Toward Realizing Talking Quantum Movies: Synchronization of Audio and Visual Content

Fei Yan, Abdullah M. Iliyasu, Kehan Chen and Huamin Yang

***Abstract*—With the ubiquity and primacy of image- and video-processing, a new subdiscipline ensuring a smooth transition from digital to quantum image-processing has emerged. This study proposes a framework to synchronize audio and visual content to realize talking quantum movies. This framework segments a movie into frame, audio, and time components. A multichannel representation for quantum images (MCQI) is used to encode the still images that make up movie frames. In addition, the amplitude content of the flexible representation of the quantum audio (FRQA) signal is recorded at each instant of time. The synchronization of audio and frames in the movie is accomplished through a quantum sequence of time. The feasibility of this framework is demonstrated by a simple simulation, and some possible applications are discussed.**

***Index Terms*—Quantum information, Quantum movie, Quantum image, Quantum audio**

I. Introduction

IN our daily lives, multimedia (text, image, video, and audio) communication and processing are indispensable and ubiquitous. As we approach the triple-s (speed, security, and size) limits of today's technologies, the search for a more powerful computing paradigm has intensified [1]. Quantum computation and quantum information science have transmuted from classroom conjectures and demonstrations to exciting subdisciplines with potentially profound impacts in areas such as information and communication security, device miniaturization, cosmology, and the internet [2]. Recent developments in both areas have further changed the landscape of computing and information-processing technology.

Quantum image processing (QIP) is an emerging framework for multimedia information processing that uses quantum computing technology to capture, manipulate, and recover quantum images, videos, and audio in different formats [3]. Since Vsalov first related quantum mechanics to image processing in 1997 [4], QIP has become an active research subdiscipline with many stimulating outcomes [5]. Most efforts have focused on developing new image formats, as well as algorithmic frameworks for the manipulation of the spatial or chromatic content of images to execute basic image-processing tasks in such applications as image watermarking [6], [7], interpolation [8], [9], and classification [10], [11].

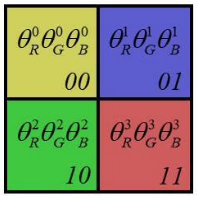
Motivated by the dominance of TV and movies and advances in the QIP subdiscipline [12], Iliyasu et al. proposed a scheme to represent and produce movies on quantum computers in 2011 [13]. It includes a quantum CD, quantum player, and movie reader in one device to meet the requirements of quantum computing, i.e., preparation, manipulation, and measurement. In Iliyasu's quantum movie scheme (QMS), multiple quantum images are stacked and encoded as frames of a movie strip. Based on this, content transition between scenes is conveyed to the audience. While conceived with sensitivity to the intricacies of quantum computing, QMS resembles early movie production, with two limitations on its usability [12]. First, disregarding decoherence and measurement issues in recovering movie content, the entire movie is encoded in greyscale, hence the result would be monochrome (or black and white). Second, QMS makes no provision for storing audio information, which limits it to the production of silent movies.

Yan et al. addressed the above shortcomings with a chromatic framework for quantum movies (CFQM) that integrates color-based descriptions with movie content. Each frame is encoded as a multichannel quantum image (MCQI) state [14]. MCQI descends from Le et al.'s flexible representation for quantum image (FRQI) model [15], which has spurred interest in the QIP subdiscipline. The CFQM framework still relied on the stack-oriented strip to encode the multiple frames that make up the registers encoding the contents of their movies. They retained QMS's structural design, conjectures, nomenclature, and descriptions, and did not address the integration of audio content into quantum movies.

The notion of quantum audio is a recent addition to QIP literature. Two quantum audio models have been proposed. These are Wang's quantum representation of digital audio (QRDA) [16] and Yan et al.'s flexible representation of quantum audio (FRQA) [17]. These developments have heightened interest in the integration of quantum video and audio to realize talking color quantum movies. However, its actualization may be daunting. First, the model must support the extra qubit sequence required to store audio information along the time points tagging the audio content to its amplitude. Similarly, the preferred MCQI-based CFQM movie format must coalesce with audio content in talking quantum movies. This imposes many requirements in terms of temporal synchronization of both types of content [2]. In this study, we consolidate the ideas presented in [1] to propose a quantum

Fei Yan, Kehan Chen and Huamin Yang are with School of Computer Science and Technology, Changchun University of Science and Technology, China (E-mail: yanfei@cust.edu.cn).

Abdullah M. Iliyasu is with Electrical Engineering Department, Prince Sattam Bin Abdulaziz University, Kingdom of Saudi Arabia.



$$\begin{aligned}|I\rangle = \frac{1}{4}[&(\cos\theta_R^0|000\rangle+\cos\theta_G^0|001\rangle+\cos\theta_B^0|010\rangle+\sin\theta_R^0|100\rangle+\sin\theta_G^0|101\rangle+\sin\theta_B^0|110\rangle)\otimes|00\rangle\\ &+(\cos\theta_R^1|000\rangle+\cos\theta_G^1|001\rangle+\cos\theta_B^1|010\rangle+\sin\theta_R^1|100\rangle+\sin\theta_G^1|101\rangle+\sin\theta_B^1|110\rangle)\otimes|01\rangle\\ &+(\cos\theta_R^2|000\rangle+\cos\theta_G^2|001\rangle+\cos\theta_B^2|010\rangle+\sin\theta_R^2|100\rangle+\sin\theta_G^2|101\rangle+\sin\theta_B^2|110\rangle)\otimes|10\rangle\\ &+(\cos\theta_R^3|000\rangle+\cos\theta_G^3|001\rangle+\cos\theta_B^3|010\rangle+\sin\theta_R^3|100\rangle+\sin\theta_G^3|101\rangle+\sin\theta_B^3|110\rangle)\otimes|11\rangle]\end{aligned}$$

Fig. 1. A 2×2 MCQI quantum image and its quantum state (figure adapted from [14])

movie framework that facilitates audio and visual synchronization. This encodes visual information with MCQI and audio with FRQA. Possible transformation and applications based on the proposed framework are discussed. The remainder of this paper is organized as follows. In section II, we review the quantum representations of image, movie, and audio, as well as their basic transformations and applications. In section III, we present and validate the reformulation and preparation of the MCQI-based movie scheme to encapsulate the audiovisual synchronization framework. In section IV, we present an animated example to demonstrate the feasibility of the proposed framework. We draw conclusions and outline possible future research in section V.

II. Quantum Representations of Image, Movie, and Audio

A. Quantum Bits and Quantum Gates

In quantum mechanics, quantum states are usually indicated using the bra-ket notation (also known as Dirac notation), which consists of vertical bars and angle brackets. In such a symbol system, each quantum state is described as a vector in Hilbert space (called a ket), which is defined as

$$|u\rangle = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{n-1} \end{pmatrix},$$

where

$$u_i \in C, \quad i = 0,1,\ldots,n-1. \tag{1}$$

The adjoint of ket is referred to as bra, which is defined as

$$\langle u| = |u\rangle^* = \{u_0^* u_1^* \cdots u_{n-1}^*\}, \tag{2}$$

where u^* is the complex conjugate of u.

