Robust Distance-Based Watermarking for Digital Video

A. S. Abdul-Ahad*, Maria Lindén†, Thomas Larsson‡
School of Innovation, Design and Engineering
Mälardalen University
Västeras, Sweden

Waleed A. Mahmoud§
Department of Electrical Engineering
Al-Isra Private University
Amman, Jordan

Abstract

In this paper, a mechanism of a distance-based algorithm is highlighted and tested to invisibly watermark a digital video (cover object) using an iconic image (watermark object). The algorithm is based on the distances among the addresses of values of the cover object. These distances are used to make the embedding. The order of manipulating these distances are specified by the values of the watermark data which is dealt with serially. The algorithm achieves a self encryption key. Each watermark object has its unique pattern of distances at different possible lengths of distance bits. This enhances the complexity of sequential embedding. The blind and non-blind algorithm are tested using direct (spatial) and first level Two Dimensional Discrete Wavelet Transform (2D DWT) embeddings. The algorithm shows resisting and withstanding against the most important attacks. Some of these include lossy compression, frame averaging, frame swapping, and frame dropping.

CR Categories: I.4.9 [Computing Methodologies]: Image Processing and Computer Vision—Applications; E.4 [Coding and Information theory]

Keywords: watermarking, digital video, iconic image, direct embedding, distance bits, 2D DWT embedding, attacks

1 Introduction

Nowadays, the e-security is becoming an urgent requirement and among its impressive weapons are watermarking techniques. The forthcoming applications like wireless broadband access, Multimedia Messaging Service (MMS), video chat, mobile TV, HDTV content, Digital Video Broadcasting (DVB), minimal service like voice and data, and other streaming services for ”anytime-anywhere” are sooner or later at the door. This prosperous world of digital multimedia communication encounters undoubtedly an outburst of malicious interventions (i.e., copyright infringement like piracy, fraud, forgery, copying, etc). This situation necessitates the design of reliable and robust systems to protect and preserve the integrity and safe passage of any form of data. Invisible watermarking is one of the techniques that are available. Sooner or later this technique will be widely used in the Internet [Hui and Yeung 2003; Cox and Miller 2001; Maes et al. 2000].

The integrity system must include authorization, authentication, privacy, encryption and copyright protection policies. These policies can be entirely or partially watermarked in the original version of any multimedia application that need to be maintained against any abusers and hackers. Watermarking applications such as signature, trade mark, logo, biometrics like fingerprint, iris, voice and so on, are forming the main tools to achieve these policies. It is trustworthy to say that the watermarking technique gives something like immunity to the multimedia object when it is watermarked. And it achieves the highest degree of privacy. So, the technique ”watermarking” has promising future and will be popularly widespread and used throughout the world, especially, in the multimedia factories, in the courts, in the banks, in the hospitals, etc, where the privacy is the highest priority [Abdul-Ahad et al. 2008; Chae and Manjunath 1999].

*amir_stephan@yahoo.co.uk
†maria.linden@mdh.se
‡thomas.larsson@mdh.se
§profwaleed54@yahoo.com

2 Important properties of digital video

The term Video is a Latin term and means “I see”. Video is basically a three-dimensional array of color pixels. Two dimensions serve as spatial (horizontal and vertical) directions of the moving pictures, and one dimension represents the time domain. A data frame (video frame image) is the set of all pixels that correspond to a single time moment. Basically, a frame is the same as a still image [Hui and Yeung 2003; Cox and Miller 2001; Doërr and Dugelay 2004].

This paper deals specifically with digital video frames having the properties of 24 bits per pixel (8 bits for each primary color red, green, and blue). Each primary color has a fixed range of 256 hexadecimal values. These allowable 256 hexadecimal numbers of each primary color can be read with 256 floating point numbers. Generally, each value is either 0 or multiples of 0.0039216 [Abdul-Ahad et al. 2008].

Figure 1: *Original digital video1 and video2 (cover objects)*

Figure 2: *Iconic image (watermark object)*

3 Direct embedding

Two cover objects, refereed to as digital video1 and video2, are used in this paper as shown in Figure 1. Both videos are of dimensions 240*320*3 with a frame rate of 30 fps and the encoding format is WMV (compressed video file format developed by Microsoft) . The watermark data may be an iconic image (the image may contain logo, signature, fingerprint, trade mark, serial number, and so on), text, or both. Figure 2 shows an iconic image with dimensions 32*32*3 (1024 pixels) to be used to watermark the cover objects.

