A Secure Fragile Video Watermarking Algorithm for Content Authentication based on ArnoldCat Map

Rinaldi Munir
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rinadi.munir@itb.ac.id

Harlili
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
harlili@informatika.org

***Abstract*—This paper presents a fragile watermarking algorithm in spatial domain to authenticate integrity of the video digital content. The watermark is a binary image and it has been replicated so that has the same size with frame size of the video. To increase security, before embedding, the watermark is encrypted by XOR-ing it with a random image. The random image is generated by using a chaos map, i.e Arnold Cat Map. The encrypted watermark is embedded by modifying pixel values of video frames. Some attacks has been done to the watermarked video. Experiment results show that the algorithm can detect and localize the modified region of video frames very well.**

Keywords—fragile watermarking, video, Arnold Cat Map

I. INTRODUCTION

Digital video is one of a kind of digital data which contains more information than an image. An single image is only a frame, while a digital video consist of frames of image and audio (if any). Data digital such as a video is easily copied, transferred, edited, altered, or manipulated. Once a digital video is manipulated, the integrity of video changes. In some cases, we need to know authenticity of the video. For example, a court need to decide if a video is genuine or has been manipulated. Fragile watermarking provides a technique to authenticate originality of a video. In the fragile watermarking technique, a data signal that called watermark is inserted into a host video become to a watermarked video without affecting its perceptual quality. The watermark can be extracted again from the video. If the watermarked video is manipulated by using a software, then the extracted watermarked is fragile or broken. Compared to the original watermark, the broken watermark is indication that the video has been altered.

Majority of research about fragile watermarking are specialized for image. However, we could also extent the image fragile watermarking schemes for video sequences. Basically a video is collection of frames where a frame is an image, therefore we could embed a watermark to each frame.

Based on domain to hide the watermark, digital watermarking schemes can be divided into spatial domain and transform domain. In spatial domain, watermarking is performed by modifying pixel values of host video directly [1, 2]. The watermark bits are embedded into pixel values. Digital watermarking in transform domain is performed by modifying of the transform coefficients of the host image. Before embedding the watermark, the host image (in spatial domain) is transformed to a transform domain by using a specific transformation (DCT, DWT, DFT, etc) [4]. The transform coefficients are modified by embedding the watermark bits [2, 3]. Watermarking in transform domain are more robust than spatial domain through non-malicious attacks like cropping, compression, scaling, rotation, etc.

Robust video watermarking in transform domain is solution to the problems of copyright protection, proving ownership, illegal copying, and transaction tracking of video. The watermark in the video is difficult to remove through malicious and non-malicious attacks. On the contrary, fragile video watermarking is suitable to solve the problem of tamper detection of video content. The watermark in the video is fragile when the video is manipulated. Robustness is not important for fragile watermarking.

Fragile video watermarking could be performed either in block-wise scheme or in pixel-wise scheme. In block-wise scheme, the host image is divided into small blocks, then the watermark is embedded into each block. This scheme makes tamper detection could be performed to tampered bocks. In pixel-wise scheme, the watermark is embedded into each pixel, therefore it can identify until pixel level [1].

Based on source of watermark, fragile video watermarking can be classified into two classes. The first is internal watermark scheme, which the watermark information is derived from gray values of host image, then it is embedded into the host image themselves. The second is external watermark scheme, which the watermark is input from the user, usually the watermark is a meaningfull binary image like a logo or something like that.

In this paper we propose a video fragile watermarking algorithm which it should has requirements as follows:

1. **Domain**: Embedding and extraction of watermark is performed in spatial domain and in pixel-wise scheme.
2. **Perceptual quality**: Embedding of watermark should not degrade quality of the host video.
3. **Watermark**: The watermark is a binary image which has the same size with the frame size of the host video. This requirement is made in order to we can identify tampering until pixel level.
4. **Security**: In order to only the authorized parties can do authentication of the received video, then the watermarking algorithm should consider security issue.
5. **Location detection**: the algorithm has capability to localize the area being tampered.

However, unfortunately, not all of the existing fragile watermarking schemes fulfil security aspect so that the

unauthorized parties can do authentication of the received video without know the secret key(s). To overcome this problem, we proposed a secure fragile video watermarking based on chaos map. Chaos system is used in security for two reasons: (1) the nature of chaos is sensitive to initial conditions of the system, (2) random chaotic behavior.

