IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 May 2023, accepted 15 June 2023, date of publication 20 June 2023, date of current version 26 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3288020

RESEARCH ARTICLE

A Reversible Multimedia Representation Method and Its Applications in Multimedia Processing

XIU HE[1,2], **YUYUN CHEN**[1], **YONG FAN**[1], **XIAOXI KONG**[1], **AND ZHANCHUAN CAI**[1], **(Senior Member, IEEE)**

[1]School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China
[2]School of Information and Intelligent Engineering, Guangzhou Xinhua University, Guangzhou (Dongguan) 523133, China

Corresponding author: Zhanchuan Cai (zccai@must.edu.mo)

This work was supported in part by the Science and Technology Development Fund of Macau under Grant 0059/2020/A2 and Grant 0052/2020/AFJ, in part by the Zhuhai Industry–University–Research Collaboration Program under Grant ZH22017002210011PWC, in part by the Guangzhou Science and Technology Project under Grant 202102080656, in part by the 2021 Ordinary University Key Research Platform and Projects of Guangdong under Grant 2021ZDZX1047, in part by the Key Discipline Project of Guangzhou Xinhua University under Grant 2020XZD02, in part by the Guangdong Key Discipline Scientific Research Capability Improvement Project under Grant 2021ZDJS144, and in part by the Guangdong Key Platform for University and Major Scientific Research Project under Grant 2018KQNCX361.

ABSTRACT The single type of multimedia data, such as text, image, video, and audio, has been extensively studied in the past decades. With the advances in multimedia applications and services, the exponential growth of multimedia data has emerged. Therefore, it is crucial to effectively represent multimedia data and capture the relationships among them in the study of multimedia applications. In the paper, we propose a novel reversible multimedia representation method based on a complex base. In the proposed multimedia representation method, we first process different modalities of data, such as text, image, and audio separately and convert them into different Gaussian integers, then place them on the same representation plane and resulting in a joint multimedia representation model. Therefore, the relationships of multimedia data are established by using the geometric operations on the same representation plane. Experimental results show that the proposed multimedia representation method is lossless and reversible. In addition, we explore the potential applications of the proposed representation method in multimedia camouflage and multimedia secret sharing, and verify that the proposed representation method has good practical applicability in multimedia security.

INDEX TERMS Multimedia representation, multimodal, complex base, multimedia camouflage, multimedia secret sharing.

I. INTRODUCTION

As for a computer system, the term multimedia refers to ‘‘a combination of multiple modalities including text, image, video, audio, etc., in such a way that can be accessed interactively’’ [1]. The unimodal data has been widely studied in the past few years. With the growth of low-cost devices, multimodal and heterogeneous multimedia data are increasing. Nowadays, multimedia has virtually become an integral part of our daily lives, which is used in ways as diverse as multimedia search, multimedia recognition, multimedia detection, artificial intelligence [2], [3], [4], [5], [6], [7], [8], and so on. How to represent multimedia data and capture the relationships among them is an important basic research in multimedia applications. Therefore, this paper focuses on multimedia representation, and explores the relationships between multimedia data.

Multimedia representation refers to describing the content of multimedia data which includes text, image, and audio in this study. In multimedia representation, information is often exchanged in a multimodal form. Multimodality can be regarded as a model that involves different modalities

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

of information (e.g., text, image, audio, video) to participate [9]. A model that is capable of handling multimodal information is better suited for multimedia interaction. As a bridge connecting multimodal information, the reversible multimedia representation method facilitates the multimedia applications, the representation and inter-conversion of multimodal data.

Therefore, it is absolutely imperative to construct a multimedia representation model. In digital computers, multimedia information is stored in the form of binary. From a mathematical point of view, the binary is a number system that uses integer 2 as a base. And a base is a natural number r, where the powers of r are used in combination to represent data in a numerical system [10]. There are lots of studies on bases. Plantz et al. [11] proposed an octal number representation system in which a base of 2^3 was adopted in place of the decimal and discussed the feasibility and some applications of the octal numerical system. Wadel [12] introduced a number system based on a negative base, and defined some rules of arithmetic operations, such as addition, subtraction, multiplication, and so on. However, the information represented in [11] and [12] are real numbers, which cannot perform geometric operations between different modal data. Dimitrov and Jullien [13] proposed a novel system that used two or more bases to represent data, and gave the applications in cryptography and digital filter implementation. However, because this system has redundant representations, the reversibility of the represented multimedia data cannot be guaranteed. Bergman [14] proposed to represent data with an irrational number as base($\frac{(1+\sqrt{5})}{2}$), and discussed some arithmetic operations. The disadvantage of this method was that it can only approximate the information to be represented infinitely, so it was not a lossless multimedia representation. Knuth and Donald [15] introduced another imaginary number system to represent data and explored potential use for manipulating complex numbers on digital computers. Heuberger et al. [16] proposed a novel positional number system, which was defined by a set of positive digits to form an arithmetic progression, and gave its application in a graph-theoretical. The systems proposed in [15] and [16] cannot represent all information, thus they were not suitable for representing multimedia information.

To break through the above limitations of representing multimedia data, we need to study a suitable multimedia representation model. Gilbert proved the uniqueness of using Gaussian integers as bases to represent the complex numbers [17]. Katai and Szabio [18] proved that Gaussian integer $-n+i$ or $-n-i$, using natural number $0, 1, 2, \ldots, n^2$ as digits, can be used as a base to represent all the Gaussian integers. Gilbert [19] further proved that all the complex numbers can be represented based on appropriate complex base. These studies in [17], [18], and [19] provided a theoretical standpoint for the application of complex bases, and we find that the representation method based on complex bases can solve the problems in [11], [12], [13], [14], [15], and [16]. Hence, we discuss the application of complex bases in multimedia representation. In [20], Sun et al. introduced a number system based on a complex base, but this complex base system was only applied to single modality data representation in digital watermarking. With the development of multimedia applications and services, such as the popular multimedia social software Facebook, WeChat, etc., the main types of data they generate are text, image, and audio, etc., this kind of multimedia data has the characteristics of heterogeneity and multimodality. Therefore, based on the study in [20], this paper focuses on seeking a unified information representation method to construct a multimodal processing model.

Additionally, we introduce two multimedia security applications to prove the practicability of the proposed multimedia representation method: multimedia camouflage and multimedia secret sharing.

Camouflage is a widely used information security technology [21]. As a quite prevalent embodiment of camouflage, reversible data hiding(RDH) has achieved well, but they also show a complex trade-off between quality and embedding capacity [22], [23], [24], [25], [26], [27], [28], [29], [30] and they are all pixel-level. With the increase of embedding capacity, the quality of the carrier becomes worse. Therefore, how to balance quality and embedding capacity is a major problem. In order to solve bottleneck of embedding capacity, multimedia camouflage scheme is designed, which employs "replacement" rather than "embedding" to camouflage secret messages.

As another important research field of multimedia security, secret sharing is to decompose the secret message into multiple shares and hand them over to participants for safekeeping, and it is stipulated that only the authorized subset of participants can be combined to recover secret information [31]. There are various secret sharing schemes [32], [33], [34], [35], [36], [37] and they have excellent performance. However, they can only share unimodal data. Diverse and comprehensive shares can greatly increase the difficulty of intercepting the shares, and reduce the transmission risks of shares. Hence, the paper proposes a multimedia secret sharing scheme (MSS) and explores the possibility of using a diverse and comprehensive multimedia data for sharing the secret message.

In this paper, the main contributions of this work can be summarized as follows:

- A novel and reversible multimedia representation method based on a complex base is proposed. This method supports different modalities of data. The lossless and reversible properties of the proposed method are verified through experiments on the original and reconstructed data.
- A novel representation relationship between multimedia data is proposed. Algebraic operations on the complex plane are applied to represent the relationships among multimedia data. Compared with the traditional elementary arithmetic of digital media, we have much richer

geometric operations, such as the addition and subtraction of vectors. In addition, we do not need to consider whether the data types involved in the operations are the same or whether the sizes are the same.

- The limitation of single modality data in traditional multimedia information security has been broken. The proposed method achieves the multimodal applications for the first time. Using the proposed representation method based on a complex base as the bridge, we realize the camouflage and secret sharing schemes of cross-modal data and build a unified multimedia information security model.

