Journal of Theoretical and Applied Information Technology
30th April 2018. Vol.96. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT DIGITAL VIDEO WATERMARKING SYSTEM ROBUST AGAINST VARIOUS SPATIAL, TEMPORAL, and SPATIO-TEMPORAL ATTACKS

MANISH K THAKUR
Assistant Professor (Sr. Grade), Department of CSE/IT, Jaypee Institute of Information Technology, India
E-mail: mthakur.jiit@gmail.com

ABSTRACT

For many years information hiding techniques are playing major roles in achieving robustness against various malicious attacks on digital multimedia data of different application areas. The induction of digital watermarking has given these practices a new way in the field of information hiding to prevent the copyrighted multimedia data against various attacks. In this paper, we propose a generalised watermarking system which efficiently identifies various attacks in spatial, temporal, and spatio-temporal domain, *viz.* frame deletion, frame swapping, and frame copying, *etc*. Based on the timing information of each frame of a video, in the proposed scheme we first generate the unique watermarks (representing the timing information) in real time and then embed the auto generated watermark into 8 x 8 discrete cosine transform blocks of corresponding frames of a video. While extracting the watermarks from each frame of the watermarked video, any alterations in the sequence of the retrieved watermarks are the indication of the temporal attacks (frame deletion, *etc*.) in the copyrighted videos. Experiments have been conducted on self captured videos to analyze the performance of the proposed watermarking model. Performance of the proposed model has been measured in terms of the recall rate, *i.e.* ability of the proposed watermarking model to correctly detect the attacks in videos, quality of the watermarked videos and extracted watermarks using the quality metrics, Peak Signal to Noise Ratio (PSNR). Experimental results show the 100% recall rate in detection of the temporal attacks, if videos are manipulated in temporal domain only. Achieved recall rate is in between the range 86% and 91%, if videos are manipulated in spatio-temporal domain and it is 93%, if videos are spatially manipulated.

Keywords: *Watermarking, Spatio-temporal attacks, Frame deletion, Frame swapping, Frame Copying*

1. INTRODUCTION

The dawn of digital epoch has touched the lives of masses imparting an easier and a quicker way of creating, processing and distribution of digital media. Popular internet software based on peer to peer architecture *viz.* Kazaa [1], BitTorren [2], eDonkey network [3], and Gnutella [4], *etc.* allows large scale dissemination of copyrighted digital data such as movies, music, software, *etc*. This massive sharing not only degrades the economic value of the artifact but also incur huge losses to the content proprietors [5]. Fretful by the upshots of illegal replication and distribution on such an enormous level, there is a need to protect the copyright of digital content [6].

Marking the initial step, various methods such as cryptographic and steganography based techniques had been proposed for the security of digital data. While cryptography contemplates on encrypting data which is only meant for the intended user [7], steganography, on the other hand conceals the message in such a fashion that it is unperceivable [8][9]. Unfortunately, both the endeavors lack in providing a robust scheme against various nasty attacks. Thus it necessitates the requirement of additional efforts to propose more efficient schemes [10] to protect the copyrighted digital data against various nasty attacks.

As detailed in [11][12], these attacks can be made in spatial domain *i.e.* spatial attacks, attacks in temporal domain, *i.e.* temporal attacks, and attacks in spatio-temporal domain *i.e.* spatio-temporal attacks.

Videos can be spatially attacked by an attacker, where he/she can manipulate the picture elements or pixel bits of a video frame.

Further, an attacker can also manipulate videos by disturbing the frame sequence (*i.e.* attack in temporal domain) through removal of frames; frames sequence reordering; and frames addition,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

etc. Some of the common attacks in temporal domain are, frame deletion, frame swapping, and frame copying.

As shown in Figure 1 (b), an attacker can delete or remove some of the video frames of the original video, resulting into tampered video with reduced frame count. Although, the count of frames gets reduced due to the attack of frame deletion, the order or sequence of frames remains unchanged. Unlike frame deletion, frame count remains unchanged during the attack of frame swapping, but the sequence or order of frames in the original video gets disordered in the tampered video. Attack of frame swapping is illustrated in Figure 1 (c). As depicted in Figure 1 (d) frame count will increase, if the video is manipulated by attacker by pasting the copied frames of the video to some other location in that video.

Lastly, an attacker can manipulate videos in both spatial and temporal domains simultaneously (*i.e.* attack in spatio-temporal domain) by manipulating pixel bits of a video frame as well as disturb the frame sequence or order.

Now a day, digital watermarking is emerging as one of the efficient solutions to proficiently identify the copyright violators and prosecute them [13]. In a digital watermarking system, we hide information or watermarks in the robust and invisible manner which helps to thwart the direct illicit copying of the information. This imperceptible information or watermark containing copyright note is entrenched into the digital media which ensures its security while transferring it over the network [14].

Today, digital watermarking which was first studied comprehensively for tranquil images has now been extended to other types of digital multimedia data, *viz.*, audio and video [15]. Major focus has now shifted to video content. While examining the videos (which was watermarked by embedding some watermark into the frames of the video), someone who embedded the watermark, may or may not be in the position to retrieve back the embedded watermark.

Non retrieval or partial retrieval of the embedded watermark may be the indication of the attacks made by attackers to misuse the copyrighted digital multimedia data, *i.e.* videos.

