Low-complexity differential modulation for high mobility MIMO-OFDM

Dhoni Putra Setiawan\textsuperscript{a}) and Hua-An Zhao

Graduate School of Science and Technology, Kumamoto University,
Kurokami, Chuo-ku, Kumamoto-shi, Kumamoto 860–0862, Japan
\textsuperscript{a}) 166d9304@st.kumamoto-u.ac.jp

Abstract: Differential Modulation is a non-coherent modulation technique which is used to solve the uncertainty channel problem. The most popular Differential Unitary Modulation technique uses a Space-Time Block Code (STBC) to improve the performance of Differential Modulation technique in multi-in multi-out (MIMO) system. However, using Differential Unitary Modulation technique in the multicarrier system become a big challenge because of the rising of the system complexity. In this paper, we introduce a low complexity differential modulation technique using the property of STBC. The proposed method is proven to provide high-quality performance with a low level of complexity.

Keywords: differential modulation, MIMO, OFDM, complexity

Classification: Wireless Communication Technologies

References


1 Introduction

High mobility communication until now still become an interesting research topic in the wireless communication. The most important problem which is occurred in high mobility condition is that the channel condition changes rapidly. This condition affects the accuracy of channel estimation while increasing the number of pilots is not a good option because it decreases the spectral efficiency. Non-coherent transmission system which does not employ the channel estimation is one of the solutions which is proposed to overcome this problem. Differential Modulation (DM) is the most interesting technique to be developed in non-coherent transmission system research. The differential modulation research aims to develop a good communication system which can be implemented even in the unpredicted channel condition.

In the era of the MIMO antenna system, the combination between DM technique and MIMO has produced some DM techniques [1, 2, 3, 4]. One of the most potential technique is called Differential Unitary Modulation (DUM) Technique which uses STBC Alamouti [5] as a Unitary Matrix [1, 2, 3]. In the papers [2, 3], Tran et al. introduce how to implement the DM technique in one of the multicarrier systems, Orthogonal Frequency Division Multiplexing (OFDM). This technique is called Unitary Differential Space-Time Frequency Modulation (UDSTFM). To arrange the symbols into Unitary Matrix, Tran et al. use diagonal matrix constellation. However, this diagonal constellation technique has a major problem, it has a very high system complexity and the worst part is the complexity increases significantly along with the increase of subcarriers. High complexity can be mean a lot, it may increase the power usage, the cost, and has a longer processing time.

In this paper, we propose a set of encoding and decoding of DM for MIMO OFDM. We use the property of STBC to create this technique, and we will show the significant difference between our method and DUM in term of complexity. We believe that our proposed method can be implemented widely in the other DUM systems to solve the complexity problem of the other systems. This paper is arranged into the following section; 1. Introduction, 2. System Models, 3. Performance Evaluation, and 4. Conclusion.

Notations: In this paper, the superscript $(\cdot)^*$ and $(\cdot)^\top$ denotes the complex conjugate operation and transpose operation respectively. We denote the notation “$\circ$” as the notation of element-wise (Hadamard) product multiplication.

2 System models

In this paper, we propose a set of Differential Modulation encoding and decoding schemes which transmit and receive the message via $2 \times 2$ MIMO OFDM system. The concept of our proposed technique is to change the multiplication between two unitary matrices into element-wise multiplication. This idea is proposed for minimizing the system complexity of Differential Unitary Modulation technique.

The original concept of DUM is shown in the equation (1). The symbols $a, b, c, d$ are random symbols outputs of the modulator. The symbols are paired in two different STBC Alamouti matrices.
The result of equation (1) is the same as the equation (2) below:

\[
\begin{bmatrix}
  a & b \\
  -b^* & a^*
\end{bmatrix}
\begin{bmatrix}
  c & d \\
  -d^* & e^*
\end{bmatrix}
= 
\begin{bmatrix}
  ac - bd^* & ad + bc^* \\
  -cb^* - a^* d^* & a^* e^* - db^*
\end{bmatrix}
\]

The element-wise calculation concept in equation (2) will be the basic concept of our proposed DM calculation process.

