Future Generation Computer Systems 160 (2024) 54–64

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Quantum Annealing for Computer Vision minimization problems

Shahrokh Heidari [a,b,*], Michael J. Dinneen [a], Patrice Delmas [a,b]

[a] School of Computer Science, The University of Auckland, Auckland, New Zealand
[b] Intelligent Vision Systems Lab, The University of Auckland, Auckland, New Zealand

ARTICLE INFO

Keywords:
Quantum Annealing
Quantum Computer Vision
Computer Vision
Discrete Minimization Models
Stereo Matching

ABSTRACT

Computer Vision (CV) labeling problems play a pivotal role in low-level vision. For decades, it has been known that these problems can be elegantly formulated as discrete energy-minimization problems derived from probabilistic graphical models such as Markov Random Fields (MRFs). Despite recent advances in MRF inference algorithms (such as graph-cut and message-passing methods), the resulting energy-minimization problems are generally viewed as intractable. The emergence of quantum computations, which offer the potential for faster solutions to certain problems than classical methods, has led to an increased interest in utilizing quantum properties to overcome intractable problems. Recently, there has also been a growing interest in Quantum Computer Vision (QCV), hoping to provide a credible alternative/assistant to deep learning solutions. This study investigates a new Quantum Annealing-based inference algorithm for CV discrete energy minimization problems. Our contribution is focused on Stereo Matching as a significant CV labeling problem. As a proof of concept, we also use a hybrid quantum–classical solver provided by D-Wave System to compare our results with the best classical inference algorithms in the literature. Our results show that Quantum Annealing can yield promising results for Stereo Matching problems, with improved accuracy on certain stereo images and competitive performance on others.

1. Introduction

Computer Vision (CV) is a field of study focusing on how computers gain high-level perception from digital images/videos, which can help decision-making in real-world environments. While humans routinely interpret the environment, enabling computers to perceive the real world from its representation through images/videos remains a largely unsolved problem. Many problems in CV are formulated as labeling problems. A CV labeling problem consists of a set of image features (such as pixels, edges, or image segments) on which we want to estimate quantities from a set of labels [1] (such as intensity in Image Restoration or disparity in Stereo Matching and Motion). Generally, CV labeling problems are modeled by a discrete minimization problem, where an objective function is defined to be optimized over a set of possible labeling solutions. When this objective function measures the badness, the optimization problem is often called energy minimization, and the objective function is referred to as an energy function [2]. Given the intrinsically tricky nature of CV minimization problems, researchers have always been looking for efficient algorithms to approximate the optimal solution as fast and accurately as possible. Thus, there has been significant development in minimization algorithms for CV problems from the classical methods in the 1990s, such as Simulated Annealing [3], Mean-field Annealing [4], and Iterated Conditional Modes (ICM) [5] to the recent state-of-the-art algorithms, such as graph-cut based [6–11] and message-passing based [12–14] approaches (we refer interested readers to the most recent comparative studies on CV minimization algorithms [15–19]). Despite being extensively researched and even considering the most recent advances using deep learning-based strategies [20], which are computationally expensive, CV labeling problems are still considered open problems with no prefect (optimal) solutions due to the extensive range of mathematics involved and the complexity of recovering unknowns from insufficient information.

Therefore, researchers have always been looking for alternatives to tackle the problem. With the advent of quantum computations which promise potentially lower-time complexity on certain problems than the best-classical counterparts [21–23], recent studies have focused on leveraging quantum properties to overcome intractable classical problems using Quantum Annealing (QA). D-Wave Systems was the first company to build a Quantum Processing Unit (QPU) that naturally approximates the ground state of a particular problem representation, namely Ising model [24]. The importance of Ising models is that one can solve a variety of NP-hard optimization problems by finding the corresponding ground state [25–27]. Despite the promising

* Corresponding author at: School of Computer Science, The University of Auckland, Auckland, New Zealand.
E-mail address: shei972@aucklanduni.ac.nz (S. Heidari).

https://doi.org/10.1016/j.future.2024.05.037
Received 11 December 2023; Received in revised form 12 April 2024; Accepted 15 May 2024
Available online 24 May 2024

experiments [21,22], D-Wave QPUs are specifically designed to solve optimization problems, making them less versatile than other quantum computation approaches. This restricts their application domain primarily to optimization and sampling tasks, while they may not be suitable for more general-purpose computing requirements. Also, D-Wave QPUs exhibit limited qubit connectivity, and the scarcity of available qubits has been consistently challenging, from the 128-qubit D-Wave One built in 2011 to the newly released 5000-qubit D-Wave Advantage. Therefore, large CV problems involving highly non-convex functions in a search space of many thousands of dimensions have not been widely explored to see if QA can provide advantages in real-world CV problems. In recent years, there has been a growing interest in Quantum Computer Vision (QCV), largely fueled by recent advancements in D-Wave QPU architectures and their capabilities in solving optimization problems, such as Classification [28–32], Synchronization [33,34], Tracking [35], Fitting [36,37], Detection [38], and Matching [39–43] problems. However, each method employs a distinct quantum model to represent the respective CV problem, allowing it to be minimized on a D-Wave QPU. A versatile framework for converting a CV problem into an appropriate quantum model holds significant value. Such a flexible solution not only simplifies the process of adapting various CV problems for quantum computation but also opens up new avenues for harnessing the power of quantum computation in addressing intricate optimization tasks.

