*Kufa Journal of Engineering*
*Vol. 6, No. 2, June, P.P. 90-100*
*Received 15 December 2014, accepted 16 June 2015*

# THREE DIMENSIONAL DCT AND TEMPORAL SECONDARY CLUSTERING BASED VIDEO STEGANOGRAPHY

**Ahmed Toman Thahab[1]**

**1 Asst. Lecturer, Department of Electric and Electronic Engineering, College of Engineering/University of Kerbala**

## ABSTRACT

Security of data transmission is one of the important factors in communication between two parties. Steganography is a part of image processing that applies to the principle that hidden data is concealed in such a way that a part from the two communicating parties knowledge the procedure of hiding method. In this paper, a steganography method using the video file as a cover is used to hide secret data video benefiting from the insignificancy exploited by the temporal DCT domain in the video cover. A two-dimensional transform is applied on the video footage; the coefficients in the transform domain between the frames are gathered into groups. Another DCT transform is applied on the one dimensional array of each group. The secret binary data are embedded using a dynamic insertion strategy. Experimental results showed that the algorithm produces a high quality stego-video whereas, the Peak signal to noise ratio is reached to 46 dB for particular parameters. The normalized correlation is in the range of (0.999-1) which means that stego-video is highly correlated to the cover video. This states that the algorithm is secure and does not arouse suspicions that secret data is concealed.

**KEYWORDS:** Discrete cosine transform, steganography, mean square error, cover video, secret data, clustering.

# ألاخفاء الفديوي بأستخدام تحويل الجيب تمام المتقطع والتجميع الثانوي الوقتي

**المدرس المساعد أحمد تومان ذهب**

**كلية الهندسة / جامعة كربلاء، قسم الهندسة الكهربائية والالكترونية**

**الخلاصة**

أن أمن المعلومات من أهم العوامل في الاتصال بين طرفين. وأن الاخفاء الصوري هو جزء من عمليات معالجة الصور الذي يؤمن البيانات المخفيه بطريقة بحيث أن خوارزمية الاخفاء تعرف فقط بين الطرفين المتصلين. في هذا البحث تم أقتراح طريقة لاخفاء معلومات سرية بالأستفادة من عدم أهمية بعض المعلومات المتوفرة في معلومات الغطاء ذات النوع فيديو. في البدء يتم تطبيق تحويل الجيب تمام المتقطع على معلومات الغطاء. وأن المعاملات في مجال التحويل تجمع بشكل مجاميع ثم يتم تطبيق تحويل الجيب تمام المتقطع على هذه المصفوفات الاحادية للبيانات السريه و يتم أدخالها بأستخدام أستراتيجية أدخال ديناميكيه . أظهرت ألنتائج العملية أن الخوارزميه تنتج فديو فيه معلومات مخفيه تبلغ فيها نسبة ذروة الاشارة الى الضوضاء 46DB. كما أن التشابه المطبع يكون بين (0.999-1) مما يجعل معلومات الفديو الذي يكون فيه المعلومات مخفية أكثر تشابه مع معلومات الغطاء. وأن هذه الخاصية تجعل الخوارزمية أكثر أمناً ولا تثير شكوك بوجود معلومات مخفية.

**الكلمات المفتاحية:** التحويل الجيب تمام المتقطع، الاخفاء الصوري، معدل مربع الخطأ، الغطاء الفديوي ، البيانات السرية، تجميع .

## 1. INTRODUCTION

Nowadays, illegal and unauthorized access of data is increased during the transferring of data between the sender and receiver. Therefore; data hiding is becoming a reviving subject for researchers to prevent hackers from intruding secret data. Steganography is a process of concealing secret data in a file cover without attracting observer attention. There are many military and industrial application used in steganography. In order to increase the security and imperceptibility of the steganography techniques, researchers have taken this technique one step further by embedding information in video as well as audio file P. Gerami, et.al, (2012). Imperceptibility means that the output cover after embedding the secret data which is called "stego" is noiseless and resembles the original cover media.

Hiding data in transform domain is more robust and suitable to human vision system. H. Liu, et al, (2005) proposed a blind video data hiding algorithm using discrete wavelet transform, the algorithm is robust and good invisibility since multiple information bits are embedded in the low-low band. The hidden information is robust against additive noise and scaling attacks. Others used video encoders such as H.264 to hide data in video output stream. K. S. Patras, et.al (2007) proposed a blind data hiding technique taking advantages of the variable block sizes used by the encoder H.264, with a capacity of 1600 bits, the PSNR for a particular inter frame is between(49.800-49.700dB).

