

Quantum algorithm for visual tracking

Chao-Hua Yu,^{1,2,*} Fei Gao,¹ Chenghuan Liu,³ Du Huynh,³ Mark Reynolds,³ and Jingbo Wang^{2,†}

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

²Department of Physics, The University of Western Australia, Perth 6009, Australia

³Department of Computer Science and Software Engineering, The University of Western Australia, Perth 6009, Australia



(Received 2 July 2018; published 1 February 2019)

Visual tracking (VT) is the process of locating a moving object of interest in a video. It is a fundamental problem in computer vision, with various applications in human-computer interaction, security and surveillance, robot perception, traffic control, etc. In this paper, we address this problem in the quantum setting, and present a quantum algorithm for VT based on the framework proposed by Henriques *et al.* [IEEE Trans. Pattern Anal. Mach. Intell. **7**, 583 (2015)]. Our algorithm comprises two phases: training and detection. In the training phase, in order to discriminate the object and background, the algorithm trains a ridge regression classifier in the quantum state form where the optimal fitting parameters of ridge regression are encoded in the amplitudes. In the detection phase, the classifier is then employed to generate a quantum state whose amplitudes encode the responses of all the candidate image patches. The algorithm is shown to be polylogarithmic in scaling, when the image data matrices have low condition numbers, and therefore may achieve exponential speedup over the best classical counterpart. However, only quadratic speedup can be achieved when the algorithm is applied to implement the ultimate task of Henriques's framework, i.e., detecting the object position. We also discuss two other important applications related to VT: (1) object disappearance detection and (2) motion behavior matching, where much more significant speedup over the classical methods can be achieved. This work demonstrates the power of quantum computing in solving computer vision problems.

DOI: [10.1103/PhysRevA.99.022301](https://doi.org/10.1103/PhysRevA.99.022301)

I. INTRODUCTION

Visual tracking (VT) is the task of locating a moving object of interest in a video. It is a fundamental problem in computer vision and has wide applications [1–5] across human-computer interaction, security and surveillance, robot perception, traffic control, and medical imaging. In recent years, a successful approach for VT, which attracts wide attention, has been tracking by detection, where discriminative machine learning classifiers are adopted to detect the object. In this approach, every pair of time contiguous frames respectively undergoes two phases: training and detection. For convenience, hereafter we call the frame used in the training phase the *training frame*, where the location of the object is determined (given in the initial frame or detected in the following frames), and we call the frame used in the detection phase the *detection frame*, where the location of the object remains to be determined. In the training phase, by using discriminative machine learning algorithms, a number of image patches (samples) around the object are selected from the training frame to train a classifier that can discriminate between the object and its background. In the detection phase, the classifier is then performed on several candidate patches selected from the detection frame to calculate their responses. The maximum response reveals the most probable position of the object. The training-detection procedure is run succes-

sively over every frame to track the object in the whole video, during which the detection frame becomes the training frame once the object is detected. The whole process is depicted in Fig. 1. Despite various advanced and fast algorithms [1,2], VT can be time consuming when the processed data size is large.

Quantum computing provides a paradigm that makes use of quantum-mechanical principles, such as superposition and entanglement, to perform computing tasks in quantum systems (quantum computers) [6,7]. The most exciting thing about quantum computing is that it can achieve significant speedup over classical computing, in solving certain problems, such as simulating quantum systems [8], factoring large numbers [9], unstructured database searching [10,11], and solving linear systems of equations by Harrow *et al.* (HHL algorithm) [12]. In recent years, the applicability of quantum computing has been extended to the fields of machine learning and data mining, resulting in a variety of related quantum algorithms [13–19] for solving various machine learning and data mining problems, such as linear regression [13–15], support vector machine [17], and association rules mining [18]. These algorithms are shown to achieve significant speedup over their classical counterparts. Overviews on the recent progress in this field can be found in the references [20,21].

Motivated by the progress in VT as well as quantum computing, and especially in machine learning, we explore whether and how quantum computing can be exploited to implement VT more efficiently than classical computing. In particular, we propose a quantum algorithm for VT based on the well-known tracking-by-detection VT framework proposed by Henriques *et al.* in 2015 (HCMB15 framework) [1].

*quantum.ych@gmail.com

†jingbo.wang@uwa.edu.au

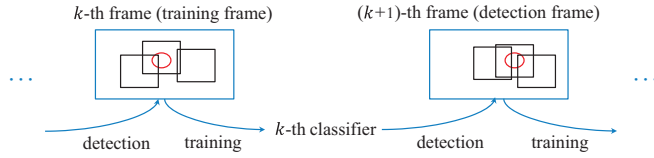


FIG. 1. Schematic of visual tracking by training and detection. The (red) circles represent the object in the training and detection frame. The (black) rectangles in the training frame denote sample image patches for training, and those in the detection frame denote the candidate image patches for detection. The training-detection procedure is run successively over contiguous frames to track the object in the video.

The earlier version of this framework is given in Ref. [2]. In the training phase, the HCMB15 framework utilizes a *base sample patch* to subtly produce a large number of *virtual sample patches* which can be represented by a circulant data matrix, and use these samples to train a ridge regression classifier. In the detection phase, it uses a *base candidate patch* to generate a large number of *virtual candidate patches*, which can also be represented by a circulant matrix, performs the classifier on those candidate patches, and obtains their corresponding responses. The most probable location of the object is revealed by the candidate patch with the maximum response. The circulant structure of the data matrix has been cleverly exploited in the HCMB15 framework and makes it extremely efficient in both training and detection phases.

Our quantum algorithm also comprises the training phase and the detection phase. In the training phase, a ridge regression classifier is trained in the quantum state form, where the optimal fitting parameters of ridge regression are encoded in the amplitudes. In the detection phase, the classifier is then utilized to generate a quantum state whose amplitudes encode the responses of all the candidate image patches. The whole algorithm is built on the proposed subroutine of simulating extended circulant Hamiltonians, which allows the estimation of the singular values of generic circulant matrices by phase estimation. Our algorithm generates both states in polylogarithmic time when the data matrices have low condition numbers, demonstrating its potential to achieve exponential speedup over the classical counterpart. Moreover, we also show that the output quantum state of our algorithm can be fully used to efficiently implement two important computer vision tasks: object disappearance detection and motion behavior matching.

The rest of the paper is organized as follows. In Sec. II, we review the HCMB15 framework in terms of its basic concepts, notations, and classical algorithmic procedures. Section III presents and analyzes our quantum algorithm. Section IV gives two important applications of our algorithm. Finally, conclusions are drawn in the last section.

II. HCMB15 FRAMEWORK

In this section, we review the basic idea and algorithms in the HCMB15 framework [1,2]. For simplicity, we only consider one-dimensional images with single channel. The generalization to two-dimensional multiple-channel images can be seen in [1,2].

The HCMB15 framework is one of the excellent candidates for implementing VT. In the training phase, it takes a base sample image patch of the training frame (where the object is usually placed at the center) with n pixels that can be represented by a vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, the size of which is, for example, two times that of the object. By cyclic shifting \mathbf{x} , it can be used to generate n virtual samples corresponding to a circulant matrix

$$X = \mathcal{C}(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_n & x_1 & x_2 & \cdots & x_{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_2 & x_3 & x_4 & \cdots & x_1 \end{bmatrix}, \quad (1)$$

where $\mathcal{C}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ is a function which generates the circulant matrix for a given vector \mathbf{x} . The i th row of X is denoted by \mathbf{x}_i with $\mathbf{x}_1 = \mathbf{x}$, and corresponds to the i th training sample.

