A Robust Blind Video Watermarking Scheme based on Discrete Wavelet Transform and Singular Value Decomposition

Amal Hammami1, Amal Ben Hamida2 and Chokri Ben Amar1

1Research Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax, Sfax, 3038, Tunisia
2Department of Computer Engineering, College of Computers and Information Technology, Taif University, Saudi Arabia

Keywords:  Blind Video Watermarking, Discrete Wavelet Transform, Singular Value Decomposition, Robustness.

Abstract: The outgrowth in technological world has massively promoted to information fraud and misappropriation by the ease of multimedia data content regeneration and modification. Consequently, security of digital media is considered among the biggest issues in multimedia services. Watermarking, consisting in hiding a signature known as watermark in a host signal, is one of the potential solutions used purposely for media security and authentication. In this paper, we propose a robust video watermarking scheme using Discrete Wavelet Transform and Singular Value Decomposition. We embed the watermark into the mid frequency sub-bands based on an additive method. The extraction process operates following a blind detection algorithm. Several attacks are applied and different performance metrics are computed to assess the robustness and the imperceptibility of the proposed watermarking. The results reveal that the proposed scheme is robust against different attacks and achieves a good level of imperceptibility.

1 INTRODUCTION

In day-to-day life, the tremendous improvement in computer technology field is leading to many problems for the multimedia industry. Indeed, many data files, such as video files, can be easily copied, distributed and tampered without compromising the quality of the data. Thus, a broad range of approaches has already been proposed to secure video sequences content. Cryptographic techniques based on secret key are used to encrypt data in order to secure the visual content (Omar and Shawkat, 2018; Vikrant and Shubhanand, 2016). Nevertheless, it is pointed out that these techniques come together in hands with some ambiguities. In fact, their main drawbacks derive from the non-preservation of the videos original formats. To tackle this problem, video watermarking techniques provide a promising security solution. A video watermarking scheme includes predominantly two steps: the embedding process which is the process of incorporating an imperceptible data (a binary image, a bits sequences, etc.), known as watermark, into cover video and the detection process which is the process of extracting the inserted data from the watermarked video (Asikuzzaman, and Pickering, 2018). Digital watermarking system has been a very interesting research area in many applications such as copyright protection, data authentication, etc (Tuan and Duan, 2015; Charfeddine et al, 2014). Hence, each watermarking system should have its intrinsic properties regarding the given application (Ben Hamida et al., 2011; Zhang et al., 2012; Ahuja and Bedi, 2015; Tarhouni et al., 2018; Jyothika and Geetharanjin, 2018; Koubaa et al, 2012). Generally, there are three important requirements considered in the most practical video watermarking systems (Asim et al., 2015; Arti and Ajay, 2017). The first one is the imperceptibility. It refers to the visual quality of the watermarked video, which should perceptually be as close to the original video as possible. The second requirement is the robustness. It denotes the ability of watermark to sustain unintentional and intentional attacks. Capacity, which is defined as the amount of the secret information concealed in the host video, is the third required property. Imperceptibility,
robustness and capacity are inversely related. Hence, it is very important to maintain a good trade-off between all the watermarking properties (Hedayath et al., 2016). The watermarking techniques are commonly divided into two main classes according to the embedding domain criterion; the spatial domain techniques and the frequency domain ones (Arti and Ajay, 2017; Amit and Navdeep, 2016; Asim et al., 2015). In the first class, the watermark is inserted by modifying the pixel values of the host video directly. For the second class, the host video frames undergone a domain transformation technique and the watermark is inserted into selective coefficients from these frames afterwards. The spatial domain-based techniques present low computational complexity. Nevertheless, they are not reliable in the presence of different image processing methods and have low-bit capacity. However, the watermarking techniques in the frequency domain are comparatively more resistant to common distortions including noise addition, lossy compression and rotation. Besides, they can effectively achieve the compromise between imperceptibility and robustness requirements of digital watermarking techniques. With much interest to this, we propose in this work a blind and robust video watermarking scheme in the frequency domain based on a cascade of two transformations: Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD).

The organization of the remainder of this paper is as follows. Section 2 provides a review on existing frequency domain video watermarking techniques. The proposed watermarking scheme is described in Section 3. Experimental results as well as a comparison with others techniques are given in Section 4. The last section summarises and concludes this work.

2 RELATED WORK

In the literature, a variety of video watermarking techniques has been proposed and can be categorized into different classes based on different criteria such as working domain, watermark visibility, watermark robustness, etc. As we previously mentioned, in frequency domain watermarking the embedding process is preceded by the application of some transformation methods to the cover video frames. Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) as well as Non-subsampled Contourlet Transform (NSCT) and Singular Value decomposition (SVD) are the most used methods in this domain (Wali et al, 2010; Guedri et al, 2011; Othmani et al, 2010). In this section, we will focus on digital video watermarking in the transform domain.