The rest of the paper is organized as follows. We discuss the proposed multimedia representation method based on complex base in Section II. Section III verifies the feasibility of the proposed method and discusses geometric properties of multimedia representation method. The practical applicability announcement of our experimental results is introduced in Section IV. Finally, the conclusion and future work are stated in Section V.

II. OUR METHOD

Assuming that A and B are real integers, then the complex number $N = A + Bi$ is called a Gaussian integer, wherein $i^2 = -1$, A and B are the real and imaginary parts of Gaussian integer N, respectively. When $r = x + yi$ is a complex base, the digits of the representation is $m_j \in \{0, 1, \ldots, x^2+y^2-1\}$, and Gaussian integer N is written in the following form:

$$N = m_j r^j + \ldots + m_1 r + m_0 = \sum_{j=0}^{n} m_j r^j \tag{1}$$

where N is said to be expressed in the complex base r. The standard algorithm for converting a number into a given integer base can be extended to these complex bases. The allowable digits m_j form a complete residue system modulo the complex base. A complete residue system modulo the complex base $r = x+yi$ is x^2+y^2. Gauss has proved if x and y are relatively prime then the natural numbers $0, 1, \ldots, x^2+y^2-1$ form a complete residue system modulo. Furthermore, if x and y have a common factor, then any complete residue system modulo r must contain some numbers with nonzero imaginary parts. Hence, when we want to represent all the Gaussian integers, a necessary condition is that $y = \pm 1$, since all the powers of the base $(x+yi)^j$ have their imaginary parts divisible by y [19]. We only consider bases of the form $x \pm i$ to represent all the Gaussian integers. On the other hand, because of the characteristics of digital computer, the digits form m_j should be $m_j \in \{0, 1\}$. Hence, we have $x^2+y^2-1 = 1$, that is $x = \pm 1, y = \pm 1$. The considered complex bases are $r = 1+i$, $r = 1-i$, $r = 1-i$ and $r = -1+i$. Katai and Szabo [18] have proved by algebraic methods that only $-x+i$ and $-x-i$ can be used as bases for all complex numbers, wherein x is a positive integer. Therefore, when each complex number can be uniquely represented by using a given base, we can choose complex bases $r = -1+i$ or $r = -1-i$, and they are all suitable for representing multimedia information. In our manuscript, we randomly choose $r = -1-i$ as complex base to represent multimedia information in all subsequent experiments.

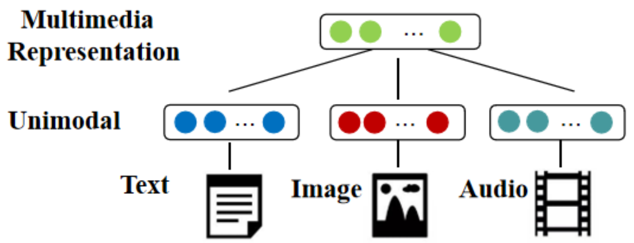


FIGURE 1. The multimedia representation model based on complex base.

In the following, we will illustrate the proposed multimedia representation method based on a complex base. As shown in Fig. 1, the input unimodal data of text, image and audio are separately represented as different Gaussian integers. Then these Gaussian integers are mapped on the same complex plane and a joint multimedia representation model can be obtained. In particular, this model is composed of two main sub-processes: Multimedia-to-Gaussian Integer and Gaussian Integer-to-Multimedia. The specific implementation details will be discussed.

A. MULTIMEDIA-TO-GAUSSIAN INTEGER

Multimedia-to-Gaussian Integer is to transform different modalities data into Gaussian integers. In first step, the input multimedia data are respectively encoded into different $0-1$ digital sequences. Then the second step is to convert each $0-1$ sequence into Gaussian integer under the given complex base. Assuming that complex base is $r = x+iy$, Gaussian integer is $N = A_0 + B_0 i$, and one $0-1$ sequence is denoted as $m = \{m_n, m_{n-1}, \cdots, m_1, m_0\}$, wherein n is the length of someone $0-1$ sequence, and $m_j \in \{0, 1\}, j \in [0, n]$. The detailed processes are as follows: Let $(x+iy)^j = s_j + it_j$. When $j = 0$, we have $s_0 = 1, t_0 = 0$. In order to construct the recursion, let's $(x+iy)^{j-1} = s_{j-1} + it_{j-1}$. Hence

$$\begin{aligned}(x+iy)^j &= (x+iy)(s_{j-1} + it_{j-1}) \\ &= xs_{j-1} - ys_{j-1} + i(xs_{j-1} + yt_{j-1}).\end{aligned} \tag{2}$$

And, it follows immediately that

$$\begin{aligned}s_j &= xs_{j-1} - yt_{j-1}, \\ t_j &= ys_{j-1} + xt_{j-1},\end{aligned} \tag{3}$$

so, we have

$$\begin{bmatrix} s_j \\ t_j \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} s_{j-1} \\ t_{j-1} \end{bmatrix}. \tag{4}$$

In case of $N = \sum_{j=0}^{n} m_j r^j$ and $m_j \in \{0, 1\}$, we take a similar substitution

$$\begin{aligned}N &= A_0 + iB_0 \\ &= m_n r^n + m_{n-1} r^{n-1} + \cdots + m_0 r^0\end{aligned}$$

Pseudo code 1 Multimedia-to-Gaussian Integer Algorithm

Input: Complex base $r = x + iy = -1 - i\ (x = -1, y = -1)$, Multimedia data
Output: Gaussian integer N

function MtoG(N)
  Converting single modality data into $0-1$ sequence $m = \{m_n, m_{n-1}, \cdots, m_1, m_0\}$
  $A_0 \leftarrow 0$
  $B_0 \leftarrow 0$
  $s_0 \leftarrow 1$
  $t_0 \leftarrow 0$
  $j \leftarrow 0$
  while $j <= n$ **do**
    $A_0 \leftarrow A_0 + m_j s_j$
    $B_0 \leftarrow B_0 + m_j t_j$
    $s_{j+1} \leftarrow -s_j + t_j$
    $t_{j+1} \leftarrow -s_j - t_j$
    $j++$;
  end while
  Get $N \leftarrow A_0 + iB_0$
end function

$$\begin{aligned} &= m_n(x+iy)^n + m_{n-1}(x+iy)^{n-1} + \cdots + m_0(x+iy)^0 \\ &= m_n(s_n + it_n) + m_{n-1}(s_{n-1} + it_{n-1}) + \cdots + m_0 \\ &= (m_n s_n + m_{n-1} s_{n-1} + \cdots + m_0) + (m_n t_n + \cdots + m_1 t_1), \end{aligned} \tag{5}$$

then we get

$$A_0 = \sum_{j=0}^{n} m_j s_j, \qquad B_0 = \sum_{j=1}^{n} m_j t_j. \tag{6}$$

In the final step, the multimedia data are represented as different Gaussian integers $N = \sum_{j=0}^{n} m_j s_j + i \sum_{j=1}^{n} m_j t_j$.

The pseudo code 1 of the Multimedia-to-Gaussian Integer algorithm gives detailed implementation procedures. Since multimedia data contain image, text, and audio in this paper, the algorithm should be repeated three times until all types of data are processed. Finally, different modalities data will be converted into different Gaussian integers.