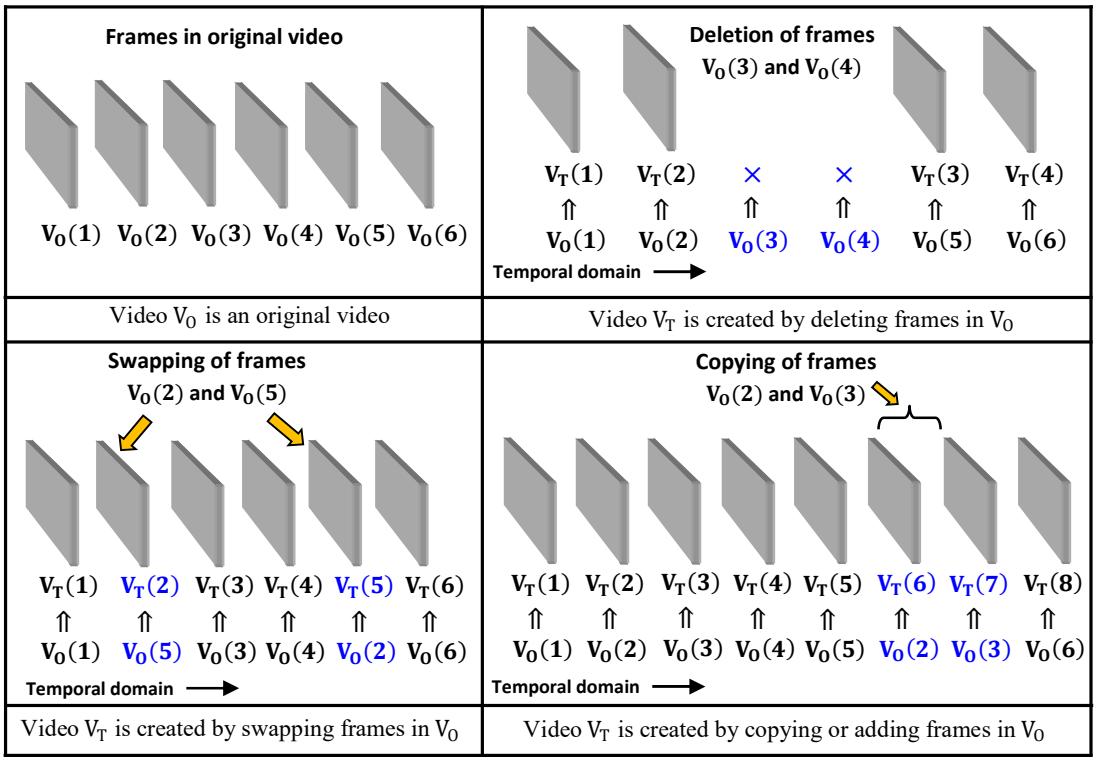

Figure 1: Examples Of Common Temporal Attacks, Original Video (V_O) Is Shown In (A), Whereas, (B), (C), And (D) Present The Frames In Manipulated Videos (V_T) After Attacks Of Frame Deletion, Frame Swapping, And Frame Copying Respectively.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Existence of non-hostile video processing, collusion attacks, attacks in temporal domain (*viz.* frame deletion, frame swapping, *etc.*), and requirement of real-time watermarking are some of the challenges that makes video watermarking distinguishable from image watermarking [16].

Watermarking can be executed either on uncompressed video (ITU-R 601) or on compressed video (*viz.* MPEG) wherein, extra redundant data is already removed and better computational results are obtained [17].

Techniques of watermarking can be broadly categorized in two domains *viz.* spatial domain and the transform domain. In precedent few years, countless authors have put forward their algorithms in the scientific literature; robust against a number of geometric attacks such as compression, filtering, geometric distortions and bit-rate reduction.

Foremost contributions in the field of watermarking started in mid 40's with the development of diverse spread spectrum techniques for military rationale [18]. Brandao et al performed the spread spectrum in uncompressed domain which not only proffers a good suppleness but was apposite for watermarking due to resemblance between watermarking and spread spectrum techniques [19]. Kim et al proposed an analogous technique for uncompressed domain watermark embedding and extraction, which was one of the practical real time approaches developed to be robust against the downscaling attack [20]. The basic pattern was scaled and embedded in an additive way while at the receiver end it was extracted blindly without the need of the original watermark pattern using secret keys.

Although such type of spatial domain watermarking techniques are predisposed to ease the hardware requirements in terms of the execution time but these techniques lack good visual models. According to Chen et al, in terms of the watermark capacity, spatial domain is the nastiest place to interleave a high capacity watermark [21]. This drawback led to the proposal of assorted embedding schemes in frequency domain.

As per Elbasi et al, the frequency domain techniques not only offer higher competence and enhanced robustness against the attacks but are also close to HVS models [22]. The watermark was adaptively applied to frequencies having vital essentials of the video or selectively applied to the middle or lower order frequency bands. This type of compressed domain watermarking, as discerned by Mobasseri et al, is more efficient than the watermarking in uncompressed domain because it doesn't necessitate copious decoded video streams and videos are amassed in the compressed formats [23]. Carrying out watermarking in the Discrete Cosine Transform (DCT) domain is the most conventional one, because all the major compression techniques *viz.* MPEG were developed in the DCT domain. Wang et al elucidated different sets to achieve the blind video watermarking for MPEG-2 videos [24]. They focused on typical geometric processing attacks such as bit-rate reduction, frame cropping, downscaling, frame dropping, *etc*. Their watermarking scheme was accomplished in the DCT domain with the limitations as follows: their algorithm cannot directly be applied in different compressed video formats *viz.* MPEG-4 or H.264 and was vulnerable to transcoding [25].