Some video watermarking schemes can operate using one single transformation technique. Indeed, a DWT based watermarking algorithm has been presented in (Mostafa and Ali, 2016). This video watermarking scheme satisfies the imperceptibility requirement exploiting the human visual system characteristics. In fact, blocks with highest motion vector magnitude are selected to hide the watermark after being transformed using the discrete wavelet transform. Another watermarking technique using the discrete wavelet transform has been proposed in (Dolley and Manisha, 2018). In this method, the watermark is embedded only on scene-changed frames chosen using a scene-change detector algorithm. Therefore, those frames are converted to the grey-scale and decomposed using the three level DWT. Then the watermark bits are inserted in the low frequency components using an additive method. The authors in (Tuan and Duan, 2015) proposed a blind approach based on discrete cosine transform. The luminance component Y extracted from the host video frame is decomposed into DCT 8x8 blocks. The watermarks bits are subsequently hidden on several randomly chosen DCT coefficients using even-odd quantization algorithm.

On the other hand, video watermarking schemes can also perform based on a combination of domain transformation techniques. An example of hybrid video watermarking technique based on non-subsampled contourlet transform and singular value decomposition is introduced in (Narasimhulu, 2017). In this approach, both of the original video frame and the watermark are decomposed by the non-subsampled contourlet transform. Then, the singular values of watermark are incorporated in those of original video using an additive non-blind algorithm. In (Naved, 2016), a robust video watermarking technique using four distinct domain transformation techniques is proposed based on DWT, SVD, FFT and DCT. The watermark embedding process is done using an additive method. A third example of watermarking schemes involving two transformation techniques is introduced in (Jeemananda and Prince, 2016). Here, authors proposed a video watermarking in four approaches all based on discrete wavelet transform and singular value decomposition. The four methods differ from each other by the number of frames to be used for the embedding process. In the third approach, authors used a scene change detector in order to select the embedding medium. In fact, the insertion occurs in the singular value corresponding...
to the low frequency sub band extracted from only detected scene change frames.

3 PROPOSED APPROACH

The proposed system is a frequency domain video watermarking using both discrete wavelet transform and singular value decomposition. Using these two domain transformation techniques allows improving the watermark robustness by exploiting their complementary properties. The proposed system consists of two stages: embedding process and extraction process. The block diagrams of these processes are shown in figure 1 and figure 2 respectively and explained separately in the two following sub-sections.

3.1 Embedding Process

As presented in figure 1, the video is split into sequences of N consecutive frames. The number N that defines the video sequence size is experimentally determined. It is used as a secret key to improve the security level of the watermarking technique. Within each video sequence, the same watermark is inserted into all frames in order to enhance the robustness against frame dropping attack. A different binary bits message is used as a watermark for every video sequence. All frames in each sequences are subject for the following process.

Since pixel values are highly correlated in RGB color space, the RGB frame is converted into YUV space. Taking into account that Human Visual System cannot notice the changes in regions of high luminance, only the luminance component Y is selected for the watermark incorporation. This component is subdivided in 4*4 non-overlapping blocks. Then, one level Discrete Wavelet Transform is applied to transform each block to the frequency domain. This operation generates four sub-bands: LL1, HL1, LH1 and HH1. The high frequency component HH1 contains the least significant parts of the video frame, a potential information loss can occur during compression. The low frequency sub-band LL1 contains the most significant information of the video frame; its modification degrades the visual quality. So, the mid frequency components LH1 and HL1 are selected as the best locations for watermarking to meet the trade-off between the robustness and the imperceptibility requirements.

The chosen sub-bands are decomposed using the singular value decomposition into three independent matrices U, S and V. Only S matrix is implied into the embedding process. Indeed, a slight modification in the singular values does not yield a large visual alteration in the host video. Besides, they exhibit attractive properties such as rotation invariance, transposition invariance and translation invariance.

Using blocks of size (4x4) to embed the watermark bits enables to obtain S matrix containing few non-zero values. Hence, the order relationship between the different singular values can be rigorously respected.

Next, the watermark is inserted in the S matrices corresponding to the two sub-band HL1 and LH1. The embedding process operates according to the following formulas:

If W = 0

\[ S_w(0,0) = S(0,0) + k_\alpha \]

\[ S_w(1,1) = S(0,0) \]  (1)

Else

\[ S_w(0,0) = S(1,1) + k_\beta \]

\[ S_w(1,1) = S(1,1) \]  (2)

With W is the watermark bit to be inserted, S_w is the modified version of the original matrix S.

\[ k_\alpha = \frac{S(0,0) + S(1,1)}{\alpha} \]

\[ k_\beta = \frac{S(0,0) + S(1,1)}{\beta} \]  (3)

According to (5) and (6), k_\alpha and k_\beta are proportional to the S coefficients, which allows avoiding the perceptual distortion and increasing the watermark robustness. The values of \( \alpha \) and \( \beta \) employed in our work will be explained in the sub section 4.1.