A. K. Bhaumik, et al (2009) proposed a data hiding scheme to enhance security which is a blind scheme and its effect is almost negligible. The technique uses the LSB replacement method but it is extended to the three planes which is less detectable than a single plane replacement. C. S.Suma et.al, (2010) proposed a steganography technique using transform method in real time in compressed domain; the method utilizes a combination of audio and text. The method produces a peak signal to noise ratio (PSNR) of approximately 44dB in one of the frames.

Sampat, et al, (2012) proposed a new stenographic system that concentrates on security of the secret data. The system generates a cover media by itself without using the conventional cover media. Steganoanalysis is difficult in this algorithm. Video steganography can be used in the spatial domain without any transform in frequency domain. H.Gupta, et.al (2014), used a hybrid combination of the least significant bit algorithm and advance encryption standard. The method uses one of the three keys 129, 192, 256 to make it difficult for the intruder to suspect the existence of data inside the video. The maximum average PSNR reached is 47dB.

Recent papers such as , V. Chantrapornchai, et.al, (2014) proposed a video steganography that hides image using the output wavelet coefficients. A discrete wavelet transform is applied on the video frames and some positions are used to hide secret pixels of the secret image.

In this paper, a video steganography technique is proposed to embed data in the temporal-transform in the three dimensional DCT domain of a cover video. The system security is evaluated on correlation and PSNR between the cover and stego-video and pixel distribution.

## 2. DISCRETE COSINE TRANSFORM (DCT)

DCT is a powerful signal tool used in many applications in image processing such as image compression, face recognition…etc. The DCT transforms the signal to frequency subbands is given by eq. 1& 2, A. R. Chadha, et.al, (2011).

$$\boldsymbol{F(u) = \sqrt{\frac{2}{N}} \sum_{i=0}^{N-1} A(i) * cos\left(\frac{u(2i+1)\pi}{2N}\right) * f(i)} \qquad (1)$$

where:

$$A(i) = \begin{cases} \frac{1}{\sqrt[2]{2}} & for\ u = 0 \\ 1 & otherwise \end{cases} \quad (2)$$

*f(i):* input signal

One of the important properties of the DCT is separation which means that a two-dimension DCT (2D-DCT) can be reconstructed from a one dimension equation. The two-dimension DCT equation for an image is given in eq. 3, 4&5, A. R. Chadha,et.al, (2011):

$$F(u,v) = \sqrt{\frac{2}{N}}\sqrt{\frac{2}{M}}\sum_{i=0}^{N-1} A(i) * cos\left(\frac{u(2i+1)\pi}{2N}\right) * \sum_{j=0}^{M-1} A(j) * cos\left(\frac{v(2i+1)\pi}{2M}\right) * f(i,j) \quad (3)$$

where:

$$A(i)=\begin{cases} \frac{1}{\sqrt[2]{2}} & for\ u = 0 \\ 1 & otherwise \end{cases} \quad (4)$$

$$A(j)=\begin{cases} \frac{1}{\sqrt[2]{2}} & for\ v = 0 \\ 1 & otherwise \end{cases} \quad (5)$$

*f(i,j):* is the input image.

The result of the 2D-DCT is approximately low frequency which is located at the upper corner of the coefficients in the transform domain and begins to increase diagonally towards the lower right corner of the transform domain.

## 3. PROPOSED VIDEO STEGANOGRAPHY USING THREE DIMENSION TEMPORAL DCT

The main principle of the proposed algorithm is to embed data in the temporal domain. Using such a domain will create less distortion. Three-dimension DCT is utilized to embed secret data in the cover video. Fig. 1 shows the block diagram of the proposed steganography system.

