UDC 651.928:654.17:517.4
Original scientific paper
Received: 24.06.2007.

Coding of information into video file

Blanca E. Carvajal-Gámez, **Marco Antonio Acevedo-Mosqueda** and **Jose Luis López-Bonilla**
Seccion de Estudios de Postgrado e Investigation, Escuela Superior de Ingenieria Mecanica y Electrica,
Instituto Politécnico Nacional, Edif. Z-4, 3er. Piso, Col. Lindavista, C.P. 07738, MEXICO D.F.
e-mail: becgamez@yahoo.com.mx; macevedo@ipn.mx; jlopezb@ipn.mx

SUMMARY

In this work we propose an algorithm for hiding information in an avi file. The avi file is splitted into its audio and video components. The audio is saved in a file and its parameters are not changed. In addition, the video is splitted into frames. The frame contains 24 images which are transformed by Discrete Wavelet Transform (DWT). By appliying the DWT we obtain four sub images each one with different band width. Once we have the frequencies splitted, places are chosen for hiding information. In this work the concealment is made inside another avi file. However, the procedure can be applied for any kind of information. The file to be hidden is splitted into audio and video. Both are concealed independently in the sub images from the original video. The video to be hidden is separated into its frequency components and only the lower ones are saved, while the others are deleted. The compressed version of the image to be concealed is inserted into the video's sub images. Once the concealment has been done the Inverse Discrete Wavelet Transform (IDWT) is applied and the video is reconstructed again. To get the hidden information the same insertion process is made, and it applies the IDWT in the recover file.

Key words: *information, wavelets, video, hiding, avi file.*

1. INTRODUCTION

The concealment of information in digital formats is a subject which is very common nowadays, due to the necessity of espionage in several communications media. A very adequate alternative for protecting the information is through the steganography, which is capable of concealing the information without perceiving its existence. In this research an avi file is hidden inside another of the same format. The form of hiding the information must provide the security that a non-authorized person can extract it. The video files are decomposed into audio and images. The audio components from both files are manipulated separately and are preserved without alteration. However, the video is broken into its 24 frames in which the Discrete Wavelet Transform (DWT) is applied for getting its different sub images. The process of hiding the original and the video is different as it will be shown down.

The proposed method is able to hide the information from a video (*Video to hide*) inside another one (*Base Video*). The first part is focused to describe the composition of the images via the DWT. Next, the process for descomposing the video into frames is described with the proposed algorithm which takes into account the places for doing the concealment. This paper refers to the recuperation process for the inserted information. Finally, several results are shown with the tests applied to the modified images and the recovered ones.

2. SPLITTING THE VIDEO INTO FRAMES

2.1 Description of the research

There are three steps in this research. The first one is based on the procedure made in the *Base Video* which is the file where the information will be hidden. The second step shows the manipulations for the video to be concealed, where the modified Video is gotten with the *Base Video* and the *Video to Conceal*. Finally, the third step shows the process of decomposing the *Modified Video* to reach the recovered one. These steps are shown in Figure 1.

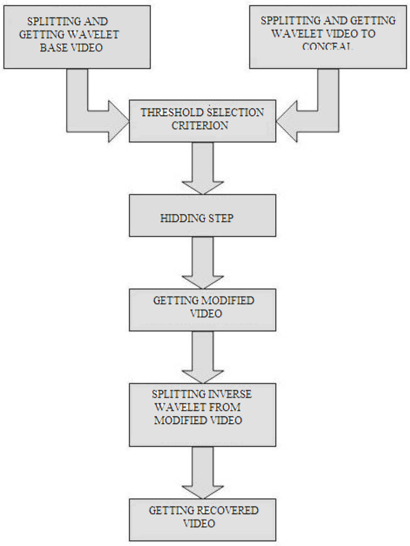

Fig. 1 Method for hiding and recovering the video

As it was mentioned, the first step is to obtain the audio components and frames from the avi files. The audio component from the *Base Video* is not altered and is kept apart from the process, while the audio component of the *Video to Conceal* is not processed and is hidden in the sub images of the *Base Video*.