The paper is organized into six sections. The first section is this introduction. The second section will review some related works about fragile video watermarking and chaos map. The algorithm to embed and extract watermark will be explained in the third section. The fourth and fifth section will present the experiment results. Finally, the last section will resume the conclusion and future works.

II. RELATED WORKS

A. Video Fragile Watermarking

Fragile watermarking for digital video has become interesting research topics. Recently digital video is very easy to be modified by using commercial software, therefore one need to prove authenticity of video content. Fragile watermarking algorithm consist of two process: embedding and extracting. Embedding process receives input such as a digital video, a digital watermark, and key(s). To extract the watermark from the video, user gives input such as a watermarked video, key(s), and original watermark. The extracted watermark is compared to original watermark and a decision is made to decide if the video has been tampered or authentic.

Elgamal et al. [1] proposed a fragile video watermarking based on block mean and modulation factor. The original video is transformed from RGB model to YCbCr model, and Cr-component is partitioned into non-overlapping blocks of pixel, depend on the number of bits of the watermarks. Watermark is a binary image. The watermark bits are embedded for each block separately. No key is used either in embedding or extraction of watermark. The proposed algorithm can detect tampering attacks such as filtering and geometric/non-geometric tranformations.

Rupali et al. [5] proposed a public-key fragile video watermarking technique to embed and extract watermark in DCT domain. There are two watermarks to embed. The first waternark is the digital signature of the frame in frequency domain, and the second watermark is numbers of blocks and frame numbers. The first watermark is used to detect tampering and the second watermark is used to localize tampered area. The watermark embedding uses the privat key, while the watermark extraction uses the public key. Anyone who knows the public key can do extraction of watermark. Experiments by changing single pixel value in a block results that tampered block can be detected. However, the proposed technique is not robust against compression.

Zhi-yu et al. [3] proposed a fragile watermarking scheme for the color video authentication. The video first is transformed from RGB model to YST model. The T-component then is partitioned into 4×4 blocks. The watermark, hence is called the authentication code, is created from the quantized DCT coefficient and is embedded into the last non-zero DCT coefficient. Embedding and extraction of watermark use a public key, so that anyone who knows the public key can do embedding and extraction of watermark The experiment results show that the scheme can detect tampering on the watermarked video.

Not all of the watermarking schemes fulfil security aspect. One scheme do not use key at all, the others use public key to extract watermark. Anyone who knows the public key can do extraction of watermark. We need a watermarking scheme so that embedding and extraction of watermark are performed by authorized party only namely the owner of video.

B. Arnold Cat Map

Arnold Cat Map (ACM) is 2-D chaos map that transforms an element from a position to another position in the same area [6]. In other words, ACM transforms coordinate (x, y) from an image $N \times N$ pixels to a new coordinate (x', y'). The iteration equation is

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & b \\ c & bc+1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \bmod(N) \qquad (1)$$

ACM is reversible, i.e the transformed image can be returned to its original image with the equation:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} 1 & b \\ c & bc+1 \end{bmatrix}^{-1} \begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} \bmod(N) \qquad (2)$$

Parameters b and c are arbitrary positive integers, and matrix determinant must be 1 so that the results of the transformation are area-preserving, that is, they remain in the same image area. ACM is repeated m times and each iteration produces an image that looks like random. Values of b, c, and m can be considered as secret keys. After being iterated p times, the image will be transformed back to the original image, as shown in Figure 1. The value of p varies for each image, depending on b, c, N. According to [6], Freeman J. Dyson's research and Harold Falk found that $T < 3N$.

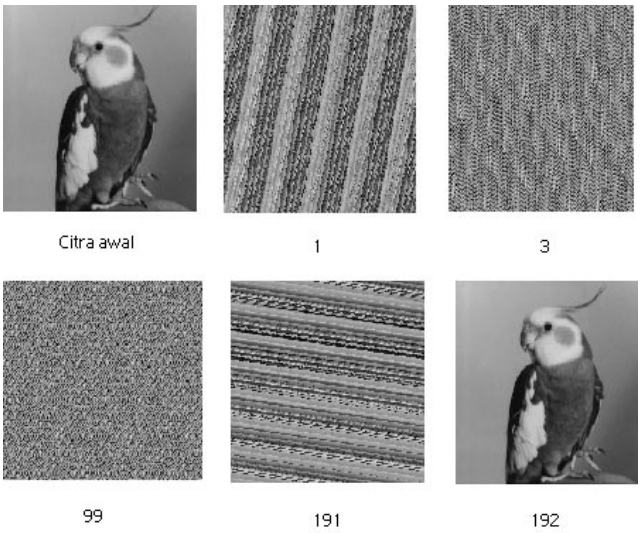


Fig. 1. Results of iteration of ACM [7]

