A Digital Video Watermarking Technique Based on Identical Frame Extraction in 3-Level DWT

Tamanna Tabassum, S.M. Mohidul Islam
Computer Science and Engineering Discipline, Khulna University
Khulna, Bangladesh
{shuktara_07, mohid_ku}@yahoo.com

***Abstract*—Digital video watermarking was introduced at the end of the last century to provide means of enforcing video copyright protection. Video watermarking involves embedding a secret information in the video. In this paper, we proposed a digital video watermarking technique based on identical frame extraction in 3-Level Discrete Wavelet Transform (DWT). In the proposed method, first the host video is divided into video shots. Then from each video shot one video frame called identical frame is selected for watermark embedding. Each identical frame is decomposed into 3-level DWT, then select the higher subband coefficients to embed the watermark and the watermark are adaptively embedded to these coefficients and thus guarantee the perceptual invisibility of the watermark. For watermark detection, the correlation between the watermark signal and the watermarked video is compared with a threshold value obtained from embedded watermark signal. The experimental results demonstrate that the watermarking method has strong robustness against some common attacks such as cropping, Gaussian noise adding, Salt & pepper noise adding, frame dropping and frame adding.**

***Index Terms*—Video Watermarking, Discrete Wavelet Transform, Video Shot, Identical Frame**

I. Introduction

Video piracy has become an increasing problem particularly with the proliferation of media sharing through the advancement of Internet services and various storage technologies. Thus, research in copyright protection mechanisms, where one of which includes digital watermarking has been receiving an increasing interest from scientists especially in designing a seamless algorithm for effective implementation [1][2][3]. Digital video watermarking involves embedding secret symbols known as watermarks within video data which can be used later for copyright detection purposes [4].

There are three factors (robustness, security, perceptual fidelity) which are necessary for video watermarking system [5]. The watermark can be visible or invisible [6]. In visible watermarking, the information is visible in the video while in invisible watermarking, information is not visible. It can be detected only by the owner. Another classification of is based on domain which the watermark is applied i.e., the spatial or the frequency domain. The easiest way to watermark a video is to change directly the values of the pixels, in the spatial domain. A more advanced way to do it is to insert the watermark in the frequency domain [7][8]. In this paper we propose an invisible video watermarking technique based on 3-Level DWT.

Many researchers around the world are working on digital watermarking for designing and implementing robust image or video watermarking technique. Some of those works are based on Discrete Wavelet Transform (DWT). A notable work is [5], in which the author resize the host video to 256×256 size blocks and apply DWT on watermark video. To extract watermark they follow just inverse of embedding process. One of the major limitation of this method is that the host video must be square in size. Image which is used as a watermark must be in square size. Moreover the quality of watermarked video is bit degraded from the original host video.

For the advent of video and neural network technology [9][10][11][12], it is possible to convert the video into video shots for watermark embedding and finally extract embedded watermark from the trained neural network. Tian Hu et al. [13] proposed an watermarking technique in 1D-DWT where the author applied 1D-DWT to the luminance of two consecutive frames to obtain low frequency image. Low frequency image is partitioned into equal sized sub image. Calculate average pixel in each block. Embed watermark in these blocks according to interval where the average pixel value is. In order to detect the watermark the author used 1D-DWT to the luminance of two consecutive frames to obtain low frequency image. Compute average pixel value of each block. Determined interval where each average value belongs. The main limitation of this work is that their system only support binary image as the watermark.

In video watermarking blind watermarking scheme is also popular. One of the notable work is [14] where the author presented a new blind watermarking scheme in which a watermark is embedded into the one level DCT. The method uses the Human Visual System (HVS) model, and neural network. The neural network is implemented while embedding and extracting watermark. The HVS model is used to determine the watermark insertion strength. The inserted watermark is a random sequence. The secret key determines the beginning position of the image where the watermark is embedded. Another novel blind video watermarking scheme is [10] which is based on pseudo-3D DCT. In this work they converted several scenes into video segment, and the frames in each scene are transformed in the 2D-DCT. Then the resulting Direct Current (DC) components are transformed along the temporal dimension. Afterwards, the normally distributed watermark is embedded into the pseudo-3D DCT Alternating Current (AC) coefficients.

