Robust Watermarking for MPEG-2 Videos

Sheng-Chuan Tai, Chuen-Ching Wang and Ying-Ru Chen
Institute of Electrical Engineering,
National Cheng-Kung University,
Tainan, Taiwan, R.O.C.
sctai@mail.ncku.edu.tw

Abstract--A robust watermarking technique is proposed for protecting the author's intellectual property rights (IPR) in MPEG-2 video sequence. The author's identification data is embedded into the Y component of the uncompressed video sequence. Employing HVS, the proposed watermarking embeds the secret information into the wavelet coefficients of the frame. Simulation results show that the robust watermarking method can against intentional attacks, such as video sequence recompression, frame blurring, frame cropping, frame noising, frame sharpening, and frame filtering.

Keyword -- intellectual property rights, robust watermarking, video watermarking, MPEG2, HVS.

The authors submit this manuscript to ***ICS Workshop on Cryptology and Information Security*** --Visual Cryptography and Watermark

The authors are with *th*e Institute of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan
Correspondence Address
Prof. Shen-Chuan Tai
Institute of Electrical Engineering
(computer / group 92533)
National Cheng Kung University
Tainan, 70101, Taiwan
Email: sctai@mail.ncku.edu.tw, wcj@rose.ee.ncku.edu.tw
TEL: 886-6-2757575-62368
FAX: 886-6-2388673

Robust Watermarking for MPEG-2 Videos

Sheng-Chuan Tai, Chuen-Ching Wang and Ying-Ru Chen
Institute of Electrical Engineering,
National Cheng-Kung University,
Tainan, Taiwan, R.O.C.
sctai@mail.ncku.edu.tw

Abstract--A robust watermarking technique is proposed for protecting the author's intellectual property rights (IPR) in MPEG-2 video sequence. The author's identification data is embedded into the Y component of the uncompressed video sequence. Employing HVS, the proposed watermarking embeds the secret information into the wavelet coefficients of the frame. Simulation results show that the robust watermarking method can against intentional attacks, such as video sequence recompression, frame blurring, frame cropping, frame noising, frame sharpening, and frame filtering.

Keyword -- intellectual property rights, robust watermarking, video watermarking, MPEG2, HVS.

1. Introduction

Multimedia production and distribution has recently gained more and more popularity due to the increasing amount of information that is stored and transmitted digitally. The growth will continue at big rate when the advanced multimedia applications such as 3G wireless mobile system, WLAN, VOD (video on demand), E-commerce, etc., will be widely available. Furthermore, these networks utilize the roaming technique, such as horizontal hand-over and vertical hand-over, to integrate the different networks into a large one. This lead to the management of digital multimedia will be toward to be complicated. Since digital multimedia can be copied easily through accessing to computer networks, this will limit the development of the multimedia service in the public environment. Fortunately, digital watermarking is an approach to protect intellectual property (IPR) rights in digital form that can add a watermark to an original image for evidence of legal ownership and detection of copyright infringement. Digital watermarking technologies allow hiding information, for example a copyright number or logo, into digital media, such as image, video and audio. The hiding process has to modify the media imperceptibly.

In the case of video watermarking, the challenge is to mark a group of images that are strongly inter-correlated and often manipulated in a compressed form, e.g. MPEG. A first group of video watermarking designing therefore directly operates on MPEG data to avoid full decompression. Hartung [1] proposed to mark only the DCT coefficient of the intra-frames (I-frames). They use a spread spectrum signal containing the copyright information, which is added to the non-zero DCT coefficients under the condition of not increasing the bit rate. Other researchers watermark MPEG-2 motion compensation vectors [2] or VLC coding [3]. The advantages of these methods are their rapidity, because they need no decompression of the MPEG data. They are not resistant to various transformations of image frames such as rescaling, change of frame-rate, compression and re-compression in a different format or GOP reorganization.

In order for the watermark to be able to resist to transformations as well as to be less dependency on the way the video compression was done, the basic approach adopted here is to mark the uncompressed video sequence in spite of the increased computation cost. Working on uncompressed video, the first possible way is to individually mark all the frames of the video using a still image watermark technique. Doing so would allow inheriting the robustness of the 2D approaches; the drawback however would be the vulnerability to averaging attacks, where consecutive frames are averaged to remove the embedded mark [4]. In this paper we plan to bring out a video watermarking techniques, which is different from image watermarking. It uses robust watermarking technique to achieve the goal of copyright protection in every frame.

