

Quantum Data Reduction with Application to Video Classification

Kostas Blekos

Computer Engineering & Informatics Dept.
University of Patras
Rio, Greece
0000-0002-6777-2107

Dimitrios Kosmopoulos

Computer Engineering & Informatics Dept.
University of Patras
Rio, Greece
dkosmo@upatras.gr

Abstract—We investigate a quantum data reduction technique with application to video classification. A hybrid quantum-classical step performs data reduction on the video dataset generating “representative” distributions for each video class. These distributions are used by a quantum classification algorithm to firstly reduce the size of the videos and then classify the reduced videos to one of k classes. We verify the method using sign videos and demonstrate that the reduced videos contain enough information to successfully classify them using a quantum classification process.

The proposed data reduction method showcases a way to alleviate the “data loading” problem of quantum computers for the problem of video classification. Data loading is a huge bottleneck, as there are no known efficient techniques to perform that task without sacrificing many of the benefits of quantum computing.

Index Terms—quantum computing, data reduction, video classification

I. INTRODUCTION

In the past few years there has been a lot of interest in exploring quantum computing applications for classical computational problems. Quantum computing, by taking advantage of the quantum mechanical properties of nature, offers a new toolbox for attacking such problems with potentially great benefits in efficiency and performance. Image and video classification algorithms are an important subset of algorithms where quantum computing applications are actively investigated [1]–[7], though, to our best knowledge, applications on video processing are only sparsely researched [8].

One important issue with quantum video processing applications is that they require the transfer of large amount of data from the classical machines to the quantum computer. This “data loading” of the quantum computer is a huge bottleneck as there are no known efficient techniques that can perform such a task without sacrificing many of the quantum-gains [9]–[12].

This problem is not unique to video processing applications [13]. For example, solving a system of linear equations through the HHL algorithm is usually stated as needing $O(\log(n))$ steps for n equations. This statement, though, hides many assumptions, one of which is that one should be able to read and store n parameters on at most $O(\log(n))$ steps [14].

In general, it is possible that near-term applications will mostly concern the cooperation of the classical hardware

with small quantum computing units [9]. This highlights the importance of reducing the classical data space before communicating with the quantum hardware. One proposal, by Harrow [9], is using hybrid classical-quantum algorithms where a “representative” subset of a problem’s data set is extracted. This subset is then used as input to the quantum computer instead of the whole data set. Here we present a different method where instead of extracting a representative subset, a representative *distribution* is generated that aims to capture the most significant features of the feature vectors of the dataset. This distribution is then used to extract specific subparts from the full dataset and is fed to the desired quantum algorithm.

A. Contribution

In this work we propose a data reduction scheme on a hybrid classical-quantum algorithm for video classification. Using a hybrid procedure, we extract the most important pixels of a video which we then use as the reduced “training set”. We classify each new video based only on these pixel distributions. We perform simulations of the proposed algorithm verifying that meaningful information can be extracted from just a small percentage of the initial video. The videos are accurately classified showing that the algorithm is capable of accurately identifying the most significant pixels using an efficient quantum procedure. To the best of our knowledge this is the first such approach.

Even though we evaluate the proposed technique using a video dataset, it should, in principle, be applicable to any other dataset where differences between successive feature vectors are correlated as, for example, is usually the case in time series data.

In Section II we provide the background knowledge. In Section III we describe the proposed method. In Section IV we present our results on a public dataset. Finally in Section V we discuss the merits and constraints of the method.

II. QUANTUM BACKGROUND

Quantum computers use quantum states of two levels (a *qubit*) to store and process information instead of using bits of 0 and 1. Abstractly, a qubit is a two-dimensional vector of complex parameters and norm 1, i.e., $|\text{qubit}\rangle = \begin{pmatrix} a \\ b \end{pmatrix}$

with $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. Similarly, a series of n quantum bits (often referred to as a *quantum register*) form a 2^n -dimensional vector of complex parameters and norm 1. We define the *computational basis* as the one-hot orthonormal basis: $|0\rangle = (1 \ 0 \ \dots \ 0)^T$, $|1\rangle = (0 \ 1 \ \dots \ 0)^T, \dots, |2^n - 1\rangle = (0 \ 0 \ \dots \ 1)^T$.

