Fragile Watermarking Scheme for Authentication on Video File

Fitrandi Ramadhan

School of Electrical Engineering and Informatics,

Bandung Institute of Technology

[1]23515050@std.stei.itb.ac.id

[*]*Jalan Ganesha No. 10, Bandung*
Indonesia.

Abstract—**Nowadays, usage of media file such as pictures, animation, and video are very high. There are some cases in Indonesia of people reporting to the law enforcement of defamation on the internet. Video as one of the media used, often act as evidence on these lawsuits. These evidence had to be the same and undergo no change in any circumstances. This research will analyze and develop a fragile watermarking scheme using SVD. This scheme will utilize the u and v matrix of the decomposition to construct a binary matrix that will be transformed using arnold's cat map. The binary matrix will then embedded as the watermark on the LSB plane of the host image. Extraction will take that LSB plane and construct a new binary matrix from the host image using the same proccess as the embedding. These two matrices will then be operated with xor and give us the difference if there is any. These scheme hoped to become one of the method to detect any tamper be it intentionally or accidental.**

Keywords— **Fragile watermarking, Video, SVD, Arnold's cat map.**

I. Introduction

In digital era nowadays, information sharing and communication more often done in the internet. There are quite many social media and message board that act as platform on this type of communication. One of many and one of the most popular message board right now is reddit. The increment of usage on reddit is illustrated on the survey on reddit submission where in 2015 reddit as gone through 150 million submission. As the message board provider reddit did not restrict any post except on racism. This polici encourage more information and knowledge sharing. On the other hand this also increase negative content around the community. On of the negative content out there is targeting defamation on some people or community.

In this message board on of many information shared is information media, likely smaller media such as image, animation, or video. Many people in these message board exchanging something called meme. By definition meme is a culture transfer or system that is spread from people to people. By this definition meme is a method, not the object itself but media used to contain these meme is usually an image, animations, or videos.

This culture not only contain positive content but also negative content at some cases. Like mention before some of it contain targeted defamation. This defamation at least some in Indonesia, end in lawsuits. As we all already know lawsuit require evidence. In this case, the evidence will be the media file, be it a video, or any digital file.

Some high profile cases in Indonesia that uses digital file as their evidence is Setya Novanto's case which uses audio file, Ahok's Pulau Seribu Speech Case which highlighted video as evidence, and Cyanide Coffee case which uses cctv video as evidence.

Authentication methods are needed to check the authenticity of any video evidence. To make it more specific, a watermarking method is needed to mark any video with high invisibility watermark and exact authentication. Video which basically are matrices of pixel value on time dimension, could be broken down with SVD. Its feature could be extracted and made to a binary image frame by frame and later embedded again on the host image as a watermark. This watermark could be recreated and inserted at the same time in the extraction process to authenticate the video. Arnold transform are used to authenticate the video on the time dimension so that frame scramble, addition or removal could be detected also.

This method hopefully could help the authentication that didn't change human perception on the video content in lawsuit trial or other similar use case on any domain that needed it.

II. Literature Study

A. *Fragile Watermarking*

Fragile watermarking is a watermarking method that used for exact authentication. File are embedded with watermark that can be checked if that particular file is the same watermarked file. Fragile watermarking usually not preferred for common use because of any change done on the file be it on purpose or accidentally will be treated as different file. Fragile watermarking have some niche usage such as checking if work on progress have any tamper or change coused by noise. In principal had any change undergo on the file watermarked.

B. *Singular Value Decomposition*

Singular value decomposition is a factorization techniques on matrices using their eigen value. SVD utilize the fact that theres always 3 unique constructing matrices, M,Σ, and N, SVD define formally as.

$$A = U \Sigma V^{\mathcal{T}} \quad \text{(II.1)}$$

with:

A = Input Matrix.
U = left singular vectors
Σ = singular values or λ's
$V^{\mathcal{T}}$ = transpose of right singular vectors

Decomposition using SVD produces a concept of relations between matrix elements. Each eigen value of the matrix will form a matrix Σ where the matrix Σ shows the strength of a concept. The U matrix is a matrix that provides information on the relationship between rows in Matrix A with a concept. Third is matrix V, this matrix provides information between the relationships between columns in matrix A and the concepts described.

C. *Arnold's Cat Map*

arnold's cat map is a matrix transformation that is defined by the following equation,

$$\Gamma\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} mod\ N \quad \text{(II.2)}$$

with:

Γ = Transformation.
x = x axis position of value.
y = z axis position of value.
N = Matrix size / rank.