The first step is referred to the getting the 24 frames per second of the *Base Video*. Each frame is splitted into its RGB components (R-Red, G-Green, B-Blue); the size of each matrix is $n \times m$. In each RGB component the DWT is applied to get four characteristic sub matrices: Approximations, Vertical Details, Horizontal Details and Diagonal Details, as it is shown in Figures 2a, 2b, 2c and 2d, respectively.

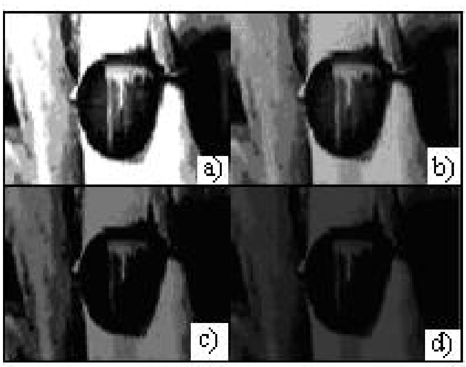

Fig. 2 RGB components in sub matrices for Base Video

Each sub matrix is sized $n/2 \times m/2$ because only one decomposition was done. The procedure is performed to the 24 frames in the *Base Video*.

In the second step the same procedure is performed in the *Video to Conceal*, with the difference that the image is decomposed twice. Figures 3a, 3b, 3c and 3d show the decomposition of the *Video to Conceal*. The sub matrices are sized $n/4 \times m/4$ because of two decompositions. The sub matrices to be hidden are Approximation (3a) and Horizontal Details (3b), into the sub matrices of Approximations (2a) and Horizontal Details (2b).

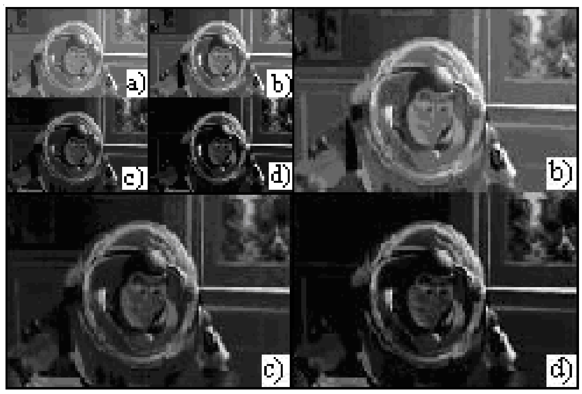

Fig. 3 RGB components in sub matrices for Video to Conceal

In the concealment step, certain places in the sub images of Approximations and Horizontal Details of the *Base Video* are identified, in which the sub images of the *Video to Conceal* will be hidden. The sub matrices of Vertical and Diagonal Details of the *Base Video* will not be altered. In the other way the sub matrices of Vertical and Diagonal Details of the *Video to Conceal* will be rejected. Once the sub matrices were concealed, the IDWT is performed to get the Modified Video.

In the *Modified Video* we add the Audio Component to get the avi file. This new file can be transmitted and/ or saved. In the transmission case the Modified Video needs the same band width for the difference to the *Base Video* if it was to be transmitted.

The last step is the recovering of the information. We perform the same procedure for the *Modified Video* for the difference to the *Base Video* for getting four sub matrices with the wavelet splitting. With the sub matrices of Approximations and Horizontal Details the information is obtained for reaching the sub matrix of Approximations and Horizontal Details, as shown in Figure 4. It can be seen that to obtain the Recovered Video, the sub matrices above mentioned are just needed, while the anothers have their values in zero.

The process with this algorithm is performed to each frame of the 24 per second that composes the video file.

In the concealment and recovering process, we handle the images in their RGB components. However, to reconstruct the avi format all images are stored in their original jpeg format [6].

