Implementing Semi-Blind Image Steganography with Improved Concealment

Karan Nair
K.J.Somaiya College of Engineering,
Vidyavihar, Mumbai

Karan Asher
K.J.Somaiya College of Engineering,
Vidyavihar, Mumbai

Jonathan Joshi
Eduvance,
Mumbai, India

ABSTRACT

There have been a number of proposed methods to conceal information using steganographic techniques.Images are the favoured carrier due to the large capacity for concealed information and relative ease to work with. A number of methods exist to conceal text as well as image data within images. This paper proposes an information embedding scheme with improved concealment of secret images within larger images using a fractional embedding scheme. It explores the natural redundancy of image data, as well as limitations of human perception and statistical attacks to provide better subjective and objective concealment.

General Terms

Steganography, security, image encryption

Keywords

Steganography, steganalysis, hiding, semi-blind, blind, image-in-image steganography

1. INTRODUCTION

Historically it has always been important for finding methods to conceal the content and existence of information. Whether it is the military plans of an army, the financial records of a company, or even personal pictures in mobile phones - ensuring the information cannot be accessed by any unauthorized persons is very important. While cryptography attempts to conceal the contents of a secret message, it still betrays the existence of a secret message. The obvious advantage of steganography is that it conceals the existence of the secret message, hiding it in seemingly harmless data. If metaphors are to be used, cryptography is like hiding a needle in a haystack, while steganography is akin to hiding a tree in a forest.In this paper, a payload embedding scheme for image in image steganography has been proposed, which utilises the statistical redundancies in images as well as limitations in human perception to improve the amount of data that can be hidden without compromising concealment.

2. HISTORY

Throughoutthe evolution of civilization, humans have alwaysaspired to more privacy and security for their communications [1]. One of the first documents describing Steganography comes from Histories by Herodotus, the Father of History. In this work, Herodotus gives us several cases of such activities. A man named Harpagus killed a hare and hid a message in its belly. Then, he sent the hare with a messenger who pretended to be a hunter [1] in order to convince his allies that it was time to begin a revolt against Medes and the Persians.

Pirate legends tell of the practice of tattooing important information, such as a map, on the head of some person, so that the hair would conceal it and keep it hidden. [2]

Kahn tells of a trick used in China of embedding a coded ideogram at a prearranged position in a message; a similar idea led to the grille system used in medieval Europe, where a wooden template would be placed over a seemingly innocuous text, highlighting an embedded secret message. [2]

During WWII the grille method or some variants were used by spies [2]. In the same period, the Germans developed microdot technology, which prints a clear, good quality photograph shrinking it to the size of a dot.This dot would then be concealed in the print of some larger image. To any person not aware of its existence, the dot would never gather any attention.

Null ciphers were also used, which can be considered a rudimentary form of text in text steganography [3]. Null ciphers effectively hide a secret message amongst the letters in a larger, innocuous message. The secret is recovered by reading the letters in a predetermined manner, like reading every 3rd letter and skipping the rest [3]

In the modern day, a majority of our information and messages are digital in nature. These messages often have to be transmitted over non-secure medium like internet, or stored on servers where several people may have access to it. Concealing digital messages is therefore of paramount importance.

3. DIGITAL STEGANOGRAPHY

In digital Steganography, electronic communications may include steganographic coding inside of a transport layer, such as a document, image, audio or video. Media files are ideal as steganographic carriers because of their large size, which allows much greater capacity for embedding information. As a simple example, a sender might start with an innocuous image file and adjust the color of every 100th pixel to correspond to a letter in the alphabet, a change so subtle that someone not specifically looking for it is unlikely to notice it. Text has too little redundancy and information to provide sufficient storage capacity. Video offers excellent storage capacity since it has large storage capacity, but is difficult and time consuming to work with. Images provide the middle ground.

A few terms that will be commonly used are as follows.

1) Payload –The message/data to be hidden.
2) Carrier –The innocuous data within which the payload is to be concealed. Carrier is selected so as to minimize suspicion.
3) Package – The output of the steganographic embedding process. The package consists of the carrier with the payload embedded.
4) Channel –It is the medium over which the package is sent. The channel is often not secure i.e. 3rd parties can eavesdrop on the content passing through the channel.

5) Key –Any information additional to the package that the recipient needs to successfully recover the payload. The key adds additional secrecy as it provides protection even in case existence of a secret message is revealed.