A matrix transformed using arnolds cat map will give results in the form of a matrix that has a collection of values that are the same as the previous matrix, but the position of these values will change and this change is chaotic when compared to Arnold's cat map with different matrix sizes. This characteristic proves that arnold's cat map gives diffusion. Another unique feature of arnold's cat map is mapping will return matrix values position to initial position.

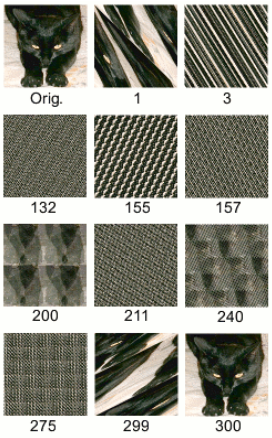


Fig. 1 Arnold's cat map iteration

D. *Related Works*

Zhang (2017) uses the SVD (singular value decomposition) scheme to get the uniqueness of a matrix. This study explains that the singular value decomposition can be used to obtain a unique binarization of a matrix. This unique matrix is then used as a matrix basis to get a binary matrix by binarizing by determining a threshold value.

It is emphasized that the use of the image will often meet values that have a small difference or have the same value as the values surrounding it. This is an important basis in this study because the proof will be the determinant of the threshold value used.

$$\lambda_1 u_1 v_1^{\mathcal{T}} = \frac{1}{\lambda_1}\begin{bmatrix} a & a & a & a \\ a & a & a & a \\ a & a & a & a \\ a & a & a & a \end{bmatrix} = \frac{1}{\mathrm{N}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{III.3}$$

with:

λ = nilai eigen.
U = eigenvector ortonormal dari AA^T.
$\mathrm{V^T}$adalah eigenvector ortonormal dari $A^{\mathrm{T}}A$

Based on the above equation, a matrix with all the values is 0.25, so the value that is considered appropriate is 0.25 and this value is influenced by the N or the size of the matrix. This threshold value determines the binarization of the matrix whether the values in the matrix will be changed to 1 or 0. This binary matrix will then be the value in the watermark matrix that will be embeded on the image.

III. Analysis

Indeed, a video is a collection of images that are displayed in a certain order within a certain time span. If it is broken down then an image is a matrix of certain color intensity values. In this case, of course, the image on the computer will be used which has three color values to form other colors. These colors are red, green, and blue. These three colors will form the colors of other derivatives that you want to display at one point. Each drop has a position so that the matrix form generated from an image is the value of the color intensity at the position of a particular point.

Singular value decomposition as an algorithm that can be used to simplify the more complex values of a matrix, if applied to the image will provide a matrix forming from the initial matrix with a simpler value. There are two important points of the existing attributes of this forming matrix which are unique matrices. The singular value decomposition of a matrix with size M × is as follows.

$$A = U\Sigma V^{\mathcal{T}} = [u_1,\ u_2,\ u_3,\ \dots\ u_m]\ \Sigma\ [v_1,\ v,\ v_3,\ \dots\ v_m]^{\mathrm{T}} \quad \text{(III.1)}$$

with:

U = eigenvector ortonormal dari $\mathrm{AA^T}$
$\mathrm{V^T}$ = eigenvector ortonormal dari$\mathrm{A^TA}$
Σ = diagonal matrix of singular values

The matrix size Σ that is formed must be valid with matrix multiplication rules if additional rows or columns are needed in this matrix multiplication then rows or columns will be added with a value of 0.

The first point of the important attribute of this matrix is that each unique matrix will have three unique forming components. This becomes a very important attribute because a fragile watermarking must be able to detect changes as small as possible and use the smallest possible sample. Besides that, size is also an important point of performance of the process carried out. With this method, it is proposed that a fragile watermarking process can be performed on a video file. The second point is the form that results from a matrix with the same values. As exemplified in the following equation.

The following equation shows that for matrices that have the same value only have λ_1. Because there is only one λ, the equation becomes as follows.

$$u_1 v_1^T = \frac{1}{\lambda_1} A = \frac{1}{\mathrm{a\ N}} \begin{bmatrix} a & a & a & a \\ a & a & a & a \\ a & a & a & a \\ a & a & a & a \end{bmatrix} = \frac{1}{\mathrm{N}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{(III.2)}$$

If we look back at the form of a matrix that is often formed by an image, it will be seen that the relationship between the matrix form shown by the above equation can be observed that in the matrix values that have the same value can be done that special operation. In an image, it is not uncommon for us to meet matrix values like this repeatedly over and over. The adjacent pixel value in an image often has the same value especially when dividing images into smaller matrix sizes. Differences will be more often found in the boundary values in the image, such as pattern boundaries or certain shapes.

