International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:8, 2009

Inter-frame Collusion Attack in SS-N Video Watermarking System

Yaser Mohammad Taheri, Alireza Zolghadr–asli, and Mehran Yazdi

***Abstract*—**Video watermarking is usually considered as watermarking of a set of still images. In frame-by-frame watermarking approach, each video frame is seen as a single watermarked image, so collusion attack is more critical in video watermarking. If the same or redundant watermark is used for embedding in every frame of video, the watermark can be estimated and then removed by watermark estimate remodolulation (WER) attack. Also if uncorrelated watermarks are used for every frame, these watermarks can be washed out with frame temporal filtering (FTF). Switching watermark system or so-called SS-N system has better performance against WER and FTF attacks. In this system, for each frame, the watermark is randomly picked up from a finite pool of watermark patterns. At first SS-N system will be surveyed and then a new collusion attack for SS-N system will be proposed using a new algorithm for separating video frame based on watermark pattern. So N sets will be built in which every set contains frames carrying the same watermark. After that, using WER attack in every set, N different watermark patterns will be estimated and removed later.

***Keywords*—**Watermark estimation remodulation (WER), Frame Temporal Averaging (FTF), switching watermark system.

I. Introduction

NOWADAYS security issue in digital media has taken more attention among researchers. Every day a large numbers of Digital media contents same as VCD and DVD's are distributed by copyright owners. Digital content can be copied rapidly, at large scale, without any limitation by malicious customers. Because of economic reasons, content owners still don't have any desire to distribute their productions. Due to this fact, paying more attention to increasing security in digital media is essential. So Digital media, like audio, video, images, and other multimedia documents, should be protected against illegitimate copy and manipulation. Although completely secure media is impossible to be produced, increasing level of security can decrease possible detriment to the acceptable amount. For this purpose, digital watermarking is used for embedding information into digital material in such a way that it is imperceptible to a human observer. One of the challenging issues in video watermarking is collusion attack in video data. In this kind of attack, colluders can gather some information about watermarking system and use them to defeat watermarking system. One kind of collusion attacks in video is inter-frame collusion attack. The basic idea behind the inter-frame collusion attack in order to estimate the redundant component is the exploitation of the redundancy, either in the host video frames or in the embedded watermark [1].

Most of the video watermarking algorithms proposed so far considered the video as a sequence of still images [2] and applied existing still image watermarking techniques to each frame. Considering this fact, frame-by-frame embedding strategies have been proposed. Two main embedding strategies based on frame-by-frame approach exist. In the first strategy, different and uncorrelated watermarks are inserted in each video frames. In this watermarking system, since such uncorrelated watermarks are in the temporal high frequency band, watermarks can be removed by temporal low-pass filtering of the watermarked frames. This attack is generally known as frame temporal filtering attack (FTF) [3] in which the attack is more relevant in static scene.

In second system, the same watermark is inserted in every frame [4]. In this system if attacker can collect visually dissimilar frames from video, a rough estimate of the watermark can be obtained using the difference between a watermarked frame and its spatial low-pass filtered version. A better estimate of the watermark can be obtained by averaging the individual estimates obtained from different frames. This attack is known as the watermark estimation remodulations (WER) attack [5] in which the attack is more relevant in dynamic scenes. Doe¨rr et.al explored Watermark modulation for each video frame. In this approach the watermark is picked out from a finite pool of reference watermark patterns. Superiority of this strategy in terms of securities demonstrated both theoretically and experimentally by them.[6]

In this paper, we propose a new inter-frame collusion attack that can defeat such system. At first, this kind of system in which watermark is picked up from N reference watermark patterns is explained and then a new algorithm for collusion would be proposed.

II. Switching Watermark System

This watermarking system was first proposed by Doe¨rr et.al.[6]. They introduced it as SS-N system. In this watermarking system, watermark is picked out from a finite pool of orthogonal watermarks for embedding in each frame. On other word, embedder should randomly switch between

F. Y. Mohammad taheri is Ms student of telecommunication engineering in University of Shiraz, Shiraz, Iran (e-mail: yasertaheri@ yahoo.com).