B. GAUSSIAN INTEGER-TO-MULTIMEDIA

Gaussian Integer-to-Multimedia is committed to transform Gaussian integers into multimedia data. We suppose that the known Gaussian integer is $N = A_0 + B_0 i$, and the given complex base is $r = x + iy$. Now we need to obtain the $0-1$ sequence $m = \{m_n, m_{n-1}, \cdots, m_1, m_0\}$ under the corresponding Gaussian integer. Since the representation of Gaussian integer in (1) is $A_0 + iB_0 = m_n r^n + m_{n-1} r^{n-1} + \cdots + m_0 r^0$. So, let

$$A_1 + iB_1 = m_n r^{n-1} + m_{n-1} r^{n-2} + \cdots + m_1, \tag{7}$$

therefore

$$\begin{aligned} A_0 + iB_0 &= r(A_1 + iB_1) + m_0 \\ &= (x+iy)(A_1 + iB_1) + m_0 \\ &= xA_1 - yB_1 + (yA_1 + xB_1)\, i + m_0, \end{aligned} \tag{8}$$

hence it follows that

$$\begin{aligned} A_0 &= xA_1 - yB_1 + m_0, \\ B_0 &= yA_1 + xB_1. \end{aligned} \tag{9}$$

Then, keep iterating continuously, finally we can get

$$\begin{bmatrix} A_j \\ B_j \end{bmatrix} = \begin{bmatrix} x & y & -x \\ -y & x & y \end{bmatrix} \begin{bmatrix} \frac{A_{j-1}}{x^2+y^2} \\ \frac{B_{j-1}}{x^2+y^2} \\ \frac{m_{j-1}}{x^2+y^2} \end{bmatrix}. \tag{10}$$

and that is

$$\begin{aligned} A_j &= \frac{xA_{j-1} + yB_{j-1} - xm_{j-1}}{x^2+y^2} \\ B_j &= \frac{-yA_{j-1} + xB_{j-1} + xm_{j-1}}{x^2+y^2}. \end{aligned} \tag{11}$$

Taking into account that A_j and B_j are both integers, and it should be defined that $m_j = |(A_j - B_j) mod\, 2|, j \in [0, n]$, it means that when A_j and B_j have the same parity, m_j is equal to 0, otherwise, the value of m_j is 1. We repeat this process, and it is terminated at the j^{th} step if $A_j = 0$ and $B_j = 0$. In the end, the corresponding digit sequence m can be obtained. After this process completes, all that remains is to reorganize the generated $0-1$ binary sequences into different visual media according to their original encoding methods.

Pseudo code 2 Gaussian Integer-to-Multimedia Algorithm

Input: Gaussian integer $N = A_0 + iB_0$, complex base $r = x + iy = -1 - i\ (x = -1, y = -1)$
Output: Multimedia(Text, image or audio)

function GtoM(m)
  $m \leftarrow \{0, 0, \cdots, 0, 0\}$
  for $j = 0; j <= n; j++$ **do**
    if $A_j == 0$ and $B_j == 0$ **then**
      break;
    else
      $m_j \leftarrow |(A_j - B_j)| mod 2$
      $A_{j+1} \leftarrow (-A_j - B_j + m_j)/2$
      $B_{j+1} \leftarrow (A_j - B_j - m_j)/2$
    end if
  end for
  Get $m \leftarrow \{m_n, m_{n-1}, m_{n-2} \cdots m_2, m_1, m_0\}$
  Convert the obtained $0-1$ sequence m into multimedia
end function

The pseudo code 2 of Gaussian Integer-to-Multimedia is used to transform Gaussian integers back to multimedia data under the complex base $r = -1 - i$. Likewise, this algorithm also needs to execute three times, until all Gaussian integers are processed. Note that different types of data have different encoding methods, so the methods of using $0-1$ sequences to reconstruct original data are also different.

C. MULTIMEDIA REPRESENTATION

Commonly used types of single modal representations, such as text, image, and audio, are often as cornerstones for learning multimodal representations. Here, a joint multimedia representation can be obtained by combining above unimodal data into a same representation space.

Digital Audio Signal Representation: In general, there are two main types of audio files: mono channel audio files and multi-channel audio files, respectively. After processed by a digital computer, audio files can be regarded as a $n \times m$ matrix. Especially, when $m = 1$, this is a mono audio file, if $m \geq 2$, it is a common multi-channel audio file. When we want to represent an audio file, the $n \times m$ matrix should be converted to a $0-1$ sequence, then the Gaussian integer is computed by using the Multimedia-to-Gaussian Integer algorithm. As for the $n \times 1$ matrix, the $0-1$ sequence can be obtained by arranging the binary representation of each number one by one. Here is an illustrative example to show the process. We suppose that there is a mono channel audio file, the format is ".flac", the sampling frequency is 16 KHZ, and the sample depth is 16 bits. As is demonstrated in Fig. 2, we randomly choose 6 consecutive sample data in this mono audio file. Fig. 2(a) shows the time-domain waveform plot of sample data. The first step is to convert the sample data from a decimal number to an integer number. As depicted in Fig. 2(b), these integer numbers can be obtained by the process, which is to multiply each sampled value by 2^{15} [see the left column in Fig. 2(b)]. The second step is to convert each integer number to a 16-bit binary coding [see the right column in Fig. 2(b)], wherein the highest bit is a sign bit, 1 and 0 mean that the data is negative and positive, respectively. For example, as for sample data $x = -0.00057983$, there is $y = x \times 2^{15} = -19$, and its binary coding is 1000000000010011. Finally, we can get the $0-1$ sequence of sample data by sequentially reading the binary coding of each number [see Fig. 2(c)]. Under the given complex base $r = -1 - i$ and its corresponding $0-1$ sequence, the sampled points can be represented as a Gaussian integer $N_1 = \sum_{j=0}^{95} m_j s_j + \sum_{j=1}^{95} m_j t_j = -145702503678976 - 140192027672834i$, wherein the real and imaginary parts of Gaussian integer $(-145702503678976, -140192027672834)$ is its coordinate on the complex plane.

For a multi-channel audio file, we only need to process each channel similarly, and concatenate them together to form a $0-1$ sequence, then achieve the audio representation. As shown in Fig. 3, we use a small portion of the data in two-channel audio file as an example to describe the processing of the multi-channel audio file. In step 1, we randomly select four sampled points from the left channel (L) and its corresponding right channel (R). In step 2, we further convert these sampled points into integers using

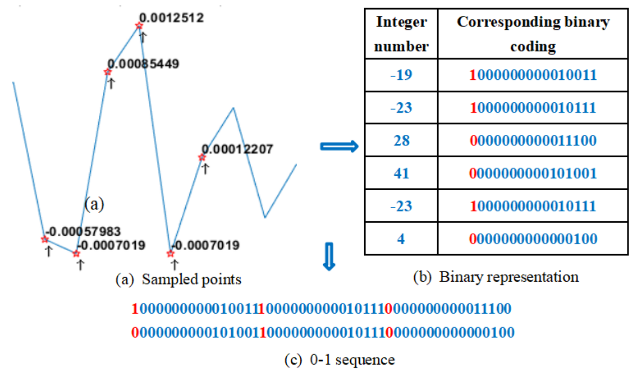


FIGURE 2. Example of a small segment of digitized audio file to $0-1$ sequence.

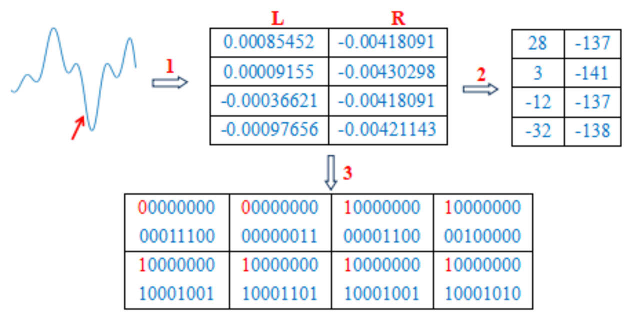


FIGURE 3. From a small portion of the data in two-channel audio file to $0-1$ sequence.

the similar single channel audio conversion method. In step 3, the left and right channels' sampled points are respectively converted to upper and lower two lines corresponding $0-1$ bit streams. A one-dimensional $0-1$ sequence can be obtained by scanning each row in turn from left to right. Especially, the $0-1$ sequence obtained in step 3 is diverse, and we can also obtain a different $0-1$ sequence by changing the scanning ways. Finally, the Gaussian integer of sampled data is $G = -144254261357020551 - 140168295577979i$. Conversely, under the given corresponding Gaussian integer, we can recover the original audio data accurately.

Image Representation: In image processing, we usually divide digital images into binary images, gray-scale images and color images. Additionally, a color image is regarded to be a three-dimensional matrix in a computer memory, and a binary image or gray-scale image is considered as a two-dimensional matrix. When we represent a digital image, the multi-dimensional matrix also needs to be transformed into a one-dimensional $0-1$ sequence. As is evident from Fig. 4, here are some typical two-dimensional $(2-D)$ to one-dimensional $(1-D)$ matrix conversion methods, wherein each smallest square block represents a pixel or a pixel block of one image. We can obtain $0-1$ sequence based on the blue line passing through the square block from one red dot to another red dot. Now supposing that an gray-scale image is given, as simply illustrated in Fig. 5(a), we randomly choose 9 pixels in this image, then the pixels form a $2-D$ matrix of

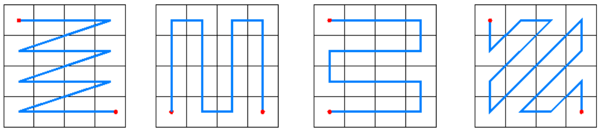


FIGURE 4. Four typical $2-D$ to $1-D$ conversion methods.

