Text Hiding in AVI Video

Assistant Lecturer: Lahieb M. Al-Yassiry
College of Information Engineering
Nahrain University
Baghdad-Iraq

Assistant Lecturer: Hamsa A. Abdullah
College of Information Engineering
Nahrain University
Baghdad-Iraq

Abstract

Steganography comes from the Greek word "steganos", roughly translated to "Covered Writing". Steganographic techniques allow one party to communicate information to another without a third party even knowing that the communication is occurring. The ways to deliver these "secret messages" is important. This paper is intended to introduce the hiding text (data) inside video as a means to communicate secret messages in away that they will not be read but by the desired person. The methods that used to hide text are the 4 Least Significant Bits (4-LSB) and Discrete Wavelet Transform (DWT), which will be hided in one color of video frames (i.e. red color).

Keywords: *Video Hiding, 4 Least Significant Bits, Steganography, AVI file, Text Hiding, Video Data Stream, Discrete Wavelet Transform.*

الخلاصة:

ان اصل كلمة "ستيكانوجرافي" هو اغريقي مستل من كلمة ستيكانوز، والترجمة الحرفية لها تعني الكتابة المغطاة. ان تقنية ستيكانوجرافيك تسمح للشخص بالاتصال بشخص اخر بدون معرفة اي طرف ثالث بهذا الاتصال. ان طرق تسليم الرسائل السرية مهمة جدا". هذا البحث تقدم "تقنية اخفاء المعلومات" في مقاطع الفيديو بطريقة لا تمكن اي شخص من قراءة الرسائل السرية ما عدا الاشخاص المخولين بذلك. الطرق المستخدمة لاخفاء الرسائل السرية هي استخدام المقاطع الرقمية الاقل اهمية (4-LSB) وتحويل المويجة المتقطعة (DWT). حيث يتم الاخفاء في لون واحد من الوان المقطع في الفيديو (اللون الاحمر) على سبيل المثال.

1.Introduction

In today's society the most practical implementation of Steganography is used in the world of computers. Data is the heart of computer communication and over the year a lot of methods have been created to accomplish the goal of using Steganography to hide data. The trick is to embed the hidden object into significantly larger objects, so the change is undetectable by the human eye. The best object up to this writing is probably a digital image. Digital images have the benefit of containing massive amounts of bytes to designate pixel color for the photo. For this reason this paper relate with video to hide information, because video is a sequence of frames and each frame is an image. Then we will have many images on the same file, which increase the size of hidden information and make it more complex than on image embedded data.

Steganography techniques are often used by copyright holders who wish to combat piracy and theft. Images, video and music can be encoded with information that can be to identify the work as being the property of an individual corporation. These encoding are watermarking. Watermarked media can be distributed on the internet while allowing copyright holders to be able to maintain their intellectual property [1].

Steganography works by replacing bits of useless or unused data in regular computer files (such as graphics, sound, text, HTML, or even Floppy disks) with bits of different invisible information. This hidden information can be plain text, cipher text, or even images and sound wave [1].

Since a video can be viewed as sequences of still images, video watermarking is an extension of image watermarking. The applications for still image watermarking can thus be executed to video watermarking by embedding data in single frame [2]. Compared with still images video watermarking present a much higher capacity or bandwidth. At the same time the computational complexity in video watermarking is higher due to the amount of data that need to be processed [2].

2. Video Data Stream

The video sequence composed of video sequence header and many Group Of Pictures (GOP) as illustrated in Figure.1. The video sequence header defines the video format, picture dimension, aspect radio, frame rate, and delivered data rate. A suggested buffer size for video sequence is also specified. GOP contains pictures that may be encoded into one of 3 supported compression format. The GOP header contains starting time for the group and it can be used as a point of random access. Each frame within GOP is numbered [3].

Each frame is consisting of 3 color plan (RGB). This research takes one color of frame to hide the text (data) using 4LSB method.