6) Steganalysis – Analysis of file suspected to contain concealed payload. Steganalysis uses statistical techniques to detect and reveal hidden data.

Digital Steganography can be broadly classified into 3 types

3.1 Non-Blind Steganography

In this method, the unaltered carrier needs to be transmitted along with the package. The amount of information to be transmitted is the most. The carrier is the key to recovering payload from the package. It is a very simple to understand system and can be implemented easily but is seldom used due to the large amount of overhead and low concealment.

3.2 Semi Blind Steganography

Semi Blind Steganography requires only some additional information besides the package. This reduces the amount of the data transmitted at the cost of increased complexity. The information is hidden in a particular manner which allows the decoding program to decipher it using a user- or payload-specific key which has to be transmitted along with the package. Semi-Blind is the preferred method as it offers key-based secrecy as well as reduced data overhead.

3.3 Blind Steganography

In Blind Steganography, only the package has to be transmitted and no additional information is required to decode it. Security in blind steganography relies solely on ensuring the existence of payload is not discovered.None of the steganographic systems that are known achieve perfect security [4], [5], and this means that they all leave hints of embedding in the package. This gives the steganalyst a useful way in to identifying whether a secret message exists or not.Jessica Fridrich [5], [6] suggests that “the ability to detect secret messages in images is related to the message length”. This statement is based on the logic that a small message embedded within a large carrier will result in a small percentage of manipulations, and therefore, it will be much harder to detect any artefacts and distortions within the package. Thus, in all the cases discussed below, the carrier used to hide the payload has to be significantly larger than the payload itself.

4.PROPOSED METHOD AND COMPARISON

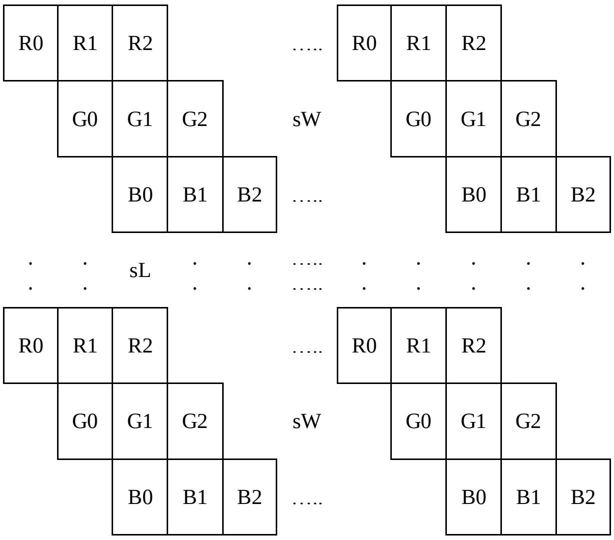
The proposed method deals with image in image embedding in a semi-blind fashion. The payload is spread out and embedded into the carrier based on size considerations of both the payload and package. The proposed method has been implemented in MATLAB for measuring performance parameters. A simple blind steganography implementation has been used to provide comparison.

4.1Semi-blind steganography with improved concealment

4.1.1Message hiding

Traditional LSB steganography restricts itself to the lowest 1 or 2 bits of a carrier to embed the payload. While this method does offer good concealment, the amount of data that can be embedded is reduced. The carrier needs to have 8 times more data than the payload for 1 bit steganography while it has to be 4 times as large if 2 bits are used (assuming 8bpp for carrier and payload).Instead of restricting the data embedding in number of bits, it is more useful to operate in base 10 and store the digits of the decimal expansion of the payload pixel value. The proposed scheme is as follows.

1. Clear the units place digit of the carrier pixel.
2. Represent payload pixel as its decimal expansion i.e. separate its digits.
3. Add digits of each payload pixel to the carrier pixels in the manner shown below. Each payload pixel requires 3 carrier pixels to be embedded. The pixels are spaced out by a spreading factor sW and sL, which act as keys to the system. The spreading factor should be as high as possible.