In addition, each sample is assigned a *label* (or *regression target*), a positive value ranging from 0 to 1, to quantify the closeness between the sample and the base sample; the value approximates to 1 if the sample is close to the base sample, and reduces to zero as the distance between them increases. The label of \mathbf{x}_i , denoted by y_i , is commonly derived using the Gaussian function

$$y_i = e^{-d_i^2/s^2}, \quad (2)$$

where d_i is the Euclidean distance of the i th sample to the base sample in the image and s is the bandwidth and is commonly taken as $s = c\sqrt{n}$ for some constant c .

In the training phase, the goal is to train a linear function $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ by *ridge regression*, which minimizes the squared error over samples \mathbf{x}_i and y_i ,

$$\min_{\mathbf{w}} \sum_{i=1}^n |f(\mathbf{x}_i) - y_i|^2 + \alpha \|\mathbf{w}\|^2, \quad (3)$$

where α is the regularization parameter that controls overfitting. The solution is given by

$$\mathbf{w} = (X^T X + \alpha I)^{-1} X^T \mathbf{y}, \quad (4)$$

where $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ is the vector of the regression targets of all n samples, and X is called the *data matrix* [1]. After the solution \mathbf{w} is obtained, one can predict the response of a new image patch $\hat{\mathbf{x}}$ by calculating $f(\hat{\mathbf{x}}) = \mathbf{w}^T \hat{\mathbf{x}}$.

The HCMB15 framework takes advantage of the property of the circulant matrix X that it can be written as $X = F \text{diag}[\mathcal{F}(\mathbf{x})] F^\dagger$. Here F is the unitary Fourier transformation matrix, $\mathcal{F}(\mathbf{x}) = \sqrt{n} F \mathbf{x}$ is the discrete Fourier transformation (DFT), and $\text{diag}(\mathbf{z})$ is the diagonal matrix formed by \mathbf{z} . As a result, the solution [Eq. (4)] can be efficiently obtained with time complexity $O(n \log(n))$ [1], which is significantly faster than the currently prevalent method by matrix inversion and products that has time complexity $O(n^3)$.

In the detection phase, a base candidate image patch of the detection frame denoted by a n -dimensional vector \mathbf{z} is given, and is used to generate n virtual candidate patches corresponding to the $n \times n$ matrix $Z = \mathcal{C}(\mathbf{z})$ with the i th row corresponding to the i th candidate patch. Then the responses

of these patches are predicted by

$$\hat{\mathbf{y}} = \mathbf{Z}\mathbf{w}, \quad (5)$$

where the i th element of $\hat{\mathbf{y}}$ corresponds to the response of i th candidate patch. The index with maximum response in $\hat{\mathbf{y}}$ reveals the target image patch that gives the best estimated position of the object and serves as the new base sample for the next training-detection procedure.

III. QUANTUM ALGORITHM

In this section, we present a quantum algorithm for VT based on the HCMB15 framework with focus on one-dimensional single-channel images. Just as the classical HCMB15 framework, our algorithm also comprises two phases: training and detection. In the first place, we propose a technique of simulating the extended circulant Hamiltonians in Sec. III A. By taking this technique as the basic subroutine, we construct our quantum algorithm to prepare the quantum state $|\mathbf{w}\rangle$ (normalized \mathbf{w}) in the training phase shown in Sec. III B, and then ultimately generate the state $|\hat{\mathbf{y}}\rangle$ (normalized $\hat{\mathbf{y}}$) in the detection phase shown in Sec. III C. We perform runtime analysis on our algorithm in Sec. III D, and finally extend it to two-dimensional single-channel images in Sec. III E.

A. Extended circulant Hamiltonian simulation

Since Eq. (1) describes a general $n \times n$ circulant matrix, we let X denote an arbitrary $n \times n$ circulant matrix. In the case where X is Hermitian, i.e., $X = X^\dagger$, Zhou and Wang [22] proposed an efficient quantum algorithm that uses the unitary linear decomposition approach [23] to implement e^{-iXt} within spectral-norm error ϵ in time $O(t \text{ polylog}(n) \log(t/\epsilon) / \log \log(t/\epsilon))$, under the assumptions that the quantum oracle $O_x |0\rangle^{\otimes \lceil \log n \rceil} = \sum_{i=1}^n \sqrt{x_i} |i\rangle$ that can be efficiently implemented in time $O(\text{polylog}(n))$ is provided, and $\sum_{i=1}^n x_i = 1$. This algorithm is based on the observation that X can be written as a linear combination of n efficient-to-implement unitary operators, namely

$$X = \sum_{j=1}^n x_j V_j, \quad (6)$$

where each $V_j = \sum_{l=0}^{n-1} |(l-j+1) \bmod n\rangle \langle l|$ for $j = 1, 2, \dots, n$, and can be implemented using $O(\log n)$ one- or two-qubit gates.

However, X in general is not Hermitian and cannot be simulated directly by the algorithm [22]. To overcome this, we take the *extended circulant Hamiltonian* of X ,

$$\begin{aligned} \tilde{X} &= |0\rangle \langle 1| \otimes X + |1\rangle \langle 0| \otimes X^\dagger \\ &= \begin{bmatrix} 0 & X \\ X^\dagger & 0 \end{bmatrix}. \end{aligned} \quad (7)$$

It is interesting and easy to see that \tilde{X} can also be written as a linear combination of simple unitary operators:

$$\tilde{X} = \sum_{j=1}^n x_j (|0\rangle \langle 1| \otimes V_j + |1\rangle \langle 0| \otimes V_j^\dagger)$$

$$\begin{aligned} &= \sum_{j=1}^n x_j (\sigma_X \otimes I) (|0\rangle \langle 0| \otimes V_j + |1\rangle \langle 1| \otimes V_j^\dagger) \\ &= \sum_{j=1}^n x_j \tilde{V}_j, \end{aligned} \quad (8)$$

where σ_X is the Pauli- X gate (or NOT gate). Therefore, following Zhou and Wang [22], we can also use the unitary linear decomposition approach [23] to design an efficient quantum algorithm that implements $e^{-i\tilde{X}t}$ within some (spectral-norm) error ϵ . We present the result in the following theorem.

Theorem 1. (Extended circulant Hamiltonian simulation). There exists a quantum algorithm that implements $e^{-i\tilde{X}t}$ within error ϵ by taking $O(t \log(t/\epsilon) / \log \log(t/\epsilon))$ calls of controlled- O_x ($|0\rangle \langle 0| \otimes I + |1\rangle \langle 1| \otimes O_x$), and $O(t \log(n) \log(t/\epsilon) / \log \log(t/\epsilon))$ one- or two-qubit gates.

Proof. This theorem can be readily proved by following the same argument as proving theorem 4.1 in [22]. In that proof, a quantum algorithm which involves a series of controlled- O_x together with controlled V_j and their inverse was constructed. To prove the above theorem, we can construct a similar quantum algorithm where \tilde{V}_j instead of V_j are used. It is easy to see from Eq. (8) that \tilde{V}_j can also be efficiently implemented by taking $O(\log n)$ one- or two-qubit gates just as V_j . ■

According to this theorem, $e^{-i\tilde{X}t}$ can be efficiently implemented within error ϵ with time complexity $O(t \text{ polylog}(n) \log(t/\epsilon) / \log \log(t/\epsilon))$, under the assumption that O_x can be implemented in time $O(\text{polylog}(n))$. This efficient implementation of $e^{-i\tilde{X}t}$ for an arbitrary circulant matrix X will serve as an elementary subroutine in both phases of our algorithm as shown in the following.