Referring to the context of copyright protection, Doerr et al cataloged several requirements for video watermarking system [18] as follows:

Perceptual Transparency- embedded watermarks should be perceptually least traceable or untraceable.

Robustness to Attacks- embedded watermarks should be robust enough and difficult to get it removed either by any premeditated operations or by inadvertent operations. If removed with malicious intention by the attacker, then it should degrade the perceived quality of the video to an extent that its commercial value is no more.

Security- embedding procedure must be secure such that any illicit intruder should not be able to spot and confiscate the watermark. Secret keys are often used to achieve this requirement where, watermark embedding and extraction process is guided by secret keys.

For Broadcast applications- mostly compressed videos are used for broadcasting and other applications, therefore, embedding process should be applicable for compressed videos.

Copyright Protection and Ownership Deadlock- watermarking scheme should be capable enough to handle the ownership issues. These issues (or deadlock) often encounter when attacker embeds his/her own watermark into the watermarked videos of someone else. Later, attacker can claim the ownership of the watermarked videos which were re-watermarked by the attacker.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Considering different requirements of the digital video watermarking system, many schemes had been proposed by researchers in last decade, specifically handling the temporal attacks [26][27]. Temporal attacks *viz.* frame drop, frame swapping, *etc*. had been often handled by embedding timing information of the frames as watermark. In [26], key based sequence was used as synchronization code and embedded in front of the watermark sequence, whereas, the scheme proposed in [27] embeds the timing information of the frames into last nonzero quantized DCT value of the blocks of each frame. Additional bits (representing the timing information) were embedded in these schemes along with the watermark bits. Instead, watermark images representing the timing information could have been embedded into frames to avoid the usages of additional bits for same. As detailed in our proposed methodology, the watermark images representing the timing information can be auto generated in real time. Also, these watermarks representing the timing information can be efficiently utilized to localize the temporal attacks (*i.e.* identification of the location of the temporal attacks, if any).

Considering different requirements of the digital video watermarking system discussed above, in this paper, we propose an efficient video watermarking scheme for compressed videos. The proposed video watermarking model is capable to handle common image based attacks (*i.e.* attacks in spatial domain), common attacks in temporal domain, and common attacks in spatio-temporal domain. In the proposed video watermarking model, we separately handled the common spatial attacks, *viz.* Gaussian noise and alteration of brightness, contrast, and luminance by embedding the watermark in middle band of the discrete cosine transform (DCT) of the decompressed video frames.

Further, the embedded watermarks in each frames of the compressed video are auto generated sequence of frame's timing information which is helpful to find the temporal attacks, if any. Jointly (embedding of timing information into the DCT blocks of each frame), the proposed scheme identifies the spatio-temporal attacks, if any, made by the attackers.

Rest of the paper is organized as follows: Section 2 presents the proposed model for auto generation of the watermarks (represented by the timing information of the frames) and the proposed watermark embedding / extraction models; in Section 3, we present the experiments conducted to analyze the performance of the models presented in Section 2 along with comparative analysis with some contemporary works; the paper is concluded in Section 4.

2. PROPOSED MODELS

In this section, we first present the auto generation model to construct the watermarks (representing the timing information) which is to be embedded into the video frames. This section also presents the generalised model to embed the constructed watermark in each frame of the video along with the watermark extraction model, where retrieval of the embedded watermark from each frame is done.

2.1 Model For Auto Generation of Watermarks

Different watermarks (in form of the timing information) for each frame is generated automatically as a sequence of decimal integers (in form of images) formed with ten decimal digits (images of 0 to 9). As shown in Figure 2, we created ten gray scale image files of fixed dimensions (say, $X \times Y$) representing the ten decimal digits, *i.e.* 0 to 9.

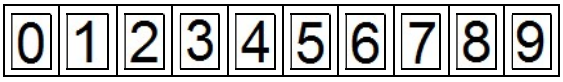

Figure 2: Images Of The Digits 0 To 9, And Are Used To Auto Create The Unique Watermarks Representing The Timing Information

Starting with the first frame of a video, we retrieved the consecutive frames of a video one by one. Before embedding the watermarks into the retrieved i^{th} frame, depending on the timing information of the i^{th} frame, we auto generated the watermark as follows (Figure 3a, Figure 3b, and Figure 3c depict some of the generated watermarks representing the timing information):

(a) If the i^{th} frame belongs to any one frame out of first ten frames of a video, then, based on the timing information (frame numbers between 0 and 9) of that frame, we picked the respective image file and created the watermark of dimension or size $X \times Y$ (depicted in Figure 3a);

(b) Watermarks for next ninety frames (*i.e.* frame numbers between 10 and 99) of the video are created by combining two image files (representing the timing information as 10 to 99) of dimension or size $X \times 2Y$ (depicted in Figure 3b);

(c) Watermarks for next nine hundred frames (*i.e.* frame numbers between 100 and 999)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of the video are created by combining three image files (representing the timing information as 100 to 999) of dimension or size X × 3Y (depicted in Figure 3c) and so on.