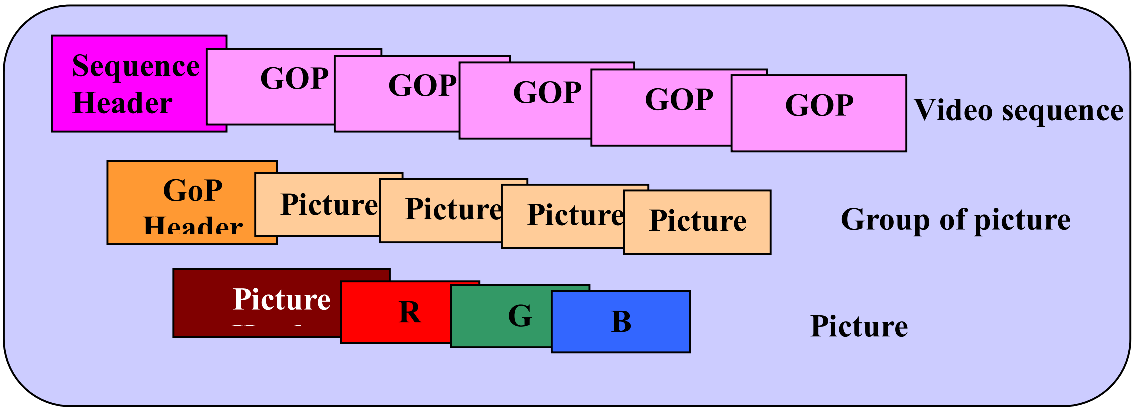


Figure 1: Video Data Stream

3. Text Hiding in LSB

Least Significant bit (LBS) insertion is a common, simple approach to embedding information in a cover image. The least significant bit (the 8th bit) of some or all of the bytes inside an image is changed to a bit of the secret message. When using a 24-bit image, a bit of each of the Red, Green, and Blue color components can be used, since they are each represented by a byte. In other words, one can store 3 bits in each pixel. An 800*600 pixel image, can thus store a total amount of 1,440,000 bit or 180,000 bytes of embedded data [4]. For example, a grid for 3 pixels of a 24-bit image can be as follows:

(00101101 00011100 11011100)

(10100110 11000100 00001100)

(11010010 10101101 01100011)

When the number 200, which binary representation is 11001000, is embedded into the LSB's of this part of the image, the resulting grid is as follows:

(0010110**1** 0001110**1** 1101110**0**)

(1010011**0** 1100010**1** 0000110**0**)

(1101001**0** 1010110**0** 01100011)

Although the number was embedded into the first 8 bytes of the grid, only the 3 underlined bits needed to be changed according to the embedded message. On average, only half of the bits in an image will need to be modified to hide a secret message using the maximum cover video [4].

In this work, the message may be a few thousand bits (8 bits per text character) embedded in million of other bits. Block diagram in Figure 2 describe the general steps of concealing text in container video and reconstructing text hiding text file.