The rest paper is organized as follows. Section 2, the watermarking embedding/extraction process are detailed. In section 3 we describe the simulation results. Finally, conclusion and possible future work are discussed in section 4.

2. The proposed robust watermarking

By exploiting HVS theorem, we embed watermarks in DWT coefficients of the Y channel (luminance) in every frame. Fig.1 shows the block diagram for the proposed robust watermarking.

2.1 Watermark Generation

We use binary images as watermark, of which the size is equal to the size of the 2nd level HH band (see Fig. 2). The original frame size is 720×480, after the two level decomposition the subband size is 180×120. For example, for a video sequence with resolution 720×480, the watermark could be an image of size 180×120, each pixel being binary (±1). The requirement is that the mean value of the watermark signals must be zero. This is because that in the watermark detection step we use the statistic model. In order to decrease the probability of error detection, the mean value of the signal is better to zero. We may also use a smaller image and repeat it several times to match the required watermark size.

2.2 Watermark Embedding

The frames are decomposed through DWT into four levels (see Fig.2) and all coefficients are quantized to *0 ~ 255*. Assume that the width and height of the *2nd* level coefficients are *M* and *N* respectively. Let the original frame be *I*, the resolution level be *l*, and the orientation be q. The decomposed image is expressed as I_l^q. We add watermark in the HH band (q = 2) of the *2nd* level (l = 1), i.e. I_1^2. We choose this subband because it has the characteristics of robustness and invisibility. The characteristic of the HH band is low energy, so we must raise the strength of watermark to resist more attacks. If we insert watermark in the first decomposition level, it may not survive under MPEG-2 DCT compression. We use the 2nd level HH band as the middle-band, which is not so sensitive to modification, to embed watermark. The watermark insertion equation can be expressed as:

$$\tilde{I}_1^2(i, j) = I_1^2(i, j) + rw(i, j)x(i, j) \qquad (1)$$

where r is a scalar factor, used to modulate the strength of the watermark, and $w(i, j)$ is a weighting function which is decided according to the HVS, and $x(i, j)$ is the watermark sequence.

According to HVS, human eye is less sensitive to high resolution and HH band, less sensitive to image areas where brightness is very high or very low, and less sensitive to highly textured areas, but more sensitive to area near the edge. Band sensitivity is a fixed value for each orientation and luminance-chrominance channel. The low-pass coefficients can provide background luminance. The spatial locality of the octave provides

distance information, while the energy of the lower frequency coefficients indicates the height of edge. The energy of lower frequency coefficients can indicate the texture activity level [5][6]. According to these characteristics, we have the equation:

$$w(i,\ j) = F(l,q)L(l,i,\ j)T(l,i,\ j)^{0.15} \qquad (2)$$

where *F(l,q)* is the noise sensitivity function which takes the orientation and level as input parameters. Let

$$F(l,\boldsymbol{q}) = \begin{Bmatrix} \sqrt{2}, & if\ \boldsymbol{q} = 2 \\ 1, & otherwise \end{Bmatrix} \bullet \begin{Bmatrix} 1.00, & if & l = 0 \\ 0.32, & if & l = 1 \\ 0.16, & if & l = 2 \\ 0.10, & if & l = 3 \end{Bmatrix} \qquad (3)$$

L(l,i,j) denotes the strength of the background brightness which affects the visual sensitivity. We use the lowest level DWT coefficients to measure this term. Recall that the eye is less sensitive to very dark or very bright areas. Let

$$L(l,i,\ j) = \begin{Bmatrix} 1 + S(l,i,\ j), & if\ S(l,i,\ j) > 0.7 \\ 2 - S(l,i,\ j), & if\ S(l,i,\ j) < 0.3 \\ 1, & otherwise \end{Bmatrix} \qquad (4)$$

where

$$S(l,i,\ j) = \frac{1}{256} I_3^3\left(1 + \left\lfloor \frac{i}{2^{3-l}} \right\rfloor, 1 + \left\lfloor \frac{j}{2^{3-l}} \right\rfloor\right) \qquad (5)$$