2.1 Encoding

As mentioned in the previous section, the proposed model uses the property of STBC Alamouti [5] to encode the symbols which come out from the PSK Modulator. The STBC Matrix can be written as the following equation:

\[
C_t = \frac{1}{\sqrt{2}} \begin{bmatrix}
  C_{t,1} & C_{t,2} \\
  -C_{t,2}^* & C_{t,1}^*
\end{bmatrix}
\]

\(C_{t,m}\) is a transpose of column matrix with the size \((N_{\text{fft}} \times 1)\), where \(N_{\text{fft}}\) is the total number of the subcarrier. The notation \(t\) represents the time of transmission and \(m\) is the number of antenna. The \(C_{t,m}\) consists of \(c_{t,m,k}\) which is the output of the PSK modulator. Notation \(k\) represents the subcarrier’s number. Then, the \(C_{t,m}\) can be written as: \(C_{t,m} = [c_{t,m,1}, c_{t,m,2}, \ldots c_{t,m,k}]^T\). Therefore, \(C_t\) is a matrix size \((2N_{\text{fft}} \times 2)\). Equation (3) is similar to the concept with UDSTFM which is proposed by LC Tran et al. in [2, 3]. However, there is a major difference between UDSTFM and our proposed system, in UDSTFM, the \(C_{t,m}\) is made into a diagonal matrix which made the matrix \(C_t\) become a matrix with the size \((2N_{\text{fft}} \times 2N_{\text{fft}})\), while in our proposed method this diagonalization process is unnecessary.

We consider \(X_t\) as the symbol of the transmitted matrix, which is the results of the DM calculation. For the first transmission, the transmitted matrix is \(X_t = C_t\) \((X_1 = C_1)\). For the next transmissions (from \(t = 2\)), we consider \(Ca_t\) and \(Cb_t\) as the encoding matrices of the proposed DM.

\[
Ca_t = \frac{1}{\sqrt{2}} \begin{bmatrix}
  C_{t,1} & C_{t,1}^* \\
  C_{t,1}^* & C_{t,1}
\end{bmatrix}
\]

\[
Cb_t = \frac{1}{\sqrt{2}} \begin{bmatrix}
  -C_{t,2}^* & C_{t,2} \\
  -C_{t,2} & C_{t,2}^*
\end{bmatrix}
\]

The multiplication between two Unitary STBC Alamouti results in another Unitary Matrix. Therefore, for our proposed system, this property also can be applied. Thus, the transmitted matrix \(X_t\) can be written as:

\[
X_t = \begin{bmatrix}
  X_{t,1} & X_{t,2} \\
  -X_{t,2}^* & X_{t,1}^*
\end{bmatrix}
\]

To encode the symbols, we need a reverse matrix of \(X_t\) which swaps the symbols between column 1 and column 2. We call this matrix as \(X_{r,t}\).
And finally, the differential calculation of the proposed technique can be written as:

\[ X_t = X_{t-1} \circ C_{A_t} + X_{r_{t-1}} \circ C_{B_t} \]  

(7)

The elements of \( X_t \) then will be transmitted by two antennas after Inverse Fast Fourier Transform (IFFT) process. The transmission model is written as:

\[ Y_t = X_t \circ H_t + N \]

(8)

Where, \( Y_t \) is the received matrix after Fast Fourier Transform (FFT), \( H_t \) is the channel coefficient and \( N \) is the Noise coefficient. Identical with the transmitter, in the receiver two antennas also employed.

### 2.2 Decoding

To decode the received matrix \( Y_t \), we consider \( Y_{a_t} \) and \( Y_{b_t} \) to be used as decoding matrices. We also consider \( \hat{C}_t \) as the result of the decoding process. For the first received transmission \( Y = \hat{C}_t \) (\( Y_1 = \hat{C}_1 \)).

\[ Y_t = \begin{bmatrix} X_{t,1} \circ H_{t,1} + N & X_{t,2} \circ H_{t,2} + N \\ -X_{r,t,2}^* \circ H_{t,3} + N & X_{r,t,1}^* \circ H_{t,4} + N \end{bmatrix} \]

(9)

For a better understanding of our proposed decoding, \( Y_t \) also can be written as the following:

\[ Y_t = \begin{bmatrix} Y_{t(1,1)} & Y_{t(1,2)} \\ Y_{t(2,1)} & Y_{t(2,2)} \end{bmatrix} \]

(10)