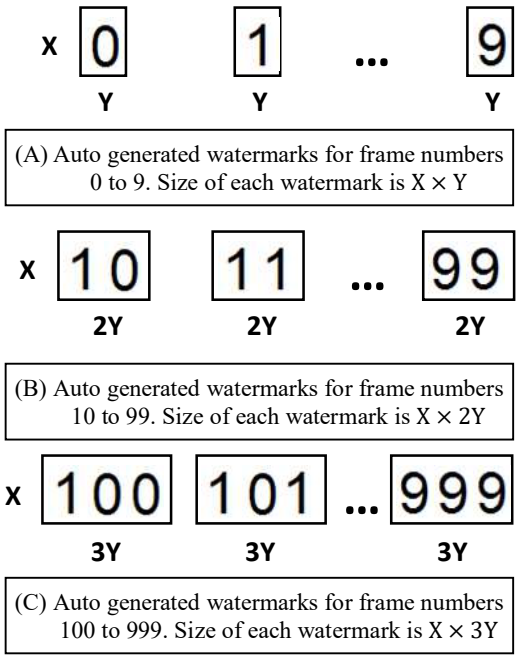

Figure 3: Some Of The Auto Generated Watermark Images Representing The Timing Information Of The First Ten Frames, Next Ninety Frames, And Next Nine Hundred Frames

2.2 Model For Watermark Embedding

On the basis of assorted researches done by various authors, the following section presents a comprehensive video watermarking model which is a hybrid of various approaches. In one of the endeavors, Kim et al proposed a blueprint to implant the watermark in uncompressed domain which employed blind detector. The model was applicable to DCT-based coding videos [20]. Wang et al intended the video watermarking scheme for compressed videos (MPEG-2). The scheme was claimed to be resistant against cropping attack [24].

Mohanty et al in his sculpt considered diverse influencing traits of HVS to aptly embed the watermark in the most perceptually significant region of sub image [28]. Chae et al used the multidimensional lattices as their key component and the host frame videos and images are transformed using the 8 × 8 block DCT [15]. In an object based watermarking plan by Swanson et al, the watermark was figured frame by frame [29]. The model deployed the author representation, visual masking and motion compensation to structure the watermarking system.

Based on the above proposed research methodologies, a generalized hybrid model for video watermarking is proposed beneath and shown in Figure 4. Referring to Figure 4, the compressed video on the incoming bit stream is syntactically parsed and data allied to headers, motion and texture are estranged out in discrete buffers. The header and motion data are set aside and simply added to the output bit stream without any alteration at the time of video reconstruction. The textured data is in the form of the static images. These images are first divided into the integral number of 8 × 8 blocks. These blocks are inversely quantized and then DCT of the ensuing matrix is computed. Each 8 × 8 DCT block consists of three types of frequency components: low band, middle band, and high band. Unlike middle frequency components, high band frequency components have highest possibility of removal through compression and noise attacks, whereas low band frequency components are perceptually most significant. Therefore, middle band components of those 8 × 8 DCT blocks which are found significant is selected as the region to embed the watermark.

After furnishing the appropriate private key in conjunction with the timing information (created using the process detailed in Section 2.1) of the frame under consideration, the basic pattern generator spawns watermarking pattern (in form of bits) which after tiling and scaling is embedded into the selected 8 × 8 DCT block X_i. The watermarked 8 × 8 blocks, Y_i thus obtained are coalesced by application of DCT and then quantized to form image frames. These frames on uniting together with the header and motion information will give the video in compressed format which is conceded on the bit stream.

2.3 Model For Watermark Extraction

In this section we present the proposed watermark extraction model. Referring to Figure 5, the preliminary process up to the frame extraction from the bit stream is akin to watermark embedding. In the watermark extractor module, blind watermark detector is used to carry out normalized correlation.

For the detector to perform successfully there is no requisite of the original video at the time of extraction. The basic random pattern is engendered using the same private key as during watermark embedding.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

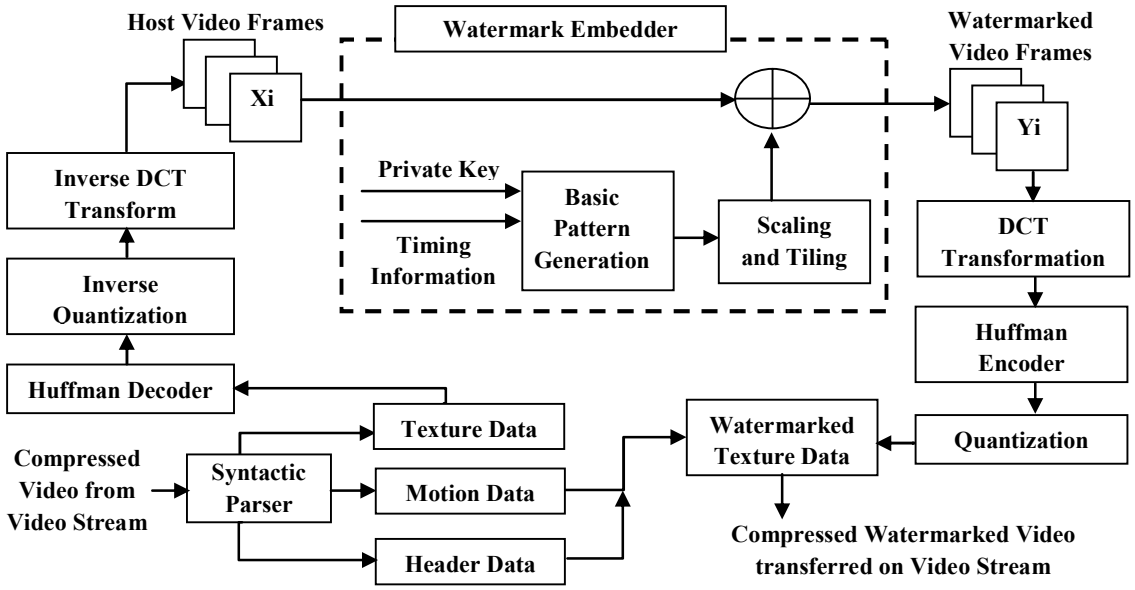

Figure 4: Model For Watermark Embedding, Where Input Is A Compressed Video. Timing Information For Each Frame Of The Input Video Is Embedded As Watermark Along With A Primary Key To Generate The Compressed Watermarked Video