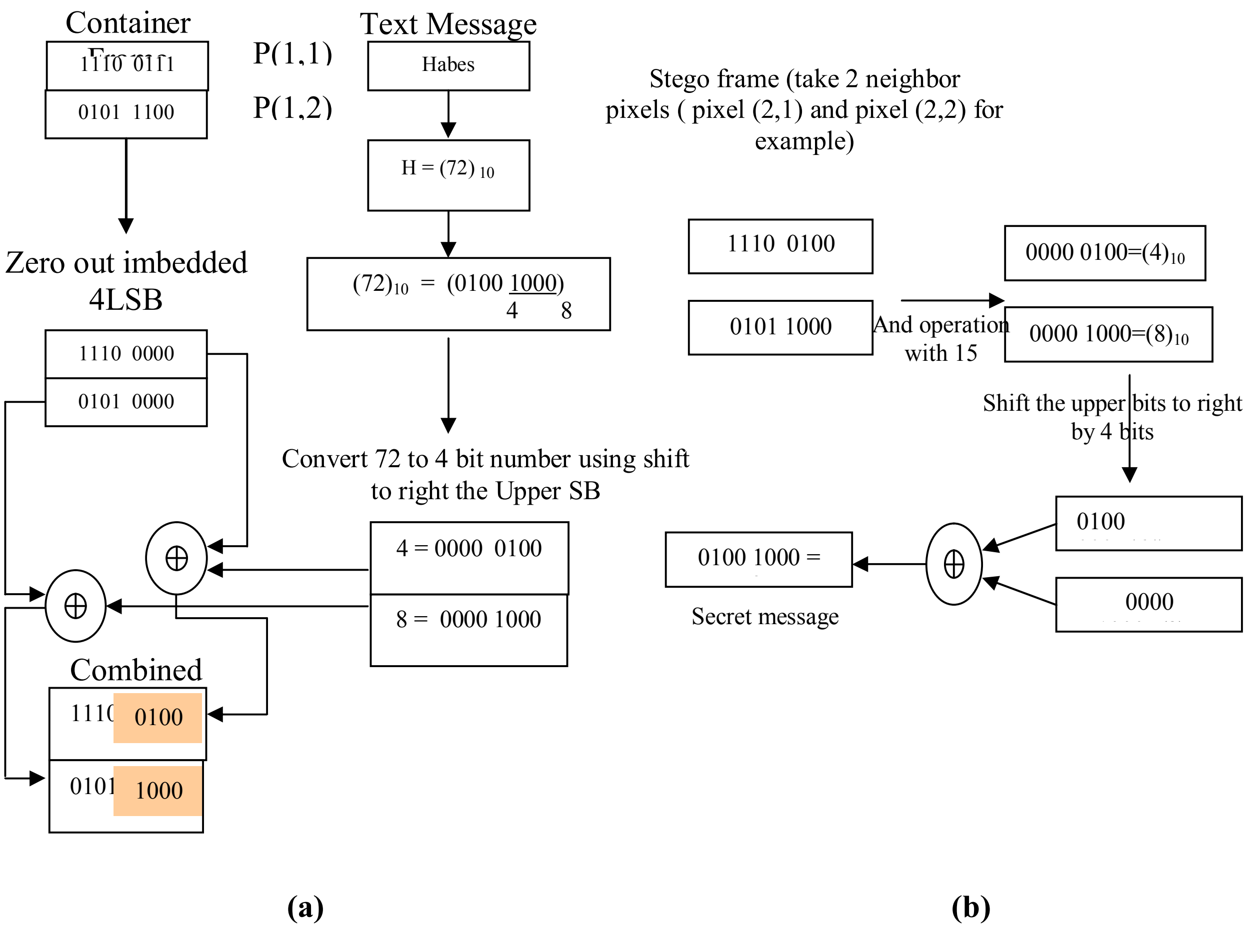


Figure 2: (a) Block Diagram of Text Hiding, (b) Block Diagram of Reconstructed Hiding Text File

3.1 Preparing Container Video

The following steps illustrate procedure of preparing video to hide text message in the LSB of one color of the video sequence frames.

1. Separate each frame of video sequence to its original color (RGB).
2. Convert one of the colored plans to stream of binary bits.
3. Use two adjacent pixels to hide one character.

3.2 Preparing Secret Text

To prepare secret text must follow the following:

1. Convert each character to decimal number.
   i. Take the 4 LSB alone, by performing **AND** operation
   ii. Take the 4 upper significant bits alone; by perform shift operation by 4.
2. Add the secret message to the cover video frames by applying **OR** operation.

Figure 2 show a block diagram to hide each character of the secret message. Two pixels are needed to hide one character. So the number of characters that can be hide in (n × m)*30 video sequence is given by the following equation.

Number of Character < = ((n × m)*30)/2-n … (1)

4. Text Hiding in Wavelet Image

Wavelet transform (WT) represents an image as a sum of wavelet functions (wavelets) with different locations and scales. Any decomposition of an image into wavelets involves a pair of waveforms: one to represent the high frequencies corresponding to the detailed parts of an Image (wavelet function) and one for the low frequencies or smooth parts of an image (scaling function).

The Wavelet transform is a technique for analyzing signals. It was developed as an alternative to the short time Fourier Transform (STFT) to overcome problems related to its frequency and time resolution properties.

DWT is a special case of the WT that provides a compact representation of a signal in spatial and frequency that can be computed efficiently [5]. One of the big discoveries for wavelet analysis was that perfect reconstruction filter banks could be formed using the coefficient sequences $a_L(k)$ and $a_H(k)$. The input sequence x is convolved with high-pass (HPF) and low-pass (LPF) filters $a_H(k)$ and $a_L(k)$ and each result is down sampled by two, constituting one level of transform, yielding the transform signals x_H and x_L. Multiple levels or "scales" of the wavelet transform are made by repeating the filtering and decimation process on the low-pass branch only. The signal is reconstructed through up sampling and convolution with high and low synthesis filters $s_H(k)$ and $s_L(k)$.