The embedding process is implemented by decomposing the pixel values of the watermark object into packages of small numbers.

Thereby, for 2 bits distance, each pixel consists of 12-packages of small numbers. The total number of watermark data is 12288 and their four possible values are ranged from 1 to 4. For 4 bits distance, each pixel consists of 6-packages of small numbers. The total number of watermark data is 6144 and their sixteen possible values are ranged from 1 to 16. And for 8 bits distance, the total number of watermark data is 3072, ranged from 1 to 256 [Abdul-Ahad et al. 2008].

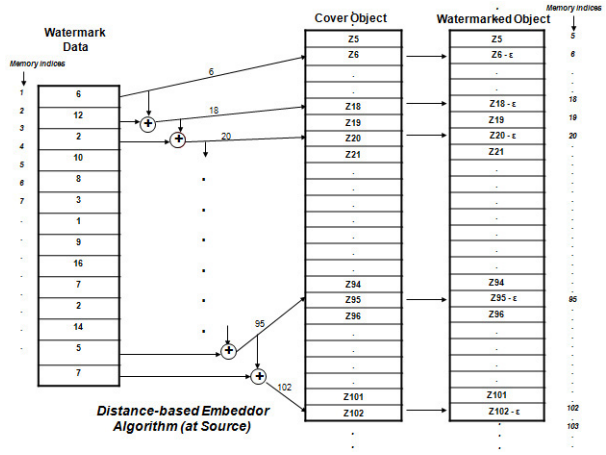


Figure 3: *Work of distance-based embeddor*

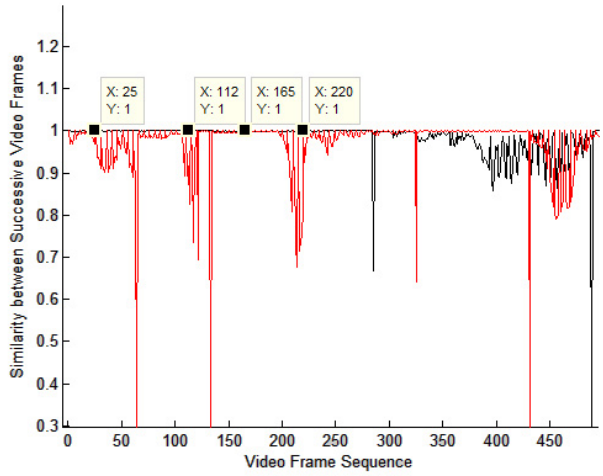


Figure 4: *Correlation between successive video frames. Red and black lines represent video1 and video2, respectively*

These values can be considered as addresses to locate the embedding positions in an uncompressed cover object (the embedding takes place before the WMV coder). The values are accumulated to get the next address. The difference in accumulated values of the watermark data can be thought of as distances among addresses of the cover object. This operation can be termed as Distance-Based Embeddor. The operation of this process is illustrated in Figure 3 and can be explained as follows: the value of the 1st watermark data points to the address of the 1st modified content. The 1st and 2nd watermark data are accumulated to get the 2nd address of the 2nd modified content and so forth. The distance between 1st address and 2nd address is equal to the value of 2nd watermark data. This accumulation goes on to the end of the last value of the watermark data. The contents of the positions which are addressed by the watermark data are modified with a factor of ε to obtain the watermarked object. This factor is selected to be the smallest value (zero is not selected) among the values of the color image of the cover object. In case of direct embedding, ε is selected to be equal to 0.0039216. This value has very little effect of distortion on the watermarked image. Thereby, the effect of uniquely distance-based deployment of watermark data in cover object helps in encryption. For example, for 2 bits distance-based embedding, the number of possible deployments of the watermark data over a cover object corresponds to the factorial of 12288 (i.e., 12288!). This reflects the complexity of the encryption per each iconic image. Besides, the possibility of embedding diversity at specific distance bits per frame (e.g., 2 bits distance, 4 bits distance and so on). It is worth mentioning here for non-blind watermarking, the embedding process can be anywhere in the cover object and doesn't matter how many times the watermark embed along the video. In case of blind watermarking, it is important to take in consideration the measurement of correlation for successive video frames to decide both the suitable place and how many times the watermark embed along the video (mostly successive video frames are highly correlated). As shown in Figure 4, the labeled video frames 25 and 112 for video1 and 165 and 220 for video2 are suitable for watermark embedding because of 100% correlation with successive video frames 26 and 113 for video1 and 166 and 221 for video2, respectively. This in turn, facilitates the blind watermark recovery. And at the same time this way of watermarking in many highly correlated video frames decreases the harming effect of frame dropping and frame swapping.