Considering the embedded watermark as noise, each frame Y_i is remitted to a denoising filter to obtain the estimated watermark patterns by subtracting the filtered video frame from Y_i. The estimated watermark is accumulated from the corresponding Y_i. After accruing pattern bits, the watermark (representing the timing information) is reconstructed by fold and sum of the extracted bits. If normalized cross correlation value C surpasses a preset threshold, the veiled messages are accurately recovered. If this correlation value C is found to be less than the preset threshold, then the video is said to have been attacked spatially.

Further, the extracted watermark of each frame is auto checked for temporal attacks, if any. Size of the extracted watermarks from a frame may be $X \times Y$ or $X \times 2Y$ or $X \times 3Y$ and so on. We stored the histogram of the extracted gray scaled watermark of the frame, F_i into a vector (EW) at EW_i. Further, we regenerated the actual gray scaled watermarks representing the consecutive timing information, computed their histograms and stored the histogram of the gray scaled watermark, W_i into a vector (AW) at AW_i. Both vectors, EW and AW were auto compared for identification of the temporal attacks (if any) as follows:

(a) if histogram value at AW_i does not have a corresponding matching value (upto a threshold) in EW, then we considered it as the attack of frame deletion, where i^{th} frame is deleted by the attacker. Identification of the i^{th} frame as the deleted or dropped frame describes the localization (*i.e.* actual location of attack) of the attack of frame drop.

(b) if histogram value at AW_i is having only one corresponding matching value (upto a threshold) in EW, say at EW_j where, $i \neq j$, then we compared the extracted watermark image and actual watermark image for the similarity using structural similarity index, SSIM [30]. If, both images are same (*i.e.* SSIM is around 1), then we considered it as the attack of frame swapping. Identification of the i^{th} frame and j^{th} frame as the swapped frames describes the localization (*i.e.* actual location of attack) of the attack of frame swapping.

(c) if histogram value at AW_i is having many corresponding matching values (up-to a threshold and also cross checked with SSIM) in EW, then we considered it as the attack of frame copying, where i^{th} frame is copied and pasted at different locations in the watermarked video by the attacker. Identification of the i^{th} frame as the copied frame describes the tampering localization.

3. EXPERIMENTAL ANALYSIS

In this section, we present the details of the conducted experiments, analysis of the performance of the proposed watermarking scheme, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

comparative analysis with some of the contemporary works.

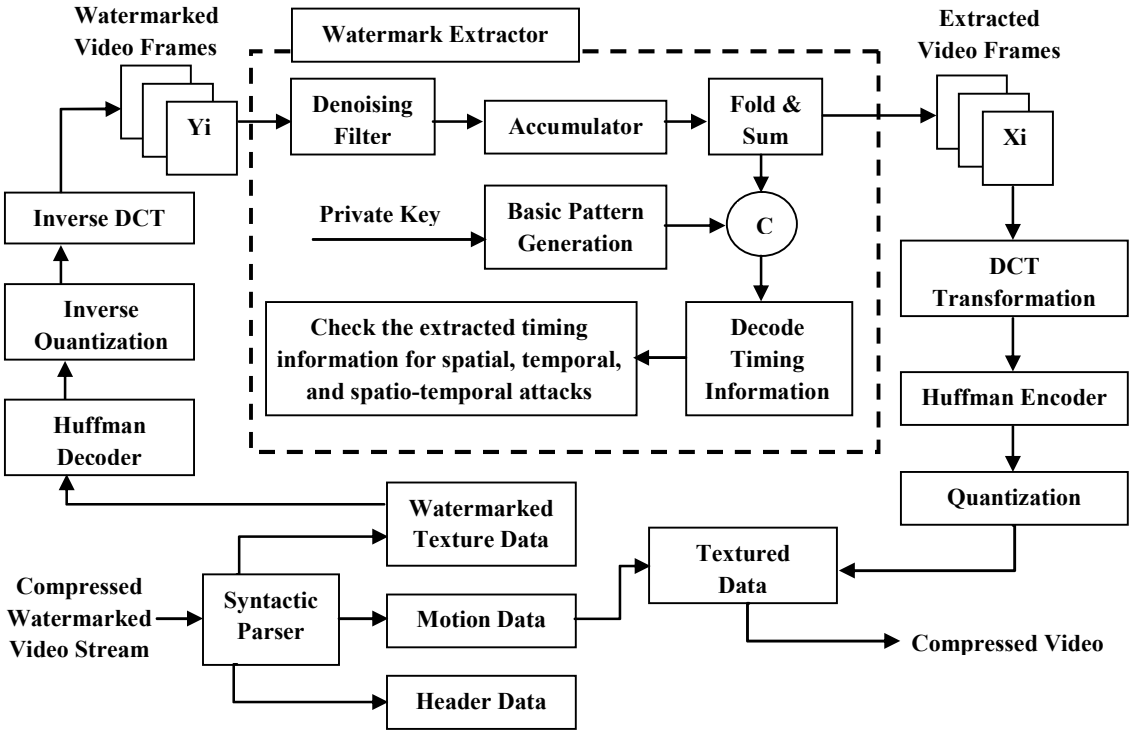

Figure 5: Model For Watermark Extraction, Where Input Is The Compressed Watermarked Video. Timing Information For Each Frame Of The Watermarked Video Is Extracted Using Primary Key

