A Collusion-Resistant Video Watermarking Scheme

Amir Houmansadr[1] and Shahrokh Ghaemmaghami[2]

[1] Electrical Engineering Department, Sharif University of Technology, Tehran, Iran
houmansadr@mehr.sharif.edu
[2] Electronics Research Center, Sharif University of Technology, Tehran, Iran
ghaemmag@sharif.edu

Abstract. A video watermarking scheme is proposed in this paper using the concept of the secret sharing scheme. The owner's mark is split into twin shares, where the shares are inserted into the video frames in the spatial domain in a simple manner. The detection algorithm uses a linear function applied to the twin shares to reconstruct the secret. This makes the watermarked video sequence robust against pirate attacks, such as frame averaging and frame swapping. Due to the compatibility of the exploited secret sharing scheme to geometrical distortions, the watermarking system is also robust to this kind of processing schemes. On account of insertion of various marks into different frames, which are linearly related, the watermarked sequence is robust to collusion attack that is a major concern in the field of video watermarking.

1 Introduction

Illegal copying and distribution of digital media has made the owner's rights to be more and more frequently violated. Traditional solutions for copyright protection, such as encryption, can no longer protect digital contents by themselves. Sooner or later, encrypted media have to be revealed for the aim of consumer's usage that may be the malicious one. At the end of 20^{th} century, digital watermarking was introduced as a complementary solution to protection of digital media ownership.

In copyright protection applications, a digital watermark is an invisible mark that is inserted into a digital media such as audio, image, or video, which is used to identify illegal distributions of copyright protected digital media and also lawbreaking customers. A digital watermark should have certain features to achieve desired functionalities in this case. The embedded mark is to be robust enough against various watermarking attacks, while keeping the perceived quality of the host image unchanged (the imperceptibility requirement). Watermarking attacks consist of deliberate attacks made maliciously to remove or change the mark sequence by lawbreakers and unintentional attacks caused as a result of different kinds of coding and compression made to the digital media prior to transmission and/or storage and also errors occurred during the transmission of the media through networks.

J. Camenisch et al. (Eds.): IH 2006, LNCS 4437, pp. 343–354, 2007.

Video contents can be mentioned as the most valuable digital media, which are increasingly used illegally, resulting in a huge damage to filmmaking industry. Video watermarking is utilized for different video applications such as copyright protection, fingerprinting, broadcast monitoring, copy protection, and so on [1]. Distinct challenges have arisen in this field, as compared to image watermarking. Because of the more possibilities to perform the collusion attack on video streams, it is a main concern in designing video watermarking systems. Collusion refers to using some watermarked data that is utilized for the aim of watermark removal.

The main goal of this paper is to design a watermarking scheme for video sequences which is robust to collusion attack. In Sect. 2, the main concept of secret sharing is introduced. Sect. 3 describes the proposed insertion and detection watermarking schemes based on the mentioned secret sharing scheme. The collusion attack, in the proposed scheme, is analyzed in Sect. 4 and simulation results are presented in Sect. 5. Finally, the paper is concluded in Sect. 6.

2 Visual Secret Sharing

A secret sharing scheme shares a secret into a number of shares so that the cooperation of a predetermined group of shareholders reveals the secret, while the secret reconstruction is impossible to any unauthorized set of shareholders. *Naor et al.* in [2] proposed a 2-dimentional secret sharing scheme which is known as visual secret sharing (VSS). Since we are using this scheme in the proposed watermarking scheme in this paper, VSS scheme is described in this section.

VSS scheme shares a binary-valued image, which is known as secret image, into two double-sized images so that reconstruction of the secret image from these twin images can be done only if both of them are available. So, a VSS system is composed of the following components:

- Secret image: a digital image composed of $M \times N$ white and black pixels, whose anonymity is the goal of the system;
- VSS sharing scheme: derives two share-images from a secret image in a pseudo-random manner;
- Share-images: digital images composed of $2M \times 2N$ white and black pixels, that are driven from the secret image in a pseudo-random manner. Two share-images are produced in every run of the VSS sharing scheme, known as twin share-images. Different runs of the VSS scheme generates different share-images, and each of these share-images reveals no information about the secret image unless its twin, i.e. the share-image generated in the same run of the VSS sharing scheme, is available;
-VSS reconstruction scheme: retrieves the secret image from every corresponding couple of share-images, i.e. twin share-images. VSS reconstruction scheme is lossless if share-images have not been distorted in any way.