Where, in the \( Y_{t(p,q)} \), \( p \) and \( q \) represent the number of row and column of the matrix \( Y_t \), respectively. As has been mentioned in the previous subsection, one symbol in this matrix represent a set of matrix size (\( N_{fft} \times 1 \)).

\[ Y_{a_t} = \begin{bmatrix} Y_{a_{t(1,1)}} & Y_{a_{t(1,2)}} \\ Y_{a_{t(2,1)}} & Y_{a_{t(2,2)}} \end{bmatrix} \]

\[ Y_{b_t} = \begin{bmatrix} Y_{b_{t(1,1)}} & Y_{b_{t(1,2)}} \\ Y_{b_{t(2,1)}} & Y_{b_{t(2,2)}} \end{bmatrix} \]

(11)

We also need the reverse matrix of \( Y_t \) which swaps the symbols between rows of \( Y_t \) we call it \( Y_{r_t} \):

\[ Y_{r_t} = \begin{bmatrix} Y_{r_{t(1,1)}} & Y_{r_{t(1,2)}} \\ Y_{r_{t(2,1)}} & Y_{r_{t(2,2)}} \end{bmatrix} \]

(12)

And finally, the decoding process can be written as:

\[ \hat{C}_t = Y_t \circ Y_{a_{t-1}} + Y_{r_t} \circ Y_{b_{t-1}} \]

(13)
Because $\hat{C}_t$ is the received $C_t$. We can write it into this following form:

$$\hat{C}_t = \begin{bmatrix} \hat{C}_{t,1} & \hat{C}_{t,2} \\ -\hat{C}^*_{t,2} & \hat{C}^*_{t,1} \end{bmatrix}$$  \hspace{1cm} (14)

In order to take the advantage of diversity, $\hat{C}_{t,m}$ which is the received form of $C_{t,m}$ can be calculated with the following equations:

$$\hat{C}_{t,1} = \frac{1}{2} (\hat{C}_{t(1,1)} + \hat{C}^*_{t(2,2)})$$

$$\hat{C}_{t,2} = \frac{1}{2} (\hat{C}_{t(1,2)} - \hat{C}^*_{t(2,1)})$$  \hspace{1cm} (15)

The received PSK symbol $\hat{c}_{t,m,k}$, which is the element of $\hat{C}_{t,m}$ can be calculated with the following maximum likelihood equation:

$$\hat{c}_{t,m,k} = \arg\min_{c_{t,m,k} \in \mathbb{C}} \{(\Re |\hat{c}_{t,m,k} - c_{t,m,k}|)^2 + (\Im |\hat{c}_{t,m,k} - c_{t,m,k}|)^2\}$$  \hspace{1cm} (16)

3 Performance evaluation

3.1 Complexity analysis

As mentioned in the two previous sections, the goal of this research is to solve the complexity problems of DUM technique in the MIMO OFDM. In this section, we show how much the improvement of the proposed system compared to the UDSTFM [2, 3].

If we use Big O notation for calculating the system complexity, the complexity of calculation between square matrices is $O(n^3)$, where $n$ is the number of row (or column). Several new ways to multiply two matrices have been proposed, with the lowest complexity is $O(n^{2.373})$ [6]. However, in this paper we choose to not to use the Big O, instead, we calculate the real complexity of our proposed model compared to DUM model.

As a comparison, we calculate the complexity of UDSTFM in the paper [2, 3]. Using conventional matrices multiplication, we can find that the number of multiplications is $(2N_{fft})^3$ and the number of additions is $(2N_{fft})^3 - (2N_{fft})^2$. If we include the total number of differential calculations which have to be done until all data decoded, then the number of multiplications is $r(2N_{fft})^3$ and the number additions is $r((2N_{fft})^3 - (2N_{fft})^2)$, where $r$ represents the number of differential calculation processes.

In the element-wise multiplication, the number of steps is the same with the total number of the elements of the matrix. In this proposed system we have two element-wise multiplication processes in one differential calculation process, in which each matrix size is $(2N_{fft} \times 2)$. Therefore, the number of multiplications can be written as $2r(4N_{fft})$ and the number of additions is $r(4N_{fft})$. The complexity decreases significantly because it removes the unnecessary multiplications between zeros which happens in the UDSTFM because of the diagonalization process.