3.1 Experiment Details

All experiments have been conducted on the self captured five videos of different duration. Count of frames in the five videos are as follows: V_1 (237 frames), V_2 (329 frames), V_3 (783 frames), V_4 (852 frames), and V_5 (944 frames).

Using the approaches presented in Section 2.1 and Section 2.2, we embedded the watermarks (auto generated timing information of the frames in a video) into each frame of the five videos, *viz.* V_1, V_2, *etc*. After embedding the watermarks, we created hundred copies of each watermarked videos as test videos, *i.e.* experimental results presented in this section are based on five hundred test videos.

Out of five hundred test videos, we spatially manipulated one hundred videos by introducing Gaussian Noise (GN) and alteration of brightness (B), contrast (C), and luminance (L), *i.e.* BCL. Another set of one hundred test videos were temporally manipulated either by deleting some frames (FD), or swapping some frames (FS), or copying some frames (FC) at different locations. We also manipulated another set of two hundred and forty videos both spatially (GN and BCL) and temporally (either by FD or by FS, or by FC) and created two hundred and forty spatio-temporally manipulated videos. Remaining sixty videos were kept un-attacked, *i.e.* the test dataset of five hundred test videos comprised of the videos which were manipulated (spatially, temporally, *etc.*) by the attackers as well as the videos which were not attacked by the attacker.

3.2 Performance Analysis

Further, performance of the proposed watermarking scheme was evaluated on the Recall parameter, visual quality degradation of the watermarked videos, and degradation of the quality of the extracted watermarks.

Recall parameter is defined as follows: we computed it as the ratio between count of correctly extracted watermarks from the watermarked videos (which has been attacked) and count of watermarks embedded into the watermarked videos. As each frame of a test video is watermarked by its timing information, obviously, count of embedded watermarks must be same as the count of frames in that video.

Peak signal to noise ratio (PSNR) is one of the most used metrics to measure the degradation of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the quality between two multimedia data [30]. In context of image data, PSNR (in dB) is computed by taking the logarithmic ratio of maximum amplitude of picture elements and mean squared error (which is mean of the pixel by pixel squared difference) and formulated in Equation 1 and Equation 2, whereas, in context of videos, it is average of the frame by frame PSNRs of the frames of two videos.

$$PSNR\ (in\ dB) = 10 \times log_{10} \frac{(n^2-1)^2}{MSE} \quad (1)$$

$$MSE(S,D) = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(S_{i,j} - D_{i,j}\right)^2 \quad (2)$$

where, n is the number of bits per pixel element; S and D are two image files between which PSNR is to be computed; M and N are the height and width of the images, S and D; and $S_{i,j}$ and $D_{i,j}$ are the pixel locations in the image files.

Table 1 presents the Recall rates achieved with different types of attacks made on the five hundred test videos. As seen in Table 1, recall rate is 100%, if, videos are only temporally attacked, *i.e.* frame deletion, frame swapping, and frame copying. As, there are no spatial attack, all the watermarks embedded into different frames have been successfully extracted from the temporally attacked videos, which later compared with the standard value of the histogram of the timing information. Hence, 100% recall rate is achieved. Further, as seen in Table 1, recall rate is around, 93%, if videos are spatially attacked.

Experimental statistics related with the extraction of watermarks, if videos are spatio-temporally attacked are also shown in Table 1. Recall rate is least (around 86%) with the combination of spatial (GN/BCL) attack and frame deletion attack. Obvious reason behind this is the loss of intermediate watermarked frames of videos, restricting the extraction mechanism and auto matching mechanism to perform accurately. Achieved recall rates are around 89% and 91%, if videos are attacked by frame swapping in combination of spatial (GN/BCL) attack and frame copying in combination of spatial (GN/BCL) attack respectively.

After embedding the timing information as watermark in each frame of test videos, test for the degradation in the perceptual quality was performed. Quality measurement using PSNR was performed over the frames of watermarked videos (five videos) and respective frames of the un-watermarked video. Average of the frame by frame PSNR (in dB) between the frames of watermarked videos and respective un-watermarked videos had been measured in the range, 48 dB and 52 dB.

Further, degradation in the quality of extracted watermarks with respect to corresponding watermarks had also been measured. Least degradation in the perceived quality was measured when videos were temporally manipulated. Obvious reason is the non-manipulation of the pixel bits of the frames of temporally manipulated videos. It had been measured in the range between 48 dB and 52 dB. Maximum degradation in the perceived quality was measured when videos were spatio-temporally manipulated. It had been measured in the range between 35 dB and 43 dB. Also, it had been measured in the range between 38 dB and 44 dB, when videos were spatially manipulated.