B. Algorithm phase I: Training

The training phase aims to produce the quantum state $|\mathbf{w}\rangle$. To attain a more concise form of \mathbf{w} (as well as $|\mathbf{w}\rangle$), we write X in the singular value decomposition form

$$X = \sum_{j=1}^n \lambda_j |\mathbf{u}_j\rangle \langle \mathbf{v}_j|,$$

where $\{\lambda_j\}_{j=1}^n$, $\{|\mathbf{u}_j\rangle\}_{j=1}^n$, and $\{|\mathbf{v}_j\rangle\}_{j=1}^n$ are respectively the singular values, the left singular vectors, and the right singular vectors of X . So, according to the definition of Eq. (7), \tilde{X} has $2n$ eigenvalues $\{\pm \lambda_j\}_{j=1}^n$ and eigenvectors $\{|\mathbf{w}_j^\pm\rangle\}_{j=1}^n$, where

$$|\mathbf{w}_j^\pm\rangle = (|0\rangle |\mathbf{u}_j\rangle \pm |1\rangle |\mathbf{v}_j\rangle) / \sqrt{2}.$$

Since $\{|\mathbf{u}_j\rangle\}_{j=1}^n$ constitutes an orthonormal basis of the \mathbb{R}^n space, \mathbf{y} can be written as a linear combination of these basis vectors, i.e., $\mathbf{y} = \sum_{j=1}^n \beta_j \|\mathbf{y}\| |\mathbf{u}_j\rangle$. In this case, the ridge regression solution \mathbf{w} [Eq. (4)] can be rewritten as

$$\mathbf{w} = \sum_{j=1}^n \frac{\beta_j \lambda_j \|\mathbf{y}\|}{\lambda_j^2 + \alpha} |\mathbf{v}_j\rangle. \quad (9)$$

It is notable that a good α , with which ridge regression can achieve good predictive performance, can be chosen efficiently by quantum cross validation [16].

The training phase of our algorithm to generate $|\mathbf{w}\rangle$ [normalized Eq. (9)] is detailed as below, and the corresponding quantum circuit is shown in Fig. 2.

(1) Prepare three quantum registers in the state $(|0\rangle|\mathbf{y}\rangle)(|0\rangle^{\otimes s_0}|0\rangle)$, where

$$\begin{aligned} |0\rangle|\mathbf{y}\rangle &= |0\rangle \left(\sum_{j=1}^n \beta_j |\mathbf{u}_j\rangle \right) \\ &= \sum_{j=1}^n \beta_j \left(\frac{|\mathbf{w}_j^+\rangle + |\mathbf{w}_j^-\rangle}{\sqrt{2}} \right), \end{aligned} \quad (10)$$

where s_0 denotes the number of qubits used for storing eigenvalues in the next step. Consequently, the regression targets (y_1, y_2, \dots, y_n) of all the samples are encoded in the amplitudes of $|\mathbf{y}\rangle$. The details of preparing $|\mathbf{y}\rangle$ are shown in the Appendix.

(2) Performing phase estimation of $e^{-i\tilde{X}t_0}$ on the first two registers, we obtain the whole state

$$\sum_{j=1}^n \beta_j \left(\frac{|\mathbf{w}_j^+\rangle |\lambda_j\rangle + |\mathbf{w}_j^-\rangle |-\lambda_j\rangle}{\sqrt{2}} \right) |0\rangle. \quad (11)$$

(3) Performing a controlled rotation on the last register (qubit) conditioned on the eigenvalue register, we have

$$\begin{aligned} \sum_{j=1}^n \beta_j \left(\frac{|\mathbf{w}_j^+\rangle |\lambda_j\rangle \left[\frac{C\lambda_j}{\lambda_j^2 + \alpha} |1\rangle + \sqrt{1 - \left(\frac{C\lambda_j}{\lambda_j^2 + \alpha} \right)^2} |0\rangle \right]}{\sqrt{2}} \right. \\ \left. + \frac{|\mathbf{w}_j^-\rangle |-\lambda_j\rangle \left[\frac{-C\lambda_j}{\lambda_j^2 + \alpha} |1\rangle + \sqrt{1 - \left(\frac{-C\lambda_j}{\lambda_j^2 + \alpha} \right)^2} |0\rangle \right]}{\sqrt{2}} \right). \end{aligned} \quad (12)$$

Here $C = O(\min_j \lambda_j)$.

In other words, the last qubit is rotated by the angle

$$\theta(\lambda) = \arcsin \left(\frac{C\lambda}{\lambda^2 + \alpha} \right)$$

conditioned on $|\lambda\rangle$, where λ is the eigenvalue. To implement this, we use the methodology presented in [24] to construct the register $|\theta(\lambda)\rangle$ from $|\lambda\rangle$, where $\theta(\lambda)$ is approximated by

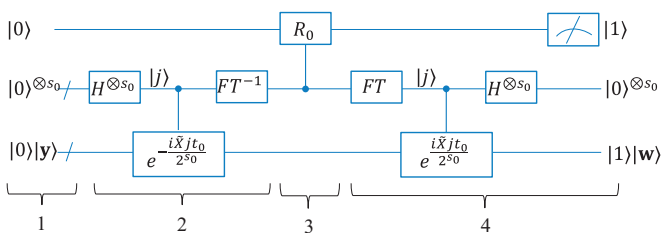


FIG. 2. Quantum circuit for the training phase. Here the numbers 1, 2, 3, and 4 denote the sequence of four steps of the training phase, “ \wedge ” denotes a bundle of wires, H denotes the Hadamard operation, FT represents the quantum Fourier transformation [6], and controlled R_0 denotes the controlled rotation in the training step (3).

truncating the Taylor series of $\theta(\lambda)$ (of λ) to some order, and $|\theta(\lambda)\rangle$ is accordingly obtained by performing a sequence of quantum multiplication and quantum addition operations as described in [24].

(4) Undoing phase estimation and measuring the last qubit to see the outcome $|1\rangle$, we obtain the state of the first register

$$|1\rangle \left[\sum_{j=1}^n \frac{C\beta_j\lambda_j}{\lambda_j^2 + \alpha} |\mathbf{v}_j\rangle \right] / \sqrt{\sum_{j=1}^n \left(\frac{C\beta_j\lambda_j}{\lambda_j^2 + \alpha} \right)^2} = |1\rangle|\mathbf{w}\rangle. \quad (13)$$

Discarding $|1\rangle$, we derive the quantum state $|\mathbf{w}\rangle$ as desired.

C. Algorithm phase II: Detection

Obtaining $|\mathbf{w}\rangle$ via the training phase of our algorithm allows us to proceed to the next phase: detection. In the detection phase, according to Eq. (5), we are to produce the final state $|\hat{\mathbf{y}}\rangle$ by performing the operation Z on $|\mathbf{w}\rangle$. To see $|\hat{\mathbf{y}}\rangle$ more concisely, let Z be written in the singular value decomposition form, $Z = \sum_{j=1}^n \gamma_j |\mathbf{p}_j\rangle \langle \mathbf{q}_j|$, where $\{\gamma_j\}_{j=1}^n$, $\{|\mathbf{p}_j\rangle\}_{j=1}^n$, and $\{|\mathbf{q}_j\rangle\}_{j=1}^n$ are respectively the singular values, the left singular vectors, and the right singular vectors of Z . Since $|\mathbf{w}\rangle$ lies in the space spanned by $\{|\mathbf{q}_j\rangle\}_{j=1}^n$, $|\mathbf{w}\rangle$ can be written as a linear combination of them, i.e., $|\mathbf{w}\rangle = \sum_{j=1}^n \delta_j |\mathbf{q}_j\rangle$. So $|\hat{\mathbf{y}}\rangle$ can be rewritten as

$$|\hat{\mathbf{y}}\rangle = \sum_{j=1}^n \gamma_j \delta_j |\mathbf{p}_j\rangle / \sqrt{\sum_{j=1}^n \gamma_j^2 \delta_j^2}. \quad (14)$$

