Performance Analysis of Digital Video Watermarking using Discrete Cosine Transform

Original scientific paper

Ashish M. Kothari
Shri Jagdishprasad Jhabarmal Tibrewala University
Jhunjhunu, Rajasthan, India
amkothari@aits.edu.in

Ved Vyas Dwivedi
Noble Group of Institutions
Junagadh, Gujarat, India
director.principal.ngi@gmail.com

Abstract *– In this paper, we have suggested the transform domain method for digital video watermarking for embedding invisible watermarks behind the video. It is used for copyright protection as well as proof of ownership. In this paper, we first extracted the frames from the video and then used frequency domain characteristics of the frames for watermarking. In this paper, we have specifically used the characteristics of the Discrete Cosine Transform for watermarking and calculated different parameters.*

Keywords *– copyright protection, digital video watermarking, discrete cosine transform, transform domain watermarking*

1. INTRODUCTION

Digital watermarking includes a number of techniques that are used to imperceptibly convey information by embedding it into the cover data. Here the cover data taken is a video sequence and watermarking is thus called Video Watermarking. Video watermarking is a field that is rapidly evolving in the area of multimedia and interest of the people in this field is increasing day by day because of the major factors [1] as stated below.

1. Privacy of the digital data is required and copying of a video is comparatively very easy.
2. Fighting against the "Intellectual property rights breach".
3. Tempering of the digital video must be concealed.
4. Copyright protection must not be eroded.

In this paper, we have focused on the transform domain watermarking method and we have specifically used the Discrete Cosine Transform. Specifically, we have embedded the messages in the R, G, and B Plane, which is a different approach compared to articles published so far. This paper is organized in seven sections. A subsequent section explains the concept of transform domain watermarking. Section 3 shows the introduction to DCT. In Section 4, we have shown a general aspect of video watermarking. Section 5 shows the formulas by which comparison of various watermarking techniques may be done. Sections 6 and 7 describe two different methods of DCT based video watermarking. Presented methods are compared in section 8.

2. TRANSFORM DOMAIN WATERMARKING

A watermarking algorithm using transform domain techniques focus on embedding information in the frequency domain of the video as opposed to the spatial domain. The most popular transforms, where the frequency domain watermarking algorithms work, are Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). These are applied to transform a frame of the video into the frequency domain where the coefficients of the digital frame are separated into different priorities in accordance with the human perception system. The watermark bits are embedded by modulating the magnitude of these coefficients.

Watermarking in the transform domain is applied in three steps as explained below.

1. *Forward Transform:* The frame of the video is first transformed to a domain that facilitates data embedding.
2. *Embedding*: The subset of the transform coefficients is modified with the prepared signature data. One can employ a model of human per-

ception to weigh the strength of the embedding modifications. By choosing a suitable frequency transform domain and selecting only certain coefficients (low- to midfrequency range), a lot of human visual system (HVS) modeling can be done implicitly. The better the transform approximates the properties of HVS, the easier it is to put more energy in the embedded signal without causing perceptible distortion.

3. *Inverse Transform:* The modified coefficients are inversely transformed to produce a watermarked frame.

3. DISCRETE COSINE TRANSFORM

Discrete cosine transformation [2, 3, 4] (DCT) transforms a signal from the spatial into the frequency domain by using the cosine waveform. DCT concentrates the information energy in the bands with low frequency, and therefore shows its popularity in digital watermarking techniques. The DCT allows a frame to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of a frame. The middle frequency bands are chosen such that they have minimized to avoid the most visual important parts of the frame (low frequencies) without over-exposing themselves to removal through compression and noise attacks (high frequencies).