The R0, R1, R2 represent the units, tens and the hundredths digit respectively of the R pixel value and so is the same in case of G and B. Here since a colour image is considered, there are 3 planes available for embedding. The minimum value of sW has to be 3 while sL has to be 1, which means the carrier has to be 3 times as long and at least as broad as the payload.Since we are using 8bpp images, the maximum deviation that can occur in the carrier pixel is by a value of 9. This amounts to a change of less than 4% of the maximum value. Further, given that the statistical distribution of the units digit of the carrier will generally follow a Gaussian distribution, the average error between the carrier pixel value will average out to zero. However this does introduce some constraints on the nature of the carrier. The carrier should have most of its pixels having medium to high values (bright image). The carrier has to be sufficiently large in the horizontal direction (at least 3 times).

4.1.2 Payload recovery

In order to recover the payload, the carrier pixels are taken mod 10, and weighted sum of the resulting values are performed as-

$$X = X_0 + X_1 * 10 + X_2 * 100$$

Where X stands for R, G or B.

The spreading factors act as a key and ensure the correct pixels are recovered. If a wrong value of sL or sW is used, it will recover random values from the carrier pixels and thus payload will remain concealed.

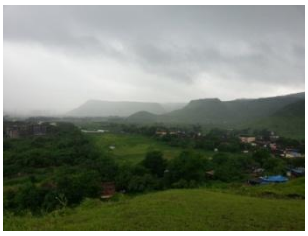


Fig.1(a) Nature (Carrier)

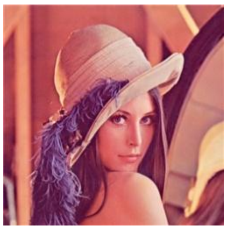


Fig.1(b) Lena (Payload)

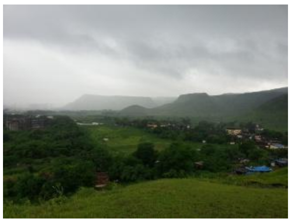


Fig.1(c) Nature (Package)

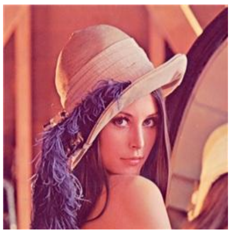


Fig 1(d) Lena (Recovered)

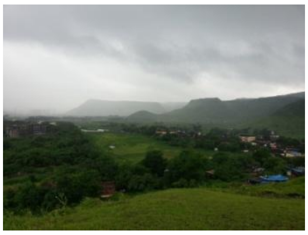


Fig.2(a) Nature (Carrier)

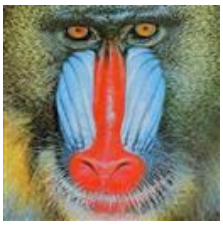


Fig.2(b) Mandrill (Payload)

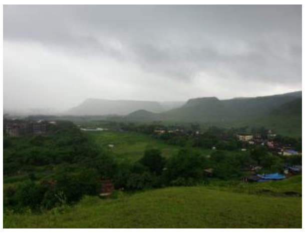


Fig.2(c) Nature (Package)

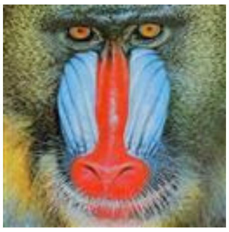


Fig.2(d) Mandrill (Recovered)

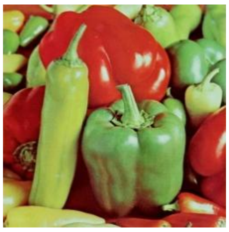


Fig.3(a) Peppers (Carrier)

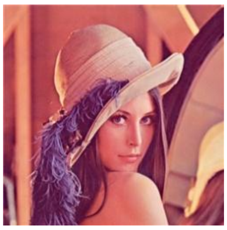


Fig.3(b) Lena (Payload)

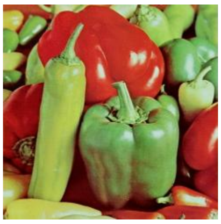


Fig.3(c) Peppers (Package)

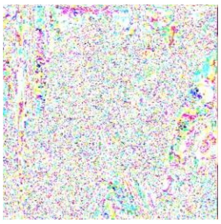


Fig.3(d) Lena (Recovered)

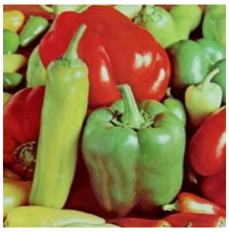


Fig.4(a) Peppers (Carrier)