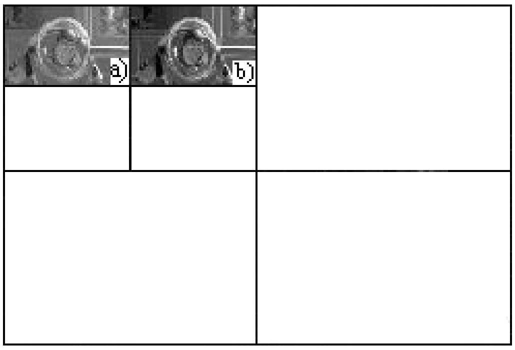

Fig. 4 RGB components once sub matrices for Recovering Video

2.2 Splitting the video into frames

From the *Base Video* and the *Video to Conceal* all their frames are taken. First the frames are splitted into its RGB components. Each one of the frames is in jpeg format [6]. You can consider then $x[n,m]$ as a tridimensional matrix which contains the RGB components, Figure 5.

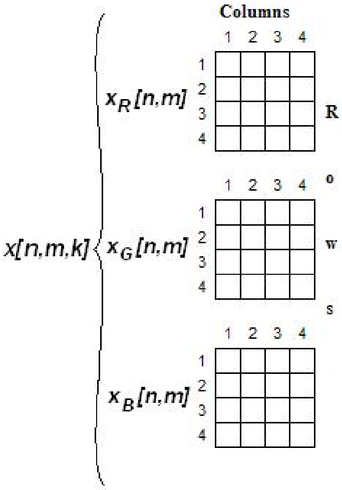

Fig. 5 RGB components

Once the RGB matrices are gotten, the information is managed from each pixel that composes each frame. Figures 6 and 7 show the splitting into frames of the *Base Video* and *Video to Conceal*, respectively.

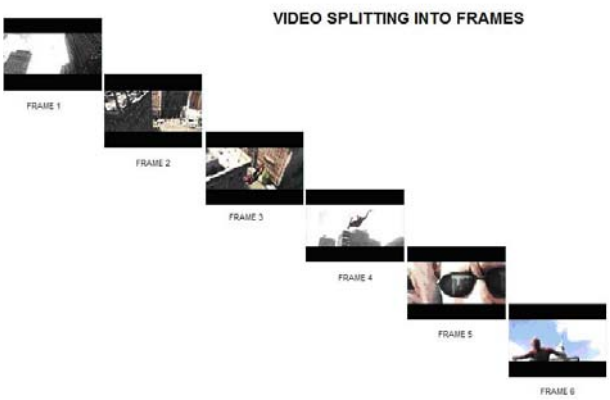

Fig. 6 Base Video splitting in frames

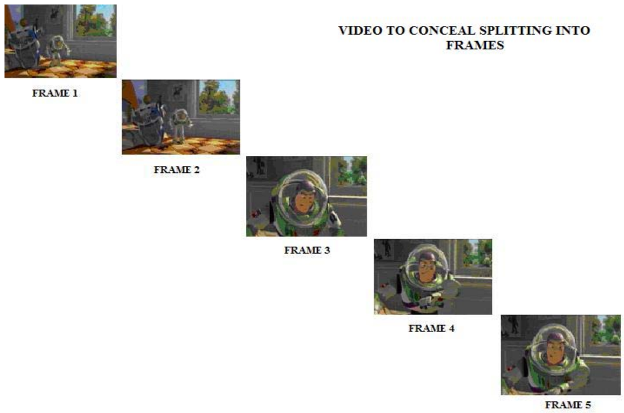

Fig. 7 Video to Conceal splitting into frames

2.2.1 Splitting the frames with *DWT*

With the frames splitted from the video, each one is splitted into its RGB components, for each one the DWT is applied to reach the sub matrices of Approximations, Horizontal Details, Vertical Details and Diagonal Details, Figure 8.