III. PROPOSED ALGORITHM

This section will explain the proposed fragile video watermarking. This algorithm is simple but it can detect manipulation in video frames until pixel level. Embedding and extracting of watermark is performed in spatial domain. To detect and localize manipulation in the watermarked video, we need the original watermark.

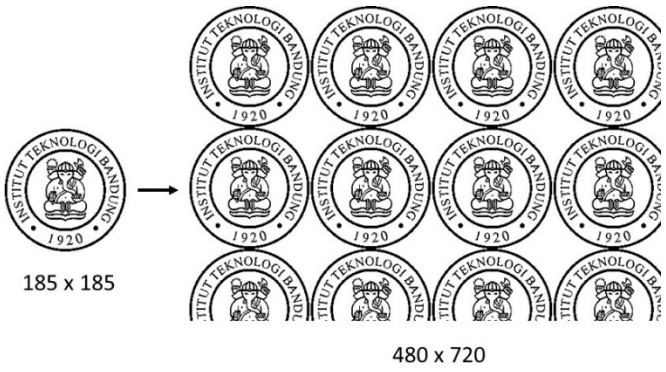


Fig. 2. Left: the original watermark; Right : the new watermark

The watermark is a binary image. However, watermark size may be less than video frame size, therefore the watermark need to be replicated by duplicating it a number of times in order to produce a new watermark that has the same size with the host video frame size. Fig. 2 shows example of replication. The original watermark 'ITB logo' has a size 185 × 185 pixels, whereas video frames have a size 480 × 720 pixels. This original watermark must be duplicated a number of times so that produce a new watermark that has size 480 × 720 pixels

Next, to increase security, before embedding, the new watermark is encrypted by XOR-ing it with a random image. A random image is generated by scrambling an arbitrary binary image by using Arnold Cat Map (ACM). For example, the new watermark above is scrambled by using ACM (see Fig. 3). The map is performed to each single watermark. Parameters of ACM are *b*, *c*, and number of iteration *m*. Different parameters will result different random image. The new watermark is encrypted with the random image by using XOR operation to produce an encrypted watermark. Next, we embed the encrypted watermark into the host video. Because of the new watermark has the same size as the frame size, then we can detect changes to the video frames to the pixel level.

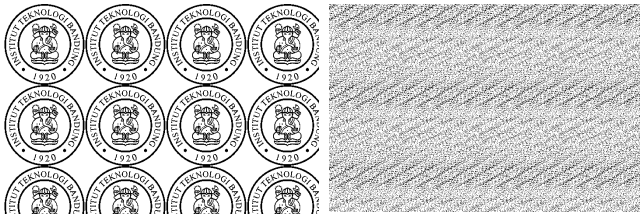


Fig. 3. Left: A new watermark; Right : A random image of the watermark

Based on the explanation above, we could design a simple, but secure, fragile video watermarking algorithm. For simplicity, the random image is generated from the watermark itself. The algorithm consist of two processes: embedding algorithm and extraction algorithm, each will be describe below.

A. Embedding Algorithm

Input: a host video file (*v*), a watermark file (*w*), ACM parameters (*b*, *c*, and *m*).
Output: a watermarked video (*v*')

Step 1: Read frames of video *v*, watermark *w*, and ACM parameters (*b*, *c*, and *m*). If the video has an audio, then separate the audio.

Step 3: Copy the single watermark to produce a new watermark *w*' which has the same size with the host video frames.

Step 2: Scrambling *w*' by iterating ACM (eq. 1) *m* times to produce a random image *r*.

Step 3: Encrypt *w*' as follows: $w'' = w' \oplus r$.

Step 4: Embed the encrypted watermark, *w*", into each frame of the video by manipulating the least significant bit (LSB) of pixels. If the frame has R, G, and B component, then perform embedding to each component.

Step 5: If the original video has a audio, combine it to the watermarked frames to produce a watermarked video.