In order to obtain the watermarked luminance component Y, the inverse SVD operator is applied on the modified matrix S_w followed by the inverse DWT. This watermarked Y is merged with the unchangeable U and V components and converted to RBG color to obtain the final watermarked video frame. The above-described process is repeated to all frames in each sequence to finally construct the watermarked video.
3.2 Extraction Process

The extraction process as depicted in figure 2 is the reverse of the embedding process. The detection algorithm is blind. Therefore, the scheme does not require the original non-watermarked video during the extraction of concealed watermark.

The watermarked video is partitioned into sequences of N frames. Then all frames of each sequence are processed as described below to extract the correspondent watermark. At first, the considered video frame is converted from the RGB to YUV color space. The luminance component Y is extracted and decomposed into 4x4 blocks. Next, the discrete wavelet decomposition is performed on each block, obtaining respectively LL1, LH1, HL1 and HH1. After applying the singular value decomposition to both LH1 and HL1 sub-bands, watermark bits are extracted from $S_{ext}$ matrices based on the following rules:

If $S_{ext}(0,0) - S_{ext}(1,1) > \frac{k_a + k_\beta}{2}$

$W_{ext} = 0$  \hspace{1cm} (7)

Else

$W_{ext} = 1$  \hspace{1cm} (8)

With $S_{ext}$ is the extracted matrix $S$, $W_{ext}$ is the extracted watermark bit, $k_a$ and $k_\beta$ are computed using (5) and (6) equations mentioned in the sub-section 3.1

4 EXPERIMENTAL RESULTS

To verify the robustness and the imperceptibility of the proposed video watermarking algorithm, simulation experiments are conducted on different standard videos. The considered input videos are presented in table 1.

The watermarks used in this work are binary sequences. For every frame in each video sequence, the maximum capacity is equal to the total number of 4x4 blocks resulting from the decomposition applied
on the corresponding luminance component Y. For a 256×256 frame, the maximum possible size of the inserted message is 4096 bits, which indicates the large capacity of proposed watermarking technique.

The number N that defines the size of each video sequence, obtained after splitting the host video, is fixed after evaluating the robustness of the watermarking technique by varying the N value. It has been notified experimentally that for each video the adequate choice for this parameter is the frame per second (FPS) value.

Table 1: Details of the used videos for simulation.

<table>
<thead>
<tr>
<th>Video Name</th>
<th>Number of Frames</th>
<th>Frame Per Second</th>
<th>First Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>stefan.avi</td>
<td>300</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>foreman.avi</td>
<td>294</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>paris.avi</td>
<td>3064</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>coastguard.ai</td>
<td>294</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>bus.avi</td>
<td>148</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Imperceptibility Results

The imperceptibility requirement is examined using the Peak Signal to Noise Ratio (PSNR), and the Structural Similarity index (SSIM) (Asim et al., 2015; Kadu et al., 2016). The Peak Signal to Noise Ratio is used to find out the quality of the watermarked video with respect to human view. This measure is defined as:

$$\text{PSNR} = 10 \times \log \left( \frac{L^2}{\text{MSE}} \right)$$  (9)

Where L is the maximum pixel value in the corresponding frame and MSE is the mean square error.

The SSIM measures the similarity between two images. It is calculated using the following formula:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c)(2\sigma_{xy} + d)}{(\mu_x^2 + \mu_y^2 + c)(\sigma_x^2 + \sigma_y^2 + d)}$$  (10)

Where $\mu_x$ and $\sigma_x^2$ are respectively the average and the variance of the intensities available in the original frame, $\mu_y$ and $\sigma_y^2$ are respectively the average and the variance of the intensities available in the original frame and $\sigma_{xy}$ is the covariance of original and watermarked frames, c and d are two variables used to stabilize the division.

For each tested video, PSNR and SSIM are determined by computing the average of PSNR and SSIM values of all video frames.

To deduce the suitable values for $(\alpha;\beta)$ used in equations (5) and (6), we measure the visual similarity between the original video and the watermarked one. For this purpose, we calculate the Peak Signal to Noise Ratio (PSNR) by varying $(\alpha;\beta)$ values. The graph shown in figure 3 reveals that the couple $(2;4)$ achieves the highest values of PSNR for all used videos. Therefore, $\alpha=2$ and $\beta=4$ are the approved values for the performances evaluation.

Figure 3: PSNR values obtained for different values of $(\alpha;\beta)$.