Peak Signal to Noise Ratio (PSNR) measures and mirrors the magnitude of distortion due to watermark data embedding in cover object. PSNR is apparently improved with an increase of distance bits. The reason is that for 2 bits distance, to embed 32*32 pixels, 12288 locations are required, while for 4 and 8 bits distances, only 6144 and 3072 locations are required, respectively. It is obvious that the number of modified values of watermarked object when using 8 bits distance is less than 4 and 2 bits distances. For 2 bits distance, the maximum and minimum distance between any consecutive addresses is 4 and 1 location(s), respectively, and the size of video frame must be at least 30498 bytes (sum the numbers of 2 bits 12288 values of iconic image) to totally embed an iconic image. For 4 bits distance, the maximum and minimum distance between any consecutive addresses is 16 and 1 location(s), respectively, and the size of video frame must be at least 46594 bytes (sum the numbers of 4 bits 6144 values of iconic image) to totally embed the same iconic image. And for 8 bits distance, the maximum and minimum distance between any consecutive addresses is 256 and 1 location(s), respectively, and the video frame size must be at least 401777 bytes (sum the numbers of 8-bits 3072 values of iconic image) to totally embed the same iconic image. Consequently, a different iconic image gives different lengths. It is clear that for the same watermark object, the 8 bits distance needs a larger cover object size to embed all watermark data in comparison to 4 and 2 bits distances. It is obvious that video file easily offers this possibility rather than audio and still image files, since, if needed, the deployment of the watermark data can be made over several video frames.

4 DWT embedding

The rapid progress of digital signal processing (DSP) techniques caused the sequential development of the electronic copyright (or proprietorship) protection. Due to this development, the DSP technique Discrete Wavelet Transform (DWT) becomes a vital and ready tool in achieving this requirement. It can be regarded as the most efficient transform dealing with multimedia applications since it provides a powerful and independent time-frequency (Multireso-

lution) representation.

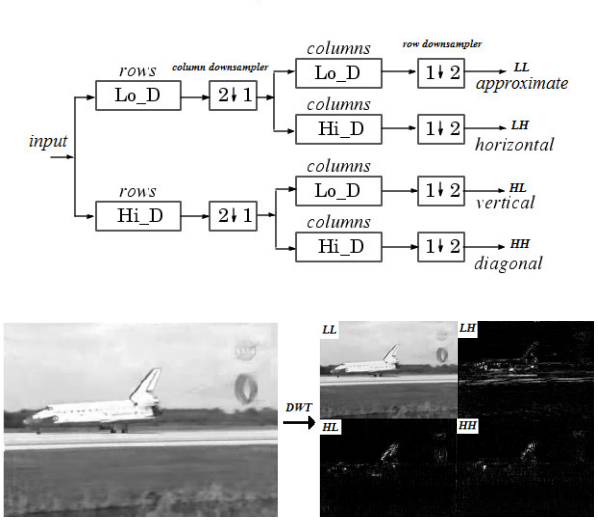


Figure 5: *Schematic diagram for the first level 2D DWT decompositions (top), and the result of the decompositions for the blue primary color (bottom)*