Similar to the training phase, this phase resorts to the simulation of the extended circulant Hamiltonian of Z , $\tilde{Z} = |0\rangle\langle 1| \otimes Z + |1\rangle\langle 0| \otimes Z^\dagger$, which has $2n$ eigenvalues $\{\pm\gamma_j\}_{j=1}^n$ and eigenvectors $\{|\mathbf{r}_j^\pm\rangle = (|0\rangle|\mathbf{p}_j\rangle \pm |1\rangle|\mathbf{q}_j\rangle)/\sqrt{2}\}_{j=1}^n$. Here we also assume that a quantum oracle O_z that efficiently implements $|0\rangle^{\log[n]} \mapsto \sum_{i=1}^n \sqrt{z_i} |i\rangle$ in time $O(\text{polylog}(n))$ and are provided in the controlled fashion, and that $\sum_{i=1}^n z_i = 1$. Under these assumptions, we are able to efficiently implement $e^{-i\tilde{Z}t}$ by Theorem 1. Armed with this ability, the detection phase of our algorithm to generate $|\hat{\mathbf{y}}\rangle$ can be achieved by the following steps, and the corresponding quantum circuit is shown in Fig. 3.

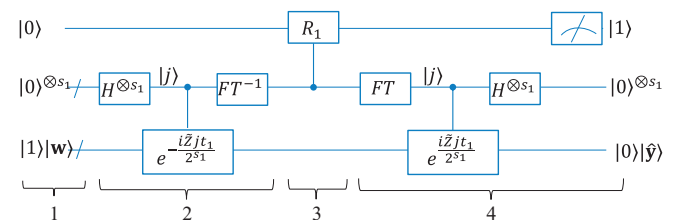


FIG. 3. Quantum circuit for generating $|\hat{\mathbf{y}}\rangle$. Here the numbers 1, 2, 3, and 4 denote the four sequential steps of the detection phase, and controlled R_1 denotes the controlled rotation in the detection step (3).

(1) Prepare three quantum registers in the state $(|1\rangle|\mathbf{w}\rangle)(|0\rangle^{\otimes s_1}|0\rangle)$, where

$$|1\rangle|\mathbf{w}\rangle = |1\rangle \left(\sum_{j=1}^n \delta_j |\mathbf{q}_j\rangle \right) = \sum_{j=1}^n \delta_j \left(\frac{|\mathbf{r}_j^+\rangle - |\mathbf{r}_j^-\rangle}{\sqrt{2}} \right), \quad (15)$$

and s_1 denotes the number of qubits for phase estimation in the following step. Note that $|\mathbf{w}\rangle$ has been produced as shown in the training phase.

(2) Performing phase estimation of $e^{-i\tilde{Z}t_1}$ on the first two registers, we have

$$\sum_{j=1}^n \delta_j \left(\frac{|\mathbf{r}_j^+\rangle |\gamma_j\rangle - |\mathbf{r}_j^-\rangle |-\gamma_j\rangle}{\sqrt{2}} \right) |0\rangle. \quad (16)$$

(3) Similar to the third step of the training phase, we also perform a controlled rotation operation on the last two registers to obtain the whole state

$$\sum_{j=1}^n \delta_j \left(\frac{|\mathbf{r}_j^+\rangle |\gamma_j\rangle [C'\gamma_j|1\rangle + \sqrt{1-(C'\gamma_j)^2}|0\rangle] - |\mathbf{r}_j^-\rangle |-\gamma_j\rangle [-C'\gamma_j|1\rangle + \sqrt{1-(C'\gamma_j)^2}|0\rangle]}{\sqrt{2}} \right). \quad (17)$$

Here $C' = O(\max_j \gamma_j)^{-1}$.

The controlled rotation can be implemented by the same approach presented in step (3) of the training phase, but the function of angle $\theta(\lambda)$ is replaced by $\theta(\gamma) = \arcsin(C'\gamma)$, where γ represents the eigenvalue stored in the second (eigenvalue) register.

(4) Undo the phase estimation and measure the last qubit to see the outcome $|1\rangle$. If successful, we obtain the state of the first register,

$$|0\rangle \left(\sum_{j=1}^n C'\delta_j \gamma_j |\mathbf{p}_j\rangle \right) / \sqrt{\sum_{j=1}^n (C'\delta_j \gamma_j)^2} = |0\rangle |\hat{\mathbf{y}}\rangle. \quad (18)$$

Then, discarding $|0\rangle$, we derive the quantum state $|\hat{\mathbf{y}}\rangle$ as desired.

D. Runtime analysis

We respectively analyze the runtime in the training phase and that in the detection phase of our algorithm, and discuss the overall runtime.

In the training phase, the error occurs in the Hamiltonian simulation of \tilde{X} and in the phase estimation in step (2). Since the complexity of Hamiltonian simulation scales sublogarithmically in the inverse of error, as shown in Theorem 1, and that of the phase estimation scales linearly [6], the source of error dominantly comes from the phase estimation. The phase estimation induces error $O(1/t_0)$ in estimating λ (singular value of X), and relative error

$$O\left(\frac{\lambda^2 - \alpha}{t_0 \lambda (\lambda^2 + \alpha)}\right) = O(1/t_0 \lambda)$$

in estimating $\lambda/(\lambda^2 + \alpha)$. Since $\sum_{i=1}^n x_i = 1$, the spectral norm of X is 1 and thus $1/\kappa_X \leq \lambda \leq 1$, where κ_X is the condition number of X . So $t_0 = O(\kappa_X/\epsilon)$ induces final error

ϵ according to the analysis in HHL algorithm [12]. In step (3), $\theta(\lambda)$ can be approximated within error ϵ by truncating its Taylor series to the order $O(\frac{\log(1/\epsilon)}{\log \log(1/\epsilon)})$, meaning that $O(\frac{\log(1/\epsilon)}{\log \log(1/\epsilon)})$ iterations of performing quantum multiplication and quantum addition operations are required [24]. Since the eigenvalue register $|\lambda_j\rangle$ (and other ancilla registers storing intermediate results) should be of $O(\log(\kappa_X/\epsilon))$ qubits to ensure λ_j being estimated within error $O(\epsilon/\kappa_X)$ via phase estimation [in step (2)] [6], each iteration takes $O(\log^2(\kappa_X/\epsilon))$ and $O(\log(\kappa_X/\epsilon))$ elementary gates to implement quantum multiplication and addition, respectively. So step (3) totally takes time $O(\frac{\log(1/\epsilon) \log^2(\kappa_X/\epsilon)}{\log \log(1/\epsilon)})$, which is relatively negligible compared to the time taken in step (2). In step (4), the success probability of obtaining $|1\rangle$ is

$$\sum_j \left(\frac{C\beta_j \lambda_j}{\lambda_j^2 + \alpha} \right)^2 = \Omega(1/\kappa_X^2),$$

since $C\lambda_j/(\lambda_j^2 + \alpha) = \Omega(1/\kappa_X)$, which means $O(\kappa_X^2)$ measurements are required to obtain $|1\rangle$ with a high probability and this can be improved to $O(\kappa_X)$ repetitions by amplitude amplification. Here α is taken for convenience to be in the range $[1/\kappa_X^2, 1]$, because from Eq. (4) it is easy to see that, when α is too small, ridge regression is reduced to ordinary linear regression and, when α is too large, it is the operation X that is performed in Eq. (4). Therefore, the total time complexity for generating $|\mathbf{w}\rangle$ in the training phase is $T_w = \tilde{O}(\text{polylog}(n)\kappa_X^2/\epsilon)$, where \tilde{O} is used to suppress the lower growing terms, the polylogarithmic factors in simulating Hamiltonian as shown in Theorem 1, and in step (3).