Using these definitions, a general quantum register is written as a linear combination of the computational basis $|q\rangle = a_0 |0\rangle + a_1 |1\rangle + \dots + a_{2^n-1} |2^n - 1\rangle$. The parameter a_i is called the *amplitude* of the state $|i\rangle$. The actual parameters of a quantum register are unknowable. When the value of a quantum register is needed a *measurement* is performed. The result of the measurement gives *one* of the computational basis vectors. In particular, the result of measuring the register $|q\rangle$ above is $|i\rangle$ with probability $|a_i|^2$.

Quantum registers are manipulated by use of quantum gates (also called quantum operators) which act as the quantum analog of the classical logical gates (AND, OR, NOT, etc). Quantum gates are represented by complex unitary matrices of appropriate dimensions and their action is calculated by simple matrix multiplication.

A. Inner product estimation

The *Inner product estimation* is an efficient subroutine for the estimation of the inner product between two quantum registers $|a\rangle, |b\rangle$. The subroutine has three steps. Firstly an ancilla qubit is entangled with the two quantum registers producing the register:

$$|\phi\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle_\alpha |a\rangle + |1\rangle_\alpha |b\rangle \right)$$

Secondly, a Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ with $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, is applied on the ancilla qubit:

$$\begin{aligned} H_\alpha |\phi\rangle &= \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} (|0\rangle_\alpha + |1\rangle_\alpha) |a\rangle + \frac{1}{\sqrt{2}} (|0\rangle_\alpha - |1\rangle_\alpha) |b\rangle \right) \\ &= \frac{1}{2} \left(|0\rangle_\alpha (|a\rangle + |b\rangle) + |1\rangle_\alpha (|a\rangle - |b\rangle) \right) \end{aligned}$$

Then, a measurement of the ancilla qubit is performed. The probability of measuring $|0\rangle_\alpha$ is its amplitude squared, thus:

$$\begin{aligned} P(|0\rangle_\alpha) &= \left\| \frac{1}{2} (|a\rangle + |b\rangle) \right\|^2 \\ &= \frac{1}{4} \sqrt{\sum_i (\bar{a}_i + \bar{b}_i)(a_i + b_i)} \\ &= \frac{1}{4} \left| \sum_i \bar{a}_i a_i + \bar{a}_i b_i + \bar{b}_i a_i + \bar{b}_i b_i \right| \\ &= \frac{1}{4} |\langle a|a\rangle + \langle a|b\rangle + \langle b|a\rangle + \langle b|b\rangle| = \frac{1 + \langle a|b\rangle}{2} \end{aligned}$$

since, by definition, $\langle \cdot | \cdot \rangle \geq 0$ and $\langle x|x\rangle = 1$. Therefore, by repeatedly measuring the ancilla qubit an estimate for the inner product $\langle a|b\rangle$ can be calculated.

B. Amplitude encoding

One often employed method to encode classical data to quantum registers is the *amplitude encoding* in which one encodes the information of a classical vector on the amplitudes of the computational basis. So, a vector $v = (v_0, v_1, \dots, v_n)$ is encoded to the quantum register

$$|v\rangle = \frac{1}{\|v\|} \sum v_i |i\rangle.$$

C. QRAM

A ‘‘Quantum RAM’’ (QRAM) is a classical or quantum data structure that outputs quantum states. Like its classical analog RAM, it is used to store and retrieve information in a ‘‘Random Access’’ model, i.e., any desired bit of information can be addressed individually at will. Various QRAM models have been proposed that provide efficient implementations for the crucial quantum state storing and retrieving procedures [15], [16]. The inner product estimation algorithm, for example, that is used in this work, can be efficiently implemented using a QRAM [17].