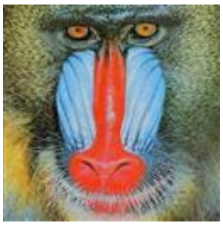


Fig.4(b) Mandrill (Payload)

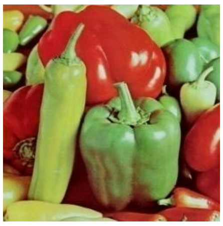


Fig.4(c) Peppers (Package)

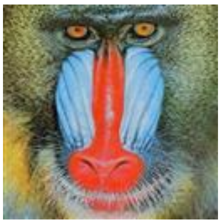


Fig.4(d) Mandrill (Recovered)

4.2 BLIND STEGANOGRAPHY

The drawback of using non blind and semi blind techniques is that both the final package and some additional information have to be transmitted over the channel and this increases the data transmitted. In blind Steganography, only the final package has to be transmitted, thus reducing the overhead. However this is at the cost of increased vulnerability as anyone who can access the package can potentially access the secret.

4.2.1 Message hiding

Blind steganography works similar to non-blind, except instead of XOR operation, the payload pixels simply replace the contents of the carrier bits. The method we used is as follows.

1. Clear the last 2 bits of the carrier pixel.
2. Store 2 bits at a time of the payload in the 2 LSBs of the carrier.

This method requires the carrier to be 4 times the size of the payload.

4.2.2Payload recovery

The principle used to recover the payload is to just read the 2 LSBs of 4 pixels together and convert them to corresponding decimal values using binary expansion i.e. multiplying the 2 bits of payload data in each pixel by an appropriate weight depending on its position in the original payload pixel. These weights are all powers of 2. Since no spreading is performed, there is no need for any additional information besides the package itself.

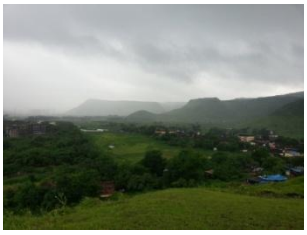


Fig.5(a) Nature (Carrier)

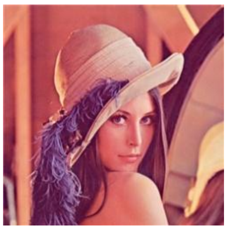


Fig.5(b) Lena (Payload)

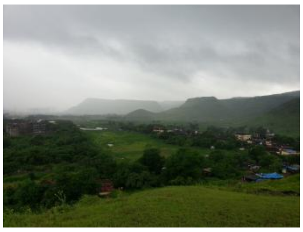


Fig.5(c) Nature (Package)

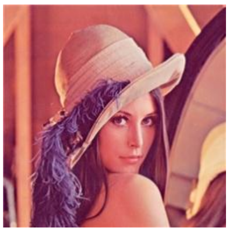


Fig 5(d) Lena (Recovered)

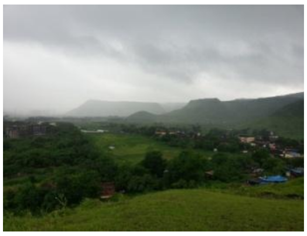


Fig.6(a) Nature (Carrier)

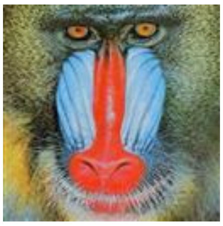


Fig.6(b) Mandrill (Payload)

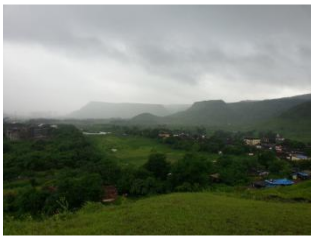


Fig.6(c) Nature (Package)

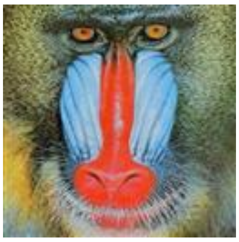


Fig.6(d) Mandrill (Recovered)

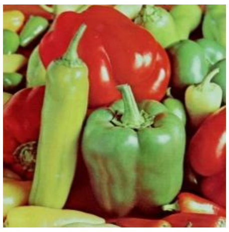


Fig.7(a) Peppers (Carrier)

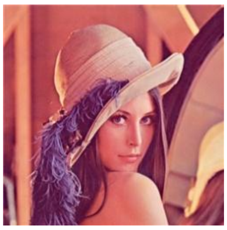


Fig.7(b) Lena (Payload)