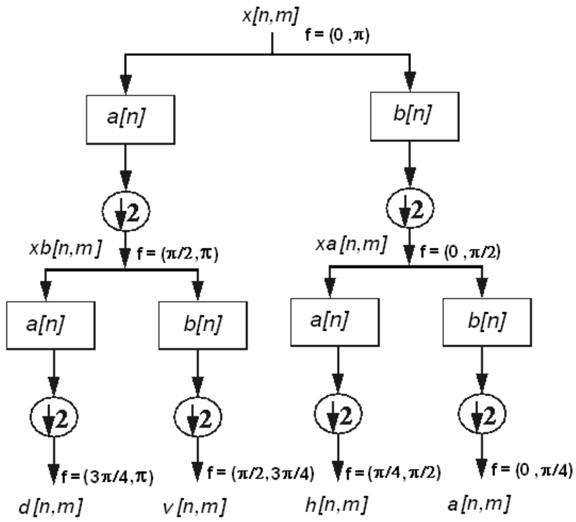

Fig. 8 Wavelet Transform of the image x[n,m]

It can be noticed that the Approximations sub matrix $a[n/2,m/2]$ contains the lower frequency components, having the half size of the original image. The Horizontal Details sub matrix $h[n/2,m/2]$ contains the medium lower frequencies of the image. Finally, the Vertical Details and the Diagonals $v[n/2,m/2]$ $d[n/2,m/2]$, have medium higher and higher frequency components, respectively [2, 3].

One important aspect in the DWT is the energy compactness in the Approximation sub matrix. In the function of DWT used the reached compactness can be higher.

The process is carried out for the frames of the *Base Video*, and twice for the *Video to Conceal*. In both cases a different DWT is selected. In the case of the *Video to Conceal*, the Approximations and Horizontal Details $a[n/4,m/4]$, $h[n/4,m/4]$ sub matrices are just needed, respectively. It must be noticed that the sub matrix size is the half of that of the *Base Video*.

2.2.2 Defining the threshold criterion

To be able to hide the sub matrices of the *Video to Conceal* in the sub matrices of the *Base Video*, a threshold criterion must be established for selecting those pixels that could be modified. The sampling values to be modified should be less or equal than the selected threshold. In this work the employed threshold consists in the standard deviation of the matrix; for each RGB component in the *Base Video*, the standard deviation is calculated.

A value of the standard deviation is obtained for each row and column of $x_R[n,m]$, $x_G[n,m]$ and $x_B[n,m]$, from which a vector is set up with the number of standard deviations which is the number of rows or columns stored in the matrix.

The equation that defines the standard deviation [9] of a series of values is given by:

$$\sigma = \sqrt{\sum_{l=1}^{m} (x_l - \bar{x})^2 / m} \tag{1}$$

where x_l is the value of each coefficient taken from the row or the column in the selected matrix, m is the total number of coefficients for the row or column and $\bar{x}$ is the average value [9] obtained from:

$$\bar{x} = \sum_{l=1}^{m} x_l / m \tag{2}$$

Lets the obtained vectors $\overset{n}{\underset{l=1}{\sigma_R(l)}}$, $\overset{n}{\underset{l=1}{\sigma_G(l)}}$ and $\overset{n}{\underset{l=1}{\sigma_B(l)}}$ be composed with n standard deviation values whose number is equal to that of the rows or columns of the sub matrices $x_R[n,m]$, $x_G[n,m]$ and $x_B[n,m]$, from which the maximun value in each vector will be considered as the threshold for such color component, and from which the boundary of the samples is defined for being modified or remained unchanged.

3. APPLICATION OF THE DISCRETE WAVELET TRANSFORMATION

The input signal is the image represented by $x[n,m]$. This discrete signal is passed through a series of filters, and is splitted into four sub images [1]. In Figure 9, it can be shown the performed procedure to the matrix $x_R[n,m]$, which is the Red component of the matrix $x[n,m,1]$.