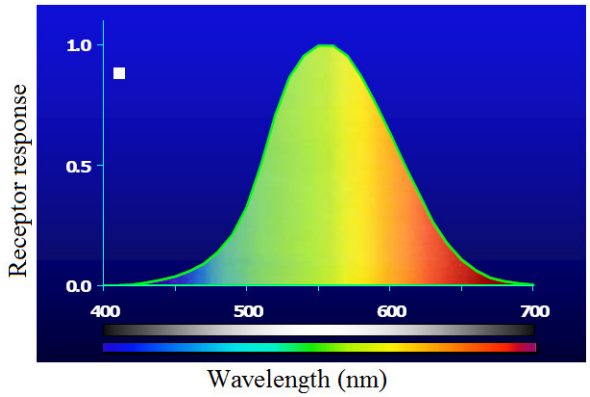


Figure 6: *Human eye sensitivity to spectral colors*

The least significant byte of a pixel's color (blue) of the video frame is decomposed into four sub-bands using 2D DWT technique as shown in Figure 5. As can be seen, any primary color channel can be transformed into four sub-bands, namely LL (approximate sub-band) where both horizontal and vertical directions have low frequencies, LH (horizontal sub-band) where the horizontal direction has low frequency and the vertical one has high frequency, HL (vertical sub-band) where the horizontal direction has high frequency and the vertical one has low frequency, and HH (diagonal sub-band) where both horizontal and vertical directions have high frequencies.

Each sub-band has dimension of 120*160 of 2D DWT coefficients. The blue color is preferred for watermark embedding, but if there is not enough space, the 2D DWT sub-bands of the red color is the next best option and the 2D DWT sub-bands of the green color is the last option. This choice is motivated by the fact that the human visual system (HVS) is subjectively less sensitive to the blue color and best sensitive to the green color as shown in Figure 6. The decomposition mechanism of the 2D DWT is very similar to the HVS mechanism. This means the higher resolution sub-bands LH, HL and HH are suitable regions to watermark embedding rather than the lower sub-band LL where the most primary color energy is concentrated. After the embedding, the 2D inverse DWT is applied to the four decompositions to get the watermarked object. Embedding watermark in these higher sub-bands increases the robustness of the embedding with no additional impact on video frame quality [Burrus et al. 1998; Kutte et al. 1998; Hunterlab 2001; Langelaar et al. 2000]. It is worthy to note that Daubechies wavelet function (db4) is used throughout this work.

5 Some models of attacking

Tables 1(a), 1(b), 2(a), 2(b), and 3 illustrate important results of attacking the watermarked object with effective and popular attacks [Doërr 2005; Alattar et al. 2003; K. Su 2001; Su et al. 2005]. The algorithm is tested by WMV compression, frame swapping, frame dropping, and frame averaging. In spite of all, signal to noise ratio (PSNR) readings are low as shown in Tables 1, 2, and 3. All the used attacks have a direct and visible effectiveness of distortion on the watermarked object. The proposed algorithm has the ability of high resilience of extracted watermark object regardless of what type of embedding technique is used. It is important to mention that all embeddings are implemented with 2 bits and 4 bits distances. The relevant results in the tables are very close to each other. This means that all embeddings are robust. However, according to the retrieving of the iconic image, it is apparent that the 1st level 2D DWT embedding is more withstandable than the direct embedding. Note that all the direct and 2D DWT embeddings used in Tables 1 and 2 are done at the same value of ε. This means that at amplified ε, Table 3 shows an improved version of the resilience of the extracted watermark shown in Table 1(a) using 4 bits distance-based direct embedding. Consequently, increasing ε increases the robustness of the watermark at the expense of the quality of the watermarked object.

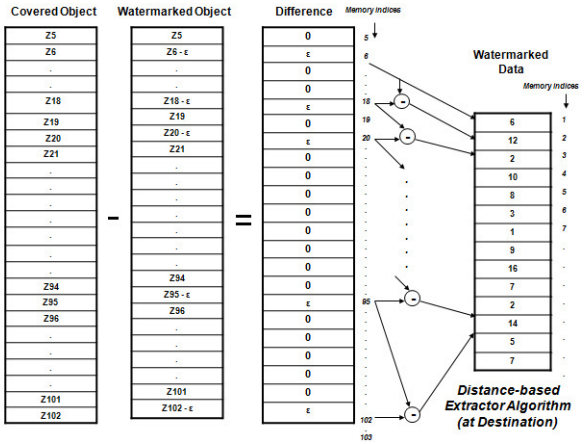
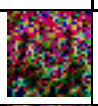
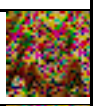
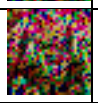
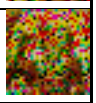
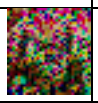
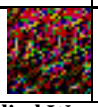
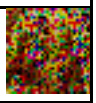
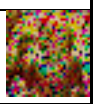
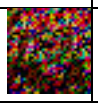