Corresponding to the definition of a bit in conventional computation, a qubit, which is a unit vector in two-dimensional Hilbert space, is regarded as the smallest unit of information in a quantum system. It is defined by

$$\psi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle, \tag{3}$$

where α and β are complex numbers.

In order to present the composition of quantum systems, the tensor product (notated as $\otimes$) is used. For example, the tensor product of two matrices:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix},$$

and

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{pmatrix},$$

is

$$A\otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{pmatrix}. \tag{4}$$

In addition, the tensor product of a matrix A for n times is denoted by $A^{\otimes n}$. Similarly, the tensor product of two vectors, $|u\rangle$ and $|v\rangle$, can be expressed as $|u\rangle\otimes|v\rangle$ and shortened to $|u\rangle|v\rangle$ or $|uv\rangle$.

The commonly used quantum gates, such as Hadamard, NOT, and C-NOT gates in Fig. 2, are often employed to break down the complex transform in the quantum circuit model into simpler ones. The gate that applies to k qubits is generally denoted as a $2^k \times 2^k$ unitary matrix. In addition, the number of qubits in the input of the gate must be equal to the number at the output end.

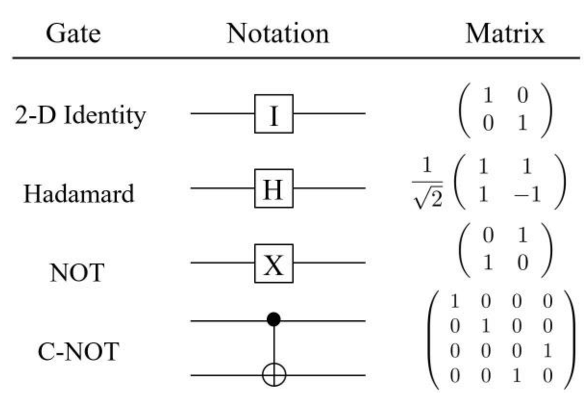


Fig. 2. Commonly used quantum gates

B. Quantum Image Representation

Inspired by the pixel representation of images in conventional computers, a representation for images on quantum computers capturing information about colors and their corresponding positions, the flexible representation of quantum images (FRQI) [15] was proposed in 2010, and has seen broad interest in QIP literature. A multichannel representation for quantum images, MCQI [14], extends the greyscale information in FRQI to color, using the R, G, and B channels to represent color information of images and keeping the normalized state. This is accomplished by assigning three qubits to encode color information of images. Based on MCQI, all of the RGB information about an image is stored simultaneously. This is represented as

$$|I\rangle = \frac{1}{2^{n+1}} \sum_{i=0}^{2^{2n}-1} |c_{RGB\alpha}^{i}\rangle \otimes |i\rangle, \tag{5}$$

where the color information $|c_{RGB\alpha}^{i}\rangle$ of the RGB channels is defined as

$$\begin{aligned} |c_{RGB\alpha}^{i}\rangle = {} & \cos\theta_R^i\,|000\rangle + \cos\theta_G^i\,|001\rangle \\ & + \cos\theta_B^i\,|010\rangle + \cos\theta_\alpha\,|011\rangle \\ & + \sin\theta_R^i\,|100\rangle + \sin\theta_G^i\,|101\rangle \\ & + \sin\theta_B^i\,|110\rangle + \sin\theta_\alpha\,|111\rangle, \end{aligned} \tag{6}$$

where $\{\theta_R^i, \theta_G^i, \theta_B^i\} \in [0, \pi/2]$ are three angles encoding the colors of the R, G, and B channels, respectively, of the i^{th} pixel, and θ_α is set to 0 to make the two coefficients constant ($\cos\theta_\alpha = 1$ and $\sin\theta_\alpha = 0$) to carry no information. An example of a 2×2 MCQI quantum image with its quantum state is presented in Fig. 1.

MCQI has the following advantages [14]:

1) MCQI provides a solution to encoding R, G, and B channel information in normalized quantum states using fewer qubits. Quantum parallelism results in MCQI quantum images requiring far fewer computing resources than RGB images.
2) MCQI facilitates the design of color image operators with lower complexity. The complexity of MCQI-based color information operators is image-size-independent, whereas similar operations on classical computers require color information to be shifted pixel by pixel.
3) Combined with a quantum-measurement-based key-generation procedure, MCQI provides the potential to design color image watermarking/encryption algorithms.

An MCQI quantum image is stored in the preparation process using the MC-PPT theorem with three vectors of angles [14]. MCQI representation shares the same method as FRQI to encode position information, i.e., the same arrangement for position qubits. The notable difference between the two representations is color qubits, because FRQI uses one qubit to encode an image color, whereas MCQI applies three qubits to carry multichannel color information. Further discussion of geometric and color transformations based on MCQI can be found in [18].

C. Quantum Movie Representation

Similar to the extension from FRQI to MCQI, to enhance the canonical elegance of Iliyasu's QMS [13], a framework integrating chromatic descriptions of individual frames (images) and merging them with time information that tags each frame and binds them into a quantum register (i.e., a movie strip) is presented in [12]. The resulting chromatic framework for quantum movies (CFQM), comprising 2^m frames, is then bound to the time tag associated with each frame in the quantum register, which evolves into the quantum movie. The CFQM framework is formulated as

$$|M\rangle = \frac{1}{2^{m/2}} \sum_{t=0}^{2^m-1} |F_t\rangle \otimes |t\rangle, \tag{7}$$

where $|t\rangle$, $t = 0,1,\cdots,2^m - 1$, are 2^m -D quantum basis states representing time information in the movie. m is the number of qubits required to encode the time information. As seen from Eq. (7), there are two essential parts in a CFQM framework: the movie frame $|F_t\rangle$ and its time tag $|t\rangle$, where $|F_t\rangle$ is an MCQI image in the form of Eq. (5).

To demonstrate the intricacies of the CFQM framework, Fig. 3 shows the mathematical representation and circuit network for a simple four-frame movie framework. In this example, each frame is a 2×2 MCQI image, which implies that two qubits are required to encode the spatial information of each frame, and another three qubits to encode the RGB chromatic information. Two more qubits are required to encode the temporal information depicting the time lapse in the movie, i.e., $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. Hence, seven qubits are needed to represent such a movie framework.