According to the VSS sharing scheme, each pixel in the secret image is split into two 2×2 blocks of pixels, which are chosen form the blocks shown in Fig. 1.

This leads to two double-sized share-images for every secret image. For the aim of sharing a white pixel from the secret image, two corresponding share blocks within the twin share-images are chosen the same. In other words, one of the six blocks in Fig. 1 is selected for both of the share-images. On the other hand, if we aim to share a black pixel from the secret image, different blocks from the same type of blocks are chosen, e.g. two different horizontal share blocks. Therefore there is 6 alternatives to share either a black or a white pixel and there are $6^{M\times N}$ solutions for the problem of sharing an $M\times N$ pixels binary-valued secret image. Fig. 2 illustrates the twin share-images corresponding to the shown secret image.

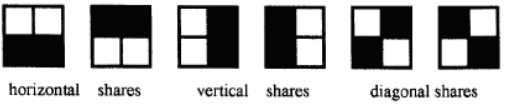

Fig. 1. Different blocks which are used to share a pixel in the secret image

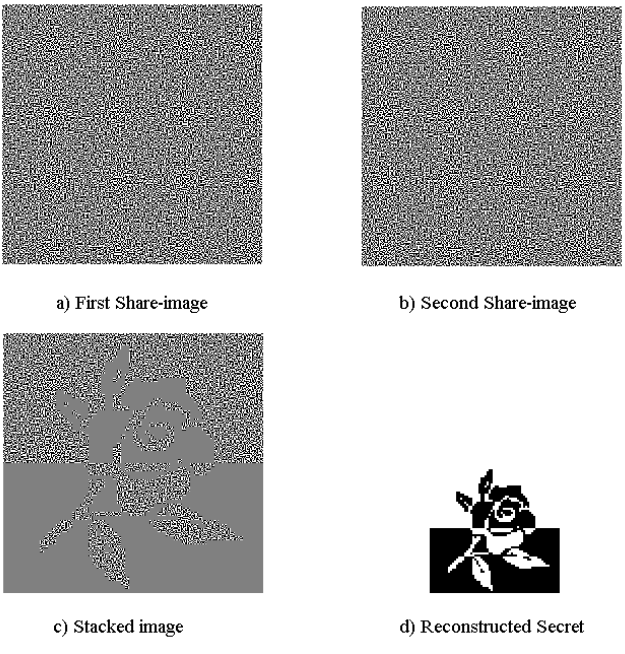

Fig. 2. Different blocks which are used to share a pixel in the secret image

Different mechanisms can be devised for the aim of reconstructing an $M\times N$ pixels secret image, S, from one of its twin share-images, *SH1*, and *SH2*. Fig. 3 shows the scheme of a simple system which we propose to be used as the VSS reconstruction scheme in this paper.

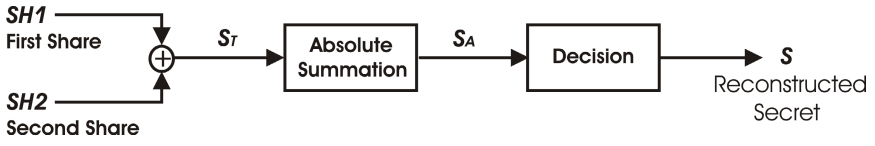

Fig. 3. Block diagram of the proposed VSS reconstruction scheme

First, twin share-images are added together to generate S_T, which we call it the stacked-image. This is because addition of twin share-images resembles printing them on two transparent sheets and then stacking them together. By allocating +1 and -1 values to white and black pixels respectively in the share-images, pixels of S_T will have one of the +2, -2, or 0 values. Recalling the VSS sharing scheme mentioned above, if S_T is divided into non-overlapping blocks of 2×2 pixels, each block corresponding to a white pixel of the secret image have two +2 and two -2 values, while every block corresponding to a black pixel have four 0 values. So, by applying an absolute summation over every block of S_T as in (2), we can decide whether the block represents a white or a black pixel in the secret image. This is done as below:

$$S(x,y) = \begin{cases} +1 & if \quad S_A(x,y) = 8 \\ -1 & if \quad S_A(x,y) = 0 \end{cases} , \tag{1}$$

where:

$$S_A(x,y) = \sum_{m=0}^{1} \sum_{n=0}^{1} |S_T(2x-m, 2y-n)| \quad x = 1..M, y = 1..N \ . \tag{2}$$