In the detection phase, the source of error for generating $|\hat{\mathbf{y}}\rangle$ is also dominated by the phase estimation in step (2). The phase estimation induces relative error $O(1/\gamma)$ in estimating γ , where γ denotes the singular value of Z . Since $\sum_i z_i = 1$, the spectral norm of Z is also one and $1/\kappa_Z \leq \gamma \leq 1$, where κ_Z is the condition number of Z . Thus $t_1 = O(\kappa_Z/\epsilon)$ induces final error ϵ for generating $|\hat{\mathbf{y}}\rangle$. Just as the step (3) of the training phase, the step (3) of the detection phase also takes relatively negligible time. In step (4), the success probability of obtaining the outcome $|1\rangle$ is $\sum_{j=1}^n (C'\delta_j \gamma_j)^2 = \Omega(1/\kappa_Z^2)$, which means $O(\kappa_Z)$ repetitions are required to obtain $|1\rangle$, with a high probability, by amplitude amplification. Taking into account the time for generating $|\mathbf{w}\rangle$ (in the training phase) in step (1), the runtime of the detection phase of our algorithm scales as $\tilde{O}(\kappa_Z[\text{polylog}(n)\kappa_Z/\epsilon + T_w]) = \tilde{O}(\text{polylog}(n)\kappa_Z(\kappa_Z + \kappa_X^2)/\epsilon)$.

The overall time complexity of our algorithm has the same scaling as that of the detection phase, because the training phase to create $|\mathbf{w}\rangle$ has been incorporated into the detection phase in the sense that $|\mathbf{w}\rangle$ is taken as the input in the detection phase (as shown in its first step). This means that, compared to the classical HCMB15 framework which takes the same $O(n \log(n))$ time to obtain both \mathbf{w} and $\hat{\mathbf{y}}$, our algorithm takes only exponentially less time for generating their quantum-state versions $|\mathbf{w}\rangle$ and $|\hat{\mathbf{y}}\rangle$, when $\kappa_X, \kappa_Z, 1/\epsilon = O(\text{polylog}(n))$. Comparisons between the classical HCMB15 framework and our quantum algorithm are detailed in Table I.

Further runtime needs to be considered if we use the output $|\hat{\mathbf{y}}\rangle$ to implement other specific tasks. For example, let us

TABLE I. Comparisons between our quantum algorithm and the classical HCMB15 algorithm, in terms of their inputs and outputs and their time complexity.

Algorithms	Inputs	Outputs	Time complexity
Classical algorithm	$\mathbf{x}, \mathbf{z}, \mathbf{y}, \alpha$	$\hat{\mathbf{y}}$	$O(n \log(n))$
Quantum algorithm	Controlled $O_{\mathbf{x}}$, controlled $O_{\mathbf{z}}, \mathbf{y}\rangle, \alpha$	$ \hat{\mathbf{y}}\rangle$	$\tilde{O}(\text{polylog}(n)\kappa_Z(\kappa_Z + \kappa_X^2)/\epsilon)$

consider the ultimate task of the HCMB15 framework to detect target candidate patch, which entails identifying the index of largest squared amplitude (denoted by p_{\max}) in $|\hat{\mathbf{y}}\rangle$ by sampling the state. Then we require $O(1/p_{\max})$ samples to reveal p_{\max} , which will be efficient if $1/p_m = O(\text{polylog}(n))$. However, in practice, when the object is clearly visible in the detection frame, the set of elements in $\hat{\mathbf{y}}$ will be approximately equal to that in \mathbf{y} [25]. In this case the largest squared amplitude in $|\hat{\mathbf{y}}\rangle$ is close to that in $|\mathbf{y}\rangle$, that is, $p_{\max} = 1/(\sum_{i=1}^n y_i^2) \approx \frac{2\sqrt{2}}{\sqrt{\pi s}} = O(1/\sqrt{n})$ according to the result of the Appendix. This means that $\Omega(\sqrt{n})$ copies of $|\hat{\mathbf{y}}\rangle$ are required to reveal the largest squared amplitude in $|\hat{\mathbf{y}}\rangle$. As a result, the overall runtime for this task would be $\Omega(\sqrt{n} \text{polylog}(n)\kappa_Z(\kappa_Z + \kappa_X^2)/\epsilon)$ by our quantum algorithm, achieving quadratic improvement at most over the classical HCMB15 framework. Hereafter we focus on using $|\hat{\mathbf{y}}\rangle$ to implement other interesting tasks related to VT, which can achieve much more significant speedup over their classical methods, as shown in the following Sec. IV, rather than sampling it exhaustively.

E. Extension to two-dimensional images

For two-dimensional images, a base sample patch of size $n \times m$ is represented by a $n \times m$ matrix \mathbf{x} with the j th ($1 \leq j \leq m$) row denoted by a n -dimensional vector \mathbf{x}_j . All the samples correspond to cyclic shifts of the base sample in both horizontal and vertical directions, and can be described by a block circulant matrix with circulant blocks [2], resulting in the $nm \times nm$ data matrix

$$X = \begin{bmatrix} \mathcal{C}(\mathbf{x}_1) & \mathcal{C}(\mathbf{x}_2) & \mathcal{C}(\mathbf{x}_3) & \cdots & \mathcal{C}(\mathbf{x}_n) \\ \mathcal{C}(\mathbf{x}_n) & \mathcal{C}(\mathbf{x}_1) & \mathcal{C}(\mathbf{x}_2) & \cdots & \mathcal{C}(\mathbf{x}_{n-1}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}(\mathbf{x}_2) & \mathcal{C}(\mathbf{x}_3) & \mathcal{C}(\mathbf{x}_4) & \cdots & \mathcal{C}(\mathbf{x}_1) \end{bmatrix}, \quad (19)$$

corresponding to the two-dimensional-image version of Eq. (1). It can be decomposed as

$$X = \sum_{j=1}^n V_j \otimes \mathcal{C}(\mathbf{x}_j) = \sum_{j=1}^n \sum_{k=1}^m \mathbf{x}_{jk} V_j \otimes V'_k, \quad (20)$$

where $V'_k = \sum_{l=0}^{m-1} |(l-k+1) \bmod m\rangle \langle l|$ and can also be efficiently implemented in time $O(\log(m))$.

Given the quantum oracle $O_{\mathbf{x}}$ such that $O_{\mathbf{x}}|0\rangle^{\otimes [\log nm]} = \sum_{i=1}^n \sum_{j=1}^m \sqrt{\mathbf{x}_{jk}} |j\rangle |k\rangle$ and under the assumption that $\sum_{jk} \mathbf{x}_{jk} = 1$, the extended circulant Hamiltonian $\tilde{X} = |0\rangle \langle 1| \otimes X + |1\rangle \langle 0| \otimes X^\dagger$ can be efficiently simulated as shown in Theorem 1 by extending the dimension n to nm . The ability of efficiently simulating \tilde{X} allows us to perform training and detection in the quantum computer as shown in the previous two subsections, except that the dimension has been extended from n to nm .

IV. APPLICATIONS

In this section, we show how the output response state $|\hat{\mathbf{y}}\rangle$ of our quantum algorithm can be fully used to efficiently implement two important tasks related to VT: (1) object disappearance detection and (2) motion behavior matching.

