An Improved Neighbouring Similarity Method for Video Steganography

Diksy M. Firmansyah, Tohari Ahmad
diksy14@mhs.if.its.ac.id, tohari@if.its.ac.id
Institut Teknologi Sepuluh Nopember
Surabaya, Jawa Timur 60111 Indonesia

***Abstract*- Video is a digital media that is often used now. Because of that, information security issues in the video have been popular now, which one of the issues is steganography. Neighbouring similarity (NS) is the steganography method for video. This method is using macro block that is generated in video encoding process, called prediction error (PE), to embed a secret message. Generated PE is able to improve the capacity of the secret message than existing method, but NS still has potency to generate greater number of PE. In this paper, we propose a method to increase the number of PE by modifying NS architecture of PE generation process. The experimental results show that the proposed method is able to improve the capacity of secret message and keep the embedded video in high quality level.**

I. Introduction

Digital media is a popular format that is used to transfer information across people in the world. It is because digital media is able to run using internet connection. In other side, internet has many issues in information security. One of them is data hiding. Data hiding is method to hide a data into another data. Data hiding is able to be used for copyright protection, sends a secret message, inserts addition data into digital media, etc.

Steganography is one branch in data hiding topic, where the other branch is watermarking [1]. Like watermarking, steganography hides a secret message by using another data as camouflage media. But, the goals of watermarking and steganography are different. Watermarking displays both secret message and camouflage media, which steganography only shows a camouflage media. The goal of steganography is how the secret message can only be read by people who have interest on it. Another data that used to cover the secret are called cover data. Additionally, the cover data that have been embedded by the secret are called stego data.

Generally, steganography can be grouped into two categories, irreversible and reversible [2]. Reversible steganography is able to restore the cover data, while irreversible steganography is not. After extracting the secret message from the stego data, reversible steganography must have an algorithm to restore the cover data similar to that before embedding process. So, for sensitive data like a medical or military data, reversible steganography is the best choice if we want to use that to transmit a secret message.

One of the reversible steganography techniques is histogram shifting [3]. It generates histogram of pixel frequency as a reference. The highest frequent, called peak point, is used to embed secret message. To get greater peak point, Tsai et al. in [4] are using the difference of pixel value in some block to generate a histogram. This proposed method is able to increase the value of peak point. For other cover media, Yeh et al. in [5] are using a macro block of video in Group of Picture (GOP) to generate a histogram, called neighbouring similarity method.

Difference expansion [6] is one of popular techniques for image steganography. It uses difference value of paired pixel to embed a secret message. In order to improve the maximum capacity, Alattar et al. in [7] and [8] modifies the architecture of paired pixel. Difference expansion uses two pixels to generate one difference point, which Alattar et al. in [7] uses three pixels to generate two difference points. Alattar et al. in [8] modifies architecture with four pixels to generate three difference points. Ahmad et al. in [9] uses [8] method to improve Reduced Difference Expansion method. The architecture has been able to are proven increase maximum capacity of secret message which can be embedded.

Inspired from [7], [8] and [9], we propose a scheme to improve neighbouring similarity capacity by modifying the technique of prediction error generating procedure. In their research, [5] uses difference of two non-overlapped frames to generate one prediction error; while in this paper proposes to use three non-overlapped frames to generate two prediction error. Experimental results show that the proposed scheme can increase maximum capacity and keep PSNR value above 40 dB.

The rest of the paper is structured as follows. Section 2 describes the research relates to the proposed method, that is Histogram Shifting and Neighbouring Similarity. Section 3 presents the proposed method whose result is provided in section 4. Finally, conclusion is drawn in section 5.

II. Existing Method

A. *Histogram Shifting*

Image is the popular media for cover data because it has a lot of redundancy data. One of the redundancy data is pixel value. In Red Green Blue (RGB) and grayscale mode, range of pixel value is 0 to 255, while the dimension of image generally is more than 255. So, we will find a lot of pixel in image have same value.