In the proposed method (i) original video need not to be squared i.e. any video can be used as original video (ii) the watermark signal can be of any size (iii) for watermark detection no original video is required and (iv) perceptually invisible robust watermarking is achieved.

The remainder of this paper is organized as the following. At first, in Section II we illustrate the various components of our proposed technique to embed and detect watermark from video content. Further, in Section III we present some key experimental results and evaluate the performance of the proposed system. At the end we provide conclusion of the paper in Section IV and state some possible future work directions.

II. Proposed Watermarking Technique

This section illustrates the overall technique of our proposed digital video watermarking technique based on 3-level DWT. At first, the formation of 3-Level DWT are presented. Then the proposed watermark embedding process, including identical frame extraction technique are discussed in detail. Finally, the watermark detection process and its different steps are discussed in detail.

A. 3-Level-DWT

Discrete wavelet transform (DWT) is a mathematical tool for hierarchically decomposing an image [15]. DWT is the multi-resolution description of an image. The decoding can be processed sequentially from a low resolution to the higher resolution [16]. The DWT splits the signal into high and low frequency parts. The high frequency part contains information about the edge components, while the low frequency part is split again into high and low frequency parts. The high frequency components are usually used for watermarking since the human eye is less sensitive to changes in edges [17]. After the first level of decomposition, there are 4 sub-bands: LL1, LH1, HL1, and HH1. For each successive level of decomposition, the LL subband of the previous level is used as the input. To perform second level decomposition, the DWT is applied to LL1 band which decomposes the LL1 band into the four subbands LL2, LH2, HL2, and HH2. To perform third level decomposition, the DWT is applied to LL2 band which decompose this band into the four sub-bands: LL3, LH3, HL3, HH3.

B. Watermark embedding

In a video, sometimes different video frames are almost identical. A continuous identical video frames is called a video shot. In order to increase the performance of watermark embedding process the proposed system will separate the video into video shots. Each video shot has one or more video frames that are almost identical. In order to determine whether two video frames are identical we compare the two image pixels. Moreover we also consider on global characteristics of the frames, which is intensity histogram. According to video standard, the intensity for a RGB frame can be calculated as,

$$I = 0.299R + 0.587G + 0.114B \tag{1}$$

Where R, G and B are Red, Green and Blue value of the pixel. Generally, the human visual system is least sensitive to the range of high frequency [18]. Among three channels of the RGB image, the blue channel has characteristic of the highest frequency range. So, for the high performance the blue channel is transformed into DWT and the watermark is embedded from HL3 sub-band of the blue channel of the host video frame. If the HL3 sub-band is fill-up then the remaining watermark signal is embedded in LH3 sub-band. Again, if the LH3 sub-band is over then HH3. If HH3 is over then the next upper level is used that is HL2, LH2, HH2 is used. In this way all the watermark is embedded into the video frame (see Figure 1). This process has the benefit of larger watermark can be embedded into the video. As we are placing the watermark into the high frequency part of the blue channel, so the greater invisibility of the watermark in the watermarked video frame is achieved.

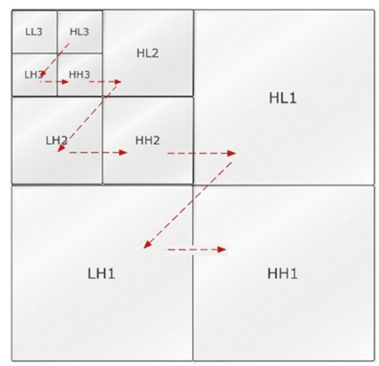


Fig. 1: Watermark embedding order in 3-level DWT sub-bands. The dotted arrow line indicating the order.