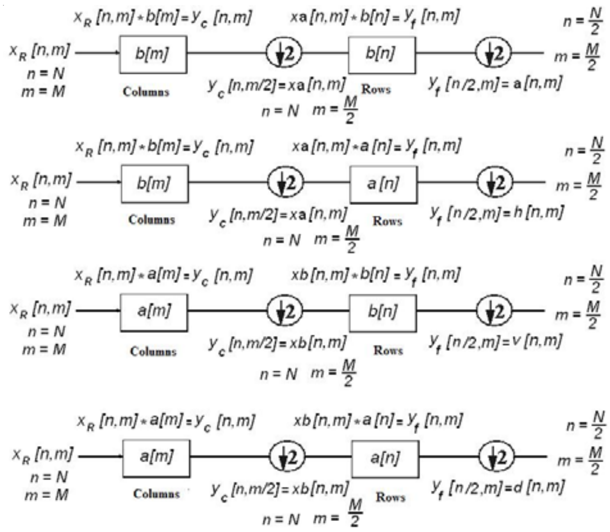

Fig. 9 Getting four sub images of the R component of the image x[n,m]

Matrix $a[n,m]$ is obtained first by filtering the lower frequencies to the columns $x_R[n,m]$, and to the resulting matrix $xa[n,m]$ the same row filtering is performed, resulting in this sub image a compression of the original containing the lower frequencies of the image.

Matrix $h[n,m]$ is reached by filtering first the lower frequencies to the columns $x_R[n,m]$, and to the resulting matrix $xa[n,m]$ the higher frequencies are filtered from the rows, remaining in this sub image the Horizontal Details of the original ones. They contained the lower medium frequencies of the image.

Matrix $v[n,m]$ is obtained by filtering first the higher frequencies to the columns $x_R[n,m]$, and the resulting matrix $xb[n,m]$ the lower frequencies are filtered from the rows, resulting in this sub image the Vertical Details of the original ones. They contained the higher medium frequencies of the image.

Finally, matrix $d[n,m]$ is set up by filtering first the higher frequencies to the columns $x_R[n,m]$, and to the resulting matrix $xb[n,m]$ the same filtering is performed to the rows, getting the Diagonal Details in this sub image, and remaining the higher frequencies of the image.

The procedure is made for the components $x_R[n,m]$, $x_G[n,m]$ and $x_B[n,m]$ of the *Base Video* and the *Video to Conceal*. After the first splitting of the *Video to Conceal*, just the matrix $a[n,m]$ is used in the RGB components. This matrix is splitted again into its four sub images and only two matrices of lower frequencies are taken, as shown in Figure 4. The two matrices of lower frequencies of the *Video to Conceal* (Approximations and Horizontal Details) have a size of *N/4* rows and *M/4* columns. Such matrix is inserted in the matrix of Approximations and Horizontal Details of the *Base Video*, having a size of *N/2* rows and *M/2* columns.

4. METHOD OF SELECTING THE RIGTH PLACE FOR CONCEALING THE VIDEO

Taking into acount the established threshold criterion, we proceed to hide the information:

i) By using threshold criterion the matrices $a[n,m]$ and $h[n,m]$ from the RGB components of the *Base Video* are analyzed for finding the places more chaotics in the images. Once such places are found, the samples of the matrices $a[n,m]$ and $h[n,m]$ of the *Base Video* are replaced by the samples $a[n,m]$ and $h[n,m]$ from the *Video to Conceal*. According to the threshold value, the amplitude of the samples of the *Video to Conceal* $a[n,m]$ and $h[n,m]$ must be changed not to affect the RGB components of the *Base Video* [5].

ii) This pseudo random form for inserting the information in the *Base Video* should produce certain security so that the *Video to Conceal* can not to be extracted.

iii) When all the samples of the matrices $a[n,m]$ and $h[n,m]$ of the *Video to Conceal* have been inserted, the IDWT is performed to the matrices $a[n,m]$, $h[n,m]$, $v[n,m]$ and $d[n,m]$ of the *Base Video* for getting the *Modified Video*.