III. METHODOLOGY

We are going to demonstrate the proposed method by means of a video classification application. The task is to classify short videos displaying different types of hand motion into one of k classes. The methodology can be easily generalized to classification tasks involving high-volume data.

We assume that each video is represented by an $N \times N \times T$ matrix (height \times width \times frames) of values in the range $[0..1]$ and belongs to one of k classes. We will reduce the information contained in this $N \times N \times T$ -sized matrix to a 2^q -sized quantum state using q qubits. When a classical-to-quantum video conversion is needed we use the amplitude encoding as described in Section II-B. We assume that a training set consists of $M \times k$ videos with M videos for each class. The overview of the algorithm is given in Algorithm 1 and schematically in Figure 1.

A. Convert frames of the training set to quantum states. Using the amplitude encoding we convert successive frames to quantum states.

B. Perform ‘‘difference’’ transform on the quantum states and store in a QRAM. We use the fact that among successive frames of continuous videos the differences are small. A *difference transform* efficiently converts two successive frames $|q_1\rangle$ and $|q_2\rangle$ to their difference $|q_1\rangle - |q_2\rangle$ by using an ancilla qubit and a Hadamard gate, similarly to the inner product estimation subroutine of Section II-A:

$$\begin{aligned} |\phi\rangle &= \frac{1}{\sqrt{2}} \left(|0\rangle_\alpha |q_1\rangle + |1\rangle_\alpha |q_2\rangle \right) \\ H_\alpha |\phi\rangle &= \frac{1}{\sqrt{2}} \left(|0\rangle_\alpha (|q_1\rangle + |q_2\rangle) + |1\rangle_\alpha (|q_1\rangle - |q_2\rangle) \right) \end{aligned}$$

Algorithm 1 Video reduction and classification with a quantum algorithm

for All training videos **do**
 Convert frames to quantum registers.
 Apply “difference” transform to consecutive frames.
 Store result in QRAM.
end for
for All classes **do**
 Calculate average video.
 Get sample distribution by measuring 2^q times each average video.
 Get training quantum register by sampling the average video using the class distribution.
end for
Reduce a new video by sampling using the class distributions.
Classify by taking the inner product estimation and setting $\text{class} = \arg \max(\langle v_1|V_1\rangle, \dots, \langle v_k|V_k\rangle)$.

We then measure the ancilla until we find it at the state $|1\rangle_\alpha$. The frame register, then, will be at the state $|q_1\rangle - |q_2\rangle$. This is now stored again in the QRAM for further processing.

C. Repeat (A–B) for all the frames of each video averaging over all training videos for each class. We repeat the previous steps until all difference frames of all training videos are stored in the QRAM. An average difference-video is then calculated for each class: $\text{diff-video}_1, \dots, \text{diff-video}_k$ by taking the average of the amplitudes of all quantum registers as they are stored in the QRAM.

D. Load average videos and perform 2^q measurements to get a distribution for each class. We load the k average difference videos to quantum registers of size $N \times N \times T$. We measure 2^q times to get a distribution—with replacement—that represents the most important pixels from all frames, as these are the pixels that have the highest amplitudes. This creates a *weighted* distribution of the most important pixels for each class of videos. Since the amplitudes have been calculated by the differences between successive frames, a low amplitude means that there is no significant change between the frames whereas a high amplitude means that there is significant difference between the successive frames. At the end of this procedure we have k distributions of 2^q pixel-coordinates.

E. Use the distributions and the average videos to produce a training quantum register for each class. Using the class distributions we convert each average video to a quantum register using q qubits. All pixels not belonging to the distributions are ignored. We produce k training quantum registers $|v_1\rangle, \dots, |v_k\rangle$.