Table 1: Performances in terms of achieved recall rate during extraction of the embedded watermarks from all the five hundred test videos. FD, FS, and FC are the temporal attacks of frame deletion, frame swapping, and frame copying respectively, whereas GN and BCL are the spatial attack of Gaussian Noise and alteration of brightness (B), contrast (C), and luminance (L).

Type of Attack	Count of attacked frames	Count of successfully extracted watermarks	Recall rate
Temporal: FD, FS, FC	1000	1000	100%
Spatial: GN, BCL	600	556	93%
Spatio-temporal: FD and GN/BCL	1063	914	86%
Spatio-temporal: FS and GN/BCL	1032	919	89%
Spatio-temporal: FC and GN/BCL	958	872	91%

3.3 Comparative Analysis

This section presents the comparative analysis between our work and some of the contemporary works. It is unfair to compare our work with some other contemporary works as the watermarking schemes are usually proposed with considered constraints or deliverables, *viz.* kind of attacks

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

simultaneously handled, perceived quality of the source data as well as extracted watermarks, *etc*. In terms of the considered deliverables, we subsequently present the comparative analysis of our scheme with some of the contemporary works.

Some of the contemporary works reported in [26][27] dealt with proposal of such watermarking scheme which is robust against various temporal attacks, *viz.* frame drop and frame swapping, *etc.*

Authors in [26] proposed temporal static based video watermarking scheme where watermark was embedded by modifying the histogram shape of average dc energy sequence and claimed that the scheme is robust against frame drop and frame swapping. Robustness of the frame drop was tested by dropping alternate frames of a video, whereas performance of the frame swapping was analyzed by changing the position of each frame. While making comparison between ours scheme proposed in this paper and the scheme proposed in [26], our scheme is having edge of embedding timing information in the video frames. This embedding facilitates us to decide the authenticity of the videos against temporal attacks as well as localization of the temporal attacks, *i.e.* location of attack (*viz.* identification of the i^{th} frame which is deleted or copied or swapped by the attacker).

In another work [27], authors proposed their scheme, where timing information (macro-blocks' and frames' indices) was embedded into the last nonzero quantized DCT value of the blocks of each frame. Authors in [27] claimed that their proposed scheme is robust against various attacks including frame drop. Comparing to the scheme proposed in [27] our scheme (proposed in this paper) is having edge of the auto generation of different watermarks (in form of timing information) to be embedded into frames of a video. Besides this, our scheme also facilitates for the auto matching of the extracted watermarks with the actual watermarks representing the timing information. These features reduce the efforts required in deciding the authenticity of the videos against temporal attacks as well as localization of the temporal attacks.

4. CONCLUSION

In this paper we proposed a digital video watermarking system. We embedded the timing information as watermark (unique watermarks) into the frames of the compressed videos. Due to large number of unique watermarks, we proposed auto generation model to generate watermarks and embedded it into the DCT blocks of a frame in a video. We also proposed watermark extraction model to extract the embedded watermarks and auto comparison to validate the timing information.

Using five watermarked videos (created by embedding timing information as watermark in each frame of respective un-watermarked video), we created a set of five hundred test videos by introducing spatial, temporal, and spatio-temporal attacks. Using the quality metrics, PSNR, quality degradation in the watermarked videos was measured in the range between 48 dB and 52 dB. Further, 100% recall rate was observed in extracting the watermarks from temporally manipulated videos, whereas it was respectively between the range 86% and 93%, if videos were spatially and spatio-temporally manipulated. Finally, degradation of quality between the extracted watermarks (from the videos whose frames were manipulated through different attacks) and actual embedded watermarks (representing the timing information) had been measured in the range between 35 dB and 52 dB.

Auto generation of the unique watermark representing the timing information and auto matching of the extracted watermarks makes the scheme efficient (achieved accuracy/recall rate is 100%) to find various temporal attacks (frame deletion, frame swapping, and frame copying) and reduces the applied human efforts.

Due to wide applicability of watermarking based systems, *viz.* authentication, copyright, *etc.* the domain of watermarking is always have the scope to enhance the schemes by incorporating many more deliverables as listed by Doerr et al in [18]. Possible extensions of the work reported in this paper is to incorporate more spatial attacks *viz.* geometrical attacks and temporal attacks *viz.* frame averaging, *etc.* along with localization of the attacks.

REFRENCES

[1] https://www.en.wikipedia.org/wiki/Kazaa, Last accessed on March, 26, 2018

[2] https://www.en.wikipedia.org/wiki/BitTorrent, Accessed on March, 26, 2018

[3] https://en.wikipedia.org/wiki/EDonkey_network, Accessed on March, 26, 2018

[4] https://en.wikipedia.org/wiki/Gnutella, Accessed on March, 26, 2018

[5] P. P. W. Chan, M. R. Lyu, and R. T. Chin, "Copyright Protection on the Web: A Hybrid Digital Video Watermarking Scheme," *Proceedings of 13th World Wide Web Conference*, May 17-22, 2004, pp. 354-355

[6] M. Seadle, M. J. R. Deller, and A. Gurijala, "Why Watermark? The Copyright Need for an Engineering Solution," *Proceedings of 2nd ACM/IEEE Joint Conference on Digital Libraries*, July 13-17, 2002, pp. 324-325

[7] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of Video Transmission over Lossy Channels," IEEE journal on selected areas in communications, Vol. 18, No. 6, June 2000, pp. 1012-1032