Based on the explanations above, we can understand that the complexity of DUM will increase exponentially along with the increase of subcarrier, while in the proposed scheme, it increases arithmetically. Moreover, another advantage of using our proposed scheme is if we transmit the same number of symbols using DUM, a system which employs a higher number of subcarriers has higher complexity.
However, with our proposed scheme the complexity is decreased. This condition happens because when the system doubles the number of subcarriers, the number of $r$ becomes $\frac{r}{2}$ which results in lower complexity.

3.2 Experimental result

We simulate our proposed scheme and compare it with UDSTFM scheme which is proposed by Tran et al. in [2, 3]. The parameters including: the carrier (2.6 GHz), subcarrier spacing (15 kHz), the number of the subcarrier (128), the velocities (200 k/h and 300 k/h), the modulator (QPSK), and the channel model (Rayleigh Fading Channel). Both schemes are simulated in the exactly same channel condition and added by the same additive noise.

![Fig. 1. UDSTFM vs. Proposed model](image)

As shown in Fig. 1, there is no single difference in term of performance between UDSTFM [2, 3] model and our proposed model. Beside the performance comparison, we also calculate the running time of both systems from the encoding process until finish the decoding process when using 128 and 256 subcarriers. By using 128 subcarriers, the average running time of UDSTFM [2, 3] is 5.5 times longer than our proposed model, while 256 subcarriers results in the average running time of UDSTFM [2, 3] 15.6 times longer than our proposed model.

4 Conclusion

In this paper, a low complexity differential modulation scheme is proposed. The experimental result shows the proposed low complexity differential modulation technique successfully matches the performance of UDSTFM technique, but with lower complexity. The complexity reduction can be achieved by removing the diagonalization process in UDSTFM and change the original matrix multiplication into the proposed element-wise multiplication. We believe that this concept also can be implemented into the other DUM techniques for multicarrier system in order to minimize their system complexity.
Novel throughput evaluation method of radio LAN systems based on occupied duration for individual UDP transmission

Kenji Kita\textsuperscript{1a)}, Masato Uchida\textsuperscript{2}, Hiroyasu Ishikawa\textsuperscript{3}, and Hideyuki Shinonaga\textsuperscript{1,2}

\textsuperscript{1} Faculty of Science and Engineering, Toyo University, 2100 Kujirai, Kawagoe, Saitama 350–8585, Japan
\textsuperscript{2} Graduate School of Science and Engineering, Toyo University, 2100 Kujirai, Kawagoe, Saitama 350–8585, Japan
\textsuperscript{3} College of Engineering, Nihon University, 1 Nakagawara, Tokusada, Tamuramachi, Koriyama, Fukushima 963–8642, Japan

\textsuperscript{a)} kita@toyo.jp

Abstract: The instantaneous throughput is evaluated based on the communication traffic volume per fixed unit time. Although the conventional method can roughly evaluate it within unit time, it is difficult to evaluate the communication situation of each packet in detail. Therefore, the authors propose a method of deriving the throughput based on the occupied duration for one packet transmission. In the proposed method, not only throughput per packet but also throughput of other signals such as beacon signals and variation of Back-off Time can be evaluated. In this paper, we define the occupation duration for deriving throughput during individual UDP transmission, and show analysis examples compared with the conventional method.

Keywords: radio LAN, occupied duration, instantaneous throughput, UDP, packet capture analysis

Classification: Network

References

1 Introduction

In recent years, demands for Radio LAN communications have increased, and communication quality is frequently evaluated. The instantaneous throughput, evaluating the throughput per unit time, is especially used to observe the communication situation and used for communication quality evaluation. However, there are cases where the throughput is not accurately evaluated by the conventional method. Therefore, the authors propose a throughput evaluation method based on an occupied duration for one packet transmission. In this paper as the simplest case, we define in various cases for individual UDP transmission in 802.11g mode, and examples applied to actual captured data are indicated.

2 Definition of throughput based on occupied duration in individual UDP transmission

The proposed throughput based on the occupied duration in individual UDP transmission is defined in detail in various cases.