Two-dimensional DCT of a frame with size MxN and its inverse DCT (IDCT) are defined in Equations 1 and 2, respectively.

$$F(u,v) = \alpha(u)\alpha(v)\sum_{y=0}^{M-1} \square \sum_{y=0}^{N-1} f(x,y)$$

$$\cos\left[\frac{(2x+1)u\cdot\pi}{2\cdot M}\right]\cos\left[\frac{(2y+1)v\cdot\pi}{2\cdot N}\right] \quad (1)$$

where

$\alpha(u) = \sqrt{1/M}$ for u=0;

$\alpha(u) = \sqrt{2/M}$ for u=1,2,3...M-1;

$\alpha(v) = \sqrt{1/N}$ for v=0;

$\alpha(v) = \sqrt{2/N}$ for v=1,2,3,...N-1;

$$f(x,y) = \sum_{v=0}^{N-1} \alpha(u)\alpha(v)F(u,v)$$

$$\cos\left[\frac{(2x+1)u\cdot\pi}{2\cdot M}\right]\cos\left[\frac{(2y+1)v\cdot\pi}{2\cdot N}\right] \quad (2)$$

where x = 0,1,2,...M-1, y = 0,1,2,...N-1

Figure 1(a) shows the three regions in the frequency domain. FL is used to denote the lowest frequency components of the block, while FH is used to denote the higher frequency components. FM is chosen as the embedding region as to provide additional resistance to lossy compression techniques, while avoiding significant modification of the cover video.

F_L	F_L	F_L	F_M	F_M	F_M	F_M	F_H
F_L	F_L	F_M	F_M	F_M	F_M	F_H	F_H
F_L	F_M	F_M	F_M	F_M	F_H	F_H	F_H
F_M	F_M	F_M	F_M	F_H	F_H	F_H	F_H
F_M	F_M	F_M	F_H	F_H	F_H	F_H	F_H
F_M	F_M	F_H	F_H	F_H	F_H	F_H	F_H
F_M	F_H	F_H	F_H	F_H	F_H	F_H	F_H
F_H	F_H	F_H	F_H	F_H	F_H	F_H	F_H

(a)

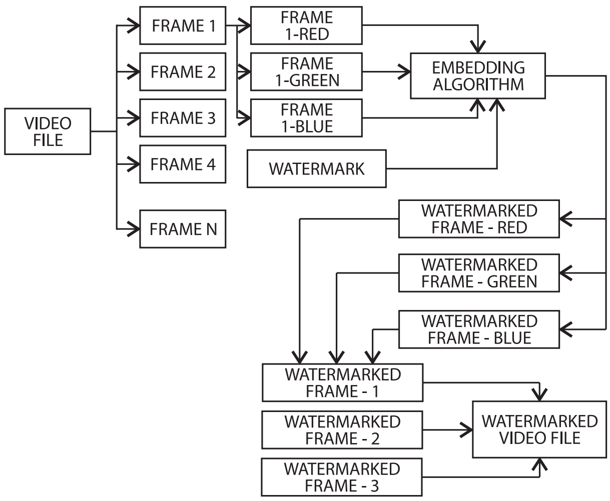
16	11	10	26	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	**22**	29	51	57	80	62
18	**22**	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(b)

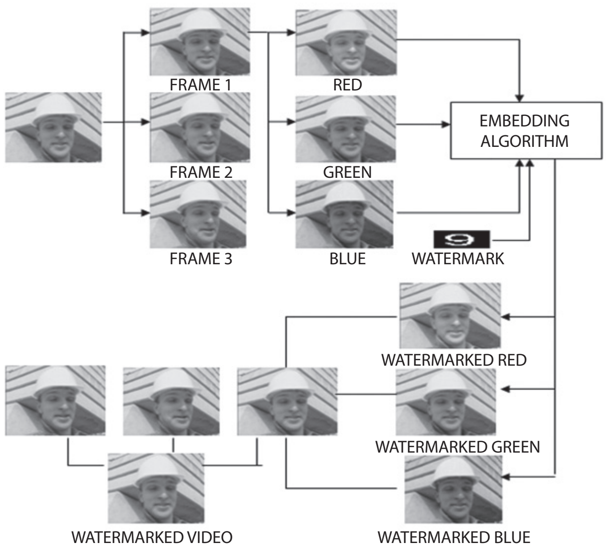
Fig. 1. Definition of DCT regions and Quantization values used in JPEG compression scheme [5]

4. IMPLEMENTATION

Figure 2 shows the idea of video watermarking at the sending end.