T(l,i,j) takes into account the activity in the neighborhood of a pixel. It can be expressed

$$T(l,i,\ j) = \sum_{k=0}^{3-l} \frac{1}{16^k} \sum_{q=0}^{2} \sum_{x=-1}^{1} \sum_{y=-1}^{1} \left[I_{k+l}^{q}\left(1 + x + \left\lfloor \frac{i}{2^k} \right\rfloor, 1 + y + \left\lfloor \frac{j}{2^k} \right\rfloor\right) \right]^2 \bullet Var\left\{ I_3^3\left(1 + p + \left\lfloor \frac{i}{2^{3-l}} \right\rfloor, 1 + q + \left\lfloor \frac{j}{2^{3-l}} \right\rfloor\right) \right\}_{q=-1,0,1}^{p=-1,0,1} \qquad (6)$$

The first term counts the distance from the edge, while the second one counts the texture. *Var* is the

variance of the 3×3 block. These two terms are multiplied because the eye is more sensitive to noise near edge, and is less sensitive to highly textured areas.

By adjusting the strength factor ρ, appropriate quality level of the watermarked image is selected. After watermark embedding, we put this watermarked frames into MPEG-2 encoder to generate MPEG-2 bitstream. Generally, to get a constant bit rate video sequence, the lower the bit rate shall imply lower image quality.

2.3 Watermark Detection

In the watermark detection step, we need not refer to the original image. All that we need to know is the watermark signal. We use the correlation between the watermarked image and watermark signals to decide whether the watermark is in the image or not. That is,

$$r = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\tilde{I}_1^2(i,j)x(i,j) \qquad (7)$$

where $\tilde{I}_1^2$ is the *2nd* level HH band coefficients of the watermarked frame. If our watermark is in the image without destroying, the value of *r* will be very large. Otherwise it may be smaller. We need a threshold value *Th* to decide if our watermark is in the image or not. If *r* is lager than *Th*, we would say that our watermark exists in that image. Otherwise, our watermark signal is not in that image. If some watermark *x'(i, j)* is in the image but it is not our watermark, the expression of the watermarked image would be:

$$\tilde{I}'^2_1(i,j) = I_1^2(i,j) + rw(i,j)x'(i,j) \qquad (8)$$

And the correlation between the image and our watermark signal would be:

$$r' = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I_1^2(i,j) + rw(i,j)x'(i,j)]x(i,j) \qquad (9)$$

The probability of false alarm P_f is selected as 10^{-8}. By statistical theory, we have:

$$P_f \le \frac{1}{2}erfc(\frac{Th}{\sqrt{2r'^2}}) \qquad (10)$$

$$10^{-8} \le \frac{1}{2} erfc(\frac{Th}{\sqrt{2r'^2}}) \quad \Rightarrow \quad Th = 3.97\sqrt{2r'^2} \qquad (11)$$

We need to find r'^2 of (11):

$$r'^2 = (\frac{1}{MN})^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} E[(I_1^2(i,j) + rw(i,j)x'(i,j))x(i,j)]^2 \qquad (12)$$

For simplification, we only count the expectation term:

$$\begin{aligned} &E[(I_1^2(i,j) + rw(i,j)x'(i,j))x(i,j)]^2 \\ &= E[I_1^2(i,j)^2 x(i,j)^2] + 2E[I_1^2(i,j)\, rw(i,j)x'(i,j)x(i,j)^2] + E[r^2 w(i,j)^2 x'(i,j)^2 x(i,j)^2] \\ &= E[I_1^2(i,j)^2]E[x(i,j)^2] + 2rE[I_1^2(i,j)]E[w(i,j)]E[x'(i,j)]E[x(i,j)^2] \\ &\quad + r^2 E[w(i,j)^2]E[x'(i,j)^2]E[x(i,j)^2] \end{aligned} \qquad (13)$$

Let $E[I_1^2(i,j)] = 0$, $E[x(i,j)] = 0$, $E[x'(i,j)] = 0$, $E[x(i,j)^2] = 1$, and $E[x'(i,j)^2] = 1$. Equation (13) becomes:

$$r'^2 = (\frac{1}{MN})^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} E[I_1^2(i,j)^2] + r^2 E[w(i,j)^2] \qquad (14)$$

Consider $E[\tilde{I}_1^2(i,j)^2]$:

$$\begin{aligned} E[\tilde{I}_1^2(i,j)^2] &= E[I_1^2(i,j)^2] + 2E[I_1^2(i,j)\, rw(i,j)x(i,j)] + E[r^2 w(i,j)^2 x(i,j)^2] \\ &= E[I_1^2(i,j)^2] + r^2 E[w(i,j)^2] \end{aligned} \qquad (15)$$