S. A. Zolghadr-asli, is Associate professor in Department of Electrical engineering, University of Shiraz, Shiraz, Iran (e-mail: zolghadr@shirazu.ac.ir).

T. M. Yazdi is Assistant professor in Department of Electrical Engineering, University of Shiraz, Shiraz ,Iran.

finite numbers of watermark patterns for each frame of video. In SS-N system, for each video frame, watermark should be chosen randomly from a collection of N reference watermarks. In this system, embedding process can be written as:

$$\hat{F}_t = F_t + \alpha W_{\phi(t)} \qquad P(\phi(t) = i) = P_i \tag{1}$$

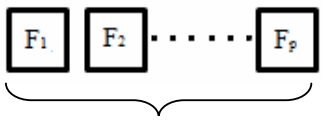
Where F_t is the luminance of t^{th} video frame, $\hat{F}_t$ the luminance of t^{th} watermarked frame and α is embedding strength. A correlation score is computed in detector side as follow:

$$C = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left| \hat{F}_t \cdot W_i \right| \tag{2}$$

T is the number of considered video frames. The correlation score is consequently compared to a threshold that is considered $\frac{\alpha}{2}$ in order to recognize presence or absence of watermark. From a security point of view, Doe¨rr and Duglay[6] showed that the SS- system effectively robust against both TFA and WER attacks but still expert attackers can adjust their approach according to this system. For example a brute force attack can be designed for defeating such system [7], however because of computational complexity; this attack is not suitable in practice. Also an attacker can remove the embedded watermark with an attack using vector quantization [6]. In the SS-N system, the security relies on the assumption that attackers can't build sets of frames carrying the same watermark. In this paper we propose a new algorithm for separating frames. Using this algorithm the attacker can build N sets of frames carrying the same watermark. So in next step, a simple watermark estimation remodulation (WER) attack can be used in each set for estimating the watermark pattern. At last, using these N sets, all N different watermark patterns can be estimated.

III. Proposed Algorithm

Using a new algorithm, all video frames are first divided into several sets so that frames of each set carry the same watermarks. Then in each group of frames, watermark is estimated. For this purpose, attacker chooses P neighboring frames. This P frames should be selected from static part of video. It means attacker should find static scene in video and takes P neighboring frames from that part of the video. Also P should be selected so large that these P frames contain all N existed watermark patterns. It should be considered that according to Kerckhoff's principle[8], watermarking system is publicly known and the attacker know that N different watermarks are randomly embedded in video.



P frames

Fig. 1 P successive frames

In each frame, one of the N watermark patterns is inserted randomly as follow:

$$\begin{cases} \hat{F}_1 = F_1 + \alpha W_{i1} \\ \hat{F}_2 = F_2 + \alpha W_{i2} \\ . \\ . \\ \hat{F}_K = F_K + \alpha W_{ik} \\ . \\ . \\ \hat{F}_P = F_p + \alpha W_{ip} \end{cases} \tag{3}$$

$$W_{ik} \in \{W_1, W_2, ..., W_N\} \qquad 1 \le k \le p$$

then, these P frames should be averaged as:

$$F_{av} = \frac{1}{P} \sum_{i=1}^{P} F_i + \alpha \frac{n_1}{P} W_1 + \alpha \frac{n_2}{P} W_2 + ... + \alpha \frac{n_N}{P} W_N \tag{4}$$