In this study, we aim to focus on a challenging labeling problem, Stereo Matching, and provide a general-purpose quantum model that can be used for any CV labeling problem (such as Image Segmentation, Image Restoration, Image Registration, Optical Flow, Object Detection, and Image Inpainting). Due to the scarcity of available qubits on the current D-Wave QPUs, we use a D-Wave hybrid quantum–classical solver to show the feasibility of the proposed quantum model once enough qubits are available. Our findings show that QA can offer promising results in CV applications compared to the state-of-the-art CV minimization inference algorithms.

The paper is organized as follows: Section 2 briefly introduces Stereo Matching, an important CV labeling problem. In Section 3, we shift the focus to QA and D-Wave QPUs. Our general-purpose quantum solution to Stereo Matching and its proof of correctness are presented in Section 4. We provide experimental results and numerical evaluation in Section 5. Finally, Section 7 concludes the paper.

2. Stereo matching

The characteristics of binocular vision in humans allow for the simultaneous observation of a singular object by both eyes. This ability significantly contributes to the understanding of depth in the brain. The distance between our eyes, often referred to as “baseline”, facilitates slight variation in the perspective captured by each eye. Despite each eye observing a nearly identical image, a marginal displacement exists. The brain uses this displacement to perceive a 3D observation from the scene. Likewise, a stereo vision system is designed to replicate human vision mechanisms. This system comprises two horizontal cameras on the left and right sides, effectively simulating human binocular perception. Each camera in the system records an image that, while fundamentally similar, features a certain degree of displacement. This displacement, often called disparity, signifies the difference in the position of a 3D point, as observed from two different viewpoints (the left and right viewpoints) [44]. The main goal of implementing a stereo vision system is to construct a 3D model using the left and right stereo images. This procedure may encompass various stages, including Camera Calibration (optional), Rectification, Stereo Matching, and 3D Reconstruction [45] as shown in Fig. 1. Camera Calibration is the process of estimating specific parameters of a camera. These parameters are used to correct image distortions and determine an accurate relationship between a 3D point in the scene and its corresponding 2D projections in the images [46]. Before Stereo Matching, rectifying a pair

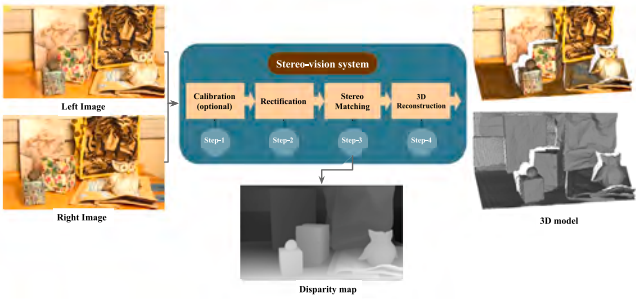


Fig. 1. Outline of stereo-vision steps: Calibration (optional), Rectification, Stereo Matching, and 3D Reconstruction.

of stereo images is essential to reduce the complexity of the underlying problem. The main goal of Stereo Matching is to match a given pixel in the left image with its corresponding pixel in the right image, where the corresponding pixels are the same projections of a 3D point in the real world. This process can be performed by searching for the corresponding pixels in a 2D search space, which is computationally expensive. Rectification transforms the 2D search space into a 1D search space. This significantly simplifies the correspondence problem, as the search for matching pixels can be reduced to a 1D search along the horizontal line of pixels rather than a 2D search in the entire image. Despite this search-space reduction, Stereo Matching represents the most computationally demanding component of a stereo vision system. A Stereo Matching algorithm estimates a disparity value for each pixel in the left image to determine its corresponding pixel in the right image. The final output is a disparity map in which regions with higher disparity values belong to real-world objects closer to the cameras, whereas those with lower disparity values belong to real-world objects farther away from the cameras. Regarding visualization, regions nearer and with greater disparity values appear brighter than those farther away with smaller disparity values (see Fig. 1).