Form equations (14) and (15), we have:

$$\begin{aligned} r'^2 &= (\frac{1}{MN})^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} E[(I_1^2(i,j) + rw(i,j)x'(i,j))x(i,j)]^2 \\ &= (\frac{1}{MN})^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} E[I_1^2(i,j)^2] + r^2 E[w(i,j)^2] \\ &= (\frac{1}{MN})^2 \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} E[\tilde{I}_1^2(i,j)^2] \end{aligned} \qquad (16)$$

From Equation (16), it is found that the threshold value *Th* can be decided without knowing the watermarking strength r.

3. Simulation Results

In our simulation we use 'Flower garden' sequence to test the performance. The texture of this sequence is

very complex, and difficult to compress. The simulation results of our proposed algorithm are listed on Table 1. It can be seen that the PSNR difference between TM5 (MPEG-2 Test Mode 5) and our proposed algorithm is tiny.

Fig.3 shows the relationship between the detector response and the threshold value. By choosing larger ρ, almost all the detector responses with r are higher than the threshold. So, if we take the average r of 12 frames (GOP length) as the response value, all the frames could successfully be recognized as having our watermark. The Figs.4~ 6 show the visual effects of the first frame of sequence 'Flower garden'. Fig. 7 shows an example of detector response of recompressed video sequence in 6 Mbps. Since the television-quality video sequences are transmitted with bit rate between 4 Mbps to 12 Mbps [7], the sequence 'Flower garden' is also tested at the minimal bit rate 4 Mbps. Table 2 shows the simulation results.

Generally speaking, if the PSNR is less than 30 dB, the video quality is not acceptable. So, when the bit rate is less than 6 Mbps, the quality of 'Flower garden' is not good enough. The result of 4 Mbps with strength r = 1.5 for 'Flower garden' is not a good choice; the watermark cannot be recognized in some frames. But when we increase the strength r to 2.0, almost all the watermark of each frame can be recognized successfully, beside the last 40 frames. If we take the r's average of 12 frames (GOP length, in fact, at least 4 frames) as the response value, all the frames could successfully be recognized as having our watermark.

To further test the robustness of our algorithm, some attacks including blurring, cropping, sharpening, filtering, and random noise are applied to the 'Flower garden' sequence. The simulation results are shown in Table 3.

4. Conclusions

In this paper, we have presented a robust watermarking applied in uncompressed video sequence. Simulation shows that the proposed watermarking method could even be survived after recompressing the video sequence and keeps an acceptable quality in watermarked image. In future work, we will make a major challenge to further design a method with fragile watermarking and robust watermarking

in order to protect the intellectual property rights for seller and buyer.

References

[1] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, 66:283-301, 1998.

[2] B. Zhu M. D. Swanson and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE Journal on Selected Areas in Communications*, 16(4): 540-550, May 1998.

[3] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Real-time labeling methods for MPEG compressed video," in *Proc. 18th Symp. Information Theory in the Benelux, Veldhoven*, The Netherlands, May 1997.

[4] I. J. Cox and J. -P. M. G. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas in Communications*, 16(4): 587-593, May 1998.

[5] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Processing*, vol.1, No. 2, pp. 244-250, Apr. 1992.

[6] Mauro Barni, Franco Bartolini, and Alessandro Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Processing*, vol. 10, No. 5, pp. 783-791, May 2001.

[7] CCIR Recommendation 601-1:1990, "Encoding parameters of digital television for studios."