For properly designed filters, the signal x is reconstructed exactly $(y = x)$. The new twist that wavelets bring to filter banks is connection between multi-resolution analysis (that, in principle, can be performed on the original, continuous signal) and digital signal processing performed on discrete, sampled signals.

DWT for an image as a 2-D signal can be derived from 1-D DWT. The easiest way for obtaining scaling and wavelet function for two dimensions is by multiplying two 1-D functions. The scaling function for 2-D DWT can be obtained by multiplying two 1-D scaling functions: $\phi(x,y)=\phi(x)\phi(y)$. Wavelet functions for 2-D DWT can be obtained by multiplying two wavelet functions or wavelet and scaling function for 1-D analysis. For the 2-D case, there exist three wavelet functions that scan details in horizontal $\psi^{(I)}(x,y)=\phi(x)\psi(y)$, vertical $\psi^{(II)}(x,y)=\psi(x)\phi(y)$, and diagonal directions $\psi^{(III)}(x,y)=\psi(x)\psi(y)$. This may be represented as a four-channel perfect reconstruction filter bank as shown in Figure 3.a. Now, each filter is 2-D with the subscript indicating the type of filter (HPF or LPF) for separable horizontal and vertical components.

The resulting four transform components consist of all possible combinations of high- and low-pass filtering in the two directions. By using these filters in one stage, an image can be decomposed into four bands. There are three types of detail images for each resolution: horizontal (HL), vertical (LH), and diagonal (HH). The operations can be repeated on the low–low band using the second stage of identical filter bank.

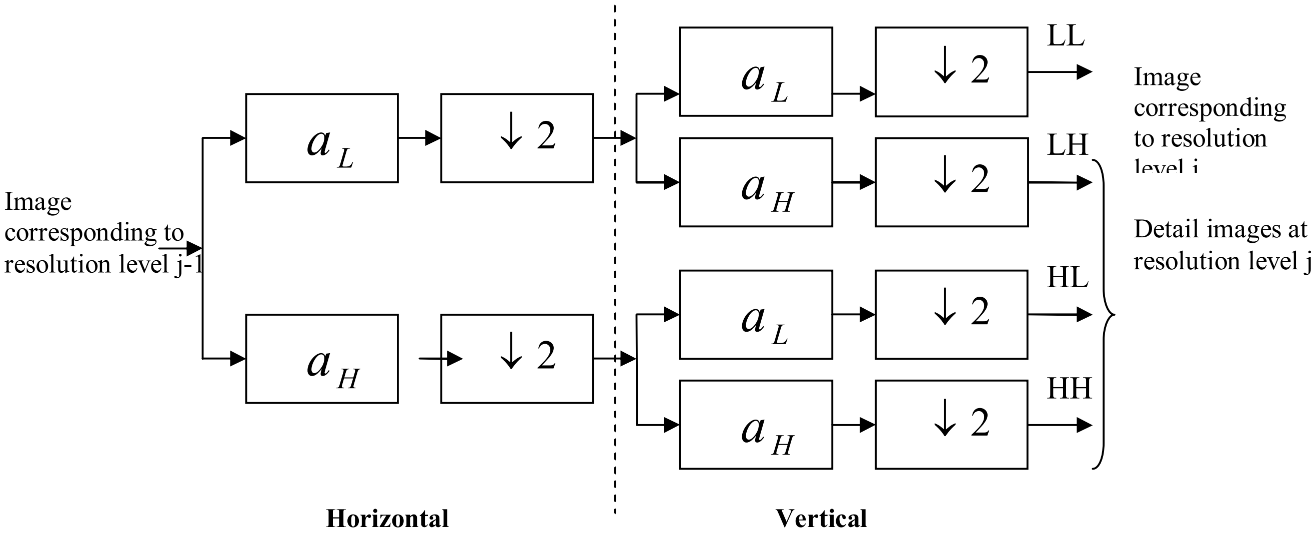
Thus, a typical 2-D DWT, used in image compression, will generate the hierarchical pyramidal structure shown in Figure 3.b. Here, we adopt the term “number of decompositions” (J) to describe the number of 2-D filter stages used in image decomposition. The DWT provides high spatial resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies.

5. The Basic Parameters

5.1 Applying Text into the Forward Wavelet Transform

On image RGB bands, two-dimensional wavelet is applied; the DWT decomposes the image into four sub bands, denoted LL (band), LH, HL, and HH (details). Convert each character to decimal number and these decimal number replaced by the coefficient in the LH, HL, and HH in red blain. The numbers of characters that can be hide in (n × m)*30 video sequences are given by the following equation.