The main concept of histogram shifting technique is utilizing greatest value of color pixel to embed a secret message [3]. Embedding process in histogram shifting is described as follows. First, the cover data is converted into histogram of color frequency, which x-axis represents range of image color

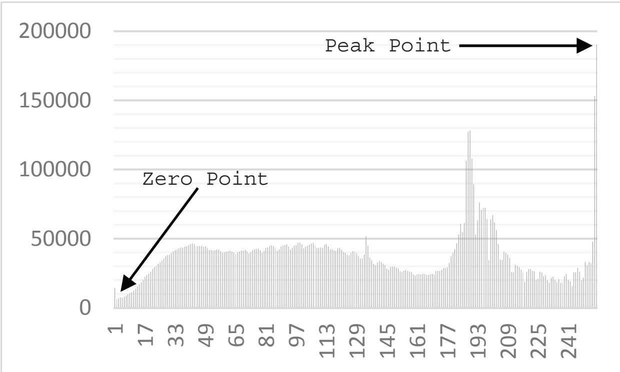

Fig. 1. Histogram of garden_sif with histogram shifting method.

(0 to 255), and y-axis represents frequency of range color. Second, the highest and the lowest value in histogram are defined. Both values are called peak and zero points respectively. An example of histogram is depicted in Fig. 1. Third, the histogram is shifted to provide an empty space beside the peak point. Shifting process can be executed by following:

$$x_{i,j} = x_{i,j} + 1, if\ p < x_{i,j} < z \quad (1)$$
$$x_{i,j} = x_{i,j} - 1, if\ z < x_{i,j} < p$$

where x_i is 8-bit pixel color value, i and j are coordinate of pixel, p is peak point, and z is zero point.

For recovery process, we must set an overhead in $x_{i,j} = z$. The overhead is an information that used in recovery process. The overhead is used because after shifting process, value in side of the zero point will be cumulative with value in the zero point. If we find the pixel value is $x_{i,j} = z$, we must know that it will be recovered to $x_{i,j} = z \pm 1$ or hold in $x_{i,j} = z$. The overhead is combined with secret message.

Embedding process of secret message will be done with this following:

$$x_{i,j} = x_{i,j} + b_k, if\ x_{i,j} = p\ and\ p < z \quad (2)$$
$$x_{i,j} = x_{i,j} - b_k, if\ x_{i,j} = p\ and\ p > z$$

where b_k is k-bit of secret message. That equation is suitable to use for stego data quality because the pixel value change only one bit for any condition.

To extract secret message, peak and zero points must be firstly known. Then, all pixels in stego data are scanned. The exactly process is done by employing Eq. (3).

$$b_k = 0, if\ x_{i,j} = p$$
$$b_k = 1, if\ x_{i,j} = p + 1\ and\ p < z \quad (3)$$
$$b_k = 1, if\ x_{i,j} = p - 1\ and\ p > z$$

B. Neighbouring Similarity

Histogram shifting method can implement in a video because the video is a sequence of image. Each image in the video is called frame. Generally, each frame in video will be similar with adjacent frame. This situation is used by Yeh et al. in [5] to implement histogram shifting technique, called neighbouring similarity method.

In histogram shifting, greater value of peak point is better because the capacity of secret message will be increased. To get greater value in a video, prediction error is used in histogram generation process. After that, prediction error is used to embed a secret message. Prediction error is macro block that is generated in prediction encoding process of a video. Neighbouring similarity is modifying prediction error generating process by following equation:

$$AE = \left|\sum RMB_{i,j} - \sum EMB_{i,j}\right|, 1 < i \leq m, 1 < j \leq n \quad (4)$$

where RMB is Reference Macro Block (that is generated from video frame which not changed), EMB is Encoded Macro Block (that is generated from video frame which changed), i and j are coordinates of row and column, and m / n is the column / row size of macro block. Neighbouring similarity is used absolute value to get greater peak point in the histogram.

Neighbouring similarity has other advantage, that is not using zero point. This method is assuming that the zero point is always in right (255). So, shifting process is always directing to the right side. Overhead information will be generated for $EMB_{i,j} = 255$.