Figure List

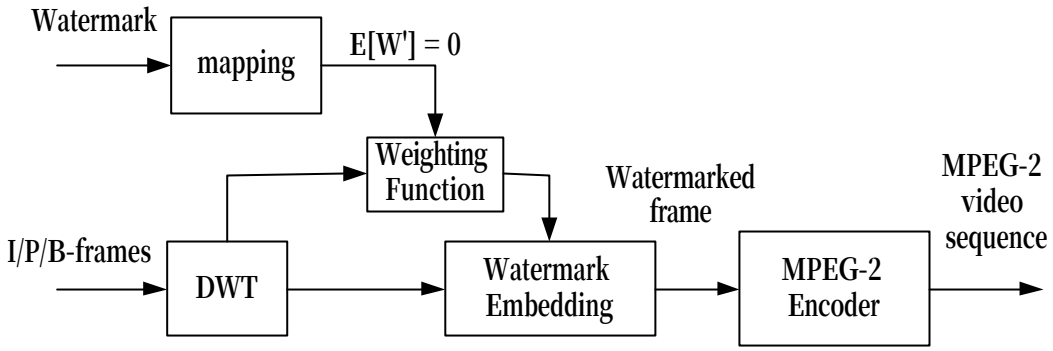


Fig.1. The block diagram of the proposed robust watermarking

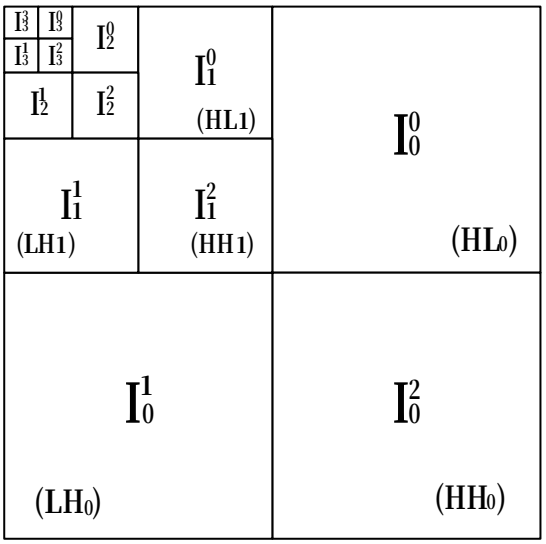
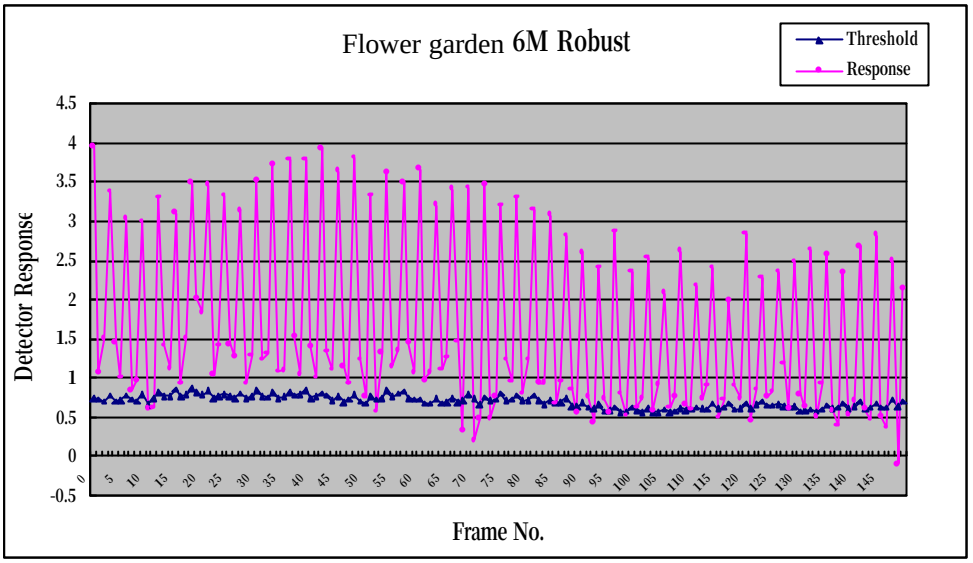
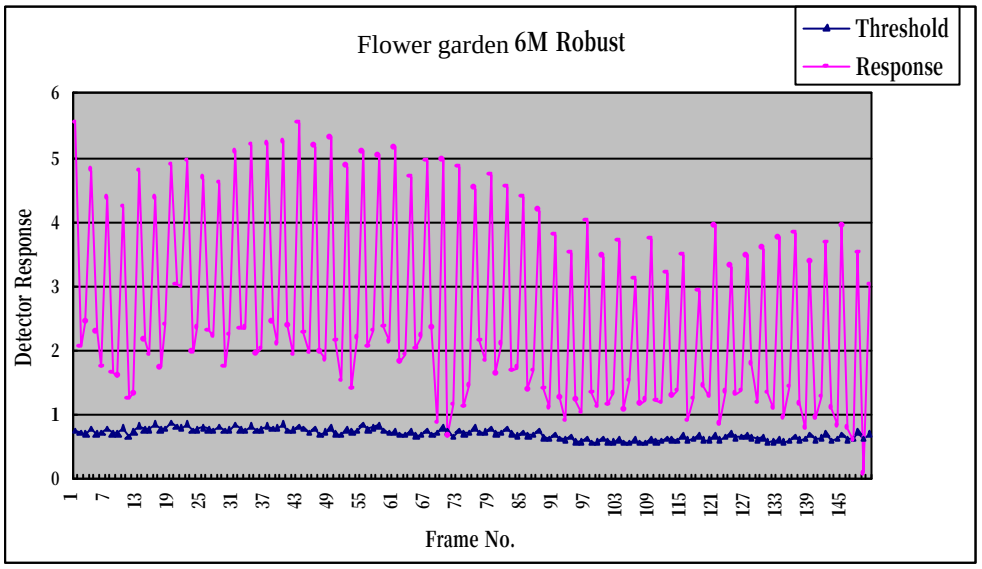