A. Object disappearance detection

The task of *object disappearance detection* is to detect whether the object has already disappeared or not, in the candidate image patch of the detection frame. If not, as discussed in the previous section, the set of elements in $\hat{\mathbf{y}}$ [Eq. (5)], namely the output in the detection phase of HCMB15 framework, will be approximately equal to that in \mathbf{y} [25]. This means that in this case the elements in $\hat{\mathbf{y}}$ also approximately follow a Gaussian distribution with a peak. However, if the object has disappeared from the video frame, the distribution of the elements in $\hat{\mathbf{y}}$ becomes much more uniform [25]. This implies that in this case the quantum state $|\hat{\mathbf{y}}\rangle$ is closer to the uniform superposition state $|\mathbf{1}\rangle := \sum_{j=1}^{n-1} |j\rangle / \sqrt{n}$. In order to clearly discriminate these two cases, we estimate the overlap between $|\hat{\mathbf{y}}\rangle$ and $|\mathbf{1}\rangle$, $P_1 = |\langle \hat{\mathbf{y}} | \mathbf{1} \rangle|^2$, using *swap test* [14,17,26], to determine and quantify the closeness between these two states. The quantum circuit of the canonical swap test for estimating the overlap of two quantum states is shown in Fig. 4, the circuit depth of which scales linearly with the qubit number of each state. The swap test has been implemented in both quantum optics [27–29] and a gate-based quantum computer [30]. The task of estimating the state overlap can also be done by recent machine-learning-found quantum algorithms [31], which have shorter and even constant depth and thus may be more promising on near-term quantum computers than the canonical swap test. These algorithms experimentally exhibit more robustness and reliability, but their performances

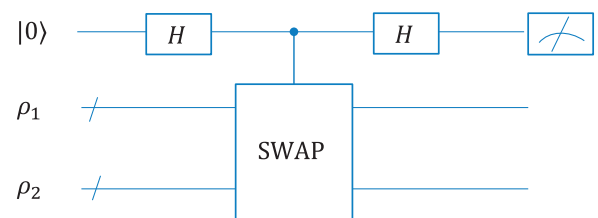


FIG. 4. Quantum circuit of swap test for estimating the overlap, $\text{Tr}(\rho_1 \rho_2)$, between two states with density matrices ρ_1 and ρ_2 , where SWAP denotes the SWAP unitary operation implemented by a series of elementary one-qubit or two-qubit gates [6]. The probability of obtaining the measurement outcome $|0\rangle$ after measuring the top ancilla qubit is $\frac{1 + \text{Tr}(\rho_1 \rho_2)}{2}$, which reveals the estimate of $\text{Tr}(\rho_1 \rho_2)$ by repeating the circuit for a sufficient number of times. In our case, we are to estimate the overlap between the two pure states $\rho_1 = |\hat{\mathbf{y}}\rangle \langle \hat{\mathbf{y}}|$ and $\rho_2 = |\mathbf{1}\rangle \langle \mathbf{1}|$, $\text{Tr}(\rho_1 \rho_2) = P_1 = |\langle \hat{\mathbf{y}} | \mathbf{1} \rangle|^2$.

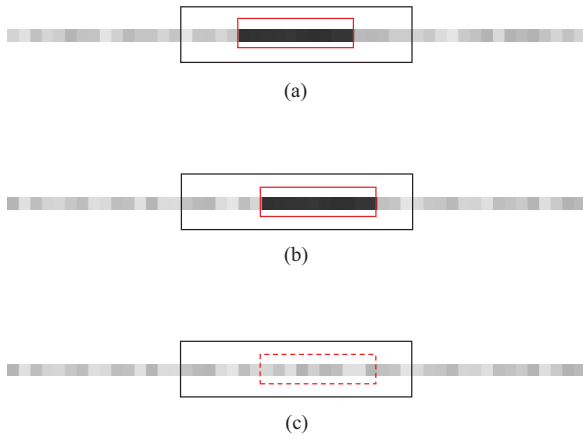


FIG. 5. Figure (a) is the training frame, where the object is placed at the center. The black (bigger) rectangle denotes the image patch and the red (smaller) rectangle denotes the object. Figure (b) is the detection frame where the object moves 3 pixels to the right relative to that in (a) but still exists in the patch. Figure (c) is the detection frame where the object disappears relative to that in (b), and the dashed rectangle denotes the position where the object should be. All the three frames are one-dimensional gray scale images of 50 pixels. The patch and the object are of 20 pixels and 10 pixels, respectively.

substantially depend on the connectivity and available quantum gates of the actual quantum computers.

By setting a threshold ϑ_1 , the object is considered as disappeared if the obtained $P_1 \geq \vartheta_1$; otherwise, the object is considered still in the video frame. Following [14,17,26], P_1 can be obtained to accuracy δ with $O(1/\delta^2)$ repetitions of swap test, so this task only takes $O(1/\delta^2)$ copies of the response state $|\hat{y}\rangle$ (as well as $|\mathbf{1}\rangle$), as opposed to $O(\sqrt{n})$ copies for sampling described in the last section. This means that, compared with the task of sampling $|\hat{y}\rangle$, the task of object disappearance detection takes exponentially fewer number of $|\hat{y}\rangle$, if $\delta = O(1/\text{polylog } n)$ is acceptable.

To show more intuitively how different the values of P_1 are in the above two cases, we carried out a numerical experiment with an illustrative example, as shown in Fig. 5. In the experiment, we manually and randomly generate one training frame, and two detection frames in which the object is either still there or disappeared. In this example, $P_1 = 0.577$ for the case where the object exists and $P_1 = 0.986$ for the case where the object disappears. We then run the experiment for 50 times to yield 50 values of P_1 for each case. These P_1 values are shown in Fig. 6. From this figure, we can see that $P_1 \leq 0.6$ for the case where the object exists, while $P_1 \geq 0.9$ for the case where the object disappears. Therefore, for this example, ϑ_1 can satisfactorily take the value 0.75.

B. Motion behavior matching

The task of *motion behavior matching* is to determine whether the motion behavior of the object in a video matches a given motion behavior template. More specifically, it is to determine whether the object in the video moves along a given path or not. To see this task visually, we present a simple example with two-dimensional images in Fig. 7. The

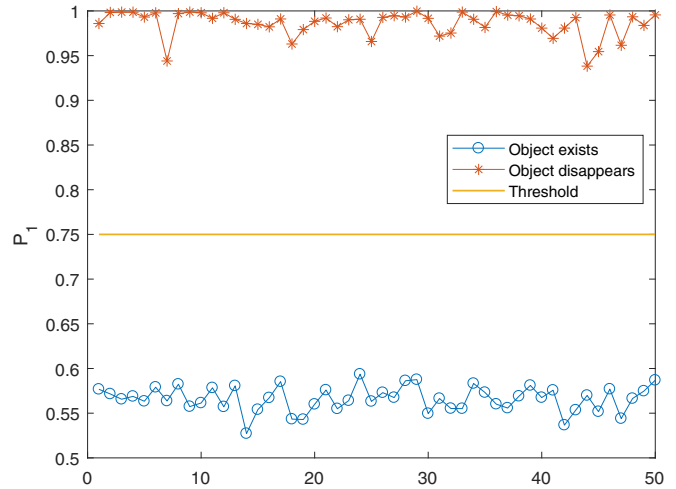


FIG. 6. Comparison between the 50 values of P_1 for the case where the object exists and those for the case where the object disappears after running the experiment in Fig. 5 for 50 times.

procedure to implement this task using our quantum algorithm is detailed in the following steps.

(1) Train the entire initial frame of the video (instead of an image patch according to the standard HCMB15 framework) by using the training part of our quantum algorithm and get the classifier $|\mathbf{w}\rangle$.

(2) Select K positions in the template including the position of the object in the initial frame, generate K “template” frames by just moving the object in the initial frame to these positions, and perform the detection part of our quantum algorithm, using the $|\mathbf{w}\rangle$ obtained on these template frames, to generate K response states, $|\hat{y}_t^1\rangle, |\hat{y}_t^2\rangle, \dots, |\hat{y}_t^K\rangle$.