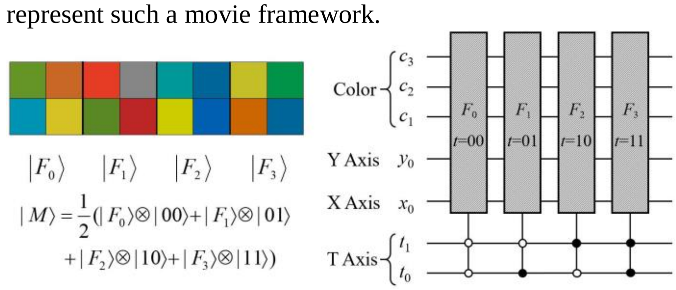


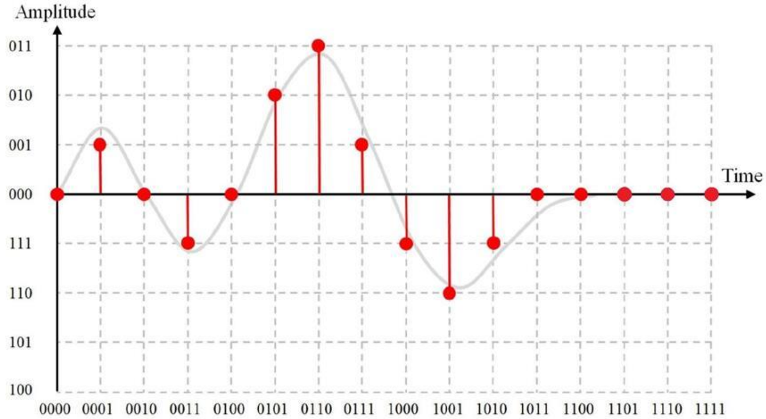
Fig. 3. A four-frame CFQM quantum movie framework (figure adapted from [12])

Quantum movie applications such as moving target detection and quantum image stabilization [19] have been proposed.

D. Quantum Audio Representation

Progress in QIP has also inspired research into quantum audio signal processing, as sound is indispensable to multimedia systems. Wang attempted to kick-start the development of quantum audio processing by proposing the quantum representation for digital audio (QRDA) [16].

The basics of QRDA representation and the outcomes of [16] are a first attempt at audio signal representation and manipulation in the quantum computing domain. The tightly-bounded unipolar encoding strategy used in QRDA may



$$\frac{1}{\sqrt{2^4}}\big(|000\rangle\otimes|0000\rangle+|001\rangle\otimes|0001\rangle+|000\rangle\otimes|0010\rangle +|111\rangle\otimes|0011\rangle+|000\rangle\otimes|0100\rangle+|010\rangle\otimes|0101\rangle +|011\rangle\otimes|0110\rangle+|001\rangle\otimes|0111\rangle+|111\rangle\otimes|1000\rangle +|110\rangle\otimes|1001\rangle+|111\rangle\otimes|1010\rangle+|000\rangle\otimes|1011\rangle +|000\rangle\otimes|1100\rangle+|000\rangle\otimes|1101\rangle+|000\rangle\otimes|1100\rangle +|000\rangle\otimes|1111\rangle\big)$$

Fig. 4. A segment of an FRQA audio signal and its representation (figure adapted from [17])

hinder accurate computation and processing in quantum audio processing, as we explain below.

1) In QRDA representation, amplitude values a_t can only represent non-negative numbers. Hence, some arithmetic operations pertaining to amplitude values are prone to errors. For instance, with two amplitudes a_m and a_n (where $a_m < a_n$), it is virtually impossible to obtain a result for the operation $a_m - a_n$.
2) As a unipolar representation, QRDA is not formulated to display or determine the midrange of the waveform in processing operations. For instance, it is difficult to execute some operations with QRDA representation because all amplitudes are positive (e.g., addition of opposite amplitude values in two waveforms will accumulate to a higher amplitude rather than offsetting each other).

The flexible representation of quantum audio (FRQA) [17] encodes amplitudes in quantum audio in a bipolar (both nonnegative and negative) manner, using binary logic arithmetic to provide tools for effective quantum audio processing.

While one-dimensional, an FRQA signal is described by amplitude and time components, and is written as

$$|A\rangle = \frac{1}{2^{l/2}}\sum_{t=0}^{2^l-1}|S_t\rangle\otimes|t\rangle, \tag{8}$$

where $|S_t\rangle = |S_t^0 S_t^1 \ldots S_t^{q-1}\rangle$ is the two's complement representation of each amplitude value, and $|t\rangle = |t_0 t_1 \ldots t_{l-1}\rangle$, $t_i \in \{0,1\}$, is the corresponding time information. The state $|A\rangle$ is normalized, i.e., $\||A\rangle\| = 1$. It is trivial that, as formalized in Eq. (8), FRQA requires $(q+l)$ qubits to represent quantum audio with 2^l samples.

Figure 4 shows a segment of an audio signal and its representation using the FRQA model. In this example, amplitude values are sampled between -2 and 3, which requires three qubits to store the amplitude information in FRQA audio. If the size of the audio is 16, then $l = 4$. Therefore, seven qubits are required to represent this quantum audio signal. Additional references for quantum audio applications, as regards its steganography, for example, can be found in [20].

III. Audiovisual Synchronization Framework and Initialization for Quantum Movies

To represent and produce a movie with audio and visual content on a quantum computer requires a series of frames, audio amplitude, and time information for each frame. In this section, we introduce this audiovisual synchronization representation of a quantum movie (ASQM) and discuss its initialization.

A. Audiovisual Synchronization Representation

A movie is a series of still images that creates the illusion of movement due to the phi phenomenon [12]. Research of the quantum movie framework is still at an early stage. The audiovisual synchronization framework for quantum movies uses the tensor product of three qubit sequences to store frame content, audio amplitude, and temporal information, synchronizing visual and audio through a time quantum sequence. The ASQM framework is formulated as

$$|M(m,n,q)\rangle = \frac{1}{2^{m/2}}\sum_{t=0}^{2^m-1}|F_t(n)\rangle\otimes|S_t\rangle\otimes|t\rangle, \tag{9}$$

where $|t\rangle$, $t = 0,1,\cdots,2^m-1$, are 2^m-D quantum basis states to represent the time information in the entire movie and m is the number of qubits required to encode the time information. Basically, $|F_t(n)\rangle$, $|S_t\rangle$, and $|t\rangle$ are the three essential parts in the ASQM framework. $|F_t(n)\rangle$ is the frame at $|t\rangle$, which is encoded in MCQI as

$$|F_t(n)\rangle = \frac{1}{2^{n+1}}\sum_{i=0}^{2^{2n}-1}|c_{t,i}\rangle\otimes|i\rangle, \tag{10}$$

where $|c_{t,i}\rangle$ encodes the color information of the image in the R, G, B, and α channel, and is defined as

$$\begin{aligned}|c_{t,i}\rangle &= \cos\theta_{t,i}^R|000\rangle+\cos\theta_{t,i}^G|001\rangle\\ &+\cos\theta_{t,i}^B|010\rangle+\cos\theta_{t,i}^\alpha|011\rangle\\ &+\sin\theta_{t,i}^R|100\rangle+\sin\theta_{t,i}^G|101\rangle\\ &+\sin\theta_{t,i}^B|110\rangle+\sin\theta_{t,i}^\alpha|111\rangle,\end{aligned} \tag{11}$$

where $|000\rangle, |001\rangle, \ldots, |111\rangle$ are 8-D computational bases;

$\{\theta^R_{t,i}, \theta^G_{t,i}, \theta^B_{t,i}, \theta^\alpha_{t,i}\} \in \{0, \pi/2\}$ are four angles encoding the colors of the R, G, B, and α channels, respectively, of the i-th pixel; and $|i\rangle = 0,1,\cdots,2^{2n}-1$ are 2^{2n}-D computational basis states.