Number of Character < = ((n × m)/4)*3)*30 … (2)



(a)

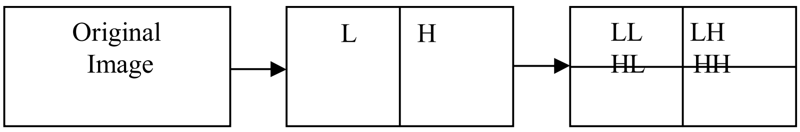


Figure 3: (a) One Filter Stage in 2-D DWT, (b) Pyramidal Structure of a Wavelet

6. Results

The original video sequence that used is uncompressed video signal operate at 30 frame/sec, consisting RGB color frame (24 bit video) with frame size (176×144). The size of video frames is very important because it specified with frame rate of the text size that need to be hide, as shown in eq.1 & eq.2. The performance of Steganographic video is measured by using MSE and PSNR. The result shows the performance of Stego. Frame is very good compared with the original frames. PSNR and MSE of each frame is calculated when the result will be described in Figure 4 and Figure 5 respectively.

The drop in PSNR was obtained by subtracting the PSNR of each frame in the error sequence from the PSNR of each frame in the error free sequence. While table 1 and 2 show the values of PSNR and MSE respectively for each frame. Finally, Figure 6 describes the original frames and the reconstructed frame of 4LSB and DWT methods.

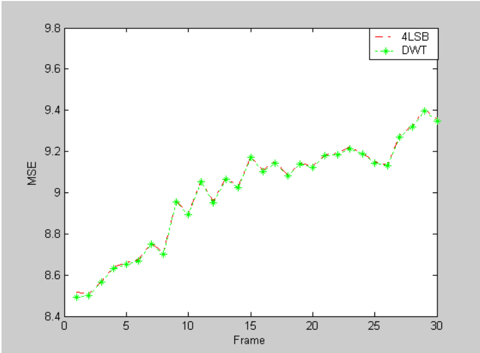


Figure 4: PSNR of 4LSB & D

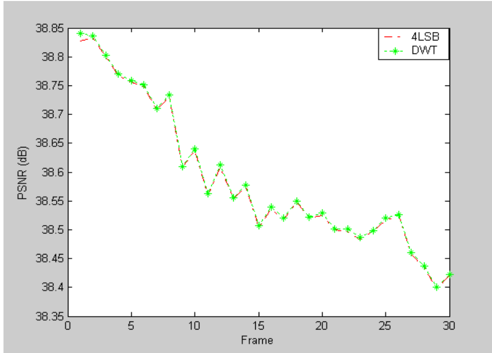


Figure 5: MSE of 4LSB & DWT

Table 1: PSNR of 4LSB & DWT

NO.	PSNR 4LSB	PSNR DWT
1.	38.828	38.841
2.	38.833	38.836
3.	38.8	38.803
4.	38.767	38.77
5.	38.756	38.759
6.	38.748	38.751
7.	38.708	38.711
8.	38.731	38.734
9.	38.608	38.61
10.	38.637	38.64
11.	38.56	38.562
12.	38.608	38.612
13.	38.554	38.556
14.	38.575	38.578
15.	38.504	38.507
16.	38.536	38.54
17.	38.517	38.52
18.	38.548	38.55
19.	38.52	38.522
20.	38.525	38.529
21.	38.5	38.502
22.	38.497	38.501
23.	38.483	38.487
24.	38.498	38.499
25.	38.516	38.52
26.	38.524	38.526
27.	38.457	38.46
28.	38.434	38.437
29.	38.397	38.401
30.	38.421	38.423

Table 2: MSE of 4LSB & DWT

NO.	MSE 4LSB	MSE DWT
1.	8.5171	8.4911
2.	8.5066	8.5018
3.	8.5716	8.5658
4.	8.6375	8.6312
5.	8.6594	8.6537
6.	8.6747	8.6698
7.	8.7559	8.7494
8.	8.7087	8.7029
9.	8.9594	8.9546
10.	8.9	8.8931
11.	9.0588	9.0542
12.	8.96	8.9509
13.	9.0718	9.0668
14.	9.0276	9.0224
15.	9.1756	9.171
16.	9.1101	9.1011
17.	9.1502	9.1428
18.	9.0841	9.0797
19.	9.1431	9.138
20.	9.1326	9.1234
21.	9.1855	9.1803
22.	9.1917	9.1836
23.	9.2221	9.2132
24.	9.1901	9.1864
25.	9.1505	9.1431
26.	9.1352	9.1307
27.	9.2771	9.2696
28.	9.3257	9.3196
29.	9.4062	9.3971
30.	9.354	9.3491