For the intensity histogram difference we're looking for, it can be expressed as,

$$|SD_i| = \sum_{i=1}^{G} |H_i(j) - H_{i+1}(j)| \tag{2}$$

Where $H_i(j)$ is the histogram value for i^{th} frame at level j. G denotes the total number of levels for the histogram. In a continuous video frame sequence, the histogram difference is small, whereas for sudden transition detection, the intensity histogram difference spikes. Even there is a notable movement or illumination changes between neighboring frames, the intensity histogram difference is relatively small compared with those peaks caused by sudden changes. Therefore, the difference of intensity histogram with a proper threshold is effective in detecting sudden transitions in video frames. The threshold value to determine whether the intensity histogram difference indicates a sudden transition can be set to,

$$T_b = \mu + \alpha\sigma \tag{3}$$

Where μ and σ are the mean value and standard deviation of the intensity histogram difference. Empirically we estimate that the value of α typically varies from 2 to 6.

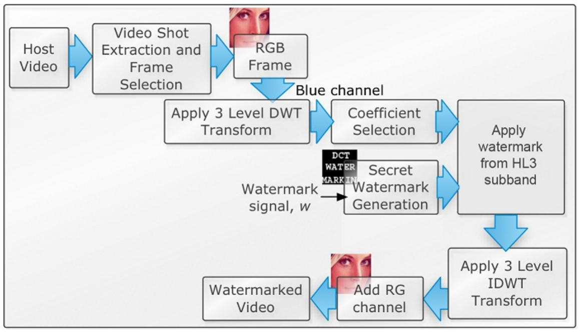


Fig. 2: Steps of proposed watermark embedding process

Before embedding the watermark it should be preprocessed. The watermark is converted into binary image form as $w'(i,j) \in \{0,1\}$, for $i,j = 0$ to M, where M is the number of binary pixel in the image to be encoded. Here the value 0 represents black and 1 represent white value. The binary form of the image $w'(i,j)$ is then transformed to obtain the vector $w(i,j) \in \{1,-1\}$, where 0 is replaced by 1 and 1 is replaced by -1. Finally two dimensional watermark $w(i,j)$ is changed into one dimensional watermark $w(l)(l = 1,2,...,L), L$ is the length of the watermark.

The proposed embedding process is shown in Figure 2. From the block diagram we see that, after separating the video into video shots the system will apply 3L-DWT on the blue channel of RGB frame. In the 3L-DWT coefficients, we embed preprocessed watermark image from the HL3 to HH1 (as mentioned earlier in Figure 1) sub-band consecutively and then it is transformed into 3-level inverse DWT form. At this stage, for RGB video frame we get the watermarked blue channel which is then combined to other two channels to obtain the watermarked video frame. The relation of embedding is given in Equation 4. Where i and j ranges over selected coefficients in the DWT $\beta'_{i,j}$ and $\beta_{i,j}$ denote the DWT coefficient of the blue channel of the original video frame and the watermarked video frame respectively, $w_{i,j}$ is the watermark signal and ζ (zeta) is the scaling parameter which value ranges from 0.2 to 0.6 (we found it from our experimental result).

$$\beta'_{i,j} = \beta_{i,j} + \zeta|\beta_{i,j}|w_{i,j} \tag{4}$$

In the case of multiple watermarking, the equation 4 can be repeated upto n times by Equation 5.

$$\beta^{n}_{i,j} = \beta^{n-1}_{i,j} + \zeta|\beta^{n-1}_{i,j}|w^{n-1}_{i,j} \tag{5}$$

C. Watermark detection

Without the original video, authorized detection of the hidden information can be easily accomplished by using the watermarked video and watermark signal. The detector detect whether the watermark is present or not in the watermarked video. Similar to the embedding process, before detecting the watermark the system need to extract the video shots and then select the appropriate identical frame from each video shot. Then the 3-level wavelet transform is performed on the blue channel of the selected frame. Finally compute the average η. The proposed watermark detection steps are illustrated in Figure 3.

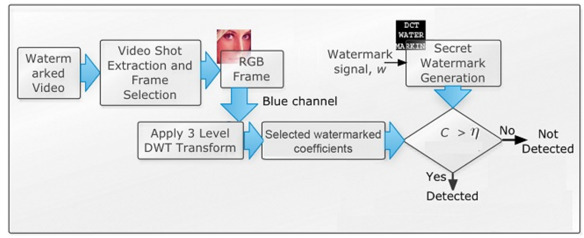


Fig. 3: Steps of proposed detection process. The system detect an watermark if the coefficient, C is larger than the average, η.