2.1 Conventional instantaneous throughput

Fig. 1(a) shows the definition of the conventional instantaneous throughput. It is represented by how much data was captured per unit time. Since the throughput is based on the fixed duration, we call it Fixed Duration (FD) throughput in this paper. For example, if $N$ packets of 1500 bytes exist in unit time $t$ seconds, FD throughput is obtained by the following equation [1]:

$$
FD \text{ throughput (Mbit/s)} = \frac{1500 \times 8 \times N \times 10^{-6}}{t}.
$$

(1)

Packet capture devices print a timestamp at the time of receiving the first or last bit. Therefore, in many analysis software, FD throughput is derived by the all information amount in which the timestamp exists in the unit time. So that means the shorter the unit time, the greater the influence of one packet. The number of packets included within a short unit time is determined for each transmission rate, and since the throughput is evaluated based on the number, the quantized result appears.

2.2 Definition of occupied duration for the proposed throughput

The authors propose a throughput evaluation method using Occupied Duration which is a duration when one packet occupies the communication path. Fig. 1(b) shows the its definition. 802.11 Data Frame (Data Frame) is one Radio LAN packet. 802.11 ACK Frame (ACK Frame) is the acknowledgement packet of the physical layer transmitted by a network interface card of a receiving device. Occupied Duration is defined by the sum of DIFS, Back-off Time with dynamic values by CSMA/CA, SIFS and Frame Length shown in the Fig. 1(b). Briefly, it is
from the last bit reception time of the previous ACK Frame to that time of ACK Frame of the target packet. Data Frame contains two physical layer headers of 802.11 and Logical Link Control headers [1], but they are excluded from Data Frame and let it be Payload. Since the proposed throughput is based on Occupied Duration for one packet transmission, it is called Occupied Duration (OD) throughput distinguished from FD throughput. OD throughput is obtained by the following equation:

\[
\text{OD throughput (Mbit/s)} = \frac{\text{Payload}}{\text{Occupied Duration}}.
\]  

Since OD throughput is derived based on Occupied Duration, it can be analyzed that the difference in the throughput at the same transmission rate is the difference in Back-off Time changing dynamically.

Fig. 1. Definition of occupied duration.
2.3 Definition of occupied duration when other signals exist

Unit time used in the conventional method includes signal transmission durations other than Data Frame such as beacon signals. Therefore, FD throughput sometimes can not indicate only Data Frame. OD throughput is derived more accurately than FD throughput by considering the other signal transmission durations. Fig. 1(c) shows the definition of Occupied Duration when other signals exist. Signal is other signal such as beacon, and it usually does not have ACK Frame. In this case, the transmission control start time can not be specified. Therefore, it is necessary to define the start time of the Occupied Duration after Signal. Back-off Time by CSMA/CA differs each packet and it is difficult to measure its value, for that reason, it is employed its average value in this case. Hereat, there may be Idle Time that is not communicated between the capture completion time of Signal and the start time of Occupied Duration. In addition, as shown in Fig. 1(d), when the start time of Occupied Duration is before capture completion time of Signal, the start time is set as capture completion time of Signal. Furthermore, Signal’s OD throughput which was impossible by the conventional method can be derived. Signal’s Occupied Duration is from the capture completion time of the preceding ACK Frame to that of the Signal, and the information volume of the Signal is all bits.

2.4 Definition of occupied duration at retransmission occurrence

During communications, ACK Frame may not be returned and physical layer retransmissions may occur. Fig. 1(e) shows Occupied Duration in the case that physical layer retransmissions occur twice as an example. In this case, the Occupied Duration is the time interval from the preceding ACK Frame to the ACK Frame of the retransmission packet normally transmitted, and the information volume is one packet. If there are many physical layer retransmissions for a specific packet, the Occupied Duration increases and OD throughput decreases.

With the above detailed definition, comparisons of FD and OD throughput is summarized in Fig. 1(f), and OD throughput shows that communication situations which can not be analyzed by FD throughput can be visualized.