Fig. 2. Four resolution levels DWT decomposition.



(a)



(b)

Fig.3. Threshold value v.s.detector response at bit rate 6 Mbps under (a) strength value ρ=1.5;(b) ρ=2.0

(a)

(b)

(c)

Fig.4 The Flower sequence for robust watermarking test; (a) first frame of video sequence ‘Flower garden’; (b) watermarked image of the first frame of sequence ‘Flower’. (PSNR is 42.2 dB with ρ=2.0); (c)MPEG-2 encoded watermarked image of the first frame of sequence ‘Flower’. (PSNR is 31.63 dB under 6 Mbps and ρ=2.0)

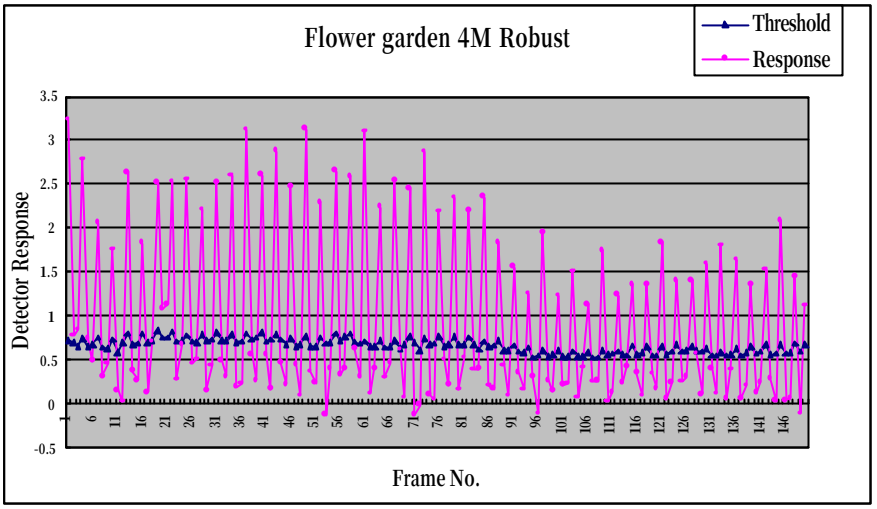


Fig.5. Threshold value and detector response at bit rate 4 Mbps under strength value ρ=1.5.