iv) A comparison between the *Base Video* and *Modified Video*, is made to verify their similarities.

v) To reach the *Recovering Video* we proceede to split the *Modified Video* by using the DWT. The threshold criterion is applied to the matrices $a[n,m]$ and $h[n,m]$ for each of the RGB components of the *Modified Video*, then the matrix elements samples $a[n,m]$ and $h[n,m]$ are identified.

vi) To the matrices $a[n,m]$ and $h[n,m]$ (all the samples of $v[n,m]$ and $d[n,m]$ have the value of zero) we applied the IDWT twice for getting the *Recovered Video*.

vii) Finally, a comparison between the *Recovered Video* and the *Video to Conceal* is made to know how much information has been lost or modified.

5. METHOD FOR TESTING THE PROPOSED ALGORITHM

To determine the similarity between the recovered images and the original ones we calculate the cross correlation.

We apply first the cross correlation between $x[n,m,k]$, which is the matrix of the original image, and $y[n,m,k]$, which is the matrix of the modified image from the *Modified Video*. Table 1 shows the cross correlation of 50 frames of the *Base Video* and the *Modified Video*, in their color components.

As it can be noticed in Table 1, the cross correlation in 50 frames of both videos is almost one, which suggests that the *Modified Video* is so similar to the *Base Video*. While forming the avi file it is not possible to appreciate any difference between them.

The cross correlation is made between the original image of the *Video to Conceal* and the matrix of the modified image of the *Recovered Video*, the results are shown in Table 2, thus it can be found out that the cross correlation between the frames of the Blue component is almost in accordance. For the Red and Green components the cross correlation is less than *98%* of similarity. Such results show the lost of information that must be rejected in the Vertical and the Diagonal Details in the first reconstruction. In the second reconstruction, three of four matrices become zero (Horizontal Details, Vertical Details, and Diagonal Details).

Table 1 Base Video correlation vs Modified Video

FRAME	R	G	B
1	0.9999	0.9999	0.9999
2	0.9999	0.9999	0.9999
3	0.9999	0.9999	0.9999
4	0.9999	0.9999	0.9999
5	0.9999	0.9999	0.9999
6	0.9999	0.9999	0.9999
7	0.9999	0.9999	0.9999
8	0.9999	0.9999	0.9999
9	0.9999	0.9999	0.9999
10	0.9999	0.9999	0.9999
11	0.9999	0.9999	0.9999
12	0.9999	0.9999	0.9999
13	0.9999	0.9999	0.9999
14	0.9999	0.9999	0.9999
15	0.9999	0.9999	0.9999
16	0.9999	0.9999	0.9999
17	0.9999	0.9999	0.9999
18	0.9999	0.9999	0.9999
19	0.9999	0.9999	0.9999
20	0.9999	0.9999	0.9999
21	0.9999	0.9999	0.9999
22	0.9999	0.9999	0.9999
23	0.9999	0.9999	0.9999
24	0.9999	0.9999	0.9999
25	0.9999	0.9999	0.9999
26	0.9999	0.9999	0.9999
27	0.9999	0.9999	0.9999
28	0.9999	0.9999	0.9999
29	0.9999	0.9999	0.9999
30	0.9999	0.9999	0.9999
31	0.9999	0.9999	0.9999
32	0.9999	0.9999	0.9999
33	0.9999	0.9999	0.9999
34	0.9999	0.9999	0.9999
35	0.9999	0.9999	0.9999
36	0.9999	0.9999	0.9999
37	0.9999	0.9999	0.9999
38	0.9999	0.9999	0.9999
39	0.9999	0.9999	0.9999
40	0.9999	0.9999	0.9999
41	1	0.9999	0.9999
42	0.9999	0.9999	0.9999
43	0.9999	0.9999	0.9999
44	0.9999	0.9999	0.9999
45	0.9999	0.9999	0.9999
46	1	0.9999	0.9999
47	0.9999	1	0.9999
48	0.9999	0.9999	0.9999
49	1	0.9999	0.9999
50	0.9999	0.9999	0.9999