The proposed reconstruction scheme acts as a lossless reverse function for the mentioned VSS sharing scheme. As we will see in the next section, share-images are inserted as digital watermarks into video frames. In a watermarking system, it is expected that the inserted marks get distorted because of different losses due to the noisy channel, watermark extraction scheme, and so on. As a result, we modify the mentioned reconstruction scheme to be used in the proposed watermarking scheme efficiently:

$$S(x,y) = \begin{cases} +1 & if \quad S_A(x,y) > 4 \\ -1 & if \quad S_A(x,y) \leq 4 \end{cases} . \tag{3}$$

3 The Proposed Watermarking Scheme

In the proposed watermarking scheme watermark, W, is a sequence of $M \times N$ bits (+1 and -1 values), where every frame of the video sequence is $2M$ by $2N$ pixels in size. The video stream is first divided into several successive GOPs (Group Of Pictures) with the length of L, where L is an even number, e.g. 12. Considering the $M \times N$ bits watermark sequence as an M by N pixels image, for the i-th GOP , i.e. $F_{i,j}$, $j=1...L$, the VSS scheme is performed $L/2$ times to split the watermark

image, W, into L sub-watermarks, i.e. $W_{i,j}$, j=1...L (two sub-watermarks are produced in every run of the VSS scheme). These sub-watermarks are inserted into the frames of the corresponding GOP as:

$$F_{i,j}^{W} = F_{i,j} + JND_{i,j}.Per_j(\{W_{i,j}|j = 1..L\}), \quad i = 1..Num \ , \tag{4}$$

where $F_{i,j}^{W}$ is the j-th frame of the i-th GOP in the watermarked video sequence, $JND_{i,j}$ is the weighting coefficient corresponding to $F_{i,j}$, Num is the number of GOPs in the video sequence, and $Per(.)$ applies a permutation to the sub-watermarks of the i-th GOP by changing their order of appearance. As a simple permutation function, modular permutation can be used as:

$$Per_j(\{W_{i,j}|j = 1..L\}) = W_{i,m}, \quad m = \mathrm{mod}(p.j, L) + 1 \ . \tag{5}$$

mod(x,y) is the modular residue of x with respect to y. Mathematically, if p is an integer number which is prime relative to L, the original video frames and the permuted video frames are related through a one-to-one relationship.

Choosing the $JND_{i,j}$ coefficients equal to a constant number leads to a simple and fast watermarking scheme, while a more robust watermarked video stream would be achieved, if the coefficients are adopted to the video frames as cited in the next section.

Fig. 4 shows the block diagram of the watermark extraction scheme. First, a noise estimator block is performed on the received possibly watermarked video sequence. Since the embedded sub-watermarks are noise-like, this leads to an efficient estimation of them as in (6). To design a noise estimator, different approaches have been suggested in the literature [3,4,5]. Our simulations show that using a simple FFT (Fast Fourier Transform) filter provides a fast and effective estimation of the inserted sub-watermarks. Fig. 5 shows the basic structure of the utilized FFT filter. The two-dimensional FFT transform of the video frame, IM, is passed through a masking stage which drops its low-frequency components and then an inverse two-dimensional FFT transform is performed.