(a)

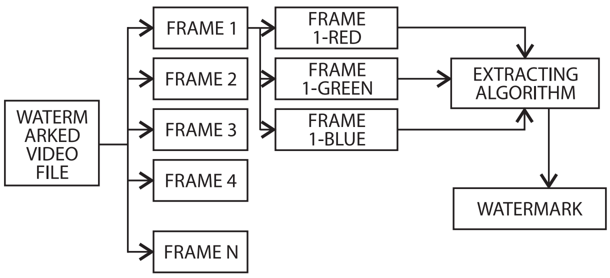


(b)

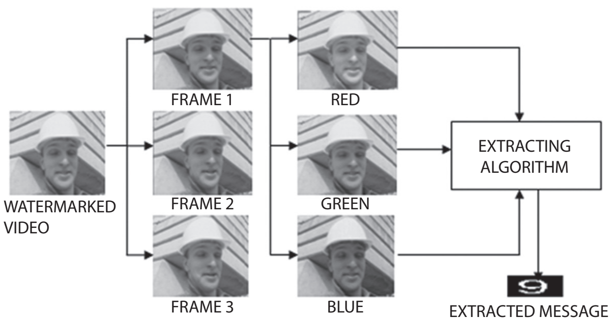
Fig. 2. Idea of Video watermarking - Embedding a watermark

Here, first of all frames is extracted from the video sequence. The next step is to divide the frame into its Red, Green and Blue part. Each part is then individually given to the embedding algorithm block where the other input is a watermark that is to be embedded. After each part is watermarked, the next frame is taken and the procedure is repeated until the last frame. After the watermark is embedded in every frame, all frames are mixed to make the watermarked video which is then transmitted in the channel.

Figure 3 shows the idea of video watermarking at the receiver end. Here, a watermarked video is divided into the frames which are divided into the Red, Green and Blue part from where the watermark is extracted. This Procedure is repeated for all frames so as to recover the watermark.



(a)



(b)

Fig. 3. Idea of Video watermarking - Extracting a watermark

The following figure (Figure 4) shows the basic block diagram of the DCT based watermarked scheme.

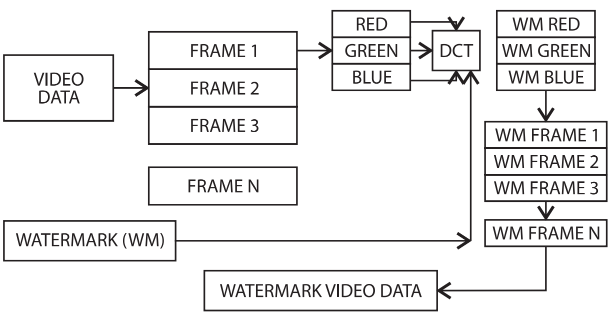


Figure 4. DCT based Watermarking technique

We have implemented two methods as explained in the following sections and we have embedded six different watermarks as shown in Figure 5.

Fig. 5. Message embedded in Frames 1 to 6 in the DCT Methods

5. VISUAL QUALITY MATRICES

We have mainly used the following visual quality metrices [6] for the sake of comparison of degradation after the watermark is added to the video.

$$MSE = \frac{1}{M \cdot N}\sum_{x=1}^{M}\sum_{y=1}^{N}\{(f(x,y) - f'(x,y))^2\} \tag{3}$$

$$PSRN = 10 \cdot \log\frac{255^2}{MSE}, \tag{4}$$

where MSE – Mean Square Error,

PSNR – Peak Signal to Noise Ratio,

f(x,y) – Original Frame of the Video,

f′(x,y) – Watermarked Frame of the Video.