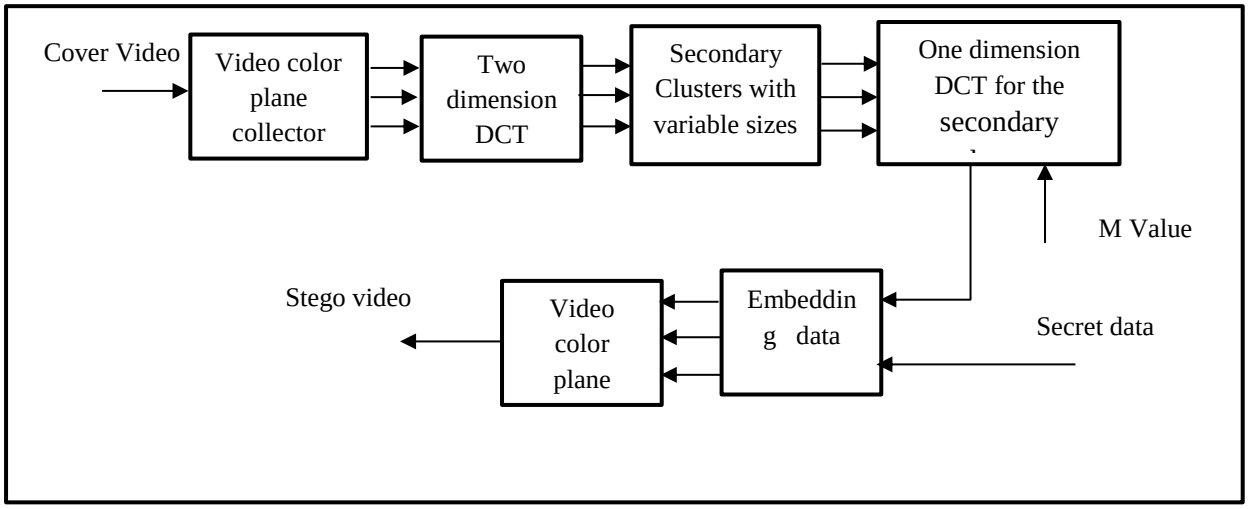



**Fig. 1. Block diagram of the proposed video steganography using three-dimension DCT**

***Color plane collector*** separates the video footage to frames; each frame consists of three color planes. All correlated frames are grouped in a three dimension matrix. Since we have three

colorplanes in single frame, all frames will be grouped with respect to the following equations:

$$Z1 = R_1 \cup R_2 \cup R_3 \ldots\ldots \cup R_n \quad (6)$$

$$Z2 = G_1 \cup G_2 \cup G_3 \ldots\ldots \cup G_n \quad (7)$$

$$Z3 = B_1 \cup B_2 \cup B_3 \ldots\ldots \cup B_n \quad (8)$$

where:
$Z_1$, $Z_2$&$Z_3$: collection matrices.
*R, G&B*: color planes.
*n*: Number of frames.
The above procedure will try to attain non-fluctuating magnitudes.
Each frame of the groups ($Z_1$, $Z_2$&$Z_3$) is transferred in the DCT domain using 2D-DCT using eq. 2. Locating arrays at a specific spatial location with an array of temporal length and applying the conventional 1D-DCT on the temporal array, the output of the transform will be an array of fluctuating magnitudes, linear and highly rippled response, at any spatial point $Z_{x(i,j)}$. This will generate high frequency AC magnitudes that will cause distortion during embedding secret data. In order to reduce the magnitude of these high frequency coefficients, the array is separated into clusters with length of (L) for each group. A one dimension DCT will be applied on the individual cluster according to eq. 9:

$$Sg(Z_x)^f = dct\ \{Z_{x_{(i,j,k+1)}}, \ldots\ldots, Z_{x_{(i,j,k+L)}}\}^f \quad (9)$$

where:
$k=0, L, 2L, 3L\ldots\ldots < n$. (Each value of *k* is associated with the 1D- cluster)
*Sg*: secondary clustering of length *L*.
$Z_x$: matrix color cluster and $x=1, 2, 3\& \in Z$
*f:* frame index.
This will result in linear response of the coefficients, in other words, the magnitudes of the high frequency coefficients are decreased in magnitudes, therefore; all the energy in the array is compacted at the low frequency coefficients.
***The embedding process*** of the coefficients resulting from the previous operation have less high frequency magnitude coefficients, therefore; Large data strings can be embedded in these coefficients more than in the low frequency magnitude coefficients. The process implies that a length of a large strings of data are embedded in the high frequency coefficients while short length of strings are embedded in the low frequency content in the three dimension DCT domain of the secondary groups. The algorithm is as follows:
**Input:** sub group in three dimension DCT domain
**Convert:** all coefficients of subgroup to binary.
**Divide:** the subgroup into M parts.
**Embed:** string1 (str1) in (1: M/2) coefficients & string2(str2) in (M/2: M) coefficients.
Whereas: length (str1) <length (str2).
**Output:** embedded coefficients in three dimension DCT domain.
After the above operation is completed, all the subgroups are collected and an inverse discrete cosine transform (IDCT) is applied to attain the temporal domain for each array of the group video frames. The arrays are ordered and spatially inverse transformed from the transform domain to the spatial domain using 2D-IDCT transform. The same embedding operation is applied on the three frame matrices.

