Digital video watermarking algorithm robust against video container format changes

S V Belim and P G Cherepanov
Dostoyevsky Omsk State University, Omsk, Russia

Abstract. A digital watermark embedding algorithm robust against any changes of a video container format with the use of a three-dimensional (3D) discrete cosine transform is proposed. For embedding the three-dimensional (3D) method based on Koch and Zhao's two-dimensional (2D) method is applied. The image of a digital watermark and its digital hologram as embedded information are used. Embedding is carried out in a YUV color scheme. Computer-simulation experiments are made. The proposed method proved to be robust against attacks of video container format changes.

1. Introduction

The problem of digital watermark (DW) robustness against any attacks on a video stream is becoming of high priority in terms of identifying the facts of unauthorized information copying. The main objective of an intruder is to destroy a digital watermark with minimum changes of the video container. The most widespread approach is to change a format of the container, i.e. its structure with the built-in identifying information. Digital watermarking is based on steganographic algorithms. In this regard the problem of developing the steganographic methods resilient against any transformations of a video container format is of utmost importance. This article deals with the algorithm of watermark embedding into a video stream therefore only video files are considered.

The most popular approach in watermarking is to break down or split a video stream into separate shots or frames, each being considered as an image and used as an independent container for a steganographic insert [1,2]. The approach exhibits high speed operation and a potential for parallelizing calculations. However, there is a problem of synchronizing the frames and it calls for putting in some additional information to arrange the extracted information precisely. Statistical characteristics of the frame [3,4], special synchronization marks [2] or excess coding [5] can be used as the additional information. Frame by frame digital watermarking has rather low resistance to any format changes since separate parts of a digital watermark are localized in a relatively small area resulting in high probability of altering a large number of bits simultaneously.

For higher resilience to any format changes of a video file the built-in data have to be evenly distributed throughout the whole video stream to reduce the chance of damaging a large number of bits of the digital watermark at the same time. Moreover, it is necessary to employ such methods of creating a digital watermark that allow restoring it with the use of incomplete data. The first problem can be solved by considering a video stream as a three-dimensional object and using three-dimensional transforms in steganographic algorithms. Applying a three-dimensional discrete cosine
transform (DCT) to a video stream was first offered by J. Roese in his work “Interframe cosine transform image coding” [6]. However, it was not utilized because of high requirements to the resources. Later, with the development of hardware, there appeared a number of researches focused on distribution of discrete cosine transform (DCT) coefficients and its influence on the quality of video stream encoding [7-10]. Several algorithms of putting steganographic inserts into the video stream being represented as a three-dimensional object have been developed. They are based on fixed and varying components of the wavelet along the temporal axis of the video [11], on three-dimensional Fourier transform [12,13], on three-dimensional wavelet-transform [14-16] and on 3D-DCT [17,18]. These works describe the methods of image brightness modulation (QIM - quantization index modulation) and image dither modulation (RDM - rational dither modulation).

To solve the problem of recovering a built-in digital watermark the methods of error correcting coding can be applied. Taking into account that the built-in information is traditionally assumed to be a graphic object the digital holography is more appropriate to employ due to its ability to restore the whole image from a small piece of a hologram [19].

This article is devoted to the development of a digital watermarking method (algorithm) robust against the video file format conversion.

2. Problem statement and solution

Let us consider the problem of embedding digital watermarks in a video stream with a YUV color model and its variants. This color model has been widely employed in video codecs and in main video standards, namely, NTSC, PAL, SECAM. Within the YUV color model each point has three defined components: one luma component (Y) and two chrominance components (U and V). The luma (brightness) component (Y) is black-and-white while chrominance components (U and V) comprise information for restoring the color of the image. We deal with only the luma component because it has minimal changes in video stream encoding.

Modified to 3D variant the Koch and Zhao method is used as an algorithm of digital watermaking [20]. A video stream is assumed to be a three-dimensional set of points where the first set defines a frame number, and the others (the second and the third ones) are the position of a point on the frame. For digital watermaking a whole video stream is split (broken down) into sets of N frames. Further each set is divided into areas of N x N pixels, from left to right and from up to down. Thus, there is a set of blocks of N x N x N size. If there are any blocks not filled completely, then they have to be omitted and not taken into account while embedding and extracting. Each obtained block is designated as \( f^i \) where \( i \) is a sequence number of this block. The spectral coefficients of discrete cosine transform are calculated for each of the blocks. The value of the spectral coefficient of discrete cosine transform for a point \((v, u, \kappa)\) is designated as \( f^i_{u,v,\kappa} \). The value of the luma component in the initial frame for a point \((x, y, z)\) is designated as \( f^i_{x,y,z} \).