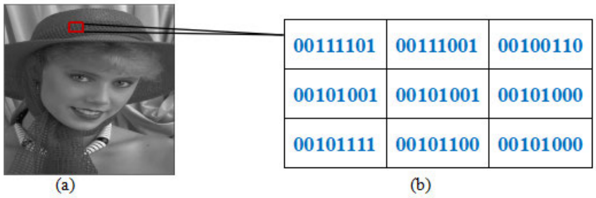


FIGURE 5. Example of gray-scale image to $0-1$ sequence. (a) gray-scale image. (b) Corresponding two-dimensional $0-1$ sequence.

24×3 [see Fig. 5(b)]. When the leftmost method shown in Fig. 4 is selected for converting the $2-D$ matrix to $1-D$ matrix, then the corresponding $1-D$ $0-1$ sequence m can be obtained. As the same, under the base $r = -1-i$ and the $0-1$ sequence m, the Gaussian integer N_2 is $N_2 = \sum_{j=0}^{71} m_j s_j + \sum_{j=1}^{71} m_j t_j = 13681256938 + 17802863426i$. Obviously, the representation method is suitable for binary and color images.

Text Representation: For a given English text, each letter or character of the text is represented by ASCII coding in a computer. For example, the ASCII coding of "Macau SAR" is: 77 97 99 97 111 32 83 65 82. Then the corresponding $0-1$ sequence can be obtained. By using the Multimedia-to-Gaussian Integer algorithm, this English text can be represented as Gaussian integer $N_3 = \sum_{j=0}^{71} m_j s_j + \sum_{j=1}^{71} m_j t_j = 1.43 \times 10^{10} - i3.5522 \times 10^{10}$ based on the complex base $r = -1-i$. It is worth mentioning that whatever the text information is expressed in any language, such as Chinese, Japanese, etc., they all can be represented as Gaussian integers.

After the above unimodal data single channel audio, text, and image are respectively represented as different Gaussian integers, we can obtain a joint multimedia representation. As depicted in Fig. 6, these unimodal data are combined into the same representation complex plane, then it is convenient to capture their relationships based on classical geometric operations on the complex plane.

III. EXPERIMENTAL RESULTS

In this section, we concentrate on perceptual measure of the proposed multimedia representation method. Signal-to-Noise Ratio(SNR) is employed to evaluate the quality of reconstructed audio and text, and Peak Signal-to-Noise Ratio (PSNR) is applied to measure the quality of reconstructed image. The higher the value of SNR and PSNR are, the better quality of reconstructed multimedia data are. Further, we discuss geometric properties of multimedia representation

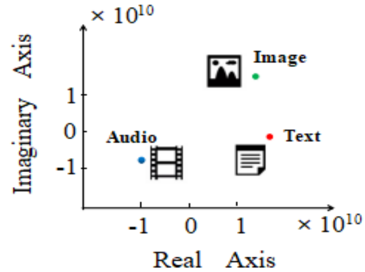


FIGURE 6. A joint multimedia representation. Each unimodal data is respectively represented as Gaussian integer, and then we combine these unimodal data and map them into a same representation complex plane.

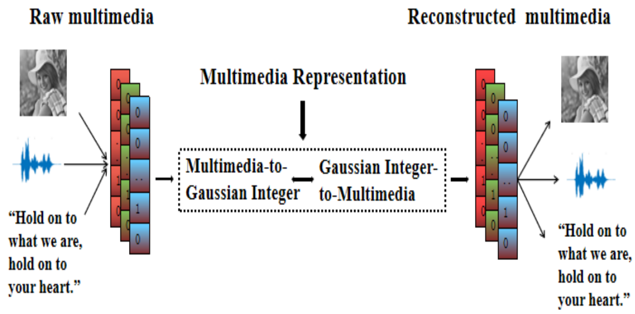


FIGURE 7. Multimedia representation. The original image, audio and text are respectively represented as Gaussian integers 1.9222×10^{4931}-$3.4222 \times 10^{4931}i$, 8.7195×10^{37859}+$6.6525 \times 10^{37854}i$, and 3.6772×10^{54}+$1.5828 \times 10^{55}i$.

method. Note that each Gaussian integer retains four decimal places for convenience to represent.

As shown in Fig. 7, it gives the results of multimedia representation based on complex base. The left column is the raw multimedia data which includes image, audio, and text, wherein the size of the image is 64×64, the utterance of audio is "You are acute" which sample frequency is 8 KHZ and sample depth is 16 bits, and the content of text is "Hold on to what we are, hold on to your heart". Firstly, the raw multimedia data are respectively coded into different $0-1$ sequences in a computer, wherein the leftmost method shown in Fig. 4 is selected to convert image data from a $2-D$ matrix to $1-D$ matrix. By using the Multimedia-to-Gaussian Integer algorithm, they are respectively represented as different Gaussian integers and mapped into a same representation space. Then Gaussian integers of each single modality data can be respectively converted back into different $0-1$ sequences by using the Gaussian Integer-to-Multimedia algorithm, and the reconstructed multimedia data can be obtained by using the appropriate methods to recode the corresponding $0-1$ sequences. Comparing raw multimedia data with reconstructed multimedia data, the value of PSNR between images is infinite($PSNR = +\infty$), and the value of SNR in audio and text are all infinite($SNR = +\infty$). These results show that reconstructed multimedia data are the same as raw multimedia, and the proposed representation method is lossless and reversible.

TABLE 1. Comparison of representation methods.

Representation Schemes	[11]	[12]	[13]	[14]	[15]	[16]	[20]	Ours
Lossless Representation	✓	✓	✓	✗	✓	✓	✓	✓
Unique Representation	✓	✓	✗	✓	✓	✗	✓	✓
Geometric Operations for Media Data with Different Dimensions or Sizes	✗	✗	✗	✗	✗	✗	✓	✓
Change Statistical Correlation	✗	✗	✗	✗	✗	✗	✓	✓
Multimedia Representation	✓	✓	✗	✗	✗	✗	✓	✓
Multimodal Application	✗	✗	✗	✗	✗	✗	✗	✓

Furthermore, in order to verify the advantages of the proposed multimedia representation method, there are some comparisons among the proposed complex number system and others. As shown in Table 1, considering the conditions that should be present when a numerical system successfully represents multimedia information and the characteristics of the arithmetic operations under the numerical systems, we compare the following metrics. (1) Lossless Representation: it refers to the ability to accurately represent the information. (2) Uniqueness Representation: it indicates the uniqueness of the represented information without redundancy. (3) Geometric Operations for Media Data with Different Dimensions or Sizes: geometric operations can be performed between media data with inconsistent dimensions or sizes. (4) Change Statistical Correlation: the inherent statistical properties of media data will change. (5) Multimedia Representation: the method can be used to represent multimedia data. (6) Multimodal Application: supporting multimodal application.

It can be seen from Table 1, reference [14] can only approximate the represented information infinitely, it is not a lossless representation. The representation of multimedia data is not unique in [13] and [16], e.g., the number 54 can be denoted as $54 = 2^1 3^3$ or $54 = 2^2 3^2 + 2^1 3^2$ in [13], so the represented data may not be inconsistent with the reconstructed data. In addition, the represented information in [11], [12], [13], [14], [15], and [16] are real numbers, so the represented media data involved in the arithmetic operations must have the same dimension or size, e.g., the sizes of a 32×32 image and a 32×64 image cannot perform traditional addition operation, and moreover if all pixels of the represented image are summed with a decimal number, the statistical correlation histogram of an image does not change, so performing the traditional arithmetic operation is also hard to scramble and protect the original data.

Overall, the proposed complex system is more suitable for multimedia representation and multimodal applications than other systems.