The two dimension spatial domain frames which are arranged in three matrices, $Z_1$, $Z_2$&$Z_3$, are separated and rearranged to form the R, G&B matrices. Each frame will consist of three color planes, each plane consists of one of the matrices $Z_1$, $Z_2$&$Z_3$ according to eq. 10:

$$\text{Frame } (n^{th}) = [Z_1 (n^{th}) + Z_2 (n^{th}) + Z_3 (n^{th})] \quad (10)$$

The frames which are created from the three matrices are sequenced and produced as a stego video which hosts the embedded data in its content.
The capacity (Cap) of the algorithm depends on the length of the video footage and the length of the secondary group is given by eq.11:

$$\text{Cap} = \left(\left(L_1 * \left\lceil \frac{L_g}{M} \right\rceil\right) + L_2 \left(L_g - \left\lceil \frac{L_g}{M} \right\rceil\right)\right) * (n/L_g) * size(frame) * 3 \quad (11)$$

where:
$\boldsymbol{L_1}$: Length of string embedded in low frequency components.
$\boldsymbol{L_2}$**:** Length of string embedded in high frequency components.
$\boldsymbol{L_g}$**:** Length of the secondary clusters which is divisible on $\boldsymbol{n}$.
$\boldsymbol{M}$**:** The barrier value of considering the coefficients as low &high frequencies.
$\boldsymbol{n}$**:** Length of video footage.
The longer the video is (number of frames), the more bits can be embedded in the video footage. The capacity to embed also depends on the barrier value (M) whereas, if the value of M is high which means considering less high frequency coefficients then, str2 which has more bits will be embedded only in few coefficients, while (str1) will be embedded in many low frequency coefficients. Since length (str1) is less than length (str2), the capacity will be decreased.

## 4. VIDEO STEGANONALYSIS

Inverse of the steganography system shown in Fig. 1 is conducted in this section, Fig. 2 shows the steganoanalysis of the proposed system.

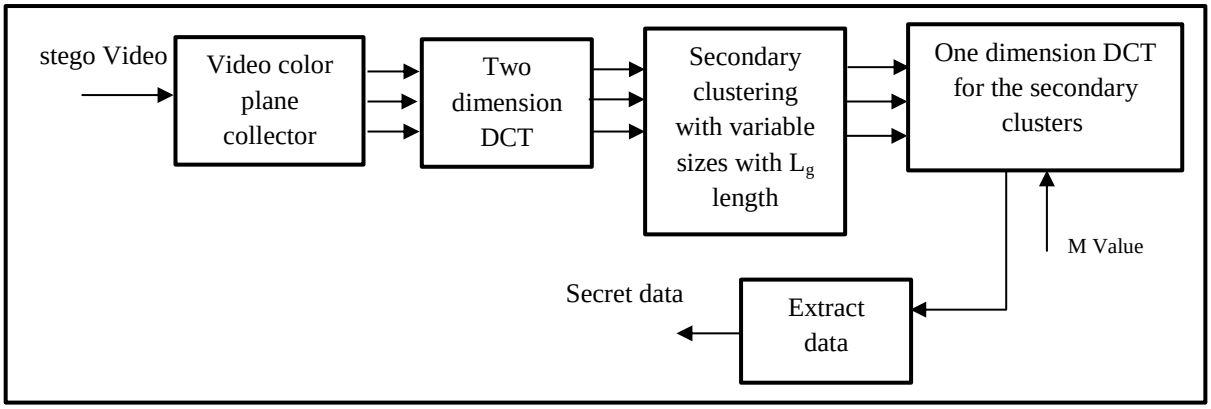



**Fig. 2. Video steganoanalysis using three dimension DCT**

A stego video is input to the algorithm where the video is separated to three matrices $Z_1$, $Z_2$&$Z_3$. A two dimension DCT is applied on each frame. Each array of the temporal DCT domain is divided at length of $L_g$ (which is given as a key) to cluster. Applying one dimension DCT and extracting the str2 from high frequency coefficients and str1 from low frequency coefficients, binary data are reordered and outputted.