The resulting equation follows:

\[
\begin{align*}
  f^i_{u,v,\kappa} &= \left( \zeta(v) \cdot \zeta(u) \cdot \zeta(\kappa) \right) \sqrt{\frac{8}{N^3}} \cdot \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sum_{z=0}^{N-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f^j_{x,y,z} \cdot \cos\left( \pi \cdot v \cdot \left( \frac{2 \cdot y + 1}{2 \cdot N} \right) \right) \cdot \cos\left( \pi \cdot \kappa \cdot \left( \frac{2 \cdot z + 1}{2 \cdot N} \right) \right) \\
  &\cos\left( \pi \cdot u \cdot \left( \frac{2 \cdot x + 1}{2 \cdot N} \right) \right) \\
\end{align*}
\]

where \( \zeta(v) \), for any point \( v \) is calculated as follows
\[ \zeta(\nu) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } \nu > 0 \\ 1, & \text{if } \nu = 0 \end{cases} \]  

(2)

As a result we obtain a set of three-dimensional matrixes of coefficients of the discrete cosine transform (DCT) for the allocated blocks in the initial video stream. The set can be used for hiding one bit of information. In a neighborhood of a point \((0,0,0)\) there are the coefficients of low-frequency components bearing information of the main part of the video stream. Digital watermarking into this area leads to noticeable visual distortions. In a neighborhood of a point \((N,N,N)\) there are the coefficients of high-frequency components corresponding to the fine details of the video stream. Digital watermarking into the area makes less noticeable visual distortions. But this set of coefficients is exposed to the greatest changes in the video stream if video format converting. To provide the maximum ratio between robustness of digital watermarking and its visibility the digital watermark (DW) is to be built into the area of mid-frequency components. Thus, two points from the relevant area are chosen \((v_1,v_1,k_1),(v_2,v_2,k_2)\).

Experimentally we determine a threshold value of coefficients modules difference which can provide the maximum ratio between visibility of a DW embedding and its robustness. A message of \(m\) bits in size will be used as a DW and the \(i\)th bit of the embedded DW is designated as \(m_i\). For embedding the message into the block \(f_{v_1,v_2,k}\) the values of additional functions have to be calculated first:

\[ \omega_i^1(v_1,v_1,k_1) = f_{v_1,v_1,k_1}^i, \]  

(3)

\[ \omega_i^2(v_2,v_2,k_2) = f_{v_2,v_2,k_2}^i, \]  

(4)

\[ Z_i^1(v_1,v_1,k_1) = \begin{cases} -1, & \text{if } f_{v_1,v_1,k_1}^i < 0 \\ 1, & \text{if } f_{v_1,v_1,k_1}^i \geq 0 \end{cases}, \]  

(5)

\[ Z_i^2(v_2,v_2,k_2) = \begin{cases} -1, & \text{if } f_{v_2,v_2,k_2}^i < 0 \\ 1, & \text{if } f_{v_2,v_2,k_2}^i \geq 0 \end{cases}. \]  

(6)

For digital watermarking the recalculation of DCT coefficients has to be completed:

\[ \sigma_i^1(v_1,v_1,k_1) = \begin{cases} P + \omega_i^2(v_2,v_2,k_2) + 1, & \text{if } \omega_i^1(v_1,v_1,k_1) \leq P \text{ and } m_i = 0 \\ \omega_i^1(v_1,v_1,k_1), & \text{if } \omega_i^1(v_1,v_1,k_1) > P \text{ and } m_i = 1 \end{cases}, \]  

(7)

\[ \sigma_i^2(v_2,v_2,k_2) = \begin{cases} P + \omega_i^1(v_1,v_1,k_1) + 1, & \text{if } \omega_i^2(v_2,v_2,k_2) \geq P \text{ and } m_i = 1 \\ \omega_i^2(v_2,v_2,k_2), & \text{if } \omega_i^2(v_2,v_2,k_2) < P \text{ and } m_i = 0 \end{cases}, \]  

(8)

\[ \tilde{f}_{v_1,v_1,k_1}^i = Z_i^1(v_1,v_1,k_1) \cdot \omega_i^1(v_1,v_1,k_1), \]  

(9)

\[ \tilde{f}_{v_2,v_2,k_2}^i = Z_i^2(v_2,v_2,k_2) \cdot \omega_i^2(v_2,v_2,k_2). \]  

(10)

To obtain the resulting video stream an inverse DCT has to be carried out:
Then all the received matrixes have to be collected in a video stream: from left to right, from up to down.

