp. 205–210, 2007.

[11] X. ting Zeng, Z. Li, and L. di Ping, “Reversible data hiding scheme using reference pixel and multi-layer embedding,” AEU - International Journal of Electronics and Communications, vol. 66, no. 7, pp. 532 539,2012.

[12] S. Cai and X. Gui, “An efficient reversible data hiding scheme basedon reference pixel and block selection,” in 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia SignalProcessing, Oct 2013, pp. 571–574.

[13] A. stbiolu, G. Uluta, and B. stbiolu, “Tamper localization of the medical images based on fragile watermarking,” in 2017 25th Signal Processing and Communications Applications Conference (SIU), May 2017, pp. 1–4.

[14] H. Zhang and M. Gao, “A semi-fragile digital watermarking algorithm for 2d vector graphics tamper localization,” in 2009 International Conference on Multimedia Information Networking and Security, vol. 1, Nov 2009, pp. 549–552.

[15] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, and T. Tan, “Feedback convolutional neural network for visual localization and segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-based convolutional networks for accurate object detection and segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142–158, Jan 2016.

[17] J. H. Bappy and A. K. Roy-Chowdhury, “Cnn based region proposalsfor efficient object detection,” in 2016 IEEE International Conference on Image Processing (ICIP), Sept 2016, pp. 3658–3662.

[18] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, “Accurate object localization in remote sensing images based on convolutional neural networks,” IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 5, pp. 2486–2498, May 2017.