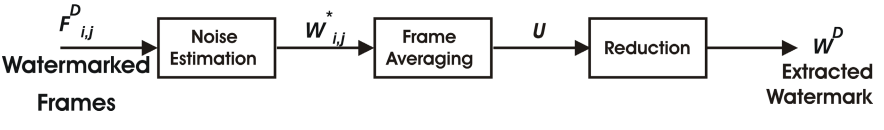

Fig. 4. Block diagram of the watermark extraction scheme

Fig. 5. Block diagram of the FFT filtering

$$W^*_{i,j} \approx W_{i,j} = \frac{1}{JND_{i,j}}(F^W_{i,j} - F_{i,j}), \quad i = 1..Num, j = 1..L \ . \tag{6}$$

After noise estimation, $W^*_{i,j}$ is an appropriate estimation of the inserted sub-watermark $W_{i,j}$. To retrieve the original watermark, W, from the sequence of estimated sub-watermarks, W^* sequence is passed from two more blocks. First, an average is computed over the frames of the resulting video sequence as:

$$U = \frac{2}{Num.L} \sum_{i=1}^{Num} \sum_{j=1}^{L} W^*_{i,j} \ . \tag{7}$$

The resulting $2M \times 2N$ pixels image, U, is then passed from a reduction function which returns the $M \times N$ pixels extracted watermark as:

$$W^D(x,y) = \begin{cases} +1 & if \quad R(x,y) > 4 \\ -1 & elsewhere \end{cases} \ , \tag{8}$$

where $W_D(x,y)$ is the (x,y)-th pixel of the extracted watermark and $R(x,y)$ is defined as:

$$R(x,y) = \sum_{i=0}^{1} \sum_{j=0}^{1} |U(2x-i, 2y-j)|, \quad 1 \le x \le M, \quad 1 \le y \le N \ . \tag{9}$$

Finally, a normalized correlation is evaluated between W_D and the watermark sequence, W, as:

$$\rho = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} W^D(x,y).W(x,y)}{\sqrt{\sum_{x=1}^{M} \sum_{y=1}^{N} W^D(x,y). \sum_{x=1}^{M} \sum_{y=1}^{N} W(x,y)}} \ . \tag{10}$$

This correlation is compared by a threshold value, TH, to decide if the watermark W exists in the video sequence received.

The main idea behind the definition of reduction function is the structure cited for VSS reconstruction scheme in the previous section. In fact, passing average of the frames, U, through the reduction function is equivalent to applying the mentioned reconstruction function to the twin share-images and then returning the average value.

4 Collusion Analyses

Collusion refers to a set of users who merge their knowledge to have access to the unwatermarked contents. Collusion can be performed in two different manners. In collusion type-I the same watermark is embedded into different data, which can be estimated by a linear combination and removed from the watermarked

contents. On the other hand, collusion type-II refers to the case where different watermarks are embedded into different copies of the same data. In this case colluders can obtain the unwatermarked data by a simple linear combination of different copies, e.g. averaging. This is because averaging different watermarks generally converges toward zero.

There are also two different approaches to implementation of collusion attack in the case of video watermarking. Inter-videos collusion refers to a number of users who have different videos containing the same watermark, or the same videos with different embedded watermarks, where collusions type-I and II could be applied respectively. Inter-videos collusion is the same as what have been considered for still images, so the solutions can be borrowed from the literature. For instance, inserting a Trusted Third Party in the watermarking system, which produces and encrypts hash of the host data, is proposed to prevent collusion type-I. Also, traditional countermeasures exist for collusion type-II which are based on the projective geometry or the theory of combinational designs [1].

In the case of video watermarking, there is another kind of collusion which is a video-specific origin. Intra-video collusion is the main threat to video watermarking, because a watermarked video alone is enough to remove the watermark. Inserting the same watermark in each frame, which is the baseline of many video watermarking schemes, makes collusion type-I feasible exploiting frames of the video sequence as watermarked images. On the other hand, by inserting different watermarks into different frames, collusion type-II can be implemented in static scenes, since there are similar frames with different watermarks. Intra-video collusion is considered in this research which is investigated in the proposed watermarking scheme in the following sections.