Stereo Matching methods are broadly categorized into global and local approaches. While local methods prioritize speed, often at the cost of accuracy due to susceptibilities like local ambiguities and occlusions, global methods comprehensively consider the entire image during disparity computation. Although computationally demanding, they effectively address challenges such as occluded and textureless regions [44]. These methods typically lean on probabilistic graphical models, a potent blend of probability and graph theory, for their formalism [47]. Based on the defined probabilistic graphical model, an energy function is modeled which can be minimized to solve the Stereo Matching problem [47, p. 1612]. In the following, we provide the general form of a global Stereo Matching energy function, which can be adapted for any CV labeling problem (see the recent comparative study on CV labeling problems [19] for more information).

Let I_l and I_r be a pair of $n \times m$ stereo images, and $D = \{d_{min}, \dots, d_{max}\}$ be a set of positive integers, where d_{min} and d_{max} are the lowest and highest possible disparity values, respectively. Considering the left image I_l as the reference for which we want to compute a disparity map, the set of pixels is defined as (1). We also initialize N as a 4-neighborhood system defined in (2).

$$P = \{(i,j) \mid i \in \{0,\dots,n-1\}, j \in \{0,\dots,m-1\}\}, \tag{1}$$

$$N = \{\{(i,j),(i',j')\} \mid (i,j) \in P, \; (i',j') \in \{(i \pm 1, j),(i, j \pm 1)\}\}, \tag{2}$$

where, $0 \le i' < n, 0 \le j' < m$. In a global Stereo Matching model, the Stereo Matching problem is modeled by a labeling problem where each pixel in P is labeled by a disparity value in D [1, p. 5]. In fact, a “labeling” involves mapping from P to D. Such a labeling problem is defined by a discrete optimization problem, where an energy function is

defined to be minimized over a set of possible labeling solutions. This energy function has two terms. The first term penalizes the solutions when inconsistent with the data, and the second term imposes some constraints on spatial coherence [18, p. 1]. Let $\mathbf{w} \in D^{n\times m}$ be a vector of variables defined as $\mathbf{w} = (w_{i,j})_{(i,j)\in P}$, where $w_{i,j} \in D$. The global Stereo Matching energy function $F : D^{n\times m} \to \mathbb{R}^+$ is defined as (3).

$$F(\mathbf{w}) = \sum_{(i,j)\in P} \theta_{\{i,j\}}(w_{i,j}) + \lambda \sum_{\{(i,j),(i',j')\}\in N} \delta(w_{i,j}, w_{i',j'}) \tag{3}$$

where,

$$\theta_{i,j}(w_{i,j}) = \left| I_l(i,j) - I_r(i - w_{i,j}, j) \right|,$$

$$\delta(w_{i,j}, w_{i',j'}) = \begin{cases} 0, & \text{if } w_{i,j} = w_{i',j'}; \\ 1, & \text{otherwise.} \end{cases}$$

The first term is the Sum of Absolute Difference (SAD) matching cost function defined by $\theta_{\{i,j\}} : D \to \mathbb{R}^+$. When $\theta_{\{i,j\}}(w_{i,j})$ is (or close to) zero, it means the pixel (i,j) in the left image matches the pixel $(i - w_{i,j}, j)$ in the right image, and they are more likely to be the same projections of a 3D point in the real world. In the second term, $\delta : D^2 \to \{0,1\}$ is the penalty function that penalizes the variation of the disparities, adding one when the allocated disparities to a pair of neighboring pixels are not equal and zero otherwise. The second term assumes that the disparities of a neighborhood of pixels present some coherence and generally do not change abruptly [48]. Furthermore, $\lambda \in \mathbb{R}^+$, known as the smoothness factor, weighs the penalties given by the second term.

We aim to provide a general-purpose quantum model for the defined global Stereo Matching problem (3), which can be adapted to any CV labeling problem. Thus, we first give the preliminaries to describe this quantum model.