[8] A. K. Khan and H. Jamal, "An Analytical Framework for Comparative Analysis of Various Watermarking and Steganographic Techniques," *Proceedings of 32nd Annual Conference on Industrial Electronics (IECON 2006)*, November, 6-10, 2006, pp. 3324-3327

[9] G. Kasana, K. Singh, and S. S. Bhatia, "Singular Value Decomposition based Steganography Technique for JPEG2000 Compressed Images," International Journal of Engineering – Transactions C: Aspects, Vol. 28, No. 12, December 2015, pp. 1720-1727

[10] D. Profrock, M. Schlauweg, and E. Muller, "Geometric Warping Watermarking Extended Concerning Geometric Attacks and Embedding Artifacts," *Proceedings of 9th Workshop on Multimedia and Security (MM&Sec'07),* September 20–21, 2007, pp. 169-174

[11] M. K. Thakur, V. Saxena, and J. P. Gupta, "Video Authentication against Set of Temporal Tampering," International Journal of Security and Its Application, Vol. 11, Issue 1, 2017, pp. 149-164

[12] M. K. Thakur, V. Saxena, and J. P. Gupta, "A Full Reference Algorithm for Dropped Frames Identification in Uncompressed Video Using Genetic Algorithm," International Journal of Digital Content Technology and its Applications, Vol. 6, Issue 20, 2012, pp. 562-573

[13] H. T. Sencar and N. Memon, "Watermarking and Ownership Problem: A Revisit," *Proceedings of 5th ACM Workshop on Digital Rights Management (DRM'05)*, November 7, 2005, pp. 93-101

[14] X. Zhu and B. Girod, "Video Streaming Over Wireless Networks," *Proceedings of European Signal Processing Conference (EUSIPCO07)*, September 2007, pp. 1462-1466

[15] J. J. Chae and B. S. Manjunath, "Data Hiding in Video," *Proceedings of International Conference on Image Processing*, October, 24-28, 1999, pp. 311-315

[16] F. Bartolini, V. Cappellini, R. Caldelli, A. D. Rosa, A. Piva, and M. Barni, "MPEG-4 Video Data Protection for Internet Distribution," *Proceedings of Thyrrhenian International Workshop on Digital Communications: Evolutionary Trends of the Internet (IWDC 2001)*, September, 17-20, 2001, pp. 713–720

[17] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information Hiding- A Survey," *Proceedings of the IEEE*, Vol. 87, Issue 7, July 1999, pp. 1062-1078

[18] G. Doerr and J. L. Dugelay, "A guide tour of video watermarking," Signal Processing: Image Communication 18 (2003), pp 263–282

[19] T. Brandao, M.P. Queluz, and A. Rodrigues, "Diversity Enhancement of Coded Spread Spectrum Video Watermarking," Wireless Personal Communications 23: 2002, pp. 93–104

[20] K. S. Kim, D. H. Im, Y. H. Suh, and H. K. Lee, "A Practical Real-Time Video Watermarking Scheme Robust against Downscaling Attack," *Proceedings of International Workshop on Digital Watermarking* (IWDW 2007), Vol. 5041, 2007, pp. 323-334

[21] T. Chen, S. Liu, H. Yao, and W. Gao, "Spatial Video Watermarking Based on Stability of DC Coefficients," *Proceedings of ICMLC 2005*, LNAI 3930, pp. 1033 – 1042, 2006

[22] E. Elbasi and A. M. Eskicioglu, "Robust Video Watermarking Scheme in Transform Domains," *Proceedings of ISC 2007*, 13-14 December 2007, pp 225-229

[23] B. G. Mobasseri and M. P. Marcinak, "Watermarking of MPEG-2 Video in Compressed Domain Using VLC Mapping," *Proceedings of MM-SEC'05*, August 1-2, 2005, pp 91-94

[24] Y. Wang and A. Pearmain, "Blind MPEG-2 Video Watermarking Robust Against Geometric Attacks: A Set of Approaches in DCT Domain," IEEE Transactions on Image Processing, Vol. 15, No. 6, June 2006, pp. 1536- 1543

[25] M. Flierl and B. Girod, "Generalized B Pictures and the Draft H.264/AVC Video Compression Standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, July 2003, pp. 587-597

[26] C. Chong, J. Ni, and J. Huang, "Temporal Statistic Based Video Watermarking Scheme

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Robust Against Geometric Attacks and Frame Dropping ," in *IWDW '09 Proceedings of the 8th International Workshop on* Digital *Watermarking*

[27] M. Fallahpour, M. Semsarzadeh, S. Shirmohammadi, and J. Zhao, "A Realtime Spatio-temporal watermarking scheme for H.264/AVC," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, Minneapolis, Mn, May 6-9, 2013, pp. 872-875

[28] S. P. Mohanty and B. K. Bhargava, "Invisible Watermarking Based on Creation and Robust Insertion-Extraction of Image Adaptive Watermarks," ACM Transactions on Multimedia Computing, Vol. 5, No. 2, Article 12, 2008

[29] M. D. Swanson, B. Zhu, B. Chau, and A. H. Tewfik, "Object-Based Transparent Video Watermarking," *Proceedings of IEEE, First Workshop on Multimedia Signal Processing*, June, 23-25, 1997, pp. 369-374

[30] M. K. Thakur, V. Saxena, and J. P. Gupta, "Performance Analysis of PSNR and SSIM against Frame Drop and its Subjective Score," International Journal of Digital Content Technology and its Applications, Vol. 7, Issue 3, February 2013, pp. 679-688