6.1 AVI file

In the first step, the system will load the header of AVI file. The Audio/Video Interleaved (AVI) file format is used with application that capture, edit and playback audio/video sequences. AVI files contain multiple streams of different types of data. Most AVI sequences use both audio and video streams. A simple variation for an AVI sequence uses video data and does not require an audio stream.

6.2 Bit Depth

Bit Depth measures how many different color shades which can be shown at the same time. The higher the bit depth, the larger is the video clip size. Normally a color depth of 16 bits is used for video clips, but today it is more common that modern codec use 24 bit color depth. The color depth that is used in this paper is 24 bit.

6.3 Objective Quality Measure

Objective video evaluation techniques are mathematical models that approximate results of subjective quality assessment, but are based on criteria and metrics that can be measured objectively and automatically evaluated by a computer program. Objective methods are classified based on the availability of the original video signal, which is considered to be of high quality (generally not compressed).

The most traditional ways of evaluating quality of digital video processing system are calculation of the signal-to-noise ratio (SNR), mean square error (MSE), and peak signal-to-noise ratio (PSNR) between the original video signal and signal passed through this system. The mean square error is found by taking the error squared divided by the total number of pixels in the frame. The PSNR is calculated by using eq.3, and the MSE is calculated by using eq.4. However, PSNR values do not perfectly correlate with a perceived visual quality due to non-linear behavior of human visual system [6]. PSNR of an image is a measure of the distortion of an image relative to a reference image. It can be used to measure the distortion of an image or frame due to transmission errors, compared with the original, error-free image or frame. PSNR is defined as:

$$\mathbf{PSNR\ (dB) = 10\ log_{10} \frac{(2n-1)^2}{MSE}} \quad \ldots \quad (3)$$

Where n is the number of bits per image sample and MSE is the Mean Squared Error between the distorted frame and the original frame. The mean square error is found by taking the error squared divided by the total number of pixels in the frame as shown bellow:

$$\mathbf{MSE = \frac{1}{W \times H} \sum_{r=0}^{H-1} \sum_{c=0}^{W-1} (I(r,c) - \hat{I}(r,c))^2} \quad \ldots \quad (4)$$

Where I(r,c) is the pixel value of the original frame at the (r,c) location, Î(r,c) is the value pixel of the Steganographic frame at the same location (r,c), H is the number of line, and W is the number of pixel per line.

7. Conclusion

The total number of hidden characters that hide in one second according to implementing of eq.1and eq.2 is **379984** for 4LSB method, while the number of hidden characters in DWT method is **570240**. The size of storage in 4LSB is **50%** of original size of video when the size of storage in DWT is **75%**. From these results the DWT is better than 4LSB to hide the character with out effect and distorted the original video. The processing time with the same (plaintext and video) using 4LSB is longer than the process of DWT where the time of the first method is **838** msec, and the time of second method is **60** msec.