F. Classify a new video by reducing it using the distributions and performing an inner product estimation. To classify a new video we use the class distributions to convert it to a quantum register of q qubits. We keep as amplitudes of the video the pixels that belong to the class

distributions and ignore all other pixels. For each video we produce k test quantum registers $|V_1\rangle, \dots, |V_k\rangle$. We use the inner product estimation subroutine to calculate the k inner products: $\langle v_1|V_1\rangle, \dots, \langle v_k|V_k\rangle$. We assign the video to the class:

$$\text{class} = \arg \max(\langle v_1|V_1\rangle, \dots, \langle v_k|V_k\rangle).$$

A. Effect of the number of qubits in performance and comparison to a classical algorithm.

The number of qubits q used in the algorithm affects the performance of the sampling steps (D, E) and the classification step (F). Performances of steps A, B and D depend mainly on the size of the videos.

Since the “difference” transform can be performed efficiently at the same time as the QRAM storing [16], the time of steps A and B collectively is $O(f(K))$, where $f(n)$ is a function that gives the time of the data-loading procedure and $K = N \times N \times T$ is the size of each video. At the moment it seems that a relation of $O(f(n)) \sim O(n)$ is unavoidable at best [9]. Step C, then, iterates steps A and B for each of the kM videos of the training dataset resulting in total runtime of $O(kMf(K))$ for steps A-C.

Steps D-E are the core of the “training” process. The sampling procedure of step D consists of loading the k average difference videos and performing $Q = 2^q$ measurements. Step E essentially just stores the reduced Q -sized versions of the k average difference videos and this can be performed concurrently with step D. Since a measurement is, in general, a linear process we can assume that the total runtime of steps D-E is $O(k(f(K) + Q))$.

Finally, step F performs the actual loading and classification of a reduced video. The runtime for this step is set as $O((k+1)f(Q) + g(Q))$, where $g(n)$ is some function that models the runtime of the classification procedure. It is presumed that this function would be much faster than a corresponding classical function ($g'(n)$), otherwise there would be no benefit on using the quantum version to begin with. The simplistic g that we use here might be $O(g(n)) \sim O(\sqrt{\log(n)})$ [18], [19]. It is also a core assumption of this work, as stated earlier, that the function $f(n)$ —converting classical data to quantum data—models a process that has worse runtime than the classification process: $O(f(n)) > O(g(n))$.

Table I summarizes the time of the algorithm and how it compares to the classical/quantum cases with/without a data reduction procedure. Classification with the quantum algorithm with reduction is much faster than one with the quantum algorithm without reduction since $Q < K \Rightarrow O(f(Q) + g(Q)) < O(f(K) + g(K))$. On the other hand classification using the quantum algorithm without reduction *might not be faster than* using the classical algorithm when the quantum data-loading is expensive, i.e., when $O(f(n)) > O(g(n))$, even if the core of the classification function is more efficient in the quantum case: $O(g(n)) < O(g'(n))$ since this condition alone can not guarantee that $O(f(K) + g(K)) < O(g'(K))$.