Extracting a digital watermak requires a similar procedure, namely, splitting the received video stream into blocks \(N \times N \times N\) and applying DCT to them. The extraction of the symbol \(m_i\) built into the DW is implemented according to the following rule:

\[
\omega_j^{i}(u_1, v_1, \kappa_1) = \left| f_{u_1, v_1, \kappa_1} \right|, \quad (12)
\]

\[
\omega_j^{i}(u_2, v_2, \kappa_2) = \left| f_{u_2, v_2, \kappa_2} \right|, \quad (13)
\]

\[
m_i = \begin{cases} 
0, & \text{if } \omega_j^{i}(u_1, v_1, \kappa_1) > \omega_j^{i}(u_2, v_2, \kappa_2) \\
1, & \text{if } \omega_j^{i}(u_1, v_1, \kappa_1) < \omega_j^{i}(u_2, v_2, \kappa_2)
\end{cases}, \quad (14)
\]

As a result the following algorithm of embedding a DW into three-dimensional area turns out:

Step 1. Allocate sets of \(N\) frames in an initial video stream.

Step 2. Divide each set of frames into blocks in \(N \times N \times N\) size.

Step 3. For each block execute DCT (1) and obtain a three-dimensional matrix of coefficients.

Step 4. Take any two points along a minor diagonal \((0,0,0) \rightarrow (N,N,0)\) i.e. from a mid-frequency component area.

Step 5. Calculate the optimum value of a threshold \(P\) providing maximum ration between the invisibility of embedding and its robustness.

Step 6. Embed each bit of the digital watermark into a relevant matrix of DCT coefficients (9)-(10).

Step 7. Apply an inverse DCT(11) to obtain a modified set of blocks.

Step 8. Collect a set of \(N\) frames from the modified set of blocks.

Step 9. Assemble a video stream from the collected frames.

The end of the algorithm.

In order to extract a digital watermark it is required to execute the first 3 steps of the algorithm and apply the equation (14) to each of the received matrixes of coefficients.

3. Obtaining a digital watermark hologram.

As it has already been stated and proved in the previous work [21] to increase the video stream robustness against any modifications it is possible to build in not a digital watermark but its digital hologram. The effect is due to a physical feature of the hologram that is each point of it comprises data on all the points of the initial image. Thus, one can restore the whole image (with some losses) having a small piece of the hologram.

To create a digital hologram a coherent beam of light has to be divided into two beams – an object beam and a reference one. The object beam has to be directed onto an initial DW so that its reflection goes onto a resulting photographic plate, and the reference beam is immediately sent directly onto this
plate. An interference pattern of the two beams of light on the photographic plate is recorded. To increase reliability of the hologram it is scaled up twice along each axe compared with the initial DW. Further the intensity of light from each point of the watermark is found on the basis of the following formula: (15)

\[ \varepsilon_{\text{rel}}(\bar{x}, \bar{y}) = \sum_{x=0}^{N} \sum_{y=0}^{N} I(x, y) \times 3 \times \cos \frac{d}{\lambda}, \]  

where the watermark point and the hologram point are meant to be coordinates of pixels in the images, \((\bar{x}, \bar{y})\) being a pixel position (coordinates) in the hologram while \((x, y)\) being a pixel position (coordinates) in the initial digital watermark;

\[ I(x, y) \] – the intensity of light of an initial watermark;

\[ \lambda \] – the wavelength;

The distance between the points of the watermark and the hologram is:

\[ d = \sqrt{(x \ast 2 - \bar{x})^2 + (y \ast 2 - \bar{y})^2 + (z \ast 2 - \bar{z})^2}, \]  

where \((z \ast 2 - \bar{z})^2\) equals unity since the hologram is two-dimensional.

The resulting algorithm of creating a digital watermark hologram is:

Step 1. For each point of a hologram put the intensity of the reference beam of light equal half the initial intensity, i.e. value of 128.

Step 2. Set light source wavelength.

Step 3. For each point of the hologram calculate the intensity of object light, namely, calculate the value of formula (15).

Step 4. Summarize the calculated intensity of object light and the intensity of reference light for each of the hologram points.

The end of the algorithm.

In order to recover an initial digital watermark the above algorithm steps have to be applied with the intensity of the reference beam equal 0.

4. Computer experiment results

Within the computer experiment the robustness of the proposed algorithm of embedding a digital watermark against color scheme changing was investigated. A video stream of .mp4 format (resolution 512 x 512, h.264 codec, YUV422p color model) was taken as a container. A monochrome image of "©" (as a watermark) was built into this video stream. A luma channel (Y) was used for embedding. The experiment was consistently carried out with embedding the digital sign and its digital hologram. Then code conversion of the color model of the video file into YUV420 was made. The size of the block was chosen \(N = 8\) since a large number of various algorithms of processing video images use the blocks in size \((8 \times 8 \times 8)\) so that resistance of the algorithm against any transformations increases. The aforementioned threshold \(P\) value will be determined experimentally.