B. Extraction Algorithm

Input: a watermarked video file (*v*'), an original watermark file (*w*), ACM parameters (*b*, *c*, and *m*).
Output: an extracted watermark, location of tampered frame (if any).

Step 1: Read the frames of video *v*', watermark *w*, and ACM parameters (*b*, *c*, and *m*).

Step 2: Copy the single watermark a number of times to produce a new watermark *w*' which has the same size with the host video frames.

Step 3: Scrambling *w*' by iterating ACM (eq. 1) *m* times to produce a random image *r*.

Step 4: For each frame, extract all of the least significant bit (LSB) of pixels. This step yields an extracted watermark *w*".

Step 5: Decrypt the watermark *w*" as follows: $w''' = w'' \oplus r$.

Step 6: Compare *w*"' with *w*'. If *w*"'= *w*', we conclude that the integrity of video is authenticated. If not, go to step 7 and 8.

Step 7: To localize tampered region, subtract *w*' to *w*"'. If a pixel is not changed, the subtraction yields 0, else the subtraction yields 1.

Step 8: Identify pixels in the watermarked video in position where the subtraction above yields 1. Those are pixels that have been manipulated.

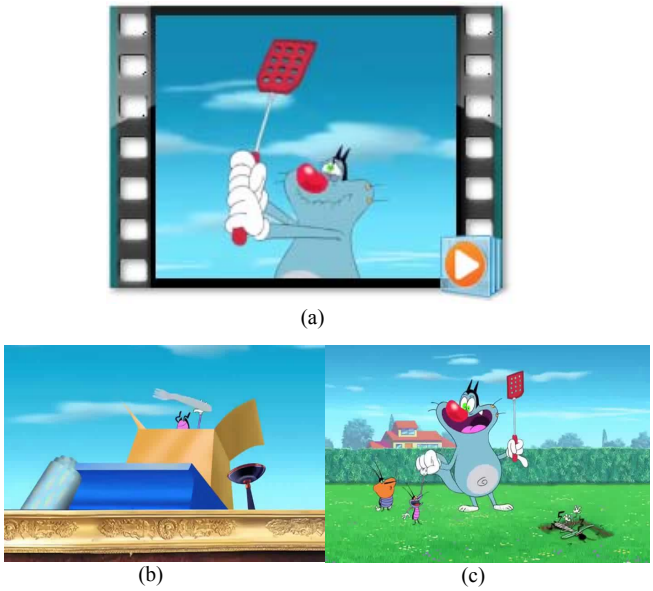
IV. EXPERIMENT RESULTS

After the watermarking algorithm has been designed, we implemented the algorithm to a computer program. We test the algorithm to a sample video. The sample video was a video clip of cartoon film which has 450 frames, each frame has a size 480 × 720 (Fig. 4a). This video contains audio inside. Fig. 4b and 4c show two frames of the video (frame 1 and frame 283).

The watermark to be embedded into the host video is 'ITB logo' as shown in Fig. 3 (after duplicated a number times so that produce a new watermark which size 480 × 720). In these experiments below we used parameters of ACM as follows:

```
b = 5; c = 7; m = 5;
```

We divide this section into two cases: (i) no attack case, and (ii) tamper detection test.



(a) (b) (c)

Fig. 4. A host video to be watermarked

A. *No-Attack Case*

In this case, we didn't manipulate anything to the watermarked video. We extracted the watermarks from the video. We could extract all watermarks in all frames or extracted watermarks only in certain frames. Fig. 5 shows the extracted watermarks from frame number 1 and frame number 283 only. There is no damage in the extracted watermarks. We conlude no tampering performed to the watermarked video.