## 5. RESULTS AND DISCUSSION

At this stage, the algorithm is tested using two inputs, the cover video and the data that will be embedded in the cover video. Various cover videos are tested with a couple lengths of secondary clusters and M value. In any steganography algorithm, the performance is mostly presented in peak signal to noise ratio (PSNR) and normalized correlation (NC). The algorithm is executed on Matlab 2010b using M-file programming. Table 1 Shows the results with various video contents and different motions with $L_g$ =16, n=80. It is obviously seen that as the value of M is increased, the PSNR is decreased for a constant $L_g$. As explained in the previous section, the selection of M means increasing the number of high frequency coefficients.

**Table 1. The effect of M value on PSNR and capacity with frame size (128*128)**

| Name of Video | Capacity in Bytes | M value | PSNR in dB | NC |
|---|---|---|---|---|
| Highway | 1228800 | 2 | 43.4168 | 1.0000 |
| | 1413120 | 7 | 41.7614 | 0.9999 |
| | 1443840 | 12 | 41.5009 | 0.9998 |
| Claire | 1228800 | 2 | 45.1420 | 1.0000 |
| | 1413120 | 7 | 43.8370 | 1.0000 |
| | 1443840 | 12 | 43.4781 | 0.9967 |
| Viptrafic | 1228800 | 2 | 43.1999 | 0.9999 |
| | 1413120 | 7 | 41.9498 | 0.9995 |
| | 1443840 | 12 | 41.7590 | 0.9995 |
| Xylophone | 1228800 | 2 | 44.2745 | 1.000 |
| | 1413120 | 7 | 43.0597 | 0.9996 |
| | 1443840 | 12 | 42.7359 | 0.9998 |

It is generally known that increasing the high frequency coefficients results in high degradation of single frame quality, but using secondary clustering of temporal DCT coefficients reduced the decadency of the frame quality which means the high frequency coefficients are trivial in magnitude, therefore; more bits can be embedded. As for the capacity, the length of string (*str2*) is larger than string (*str1*) in low frequency interpreting the dynamic strategy of embedding. Since the numbers of high frequency coefficients are more than low frequency, then the capacity of the algorithm is higher than others without degrading the video quality of the embedding data.

The reduction in PSNR in Table 1, for worst case between M=2 (number of high frequency coefficients are 8)& M=12 (number of high frequency coefficients are 13) when $L_g$=16 is 4% but an increase in capacity of 15% . Correlation reduction is almost negligible.

**Table 2. The effects of frame length on capacity with frame size (128*128)**

| Name of Video | Capacity in Bytes | M value | PSNR in dB | NC |
|---|---|---|---|---|
| Highway | 1474560 | 2 | 43.4330 | 1.0000 |
| | 1695744 | 7 | 41.8104 | 0.9993 |
| | 1720320 | 12 | 41.6337 | 0.9991 |
| Claire | 1474560 | 2 | 45.2254 | 1.0000 |
| | 1695744 | 7 | 43.8297 | 1.0000 |
| | 1720320 | 12 | 43.5727 | 0.9999 |
| Viptrafic | 1474560 | 2 | 43.0913 | 0.9999 |
| | 1695744 | 7 | 41.7953 | 0.9996 |
| | 1720320 | 12 | 41.6654 | 0.9995 |
| Xylophone | 1474560 | 2 | 44.2173 | 1.0000 |
| | 1695744 | 7 | 43.0466 | 0.9999 |
| | 1720320 | 12 | 42.8450 | 0.9998 |

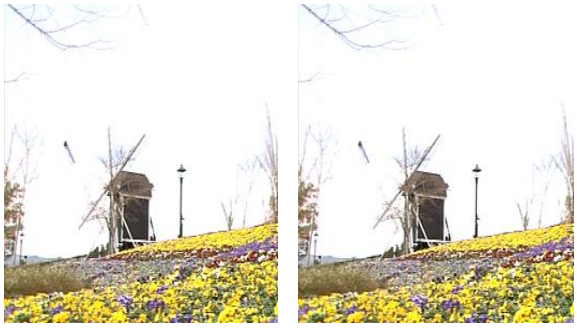
Table 2 Illustrates the capacity of the proposed algorithm by increasing the number of video frames. At this experiment n= 96, $L_g$=24 and frame size 128*128. Comparing Table 2 with Table 1, it is obvious that increasing the length of the video footage increases the capacity of the algorithm since we are embedding data in temporal domain through a spatial dimension. Altering the spatial dimension of the video frame can also increase the embedding capacity of the algorithm.