Neighbouring similarity is able to be applied in frame level, not only in macro block encoding level. We can generate prediction error from difference value of two non-overlapped frames, but we must prevent the frame from underflow / overflow problem. We can adopt a method from Tsai et al. in [4] by shifting the histogram into two zero point. First, two zero point (z_1 and z_2) are selected from all frame, where $z_1 < z_2$. Then, the histogram will be shifted by this equation:

$$x_{i,j} = x_{i,j} + 1, if\ x_{i,j} < z_1 \quad (5)$$
$$x_{i,j} = x_{i,j} - 1, if\ x_{i,j} > z_2$$

For recovery process, overhead information for $x_{i,j} = z_1$ and $x_{i,j} = z_2$ must be created. Example of the generated histogram in neighbouring similarity is depicted in Fig. 2.

III. Proposed Method

Neighbouring similarity method is able to produce greater value of peak point because difference value between frames in a video sequence are not very contrast. Frame 1 will be similar with frame 2. Frame 3 will be similar with frame 4, etc. Because of that, greater value of histogram will be on the left. The problem is this method does not consider that frame 2 will be similar with frame 3, frame 4 will be similar with frame 5, etc.

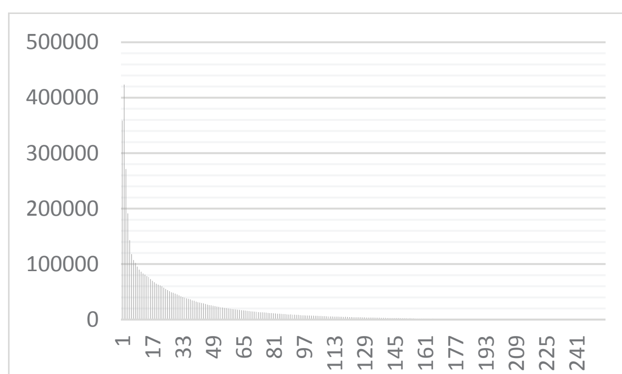

Fig. 2. Histogram of garden_sif with neighbouring similarity method.

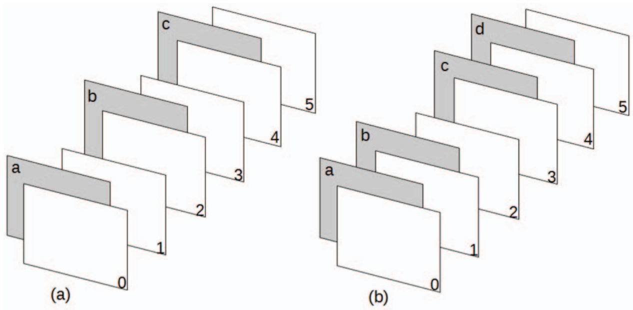

Fig. 3. Illustration of PE generation process; (a) Neighbouring Similarity, (b) Proposed Method

In this paper, we propose to use three frames for generating two prediction error. Each three non-overlapped frames in video sequence, our proposed method will be use the second frame for reference, while the first and the third frame for encoding. So, this method is able to preserved the advantage of neighbouring similarity, but can generate more prediction error. The illustration of prediction error generation process is shown in Fig. 3.

A. *Embedding Process*

First, histogram of the frames is shifted with Eq. (5) to prevent underflow / overflow problem. The shifting method is applied to all frames even though not all frames will be change. It is done to keep the advantage of similarity in video sequence. From the shifting process, this method will be get two overhead information, that is a pixel on the zero point 1 and 2. Then, the overhead information will be combined with the secret message.

Second, prediction error is generated by following equation:

$$PE_{k,i,j} = \left| F_{\left(k+\left\lfloor\frac{k}{2}\right\rfloor,i,j\right)} - F_{\left(k+\left\lfloor\frac{k}{2}\right\rfloor+1,i,j\right)} \right| \tag{6}$$

where PE is Prediction Error, F is Frames, k is sequence number, i and j are coordinate of pixel. After all prediction error has been generated, the histogram is created from all prediction error that is generated. Actually, histogram could be created by two ways. The first way is created from PE individually (one by one), while the second way is accumulated from all PE. In this paper, we choose the second way to generate the histogram.