Particularly, after each single modality data in Fig. 7 are represented as Gaussian integers, a joint multimedia representation can be obtained by mapping them into the same representation complex plane. Then, the relationships between multimedia data can be established on the representation space. Supposing that the unimodal data audio, text, and image are respectively represented as Gaussian integers $N_1 = A_1 + iB_1$, $N_2 = A_2 + iB_2$ and $N_3 = A_3 + iB_3$, and their positions can be respectively marked as coordinates (A_1, B_1), (A_2, B_2) and (A_3, B_3), wherein the abscissa and ordinate of each point are corresponding to the real and imaginary parts of its Gaussian integer. Now we randomly choose two Gaussian integers N_1 and N_2 among them to discuss their relationships. Based on geometric operations on the complex plane, the two Gaussian integers N_1 and N_2 can have the following geometric operations:

FIGURE 8. Subtraction operation between multimedia data. The subtraction operation can be performed between multimedia data which have different data types and sizes.

$$
\begin{aligned}
N_1 + N_2 &= (A_1 + A_2, B_1 + B_2) \\
N_1 - N_2 &= (A_1 - A_2, B_1 - B_2) \\
N_1 \times N_2 &= (A_1A_2 - B_1B_2, A_1B_2 + B_1A_2) \\
N_1 \div N_2 &= \left(\frac{A_1A_2 + B_1B_2}{A_2^2 + B_2^2}, \frac{B_1A_2 - A_1B_2}{A_2^2 + B_2^2}\right)
\end{aligned} \tag{12}
$$

Then, based on represented multimedia data in Fig. 7, we select the subtraction operation on the complex plane as an example to illustrate the relationships between multimedia data, and the calculated results are shown in Fig. 8. It can be found that the subtraction operation of multimedia data can be performed, the first row presents geometric operation of image minus audio, the second row presents geometric operation of text minus image, and it is also feasible for geometric operations on other data types between multimedia. Furthermore, when the selected geometric operation is known, different data types can be represented by each other. For example: the second row in Fig. 8 demonstrates the result in which Gaussian integer of text minus the Gaussian integer of image, we assume that Gaussian integer of this result is equal to G_k, and Gaussian integer of participated text is G_t. Then Gaussian integer G_i of participated image can be

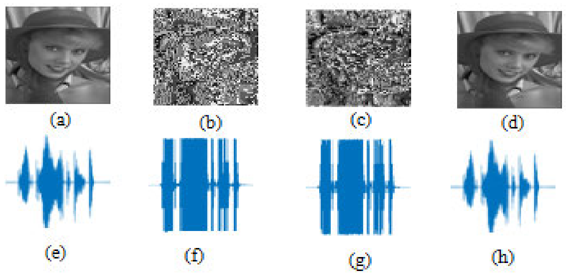


FIGURE 9. Inverse and conjugate operations for unimodal data. (a)Original image. (b)Inverse operation of image. (c)Conjugate operation of image. (d)Reconstructed image. (e)Original audio. (f)Inverse operation of audio. (g)Conjugate operation of audio. (h)Reconstructed audio.

deduced and there is $G_i = G_t - G_k$. Finally, the participating image can be reconstructed by employing Gaussian integer-to-Multimedia algorithm. This process completes the mutual representation from text to image. Furthermore, the addition, multiplication, and division have the same property between multimedia data. For traditional geometric operations, the participating data types must be consistent and have the same length.

In addition, there are inverse and conjugate operations for the geometric operation of one Gaussian integer $N_1 = A_1 + iB_1$, and the equations are as follows:

$$\begin{aligned} -N_1 &= (-A_1, -B_1) \\ \overline{N_1} &= (A_1, -B_1) \end{aligned} \tag{13}$$

As depicted in Fig. 9, it demonstrates inverse and conjugate geometric operations on unimodal data, where the reconstructed data are obtained by twice inverse or conjugate operations. It can be seen that once inverse or conjugate operation has scrambled original data, but the reconstructed data are the same as original data. This is because after one transformation, Gaussian integer of unimodal data has changed, and Gaussian integer of the data obtained by twice inverse or conjugate operations is equal to Gaussian integer of the original data. Multimedia data composed of these unimodal data also has the same properties, and these geometric operations have certain research value in multimedia scrambling or encryption.

From the above experiments, we know that the proposed multimedia representation method is lossless and reversible. After multimedia data are represented as Gaussian integers and marked on the same complex plane, the mutual representation between multimedia data can be performed by using geometric operations on the complex plane. Especially there is a limitation in this study, when we want to represent large-size multimedia data, the order of magnitude calculations of Gaussian integers are usually large. Note that our experiments need to use multiprecision computing toolbox for MATLAB to achieve the computation of Gaussian integer. Theoretically, the toolbox is capable of working with any level of precision and hardware configuration is the only limiting factor. In practice, since our computer resources is limited, the $0-1$ sequence length of multimedia data can not be infinite.

IV. POTENTIAL APPLICATION AND DISCUSSION

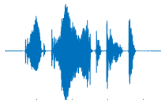
In this section, based on the geometric characteristics of the proposed multimedia representation method, we apply the proposed method to multimedia camouflage and multimedia sharing in multimedia security. In all experiments, the selected audio signals are LibriSpeech ASR corpus (test-clean) and downloaded from: https://www.openslr.org/12/, wherein the sampled frequency is 16 KHZ and sampled depth is 16 bits. The original version of selected images is chosen from the ''Kodak image set'', which is downloaded from: http://r0k.us/graphics/kodak/.

A. MULTIMEDIA CAMOUFLAGE

Multimedia camouflage based on the proposed representation method is a multimedia security technology that replaces secret information with irrelevant information for communication, which can realize the transformation between cross-modal data. The camouflaged process is as follows: Supposing that one needs to transmit a secret image to a receiver in a communication process. In order to ensure the confidentiality of the information, an irrelevant text or audio will be sent. On the receiving end, when one knows the selected secret key and geometric operation, then secret data can be reconstructed. As an important branch of camouflage technology, Reversible data hiding (RDH) is a way to embed data into the cover medium for the purposes of ownership protection, authentication, fingerprinting, secret communication, annotation, etc., and perfectly recover both the original content and the hidden data [38]. Although various RDH schemes have achieved well, they also exist the bottleneck of embedding capacity, multimedia camouflage scheme can well solve the current problem.

It can be found from Table 2, there are three camouflaged experiments, wherein the size of each image is 64×64, and the content of selected audio signals are respectively ''She was alone that night'', ''The lad had checked him then'', ''It was the beauty of it'' from experiment 1 to experiment 3. In each experiment, the secret message is camouflaged as different modalities of message 1 or message 2 to transmit, they all can be represented as different Gaussian integers by using the Multimedia-to-Gaussian Integer, and the reconstructed message can be recovered through the secret key and camouflaged message. Here we illustrate an image camouflaged example to explain the experimental process. Now assuming that we need to send a secret image to the receiver, in order to protect the security of the secret image, we choose a camouflaged text or audio file as a public message to send. Because the readable text or audible audio which does not carry any secret information is difficult to arouse suspicion by attackers and is safe enough in the transmission phase. We denote that the secret image and public camouflaged message

TABLE 2. Multimedia camouflage based on the proposed multimedia representation method.

Experiment	Secret message	Camouflaged message 1	Camouflaged message 2	Reconstructed message
Experiment 1			"Life is a song, sing the life rhythm and melody"	
Gaussian integer	4.4181×10^{4931} $+3.2191 \times 10^{4931}i$	2.9958×10^{75835} $-1.7195 \times 10^{75794}i$	9.6138×10^{56} $+3.2904 \times 10^{57}i$	4.4181×10^{4931} $+3.2191 \times 10^{4931}i$
Experiment 2		"Online shopping is becoming more and more common"		
Gaussian integer	$-1.0790 \times 10^{77834}$ $-1.0788 \times 10^{77834}i$	9.1713×10^{56} $+3.6602 \times 10^{57}i$	-2.2835×10^{4931} $+7.0977 \times 10^{4931}i$	$-1.0790 \times 10^{77834}$ $-1.0788 \times 10^{77834}i$
Experiment 3	"The event will be held on Sunday"			"The event will be held on this Sunday"
Gaussian integer	-8.0381×10^{43} $+1.4232 \times 10^{44}i$	-2.0788×10^{4931} $-7.1264 \times 10^{4931}i$	6.4271×10^{77860} $-7.4821 \times 10^{77850}i$	-8.0381×10^{43} $+1.4232 \times 10^{44}i$

(audio or text) are represented as Gaussian integers N_{secret} and N_{public}, respectively. In particular, various geometric operations in (12) are selected to encrypt the camouflaged process. Supposing that we choose an addition operation to encrypt, there is $N_{key} = N_{secret} + N_{public}$. After receiving the camouflaged message and secret key, we can calculate the Gaussian integer N_{secret} of the secret image, and reconstruct the secret image by using the Gaussian Integer-to-Multimedia algorithm. In addition, there is $SNR = +\infty$ between the secret image and the reconstructed image. The experimental result shows that the recovered secret image is lossless, and the camouflage scheme is successful. Similarly, the text and audio can also be camouflaged as other different types of data, the process will not repeat again.