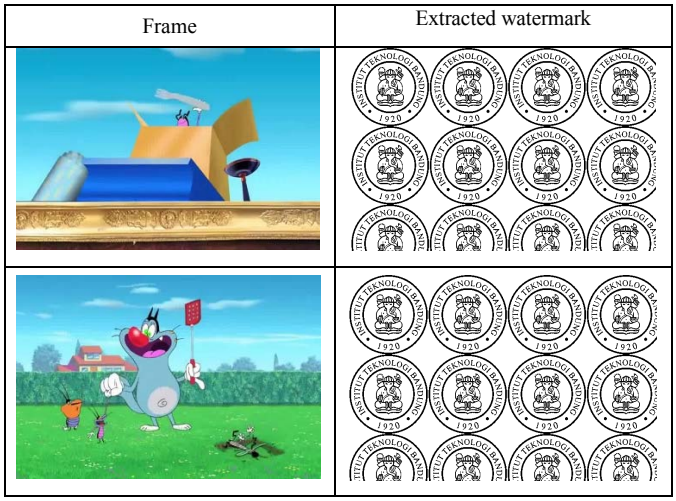


Fig. 5. The watermarked frames and the extracted watermarks

In this algorithm, parameters of ACM behave as keys. Embedding and extraction of the watermark could be done by the authorized party. If the receiver didn't have the keys, then the extracted watermark is not the same as the original watermark. Therefore this algorithm provides security aspect. For example, the receiver used b = 10, c = 8; m = 5 to extract the watermark. Fig. 6 shows the extracted watermarks from two frames. Compared to the original watermark, this watermark is not same.

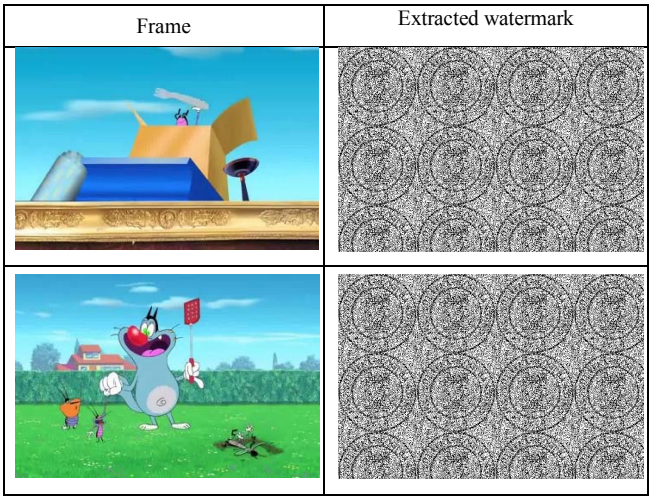


Fig. 6. The watermarked frames and the extracted (wrong) watermarks.

B. *Tamper Detection Test*

Main goal of fragile watermarking is to determine wheater the video has been manipulated or not. If the video has been manipulated, it also able to locate where the alteration made on the video frames. In these experiments, we performed some attacks to the watermarked video.

1. Detection Test Againts Text Addition

On this test, we modified the watermarked video by adding a text **'(C) Cartoon Network**' at the left top of the frames (Fig. 7a). Next, we extracted the watermark from a frame and we got an extracted watermark (after decrypted) contains the adding text (Fig. 7b). Fig. 7c shows detection of pixels that have been manipulated by adding a text **'(C) Cartoon Network**'. Fig 7d shows the tampered pixels in the correspondence frame.

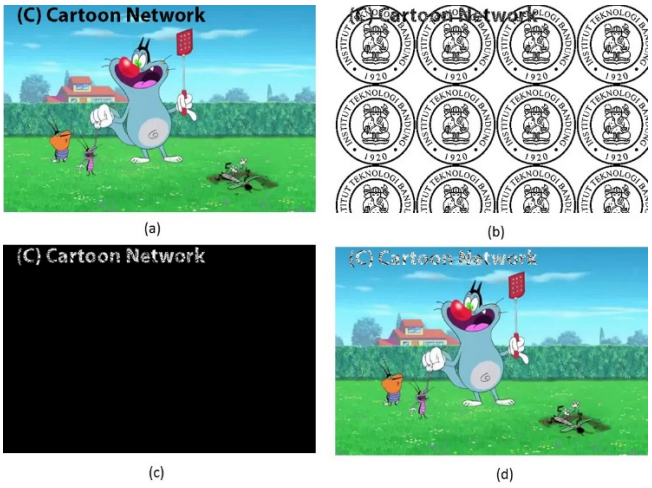


Fig. 7. (a) Watermarked frame after adding a text, (b) extracted watermark; (c) and (d) detected tampering region

2. Detection Test Againts Copy-Paste Attack

On this test, figure of character 'Doraemon' was copied and pasted into the watermarked video (Fig. 8a). We extracted the watermarks from the video and got an extracted watermark as shown in Fig. 8b. The watermark contains a

silhouette of strange object. Detection of copy-paste object is shown in Fig. 8c and 8d. We concluded that the video has been tampered.