Third, the peak point is chosen from the histogram and the histogram is shifted to provide an empty space on the right side of the peak point. Then, the overhead information is generated and combined in the secret message. Secret message embedding process is done by Eq. (2) like histogram shifting and neighbouring similarity methods. Lastly, modification of the prediction error is applied to all pixels in the frame. For each odd prediction error, the pixel changes will be applied into $F_{\left(i+\left\lfloor\frac{i}{2}\right\rfloor\right)}$ frame. For each even prediction error, the pixel changes are applied into $F_{\left(i+\left\lfloor\frac{i}{2}\right\rfloor+1\right)}$. This process can be done by following equation:

$$\begin{aligned} x_{i,j} &= x_{i,j} + \left(\alpha_{i,j} - \beta_{i,j}\right), if\ x_{i,j} > y_{i,j} \\ x_{i,j} &= x_{i,j} - \left(\alpha_{i,j} - \beta_{i,j}\right), if\ x_{i,j} < y_{i,j} \end{aligned} \tag{7}$$

where x is pixel in encoded frame, y is pixel in reference frame, α is prediction error after embedding process, and β is prediction error before embedding process. Illustration of embedding process is depicted in Fig. 4.

B. *Extraction Process*

First, the prediction error is generated using Eq. (5) like the embedding process. But in extraction process, histogram shifting in frame level is not needed. Additionally, amount of prediction error generated in extraction process will be same with embedding process. After all prediction error has been generated, the histogram is created with Eq. (6).

Second, all of prediction error values are scanned. The secret message is extracted with this following equation:

$$\begin{aligned} b_l &= 0, if\ x_{k,i,j} = p \\ b_l &= 1, if\ x_{k,i,j} = p+1 \end{aligned} \tag{8}$$

where b_l is l-bit secret message. Then, the overhead information is extracted from the secret message.

To recover the cover data, prediction error is shifted by using Eq. (9).

$$\begin{aligned} PE_{k,i,j} &= PE_{k,i,j} - 1, if\ PE_{k,i,j} > p\ and\ PE_{k,i,j} < 255 \\ PE_{k,i,j} &= PE_{k,i,j} - \theta_h, if\ PE_{k,i,j} = 255 \end{aligned} \tag{9}$$

where θ_h is h-overhead information. After this process, stego data will be back to shifted cover data. Shifted cover data is original cover data that shifted with Eq. (5) to prevent underflow / overflow problem. Lastly, the original cover data is recovered from the shifted cover data with Tsai et al. in [4] method by this following equation:

$$\begin{aligned} x_{i,j} &= x_{i,j} + 1, if\ x_{i,j} < z_1 \\ x_{i,j} &= x_{i,j} - 1, if\ x_{i,j} > z_2 \end{aligned} \tag{10}$$

The extraction process is shown in Fig. 5.

IV. Experimental Result

Ten videos from Xiph [10] are used in this experiment to evaluate performance of proposed method. That videos are used because related work in [5] use the videos too. Most of the videos are choosen in qcif resolution (176 x 144 pixel) and all video sequences are formatted in YUV4MPEG by using

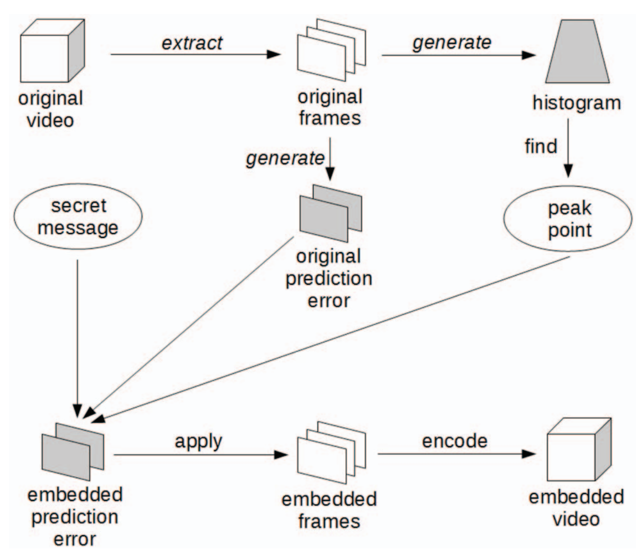

Fig. 4. Illustration of embedding process in proposed method.