4.1 Linear Collusion

For a set of watermarked frames $F_k^W = F_k + \beta_k W_k$, $k=1,..,(Num.L)$, and their corresponding raw video frames, F_k , the linear collusion attack is made as:

$$\overline{X} = \sum_{k=1}^{L} \beta_k F_k^W = \sum_{k=1}^{L} \beta_k F_k + \sum_{k=1}^{L} \beta_k \alpha_k W_k^* \ , \tag{11}$$

where W_k^* is the possibly distorted watermark sequence, and β_k is a weighting coefficient. $\overline{X}$ gives an optimal MSE (Mean Squared Error) estimate of the watermark or the host signal in the case of collusions type-I or type-II, respectively [6].

In the proposed watermarking scheme, different sub-watermarks are inserted into different frames. As a result, collusion type-I is entirely infeasible. In fact, collusion type-I needs some video frames containing the same watermark to be estimated by some linear combination such as frame averaging. Even if the original watermark, W, is estimated by attacker in some way, it can not be used to produce the unwatermarked video sequence; this is because what is inserted into video frames is not the original watermark, W, itself but sub-watermarks, $W_{i,j}$, which has been obtained from it in a pseudo-random manner

during different runs of the VSS scheme. So, we just have to investigate collusion type-II on the proposed scheme.

The main idea in this research to defeat collusion is to insert different sub-watermarks into video frames so that a linear combination of them results in the main watermark sequence. As we mentioned in Sect. 3, the watermark is extracted by performing a linear combination on the video frames, i.e. averaging (see (7) and (8)).

Collusion type-II, e.g. averaging, is performed by modifying a number of successive frames in still regions of the watermarked video sequence, F_i^W , *i=1,..,k*, as:

$$\begin{aligned} \overline{F_i^W} &= \tfrac{1}{k}\sum_{j=1}^{k} F_j^W = \tfrac{1}{k}\sum_{j=1}^{k} F_j + \tfrac{1}{k}\sum_{j=1}^{k} JND_j W_j^* \\ &\approx F_i + \tfrac{1}{k}\sum_{j=1}^{k} JND_j W_j^* \end{aligned} \quad , \tag{12}$$

where the second line of the above equation is valid in still regions of video sequence. So, evaluating U from (6), (7), and (12) is as follows:

$$\begin{aligned} U &= \tfrac{2}{Num.L}\sum_{i=1}^{Num}\sum_{j=1}^{L} W_{i,j}^* = \tfrac{2}{Num.L}\sum_{p=1}^{Num.L} W_p^* \\ &= \tfrac{2}{Num.L}\sum_{p=1}^{k} W_p^* + \tfrac{2}{Num.L}\sum_{p=k+1}^{Num.L} W_p^* \\ &= \tfrac{2}{Num.L}\sum_{p=1}^{k} \tfrac{1}{JND_p}\tfrac{1}{k}\sum_{j=1}^{k} JND_j W_j^* + \tfrac{2}{Num.L}\sum_{p=k+1}^{Num.L} W_p^* \end{aligned} \quad . \tag{13}$$

Even if the *JND* coefficients are not constant, they are very similar in still regions, because they depend on the host frames. So, U is evaluated as:

$$\begin{aligned} U &= \tfrac{2}{Num.L}\sum_{p=1}^{k}\tfrac{1}{k}\sum_{j=1}^{k} W_j^* + \tfrac{2}{Num.L}\sum_{p=k+1}^{Num.L} W_p^* \\ &= \tfrac{2}{Num.L}\sum_{j=1}^{k} W_j^* + \tfrac{2}{Num.L}\sum_{p=k+1}^{Num.L} W_p^* \\ &= \tfrac{2}{Num.L}\sum_{j=1}^{Num.L} W_j^* \end{aligned} \quad , \tag{14}$$

which is the same as (7). So, linear collusion has no effect on the detection process of the proposed scheme.