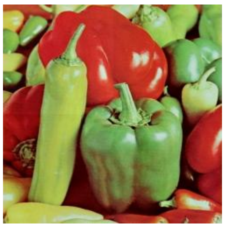


Fig.7(c) Peppers (Package)

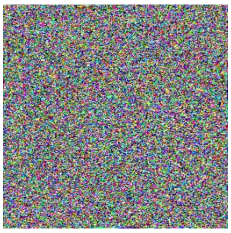


Fig.7(d) Lena (Recovered)

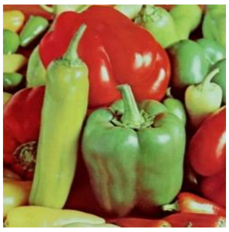


Fig.8(a) Peppers (Carrier)

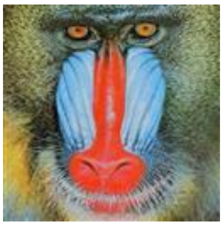


Fig.8(b) Mandrill (Payload)

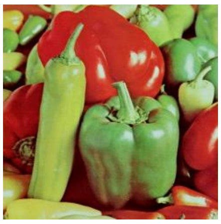


Fig.8(c) Peppers (Package)

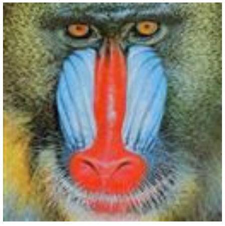


Fig.8(d) Mandrill (Recovered)

5. RESULTS &STATISTICAL ANALYSIS

The 3 techniques were performed using 2 different carriers and payloads. The image details were as follows,

Table 1. Details Of The Images Used

Image name	Type	Resolution
Nature	Carrier	3264x2448
Peppers	Carrier	512x512
Lena	Payload	256x256
Mandrill	Payload	128x128

The image size selection has been made to demonstrate the most extreme cases during image embedding-

i) When the carrier is much larger than the payload

ii) When the carrier is not sufficiently larger than the payload.

Image embedding and then recovery were done for each carrier-payload pair with both techniques. It is imperative to use lossless formats for storage and transmission of the package. As the payload is stored in the LSBs of the package, lossy methods would damage the concealed payload and make it irretrievable.As Ismail Avcibas et al [7] note, performance parameters for objective quality assessment can also be used for measuring steganographic security. Two performance metrics were used to determine the extent of Steganography. These performance metrics are also used for image quality measurements.Peak signal-to-noise ratio (PSNR) gives a measure of the Mean Square Error(MSE) [8]. Typical lossy compression algorithms produce values of PSNR in the range of 30-50 dB for an image with 8-bits per pixel. However values greater than 60 dB ensure the statistical differences between carrier and package is negligible. This is empirically considered an acceptable value as human eye finds it difficult to perceive any differences in the images. Higher PSNR values are desirable. The second metric used is the Structural Similarity index(SSIM) [9], which measures structural errors rather than simple pixel value errors. It has been observed that PSNR alone is not a good metric of image quality as 2 images with identical PSNR values can have major difference in their structure. SSIM is a statistical analysis of the structure of the 2 images and produces a value in the range(-1,1), where higher values are desirable.The methods discussed take one pixel from the payload and spread it across a number of pixels in the carrier. Each method has a minimum size ratio required between the carrier and payload. If the actual ratio is not greater or equal to this ratio, then the image cannot be recovered properly. For both techniques, only if the minimum size considerations are met, the images are recovered without any loss. Having the desired size ratios ensures multiple payload pixels are not embedded into the same carrier pixel. The spread calculations (sL, sW) are as follows-

Semi Blind,

$$sL = \frac{Lc}{Lp} \qquad sW = \frac{Wc - 3}{Wp}$$

Blind,

$$sL = \frac{Lc - 2}{Lp} \qquad sW = \frac{Wc - 2}{Wp}$$

WhereLc, Wc are the vertical and horizontal dimensions of the carrier, $Lp, Wp$are the vertical and horizontal dimensions of the payload and sL, sW are the spread parameters for the vertical and horizontal dimensions respectively.The minimum size ratio using these considerations is as follows,

Table 2. Minimum Size Ratio.