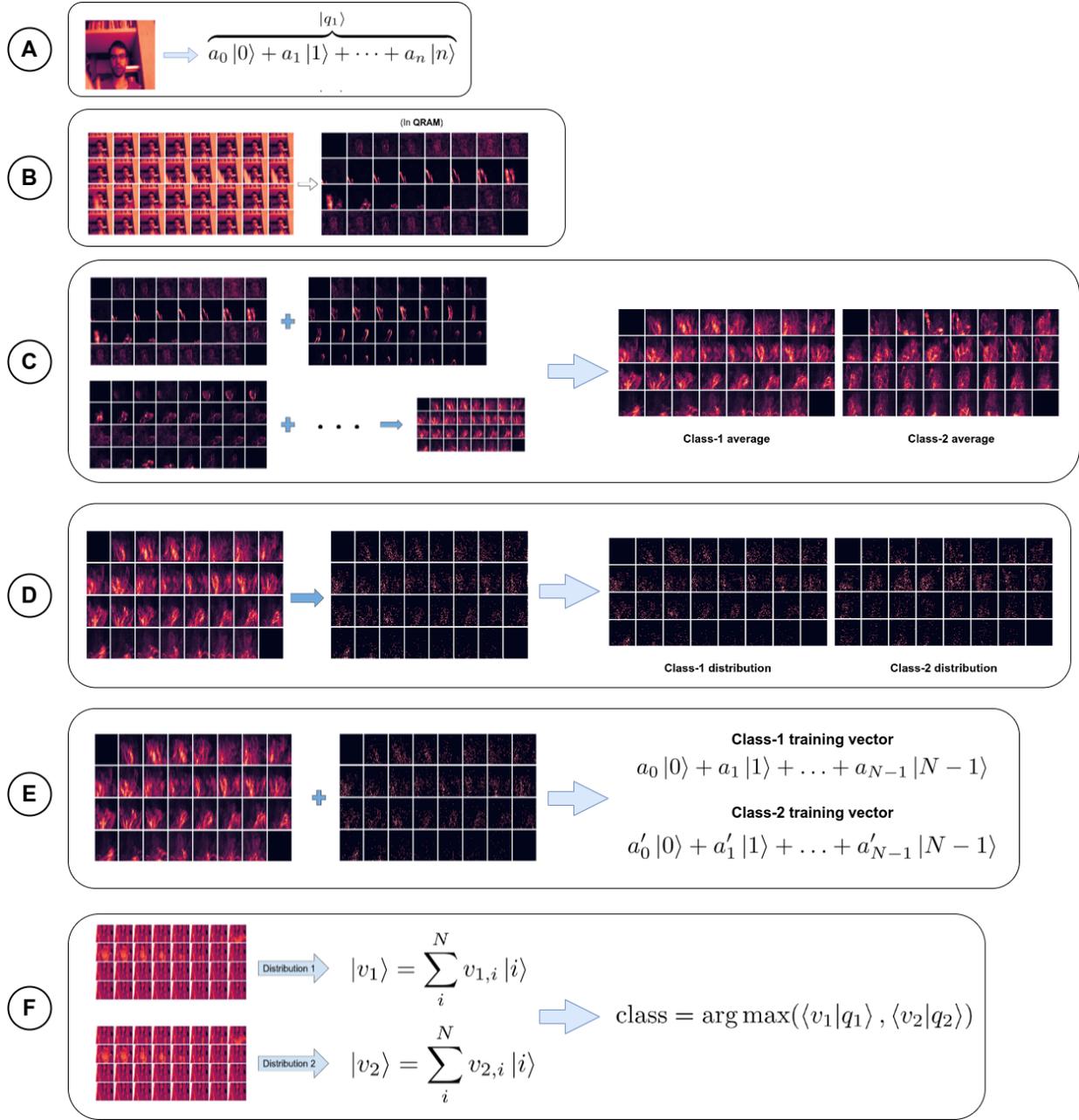


Fig. 1. Schematic of the algorithm. (A) Convert training set to quantum states. (B) Transform and store in a QRAM. (C) Repeat (A–B) and keep average of each class. (D) Perform 2^q measurements on the averages of each class. (E) Use measurement results to produce a training quantum register for each class. (F) Perform inner product estimation. Classify to best match.

Step	Quantum w/ reduction	Classical w/ reduction	Quantum	Classical
A-B	$O(f(K))$	-	-	-
C	$O(kM) * (\text{A-B})$	-	-	-
D-E	$O(k(f(K) + Q))$	$O(kM(Q + K))$	$O(kMf(K) + \text{training})$	$O(\text{training})$
F	$O((k+1)f(Q) + g(Q))$	$O(g'(Q))$	$O(f(K) + g(K))$	$O(g'(K))$
Total	$O(Mf(K) + Q + f(Q) + g(Q))$	$O(M(Q + K) + g'(Q))$	$O(Mf(K) + g(K))$	$O(g'(K))$