As for the length of group $L_g$, and observing tables that in 1&2, in some cases (Highway & Claire videos), the PSNR parameter in Table 1 is higher than Table 2 while the other videos, PSNR is decreased in Table 2. It is been demonstrated during experiments that in high motion object video, using small length groups preserves frame quality than using large length groups since motion will be distorted.

All parameters used in tables are image metric parameters which have been modified to be applied on video quality assessment. Another parameter to assess quality is the human vision system. Embedded data video frames which is called stego frames are shown in Figs. 3, 4 and 5 for various cases.

Reconstructed frames from experiments show that the stego video frame is approximately identical to the original frame of the videos, this supports the property that steganography does not attract attention that secret data is embedded in the cover video.

The value of the normalized correlation for the tables of experiments are in the range of (0.9995-1.000) which is around one, this reflects the sense that all frames are highly correlated through the video footage. Increasing the M value with a particular $L_g$, decreases NC parameter for the video since the numbers of high frequency coefficients are increased. Observing closer the NC value, the increase of high frequency coefficients does not considerably decrease the NC value because of applying the secondary clustering. The reduction in NC is 0.02% for the case M=12 and M=2 for $L_g$=24 but this reduction is insensitive for human eye to recognize. If one of frames was not correlated with the cover frame, the stego video may trigger hacker attention. Fig. 6 Illustrates the NC value of each frame against the number of frames, the frames are highly correlated to the cover frame.



a. original Frame b. Stego frame

**Fig. 3. frame No.6 from original and stego video with Lg=24, M=7**

a. original Frame b. Stego frame

**Fig. 4. frame No.56 from original and stego video with Lg=16, M=12**

a. original Frame b. Stego frame

**Fig. 5. frame No.70 from original and stego video Lg=16, M=12**

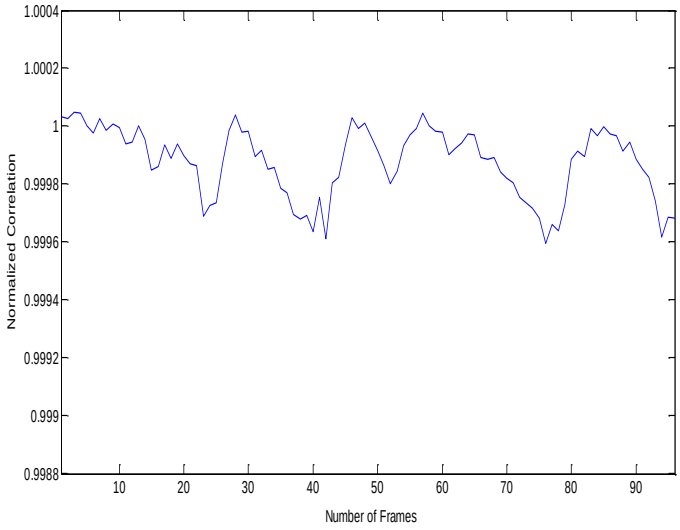



**Fig. 6. The normalized correlation for each frame for viptrafic Lg=16, M=12**

## 6. SECURITY OF THE PROPOSED ALGORITHM

The essential goal of any video steganography algorithm is security. In order to demonstrate the algorithm's security, researchers observe the histogram. Fig. 7 shows the pixel distribution of a single frame for experiment $L_g$=24 &M=7.