Technique	Horizontal ratio	Vertical Ratio
Semi Blind	3:1	1:1
Blind	2:1	2:1

The size ratios for the carrier-payload pairs used are as follows,

Table 3. Size Ratio (Semi Blind)

Carrier	Payload	Horizontal Ratio	Vertical Ratio
Nature	Lena	9.56:1	12.74:1
Nature	Mandrill	19.13:1	25.48:1
Peppers	Lena	2:1	1.99:1
Peppers	Mandrill	4:1	3.98:1

Table 4. Size Ratio (Blind)

Carrier	Payload	Horizontal Ratio	Vertical Ratio
Nature	Lena	9.55:1	12.74:1
Nature	Mandrill	19.11:1	25.48:1
Peppers	Lena	1.99:1	1.99:1
Peppers	Mandrill	3.98:1	3.98:1

As we can see, for the case of Peppers-Lena combination, the size ratio is not met. As a result, the payload is irrecoverably lost. In all other cases, the payload is recovered without any error.The PSNR and SSIM values for the 2 methods are as follows,

Table 5. PSNR And SSIM (Semi Blind)

Carrier	Payload	PSNR	SSIM
Nature	Lena	60.1538	0.9999
Nature	Mandrill	65.8754	0.9999
Peppers	Lena	46.6453	0.9873
Peppers	Mandrill	50.8876	0.9989

Table 6. PSNR And SSIM (Blind).

Carrier	Payload	PSNR	SSIM
Nature	Lena	70.691	0.9980
Nature	Mandrill	76.7153	0.9980
Peppers	Lena	41.982	0.9916
Peppers	Mandrill	59.3441	0.9954

The PSNR and SSIM values indicate that in terms of performance, the larger the carrier with respect to the payload, the better the PSNR and SSIM values. Between semi-blind and blind techniques, we see a trade-off. Blind produces better PSNR but loses out on SSIM. And while Semi Blind has lower PSNR values, its performance still lies well within acceptable limits. The major advantage in semi blind is that it produces better SSIM which means statistical differences between carrier and package are less, which provides better concealment against statistical attacks against the method.

6. CONCLUSION

The techniques discussed in this paper achieve distortion-less payload embedding and recovery into the carrier, provided the carrier satisfies the minimum size requirements. The Blind technique has the higher size requirements, with a 4:1 pixel count ratio. The proposed semi-blindtechnique achieves its results with a 3:1 ratio, which gives higher capacity for storing the payload. Detectability of the payload in the package is lower in blind based on visual differences alone. However semi-blind technique produces greater statistical similarity which gives us better resistance to statistical attacks. All this indicates that the semi-blind technique proposed in this paper, using fractional embedding of the payload in the carrier, produces 33% higher efficiency of embedding while retaining near similar PSNR and better SSIM characteristics as compared to prevailing blind embedding techniques.

7. REFERENCES

[1] Bruce Norman, Secret warfare, the battle of Codes and Ciphers. Acropolis Books Inc., first edition, 1980. ISBN 0-87491-600-3.

[2] Neil F. Johnson, Sushil Jajodia, George Mason University, *"Exploring Steganography: Seeing the Unseen"*, IEEE Computers, February 1998, pp. 26-34.

[3] A. Joseph Raphael, Dr.V Sundaram, “Cryptography and Steganography – A Survey’’ , Int. Journal of Comp. Tech. Appl., Vol 2 (3), 626-630.

[4] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, "Digital Watermarking and Steganography (Second Edition)", Morgan Kaufmann Publishers, ISBN: 978-0-12-372585-1, 2007.

[5] Philip Bateman, Dr. Hans Georg Schaathun,“Image Steganography and Steganalysis”, Department of Computing, University of Surrey, 2008.

[6] J. Fridrich, M. Goljan, and D. Hogea. "Attacking the OutGuess", Proceedings of the 3rd Information Hiding Workshop on Multimedia and Security 2002, Juan-les-Pins, France, 2002.

[7] Ismail Avcıbas¸Nasir Memon and Bülent Sankur “Steganalysis Using Image Quality Metrics”, IEEE Transactions on Image processing Vol.12

[8] Huynh-Thu, Q. ;Psytechnics Ltd., Ipswich ; Ghanbari, M. “Scope of validity of PSNR in image/video quality assessment”, IEEE xplore.

[9] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli “The SSIM Index for Image QualityAssessment”,http://www.cns.nyu.edu/lcv/ssim/