(3) Select K “actual” frames appropriately from the video, where the object is supposed to be located at the selective K positions if the object motion behav-

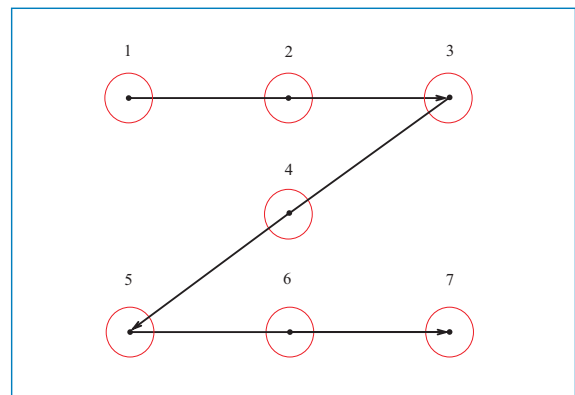


FIG. 7. Simple example for motion behavior matching. Here the rectangle with blue border denotes the frame, the red circles denote the object, the black arrowed lines correspond to the Z-shape motion template, and the numbers 1, 2, . . . , 7 are to mark seven selected positions in the template. The initial frame of the video contains the object located in position 1. If the object motion behavior in the video matches the template, the object in seven appropriately selected frames of the video should be located at these seven positions.

ior matches the template well, and perform the detection part of our quantum algorithm with the classifier $|\mathbf{w}\rangle$ on these frames, to generate K response states, $|\hat{y}_a^1\rangle, |\hat{y}_a^2\rangle, \dots, |\hat{y}_a^K\rangle$. Here we assume that the information of the object moving speed is known so that these actual frames can be appropriately selected.

(4) Evaluate the closeness between $|\psi_t\rangle$ and $|\psi_a\rangle$ by estimating the value of $P_2 = |\langle\psi_t|\psi_a\rangle|^2$ using swap test as shown in Fig. 4, where $|\psi_t\rangle = \otimes_{k=1}^K |\hat{y}_t^k\rangle$ and $|\psi_a\rangle = \otimes_{k=1}^K |\hat{y}_a^k\rangle$. Just as P_1 , P_2 can also be estimated by shorter-depth quantum algorithms [31]. Setting another threshold ϑ_2 , if $P_2 \geq \vartheta_2$, we regard the motion behavior of the object in the video matches the template well; otherwise, the matching fails.

In step (4), P_2 can be obtained to accuracy δ with $O(1/\delta^2)$ copies of $|\psi_t\rangle$ and $|\psi_a\rangle$ and each copy takes $2K$ response states of the same size. Therefore, $O(K/\delta^2)$ copies of response states ($|\psi_t\rangle$ and $|\psi_a\rangle$) are taken in total for implementing this task. Moreover, $O(K/\delta^2) = O(\text{polylog}(n))$ if $K, 1/\delta = O(\text{polylog}(n))$ is acceptable. In practice, K can be chosen to be very small relative to the number of frames in the video, if we just want to roughly know the object's motion in the video. The choice for the threshold ϑ_2 depends on how well we require the object's moving motion to match the template, but practically and reasonably ϑ_2 should be chosen to be close to 1, e.g., $\vartheta_2 = 0.9$. In addition, the accuracy δ is generally chosen to be $O(\vartheta_2)$, e.g., $\delta = \vartheta_2/10$. Therefore, the scaling $O(K/\delta^2)$ can be small in practice.

In addition to the above two applications, we expect that this algorithm can be implemented for other tasks of practical interest.

V. CONCLUSIONS

We have presented a quantum algorithm for visual tracking based on the well-known classical HCMB15 framework, which can be applied to detect the object position as done in the HCMB15 framework with quadratic speedup. Our algorithm first trains a quantum-state ridge regression classifier, where the optimal fitting parameters of ridge regression are encoded in the amplitudes. The classifier is then performed on the detection frame to generate a quantum state whose amplitudes encode the responses of all the candidate image patches. Taking full advantage of efficient extended circulant Hamiltonian simulation, both states can be generated in time polylogarithmic in their dimensionality, when the data matrices have low condition numbers. This demonstrates that our quantum algorithm has the potential to achieve exponential speedup over the classical counterpart. Furthermore, we have also shown how our algorithm can be applied to efficiently implement two important tasks: object disappearance detection and motion behavior matching.

We expect the techniques used in our algorithm, such as extended circulant Hamiltonian simulation, to be helpful in designing more quantum algorithms requiring manipulating circulant matrices. We also hope this algorithm may inspire more quantum algorithms for visual tracking as well as other computer vision problems.

ACKNOWLEDGMENTS

The authors would like to thank J. F. Henriques for helpful discussions and Bruce Hartley for proofreading the manuscript. This work is supported by NSFC (Grants No. 61572081, No. 61672110, and No. 61671082). C.-H.Y. is supported by China Scholarship Council.

APPENDIX: PREPARING THE QUANTUM STATE $|\mathbf{y}\rangle$ IN THE TRAINING PHASE

According to the definition of \mathbf{y} [Eq. (2)], for one-dimensional images, it is easy to see that $y_i = e^{-(i-1)^2/s^2}$ for $i = 1, 2, \dots, \lfloor(n+1)/2\rfloor$, and $y_i = e^{-(n+1-i)^2/s^2}$ for $i = \lfloor(n+1)/2\rfloor + 1, \lfloor(n+1)/2\rfloor + 2, \dots, n$. Since the elements of \mathbf{y} are generally not uniformly distributed, it is time consuming to create the quantum state $|\mathbf{y}\rangle = \sum_{i=1}^n y_i |i\rangle / \|\mathbf{y}\|$ in the common way: y_i are loaded into a quantum register in parallel and a controlled rotation and measurement are then conducted on an ancilla qubit so that $y_i / \|\mathbf{y}\|$ can be written in the amplitudes. Another efficient way is referred to [32], where $\sum_{i=i_1}^{i_2} y_i^2$ for any two i_1 and i_2 , with $1 \leq i_1 \leq i_2 \leq n$, is required to be efficiently computable to create $|\mathbf{y}\rangle$ [12]. However, y_i cannot satisfy this condition because there is no efficient formula for calculating $\sum_{i=i_1}^{i_2} y_i^2$ for any two i_1 and i_2 with $1 \leq i_1 \leq i_2 \leq n$. In the following, we present an approach that combines both of the two ways to efficiently create $|\mathbf{y}\rangle$.

Our approach is based on the observation that, for $i = 2, \dots, \lfloor(n+1)/2\rfloor$, y_i can be approximated by integrating some Gauss function in some appropriate range, that is,

$$y_i^2 \approx \tilde{y}_i^2 := s \int_{\frac{i-2}{s}}^{\frac{i-1}{s}} e^{-2t^2} dt, \quad (\text{A1})$$

but $y_i^2 \leq \tilde{y}_i^2$. This is related to the error function $E(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ ($x > 0$) that can be approximated by some elementary functions. For example, $E(x) \approx G(x) := 1 - (a_1 t + a_2 t^2 + a_3 t^3) e^{-x^2}$ within error 2.5×10^{-5} [33], where $t = \frac{1}{1+px}$, $p = 0.47047$, $a_1 = 0.3480242$, $a_2 = -0.0958798$, and $a_3 = 0.7478556$. Since this error is very small and thus negligible, we replace $E(x)$ with $G(x)$ and have

$$\tilde{y}_i^2 = \frac{\sqrt{\pi}}{2\sqrt{2}} s \left[G\left(\frac{\sqrt{2}(i-1)}{s}\right) - G\left(\frac{\sqrt{2}(i-2)}{s}\right) \right]$$

for $i = 2, \dots, \lfloor(n+1)/2\rfloor$. This means that $\sum_{i=i_1}^{i_2} \tilde{y}_i^2$ for any two $2 \leq i_1 \leq i_2 \leq \lfloor(n+1)/2\rfloor$ is efficiently computable. Furthermore, for $i = 2, \dots, \lfloor(n+1)/2\rfloor$, since $y_1 = 1$ and $y_i = y_{n+2-i}$, we can set $\tilde{y}_1^2 = 1$ and $\tilde{y}_i^2 = \tilde{y}_{n+2-i}^2$; thus $\sum_{i=i_1}^{i_2} \tilde{y}_i^2$ for any two $1 \leq i_1 \leq i_2 \leq n$ is also efficiently computable. Consequently, using the approach of [32], we can create the state

$$|\tilde{\mathbf{y}}\rangle := \sum_{i=1}^n \frac{\tilde{y}_i}{\sqrt{\sum_{i=1}^n \tilde{y}_i^2}} |i\rangle \quad (\text{A2})$$

efficiently in time $O(\log n)$ [32].