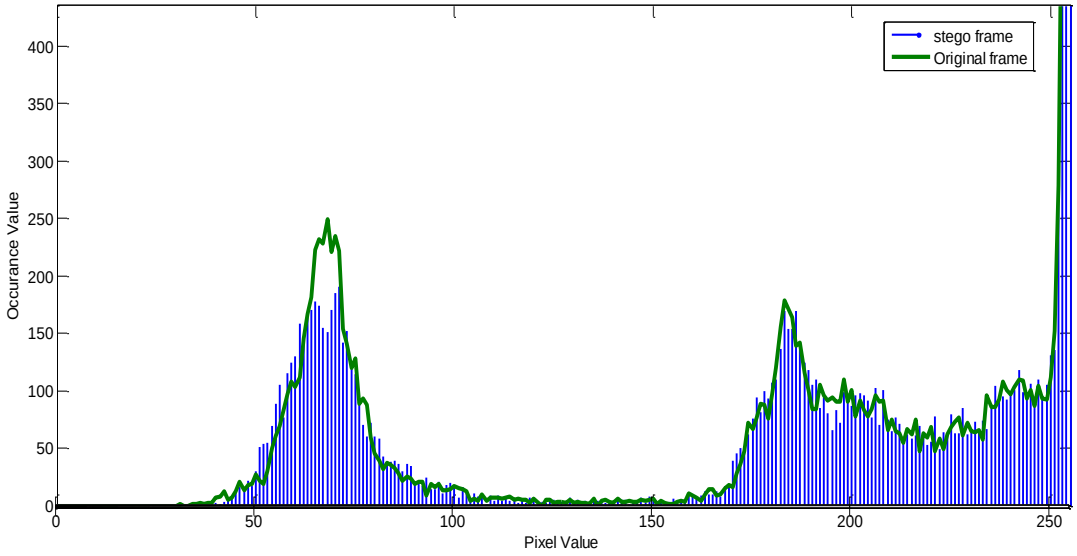



**Fig. 7. Pixel distribution of frame. 40 for Claire experiment Lg=16, M=7**

It can be seen that the occurrence of the pixels are highly correlated between the stego and cover video frames. The distributions of pixel values are approximately identical except for pixels at values around 70. These values appear more than 249 times, while in the original video they appear less frequent. But these pixels are not equal to zero to produce high distortion. This means stego video does not attract attention that secret data is embedded in cover video media.

## 7. COMPARISON WITH OTHER ALGORITHMS

In order to demonstrate the superiority of the proposed algorithm, a comparison is set to other corresponding algorithms as shown in Table 3.

**Table 3. Comparison between other algorithms in terms of PSNR and NC**

| Name of Algorithm | Maximum value of PSNR in dB | Range of NC |
|---|---|---|
| Proposed Algorithm | 46.7508 | 1-0.9995 |
| K. Dasgupta et.al[2013] | 41.613 | Not mentioned |
| S. Suma | 43.89 | 1-0.1 |

In paper K. Dasgupta et.al, (2013), the range of PSNR attained on three video is (39.374-41.613)dB with a constant payload of 204288(2.66 bPB) while in the proposed algorithm if $L_g$=10, $n$=30, frame size=128*128 which is less than frame size used in P. Gerami, et al, (2012), we gain a capacity of (313344) bits and a PSNR of 46.7508dB.

The work proposed in S. Suma which used a compressed bit stream, the range of correlation value is (1-0.1) for the video footage. While in the proposed algorithm, the correlation is in the range of (1-0.9995) which means that the proposed algorithm produces highly correlated video frames than work in S. Suma et.al, (2010). The proposed algorithm produced a PSNR for a particular experiment (41.5-46dB) as shown in Fig. 8, while in S. Suma et.al, (2010), produced a PSNR of range (35.5-43.89dB). This shows superior performance in quality and security.