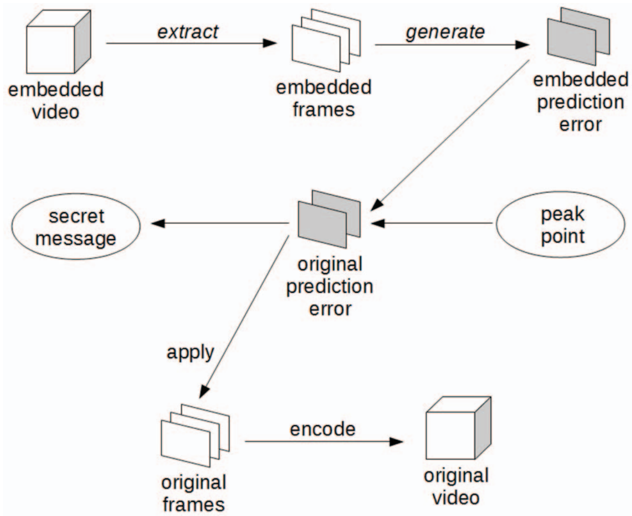

Fig. 5. Illustration of extraction process in proposed method.

mjpegtools. To make simplify and ensure the videos is able to encoded without any compression, we reformat all sample videos to Matroska Video (.mkv) format using mjpegtools in FFmpeg [11].

Before starting the steganography process, all frames in each video will be extracted using FFmpeg first. Extracted frames are formatted in Portable Network Graphics (.png). Then, each frame is converted to grayscale scheme and re-encoding using FFmpeg - mjpeg codec with Matroska Video format.

We are using lorem ipsum text for the secret message and also implements neighbouring similarity method for comparative. Lorem ipsum is used because it is a popular replacement text that is used in DTP (Desktop Publishing) industry [12]. Whereas neighbouring similarity is used because this method is a basic concept of our proposed method.

Sample frames of each cover and stego data are depicted in Fig. 6, while the statistical results are shown in Table 1. Both neighbouring similarity and proposed method in Fig. 6 looks similar with original frame if we use naked eyes for compare that. We cannot see difference about original frame, frame after neighbouring similarity embedding process, and frame after proposed embedding process. In statistical results that shown in Table 1, we know if the proposed method is able to increase the capacity and preserve PSNR above 40 dB. Average of capacity increment is 33.94% and average of PSNR value is 50.82 dB.

Capacity is the maximum bit of secret message that can be embedded into cover data, while PSNR is a method to measure the quality of stego data. Capacity measurement is performed by reducing the peak point value with a length of overhead data. For calculating PSNR, we are using feature in FFmpeg and take the average value. Greater value of PSNR is indicating better quality because the stego data is more similar with the cover data. The stego data can be categorized into high-quality if the PSNR value is 40 dB and above [13].

(a) (b) (c) (a) (b) (c)

Fig. 6. First frame at each dataset; (a) original, (b) neighbouring similarity, (c) proposed method.

TABLE 1.
COMPARISON OF NEIGHBOURING SIMILARITY AND PROPOSED METHODS.

Cover Data	Neighbouring Similarity		Proposed		Increment	
	Capacity (bit)	*PSNR (dB)*	*Capacity (bit)*	*PSNR (dB)*	*Capacity*	*PSNR*
bridge_close_qcif	6,792,960	52.28	8,880,727	51.93	30.73%	-0.35
bus_qcif_15fps	62,470	52.05	85,082	51.69	36.20%	-0.36
coastguard_qcif_mono	759,887	43.63	1,011,318	43.33	33.09%	-0.30
flower_cif	4,834,512	52.27	6,435,679	52.03	33.12%	-0.24
football_qcif_15fps	145,730	52.19	191,987	51.85	31.74%	-0.34
foreman_qcif_mono	1,048,019	42.31	1,386,415	41.98	32.29%	-0.33
garden_sif	388,587	52.12	530,436	51.83	36.50%	-0.29
hall_monitor_qcif	1,019,010	52.56	1,359,313	52.29	33.40%	-0.27
highway_qcif	5,295,302	52.18	7,356,312	51.85	38.92%	-0.33
mother_daughter_qcif	1,406,333	60.44	1,876,178	59.41	33.41%	-1.03