The phrase peak signal-to-noise ratio, often abbreviated as PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. A higher PSNR would normally indicate that the reconstruction is of higher quality at the receiver end.

PSNR is most easily defined via the Mean Square Error (MSE) which holds for two $m \times n$ Frames f and f' where one of the Frames is considered a noisy approximation of the other.

6. METHOD 1 - COMPARISON OF MIDDLE BAND DCT COEFFICIENTS

In this method, two locations Bi (u1, v1) and Bi (u2, v2) are chosen from the FM region for comparison. Rather than choosing arbitrarily these locations, extra robustness to compression can be achieved if one bases the choice of coefficients on the recommended JPEG quantization table shown below in Figure 1(b). If two locations are chosen such that they have identical quantization values, we can feel confident that any scaling of one coefficient will scale the other by the same factor, preserving their relative size.

Based on the table, one can observe that coefficients (5, 2) and (4, 3) would make suitable candidates with the same values.

A Embedding Algorithm (Method 1)

The following steps describe the way in which the watermark is embedded in the video.

1. The video is converted into the number of frames.
2. The frame is divided into blocks and two-dimensional DCT is applied to the first block.
3. If the message bit is a '1', then it is checked whether (5, 2) is less than (4, 3) or not and if it is not, the two blocks are swapped so as to make (5, 2) < (4, 3).
4. If the message bit is a '0', then it is checked whether (5, 2) is greater than (4, 3) or not and if it is not, the two blocks are swapped so as to make (5, 2) > (4, 3).
5. If (5,2) - (4,3) < k, the coefficients are modified such that (5,2) – (4,3) > k. Here, k is a constant and it is used to improve the robustness of the watermark.
6. Move to the next block and repeat the procedure.
7. Perform Inverse DCT to have the final watermarked Frame.
8. The next frame is taken and Steps 3 to 6 are repeated until the last frame comes.
9. All the watermarked frames are combined to make the watermarked video.

Table I shows the values of PSNR and MSE for the first six frames for six different messages considering the block size = 8 and k = 10.

Table 1. Visual Quality Matrices Values

Frame No.	PSNR (dB)	MSE
1	37.0509	12.8230
2	37.0408	12.8529
3	37.0077	12.9513
4	36.9879	13.0103
5	37.0293	12.8869
6	37.0126	12.9367

B Extraction Algorithm (Method 1)

The stepwise execution of the extraction process for the watermark recovery from the video is shown below.

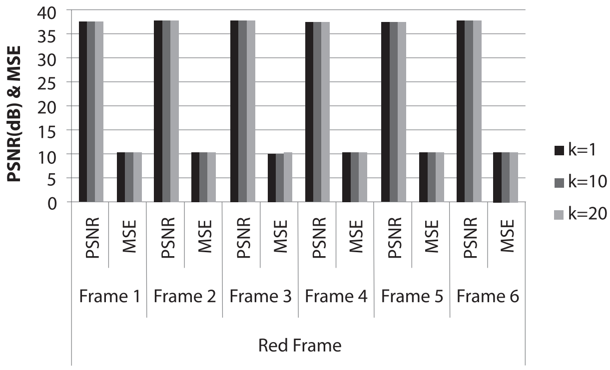
1. The Video is converted into the number of frames.
2. The frame is divided into blocks and two-dimensional DCT is applied to the first block.
3. If (5,2) > (4,3), then the message bit is 1.
4. If (5,2) < (4,3), then the message bit is 0.
5. Move to the next block and repeat the procedure.
6. The next frame is taken and Steps 2 to 5 are repeated until the last frame comes.

What follows are the watermarks recovered from frames 1 to 6, respectively (Figure 6).