In the above expression, n_i shows number of frames that are inserted by the same watermark W_i. Then ΔF_i is defined as deference between $\hat{F}_i$ and F_{av}. For instance, if W_k and W_h were respectively embedded in k^{th} and h^{th} frame of these selected frames as randomly selected watermark signals, ΔF_k and ΔF_h are computed as fellow:

$$\Delta F_k = F_k - F_{av} = (F_k - \frac{1}{P} \sum_{i=1}^{P} F_i) - \alpha \frac{n_1}{P} W_1 - \alpha \frac{n_2}{P} W_2 - ..... + \alpha (1 - \frac{n_k}{P}) W_k - ... - \alpha \frac{n_N}{P} W_N$$

$$W_k \in \{W_1, W_2, ..., W_N\} \tag{5}$$

If selected frames are static enough which means:

$$F_k - \frac{1}{P} \sum_{i=1}^{P} F_i \approx 0. \tag{6}$$

Then:

$$\Delta F_k = -\alpha \frac{n_1}{P} W_1 - \alpha \frac{n_2}{P} W_2 - ..... + \alpha(1-\frac{n_k}{P}) W_k - ... - \alpha \frac{n_p}{P} W_N \quad (7)$$

Also for ΔF_h we have:

$$\Delta F_h = -\alpha \frac{n_1}{P} W_1 - \alpha \frac{n_2}{P} W_2 - ..... + \alpha(1-\frac{n_h}{P}) W_h - ... - \alpha \frac{n_p}{P} W_N \quad W_h \in \{W_1, W_2, ..., W_N\} \quad (8)$$

As it is clear, W_{ik} and W_{ih} were selected from a set of N watermarks included $W_1, W_2, ..., W_N$.

In this step, linear correlation between ΔF_k and ΔF_h is computed. So if the same watermark W_k is embedded in k^{th} and h^{th} frame of P selected frames, correlation between ΔF_k and ΔF_h is computed as follow:

$$C_{hk} = \Delta F_k . \Delta F_h = \alpha^2 \left(\frac{n_1}{P}\right)^2 + \alpha^2 \left(\frac{n_2}{P}\right)^2 + ... + \alpha^2 \left(1-\frac{n_k}{P}\right)^2 + ... + \alpha \left(\frac{n_k}{P}\right)^2 = \sum_{i=1}^{N} \alpha^2 \left(\frac{n_i}{P}\right)^2 + \alpha^2 (1-\frac{2n_k}{P}) = K_k \quad (9)$$

Also if two different watermarks W_k and W_h are embedded in k^{th} and h^{th} frames respectively, linear correlation between ΔF_k and ΔF_h is computed as follow:

$$C_{hk} = \Delta F_k . \Delta F_h = \alpha^2 \left(\frac{n_1}{P}\right)^2 + \alpha^2 \left(\frac{n_2}{P}\right)^2 + ... - \alpha^2 \left(1-\frac{n_{ik}}{P}\right)\left(\frac{n_{ih}}{P}\right) - \alpha^2 \left(1-\frac{n_{ih}}{P}\right)\left(\frac{n_{ik}}{P}\right) + ... + \alpha \left(\frac{n_k}{P}\right)^2 = \sum_{i=1}^{N} \alpha^2 \left(\frac{n_i}{P}\right)^2 - \alpha^2 \left(\frac{n_h}{P}\right) - \alpha^2 \left(\frac{n_k}{P}\right) = K_k - \alpha^2 (1-\frac{2n_k}{P}) - \alpha^2 \left(\frac{n_h}{P}\right) - \alpha^2 \left(\frac{n_k}{P}\right) \quad (10)$$

Now, we use the method that is mentioned in previous part for an algorithm to separate frames with different watermarks. In this algorithm the attacker should first compute $\Delta F_1 . \Delta F_1$, $\Delta F_1 . \Delta F_2$ ..... $\Delta F_1 . \Delta F_P$ as fellow:

If 1st and j^{th} frames have the same watermark pattern:

$$\Delta F_1 . \Delta F_j = K_1$$

If 1st and j^{th} frames have different watermark patterns:

$$\Delta F_1 . \Delta F_j = K_1 - \beta_{1j} \quad \beta_{1j} = \alpha^2 (1-\frac{2n_1}{P}) + \alpha^2 \left(\frac{n_1}{P}\right) + \alpha^2 \left(\frac{n_j}{P}\right) \quad (11)$$