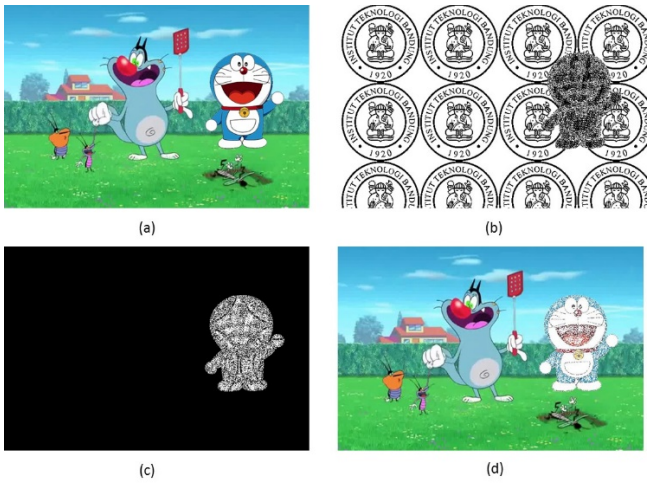


Fig. 8. (a) Watermarked frame after copy-paste attack, (b) extracted watermark; (c) and (d) detected tampering region

3. Detection Test Againts Adding Noise

Some videos maybe contain noise. There are some kind of noise such as gaussian noise, salt and pepper noise, poisson noise, etc. In this test, we added salt and pepper noise with density 0.1 into the watermarked video (Fig. 9a). We found that the extracted watermark also contained noise that indicated the video has been altered (Fig. 9b). The tampering region is entire of frame (Fig 9c and 9d).

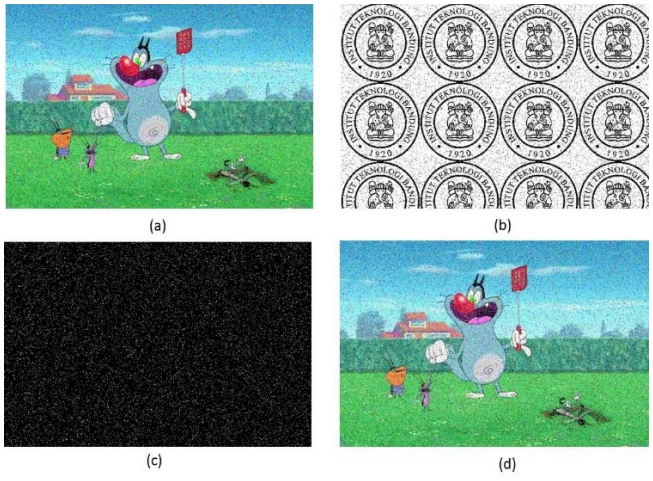


Fig. 9. (a) The watermarked frames after adding noise 'salt and pepper'; (b) the extracted watermark; (c) and (d) detected tampering region

4. Detection Test Againts Contrast Change

One of the common manipulation of video is filtering. The video is manipulated by changing its contrast or brightness. On this test we changed contrast of the watermarked video and then extracted the watermark (Fig 10, left). The extracted watermarks contain noises (Fig 10, right).

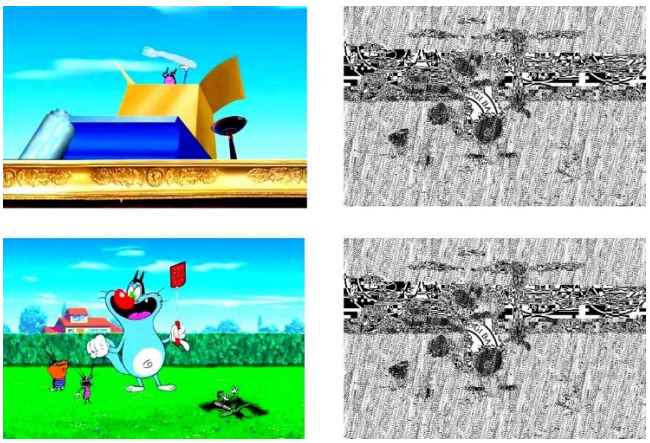


Fig. 10. The watermarked frames after changing the contrast and the extracted watermarks

5. Detection Test Againts Cropping

Video cropping is one of geometrical attack. We manipulated the watermarked video by cropping the certain region, horizontally or vertically. In this test we cropped a low part of the video. Before we extracted the watermark, we returned the frames size into original size by adding white or black pixels. We found the extracted watermarks contained the black region that indicated the cropped region in the frames (Fig. 11).