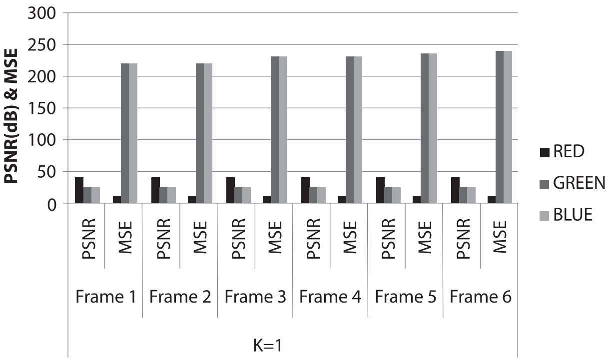
Fig. 6. Extracted Watermarks from Frames 1 to 6 in Method 1

C Results (Method 1)

Figure 7 shows the charts that can be used for the conclusion.



(a)



(b)

Fig. 7. Comparative Analysis of Method 1
(a) PSNR and MSE value for RED frame with various values of k,
(b) PSNR and MSE value for individual frame with k=1

7. METHOD 1 - COMPARISON OF MIDDLE BAND DCT COEFFICIENTS

Another possible technique is to embed a PN (Pseudo Random) sequence W into the middle frequencies of the DCT block. One can modulate a given DCT block (x, y) using equation (5)

$$f'(u,v) = \begin{Bmatrix} f(u,v) + k \cdot W(u,v) & u,v \in FM \\ f(u,v) & u,v \notin FM \end{Bmatrix} \quad (5)$$

where f (u, v) is two-dimensional DCT of f (x, y).

A Embedding Algorithm (Method 2)

The following steps describe the way in which the watermark is embedded in the video.

1. The Video is converted into the number of frames.
2. The frame is divided into blocks and two-dimensional DCT is applied to the first block.

3. A PN sequence is generated using an independent state.
4. If the message bit is 0, the middle frequency components F_M are added to the PN sequence W which was previously multiplied by a gain factor k.
5. Move to the next block and repeat the procedure.
6. Perform Inverse DCT to have the final watermarked Frame.
7. The next frame is taken and Steps 3 to 6 are repeated until the last frame comes.
8. All the watermarked frames are combined to make the watermarked video.

Table II shows the values of PSNR and MSE for the first six frames for six different messages considering the block size = 8 and k = 10.

Table 2. Visual Quality Matrices Values

Frame No.	PSNR (dB)	MSE
1	35.9592	16.4876
2	36.0479	16.1543
3	36.1212	15.8840
4	36.0068	16.3080
5	36.0585	16.1150
6	36.0948	15.9810

B Extraction Algorithm (Method 2)

The stepwise execution of the extraction process for the watermark recovery from the video is shown below.

1. The Video is converted into the number of frames.
2. The frame is divided into blocks and two-dimensional DCT is applied to the first block.
3. A PN sequence is generated using an independent state.
4. For each block, determine its correlation with base PN sequence. Set correlation to 1 when patterns are identical; otherwise calculate correlation. If correlation exceeds average correlation when set message vector is zero.
5. Move to the next block and repeat the procedure.
6. The next frame is taken and Steps 2 to 5 are repeated until the last frame comes.

Figure 8 shows the watermarks recovered from Frames 1 to 6, respectively.

Fig. 8. Extracted Watermarks from Frames 1 to 6 in Method 2

C Results (Method 2)

Figure 9 shows the charts that can be used for the conclusion.