In the above expression, n_1 is number of frames that carry the same watermark as the first frame while n_j is number of frames that contain the same watermark as j^{th} frame. It is clear that $\Delta F_1 . \Delta F_1 = K_1$, so as it has been shown, if $\Delta F_1 . \Delta F_j \approx \Delta F_1 . \Delta F_1 = K_1 \quad (2 \le j \le P)$, it can be concluded that 1^{st} and j^{th} frames have the same watermark patterns. Also if $\Delta F_1 . \Delta F_1 - \Delta F_1 . \Delta F_j = \beta_{1j}$, j^{th} and 1^{st} frames have different watermark patterns. So the frames that contain the same watermarks as first frame and also the number of these frames (n_1) are easily obtained. In the next step, the attacker put away these n_1 frames obtained in previous step and repeats proposed algorithm for $P - n_1$ remaining frames. It is enough to select first frame from these $P - n_1$ frames and find frames that have same watermark patterns as this frame. After localizing these frames and obtaining the number of them (n_2), these (n_2) frames should be put away again. If N is number of watermark patterns that are used, this algorithm should be done for $N-1$ times. After doing this algorithm for $N-1$ times, the attacker has N sets of frames which each set contains n_j frames $(1 \le j \le N)$ that carry same watermark

After obtaining n_1, n_2, ..., n_N, the attacker can estimate embedding strength as follow:

$$K_1 = \sum_{i=1}^{N} \alpha^2 \left(\frac{n_i}{P}\right)^2 + \alpha^2 (1-\frac{2n_1}{P})$$

So:

$$\alpha = \sqrt{\frac{K_1}{\sum_{i=1}^{N} \left(\frac{n_i}{P}\right)^2 + (1-\frac{2n_1}{P})}} \quad (12)$$

Now, the attacker can compute correlation between every frame of video and a frame from each group to find out which watermark pattern embedded in that frame. If correlation value between that frame and a frame from j^{th} set containing W_j as watermark is largest, that frame contains

W_j as watermark too. So the attacker can collect frames carrying same watermark pattern and use watermark estimation remodulation attack (WER) in every set to estimate watermark pattern and subtracted the estimated watermark from each video frame of that set.

IV. Experimental Result

The performance of the proposed algorithm for collusion attack was verified on a video sequences. A video sequence containing 150 frames is considered. Here, we use 'Suzie' video for test. SS-N watermarking system is used for embedding N watermarks randomly in each frame of video. We use N=3 different watermark patterns with global embedding strength $\alpha = 3$ for embedding watermarks. For start of attack, at first P successive frame should be selected from static part of video, P=12 successive frames are chosen from static scene of video. For this purpose, 12 frames from first part of video that is static enough are selected.

Fig. 2 A scene of "Suzie" video that is used

Now, ΔF is obtained for each frame and $\Delta F_1.\Delta F_j$, $(1 \le j \le 12)$ are calculated as it is shown in Table I:

TABLE I
Obtaining Value for 12 Selected Frames

$\Delta F_1.\Delta F_i$	i=1	i=2	i=3	i=4
	4.642	-- 3.044	4.598	-3.624
	i=5	i=6	i=7	i=8
	-3.702	-3.683	4.569	-3.121
	i=9	i=10	i=11	i=12
	-3.176	4.553	4.541	-3.136

So, as it has been seen in above table, for $i = 1,3,7,10,11$, $\Delta F_1.\Delta F_j$ s are almost equal and clearly larger than others. So, these frames contain the same watermark patterns, and $n_1 = 5$ is number of these frames. After putting away these frames, $\Delta F_1.\Delta F_j$ should be obtained for remaining frames.