TABLE I

COMPARISON OF EXECUTION TIME FOR A CLASSICAL/QUANTUM CLASSIFICATION ALGORITHM WITH/WITHOUT A DATA REDUCTION STEP. STEPS A-E ARE THE DATA LOADING, REDUCTION AND TRAINING; STEP F IS PREDICTION. $K = N \times N \times T$ IS THE VIDEO SIZE OF THE INITIAL DATASET, $Q = 2^q$ IS THE “VIDEO” SIZE OF THE REDUCED DATASET AND M IS THE SIZE (NUMBER OF “VIDEOS”) OF THE TRAINING SET FOR EACH ONE OF THE k CLASSES.

f, g, g' ARE EXECUTION TIME FUNCTIONS f : QUANTUM DATA-LOADING, g, g' : QUANTUM/CLASSICAL CLASSIFICATION. $O(g(n)) < O(g'(n))$ IS ASSUMED, OTHERWISE THE CLASSICAL WOULD BE THE CLEAR CHOICE.

IV. EXPERIMENTAL RESULTS

To showcase that the data reduction step retains enough information so that a quantum algorithm may *then* efficiently process the reduced data we evaluated the proposed algorithm on a sample video dataset.

Dataset and pre-processing We evaluate the proposed algorithm using a small subset of the “20BN-jester Dataset V1” that contains labeled video clips showing humans performing predefined hand gestures [20]. We crop and down-scale each video so that all frames are 64×64 pixels and all videos 32 frames long $N = 64, T = 32$. We perform the simulations using two (“Swiping Left”, “Pulling Hand In”), three (“Swiping Left”, “Pulling Hand In”, “Pushing Hand Away”) and four (“Swiping Left”, “Pulling Hand In”, “Pushing Hand Away”, “Swiping Right”) of the available classes ($k = 2, 3, 4$). We used training sets of sizes $M \times k$ for values $M = 20, 40, 60, 80, 120$. We encode the reduced data using q qubits for values $q = [4..17]$. We stress that the actual data reduction is logarithmic with the number of qubits as shown in Figure 2a. Encoding a video using $q = 4$ qubits corresponds to a $(64 * 64 * 32) / 2^4 = 8192$ reduction.

Simulation For each combination of the parameters (training set size, number of classes, number of qubits that encode the reduced data) we simulated the quantum procedures. We run the simulations for at least 100 iterations and averaged out the results.

Results In Figures 2b,2c and 2d we report the accuracy achieved for each simulation case. We observe that in all cases there is enough information extracted for meaningful video classification even for as few as 10 qubits, for appropriate training sizes. A quantum register of $q = 10$ qubits encodes $2^{10} = 1024$ pixels, corresponding to just $2^{10} / (64 \times 64 \times 32) \approx 0.8\%$ of the initial video. This is a significant reduction. It appears that after approximately 13 qubits no much more information is extracted. We have to assume that this is at least partly due to our use of a naive classification method as one would expect that for a non-reduced dataset (corresponding to $q = 17$) higher values of accuracy should be achievable.

The training size also plays a crucial role. For training sizes of less than 40 videos per class the classification accuracy was very low. As the training size increased so did, in general, the classification accuracy. Exceptions to this were observed as k became larger and we hypothesize this is due to the large similarities between the classes. The classes (“Swiping Left”, “Swiping Right”) and (“Pulling Hand In”, “Pushing Hand Away”) are—in practice—time-reversed versions of each other (i.e. a “Pulling Hand In” video played in reverse could be classified as “Pushing Hand Away”) On the other hand, it is a strong point of the proposed method that it can accurately discriminate between these very time-symmetric classes of videos.

V. DISCUSSION

The results obtained suggest that a very small number of qubits are able to contain enough information so that videos can be successfully classified. Even with the naive

classification method that we employed at the classification step, the procedure achieved high accuracy values, with a large data reduction and high performance. Even at $q = 5$ qubits there is a statistically significant deviation from random-choice classification showing that meaningful information was extracted using just 0.02% of the initial video, corresponding to reduction of the video size by a factor of 4000. The quantum method used is therefore capable of accurately identifying the most significant pixels using a very efficient procedure.