From the above experiments, we can conclude that multimedia camouflage based on the proposed multimedia representation method can well achieve mutual camouflage between different modal data. The evaluated measurements in RDH schemes are embedding capacity (EC) and the quality of cover data (PSNR) [30]. Compared with the existing RDH schemes, there are some advantages:

(1) No size limitation between secret messages and camouflaged messages and the quality of cover data is lossless. The recent RDH studies focus on increasing the image fidelity and the embedding capacity [26], [27], [28], [29], [30], wherein the best study in image "lena" is $EC = 30000$ bits and $PSNR = 55.17$. When we continue to increase the embedding capacity in RDH schemes, the quality of the cover data will be degraded and even the reconstructed secret data is affected. In our camouflaged scheme, although different modalities data are used in Table 1, we can also achieve camouflage in the same modality data. For instance: the size of a 64×64 secret "lena" image can be camouflaged by using a cover 128×128 image in our scheme, the calculated embedding capacity is $EC = 32768$ bits. Therefore, the camouflaged scheme has an advantage in embedding capacity. On the other hand, our scheme uses "replacement" instead of "embedding" to protect secret information, so the cover data

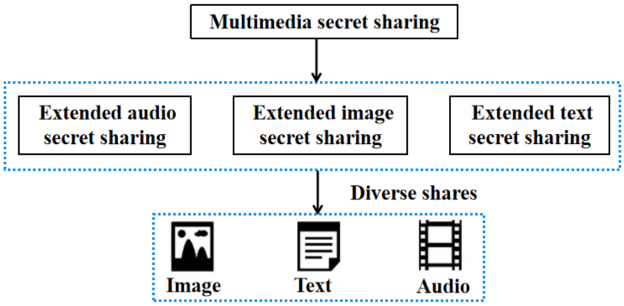


FIGURE 10. A unified multimedia secret sharing(MSS) framework, which can be used to share multimedia into multimedia. From the data types of secret and shared messages, MSS includes extended audio, image or text secret sharing. And each secret sharing scheme can apply diverse multimedia as shares.

before and after the camouflaged scheme executed is no changed, which means that the quality of cover data is lossless on the camouflaged processes and the value of PSNR is $+\infty$.

(2) The application is to flexibly achieve the camouflage of cross-modality data. However, current studies are nearly single modal data and the cross-modality camouflaged strategy are not present in the state-of-the-art works.

(3) The camouflaged messages are meaningful and diverse, and have nothing to do with secret information. Hence, it enhances confidence in the transmission phase.

B. MULTIMEDIA SECRET SHARING

Multimedia secret sharing(MSS) is the application of secret sharing in multimedia information security, which can not only protect secret information, but also realize decentralized management of secret information. As seen in Fig. 10, from the perspective of the data types of secret messages and share messages, MSS can be regarded as a unified secret sharing model, which includes extended image secret sharing(EISS), extended audio secret sharing(EASS), and extended text secret sharing(ETSS). Compared with traditional secret sharing, the generated shares in each extended secret sharing scheme are diverse multimedia data, and are not just the same data type as the secret message.

To better understand sharing and reconstruction processes of such new multimedia sharing strategy, an illustration is given. For such sharing process, supposing that one wants to share a secret multimedia M_1 into two other information M_2 and M_3. Under geometric operations on the complex base, we know that when multimedia data are separately represented as Gaussian integers N_1, N_2 and N_3, they also obey the parallelogram rule in a two-dimensional vector space, and one vector can be a linear combination of the other two vectors. As shown in (14-16), we can apply one of these simple formulas to implement sharing process and establish the relationships among multimedia data. When one wants to reconstruct secret multimedia M_1, we need to obtain two other shares and corresponding correlation coefficient λ_i, then calculate its Gaussian integer of secret multimedia. Finally, the secret multimedia is reconstructed

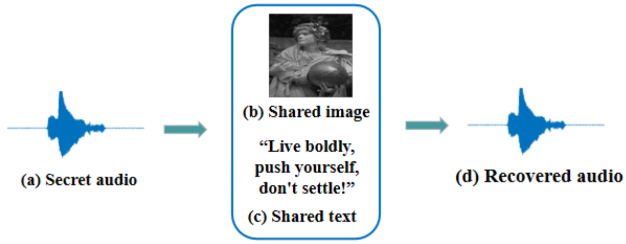


FIGURE 11. Multimedia secret sharing for audio (MSS-Audio(2, *n*)). Sharing secret audio "VENICE" into *n* shares, and we choose an image size of 64 × 64 and one text to recover secret audio. The Gaussian integers of them are respectively are $N_1 = -1.2646 \times 10^{69735} + 1.6705 \times 10^{69740} i$, $N_2 = 3.3809 \times 10^{4930} + 6.5844 \times 10^{4931} i$, $N_3 = 3.5586 \times 10^{48} + 1.2258 \times 10^{49} i$, and $N_4 = -1.2646 \times 10^{69735} + 1.6705 \times 10^{69740} i$. Based on the sharing (14) $N_1 = \lambda_1 N_2 + \lambda_2 N_3$, we suppose that $\lambda_1 = 1$, and then $\lambda_2 = 1.2659 \times 10^{69691} + 3.6489 \times 10^{69690} i$, hence they are sharing keys which are used to recover the original secret audio. Note that the sharing keys are not unique, any λ_1 and λ_2 can be used as secret keys, when they can make the above formula true. The SNR between secret audio and recovered audio is infinite($+\infty$), it indicates that the secret audio can be reconstructed accurately.

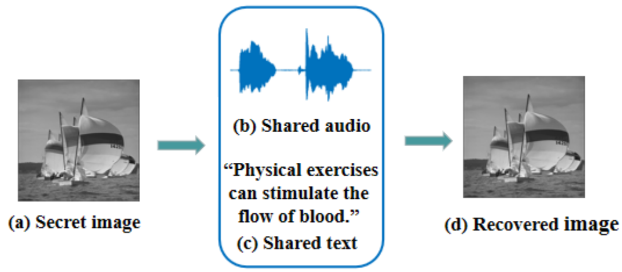


FIGURE 12. Multimedia secret sharing for image (MSS-Image (2, *n*)). Secret image with size of 64 × 64 is shared into *n* shares, wherein the audio"Fine glorious" and one text are used to recover the secret image. The Gaussian integers of them are respectively $N_1 = 3.9679 \times 10^{4931} + 5.4528 \times 10^{4931} i$, $N_2 = -7.6490 \times 10^{81301} - 7.4702 \times 10^{81301} i$, $N_3 = -5.8061 \times 10^{60} + 1.3498 \times 10^{61} i$, and $N_4 = 3.9679 \times 10^{4931} + 5.4528 \times 10^{4931} i$. Based on the sharing (15) $N_1 = \lambda_3 N_2 + \lambda_4(-N_3)$, we suppose that $\lambda_3 = 2$, then $\lambda_4 = 1.2340 \times 10^{81240} + 1.8010 \times 10^{81241} i$. And they are sharing keys which can be used to recover the secret image from the shared audio and text. The PSNR between the secret image and recovered image is infinite($+\infty$).

by using Gaussian integer-to-Multimedia algorithm.

$$N_1 = \lambda_1 N_2 + \lambda_2 N_3 \tag{14}$$

$$N_1 = \lambda_5 N_2 + \lambda_6 (-N_3) \tag{15}$$

$$N_1 = \lambda_3 N_2 + \lambda_4 \overline{N_3} \tag{16}$$

wherein λ_i, $i \in [1, 6]$ is the correlation coefficient which can be used to reconstruct secret information.