3 Application examples in individual UDP transmission

3.1 Experiment system

Fig. 2 shows an experiment system. UDP packets are transmitted from the Tx Personal Computer (PC) on which Ubuntu 14.04 [2] is installed to Rx Smartphone via the access point. iPerf 3 for iPhone [3] is installed in the Rx Smartphone. This can set a transmission data amount of UDP packets per unit time at transmitting devices, preset UDP transmission rate is set by 30 Mbit/s in this experiment. We employed iPhone 6 as the Rx Smartphone, the iOS was version 9.3.4, iPerf 3 for iPhone was version 3.0.9. In the wireless section, the Capture PC connected with AirPcap Nx [4] to capture the packets. Wireshark [5] is used for packet analysis. Occupied Duration and the throughputs were obtained from this capture data. Experiments were conducted at a laboratory of Toyo University.
3.2 Comparison between proposed and conventional method

Figs. 3(a) and (b) are examples in the other signals and the retransmission existences respectively. Both show the results of the proposed, conventional method unit time 10 msec and 1 msec in order from the top. The horizontal axis is time and the vertical axis is the FD or OD throughput. Black lines are the throughput of Data Frame, and reds are that of Signals. First, we discuss Fig. 3(a) in the other signals existence. Since OD throughput in the proposed method at the top is derived based on Occupied Duration per packet, the area of the rectangle indicates the communication traffic volume and the width indicates Occupied Duration. In other words, OD throughput becomes higher as Occupied Duration is shorter at the same transmission rate and Payload. In addition, if the transmission rate is the same, the Back-off Time dynamically changing can be evaluated by the difference in OD throughput. Furthermore, the Signal is visualized and Idle Time is confirmed between the Signal and the Data Frame. The middle at unit time 10 msec shows the change of the FD throughput is roughly understood, but compared to the proposed method, it is not shown in packet unit. The bottom at unit time 1 msec shows the FD throughput is shown only four values of 0, 12, 24 and 36 Mbit/s. Considering one packet transmission duration based on the 802.11g mode transmission rate at 54 Mbit/s, three packets are transmitted at most in 1 msec as described in section 2.1. Therefore, it is evaluated with four values. Next, we discuss Fig. 3(b) in the retransmission existence. As the result of the packet capture analysis, there was some influence on the transmission path in this section, and the transmission rate was 36 Mbit/s. The top in proposed method, the retransmission occurred once in the duration where the OD throughput was around 10 Mbit/s, and occurred twice around 5 Mbit/s. In the middle and bottom, FD throughput is declining, but it is difficult to analyze the communication situation such as the number of retransmissions. Particularly, in the vicinity of 31.892 and 31.895 seconds at unit time 1 msec, the FD throughput is 0 because one packet was transmitted across the unit time, since the OD throughput is obtained with the Occupied Duration, it is possible to accurately evaluate the communication situation such that retransmission occurs and it takes time to transmit one packet. In this way, OD throughput can be accurately analyzed even when other signals or retransmission occurs, and Back-off Time and communication situations can be estimated.
4 Conclusion

We proposed OD throughput based on Occupied Duration for one packet transmission in individual UDP Radio LAN transmission. Detailed analysis of other signals and retransmission existence that was difficult with the conventional method was facilitated by the proposed method. As further study, we will extent to 802.11n or ac that needs to consider frame aggregation etc., TCP and competitive transmissions.

Fig. 3. Application and analysis example.
A video quality improvement technique for a fast combined watermark embedding method

Hajime Matsunaga¹, Tomoko Sawabe², and Masami Kihara²a)

¹ Graduate School of Science and Engineering, Nihon University,
7–24–1 Narashinodai, Funabashi 274–8501, Japan
² Department of Computer Engineering, College of Science and Technology, Nihon University, 7–24–1 Narashinodai, Funabashi 274–8501, Japan
a) kihara.masami@nihon-u.ac.jp

Abstract: The use of watermarks to embed user information is expected to deter the unauthorized use of video content in VOD (Video On-Demand) streaming services. This requires that watermark embedding and video encoding for each video must be finished so rapidly as to not create excessive VOD delay between the user’s video request and video download to the user. This paper describes a watermark embedding and encoding process for a fast combined watermark embedding method, and proposes a new encoding technique that minimizes video quality degradation.

Keywords: VOD service, video, watermark embedding, encoding

Classification: Multimedia Systems for Communications

References

1 Introduction

VOD (Video on-demand) services such as dTV, hulu, Netflix and Amazon Video are increasing. These services provide movies, dramas, and TV programs, and interest is growing in 4K UHD (Ultra High Definition) content with the high resolution of 3840 x 2160. While the encryption systems such as HDCP (High-bandwidth Digital Content Protection system) used in HDMI (High Definition Multimedia Interface) and the new DRM (Digital Rights Management) mechanism called a CDM (Content Decryption Module) solutions such as Google Widevine, Apple Fair Play and Microsoft PlayReady in video service systems attempt to prevent the unauthorized usage of video contents, the illegal upload and download of video content will not disappear. To supplement existing DRM techniques, we can use watermark technology to invisibly embed user information into the video content. The embedded user information can be extracted from the video content released illegally.