Armed with the capability of efficiently creating $|\tilde{y}\rangle$, we can create the state $|\mathbf{y}\rangle$ efficiently as well, by the following procedure.

(1) Add two registers to the register of $|\tilde{y}\rangle$ and load the \tilde{y}_i and y_i in parallel to get the state

$$\sum_{i=1}^n \frac{\tilde{y}_i}{\sqrt{\sum_{i=1}^n \tilde{y}_i^2}} |i\rangle |y_i\rangle |\tilde{y}_i\rangle.$$

Note that \tilde{y}_i and y_i can be computed efficiently.

(2) Add another qubit and perform the controlled rotation to have the state

$$\sum_{i=1}^n \frac{\tilde{y}_i}{\sqrt{\sum_{i=1}^n \tilde{y}_i^2}} |i\rangle |y_i\rangle |\tilde{y}_i\rangle \left(\frac{y_i}{\tilde{y}_i} |1\rangle + \sqrt{1 - \frac{y_i^2}{\tilde{y}_i^2}} |0\rangle \right).$$

(3) Undo the first step and get the state

$$\sum_{i=1}^n \frac{\tilde{y}_i}{\sqrt{\sum_{i=1}^n \tilde{y}_i^2}} |i\rangle \left(\frac{y_i}{\tilde{y}_i} |1\rangle + \sqrt{1 - \frac{y_i^2}{\tilde{y}_i^2}} |0\rangle \right).$$

(4) Measure the last qubit to ensure the outcome of $|1\rangle$ and the first register will be in the state $|\mathbf{y}\rangle$ as desired. The success probability of this postselection is

$$\frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n \tilde{y}_i^2} = \Theta(1),$$

because

$$\sum_{i=1}^n \tilde{y}_i^2 \approx \sum_{i=1}^n y_i^2, \quad \sum_{i=1}^n y_i^2 \approx \frac{\sqrt{\pi}}{2\sqrt{2}} s$$

when n is sufficiently large. Consequently the time complexity for creating $|\mathbf{y}\rangle$ scales as $O(\log n)$.

-
- [1] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, High-speed tracking with kernelized correlation filters, *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 583 (2015).
 - [2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels, in *Computer Vision—ECCV 2012*, edited by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Lecture Notes in Computer Science Vol. 7575 (Springer, Berlin, Heidelberg, 2012), pp. 702–715.
 - [3] P. Mountney, D. Stoyanov, and G.-Z. Yang, Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images, *IEEE Signal Process. Mag.* **27**, 14 (2010).
 - [4] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, Recent advances and trends in visual tracking: A review, *Neurocomput.* **74**, 3823 (2011).
 - [5] A. Smeulders, D. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, Visual tracking: An experimental survey, *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1442 (2014).
 - [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2010).
 - [7] A. Montanaro, Quantum algorithms: an overview, *npj Quantum Inf.* **2**, 15023 (2016).
 - [8] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, *Proc. Natl. Acad. Sci. USA* **115**, 9456 (2017).
 - [9] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, edited by S. Goldwasser (IEEE, Los Alamitos, CA, 1994), pp. 124–134.
 - [10] L. K. Grover, Quantum Mechanics Helps in Searching for a Needle in a Haystack, *Phys. Rev. Lett.* **79**, 325 (1997).
 - [11] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, *Quantum Amplitude Amplification and Estimation*, Contemporary Mathematics Series Millennium Vol. 305 (AMS, New York, 2002).
 - [12] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
 - [13] N. Wiebe, D. Braun, and S. Lloyd, Quantum Algorithm for Data Fitting, *Phys. Rev. Lett.* **109**, 050505 (2012).
 - [14] M. Schuld, I. Sinayskiy, and F. Petruccione, Prediction by linear regression on a quantum computer, *Phys. Rev. A* **94**, 022342 (2016).
 - [15] G. Wang, Quantum algorithm for linear regression, *Phys. Rev. A* **96**, 012335 (2017).
 - [16] C.-H. Yu, F. Gao, and Q.-Y. Wen, Quantum algorithms for ridge regression, [arXiv:1707.09524](https://arxiv.org/abs/1707.09524).
 - [17] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, *Phys. Rev. Lett.* **113**, 130503 (2014).
 - [18] C.-H. Yu, F. Gao, Q.-L. Wang, and Q.-Y. Wen, Quantum algorithm for association rules mining, *Phys. Rev. A* **94**, 042311 (2016).
 - [19] N. Liu and P. Rebentrost, Quantum machine learning for quantum anomaly detection, *Phys. Rev. A* **97**, 042315.
 - [20] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, *Rep. Prog. Phys.* **81**, 074001 (2018).
 - [21] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature (London)* **549**, 195 (2017).
 - [22] S. S. Zhou and J. B. Wang, Efficient quantum circuits for dense circulant and circulant-like operators, *R. Soc. Open Sci.* **4**, 160906 (2017).
 - [23] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Simulating Hamiltonian Dynamics with a Truncated Taylor Series, *Phys. Rev. Lett.* **114**, 090502 (2015).
 - [24] I. Cong and L. Duan, Quantum discriminant analysis for dimensionality reduction and classification, *New J. Phys.* **18**, 073011 (2016).
 - [25] J. F. Henriques (private communication).

- [26] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum Fingerprinting, *Phys. Rev. Lett.* **87**, 167902 (2001).
- [27] J. C. Garcia-Escartin and P. Chamorro-Posada, Swap test and Hong-Ou-Mandel effect are equivalent, *Phys. Rev. A* **87**, 052330 (2013).
- [28] R. B. Patel, J. Ho, F. Ferreyrol, T. C. Ralph, and G. J. Pryde, A quantum Fredkin gate, *Sci. Adv.* **2**, e1501531 (2016).
- [29] F. Ferreyrol, T. C. Ralph, and G. J. Pryde, Implementation of a quantum Fredkin gate using an entanglement resource, *2013 Conference on Lasers and Electro-Optics Europe (CLEO EUROPE/IQEC) & International Quantum Electronics Conference* (IEEE, Munich, Germany, 2013), pp. 1-1.
- [30] N. M. Linke, S. Johri, C. Figgatt, K. A. Landsman, A. Y. Matsuura, and C. Monroe, Measuring the Renyi entropy of a two-site Fermi-Hubbard model on a trapped ion quantum computer, *Phys. Rev. A* **98**, 052334 (2018).
- [31] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, Learning the quantum algorithm for state overlap, *New J. Phys.* **20**, 113022 (2018).
- [32] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, [arXiv:quant-ph/0208112](https://arxiv.org/abs/quant-ph/0208112).
- [33] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover Publications Inc., Mineola, NY, 1965).