Here the feasibility of the proposed MSS is verified by subsequent experiments. As depicted in Figs. 11-13, the three experiments show the multimedia secret sharing for audio (MSS-Audio(2, n)), multimedia secret sharing for image (MSS-Image(2, n)) and multimedia secret sharing for text (MSS-Text(2, n)), respectively. In particular, the choices of generated n shares are diverse in each experiment, and they can be the same or different data type as the secret message. In order to highlight the diversity of generated shares, we deliberately select 2 shares that are different from the data type of secret message to share in each experiment. Here, an illustrative example MSS-Audio $(2, n)$ is given. As depicted in Fig. 11, it shows that secret audio M_a is divided into n shares which include different types of data, and any 2 shares can be used to reconstruct the secret audio, such as two audio signals, or one text and one image, or two text messages, or two images and so on. Here we only analyze one of the shared combinations to discuss. Firstly, we select one text and one image that are different from the data type of secret audio as the shares, and mark them as M_b and M_c, respectively. Secondly, we represent multimedia data M_a, M_b and M_c as Gaussian integers N_1, N_2 and N_3, respectively. Based on the previous (14-16), we assume that $N_1 = \gamma_1 N_2 + \gamma_2 N_3$. Finally, when one knows the two secret keys γ_1 and γ_2, the secret audio M_d can be reconstructed. Throughout the application, the multimedia representation, as well as their geometric operations of Gaussian integers are a lossless and reversible process, it means that there is no loss between the recovered information and the original information, and they are numerically identical. Therefore, the original audio M_a and the recovered audio M_d are the same, the SNR between them is infinite($+\infty$). The result presents that the secret audio is reconstructed accurately, and MSS-Audio$(2, n)$ is successful.

The current studies on secret sharing mainly focus on image secret sharing. Because audio and text are very sensitive to information changes, there are few studies on audio secret sharing, and text secret sharing is even less. The implementation process of MSS-Image $(2, n)$ and MSS-Text $(2, n)$ are similar to the MSS-Audio $(2, n)$, so the detailed implementation process is not discussed in this section. As we can see in Fig. 12, we choose an audio and text as shares to recover the secret image, the sharing result indicates that original secret image can be reconstructed accurately. As for MSS-Text $(2, n)$, it can be found in Fig. 13, the generated image and audio are used to share the sensitive secret text, and the experimental result demonstrates that secret text can also be recovered perfectly.

Table 3 and Table 4 demonstrate the performance of our scheme, wherein the key measurements [35] in the secret sharing scheme are discussed.

As shown in Table 3, compared with the current audio secret sharing [39], [40], [41], [42], our scheme exhibits some good features, e.g., the quality of generated shares is high, not meaningless or noise-like, and the types of generated shares are also diverse such that the attackers are difficult to determine which shares are used to reconstruct secret message. Furthermore, our scheme has no any constraints on secret or cover audio, and some schemes need the size of the secret audio and generated shares to be the same, or the amplitude of secret audio and generated share should be uniform distribution.

And feature comparisons between our MSS scheme and image secret sharing methods are demonstrated in Table 4, our scheme is no pixel expansion, and can recover the secret

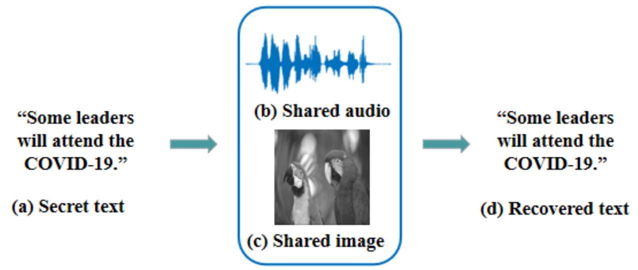


FIGURE 13. Multimedia secret sharing for text (MSS-Text $(2, n)$). Sharing the secret text to n shares, wherein the selected audio "We're leaving on the Abraham Lincoln" and an image size of 64 × 64 are used to recover the secret text. The Gaussian integers of them are respectively are $N_1 = -1.2832 \times 10^{45} + 3.3285 \times 10^{45} i$, $N_2 = -1.5223 \times 10^{90742} - 1.5582 \times 10^{90742} i$, $N_3 = -5.8611 \times 10^{4931} - 5.5898 \times 10^{4931} i$, and $N_4 - 1.2832 \times 10^{45} + 3.3285 \times 10^{45} i$. Based on the sharing (16) $N_1 = \lambda_5 N_2 \times + \lambda_6 \overline{N_3}$, then $\lambda_5 = 1$, $\lambda_6 = 1.8776 \times 10^{85799} + 3.4994 \times 10^{88600} i$. Therefore, λ_5 and λ_6 are used to recover the original secret audio from the shared text and image. The SNR between the secret text and recovered text is infinite($+\infty$).

TABLE 3. Comparison of audio secret sharing schemes.

Scheme	Threshold	The quality of shares	Share type	Constraints
[39]	2	Low	Audio	Yes
[40]	k	Low	Audio	No
[41]	n	High	Audio	Yes
[42]	n	Low	Audio	No
Ours	2	High	Audio,image and text	No

TABLE 4. Comparison of Image secret sharing schemes.

Scheme	Pixel expansion	Recover cover image losslessly	Recover secret image losslessly	The quality of shares	The data type of shares
[33]	Yes	No	Yes	High	Image
[34]	No	Yes	Yes	High	Image
[35]	No	Yes	Yes	Low	Only binary image
[36]	Yes	No	No	High	Image
Ours	No	Yes	Yes	High	Image, audio and text

and cover images accurately. Furthermore, the generated shares are also comprehensible and diverse.

From the above-mentioned experimental results in Figs. (11-13) and comparative evaluations in Table 3 and Table 4, it can be concluded that MSS is successful and reversible. Furthermore, MSS has also the following advantages:

(1) MSS offers a unified secret sharing model, which is suitable for image secret sharing, audio sharing and text sharing. The types of generated shares are various and multimodal.

(2) MSS is a lossless secret sharing strategy. In the sharing process, the quality of generated shares and recovered secret message are high, and their value of SNR or PSNR are all $+\infty$.

(3) The application of MSS is based on geometric operations, and there is no limitation between the secret message and generated shares.

Moreover, this representation method can also be used in other multimodal application, such as scrambling. Fig. 9 is the inverse and conjugate operations for different modalities data. As shown in Fig. 9, it can be found that the Gaussian integer of reconstructed data can be recovered by twice inverse or conjugate operations. Therefore, the multimodal scrambling scheme can be achieved, which supports different modalities of data scrambling. At present, the discussed multimodal applications focus on the field of multimedia information security, how to extend the proposed multimedia representation method to other areas needs to be further explored.

V. CONCLUSION

In this paper, we propose a reversible multimedia representation method based on a complex base. In the proposed method, the unimodal data are respectively represented as Gaussian integers, and mapped into different points on the same representation plane, then a joint multimedia representation model is achieved. Afterwards, based on the geometric operations on the complex plane, the relationships among multimedia data are established. We can perform addition, subtraction, multiplication and other operations on multimedia data. In particular, compared with the traditional elementary arithmetic of digital multimedia, there is no need to consider whether the data types involved are the same and whether they are of the same size.

Experimental results show that the proposed representation method is lossless and reversible. Furthermore, the applications in multimedia camouflage and multimedia secret sharing verify the potential practicability value of the proposed method. More importantly, the proposed representation method breaks the limitation of traditional multimedia information security, which only focuses on single modality data. We replace it with heterogeneous and multimodal multimedia data to achieve a unified security model. In future work, we will consider how to solve the problem of the large order of magnitude of the proposed multimedia representation method, and continue to explore new potential research value.

REFERENCES

[1] M. S. Z.-N. Li, *Fundamentals of Multimedia*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.

[2] X. Shen, F. Shen, L. Liu, Y.-H. Yuan, W. Liu, and Q.-S. Sun, "Multiview discrete hashing for scalable multimedia search," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 5, pp. 1–21, Sep. 2018.

[3] X. Lu, L. Zhu, J. Li, H. Zhang, and H. T. Shen, "Efficient supervised discrete multi-view hashing for large-scale multimedia search," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 2048–2060, Aug. 2020.

[4] Y. Gao, J. Sang, C. Fu, Z. Wang, T. Ren, and C. Xu, "Metadata connector: Exploiting hashtag and tag for cross-OSN event search," *IEEE Trans. Multimedia*, vol. 23, pp. 510–523, 2021.

[5] S. Tian, X. Yin, Y. Su, and H. Hao, "A unified framework for tracking based text detection and recognition from web videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 542–554, Mar. 2018.