There are many ways that the above results could be improved. A better classification method, either classical or quantum, is an obvious first step as the method we employed as proof of concept is very simplistic. A more accurate extraction of the significant bits could also be achieved by more detailed methods; for example, by using cross-validation in place of the simple video averaging.

Extending this method to different datasets (e.g. time series data or 3D image data) or to different outcomes (e.g. forecasting or object detection) should also be straightforward. We believe that applicability of this method to different domains is worth investigating.

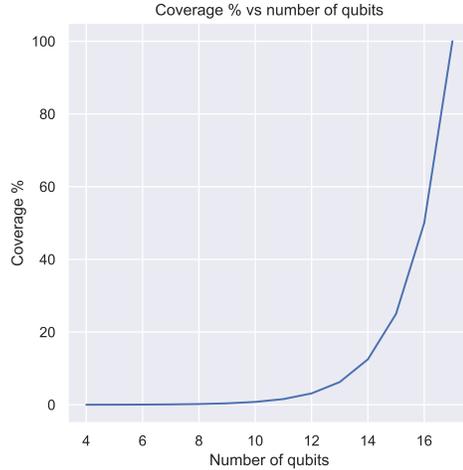
Another aspect worth exploring is the accuracy plateau that is observed after using 12 qubits. This might be coincidental to the naive classification method that we employ here but other sources should be investigated (e.g. the existence of barren plateaus).

Finally, as this method uses very few quantum operations and a few quantum bits, it would be straightforward to implement it on a real quantum device such as on an IBM qpu. Thus, empirical timing results could be considered and compared to their classical counterparts.

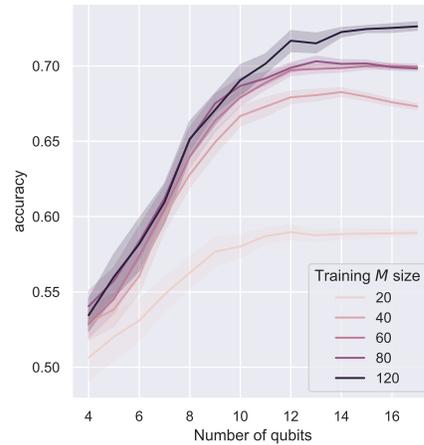
Acknowledgment This work is partially supported by the Greek Secretariat for Research and Innovation and the EU, Project SignGuide: Automated Museum Guidance using Sign Language T2EDK-00982 within the framework of “Competitiveness, Entrepreneurship and Innovation” (EPAnEK) Operational Programme 2014–2020.

REFERENCES

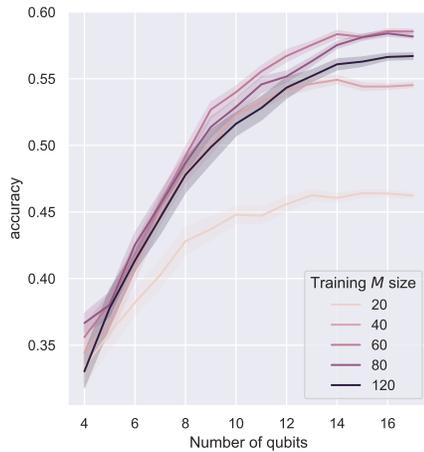
- [1] Yijie Dang, Nan Jiang, Hao Hu, Zhuoxiao Ji, and Wenyin Zhang. “Image classification based on quantum K-Nearest-Neighbor algorithm”. *Quantum Information Processing* **17**, 239 (2018).
- [2] Nan-Run Zhou, Xiu-Xun Liu, Yu-Ling Chen, and Ni-Suo Du. “Quantum K-Nearest-Neighbor Image Classification Algorithm Based on K-L Transform”. *International Journal of Theoretical Physics* **60**, 1209–1224 (2021).
- [3] Maria Schuld, Alex Bocharov, Krysta Svore, and Nathan Wiebe. “Circuit-centric quantum classifiers”. *Physical Review A* **101**, 032308 (2020). arXiv:1804.00633.
- [4] Mateusz Ostaszewski, Przemysław Sadowski, and Piotr Gawron. “Quantum image classification using principal component analysis”. *Theoretical and Applied Informatics* **27**, 1–12 (2015). arXiv:1504.00580.
- [5] Edward Farhi and Hartmut Neven. “Classification with Quantum Neural Networks on Near Term Processors”. *Quantum Review Letters* **Vol. 1 No. 2** (2020), 10.37686/qr.l.v1i2.80 (2020).
- [6] Tak Hur, Leeseok Kim, and Daniel K. Park. “Quantum convolutional neural network for classical data classification” (2021). arXiv:2108.00661.
- [7] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. “Quantum embeddings for machine learning” (2020). arXiv:2001.03622.