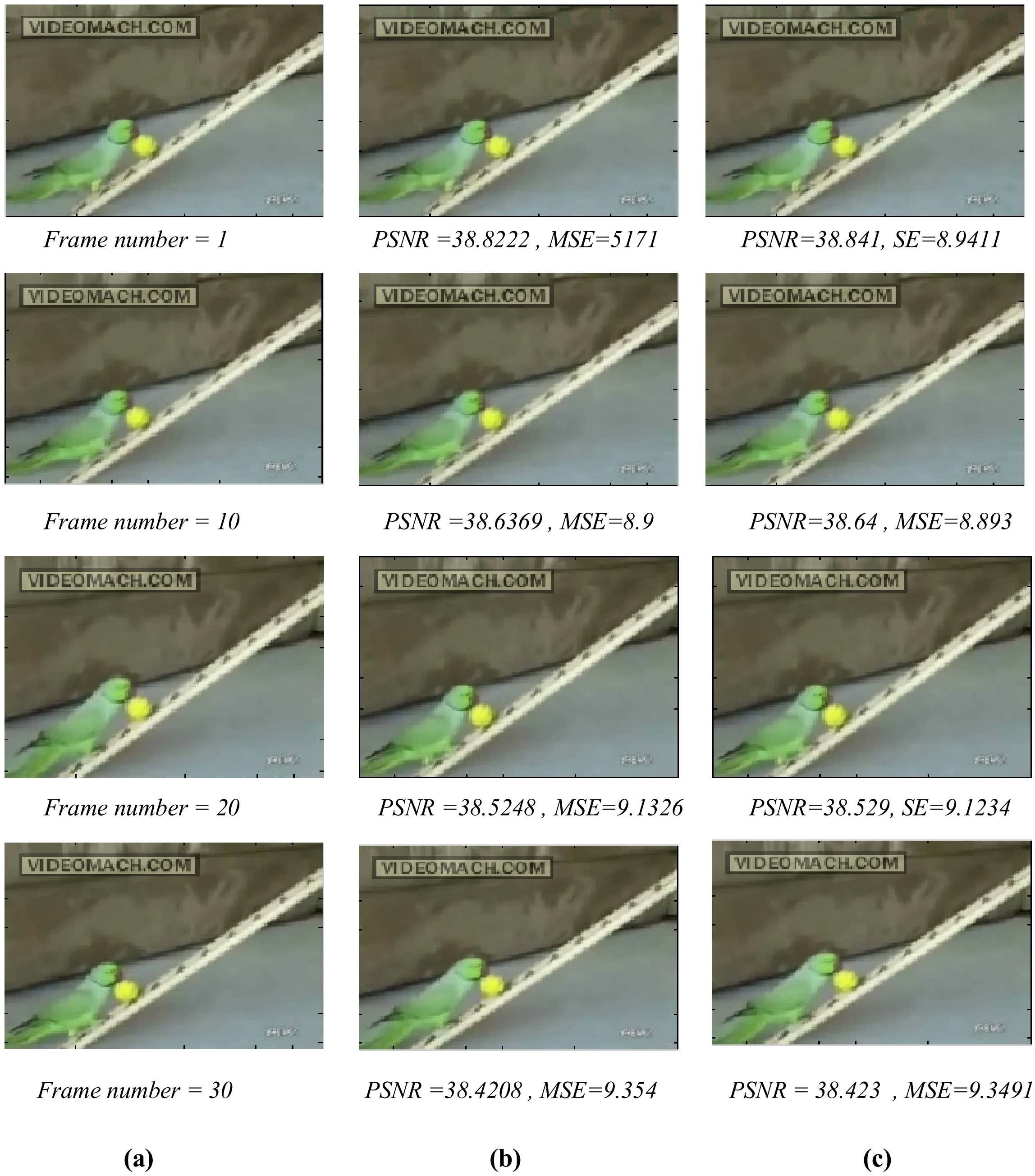


Figure 6: The Original and Steganography frames

(a) Original frames number 1, 10, 20, and 30, (b) Steganography frames number 1, 10, 20, and 30 using 4LSB, (c) Steganography frames number 1, 10, 20, and 30 using DWT

REFRENCES

[1] Alkhraisat Habes, "Information Hiding IN BMP Image Implementation, analysis and evaluation" ,information transmission in computer networks, 2006.

[2] Panos Liatsis," Recent Trends in Multimedia information Processing: Proceedings of 9th International workshop on system, signals, and image processing", world Scientific, 2003.

[3] R. Westwater, & B.Furht, " Real Time Video Compression Techniques and Algorithms", Kluwer acadimic published, 1997.

[4] T. Morkel, J.H.P. Eloff, M.S. Oliver, "An overview of image steganography", proceeding of fifth annual information security south africa conference, ISSA, PP:1-11 (Published Electronically), 2005.

[5] Panrong Xiao P.,Image Compression By Wavelet Transform, MSc. thesis, Computer Science, East Tennessee and state University, August 2001.

[6] I.E.G.Richardson, "H. 264 and MPEG-4 Video Compression", John Wiley and Sons, 2003.

[7] J. J. Chane and B.S. Manjunath, "Data Hiding in Video", 2000.

[8] Knlghtord, "Steganography: the world of data Hiding", 2005.