We simulated the collusion type-II on *hawk3* video sequence which was watermarked using the proposed scheme. As mentioned earlier, watermark detection in the proposed scheme is performed by evaluating a normalized correlation and comparing it by an appropriate detection threshold. Choosing this threshold is a tradeoff between minimizing wrong rejection and wrong confirmation of the watermark. This threshold should be chosen in respect to the average True to False detection Ratio (TFR) which is opted to 0.15 in our simulations. A watermarking attack to be effective should decrease the correlation coefficient below this detection threshold, making the watermark signal undetectable. So, we investigated

the effect of collusion on the watermarked video by surveying the amount of decrement enforced to the correlation coefficient. The mentioned video sequence has also been watermarked by CDMA scheme proposed by *Mobasseri* [7], which is a well-known similar video watermarking scheme, and the effect of collusion type-II on two schemes has been compared. To make a fair judgment, both watermarked sequences have the same watermark energy. Fig. 6 illustrates the effect of collusion type-II on the watermarked sequences versus number of frames exploited in performing the collusion attack. Simulations show that CDMA scheme is clearly vulnerable to collusion attack and the watermark is undetectable as the number of colluded frames increases. In contrary, the proposed scheme which is fundamentally similar to CDMA scheme shows a great amount of robustness to this kind of attack. As the number of colluded frames grows, detection coefficient in the proposed scheme varies around a fixed value near the correlation coefficient of the collusion-free detection.

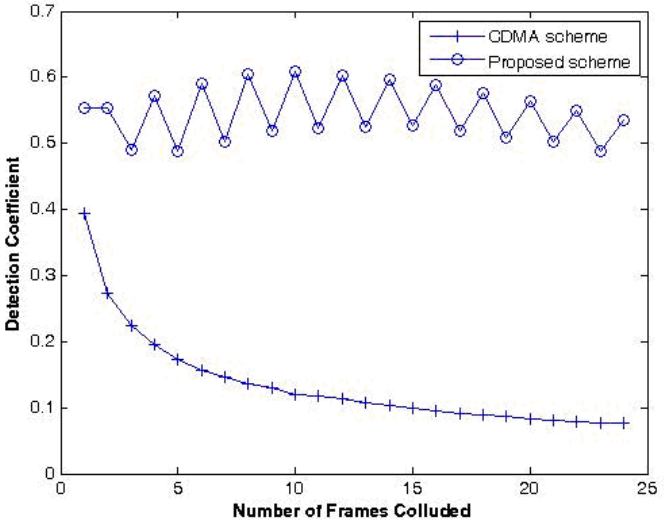

Fig. 6. Block diagram of the FFT (Detection coefficient vs. number of frames used for making collusion type-II for proposed scheme and CDMA watermarking scheme. Watermarks have the same energy in two schemes.

As it can be seen, simulation results are in conformity with mathematical analysis presented earlier regarding robustness of the proposed watermarking scheme against collusion attack. The alternating behavior of the proposed scheme in Fig. 6 is due to the fact that for odd number of frames one of the frames missing its twin frame acts as noise which reduces the system functionality; this effect decreases as the number of colluded frames increases.

4.2 Generalized Collusion

Even if the collusion is not linear, the watermark can be extracted efficiently. As described in the previous section, collusion type-I is infeasible due to inserting different watermarks into different frames. So, we just have to investigate collusion type-II.

As mentioned previously, collusion type-II is performed over the still regions. So, we propose to use only moving objects of video frames in evaluating U from (7) because colluders cannot change the sub-watermarks in these regions. As described in Sect. 3, corresponding shares of the watermark are inserted in the frames belonging to the same GOP. It is supposed that there are common moving areas in the frames belonging to a GOP, so a part of the main watermark can be extracted by superimposing the moving parts of every twin sub-watermark.

According to visual models, human eye decreases its sensitivity in high entropy regions, i.e. moving areas in the video sequences. So, by evaluating *JND* coefficients in an adaptive manner as in [8], the system robustness to collusion and other attacks will be elevated.

5 Other Attacks

We simulated the proposed watermarking scheme using *Matlab7* software. A constant value of 3 is chosen for *JND* coefficients, which preserves the quality of watermarked sequences according to subjective experiments. Using adaptive *JND* values leads to a more robust watermarking system at the expense of more computational complexity. According to mathematical analysis and simulations presented in the previous section, linear collusion makes an ignorable difference to the extracted watermark. Also, other watermarking attacks have been considered in the proposed scheme.