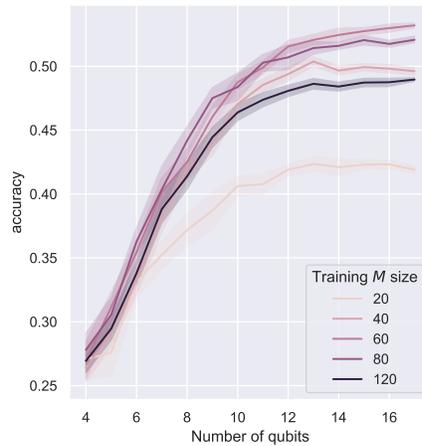
(a) Percentage of the initial video that a quantum register of q qubits encodes.



(b) Accuracy vs number of qubits for 2 classes



(c) Accuracy vs number of qubits for 3 classes



(d) Accuracy vs number of qubits for 4 classes

- [8] Kostas Blekos and Dimitrios Kosmopoulos. “A Quantum 3D Convolutional Neural Network with Application in Video Classification”. In *Advances in Visual Computing*. Pages 601–612. Lecture Notes in Computer Science Cham (2021). Springer International Publishing.
- [9] Aram W. Harrow. “Small quantum computers and large classical data sets” (2020). arXiv:2004.00026.
- [10] Tobias Haug, Chris N. Self, and M. S. Kim. “Large-scale quantum machine learning” (2021). arXiv:2108.01039.
- [11] Jin-Min Liang, Shu-Qian Shen, Ming Li, and Lei Li. “Variational quantum algorithms for dimensionality reduction and classification”. *Physical Review A* **101**, 032323 (2020). arXiv:1910.12164.
- [12] John A. Cortese and Timothy M. Braje. “Loading Classical Data into a Quantum Computer” (2018). arXiv:1803.01958.
- [13] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. “Quantum machine learning”. *Nature* **549**, 195–202 (2017). arXiv:1611.09347.
- [14] Scott Aaronson. “Read the fine print”. *Nature Physics* **11**, 291–293 (2015).
- [15] Iordanis Kerenidis and Anupam Prakash. “Quantum Recommendation Systems” (2016). arXiv:1603.08675.
- [16] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. “Quantum Random Access Memory”. *Physical Review Letters* **100**, 160501 (2008).
- [17] Iordanis Kerenidis, Jonas Landman, Alessandro Luongo, and Anupam Prakash. “Q-means: A quantum algorithm for unsupervised machine learning” (2018). arXiv:1812.03584.
- [18] Anurag Anshu, Zeph Landau, and Yunchao Liu. “Distributed quantum inner product estimation” (2021). arXiv:2111.03273.
- [19] Ryan O’Donnell and John Wright. “Efficient quantum tomography”. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. Pages 899–912. STOC ’16 New York, NY, USA (2016). Association for Computing Machinery.
- [20] Joanna Materzynska, Guillaume Berger, Ingo Bax, and Roland Memisevic. “The jester dataset: A large-scale video dataset of human gestures”. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. Pages 2874–2882. IEEE Computer Society (2019).