In the FRQA quantum audio model [17], amplitude values in quantum audio are bipolar. Studies demonstrate the utility and lucidity of extending the use of two's complement arithmetic to encode amplitude and other details of quantum audio in the ASQM configuration. Consequently, we define the audio amplitude $|S_t\rangle$ of a frame at time $|t\rangle$ as

$$|S_t\rangle = |S_t^0 S_t^1 \ldots S_t^{q-1}\rangle,$$

where

$$S_t^i \in \{0,1\},$$

and

$$i = 0,1,\ldots,q-1,$$

and

$$t = 0,1,\ldots,2^m-1, \tag{12}$$

denotes the time information of the 2^m-sized quantum audio signal, and $|S_t^0 S_t^1 \ldots S_t^{q-1}\rangle$ is the binary sequence encoding the two's complement notation of the amplitude value. Two cases that arise depending on the binary sequence S_t are as follows:

1) If the amplitude is non-negative, then $S_t^0 = 0$, and S_t is simply represented as a binary sequence of the value itself.
2) If the amplitude is negative, then $S_t^0 = 1$, and S_t is represented by the two's complement mode of its absolute value.

As the established norm in QIP, the state $|M(m,n,q)\rangle$ is a normalized one, as seen in Eq. (13):

$$\begin{aligned}
&\||M(m,n,q)\rangle\| \\
&= \frac{1}{2^{m/2}} \sqrt{\sum_{t=0}^{2^m-1} \||F_t(n)\rangle\|^2} \\
&= \frac{1}{2^{m/2+n+1}} \Big[\sum_{t=0}^{2^m-1} \sum_{i=0}^{2^{2n}-1} \cos^2\theta^R_{t,i} + \cos^2\theta^G_{t,i} \\
&\quad + \cos^2\theta^B_{t,i} + \cos^2\theta^\alpha_{t,i} + \sin^2\theta^R_{t,i} + \sin^2\theta^G_{t,i} \\
&\quad + \sin^2\theta^B_{t,i} + \sin^2\theta^\alpha_{t,i}\Big]^{\frac{1}{2}} \\
&= 1.
\end{aligned} \tag{13}$$

Further details regarding the preparation of a quantum image or audio state can be found in [14] and [17]. Figure 5 shows the example of a 2×2 MCQI image (at time $|t_0\rangle = |00\rangle$) that undergoes three sequential 45° clockwise rotations at $|t_1\rangle = |01\rangle$, $|t_2\rangle = |10\rangle$, and $|t_3\rangle = |11\rangle$. Suppose this moving image includes audio with four possible amplitudes. In this case, the corresponding audio amplitudes for $|t_0\rangle$, $|t_1\rangle$, $|t_2\rangle$, and $|t_3\rangle$ are supposed to be $|001\rangle$, $|011\rangle$, $|010\rangle$, and $|111\rangle$, respectively.

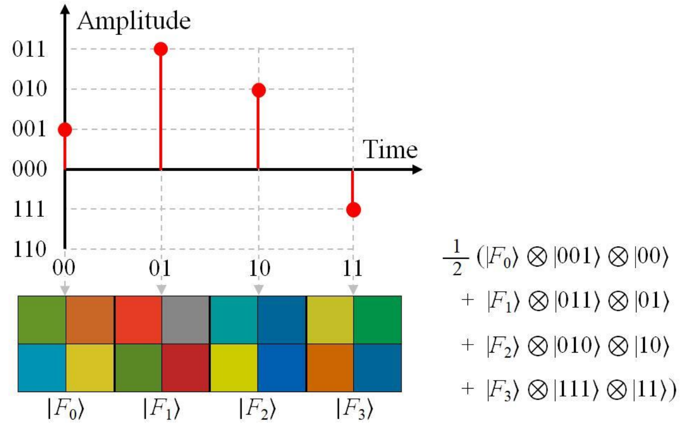


Fig. 5. A four-frame ASQM quantum movie framework and its quantum state.

B. *Initialization of ASQM Framework*

As in other areas of quantum information science, following the preparation of the ASQM format, the initial state must be steered into a desired state. In quantum computation, all of the transforms are unitary and can be described by unitary matrices. A matrix is said to be unitary if its Hermitian conjugate or its adjoint is the same as its inverse. We assume the initialized quantum state is the $m+q+2n+3$ basis state $|0\rangle$ (denoted by $|0\rangle^{\otimes m+q+2n+3}$), where m is the number of qubits used to encode the time information, q is the number of qubits needed to encode the amplitude information, and $2n$ qubits are used to encode the position information of the frame, where each is essentially a $2^n \times 2^n$ sized MCQI image. The remaining three qubits are used to retain chromatic details (RGB) of the movie content. Similar to [12], the preparation of the ASQM framework has the following three steps:

Step 1: Apply the transformation $G = I \otimes H^{\otimes 2n+2} \otimes I^{\otimes q} \otimes H^{\otimes m}$ to the initialized state $|0\rangle^{\otimes m+q+2n+3}$ to generate an intermediate state $|K\rangle$:

$$\begin{aligned}
&G\left(|0\rangle^{m+q+2n+3}\right) \\
&= \left(I \otimes H^{\otimes 2n+2} \otimes I^{\otimes q} \otimes H^{\otimes m}\right)\left(|0\rangle^{m+q+2n+3}\right) \\
&= |0\rangle \otimes \left(H^{\otimes 2}|0\rangle^{\otimes 2}\right) \otimes \left(H^{\otimes 2n}|0\rangle^{\otimes 2n}\right) \\
&\quad \otimes |0\rangle^{\otimes q} \otimes \left(H^{\otimes m}|0\rangle^{\otimes m}\right) \\
&= |0\rangle \otimes \frac{1}{2}\sum_{c=0}^{3}|c\rangle \otimes \frac{1}{2^n}\sum_{i=0}^{2^{2n}-1}|i\rangle \\
&\quad \otimes |0\rangle^{\otimes q} \otimes \frac{1}{2^{m/2}}\sum_{t=0}^{2^m-1}|t\rangle \\
&= |K\rangle,
\end{aligned} \tag{14}$$

using two kinds of unitary matrices (the 2-D identity matrix I and Hadamard matrix H as presented in Fig. 2). $H^{\otimes 2n+2}$ and $H^{\otimes m}$ indicate the tensor product of $2n+2$ and m Hadamard matrices, respectively.