On different values the matrix will also be able to represent the histogram of the appearance of the pixels contained in the matrix. So that in a matrix where there is a value that is still dominated by certain values, the forms of these properties can still be maintained.

Then a binarization process can be done in the image. The purpose of this binarization is to facilitate the watermark checking process from an image. In this matrix factorization, certain values can be found that have high occurrence values. As in the previous paragraph these high values are values that generally have an adjacent position in the image. The method developed by Zhang et al. (2017), because the matrix size used is always the same, namely n = 4, and based on the proof of the equation y he chooses the method of selecting the same threshold value of 2.5. The proof is continued by an experiment and an invisible assessment that the use of a threshold value of 0.25 gives the best form of texture than the other values.

A value of 0.25 determined by Zhang et al. (2017) is not only proven on the basis of visible differences. For most of the images tested the matrix shows that the distribution of values in matrix u_1 v_1 ^ T follows the normal distribution and is distributed at a value of around 0.25. This threshold corresponds to the equation [X] which shows an example for a 4 × 4 matrix. If for all matrix block matrices selected are other sizes N × N then the distribution of the normal distribution will be 1⁄N.

Unlike the watermarking scheme in the picture. The watermarking scheme in the video also needs to deal with changes in information that moves in 3 axes with the third dimension is time. As explained in [1] information is only encoded in one image while to get relevant results and can detect changes in the video processing needs to be carried out involving each frame and forming an equivalent dependency between frames.

The scheme that would be proposed in this study to get bonds between frames is a diffusion operation that involves the values in the frame as input from the process. In the scheme proposed in [1], matrix permutation operations are added namely arnolds transform or arnolds cat map with keys as input to improve the security of the algorithm. This scheme utilizes the diffusion capability of the arnold's cat map to scramble images. Utilizing the need for a key from arnolds cat map can be used as a shared value of the frame in the video as the key value. In one video there will only be one value of the total number of frames. so this value must be consistent with the original image and the watermarked image. Arnold transform has a property where this permutation will return the image back to its original position with a certain number of repetitions depending on the size of the matrix being operated.

The number of repetitions needed to return an image will vary according to the rank matrix as an example of the number of repetitions needed to get the same image again.

Table III.1 Arnold's Cat Map Iteration by Size

Matrix Size	Number of iteration
300 × 300	300
257 × 257	258
183 × 183	60
157 × 157	157
150 × 150	300
147 × 147	56
124 × 124	15
100 × 100	150

From the above values we can conclude that the change from the number of arnold transform operations needed for each matrix size is irregular series. according to [1] there is no method that can correctly get the number of iterations needed and only in the form of the approximation value, for that arnold transform operation cannot be done directly due to the deficiency of the decryption algorithm. For this reason, it is proposed to use one of the specified quantities of this arnold transform operation and do nothing more than the required repetition. For example, one example of an operation is a matrix size of 256x256. The number of repetitions required by the to restore the original image is 192 repetitions. For this reason, it is necessary to ensure that the number of repetitions made on the matrix does not exceed that value. To ensure that the value is the upper limit of the number of repetitions, modulo operations can be performed on the value of the number of frames before becoming a key in the map arnolds. To ensure that the arnolds cat map operation performed will meet the requirements of having 192 number of repetitions,

the matrix will be divided into blocks of 256x256 size for handling matrices that do not have the right size or multiples of 256x256 can be added padding so that the image reaches a certain number.

Of course, this operation has a disadvantage that can be the target of an attack on this scheme, that this scheme cannot detect changes in frames with the exact number of changes multiplied by 192 or whatever value becomes the number of repetitions based on the size of the operating matrix of Arnold's cat map

IV. Proposed Scheme

Based on the analysis of the algorithm above it can be seen that an image can be its own binary texture from the decomposition process using SVD. Process output will be used as a watermark.

Decomposition of each frame from the video gets spatial domain from the image of the video file. Video itself sometimes has multiple spatial domains in one frame. Grayscale-based images will only have 1 spatial domain, whereas for images that have three spatial domains, generally the three red, green, and blue domains will be maintained.