Consider the size of the selected coefficient blocks is $M \times N$ and the total length of the watermak is L then the average of the selected coefficients in the sub-bands is η. Where η can be calculated as:

$$\eta = \begin{cases} \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} F'(i,j), if\ MN \leq L \\ \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N=\frac{L}{M}-1} F'(i,j), if\ MN > L \end{cases} \tag{6}$$

Now determine the correlation C between the selected DWT coefficients $F'(i,j)$ and the provided watermark vector w and compare C with η.

$$C = \begin{cases} \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0,m=i*M}^{N-1} F'(i,j)w_m, if\ MN \leq L \\ \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0,m=i*M}^{N=\frac{L}{M}-1} F'(i,j)w_m, if\ MN > L \end{cases} \tag{7}$$

If the provided watermark signal and the embedded watermark signal are similar then the value of the correlation is larger than the average or threshold value otherwise not, i.e. if $C > \eta$ then we can say that the provided watermark is detected, otherwise not. As $F'(i,j)$ may be negative and w has value -1 or 1, so η always greater than C. As a result a scaling parameter k is required. So the adjusted threshold is in equation 8.

$$\eta = \begin{cases} \frac{k}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} F'(i,j), if\ MN \leq L \\ \frac{k}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N=\frac{L}{M}-1} F'(i,j), if\ MN > L \end{cases} \tag{8}$$

The value of k is estimated empirically through experiments.

III. Evaluation and Results

To verify the effectiveness (qualities and robustness) of the proposed video watermarking technique, we conduct several experiments with this procedure on several uncompressed video clip that are listed in Table I. Most of the video are downloaded from the YouTube [1]. The description of videos that we used in our experiment are as follows:

[1] YouTube, http://www.youtube.com

- *Baboon (BA)* represents a video that is constructed by several still images. This particular one has been made with Camtasia studio [2] software.
- *Clock (CL)* is a low-resolution sample video which has 5010 video frames.
- *BBC Life 2009 [Ep1] (BB)* brings real images from nature with minimum camera movement.
- *Tom And Jerry (TJ)* is a famous animated TV series developed in 1951.
- *Baby's Day Out (BD)* is a famous live-action comedy film.
- *Boundin' (BO)* is a short film developed by Pixar [19].

TABLE I: Properties of the videos that are used in the experiment.

	Resolution	Length [min]	#of frames	#of identical frames	PSNR (dB)
BA	128x128	6.25	3850	55	+43.69
CL	256x128	5.34	5010	214	+45.02
BB	480x270	9.58	14950	47	+49.81
TJ	320x240	7.03	12275	358	+53.01
BD	320x240	10.00	16800	68	+54.21
BO	256x128	3.43	3353	49	+46.53

A. Perceptibility

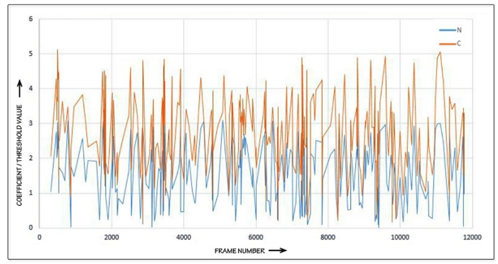
Perceptibility expresses amount of distortion caused by watermark embedding. In other words, it indicates how visible the watermark is. It is measured by peak signal-to-noise ratio (PSNR). The less the value of PSNR is the more perceptible the watermark is. Table I demonstrating the PSNR value of each video. It should be noted here that as we only embed the watermark in the identical frame of each video shots so here the PSNR value represent the average of total identical frames from all video shots. If the PSNR value is 41db or more then it can be said that the quality of the video is almost same to the original video.In our proposed method the PSNR value for all the videos is more than 41db.

TABLE II: Time consumption test results.