Step 2: Initialize the movie frame, a register of 2^m MCQI

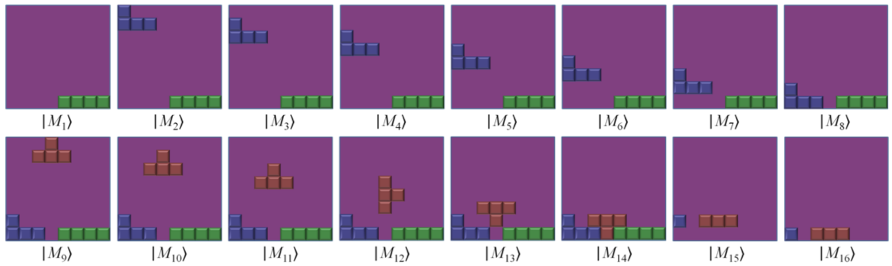


Fig. 6. A short (16-frame) tetris tile-matching puzzle game

images, at time $|t\rangle$. Rotation matrices (rotations around the y-axis of a Bloch sphere by angle 2θ) are utilized in the operation, which is formalized as

$$R_y(2\theta)=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix},$$

where

$$\theta\in\{\theta_{t,i}^R,\theta_{t,i}^G,\theta_{t,i}^B\}. \tag{15}$$

Based on the transform in Eq. (15), $R_{t,i}^R, R_{t,i}^G, R_{t,i}^B$ are three 8×8 controlled rotation operations required to initialize the color information in the R, G, and B channels at position $|i\rangle$, as follows:

$$\begin{aligned} R_{t,i}^R &= I\otimes\sum_{c=1}^{3}|c\rangle\langle c|+R(2\theta_{t,i}^R)\otimes|0\rangle\langle 0|,\\ R_{t,i}^G &= I\otimes\sum_{c=0,c\neq 1}^{3}|c\rangle\langle c|+R(2\theta_{t,i}^G)\otimes|1\rangle\langle 1|,\\ R_{t,i}^B &= I\otimes\sum_{c=0,c\neq 2}^{3}|c\rangle\langle c|+R(2\theta_{t,i}^B)\otimes|2\rangle\langle 2|. \end{aligned} \tag{16}$$

Each of the $R_{t,i}^R$, $R_{t,i}^G$, and $R_{t,i}^B$ operations in Eq. (16) is a three-qubit gate that can be constructed by elementary quantum gates (controlled rotation and C-NOT gates). When we consider these three rotations as a whole, the RGB information of a pixel $|i\rangle$ in a frame can be initialized as

$$R_{t,i}=I^{\otimes 3}\otimes\sum_{j=0,j\neq i}^{2^{2n-1}}|j\rangle\langle j|+(R_{t,i}^R R_{t,i}^G R_{t,i}^B)\otimes|i\rangle\langle i|. \tag{17}$$

As noted earlier, an MCQI image contains $2^n\times 2^n$ pixels. Therefore, executing the operation $R_{t,i}$ in Eq. (17) requires the traversal of all 2^{2n} pixels in the images. Consequently, the operation can be generalized as

$$\Re_t=\prod_{i=0}^{2^{2n}-1}R_{t,i}. \tag{18}$$

The value-setting operation for quantum audio amplitude Ω_t can be defined as

$$\Omega_t=\bigotimes_{i=0}^{q-1}\Omega_t^i,$$

where

$$\Omega_t^i(|0\rangle)=\begin{cases}|\overline{0\oplus B_t^i}\rangle, & i=0\\ |0\oplus B_t^i\rangle, & i\neq 0\end{cases}, \tag{19}$$

where $\oplus$ is the XOR operation. It is apparent that the operation Ω_t^i works by means of an additional C-NOT gate to negate the MSQb (most significant qubit) only when $B_t^i=1$, and otherwise remains unchanged. Hence the amplitude for each sample is

$$\begin{aligned}\Omega_t|0\rangle^{\otimes q} &= \bigotimes_{i=0}^{q-1}(\Omega_t^i|0\rangle)\\ &= |\overline{0\oplus B_t^0}\rangle\otimes\Big(\bigotimes_{i=0}^{q-1}|0\oplus B_t^i\rangle\Big)\\ &= \bigotimes_{i=0}^{q-1}|S_t^i\rangle=|S_t\rangle.\end{aligned} \tag{20}$$

Based on $\Re_t$ and Ω_t, in the ASQM framework, a frame and corresponding audio amplitude tagged at time $|t\rangle$ in the strip can be initialized as

$$H_t=I^{\otimes 2n+q+3}\otimes\sum_{s=0,s\neq t}^{2^m-1}|s\rangle\langle s|+\Re_t\otimes\Omega_t\otimes|t\rangle\langle t|, \tag{21}$$

where H_t is a unitary matrix, since $H_tH_t^{\dagger}=I^{\otimes m+2n+3}$. Specifically, when we initialize the p-th frame in ASQM, the H_p operation is applied on the intermediate state $|K\rangle$ in Eq. (14), and we can obtain

$$\begin{aligned} &H_p(|K\rangle)\\ &=H_p\Big(|0\rangle\otimes\frac{1}{2}\sum_{c=0}^{3}|c\rangle\\ &\quad\otimes\frac{1}{2^n}\sum_{i=0}^{2^{2n}-1}|i\rangle\otimes|0\rangle^{\otimes q}\otimes\frac{1}{2^{m/2}}\sum_{t=0}^{2^m-1}|t\rangle\Big)\\ &=\frac{1}{2^{m/2+n+1}}\Big[|0\rangle\otimes\sum_{c=0}^{3}|c\rangle\otimes\sum_{i=0}^{2^{2n}-1}|i\rangle\otimes|0\rangle^{\otimes q}\\ &\quad\otimes\sum_{t=0,t\neq p}^{2^m-1}|t\rangle+\Re_p\Big(|0\rangle\otimes\sum_{c=0}^{3}|c\rangle\\ &\quad\otimes\sum_{i=0}^{2^{2n}-1}|i\rangle\Big)\otimes\Big(\Omega_p|0\rangle^{\otimes q}\Big)\otimes|p\rangle\Big]\\ &=\frac{1}{2^{m/2+n+1}}\Big[|0\rangle\otimes\sum_{c=0}^{3}|c\rangle\otimes\sum_{i=0}^{2^{2n}-1}|i\rangle\otimes|0\rangle^{\otimes q}\\ &\quad\otimes\sum_{t=0,t\neq p}^{2^m-1}|t\rangle+\prod_{i=0}^{2^{2n}-1}R_{p,i}\Big(|0\rangle\otimes\sum_{c=0}^{3}|c\rangle\\ &\quad\otimes\sum_{i=0}^{2^{2n}-1}|i\rangle\Big)\otimes\Big(\bigotimes_{d=0}^{q-1}\Omega_p^d|0\rangle\Big)\otimes|p\rangle\Big]. \end{aligned} \tag{22}$$