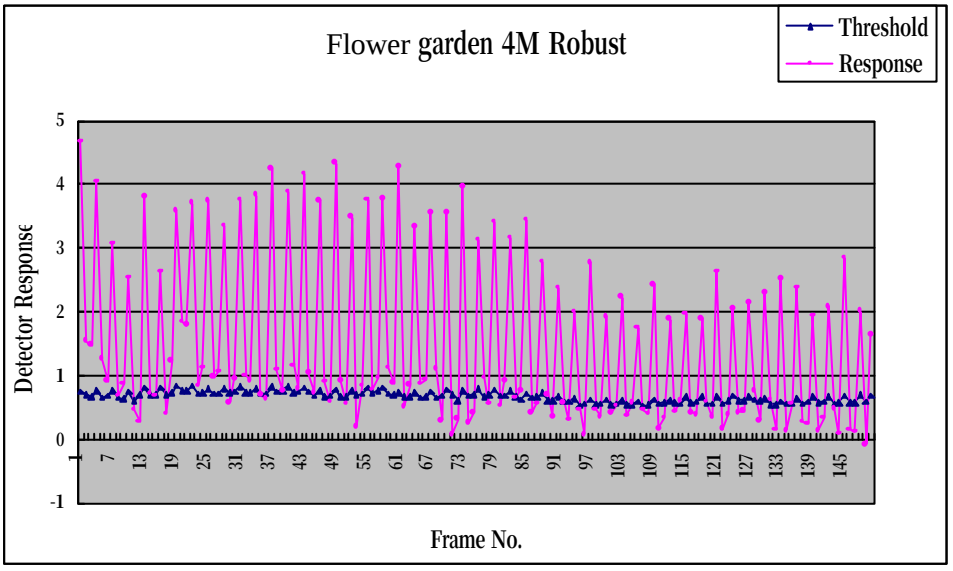


Fig.6. Threshold value and detector response at bit rate 4 Mbps under strength value ρ=2.0.

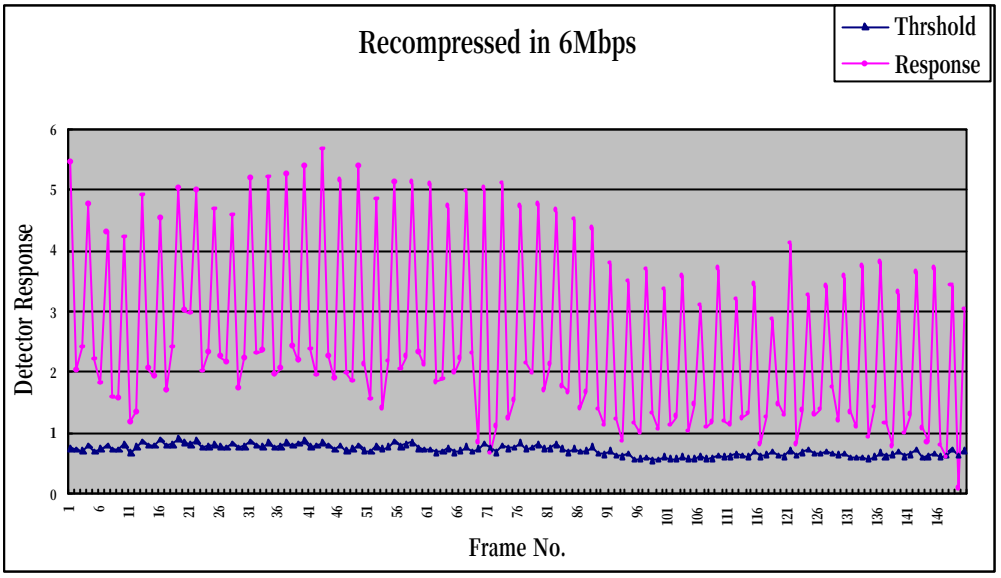


Fig. 7. Detector response of recompressed video ‘ Flower Garden’ in **6 Mbps**, GOP=12, M=3, N=12, the average of PSNR in this sequence is 29.4 dB.

Table List

Table 1. PSNR comparisons between TM5 and the proposed algorithm of sequence ‘ Flower garden’ .

Environment	**Algorithm**	**PSNR (dB)**	
Data rate = 6 Mbps, N = 12, M = 3, Frame rate = 30fps, 720×480, 4:2:0	TM5	30.40	
	Proposed Algorithm	ρ =1.5	30.16
		ρ =2.0	30.01

Table 2. PSNR comparison between TM5 and the proposed algorithm of sequence ‘ Flower garden’ .

Environment	**Algorithm**	**PSNR (dB)**	
Data rate = 4 Mbps, N = 12, M = 3, Frame rate = 30fps, 720×480, 4:2:0	TM5	28.13	
	Proposed Algorithm	ρ =1.5	28.03
		ρ =2.0	27.93

Table 3. Various attacks for the proposed robust algorithm. (ρ=2.0, GOP=12, M=3, N=12, 6Mbps).

Attack	PSNR(dB)	Th	Response
Blurring	22.1063	0.1700	0.8680
Cropping	4.7123	0.2672	0.3949
Sharpen	26.3468	1.0153	7.7198
3x3 Filter	22.6139	0.3226	1.8256
25% Random Noise	17.7488	1.4265	5.2645
50% Random Noise	14.8307	1.8621	5.0060