Figure 7: *Work of distance-based extractor*

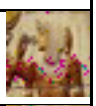
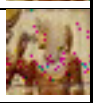
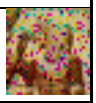
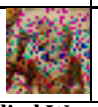
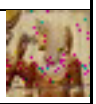
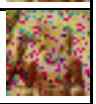
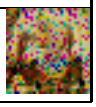
6 Recovering process

Figure 7 illustrates the work of the distance-based extractor. The watermark data is extracted from the difference between the cover and watermarked video frames in case of non-blind extraction, or between successive video frames in case of blind extraction. The difference contains numbers of zeros and εs in case of direct embedding or 1st level of 2D DWT embedding. The nonzero values indicate the addresses where the modification has taken place. The difference between any consecutive addresses of nonzero content is recovering the data of the iconic image. The recovered data must be

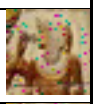
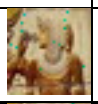
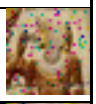
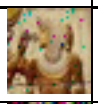
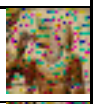
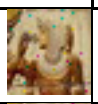
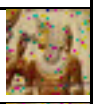
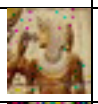
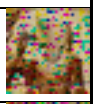
A) 4 bits distance-based direct embedding

Non-Blind Watermarking	
Video1	**Video2**
WMV Compression PSNR = 41.1421 Corr = 1 	WMV Compression PSNR = 40.6394 Corr = 1 
Frame swapping PSNR = 39.6156 Corr = 1 	Frame swapping PSNR = 38.2422 Corr = 0.9939 
Frame dropping PSNR = 37.8392 Corr = 0.9999 	Frame dropping PSNR = 25.6583 Corr = 0.9801 
Frame averaging PSNR = 41.5083 Corr = 1 	Frame averaging PSNR = 36.2609 Corr = 1 
Blind Watermarking	
Video1	**Video2**
Frame swapping PSNR = 41.7833 Corr = 1 	Frame swapping PSNR = 40.9564 Corr = 0.9941 
Frame dropping PSNR = 38.7901 Corr = 1 	Frame dropping PSNR = 25.6615 Corr = 0.9806 
Frame averaging PSNR = 42.7947 Corr = 1 	Frame averaging PSNR = 36.6966 Corr = 0.9941 

A) 4 bits distance-based first-level DWT embedding

Non-Blind Watermarking	
Video1	**Video2**
WMV Compression PSNR = 38.2807 Corr = 1 	WMV Compression PSNR = 38.1235 Corr = 1 
Frame swapping PSNR = 37.1614 Corr = 0.9998 	Frame swapping PSNR = 36.0716 Corr = 1 
Frame dropping PSNR = 35.9163 Corr = 0.9998 	Frame dropping PSNR = 25.5111 Corr = 0.9975 
Frame averaging PSNR = 37.9260 Corr = 0.9998 	Frame averaging PSNR = 24.4765 Corr = 0.9974 
Blind Watermarking	
Video1	**Video2**
Frame swapping PSNR = 37.1522 Corr = 0.9998 	Frame swapping PSNR = 35.8883 Corr = 1 
Frame dropping PSNR = 36.0045 Corr = 0.9998 	Frame dropping PSNR = 25.5111 Corr = 0.9975 
Frame averaging PSNR = 37.7667 Corr = 0.9997 	Frame averaging PSNR = 26.6993 Corr = 0.9978 