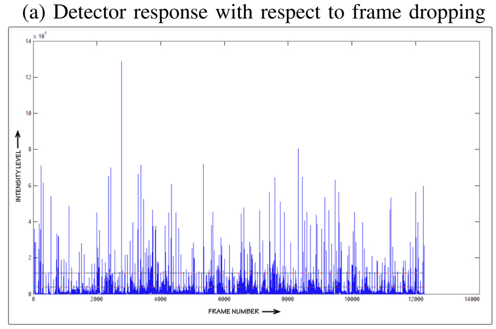
	Resolution	Embedding time (s)	Detection time (s)
BA	128x128	378	219
CL	256x128	340	221
BB	480x270	490	185
TJ	320x240	412	175
BD	320x240	569	239
BO	256x128	199	49

B. Time consumption

This section discusses time consumption for both the embedding and the detection process. The time has been measured by MatLab function called tic and toc. Average results of three iterations are listed in Table II. The test has been executed on a machine with the following configuration: Intel Core i-3 processor, 2.50GHz, 2GB memory, windows 7 operating system. From table II we can say that both the embedding and detection time depends on the resolution and number of frames in the video. When the resolution and the number of frame increases then the embedding and the detection time also increases. Besides this the detection process is faster than the embedding process.



(a) Detector response with respect to frame dropping



(b) Frames intensity levels

Fig. 4: Detector response (threshold and correlation) with respect to frame dropping

C. Frame dropping

For the existence of the inherent redundancy in video data, there is little change between frames in a shot. So, the frame dropping is often used as an effective video watermark attacks, since it leads little or no damage to the video signal. To test the performance of video watermarking procedure against the video frame dropping, we choose Tom And Jerry (TJ) video which has 12275 frames and randomly dropped 500 frames and finally execute the detector. Here Response of the detector to the frame number that are already dropped is depicted in figure 4a. In this figure the red line indicating the correlation and the blue color indicating the threshold value. As we embed watermark into the identical frame of each video shot if no identical frame is deleted the system can easily detect the watermark after a large number of frames are dropped. But if the identical frame is deleted then the system can not detect the watermark for that shot, but it does not affect the final result. The system can detect the watermark if 40% of the identical frames remains in the video.

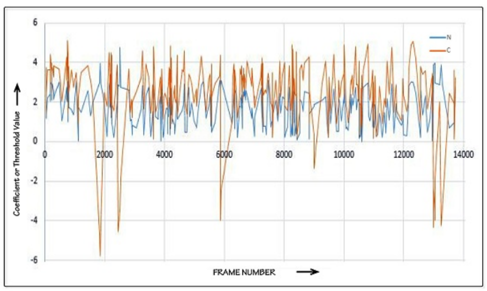
So our technique perform well in case of frame dropping in a certain level. Figure 4b demonstrating the identical frame

[2] Camtasia studio, http://www.techsmith.com

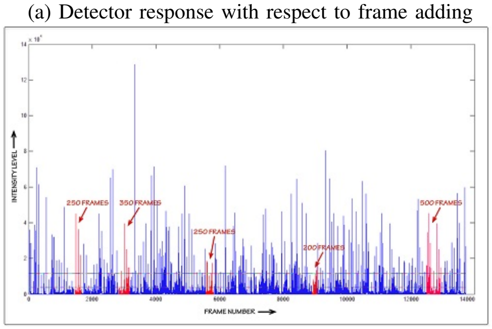
positions and their intensity level. In this figure the upper horizontal line indicating the cut threshold.

D. Frame adding

To test the performance of our method in the case of frame adding we added few frames in random selected position in the Tom And Jerry (TJ) video which has 12275 frames and finally execute our detector and plot in a graph which is demonstrating in Figure 5a. Figure 5b demonstrating the frame adding position. In this figure the red vertical line indicating those positions. The number of frame that are added also mentioned in figure 5b. From figure 5a we can say that at any point the correlation value is always larger than the threshold except those positions where the watermark frames are added. In those positions the correlation is lower than the threshold as our watermark is not available in that frame. If the number of frame added increases the proposed system can successfully detect the watermark upto 65% of addition of new frame .



(a) Detector response with respect to frame adding



(b) Frames intensity levels

Fig. 5: Detector response (threshold and correlation) with respect to frame adding

E. Scaling

The scaling test scales down the watermarked video sequences to the specified resolution. The resolution is given by scaling factor (column in Table III named SF) which determines how much the video frame of the output raw stream are scaled down e.g. scaling factor of 1/4 means that the image area size is reduced 4-times, i.e. both width and height are halved. The raw stream is compressed at the scaling factor fragment of the former bit-rate, i.e. for example scaling factor of 1/4 means 1/4 of the former bit-rate.We experimented on 4 different videos with 3 different scaling factor for all videos the system can easily detect the watermark after scaling the video 1/2 and 1/4 to its original size but if the video size decreased to 1/6 the system fail to detect watermark.