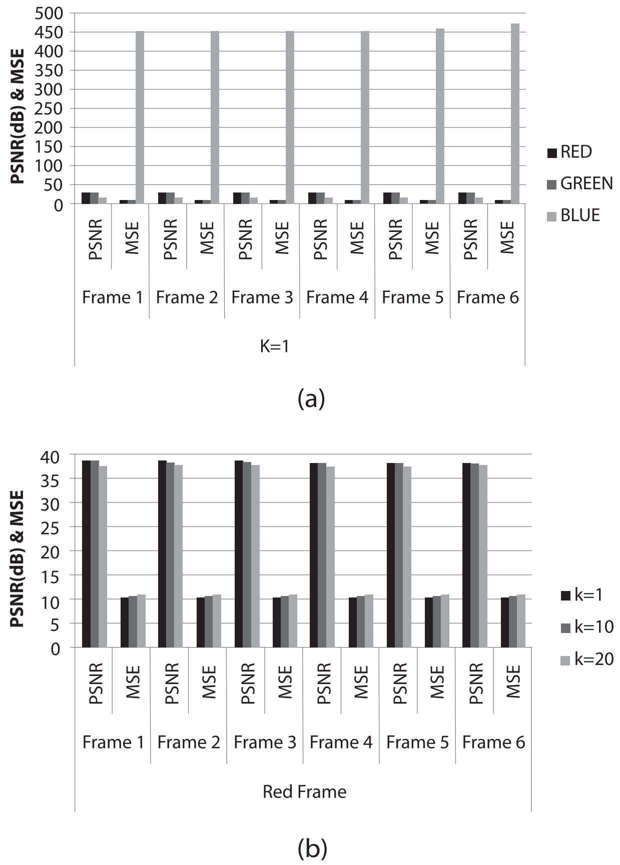


Fig. 9. Comparative Analysis of Method 2
(a) PSNR and MSE value for individual frame with k=1, (b) PSNR and MSE value for RED frame with various values of k

8. COMPARISION

The chart shown in Figure 10 shows the comparative analysis between the above two methods.

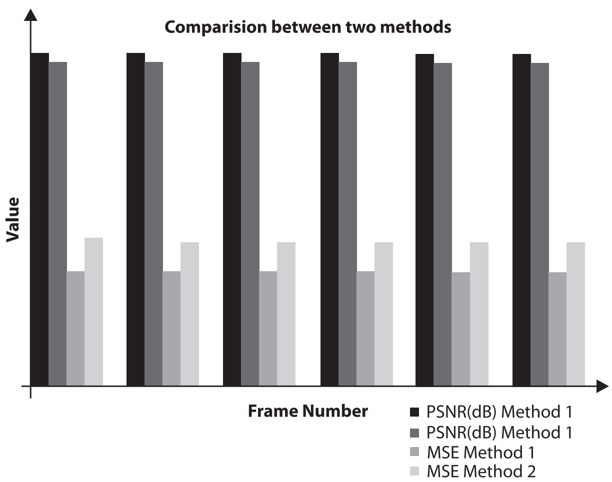


Fig. 10. Comparison chart of discussed methods

9. CONCLUSION

We have reached the conclusion that for the same block size and constant value the first method gives better results at the receiver end compared to the second. As the value of the constant increases, the watermark recovery is better at the cost of perceptibility of the watermarked video in both methods.

10. REFERENCES

[1] P.T. Rini, Review of Robust Video Watermarking Techniques, IJCA Special Issue on Computational Science - New Dimensions & Perspectives, No. 3, pp. 90-95, 2011.

[2] E. Koch, J. Zhao, Towards Robust and Hidden Image Copyright Labeling, Proc. of IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, Halkidiki, Greece, 20-22 June 1995, pp. 452-455, 1995.

[3] C.T. Hsu, J.L. Wu, Hidden Signatures in Images, Proc. of IEEE Int. Conf. on Image Processing, Lausanne, Switzerland, 16-18 September 1996, pp. 223-226, 1996.

[4] I.J. Cox, J. Kilian, T. Leighton, T. Shamoon, Secure Spread Spectrum Watermarking for Multimedia, Technical Report, 95-10, NEC Research Institute, 1995.

[5] H. Chiou-Ting Hsu, Ja-Ling Wu, DCT-based Watermarking for Video, IEEE Transactions on Consumer Electronics, Vol. 44, No. 1, February 1998, pp. 206-216, 1998.

[6] M. Kutter, F.A.P. Petitcolas, A Fair Benchmark for Image Watermarking Systems, Proc. of SPIE Conference, Electronic Imaging '99. Security and Watermarking of Multimedia Contents, Vol. 3657, San Jose, CA, USA, 25-27 January, pp. 226-239, 1999.