In each domain, the LSB removal (least significant bit) will be done. To get consistency and watermark insertion space. This image will be called the image host. Determined the N value that will be used as the independent variable and input to the SVD. Keep in mind that different N values will produce different outputs. In each spatial domain, SVD will be done using the selected N value and one λ will be taken as the threshold that will be used in binarization.

At each value in the domain binarization will be carried out based on the threshold value that is used is the largest λ value. Each binary value will be transformed using a arnold's cat map. Transformation is done as much as i = nFrame mod 192. The transformation matrix will be inserted in the LSB plane host image. This result image will be the image that has been affixed with the watermark.

By using the same process we can do watermark extraction to check the authenticity of the file. Decomposition for each frame of the video gets spatial domain from the image of the video file. Video itself sometimes has multiple spatial domains in one frame. Grayscale-based images will only have 1 spatial domain, whereas for images that have three spatial domains, generally the three red, green, and blue domains will be maintained. The LSB plane of the image will be saved and removed from the image to be examined.

SVD is performed on an image with an N value that is the same as the N value used in the encoding process and retrieves the same value of λ with the encoding process. Re-binarization of images is based on the value of λ. The results of binarization will be transformed using arnold's cat map Transformation is done by the number of iterations as much as i = nFrame mod 192. The absolute differentiation between the values extracted in the LSB Plane is a watermarked image with the results of the binarization carried out. This differentiation can use XOR operations in both binary images If there are differences in these results, the image has undergone changes in different parts of the LSB.

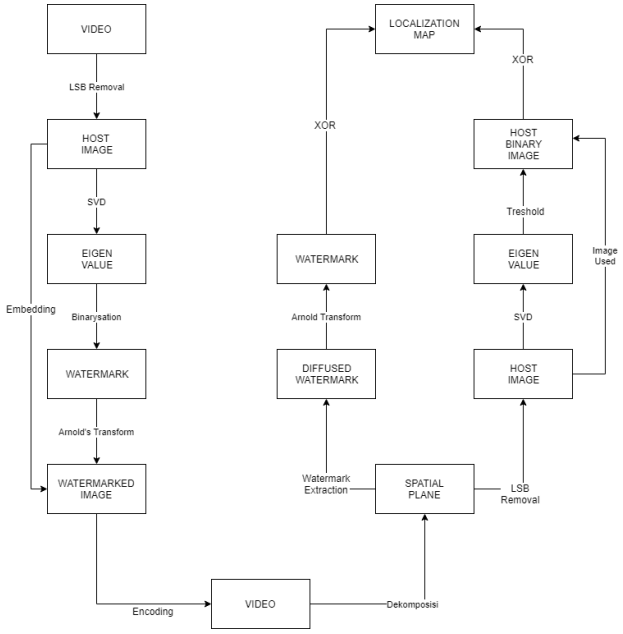


Fig. 2 Proposed Scheme

V. Evaluation

Testing is done by making a video with the existence of MP4 assisted by additional tools namely vegas pro 13. The use of this tool aims to obtain images that are consistent with relatively small and fixed pixel counts of 256 x 256 and ease of processing and a consistent number of frames which is 12 frames per seconds. The duration of the resulting film is also determined in advance, which is 5 seconds.

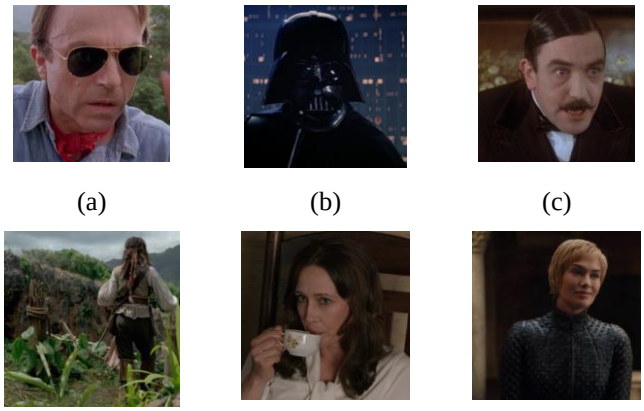


Fig. 3 Original Frame

The test is done on the video image with the image without changes or tempering and the image that is tempered and then examined the difference in the image. Fragile watermarking conducted in this study is watermarking which has a high level of invisibility.