TABLE III: Scaling test results.

	Video Scaling Factor (SF)	Threshold, η	Correlation, C	Detected?
BA	1/2	2.4753	4.1230	Yes
	1/4	3.1381	7.4546	Yes
	1/6	2.4753	-1.1387	No
CL	1/2	4.4452	10.1113	Yes
	1/4	3.0774	7.2104	Yes
	1/6	4.4452	-4.5801	No
BB	1/2	2.4753	4.1230	Yes
	1/4	3.0308	6.6890	Yes
	1/6	4.4452	-4.5801	No
TJ	1/2	4.4452	10.1113	Yes
	1/4	2.9810	6.1694	Yes
	1/6	3.1341	3.5335	Yes

F. Cropping

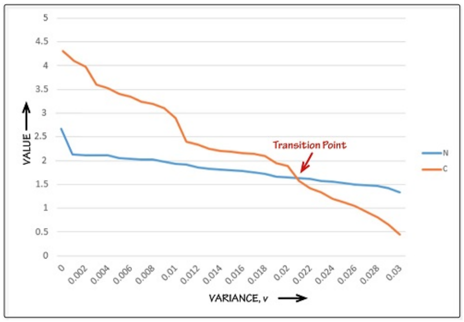
In the cropping test, we embed watermark in the BBC Life 2009 [Ep1] (BB) video to produce watermarked video. The BB video has total 14950 frames and the resolution was 480×270. After embedding we cropped the watermarked video frames by different range listed in the column $Cropping$ of Table IV. When the frame was cropped by 64×64 the correlation value is less than threshold value, so the watermark cannot be detected because in this case most of the image frame part were lost. But if the cropping is in a reasonable range like 80×80 or more, few of the image part was lost, so it can be detected by our proposed technique.

TABLE IV: Cropping test results on BB video.

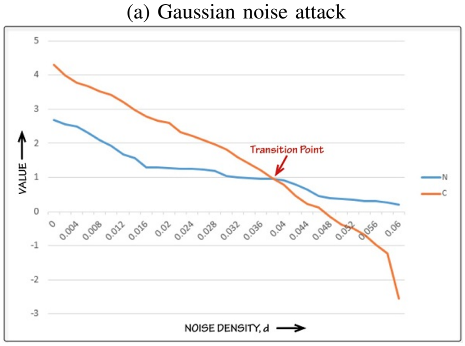
Cropping	Threshold, η	Correlation, C	Detected?
256×256	2.4353	5.1230	Yes
128×128	3.2331	5.0524	Yes
100×100	3.7753	4.9334	Yes
80×80	3.8432	4.8113	Yes
75×65	4.0739	4.3184	Yes
64×64	4.2452	3.9501	No

G. Noise attack

In order to check the performance of our technique in different noise attack we performed some experiment in MatLab. In this test at first we embedded an watermark in the Baboon video (BA), then by using MatLab function we added Gaussian noise and salt & pepper noise separately to the video. Finally we tested whether the watermark exists in the modified watermarked video. The test result are demonstrating in Figure 6. From Figure 6a we can say that for a small change of variance, the threshold and the correlation value changes and our system can support at most 0.021 for variance to detect



(a) Gaussian noise attack



(b) Salt & pepper attack

Fig. 6: Noise attack and the corresponding watermark detector's threshold and correlation value.

the watermark. From Figure 6b we observed similar response, the detector result changes based on how much salt & pepper noise are available in the video. This figure also demonstrating that when the noise density in the range between 0 to 0.038 then the detector can detect the watermark properly.