Figure 4 displays the first frames of two host videos and their corresponding watermarked versions. It can be seen that the watermark is completely transparent and there is no visual distinction between frames of non-watermarked and watermarked videos. The obtained PSNR values, which are illustrated in figure 5, vary between 33.9109 dB and 48.6773 dB. It indicates that the watermarked videos have a good visual quality. Likewise, the SSIM values are approaching towards 1, as illustrated in figure 6, which proves the high similarity between the host videos and the watermarked ones. Hiding the watermark into the S component of SVD applied to the mid frequency sub bands of DWT presents a reasonable choice that enables to obtain this good level of imperceptibility.

4.2 Robustness Results

The robustness requirement is scrutinized computing the normalized correlation (NC) (Asim et al., 2015; Kadu et al., 2016).
The NC assesses the similarity between the original watermark and the watermark extracted from the attacked frame. It formula is provided below:

$$ NC = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} W(i,j)W'(i,j)}{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} W^2(i,j)} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} W'^2(i,j)}} $$ (11)

Where $W(i,j)$ and $W'(i,j)$ are respectively the original watermark bit and the extracted one and $m$ and $n$ are the watermark dimensions.

To test the performance of the proposed watermarking scheme in term of robustness, different distortions and attacks are firstly performed on the watermarked videos. Secondly, the NC is calculated after the extraction process.

Figure 7 exhibits the NC values of attacked videos. According to this figure, our watermarking scheme is robust to salt and pepper and Gaussian noise attacks. In fact, the NC reaches 0.99364 and 1 after applying respectively salt and pepper and different white Gaussian noise mean. This high correlation is guaranteed by involving the discrete wavelet transform, which is resilient to noise adding.

Moreover, it is noticed that the proposed scheme is immune to rotation attack. In fact, by varying the rotation degrees the watermark is successfully extracted. The obtained NC values against this attack are up to 0.99976. This robustness is achieved due to the use of the singular value decomposition SVD that is invariant to geometric attacks and especially to rotation attack.

Besides, results in this figure demonstrate the robustness against median filter attacks and cropping attacks. Indeed, the NC corresponding to cropping...
attack is superior than 0.94 and attains 0.99988. Concerning the median filter attack, the obtained NC values vary between 0.941314 and 1.

Experimental results prove that the proposed scheme is compression resilient by providing NC values up to 0.99987. This high level of robustness is reached thanks to the embedding of the watermark in the mid frequency sub bands of the discrete wavelet transform.

The redundant embedding of a full watermark in every frame of each video sequence reinforces the resilience of the proposed scheme against the frames dropping attack. Indeed the minimal NC value after removing 32% of the frames in every sequence is 0.95335.

4.3 Comparison with Existing Methods

In this section, we will compare the performance of our proposed technique with three existing ones; (Dolley and Manisha, 2018), (Narasimhulu, 2017) and (Jeebananda and Prince, 2016).

According to NC results presented in table 2, it is noticed that our algorithm and the technique proposed in (Jeebananda and Prince, 2016) withstand Gaussian noise attack. In fact, our proposed scheme efficiently resists to this attack with the highest value of NC that is equal to 1. However, the method presented in (Dolley and Manisha, 2018) has a poor robustness to Gaussian noise attack. Besides, both our technique and (Jeebananda and Prince, 2016) one successfully survive the median filter attack and our technique shows the better NC value that is 1. All approaches cited in the table 2 demonstrate their robustness against salt and pepper and rotation and the best NC values are provided by our technique.

Moreover, it is proven that our scheme provides high level of robustness to MJPEG compression compared to the techniques proposed in (Jeebananda and Prince, 2016) and (Dolley and Manisha, 2018) by offering a NC value that reaches 0.99987. Regarding cropping attack, our method shows better robustness than (Narasimhulu, 2017) with a NC value equal to 0.99976. On the other hand, the method presented in (Jeebananda and Prince, 2016) and the proposed one have a similar resilience to frames dropping attacks.

5 CONCLUSION

In this paper, a blind robust video watermarking scheme based on discrete wavelet transform and singular value decomposition was proposed. The S component of the singular value decomposition (SVD), which is applied to the mid frequency sub bands of the discrete wavelet transform (DWT), is used in the embedding process in order to achieve the best trade-off between the imperceptibility and the robustness requirements. Normalized Correlation Coefficient (NC), as well as Peak Signal to Noise Ratio (PSNR) and Structural Similarity index (SSIM) are computed to scrutinize the proposed technique performance. Experimental results prove that the proposed scheme successfully sustains several attacks namely geometrical, image processing and compression. Comparing it with others techniques, the proposed one shows high level of robustness. In term of imperceptibility, the quality of the video is maintained. Hence, it can be concluded that the proposed approach is efficiently suitable for applications, which require more robustness than imperceptibility. Thereby, the future works will focus on exploiting the developed technique for video authentication goal in video surveillance context.

ACKNOWLEDGMENTS

The research leading to these results received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

REFERENCES