Table 2 Video to Conceal correlation vs Recovered Video

FRAME	R	G	B
1	0.9744	0.9720	0.99999
2	0.9739	0.9717	0.99999
3	0.9737	0.9716	0.99999
4	0.9735	0.9713	0.99999
5	0.9740	0.9721	0.99999
6	0.9744	0.9720	0.99999
7	0.9739	0.9717	0.99999
8	0.9737	0.9716	0.99999
9	0.9735	0.9713	0.99999
10	0.9740	0.9721	0.99999
11	0.9744	0.9720	0.99999
12	0.9739	0.9717	0.99999
13	0.9737	0.9716	0.99999
14	0.9735	0.9713	0.99999
15	0.9740	0.9721	0.99999
16	0.9744	0.9720	0.99999
17	0.9739	0.9717	0.99999
18	0.9737	0.9716	0.99999
19	0.9735	0.9713	0.99999
20	0.9740	0.9721	0.99999
21	0.9744	0.9720	0.99999
22	0.9739	0.9717	0.99999
23	0.9737	0.9716	0.99999
24	0.9735	0.9713	0.99999
25	0.9740	0.9721	0.99999
26	0.9744	0.9720	0.99999
27	0.9739	0.9717	0.99999
28	0.9737	0.9716	0.99999
29	0.9735	0.9713	0.99999
30	0.9740	0.9721	0.99999
31	0.9744	0.9720	0.99999
32	0.9739	0.9717	0.99999
33	0.9737	0.9716	0.99999
34	0.9735	0.9713	0.99999
35	0.9740	0.9721	0.99999
36	0.9744	0.9720	0.99999
37	0.9739	0.9717	0.99999
38	0.9737	0.9716	0.99999
39	0.9735	0.9713	0.99999
40	0.9740	0.9721	0.99999
41	0.9744	0.9720	0.99999
42	0.9739	0.9717	0.99999
43	0.9737	0.9716	0.99999
44	0.9735	0.9713	0.99999
45	0.9740	0.9721	0.99999
46	0.9744	0.9720	0.99999
47	0.9739	0.9717	0.99999
48	0.9737	0.9716	0.99999
49	0.9735	0.9713	0.99999
50	0.9740	0.9721	0.99999

The cross correlation between the *Base Video* and the *Modified Video* is *99.68%* while for the *Video to Conceal* and the modified images of the *Recovered Video* is *98.33%*. Despite the fact that the cross correlation is greater than *98 %* of similarity, once we obtained the *Recovered Video* it is not possible to find out any difference by the *Video to Conceal*.

6. PICTURE QUALITY

The treatment of pictures by compression and its inverse treatment it can happen that original pictures are affected, i.e., pictures can lose their characteristics of each pixel. Sometimes this loss can not be really important because if we take a look at the picture we can not recognize defects when it is regenerated. It can be called "compression loss" [4].

To calculate a "compression loss" between many pictures we can use: Mean Square Error (MSE) and Peak Signal Noise Relationship (PSNR) [4]. MSE is a square accumulated error between the original picture and compressed picture, this is given as follow:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\| I(i,j,k) - K(i,j,k)\right\|^2 \quad (3)$$

This formula is given just for 8 bit pictures or in scale of grays. In this application for RGB pictures we use next formula:

$$MSE = \frac{1}{3mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\sum_{k=0}^{2}\left\| I(i,j,k) - K(i,j,k)\right\|^2 \quad (4)$$

where I and K are two pictures given, the first one is the original and the second one was modified, m and n are the rows and columns conforming picture matrix.

PSNR is the relation between the maximum value possible and the power of noise. This tool shows the fidelity of compressed picture vs original picture. PSNR is usually expressed in *dB*, given by the next formula:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \quad (5)$$

where MAX_I^2 is the maximum value taken by one pixel from the picture frame. Typical PSNR values for high resolution pictures are between *30 dB* and *40 dB*.