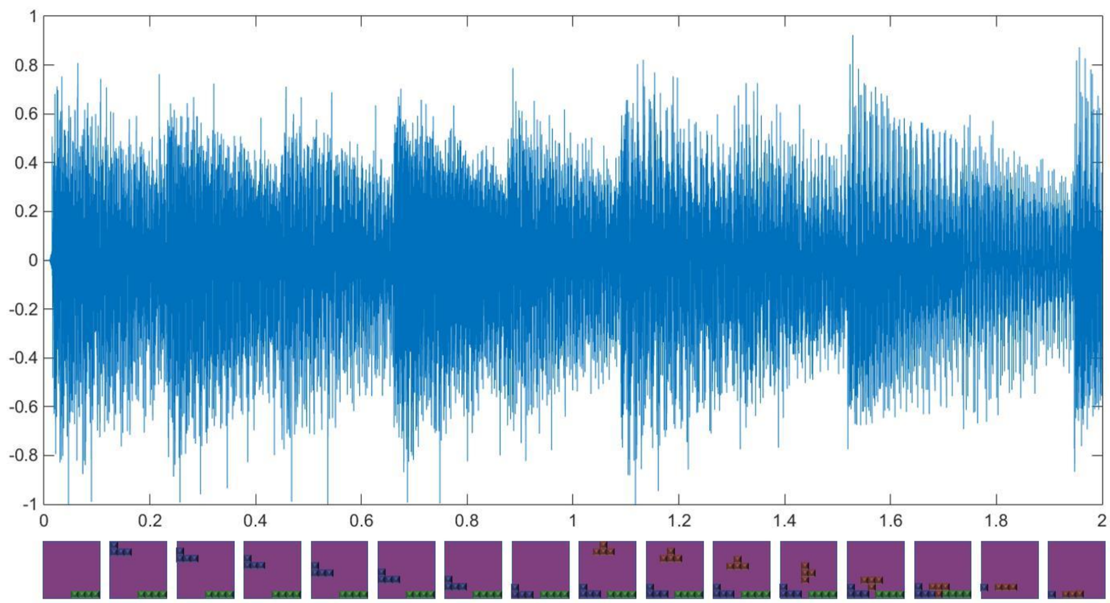


Fig. 7. Presentation of two-second, 16-frame ASQM movie segment

So far, in the strip of the entire ASQM, the image and audio amplitude of the p-th frame have been initialized, while the remaining frames are uninhabited, i.e., they contain position and time information without the color and audio content.

Step 3: As mentioned earlier, an ASQM framework is composed of 2^m MCQI images and their temporally corresponding audio amplitudes. Therefore, the preparation of the proposed ASQM talking quantum movie can be divided into 2^m sub-operations, and each MCQI image and audio amplitude can be prepared through a unique sub-operation. Like the operation in Eq. (22), when we continue to initialize the second through r-th frame (and audio amplitude) in the ASQM format, an H_r operation is applied to the quantum states in Eq. (22) to produce

$$\begin{aligned}
&H_r H_p |K\rangle \\
&= H_r\left(H_p |K\rangle\right) \\
&= \frac{1}{2^{m/2+n+1}}\Big[|0\rangle \otimes \sum_{c=0}^{3}|c\rangle \otimes \sum_{i=0}^{2^{2n}-1}|i\rangle \otimes |0\rangle^{\otimes q} \\
&\otimes \sum_{\substack{t=0,\\ t\neq r, t\neq p}}^{2^m-1}|t\rangle + \Re_r\Big(|0\rangle \otimes \sum_{c=0}^{3}|c\rangle \otimes \sum_{i=0}^{2^{2n}-1}|i\rangle\Big) \\
&\otimes \left(\Omega_r |0\rangle^{\otimes q}\right) \otimes |r\rangle + \Re_p\Big(|0\rangle \otimes \sum_{c=0}^{3}|c\rangle \otimes \sum_{i=0}^{2^{2n}-1}|i\rangle\Big) \\
&\otimes \left(\Omega_p |0\rangle^{\otimes q}\right) \otimes |p\rangle\Big].
\end{aligned} \tag{23}$$

Since ASQM comprises 2^m frames, the procedure to transform the quantum computer from the initialized state ($|0\rangle^{\otimes m+q+2n+3}$) to the desired ASQM state $|M(m,n,q)\rangle$ is formalized as

$$\begin{aligned}
&\prod_{t=0}^{2^m-1} H_t G(|0\rangle^{\otimes m+q+2n+3}) \\
&= \frac{1}{2^{m/2}} \sum_{t=0}^{2^m-1} |F_t(n)\rangle \otimes |S_t\rangle \otimes |t\rangle \\
&= |M(m,n,q)\rangle.
\end{aligned} \tag{24}$$

The initialization procedure outlined thus far precedes the ASQM state storage and retrieval. Whereas the storage process is an intricate part of the initialization process, the retrieval procedure is not that straightforward. A measurement operation uses a set of measurement basis vectors to produce one outcome (also a basis vector). However, in quantum computations, between the quantum system and its surrounding environment or attempts to observe its content leads to its collapse and, with that, the loss of any superposition between its states.

A suggested solution is to apply multiple measurements on many identical copies of the quantum states [21]. Alternatively, ancilla-driven techniques have been suggested to steer the needed interaction with the movie content and, as such, the final measurements leave the movie state undisturbed [13].

In ASQM, an adjusted version of the latter measurement technique is assumed. The color information of each frame is encoded in a three-qubit state $|c_3c_2c_1\rangle$. Hence to retrieve the color information, i.e. the eight coefficients of $|c_3c_2c_1\rangle$, requires multiple measurements on these three qubits to obtain the probability for each state. Specifically, the angle $\theta_{t,i}^X$ encodes the greyscale value in the X channel, and is controlled by c_2 and c_1 when they are in the states $|00\rangle$, $|01\rangle$, and $|10\rangle$ (where X is used to represent R, G, and B channels, respectively). Multiple measurements on the state of c_3 will retrieve a result of either 0 with probability $\cos^2\theta_{t,i}^X$ or

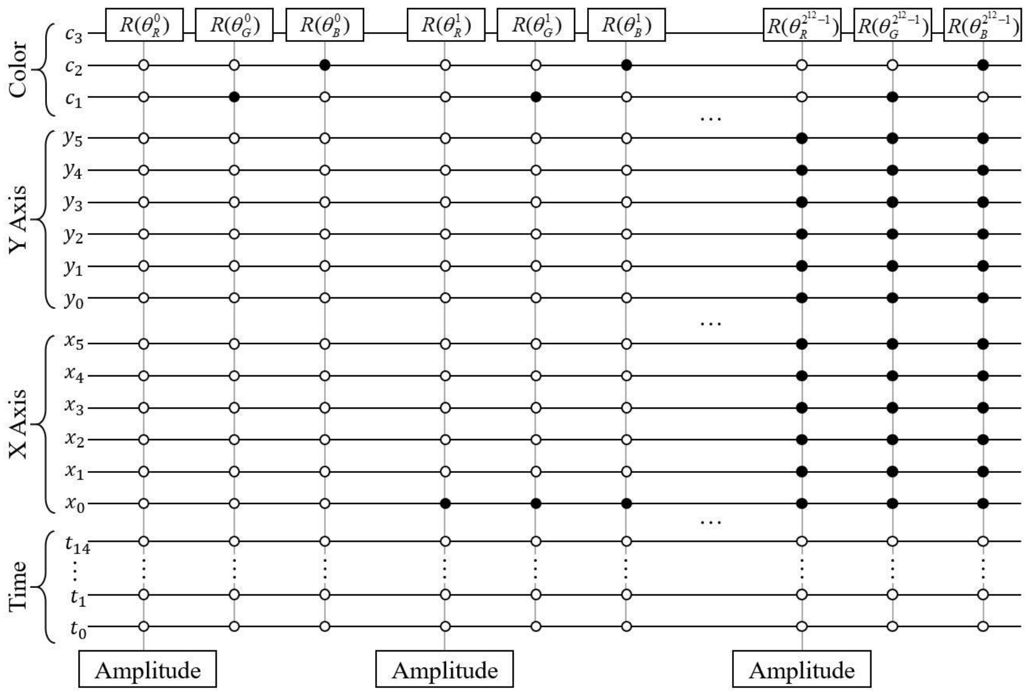


Fig. 8. Quantum circuit representation of a single frame in the movie segment in Fig. 7

1 with probability $\sin^2 \theta_{t,i}^X$. Based on the probability, we can retrieve the greyscale value of the color information in the X channel. In terms of the audio content, since two's complement arithmetic was used to compile it as binary gate sequences, where each measurement produces a special temporally-labelled amplitude information. By performing both measurement operations iteratively, all of the quantum information in the ASQM framework can be retrieved.