Unfortunately, the real-time processing needed to embed watermarks into content demands high performance processors [1] or specialized custom LSIs [2]. The watermark combination method was developed to solve this problem [3, 4, 5]. The method well supports VOD services by minimizing the watermark embedding time and encoding time.

This paper proposes a fast watermark embedding method based on combining information that is embedded in the IDR (Instantaneous Decoder Refresh) frames of h.264. Moreover, it introduces an encoding technique that minimizes the degradation in video quality achieved by the proposed combined watermark embedding method.

2 Conventional watermark embedding

2.1 Basic watermark embedding method

The basic watermark method directly embeds user-specific information into video frames; the example of “101” is shown in Fig. 1(a-1). The video content is encoded after the user-specific information is embedded (Fig. 1(a-2)). The encoding is executed using the IDR (Instantaneous Decoder Refresh) frames of h.264 or GOP (Group of Picture) frames in MPEG-2. These units are called “closed units” in this paper. While this approach offers robust embedding of user-specific information, its time overhead is a barrier to adoption given the real-time processing requirements of VOD services.

2.2 Watermark combination method

The watermark combination method [3, 4, 5] was developed to shorten the watermark embedding and encoding time, i.e. the barrier described in Section 2.1. The watermark combination method does not directly embed the complete user-specific...
information but uses a sequence of basic identifiers. The user information of “101” can be specified by the sequence of the three basic identifiers “1”, “0” and “1” as shown in Fig. 1(b-1). The unit of the sequence as shown in Fig. 1(b-1) is called the combination unit in this paper. Basic identifiers are embedded in the head frames of the closed unit, called ID frames in this paper. In the example given, the combination unit is a sequence of 3 head frames, and the number of basic identifiers in each ID frame is 2, so this example can distinguish up to 8 users.

After a user request is received, the video data to be sent to the user is appropriately formed by simply selecting the corresponding stock sequential frames as shown in Fig. 1(b-1). This watermark embedding process is thus very fast.

![Diagram](https://via.placeholder.com/150)

(a-1) Decoded video sequence containing watermarked frames; conventional scheme

(a-2) Closed units in encoded video stream; conventional scheme

(b-1) Combined encoded video data in conventional watermark combination method

**Fig. 1.** Watermark embedding and encoding in conventional watermark combination method

### 3 Proposed combined watermark embedding and encoding method

Fig. 2 shows an example of the proposed method. The original video frames, that are decoded from compressed video data, are shown in Fig. 2(c-1). The ID frames are extracted from the decoded video frames in Fig. 2(c-1), and as many copies as there are basic identifiers are made. The basic identifiers are then embedded into the copied ID frames by conventional watermark software (Fig. 2(c-2)). In the other watermark embedding approach, all video frames are prepared from the decoded video frames, and basic identifiers are embedded to the ID frames as an initial watermark (indicated by “X” in Fig. 2(c-3)). The initial watermark embedding
operation is very important in determining the final video quality (the advantage of the initial watermark is described in Section 5).

Each ID frame with its basic identifier (Fig. 2(c-2)) is encoded one by one (Fig. 2(c-4)) to yield the IDR frames. All frames with the initial watermark (Fig. 2(c-3)) are also encoded based on the closed unit (Fig. 2(c-5)). This is performed in a preprocessing step before a user requests a video.

When a user request is received, the sequence of the encoded ID frames with the appropriate basic identifier is decided. The selected encoded ID frames with the basic identifier (“1”, “0” and “1” in Fig. 2(c-6)) and all encoded frames in closed units excluding ID frames are combined which yields the video file to be downloaded.
4 Processing time estimation

Two 4 second 4K video sequences, 30 frames per second and compressed to yield 50 Mb/s, are used to estimate the processing time of the proposed method. Since the video data reconfigured as shown in Fig. 2(c-6) is not held as sequential data in the storage media, random access is performed.