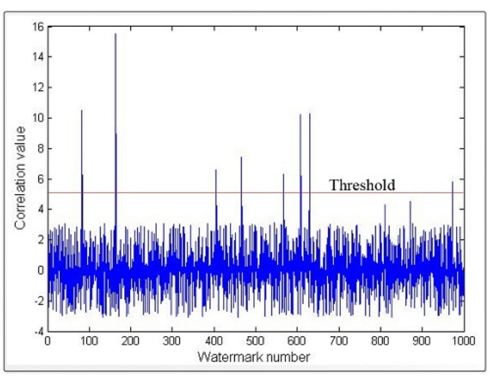


Fig. 7: Detector response to 1000 randomly generated watermarks. The ten different watermark detected correctly by the system.

H. Multiple watermark embedding

Multiple watermark embedding test influence different watermarks embed into one video sequences. To test the performance of our proposed technique, we randomly choose a video frame from the BBC Life 2009 [Ep1] (BB) video and finally watermarked the frame with ten different watermarks. Figure 7 shows the detector response, which indicates the presence of all the ten watermarks embedded in the video frame.

IV. Conclusion

In this paper, identical frame based video watermarking technique on 3-level DWT is proposed which is perceptually invisible.This proposed method shows little bit time complexity in watermark embedding process. Though the system has some limitations but it shows better results in various attacks. In future our plan is to minimize the watermark embedding time to improve the performance of the proposed system.

References

[1] G. Doërr and J. Dugelay, "A guide tour of video watermarking," *Signal processing: Image communication*, vol. 18, no. 4, pp. 263–282, 2003.

[2] S. Bhattacharya, T. Chattopadhyay, and A. Pal, "A survey on different video watermarking techniques and comparative analysis with reference to h. 264/avc," in *IEEE Tenth International Symposium on Consumer Electronics, ISCE'06*. IEEE, 2006, pp. 1–6.

[3] V. Agrawal, "Perceptual watermarking of digital video using the variable temporal length 3d-dct," Ph.D. dissertation, Citeseer, 2007.

[4] I. J. Cox, M. L. Miller, K. Tanaka, and Y. Wakasu, "Digital data watermarking," Patent EP0 840 513, May, 1998. [Online]. Available: http://www.freepatentsonline.com/EP0840513A2.html

[5] S. Gandhe, U. Potdar, and K. Talele, "Dual watermarking in video using discrete wavelet transform," in *Second International Conference on Machine Vision, ICMV '09.*, dec. 2009, pp. 216 –219.

[6] I. Cox and M. Miller, "Electronic watermarking: the first 50 years," in *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 225 –230.

[7] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[8] M. Shensa, "The discrete wavelet transform: Wedding the a trous and mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, 1992.

[9] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560 –576, july 2003.

[10] Y. Zhang and H. Bi, "A robust blind video watermarking scheme in the 3d-dct domain," in *International Conference in Electrics, Communication and Automatic Control Proceedings*, R. Chen, Ed. Springer New York, 2012, pp. 1009–1015. [Online]. Available: http://dx.doi.org/10.1007/978-1-4419-8849-2_128

[11] M. Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464 –2482, oct 1992.

[12] R. Lippmann, "An introduction to computing with neural nets," *ASSP Magazine, IEEE*, vol. 4, no. 2, pp. 4 –22, apr 1987.

[13] T. Hu and J. Wei, "A digital video watermarking scheme based on 1d-dwt," in *International Conference on Biomedical Engineering and Computer Science*, april 2010, pp. 1 –3.

[14] S. Islam, R. Debnath, and S. Hossain, "Dwt based digital watermarking technique and its robustness on image rotation, scaling, jpeg compression, cropping and multiple watermarking," mar. 2007, pp. 246 –249.

[15] W. Lu, H. Lu, and F. Chung, "Robust digital image watermarking based on subsampling," *Applied mathematics and computation*, vol. 181, no. 2, pp. 886–893, 2006.

[16] X. Xia, C. Boncelet, and G. Arce, "A multiresolution watermark for digital images," in *International Conference on Image Processing*, vol. 1. IEEE, 1997, pp. 548–551.

[17] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5. IEEE, 1998, pp. 2969–2972.

[18] R. Gonzalez and E. Richard, *Digital image processing*, 3rd ed. Prentice Hall Press, 2002.

[19] B. Luckey, "Boundin'," Pixar, Short film, November 2004.