TABLE II
Obtaining Value for 7 Remained Frames

$\Delta F_1.\Delta F_i$	i=2	i=4	i=5	i=6
	7.591	-4.429	-4.391	7.476
	i=8	i=9	i=12	
	-4.342	7.423	-4.179	

So, looking at the above table, it is clear that for $i = 2,6,9$, $\Delta F_2.\Delta F_j$ are almost equal and larger than others. It can be concluded that 2nd ,6th and 9th frame have same watermark pattern and $n_2 = 3$ is the number of frames that contain this watermark pattern. At last, $n_2 = 4$ remained frames contain the same frame. So, all these 12 frames are separated based on the watermark patterns embedding in them as follow:

TABLE III
N=3 Sets of Fame Based on Watermark Patterns

watermark	W1	W2	W3
frames	1,3,7,10,11	2,6,9	4,5,8,12
number	n1=5	n2=3	n3=4

Then attacker can estimate modulation strength α using expression (12) as follow:

$$K_1 = 4.642$$

$$\alpha = \sqrt{\frac{4.642}{\left(\frac{5}{12}\right)^2 + \left(\frac{3}{12}\right)^2 + \left(\frac{4}{12}\right)^2 + (1 - \frac{2 \times 5}{12})}}$$

$$= 3.003$$

Correlation between each frame of video and a frame from each set can be computed and if correlation between that frame and a frame from set i is larger than correlation between that frame and frame from other sets, that frame contains same watermark pattern as frames in set i. Therefore we can separate all video frames. Now a simple WER attack can be used for each set to estimate watermark in each set and subtract the estimated watermark from each frame in that set.

For evaluating the attack in our selected video sequence, the correlation score in detector side is obtained before and after the attack as follow:

TABLE IV
Correlation Score Before and After Attack in Detector Side

	Before attack	After attack
Correlation score	3.12	0.34

As it is clear the correlation score after the attack is smaller

than $\frac{\alpha}{2}$, so the detector can't detect the watermark and the attack is successful.

V. Conclusion

Inter-frame Collusion attack in frame-by-frame video watermarking is very critical; using the same or redundant watermark in every frame can lead to weakness against WER attack. Also uncorrelated watermarks for video frames can be removed by FTF attack. Switching watermark system or SS-N system is more secure against WER and FTF attacks. In this system, for each frame a watermark is randomly selected from N existed watermark patterns. The security of SS-N system relies on the assumption that attackers can't build sets of frames containing the same watermark pattern. . A new algorithm for separating frames and building sets of frame carrying the same watermark was proposed and it has been shown that the attacker can use this algorithm to build sets of frames carrying the same watermark. Then using WER in each set, one can estimate the watermark and subtract the watermark from each frame of that set.

.

Acknowledgment

This research is supported financially by ITRC (Iranian Telecommunication research center).

References

[1] Su, K., Kundur, D., and Hatzinakos, D.: 'Statistical invisibility for collusion-resistant digital video watermarking'. IEEE Trans. Multimedia, 2005, pp. 43–51

[2] G. Doërr and J.-L. Dugelay, "A guide tour of video watermarking," Signal Processing: Image Commun., vol. 18, no. 4, pp. 263–282, Apr.2003

[3] Doe¨ rr, G., and Dugelay, J.-L.: 'New intra-video collusion attack using mosaicing'. Proc. IEEE Int. Conf. Multimedia and Expo, 2003, vol. II, pp. 505–508

[4] T. Kalker, G. Depovere, J. Haitsma, and M. Maes. A video watermarking system for broadcast monitoring. In Security and Watermarking of Multimedia Contents, volume 3657 of Proceedings of SPIE, pages 103–112, January 1999.

[5] Voloshynovskiy, S., Pereira, S., Herrigel, A., Baumgartner, N., and Pun, T.: 'Generalized watermarking attack based on watermark estimation and perceptual remodulation', Proc. SPIE, 2000, 3971, Security and Watermarking of Multimedia Content II, pp. 358–370

[6] G. Do¨err and J.-L. Dugelay. Security pitfalls of frame-by-frame approaches to video watermarking. IEEE Transactions on Signal Processing, Supplement on Secure Media, 52(10):2955–2964, October 2004.

[7] G. Doërr and J.-L. Dugelay. (2003) Switching between orthogonal watermarks for enhanced security against collusion in video. Eurécom Inst.,Sophia-Antipolis,France.[Online].Available: http://www.eurecom.fr/~doerr

[8] A. Kerckhoffs. La cryptographie militaire. Journal des sciences militaries, IX:5–83, January 1883.

.