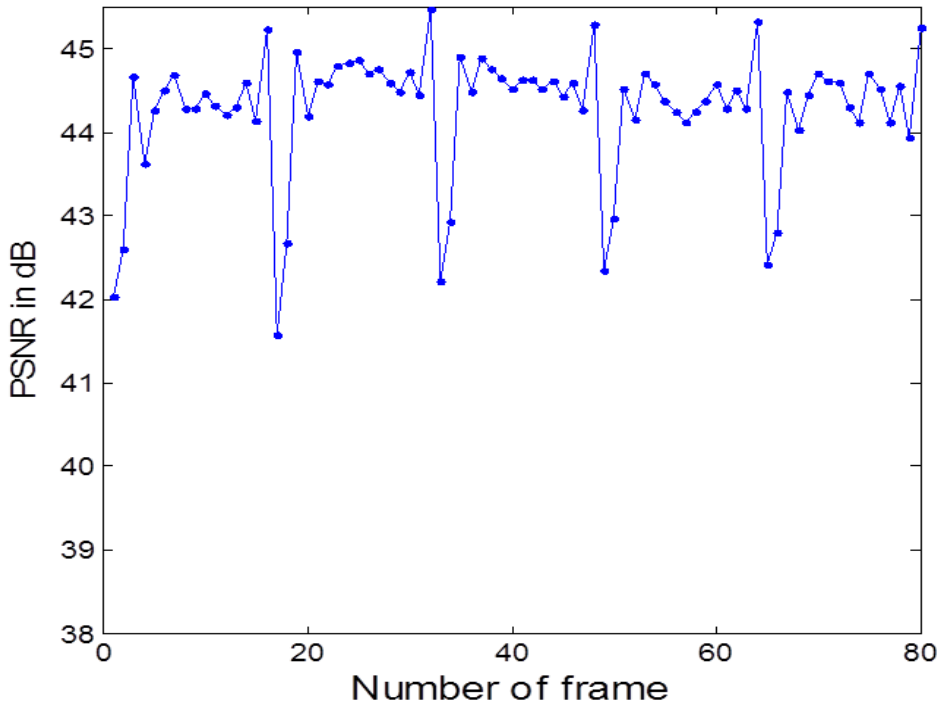



**Fig. 8. Illustrates the PSNR for each frame of the video footage**

## 8. CONCLUSIONS

A new video steganography is proposed using three dimensions DCT. Splitting the one dimension array in to clusters of length ($L_g$) will produce small magnitudes of high frequency coefficients which will permit to embed more data bits in these coefficients. The length of the secondary clusters influences the frame quality of the video; $L_g$ should be small in high motion objects and considerably increased in slow motion object videos. Values of M effects frame video quality but more bits are embeded in cover media without highly degrading video quality. The capacity of embedding depends on the dimension of the frame, temporal length of the video and length of cluster and M value. Since the video footage is highly correlated to the cover media, therefore; the algorithm applies to the principle of security which most steganography satisfy. The algorithm also produces superior results than other algorithms. For future work suggestion, three dimension DCT can be replaced by discrete wavelet transform.

## REFERENCES

Aman R. Chadha, Pallavi P. Vaidya and M. Mani Roja.” Face Recognition using Discrete Cosine Transform for Local and global Features”. Proceedings of the International Conference on Recent Advancements in Electrical, Electronics and Control Engineering, IEEE Xplore, CFP1153R-ART; 15 , India, 15- 17 Dec 2011, pp.502-505.

Arup Kumar Bhaumik, Minkyu Choi, Rosslin J.Robles and Maricel O. Balitanas, “ Data Hiding in Video” International Journal of database theory and Application, Vol. 2, No.2, June 2009, pp. 9-15.

C. Chantrapornchai, Kornkanok Churin, Jitdamrong Preechasuk and Suchitra Adulkasem. “ Video Steganography for Hiding Image with Wavelet Coefficients”, International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.6, 2014, pp.385-396.

Hemant Gupta and Setu Chaturvedi. "Video Steganography through LSB Based Hybrid Approach" International Journal of Computer Science and Network Security (IJCSNS) VOL.14, No.3, March 2014, pp 99-106.
Hongmei Liu, Jiwu Huang and Yun Q. Shi. "DWT-Based Video Data Hiding Robust to MPEG Compression and Frame Loss". International Journal of Image and Graphics, Volume 05, Issue 01, January 2005, pp. 111-134.
Parisa Gerami, Subariah Ibrahim and Morteza Bashardoost. "Least Significant bit Image steganography using Particle Swarm Optimization and Optical Pixel Adjustment" International Journal of Computer Applications vol 55-No.2, October 2012, pp. 131-137.
Kousik Dasgupta, Jyotsna Kumar Mondal and Paramartha Dutta." Optimized Video Steganography using Genetic Algorithm (GA) ". International Conference on Computational Intellegence: Modeling, Techniques and Applications (CIMTA) 2013 Elsevier, Procedia Technology, India, 27-28 Sep 2013, pp. 131 – 137.
S K. Kapotas, E E. Varsaki and A. N. Skodras." Data Hiding in H. 264 Encoded Video Sequences" published in Multimedia Signal Processing (MMSP), IEEE 9th Workshop in 1-3 October, 1-3 Oct. 2007, pp. 373 - 376.
S.Suma Christal Mary." Improved Protection in video steganography used compressed video Bit Streams" International Journal on Computer Science and Engineering, Vol. 02, N0.03, 2010, pp. 764-766.
Vivek Sampat, Kapil Dave, Jiger Madia and Parag Toprani. "A Novel Video Steganography Technique using Dynamic Cover Generation", National Conference on Advancement of Technologies-Information Systems &Computer Networks (ISCON) Proceedings published in International Journal of Computer Applications (IJCA), 2012, pp.26-30.

## LIST OF SYMBOLS

bpB: bit per byte.
C: Capacity of embedding.
$L_g$: Length of secondary clustering.
M: Value of barrier between the low and high frequency coefficients.
n: length of frame footages.
NC: Normalized Correlation.
PSNR: Peak Signal to Noise Ratio.
Str1: the binary string of data embedded in low frequency coefficients.
Str2: the binary string of data embedded in high frequency coefficients.
$Z_1$, $Z_2$&$Z_3$: Color matrices of video frames.