Table 1 have some attribute, that is cover data, neighbouring similarity, proposed, and increment. Cover data means name of the video that used for embedding process. Neighbouring similarity means the experimental result values for neighbouring similarity method. Proposed means experimental result values for proposed method of this paper. Increment means comparison of proposed method and neighbouring similarity method. Both neighbouring similarity, proposed and increment, we separate into two columns respectively. First column is capacity that means maximum length (in bit) of secret message that can be embedded into cover data. Second column is PSNR that show the result of quality measurement with PSNR method in FFmpeg tools.

Both proposed and existing methods, the secret message and the cover data is able to recover completely. We are checked the recovery process by using PSNR. All of recovered data are producing infinity value, that indicate if no noise is founded in recovered data. Additionally, we can extend the capacity of secret message by adding other value on histogram to embed a secret message like neighbouring similarity.

V. CONCLUSION

In this paper, a reversible video data hiding based on neighbouring similarity is presented. The proposed method is useful to increase maximum capacity that can be produced on neighbouring similarity. Additionally, it only decreases a small number of quality.

In the future, we will try to use overlapping scheme for neighbouring similarity. With overlapping scheme, the prediction error that generated will be greater than our proposed method in this paper, but still preserve similarity in prediction error generation process. So, the capacity could be increased because greater prediction error is greater value of peak point. We also want to try other scheme to modify the generation process of prediction error. Modification is expected to increase capacity or quality of stego data.

REFERENCES

[1] M. S. Subhedar and V. H. Mankar, "Current status and key issues in image steganography: A survey," *COMPUTER SCIENCE REVIEW,* Vols. 13-14, pp. 95-113, 2014.

[2] M. Holil and T. Ahmad, "Secret Data Hiding by Optimizing General Smoothness Difference Expansion-Based Method," *Journal of Theoretical and Applied Information Technology,* vol. 72, no. 2, pp. 155-163, 2015.

[3] Z. Ni, Y.-Q. Shi, N. Ansari and W. Su, "Reversible Data Hiding," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,* vol. 16, no. 3, pp. 354-362, 2006.

[4] P. Tsai, Y.-C. Hu and H.-L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Processing,* vol. 89, pp. 1129-1143, 2009.

[5] H.-L. Yeh, S.-T. Gue, P. Tsai and W.-K. Shih, "Reversible video data hiding using neighbouring similarity," *IET Signal Processing,* vol. 8, no. 6, pp. 579-587, 2014.

[6] J. Tian, "Reversible Data Embedding Using a Difference Expansion," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,* vol. 13, no. 8, pp. 890-896, 2003.

[7] A. M. Alattar, "Reversible watermark using difference expansion of triplets," in *Image Processing*, 2003.

[8] A. M. Alattar, "Reversible watermark using difference expansion of quads," in *Acoustics, Speech, and Signal Processing*, 2004.

[9] T. Ahmad, M. Holil, W. Wibisono and R. Muslim I, "An Improved Quad and RDE-based Medical Data Hiding Method," in *Computational Intelligence and Cybernetics (CYBERNETICSCOM)*, Yogyakarta, 2013.

[10] "Xiph.org :: Derf's Test Media Collection," Xiph, 14 May 1999. [Online]. Available: https://media.xiph.org/video/derf/. [Accessed 10 January 2016].

[11] F. Bellard, "FFmpeg," [Online]. Available: https://www.ffmpeg.org/. [Accessed 10 January 2016].

[12] "Lorem ipsum - download text and word document examples," Signale, [Online]. Available: http://www.loremipsum.de/download_lorem_ipsum.html. [Accessed 17 February 2016].

[13] A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, "Digital Image Steganography: Survey and Analysis of Current Methods," *Signal Processing,* vol. 90, no. 3, pp. 727-752, 2010.