(a) (b) (c)

(d) (e) (f)

Fig. 4 Watermarked Frame

It can be seen from the results shown in Figure IV.1.2 that the image obtained after the embedding process has a relatively high level of invisibility. Thisvisibility in a subjective manner will not affect the human perception of the image. It is hoped that this level of invisibility can increase the use of this scheme in everyday life.

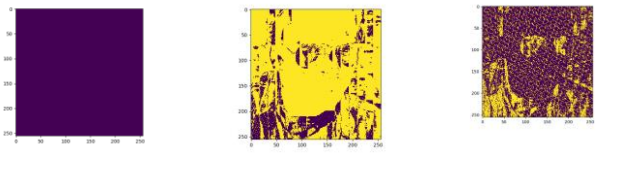
(a) (b) (c)

(d) (e) (f)

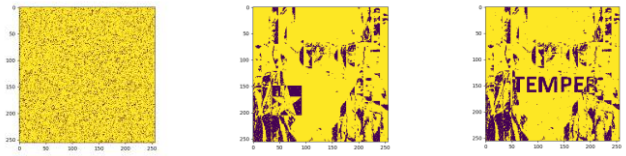
Fig. 5 (a) untouced (b) crop (c) frame addition (d) kompresi (e) object addition (f) text addition

It can be seen from the results shown in Figure IV.1.2 that the image obtained after the embedding process has a relatively high level of invisibility. Thisvisibility in a subjective manner will not affect the human perception of the image. It is hoped that this level of invisibility can increase the use of this scheme in everyday life.

Then the xor results will be displayed from each image in Figure IV.2. these results will be expected to show a binary image of the comparison between the watermark and the tempered host image. This result will also show the position of the attack carried out with the expectation of providing information about the position of the attack carried out on the image.



(a) (b) (c)



(d) (e) (f)

Fig. 6 (a) untouced (b) crop (c) frame addition (d) kompresi (e) object addition (f) text addition

From the results of the XOR image testing results from the watermark without any attack it can be seen that binary images display false values on all pixels which shows that there is no difference at this point, but the frames that do not always change the pixel tampering will show a value of 1 and another pixel is 0. The XOR image shows that the tampering position cannot be detected mathematically. Comparison of false and true values on the frame is obtained below.

Table IV.1 Count of XOR

Citra	Serangan	Jumlah Pixel (0)	Jumlah Pixel (1)
(a)	Tidak ada	262.144	0
(b)	Crop	209.638	52.506
(c)	Frame Removal	174.962	21.646
(d)	Kompresi	164.578	97.566
(e)	Object Addition	205.701	56.443
(f)	Text Addition	209.059	53.085

In the results of this test can be compared the number of changes given to the video frame with the amount detected for changes. Without calculating the area of change or calculating accuracy only from the number of detection changes in the entire frame, changes can be detected for frame (b), (e), and (f) using the following F1-score.

Table IV.2 F-Score

Citra	Serangan	Jumlah Pixel (0)	Jumlah Pixel (1)	F-Score
(b)	Crop	209638	52506	0.4%
(e)	Object Addition	205701	56443	4.4%
(f)	Text Addition	209059	53085	0.6%

In the F-score can be seen from the three calculations above that the resulting accuracy is very unreliable to be used as a basis for localization. The frame can visibly see its temper position and each frame that gets a change can be localized when using the human perception of the object. unfortunately

this perception cannot be proven formally and proven by the results of the F-score produced.

Analysis with visible perceptions shows that it can be compared with the image in (b) where the neck part of the actor looks entirely white due to crop section scarf although the differences are still difficult to perceive without attention.

In figure (c) is shown the result of an additional frame attack. The addition of the frame does not give a specific mark on xor but still shows a considerable difference in the xor results even though the image produced in the xor results still has the texture of the frame. This is possible because of the similarity of the frame that is embedded in the original frame sequence. In figure (d) shown the result of compression. These results provide a much different texture. This is possible because of significant color map changes between original images and compressed images. In figure (c) and (e) a significant difference is displayed where the object can be seen clearly in the xor image. In figure (e) displayed a star-shaped object that is quite clear and in the image (f) the text "TEMPER" is displayed quite clearly.

This test proves that for the five attacks can be shown the difference between the XOR results of the watermarked image and XOR results from the image of the watermark that has been tempered. Xor images can show the difference between the distemper parts for certain attacks. For crop attacks and object addition xor images still depend on the color of the object that was deleted or added.