We estimate that the encoded ID frames and closed units excluding ID frames for 1 closed unit (30 frames for 1 second) have the read times of approximately 42 ms and 62.5 ms, respectively, and that the maximum total read time is 144.5 ms including average access time of 20 ms. Since this read time is shorter than the length of 1 closed unit, this implies that the video data assumed here can be smoothly played with an adequate (several seconds) video buffer.

In this estimation we used a HDD (Hard Disk Drive) as the storage medium. Its I/O data-transfer rate is 600 MB/s max, spindle speed is 5400 rpm and cache buffer is 64 MB.

On the other hand, in the conventional method as shown in Fig. 1(a-1), we have to consider that the watermark embedding time and the encoding time for 4K video data processing are around 3 seconds per frame using the conventional watermark software and 2 to 7 seconds for a closed unit (30 frames for 1 second), respectively. HDD read and write times are excluded.

These times mean that, without the proposed method, the user has to wait at least 5 to 10 seconds until the video data is played, even if the video has just one-second of screen time. This waiting time is seen as excessive in normal video services.

The computer used in the estimation and experiment had an Intel Core i7-3770 3.4 GHz processor and DDR3 32 GB memory with Windows 10 OS for the encoding and CentOS Linux 7.1 for watermark embedding. The encoding process fully uses the computer’s 4 cores and 8 threads.

5 Video quality evaluation

While the proposed method can eliminate the watermark embedding time and the encoding time, the reconfiguration demanded by inserting the encoded ID frames degrades the video quality. This section describes the results of a video quality evaluation when the proposed method is applied to 4K video.

The video quality in decoded frames was evaluated by the typical measures of SSIM (Structural Similarity) [6] and PSNR (Peak Signal-to-Noise Ratio) [7]. These measures were calculated by FFmpeg, a library of multimedia data transcoding functions [8].

Fig. 3 shows the results of a video quality evaluation using SSIM and PSNR. Two 4 second 4K video data (30 frames per second, 50 Mb/s) were used in this evaluation. Video-A shows a slow panning shot in a room with plain color walls. Video-B shows a landscape including cherry trees with many cherry blossom petals flying around. Video-A and Video-B were encoded with the h.264 codec with FFmpeg and their encoded data rate was around 50 Mb/s.

The characteristics of three encoding methods are compared in Fig. 3: the method with the initial watermark embedding technique described in Section 3
(blue points in Fig. 3), the method without initial watermark embedding (orange points in Fig. 3) and additionally the method that encodes all frames including all ID frames configuring the closed unit (gray points in Fig. 3). Horizontal axis in Fig. 3 is frame number. Frame numbers 1, 31, 61 and 91 are IDR frames.

![Fig. 3. Results of video quality evaluation](image)

Video data encoded with all frames (gray points) has the highest SSIM and PSNR in both Video-A and Video-B, since all frames including IDR frames are encoded. The method that encodes all frames is the best approach considering video quality, however the prepared ID frames, which equals all frames of the closed unit, generates and stores a large amount of data in the preprocessing stage. That is why we propose the method with the initial watermark; it can minimize the volume of the encoded ID frames including user-specific information that needs to be held in the video data storage media.

In the encoding method without initial watermark embedding, SSIM and PSNR are greatly degraded for about 10 frames from the IDR frame in both Video-A and Video-B. However, in the encoding method with initial watermark embedding, SSIM is improved to around 0.98 and an increase of about 1 to 4 dB in PSNR was attained. The proposed method, described in Section 3, is very effective in improving decoded video quality.

We also carried out a subjective assessment based on 15 participants making dichotomous choices. The test, which compared the method of encoding all frames with the method with the initial watermark, found no significant difference between the resulting videos, though it is obvious that the method without the initial watermark yielded worse quality than the method of encoding all frames. Qual-
itative assessments on a 4K display also confirmed that the proposed method created no perceptible degradation in video quality.

In comparisons with and without the initial watermark in Video-A, while 73 percent of the participants found the proposed method to be superior to the method without the initial watermark, no significant improvement appeared in the evaluation results for Video-B. While the proposed method is effective, the degree of video quality improvement varies with the video content.

6 Conclusion

This paper proposed a method for embedding watermarks into video data by the technique of embedding initial watermarks in ID frames. The method can minimize the video quality degradation triggered by the video frame reconfiguration that results from the insertion of ID frames in the combined watermark combination method. Tests on two 4K videos showed that the proposed technique is very effective.