We applied different geometric distortions to the watermarked sequence to see how the detection response alters. In the case of video watermarking, the attacker has to perform the same geometric distortion on all of the frames to keep the continuity of the video sequence. By performing spatial synchronization prior to detection, output of the detection algorithm following various amounts of frame cropping, frame rotating, and changing the Aspect Ratio (AR) showed a high resilience against such distortions. As discussed in Sect. 3, decision on the watermark existence is made by evaluating a correlation coefficient. Tables 1 to 3 show the decrement of this correlation coefficient after performing frame cropping, frame rotating, and changing the AR of the watermarked video sequence, respectively. This high resistance to geometric attacks is due to the VSS compatibility with this kind of distortions which is further discussed in [9].

Also, the proposed scheme has brilliant robustness against some common pirate attacks. Changing the video bit rate, which is usually performed by a linear combination of frames, has little effect on the correlation coefficient. Because detection scheme is independent from the order of frames, frame swapping makes

Table 1. Decrement of ρ after frame cropping

Cropping Percentage	10	20	30	40	50
Decrement of ρ (%)	3	5	3	6	4

Table 2. Decrement of ρ after frame rotating

Rotation Angle (*degrees*)	5	10	15	20	25
Decrement of ρ (%)	7	6	8	12	15

Table 3. Changing the AR of 240*360 pixels watermarked frames

New Size (*pixels*)	$240 * 180$	$240 * 90$	$480 * 360$
Decrement of ρ (%)	21	18	23

nothing to the extracted watermark. Also, frame dropping makes little changes to the extracted watermark, which is evaluated by averaging a pool of share-images.

Finally, temporal synchronization, which is crucial in the detection stage of many video watermarking schemes, is not needed in the proposed scheme because detection is independent from the order of frames.

As expected from its simple structure, the used VSS scheme is very fast. This leads the proposed watermarking system to be implemented in real-time using *Matlab7* software.

6 Conclusions

In this paper, we have proposed a novel video watermarking scheme, based on the concept of visual secret sharing. It is shown that the watermarked video sequence is robust to linear collusion and, by performing a more complex detection scheme, i.e. using moving areas, the watermark can be extracted in the presence of any kind of collusion. This robustness is based on the fact that the embedded watermark can be extracted by a linear combination between different share-images, i.e. sub-watermarks, which are inserted into different frames of the watermarked sequence. This linear combination also makes the watermarking system robust to pirate attacks, such as frame dropping, frame swapping, and changing the rate of video frames. No temporal synchronization is needed for the aim of watermark extraction due to this linear combination. Also the watermarked sequence is robust to geometrical distortions, which is due to compatibility of the VSS scheme with this kind of distortions. The proposed watermarking system is fast and is implemented in real-time.

References

1. Doer, G., Dugelay, J.L.: A guide tour of video watermarking. In: Signal processing: Image communications, vol. 18, pp. 263–282. Elsevier Science, North-Holland, Amsterdam (2003)
2. Naor, M., Shamir, A.: Visual cryptography. In: De Santis, A. (ed.) EUROCRYPT 1994. LNCS, vol. 950, pp. 1–12. Springer, Heidelberg (1995)
3. Voloshynovskiy, S., Herrigel, A.N.B., Pun, T.: A stochastic approach to content adaptive digital image watermarking. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 212–236. Springer, Heidelberg (2000)
4. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7), 629–639 (1990)
5. Torkamani-Azar, F., Tait, K.E.: Image recovery using the anisotropic diffusion equation. IEEE Trans. on Image Processing 5(11), 1573–1578 (1996)
6. Su, K., Kundur, D., Hatzinakos, D.: A novel approach to collusion-resistance video watermarking. In: Proceedings of SPIE Security of Watermarking of Multimedia contents IV. San Jose, CA, pp. 491–502 (2002)
7. Mobasseri, B.G.: Exploring CDMA for watermarking of digital video. In: Proceedings of the SPIE 3657, pp. 96–102 (1999)
8. Podilchuk, C.I., Zeng, W.: Image-adaptive watermarking using visual models. IEEE Journal on selected areas in communications 16(4), 525–539 (1998)
9. Houmansadr, A., Ghaemmaghami, S.: A digital image watermarking scheme based on the visual cryptography. In: Proc. 3rd Int'l Symposium on Telecommunications, pp. 843–848 (2005)