In this research for *Video to Conceal* and *Recovered Video* we find a PSNR value of *28.445 dB*.

7. CONCLUSION

The method proposed in this research allow us to hide a great amount of information in a *Video File*, and its samples are not greatly altered. On the other hand, the method of inserting data into original file is so secure, since we chose the place where the information will be hidden depending on the image characteristic. This pseudo random form to conceal information produces a great security, since the most adequate places in the original image for inserting information are found. In addition, just by changing the threshold value, the places in the original image where the information will be hidden, change too. After performing the tests we can notice that in the *Base Video* a great deal of information can be hidden, and the quality of the *Modified Video* remains without changes.

In addition, we set that the *Video to Conceal* should be splitted twice, therefore the matrix of Approximations and Horizontal Details are needed for the reconstruction. Although in this work only the tests for hiding information were done in this two matrices, the Vertical and Horizontal Details can be added to grow the cross correlation between the *Video to Conceal* and the *Recovered Video*. However, it increases the processing time and hence the possibility of implementing it in real time applications.

8. REFERENCES

[1] J.S. Walker, *A Primer on Wavelets and Their Scientific Applications*, Chapman & Hall / CRC Press, Boca Raton, 2003.

[2] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, Englewod Cliffs, New Jersey, 1995.

[3] D.F. Walnut, *An Intoduction to Wavelet Analysis*, Birkhäuser, Boston, 2004.

[4] V.P. Tuzlukov, *Signal Processing Noise*, CRC Press, Boca Raton, 2002.

[5] I. Orea-Flores, M.A. Acevedo and J.L. López-Bonilla, Wavelet and discrete cosine transform for inserting information into BMP image, *J. Anziam*, Vol. 48, No. 1, pp. 23-35, 2006.

[6] R. Neelamani, R.L. Quiroz, Zh. Fan, S. Dash and R.G. Baraniuk, JPEG compression history estimation for color images, *IEEE Trans. Image Processing*, Vol. 15, No. 6, pp. 1365-1378, 2006.

[7] G.C. Canavos, *Probabilidad y Estadística: Aplicaciones y Métodos*, McGraw Hill, Mexico, 1993.

ŠIFRIRANJE INFORMACIJA U VIDEO ZAPISU

SAŽETAK

U ovom radu predlažemo jedan algoritam za skrivanje informaciju u avi dokumentu. Taj avi dokument je podijeljen na audio i video komponente. Audio komponenta se pohranjuje u jednom dokumentu čiji se parametri ne mijenjaju. Video komponenta je podijeljena u okvire. Okvir sadrži 24 slike koje se transformiraju pomoću diskretne kratkovalne transformacije - Discrete Wavelet Transform (DWT). Pomoću DWT dobijemo 4 podslike, a svaka od njih ima drugačiju širinu trake. Čim smo podijelili frekvenciju, odabrali smo mjesta za skrivanje informacije. U ovom radu skrivanje informacija je napravljeno u drugom avi dokumetu. Ova se metoda može primijeniti za bilo koju informaciju. Dokument, koji treba sakriti, dijeli se na audio i video. Oba se sakrivaju u tim podslikama iz izvornog videa, ali neovisno jedan od drugoga. Skrivani video se odvaja od komponente frekvencije, a samo one donje se pohranjuju, dok se ostale brišu. Sažeta verzija slike, koju treba sakriti, ubacuje se u podsliku videa. Nakon što se obavi skrivanje, primijeni se inverzna diskretna kratkovalna transformacija - Inverse Discrete Wavelet Transform (IDWT), tako da se video ponovo rekonstruira. Da bi se izvukla skrivena informacija, napravi se isti postupak u kojem se primjenjuje IDWT na obnovljeni dokument.

Ključne riječi: *šifriranje informacija, video zapisi, audio zapisi, podslike.*