B) 2 bits distance-based direct embedding

Non-Blind Watermarking	
Video1	**Video2**
WMV Compression PSNR = 37.8241 Corr = 1 	WMV Compression PSNR = 37.4899 Corr = 1 
Frame swapping PSNR = 37.1019 Corr = 0.9999 	Frame swapping PSNR = 36.2532 Corr = 1 
Frame dropping PSNR = 35.9362 Corr = 0.9999 	Frame dropping PSNR = 25.5132 Corr = 0.9976 
Frame averaging PSNR = 40.4198 Corr = 1 	Frame averaging PSNR = 33.9039 Corr = 0.9997 
Blind Watermarking	
Video1	**Video2**
Frame swapping PSNR = 38.7691 Corr = 1 	Frame swapping PSNR = 38.0842 Corr = 1 
Frame dropping PSNR = 36.5731 Corr = 0.9999 	Frame dropping PSNR = 25.5468 Corr = 0.9975 
Frame averaging PSNR = 40.9573 Corr = 1	Frame averaging PSNR = 34.1585 Corr = 0.9997

B) 2 bits distance-based first-level DWT embedding

Non-Blind Watermarking	
Video1	**Video2**
WMV Compression PSNR = 37.9244 Corr = 1	WMV Compression PSNR = 37.1998 Corr = 1
Frame swapping PSNR = 37.0914 Corr = 0.9999	Frame swapping PSNR = 36.0781 Corr = 1
Frame dropping PSNR = 35.9735 Corr = 0.9998	Frame dropping PSNR = 25.5779 Corr = 0.9975
Frame averaging PSNR = 39.9828 Corr = 1	Frame averaging PSNR = 28.9937 Corr = 0.9987
Blind Watermarking	
Video1	**Video2**
Frame swapping PSNR = 38.0546 Corr = 0.9999	Frame swapping PSNR = 37.4842 Corr = 1
Frame dropping PSNR = 36.6066 Corr = 0.9998	Frame dropping PSNR = 25.8397 Corr = 0.9978
Frame averaging PSNR = 39.7681 Corr = 1	Frame averaging PSNR = 29.4287 Corr = 0.9991

Table 1: *Attacks for distance-based direct embedding. Video1: compression (frame 25), swapping (frames 25, 26), dropping (frames 25, 27), and averaging (frames 23–26). Video2: compression (frame 165), swapping (frames 165, 166), dropping (frames 165, 167), and averaging (frames 163–166)*

Table 2: *Attacks for distance-based DWT embedding. Video1: compression (frame 25), swapping (frames 25, 26), dropping (frames 25, 27), and averaging (frames 23–26). Video2: compression (frame 165), swapping (frames 165, 166), dropping (frames 165, 167), and averaging (frames 163–166)*

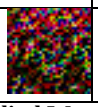
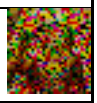
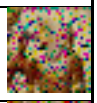
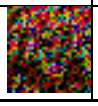
Non-Blind Watermarking			
Video1		**Video2**	
WMV Compression PSNR = 30.1508 Corr = 1		WMV Compression PSNR = 39.6619 Corr = 1	
Frame swapping PSNR = 38.9333 Corr = 1		Frame swapping PSNR = 37.8506 Corr = 1	
Frame dropping PSNR = 37.3505 Corr = 0.9999		Frame dropping PSNR = 25.6290 Corr = 0.9976	
Frame averaging PSNR = 41.5257 Corr = 1		Frame averaging PSNR = 36.2017 Corr = 0.9999	
Blind Watermarking			
Video1		**Video2**	
Frame swapping PSNR = 40.6039 Corr = 1		Frame swapping PSNR = 40.3088 Corr = 1	
Frame dropping PSNR = 38.1280 Corr = 0.9999		Frame dropping PSNR = 25.8901 Corr = 0.9978	
Frame averaging PSNR = 41.8660 Corr = 1		Frame averaging PSNR = 36.5163 Corr = 0.9999	

Table 3: *Attacks for distance-based direct embedding using amplified ε. Video1: compression (frame 25), swapping (frames 25, 26), dropping (frames 25, 27), and averaging (frames 23–26). Video2: compression (frame 165), swapping (frames 165, 166), dropping (frames 165, 167), and averaging (frames 163–166)*

12288 packages for 2 bits distance embedding and 6144 packages for 4 bits distance embedding.