Tabel IV.3 Time elapsed

Citra	Serangan	Waktu
(a)	Tidak ada	7.01s
(b)	Crop	6.74s
(c)	Frame Removal	7.74s
(d)	Kompresi	6.74s
(e)	Object Addition	6.89s
(f)	Text Addition	6.75s

From the calculation of this time it was found that there were no significant differences from different attack checks. However, because this inspection uses videos with a size of 256 x 256 with 25 frames per second, each pixel added either from the side of the frame or resolution will add the number of comparisons made. In addition to the frame or layer will add time linearly while the addition of resolution will add a quadratic time compared to the resolution used.

VI. CONCLUSIONS

The conclusions obtained based on the results of research that has been carried out are as follows:

1. Fragile watermarking can be done on video by doing SVD on video then binarized with certain threshold.
2. The fragile watermarking method on video using SVD can be done by dividing each image into matrices for each frame and comparing binarization between SVD of watermarked images and the watermark itself.
3. arnold's cat map has an empirical number of round iterations so that a large determination of the image is needed before transformation. To do this, additional processes such as padding are needed which can reduce the equation between the original image and the watermarked image.
4. Rotation iteration in arnold's cat map limits the keys obtained from the number of video frames so that the tempering performed on the number of frames when done exactly in accordance with the arnold's transform's iteration number cannot be detected.
5. Fragile watermarking using SVD can provide authentication for video images from the tempering side of the image area and also tempering which is done by using arnold's cat map.
6. 6. The scheme does not provide good accuracy in terms of formal verification of localization as indicated by the F-score generated by the extraction.
7. 7. The performance of time is not affected by the type of attack carried out, but the number of checks carried out on videos of a larger size will be greater..

REFERENCES

[1] Heng Zhang, Chengyou Wang, Xiao Zhou. 2017. School of Mechanical, Electrical and Information Engineering, Shandong University China. FragileWatermarking for Image Authentication Using the Characteristic of SVD.

[2] Munir, R. (2017): School of Electrical Engineering and Informatics, Institut Teknologi Bandung. A Fragile Watermarking Scheme for Authentication of GIF Images.

[3] Ait Sadi, K. Guessoum, A. Bouridane, A. Khelifi, F. 2012. Content fragile watermarking for H.264/AVC video authentication.

[4] Alomari, Raja' S., Al-Jaber, Ahmed. 2004. Computer Science Department, King Abdullah II School for Information Technology. University of Jordan. A Fragile Watermarking Algorithm for Content Authentication.

[5] Baharav, Zachi. Shaked, Doron. 1999. HP Laboratories Israel. Watermarking of Dither Halftoned Images.

[6] Baretto, P. Kim, H. 1999. Pitfalls in Public Key Watermarking, Proceedings of Sibgrapi-Brazilian Symposium on Computer Graphics and Image Proccessing. pp 241-242

[7] Busiri, Moch Ginanjar. Munir Rinaldi. 2017. School of Electrical Engineering and Informatics. Institut Teknologi Bandung. Mobile Application of Video Watermarking Using Discrete Cosine Transform on Android Platform.

[8] Cox, Ingemar J., Miller, Matthew L., Bloom, Heffrey A., Fridrich, Jessica., Kalker, Ton. 2008. Morgan Kauffman Publisher. Digital Watermarking and Steganography.

[9] Doig, Jeremy; Jazayeri, Mike (19 May 2010), Introducing WebM, an open web media project, WebM Project, retrieved 15 November 2017.

[10] Graphics Interchange Format. A Standard Defining a Mechanism for the Storage and Transfmission of Raster-Based Graphics Information. 1987. CompuServe Incorporated.

[11] https://github.com/webmproject/libvpx/releases/tag/v0.9.0. 15 November 2017

[12] https://permadi.com/2010/06/webm-file-structure/. 15 November 2017

[13] https://www.webmproject.org/docs/container/. 16 November 2017.

[14] Linnatrtz, J. P. M. G., Talstra, J. C. 2006. Philips Research. MPEG PTY-Marks: Cheap detection of embedded copyright data in DVD-video.

[15] R. Machado, EZStego, http://www.stego.com.Tseng Y., Chen Y., and Pan H., 2002. A Secure Data Hiding Scheme for Binary Images. IEEE Transactions on Communications. Vol 50. No. 8. pp 1227-1231