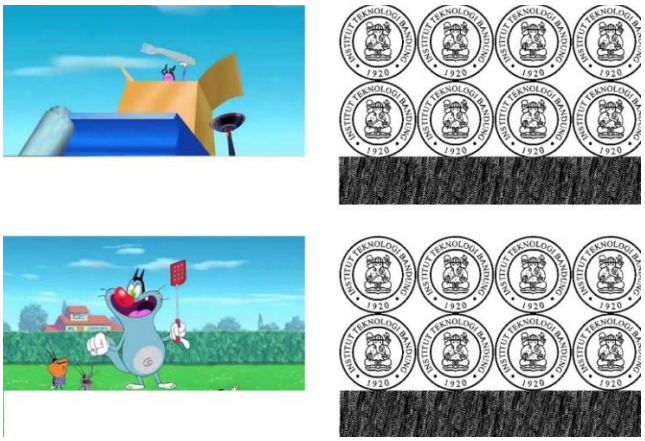


Fig. 11. The watermarked frames after cropping and the extracted watermarks

V. DISCUSSION

We have run some experiments to test performance of the fragile video watermarking algorithm. If the watermarked video is not manipulated, altered, or tampered, then we will get the extracted watermark is same exactly to the original watermark. It is indication that the integrity of the video is authenticated and we conclude that the video is genuine.

Next we manipulated the watermarked video by adding a text, inserting a new object into the video, changing contrast, and cropping some pixels. Adding a text or inserting an object into the video result the extracted watermarks that contain silhouette of the object or text. We can detect them visually. Compared to the original watermark, the extracted watermark is not same. We conclude the video has been altered. The algorithm could detect the tampered region very well.

Changing contrast and brightness of video mean changing all pixel values. As the result, the extracted watermark also change entirely. It is broken totally. We conclude the vide has been manipulated.

Cropping a block area in the watermarked video results the extracted watermark is also cropped in the correspondence block. The watermark has a black region in the cropped area of the correspondence frame.

VI. CONCLUSION AND FUTURE WORKS

A secure fragile video watermarking on spatial domain based on Arnold Cat Map has been presented. Some experiments has been run to test the performance of the algorithm. The experiment results showed that the algorithm could detect tampering on the watermarked video. The algorithm could also locate the tampering region.

For future works, the algorithm can be developed for the compressed video. Embedding of watermark is performed in encoding and decoding stage of the video. Of course watermark embedding and extraction must be operated in transform domain.

ACKNOWLEDGMENT

Thank to Institut Teknologi Bandung (ITB), Indonesia. This research is funded by *Program Penelitian dan Pengabdian Masyarakat ITB (P3MI) 2019*.

REFERENCES

[1] A.F. Elgamal, N.A. Mosa, W.K., ElSaid, "A Fragile Video Watermarking Algorithm for Content Authentication based on Block Mean and Modulation Factor", International Journal of Computer Applications (0975 – 8887) Vol. 80 – No.4, October 2013.

[2] T. Jayamalar, V. Radha, "Survey on Digital Watermarking Techniques and Attacks watermark", International Journal of Engineering Science and Technology, Vol. 2, No. 12, pp 6963-6937, 2010.

[3] H. Zhi-yu, T. Xiang-Hong, "Integrity Authentication Scheme of Color Video Based on the Fragile Watermarking", Proc. of 2011 International Conference on Electronics, Communications and Control (ICECC).

[4] Maryam A., Mansoor R., Hamidreza A., "A novel robust scaling image watermarking scheme based on Gaussian Mixture Model" in Expert Systems with Applications 42, 2015, pp 1960–1971.

[5] Rupali D. P., Shilpa M., "Fragile Video Watermarking for Tampering Detection and Localization", Proc. of 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.

[6] Katherine S., "A Chaotic Image Encryption", Mathematics Senior Seminar, 4901, University of Minnesota, Morris, 2009.

[7] Rinaldi M., "Algoritma Enkripsi Citra dengan Kombinasi Dua Chaos Map dan Penerapan Teknik Selektif Terhadap Bit-bit MSB", Proc. of Seminar Nasional dan Aplikasi Teknologi Informasi (SNATI), Universitas Islam Indonesia Yogyakarta, 2012.