IV. Examples to Illustrate the Feasibility of the Proposed ASQM Framework

A short (16-frame) tetris tile-matching puzzle game (shown in Fig. 6) is used to depict the movie content of a simulation-based execution of the ASQM framework. All of these frames have 64×64 resolution, and to represent them in the MCQI format requires 15 qubits; 12 qubits record their spatial location in the movie strip, and three qubits record their chromatic details.

A two-second audio with 88176 samples is shown in Fig. 7. Under the audio are 16 images of a tetris tile-matching puzzle game corresponding to the audio. To complete the algorithm, the two-second audio should be divided into 88176 subintervals. Therefore, 17 qubits are the minimum requirement to represent the temporal details of the clip. During the two-second interval, the audio varies in the range 1 to -1. Consequently, using four decimal places, 15 qubits are required to represent the amplitude information of the clip. Factoring in terms of the movie content, the 16 frames require 47 qubits to encode and manipulate the information.

The audio content for this movie clip consists of 88176 samples, each encoded in the FRQA quantum audio format. Its audio and movie content are presented in Fig. 7. As outlined earlier, the audio content is divided into 88176 sub-intervals, one for each quantum audio sample. Since each subinterval should be placed an image, and the number of amplitudes as well as images to build this movie are 88176 and 16, respectively, each image should be repeated 5511 times.

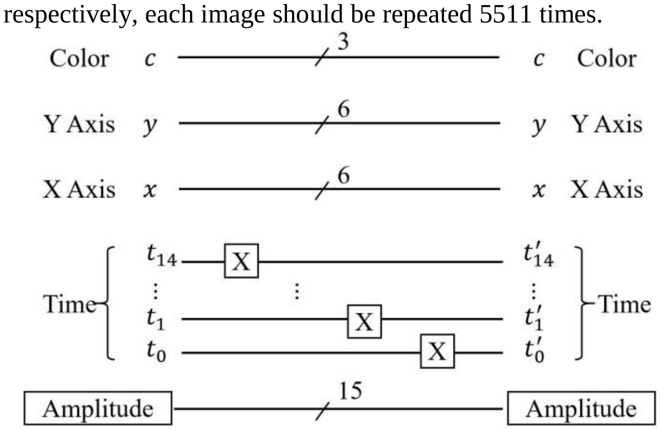


Fig. 9. Quantum circuit implementation of reversal operation on the movie segment in Fig. 7

Putting it all together, the quantum circuit to present a single frame of the tetris tile-matching puzzle game is shown in Fig. 8. Since the circuit only shows a single frame, the "time" qubits $|t_0 t_1 \ldots t_{14}\rangle$ and the amplitude value are unchanged during its implementation. In addition, $|x_0 x_1 x_2 x_3 x_4 x_5\rangle\ |y_0 y_1 y_2 y_3 y_4 y_5\rangle$, ranging from $|000000\rangle|000000\rangle$ to $|111111\rangle|111111\rangle$, encodes all the position information along the X axis and Y axis of the frame, and $|c_1 c_2 c_3\rangle$ stores the color information of each pixel. Three cases of $|c_1 c_2\rangle$ ($|00\rangle$, $|10\rangle$, and $|01\rangle$) are taken as the control qubits of the "$R(\theta_R)$", "$R(\theta_G)$", and "$R(\theta_B)$"

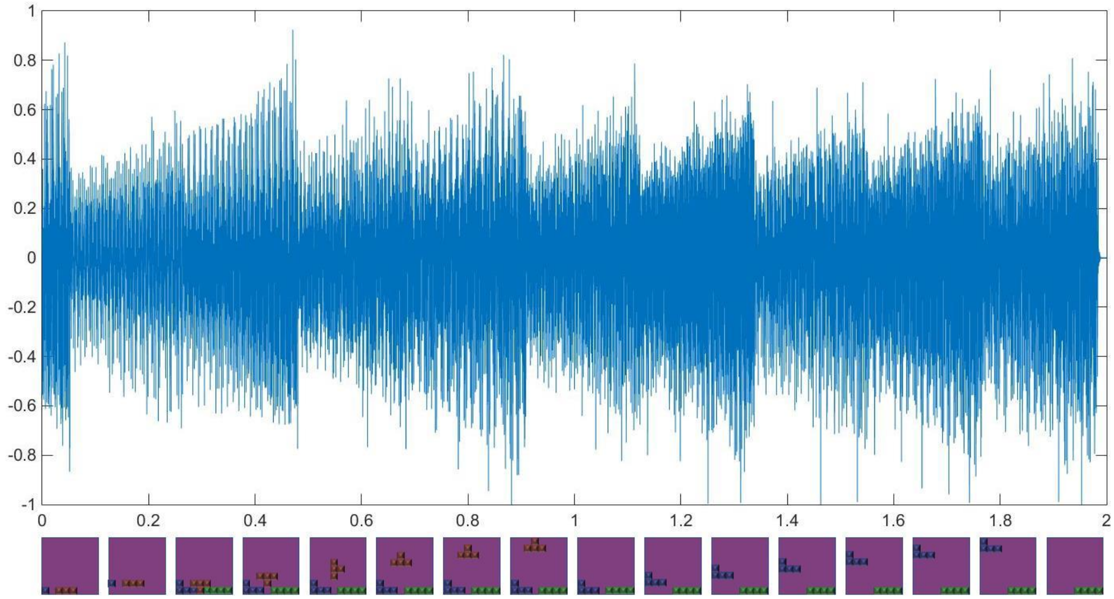


Fig. 10. Presentation of the movie segment in Fig. 7 after the reversal transformation

operations on $|c_3\rangle$. After 2^n controlled rotation operations (discussed in Eq. (16)) traversing all the pixels in a frame, the color information is initialized.