The direct embedding algorithm is directly implemented on the content of the cover image and the amount of distortion depends upon ε and length of distance bits. While in distance-based 2D DWT embedding, in addition to ε and length of distance bits, the amount of distortion is affected by the order and level of wavelet selection.

7 Conclusion

The lossy compression constitutes the main obstacle and challenge to the technical watermarking but the algorithm was able to overcome it completely. Especially when using the technique DWT, this technique provides a kind of protection for the watermark data, particularly at higher resolution sub-bands, compared with direct embedding. Direct embedding is a pixel-wise process while DWT embedding is a coefficient-wise (indirect) process. Results written in the tables confirm this protection. In other words, the importance and benefits of watermarking technique will stimulate researchers to discover ways for lossless video compression.

Despite that PSNR is large, the resilience of the watermark image is vulnerable in front of the attacks particularly at 4 bits distance-based direct embedding, while at 2 bits distance-based direct embedding, the resilience is improved and robust. This is due to the wide deployment of 4 bits distance-based direct embedding over the frame, where the effect of lossy WMV compression is huge and devastating. This situation exists, but with little effect in case of 2D DWT embedding, where both 2 bits and 4 bits distance-based embeddings are robust.

Future research need to focus on decreasing the distortions and artifacts of the extracted watermark by extracting and embedding the most representative features of the watermark object. The goal would be to extend the proposed method to be resilient against other attacks such as geometric distortion of the images.

References

ABDUL-AHAD, A. S., CÜRÜKLÜ, B., AND MAHMOUD, W. A. 2008. Robust distance-based watermarking for digital image. In *Proceedings of the 2008 International Conference on Security and Management (SAM'08)*, 404–409.

ALATTAR, A., LIN, E., AND CELIK, M. 2003. Digital watermarking of low bit-rate advanced simple profile MPEG-4 compressed video. *IEEE Transactions on Circuits and Systems for Video Technology 13*, 8 (August), 787–800.

BURRUS, C. S., GOPINATH, R. A., AND GUO, H. 1998. *Introduction to Wavelets and Wavelet Transforms.* Prentice Hall.

CHAE, J., AND MANJUNATH, B. 1999. Data hiding in video. In *International Conference on Image Processing*, 311–315.

COX, I., AND MILLER, M. 2001. Electronic watermarking: the first 50 years. *IEEE Fourth Workshop on Multimedia Signal Processing*, 225–230.

DOËRR, G., AND DUGELAY, J.-L. 2004. Security pitfalls of frame-by-frame approaches to video watermarking. *IEEE Transactions on Signal Processing 52*, 10 (October), 2955–2964.

DOËRR, G. 2005. *Security Issue and Collusion Attacks in Video Watermarking*. PhD thesis, de lUniversit e de Nice-Sophia Antipolis.

HUI, S. Y., AND YEUNG, K. H. 2003. Challenges in the migration to 4G mobile systems. *IEEE Communications Magazine 41*, 12 (December), 54–59.

HUNTERLAB, 2001. The basics of color perception and measurement, version 1.4. www.hunterlab.com/pdf/color.pdf.

K. SU, D. KUNDUR, D. H. 2001. A content-dependent spatially localized video watermarked for resistance to collusion and interpolation attacks. In *IEEE International Conference on Image Processing*.

KUTTE, M., JORDAN, F., AND BOSSEN, F. 1998. Digital watermarking of color images using amplitude modulation. *Journal of Electronic Imaging 7*, 2 (April), 326–332.

LANGELAAR, G., SETYAWAN, I., AND LAGENDIJK, R. 2000. Watermarking digital image and video data. a state-of-the-art overview. *IEEE Signal Processing Magazine 17*, 5 (September), 20–46.

MAES, M., KALKER, T., LINNARTZ, J.-P., TALSTRA, J., DEPOVERE, F., AND HAITSMA, J. 2000. Digital watermarking for DVD video copy protection. *IEEE Signal Processing Magazine 17*, 5 (September), 47–57.

SU, K., KUNDUR, D., AND HATZINAKOS, D. 2005. Statistical invisibility for collusion-resistant digital video watermarking. *IEEE Transactions on Multimedia 7*, 1 (February), 43–51.