Being a standard method of assessing correlation between two images a normalized correlation (NC) was used as a criterion of reliability of a digital watermark [22]. This measure accepts values from 0 to 1 where 1 turns out only in case the images coincide totally. The measure is defined as follows:

\[ \text{NC} = \frac{\sum \sum W(i, j) \cdot \tilde{W}(i, j)}{\sum \sum [W(i, j)]^2}, \]
where $W(i,j)$ and $\tilde{W}(i,j)$ are values of pixel brightness in the built-in and the extracted watermarks respectively.

Being a standard of measuring distortions a *peak signal to noise relation* (PSNR) between an initial video stream and a coded one was calculated as a measure of invisibility (obsccurity) of the watermark [23]. It is measured in decibels. The more the value, the fewer distortions have been brought into the image. The distortions of more than 40 dbs are virtually invisible to the human eye. This measure is calculated as follows:

$$
\text{PSNR} = 20 \log_{10} \frac{255}{\sqrt{\text{MSE}}},
$$

where

$$
\text{MSE} = \frac{1}{MK} \sum_{i} \sum_{j} (P(i,j) - \tilde{P}(i,j))^2,
$$

$P(i,j)$ and $\tilde{P}(i,j)$ are brightness of a pixel in an initial video stream and in a stream with a watermark respectively, while $M$ and $K$ are frame width and height in the video stream.

The averaged results of PSNR and NC ratio for various values of threshold $P$ for the proposed method with direct embedding a digital watermark and with embedding its hologram are presented in table 1.

**Table 1. Results of PSNR and NC ratio for the proposed methods**

| Threshold value/size | PSNR   | NC    |
|----------------------|--------|-------|
|                      | Method without hologram |     |
| $P = 36$            | 36.2609db | 0.568 |
| $P = 37$            | 36.1839db | 0.5897|
| $P = 38$            | 36.1064db | 0.6132|
|                      | Method with hologram |     |
| $P = 36$            | 36.2609db | 0.9589|
| $P = 37$            | 36.1839db | 0.9639|
| $P = 38$            | 36.1064db | 0.9656|

As seen from the results (Table 1) the most optimum value for the threshold is $P = 37$ since at given value the greatest product of the normalized correlation and the peak relation of the signal to the noise for the proposed method with hologram is reached.

**Table 2. Results of PSNR values for the considered methods**

| Method                        | PSNR value |
|-------------------------------|------------|
| Method with hologram          | 36.1839db  |
| Method without hologram       | 36.1839db  |
| Work [6]                      | 32.5876db  |
| Work [7]                      | 35.4263db  |
The results of the experiments with different methods show that at value of the threshold \( P = 37 \) the proposed method brings fewer distortions than the methods based on wavelet-transforms (Table 2). Nevertheless, the distortions brought by the proposed method are still visually distinguished in bright fragments of the video frame.

**Table 3.** Values of normalized correlation after changing a color model of the video file with the embedded digital watermark

| Color model | Value of normalized correlation with hologram | Value of normalized correlation without hologram |
|-------------|---------------------------------------------|-----------------------------------------------|
| YUV420      | 0.3738                                      | 0.2424                                        |

The obtained results demonstrate a visible gain of normalized correlation value in the method with a hologram in comparison with the method without it. The gain is accounted for the hologram’s physical property allowing to reconstruct the whole initial image from the piece of a hologram.

Visual distortions of the digital watermark after changing a color format/model are presented in Figure 1.

![Figure 1](image)

**Figure 1.** Visual distortions of the digital watermark after changing a format of the video file:

a) the initial image of the digital watermark, b) the extracted digital watermark without the use of the hologram, c) the extracted digital watermark with the use of the hologram, d) the extracted digital watermark after transformation into YUV420 model without the use of the hologram, e) the extracted digital watermark after transformation into YUV420 model with the use of the hologram.

The graphic results prove that the proposed method with hologram provides both formal and visually distinguishable gain of reliability of the watermark extracted encoding the video stream into other formats.