Some kinds of operations can be designed based on the ASQM quantum movie framework to generate more practical and interesting movie applications. Movie reversal is the process of reversing a selected movie segment such that the end of the movie is watched and heard first, and the beginning last. This can create interesting movie effects or make small portions of inappropriate visual content unintelligible. The reversal operation can be achieved by inverting the time line of a movie segment. If the time content includes *m* qubits, then the reversal of a movie is a series of exchange operations on the time information, i.e., $|0\rangle \leftrightarrow |2^m-1\rangle$, $|1\rangle \leftrightarrow |2^m-2\rangle$, …, $|2^{m/2}-1\rangle \leftrightarrow |2^{m/2}\rangle$, where "$\leftrightarrow$" indicates the exchange operation. It is noteworthy that the quantum NOT gate can change the state of a qubit from $|0\rangle$ to $|1\rangle$, and vice versa, to easily achieve the exchange operation. The new and original positions of the time information follow the relation $p_{new}+p_{old}=2^m-1$, which is achieved by applying 2^m quantum NOT gates on the wires encoding the time information of the ASQM framework.

The state $|t_0t_1\ldots t_{14}\rangle$ in Fig. 9 shows the time information of an episode of a movie, where a sequence of quantum NOT gates (i.e., "X" operations in Fig. 2) is applied on each qubit of the time information $|t_0t_1\ldots t_{14}\rangle$. During the circuit flows, the color information $|c\rangle$, position information $|y\rangle$ and $|x\rangle$, and amplitude information are unchanged.

A simulation experiment shows the quantum reversal operation to the quantum movie represented by the ASQM framework. The movie shown in Fig. 7 is reversed using the circuit shown in Fig. 9. Each qubit of time information in Fig. 9 is operated by a quantum NOT gate. If the qubit state is $|0\rangle$, then it will be transformed to $|1\rangle$, and if the qubit state is $|1\rangle$, then it will be reset to $|0\rangle$. After the conversion of each qubit, all the time information will be inverted. The simulation result is shown in Fig. 10. The order of the quantum audio and images are completely opposite to that in Fig. 7, which proves that the reversal operation works well.

V. Concluding Remarks

In this study, we proposed an audiovisual synchronization framework for quantum movies (ASQM) and discussed the requirements for its initialization and for the retrieval of both media types. Compared to previous studies, the proposal provides a first step in the quest to extend the separate quantum audio models and CFQM quantum movie formats into a single framework that encapsulates talking quantum movies.

In future work, more operations should be developed based on the audio contents in the ASQM framework to facilitate multimedia applications. First, like digital watermarking, quantum watermarking aims to protect the copyright of a quantum image and authenticate its ownership using visible or invisible quantum signals embedded in the quantum carrier image. Watermarking technologies may be applied on the ASQM movie framework in terms of its frame, audio, or both. Second, in quantum computing, the complexity of quantum algorithms is usually computed in terms of quantum gates. We may consider using fewer quantum gates or operations to prepare and manipulate the ASQM movie framework, or provide a compression strategy to use fewer quantum computing resources.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 61502053) and the Science & Technology Development Program of Jilin Province, China (No. 20170520065JH).

REFERENCES

[1] F. Yan, A. M. Iliyasu, S. Jiao, and H. Yang, "Audio-visual synchronisation in quantum movies," IEEE Congress on Information Science and Technology (CIST), pp. 274-278, 2018.
[2] A. M. Iliyasu, "Roadmap to talking quantum movies: a contingent inquiry," IEEE Access, doi: 10.1109/ACCESS.2018.2871107, 2018.
[3] F. Yan, A. M. Iliyasu, and S. E. Venegas-Andraca, "A survey of quantum image representations," Quantum information Processing, vol. 15, no. 1, pp. 1-35, 2016.
[4] A. Y. Vlasov, "Quantum computations and images recognition," arXiv: quant-ph/9703010, 1997.
[5] F. Yan, A. M. Iliyasu, and P. Q. Le, "Quantum image processing: a review of advances in its security technologies," International Journal of Quantum Information, vol. 15, 1730001, 2017.
[6] P. Li, H. Xiao, and B. Li, "Quantum representation and watermark strategy for colour images based on the controlled rotation of qubits," Quantum Information Processing, vol. 15, no. 11, pp. 4415-4440, 2016.
[7] R. Zhou, P. Yang, X. Liu, and H. Ian, "Quantum colour image watermarking based on fast bit-plane scramble and dual embedded," International Journal of Quantum Information, vol. 16, no. 7, 1850060, 2018.
[8] N. Jiang and L. Wang, "Quantum image scaling using nearest neighbor interpolation," Quantum Information Processing, vol. 14, no. 5, pp. 1559-1571, 2015.
[9] R. Zhou, W. Hu, G. Luo, X. Liu, and P. Fan, "Quantum realization of the nearest neighbor value interpolation method for INEQR," Quantum Information Processing, vol. 17, 166, 2018.
[10] H. Li, P. Fan, H. Xia, H. Peng, and S. Song, "Quantum implementation circuits of quantum signal representation and type conversion," IEEE Transactions on Circuits and System I: Regular Papers, vol. 16, no. 1, pp. 341-354, 2018.
[11] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features," Knowledge-Based Systems, vol. 141, pp. 80-91, 2018.
[12] F. Yan, S. Jiao, A. M. Iliyasu, and Z. Jiang, "Chromatic framework for quantum movies and applications in creating montages," Frontiers of Computer Science, vol. 12, no. 4, pp. 736-748, 2018.
[13] A. M. Iliyasu, P. Q. Le, F. Dong, and K. Hirota, "A framework for representing and producing movies on quantum computers," International Journal of Quantum Information, vol. 9, no. 6, pp. 1459-1497, 2011.
[14] B. Sun, A. M. Iliyasu, F. Yan, F. Dong, and K. Hirota, "An RGB multi-channel representation for images on quantum computers," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 17, no. 3, pp. 404-417, 2011.
[15] P. Q. Le, F. Dong, and K. Hirota, "A flexible representation of quantum images for polynomial preparation, image compression, and processing operations," Quantum Information Processing, vol. 10, no. 1, pp. 63-84, 2011.
[16] J. Wang, "QRDA: quantum representation of digital audio," International Journal of Theoretical Physics, vol. 55, no. 3, pp. 1622-1641, 2016.
[17] F. Yan, A. M. Iliyasu, Y. Guo, and H. Yang, "Flexible representation and manipulation of audio signals on quantum computers," Theoretical Computer Science, vol. 752, pp. 71-85, 2018.
[18] B. Sun, A. M. Iliyasu, F. Yan, J. A. Garcia, F. Dong, A. K. Al-Asmari, and K. Hirota, "Multi-channel information operations on quantum images," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 18, no. 2, pp. 140-149, 2014.
[19] F. Yan, A. M. Iliyasu, H. Yang, and K. Hirota, "Strategy for quantum image stabilization," Science China Information Sciences, vol. 59, 052102, 2016.
[20] K. Chen, F. Yan, A. M. Iliyasu, and J. Zhao, "Exploring the implementation of steganography protocols on quantum audio signals," International Journal of Theoretical Physics, vol. 57, no. 2, pp. 476-494, 2018.
[21] F. Yan, A. M. Iliyasu, P. Q. Le, B. Sun, F. Dong, and K. Hirota, "A parallel comparison of multiple pairs of images on quantum computers," International Journal of Innovative Computing and Applications, vol. 5, no. 4, pp. 199-212, 2013.