[6] X. Yun, Y. Zhang, F. Yin, and C. Liu, "Instance GNN: A learning framework for joint symbol segmentation and recognition in online handwritten diagrams," *IEEE Trans. Multimedia*, vol. 24, pp. 2580–2594, 2022.

[7] M. Koyuncu, A. Yazici, M. Civelek, A. Cosar, and M. Sert, "Visual and auditory data fusion for energy-efficient and improved object recognition in wireless multimedia sensor networks," *IEEE Sensors J.*, vol. 19, no. 5, pp. 1839–1849, Mar. 2019.

[8] W. Zhu, X. Wang, and W. Gao, "Multimedia intelligence: When multimedia meets artificial intelligence," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1823–1835, Jul. 2020.

[9] C. Jewitt, J. Bezemer, and K. O'Halloran, *Introducing Multimodality*. Evanston, IL, USA: Routledge, 2016.

[10] S. Chrisomalis, *Numerical Notation: A Comparative History*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[11] A. R. Plantz and M. Berman, "Adoption of the octal number system," *IEEE Trans. Comput.*, vol. C-20, no. 5, pp. 593–598, May 1971.

[12] L. B. Wadel, "Negative base number systems," *IRE Trans. Electron. Comput.*, vol. 6, no. 2, p. 123, Jun. 1957.

[13] V. S. Dimitrov and G. A. Jullien, "Loading the bases: A new number representation with applications," *IEEE Circuits Syst. Mag.*, vol. 3, no. 2, pp. 6–23, Nov. 2003.

[14] G. Bergman, "A number system with an irrational base," *Math. Mag.*, vol. 31, no. 2, pp. 98–110, 1957.

[15] E. Donald, "An imaginary number system," *Commun. ACM*, vol. 3, no. 4, pp. 245–247, 1960.

[16] C. Heuberger, H. Prodinger, and S. G. Wagner, "Positional number systems with digits forming an arithmetic progression," *Monatshefte Math.*, vol. 155, nos. 3–4, pp. 349–375, Dec. 2008.

[17] W. J. Gilbert, "Complex numbers with three radix expansions," *Can. J. Math.*, vol. 34, no. 6, pp. 1335–1348, Dec. 1982.

[18] I. Katai and J. SzabIO, "Canonical number-systems for complex integers," *Acta Sci. Math.*, vol. 37, nos. 3–4, pp. 255–260, 1975.

[19] W. J. Gilbert, "Fractal geometry derived from complex bases," *Math. Intelligencer*, vol. 4, no. 2, pp. 78–86, Jun. 1982.

[20] W. Sun and D. X. Qi, "Complex number system based algorithm for engineering graph digital watermarking," *Int. J. Image Graph.*, vol. 8, no. 1, pp. 626–630, 2003.

[21] S. Katzenbeisser, *Information Hiding Techniques for Steganography and Digital Watermarking*. Norwood, MA, USA: Artech House, 2000.

[22] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[23] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, Mar. 2006.

[24] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 721–730, Mar. 2007.

[25] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3524–3533, Dec. 2011.

[26] B. Ou, X. Li, W. Zhang, and Y. Zhao, "Improving pairwise PEE via hybrid-dimensional histogram generation and adaptive mapping selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 7, pp. 2176–2190, Jul. 2019.

[27] H. Wu, X. Li, Y. Zhao, and R. Ni, "Improved reversible data hiding based on PVO and adaptive pairwise embedding," *J. Real-Time Image Process.*, vol. 16, no. 3, pp. 685–695, Jun. 2019.

[28] R. Kumar and K.-H. Jung, "Enhanced pairwise IPVO-based reversible data hiding scheme using rhombus context," *Inf. Sci.*, vol. 536, pp. 101–119, Oct. 2020.

[29] W. He, G. Xiong, and Y. Wang, "Reversible data hiding based on adaptive multiple histograms modification," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3000–3012, 2021.

[30] W. He and Z. Cai, "Reversible data hiding based on dual pairwise prediction-error expansion," *IEEE Trans. Image Process.*, vol. 30, pp. 5045–5055, 2021.

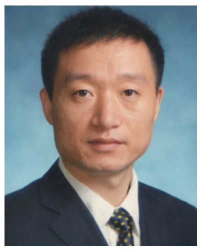
[31] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979.
[32] P.-Y. Lin, J.-S. Lee, and C.-C. Chang, "Distortion-free secret image sharing mechanism using modulus operator," *Pattern Recognit.*, vol. 42, no. 5, pp. 886–895, May 2009.
[33] Y. Liu, C. Yang, Y. Wang, L. Zhu, and W. Ji, "Cheating identifiable secret sharing scheme using symmetric bivariate polynomial," *Inf. Sci.*, vol. 453, pp. 21–29, Jul. 2018.
[34] L. Bao, S. Yi, and Y. Zhou, "Combination of sharing matrix and image encryption for lossless (k, n)-secret image sharing," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5618–5631, Dec. 2017.
[35] X. Yan, Y. Lu, L. Liu, and X. Song, "Reversible image secret sharing," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3848–3858, 2020.
[36] X. Wu and C.-N. Yang, "Partial reversible AMBTC-based secret image sharing with steganography," *Digit. Signal Process.*, vol. 93, pp. 22–33, Oct. 2019.
[37] C.-C. Chang, Y.-P. Hsieh, and C.-H. Lin, "Sharing secrets in stego images with authentication," *Pattern Recognit.*, vol. 41, no. 10, pp. 3130–3137, Oct. 2008.
[38] Y. Q. Shi, G. Xuan, and W. Su, "Lossless data hiding: Fundamentals, algorithms, and applications," in *Multimedia Security Handbook*. Boca Raton, FL, USA: CRC Press, 2004, pp. 531–547.
[39] M. Ehdaie, T. Eghlidos, and M. Reza Aref, "A novel secret sharing scheme from audio perspective," in *Proc. Int. Symp. Telecommun.*, Aug. 2008, pp. 13–18.
[40] S. Vyavahare and S. Patil, "Analysing secret sharing schemes for audio sharing," *Int. J. Comput. Appl.*, vol. 137, no. 11, pp. 39–42, Mar. 2016.
[41] J. Wang, T. Wu, and T. Sun, "An audio secret sharing system based on fractal encoding," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Sep. 2015, pp. 211–216.
[42] S. S. Bharti, M. Gupta, and S. Agarwal, "A novel approach for verifiable (n, n) audio secret sharing scheme," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25629–25657, Oct. 2018.

XIU HE received the B.S. degree from Xinzhou Teachers University, Xinzhou, China, in 2010, and the M.S. degree from the Taiyuan University of Technology, Taiyuan, China, in 2013. She is currently pursuing the Ph.D. degree with the Faculty of Innovation Engineering, School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macau, China. She is also a Lecturer with the School of Information Science, Guangzhou Xinhua University, Guangzhou (Dongguan), China. Her research interests include signal and image processing and multimedia information security.

YUYUN CHEN received the B.S. degree in electrical engineering and automation engineering from the Beijing Institute of Technology, Zhuhai, China, in 2018, and the M.S. degree in computer and information systems from the School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macau, China, in 2020, where she is currently pursuing the Ph.D. degree in computer technology and application with the Faculty of Innovation Engineering, School of Computer Science and Engineering. Her research interests include remote sensing image inpainting and deep learning.

YONG FAN received the bachelor's degree in education technology from Southwest Normal University, in 2002, and the M.Sc. degree in computer application technology from Southwest Petroleum University, in 2014. He is currently pursuing the Ph.D. degree with the Faculty of Innovation Engineering, School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macau, China. His research interest includes image and video processing.

XIAOXI KONG received the M.S. degree in information technology and computer science from the Macau University of Science and Technology, Taipa, Macau, China, in 2020, where she is currently pursuing the Ph.D. degree in computer technology and application. Her current research interests include multimedia information security and image processing.



ZHANCHUAN CAI (Senior Member, IEEE) received the Ph.D. degree in computer software and theory from Sun Yat-sen University, Guangzhou, China, in 2007. He is currently a Professor with the School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China, where he is also with the State Key Laboratory of Lunar and Planetary Sciences. He has authored over 100 articles in refereed journals and conferences. His research interests include image processing and computer graphics, intelligent information processing, multimedia information security, and remote sensing data processing and analysis.