5. Conclusion

These results show that the proposed method is robust against the attacks of changing a color model and a video stream encoding format. This effect is due to embedding a DW into the channel of brightness (luma component) containing "black-and-white" image. The DW embedding into U or V components (chrominance components) brings fewer visual distortions, but it doesn't make sense since this approach is too vulnerable (these components are subjected to compression in various color models leading to losses in the extracted DW. Moreover, comparing with other works [6], [7], the proposed method provides higher robustness and obscurity (invisibility) of embedding (digital watermarking). Nevertheless, the DW embedding is visually distinguished in the bright fragments (areas) of images (video frames) resulting in limitations of the algorithm application.
References

[1] E T Lin and E J Delp 2004 Temporal synchronization in video watermarking *IEEE Transactions on Signal Processing* v 52(10) p 3007-3022

[2] D Delannay and B Macq 2004 Classification of watermarking schemes robust against loss of synchronization *Proceedings of SPIE* 5306, *Security, Steganography and Watermarking of Multimedia Contents VI* p 581-591

[3] C Chen, J Ni and J Huang 2009 Temporal statistic based video watermarking scheme robust against geometric attacks and frame dropping *Digital Watermarking* Springer p 81-95

[4] S W Sun and P C Chang 2003 Video watermarking synchronization based on profile statistics *Proceedings of IEEE 37th Annual 2003 International Carnahan Conference on Security Technology* p 410-413

[5] V A Mitekin and V A Fedoseev 2014 Method of embedding information in a video that is resistant to sync loss errors *Computer optics.* v38 p564-573

[6] J Roese, W Pratt and G Robinson 1977 Interframe cosine transform image coding *IEEE Transaction on Communication* v 25(11) p 1329–1339

[7] M Bauer and K Sayood 1995 Video coding using 3 dimensional DCT and dynamic code selection *Proceedings of Data Compression Conference* p 451.

[8] M Servais and G de Jager 1997 Video compression using the three dimensional discrete cosine transform (3D-DCT) *Proceedings of the South African Symposium on Commun. and Signal Process* p 27–32

[9] R K W Chan and M C Lee 1997 3D-DCT quantization as a compression technique for video sequences *Proc. of International Conf. on Virtual Sys. and MultiMedia* p 188–196

[10] N Bozinovic and J Konrad 2005 Motion analysis in 3D DCT domain and its application to video coding *Signal Processing: Image Communication* v 20 p 510–528

[11] M Swanson, B Zhu and A T Tewfik 1998 Multiresolution scene-based video watermarking using perceptual models *IEEE Journal on Sel. Areas in Comm* v 16(4) p 540-550

[12] F Deguillaume, G Csurka, J O’Ruanaidh and T Pun 2000 Robust 3D DFT video watermarking *Proc. SPIE, Security and Watermarking of Multimedia Content II* v 3971 p 346-357

[13] H Liu, N Chen, J Huang, X Haung and Y Q Shi 2002 A robust DWT-based video watermarking algorithm *IEEE International Symposium on Circuits and Systems* p 631-634

[14] M Kucukgoz, O Harmanci, M K Miheak and R Venkatesan 2005 Robust Video Watermarking via Optimization Algorithm for Quantization of Pseudo-Random Semi-Global Statistics *Proc. SPIE, Security, Steganography, and Watermarking of Mult. Cont. VII* v 5681

[15] P Campisi and A Neri 2005 Video watermarking in the 3DDWT domain using perceptual masking *IEEE Int. Conference on Image Processing* (Genou, Italy)

[16] P Campisi 2005 Video watermarking in the 3D-DWT domain using quantization-based methods *IEEE International Workshop on Multimedia Signal Processing* (Shanghai, China)

[17] J H Lim, D J Kim, H T Kim and C S Won 2001 Digital video watermarking using 3D-DCT and Intra-Cubic Correlation *Proc. SPIE, Security and Watermarking Contents III* v 4314 p 54-72

[18] P Campisi and A Neri 2007 3D-DCT video watermarking using quantization-based methods *15th European Signal Processing Conference (EUSIPCO 2007)* (Poznan, Poland) copyright by EURASIP

[19] M V Smirnov 2005 Holographic approach to embedding holographic watermarks into a photo image *Optical magazine* v 72(6) p 51–56

[20] E Koch and J Zhao 1995 Towards robust and hidden image copyright labeling *IEEE Workshop on Nonlinear Signal and Image Processing* p 452-455

[21] S V Belim and P G Cherepanov 2017 Embedding digital watermarks in a video stream based on a 3-d discrete cosine transform *VKIT* v 5 p 28–34
[22] K A Navas M Sasikumar and S Sreevidya 2007 A benchmark for medical image watermarking
Proceedings of 14th International Workshop on Systems, Signals and Image Processing (Maribor, Slovenia) p 237-240

[23] Q Huynh-Thu and M Ghanbari 2008 Scope of validity of PSNR in image/video quality assessment
Electronics Letters v 44 p 800-801