3. Quantum annealing

QA [49] is a specialized optimization technique that leverages principles from quantum mechanics to solve complex computational problems. In this model, quantum bits (qubits) are particles in a quantum dynamical system that evolve based on special forces acting on them. These forces are either internal (from interactions among qubits) or external (from other sources). Each state of a register of qubits has energy based on the applied forces. A time-dependent Hamiltonian is a mathematical representation of a system, providing information about the system's energy and detailing the forces acting upon it at any given time [24]. QA is a computational technique employed to discover the state of the system with the minimum energy as determined by the time-dependent Hamiltonian. Consequently, QA constitutes a computational paradigm known for its efficiency in addressing optimization problems and providing approximations to the optimal solutions. It is inspired by the concept of annealing in metallurgy, where a material is slowly cooled to minimize defects and reach a low-energy state. In QA, this cooling process is simulated by a QPU known as a quantum annealer which is based on a time-dependent Hamiltonian $H(t)$ that has three components [24]: *Initial Hamiltonian* H_I, where all qubits are in a superposition state. *Problem Hamiltonian* H_p, where the specific forces are defined to encode the objective function. The lowest-energy state of H_p is the solution that minimizes the objective function. *Adiabatic path* $s(t)$, which is a smooth function that decreases from 1 to 0, such as $s(t) = 1 - \frac{t}{t_f}$, where $s(t)$ decreases from 1 to 0 as t increases from 0 to some elapsed time t_f. During QA, the *Initial Hamiltonian* is slowly evolved along the *Adiabatic path* to the *Problem Hamiltonian* as $H(t) = s(t)H_I + (1-s(t))H_p$ [24], decreasing the influence of H_I over time to reach H_P as $s(t)$ goes from 1 to 0. D-Wave Systems was the first company to build a quantum annealer. To minimize/maximize an objective function using QA and a D-Wave QPU, it should be in a standard model like Ising or Quadratic Unconstrained Binary Optimization (QUBO) models [24]. Given a vector of n binary variables as $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0,1\}^n$, a QUBO model is represented as $H_{qubo}(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}$, where $\{0,1\}^n$ is a set of n binary values, and $\mathbf{Q}$ is an $n \times n$ matrix that can be chosen to be upper-diagonal. Therefore, $H_{qubo}(\mathbf{x})$ can be reformulated as (4).

$$H_{qubo}(\mathbf{x}) = \sum_i \mathbf{Q}_{i,i} x_i + \sum_{i<j} \mathbf{Q}_{i,j} x_i x_j. \tag{4}$$

The diagonal terms $\mathbf{Q}_{i,i}$ are the linear coefficients acting as the external forces, and the off-diagonal terms $\mathbf{Q}_{i,j}$ are the quadratic coefficients for the internal forces [24].

4. Quantum stereo matching

We introduce an equivalent QUBO model to the global Stereo Matching minimization problem (3) and provide proof of its correctness. Our idea draws inspiration from the approach employed by the D-Wave Ocean Software Development Kit (SDK) when handling discrete objective functions [50]. We first allocate $|D|$ binary variables to each pixel $(i,j) \in P$, where $|D|$ is the number of elements in D, $0 \le i \le n-1$, and $0 \le j \le m-1$. Therefore, we define $\mathbf{x} \in \{0,1\}^{nm|D|}$ as a vector of $nm|D|$ binary variables such that $\mathbf{x} = (x_{i,j,d})$ for all $(i,j) \in P$ and $d \in D$. Let our QUBO model be defined as (5).

$$\begin{aligned} H(\mathbf{x}) = \alpha \sum_{(i,j)\in P} \left(1 - \sum_{d\in D} x_{i,j,d}\right)^2 + \sum_{(i,j)\in P} \sum_{d\in D} \theta_{\{i,j\}}(d) x_{i,j,d} \\ + \lambda \sum_{\{(i,j),(i',j')\}\in N} \sum_{d_1\in D} \sum_{d_2\in D} \delta(d_1,d_2) x_{i,j,d_1} x_{i',j',d_2}, \end{aligned} \tag{5}$$

where $\alpha > \left(\sum_{(i,j)\in P} \max\{\theta_{\{i,j\}}(d) \mid d \in D\}\right) + \lambda|N|$, and $|N|$ is the number of elements in N. We set $\mathbf{x}^* = \arg\min_{\mathbf{x}} H(\mathbf{x})$ and define a vector of nm integer values as $\mathbf{w}^* = (w^*_{i,j})_{(i,j)\in P}$, where $w^*_{i,j} = d$ if $x^*_{i,j,d} = 1$. Then, $\mathbf{w}^*$ minimizes the global Stereo Matching energy function (3).

Proof of correctness

Eq. (5) has three parts. The first part guarantees each pixel is assigned a unique disparity value from D. The second calculates the cost of the assigned disparity values to the pixels. The third part encodes the defined contextual constraint.

Definition 1. $\mathbf{x}$ is called feasible if and only if $\sum_{d\in D} x_{i,j,d} = 1$ for all pixels $(i,j) \in P$. We denote a feasible $\mathbf{x}$ by $\mathbf{x}'$.

Definition 1 states that given a pixel $(i,j) \in P$, its corresponding vector of binary variables $(x'_{i,j,d_{min}}, \dots, x'_{i,j,d_{max}})$ has only one value of "1" in its values, making it possible to label each pixel uniquely by a disparity $d \in D$. Hence, the allocated disparity to a pixel $(i,j) \in P$ is d if $x'_{i,j,d} = 1$.

Definition 2. Given $\mathbf{x}'$, the corresponding integer vector $\mathbf{w}' = (w'_{i,j})_{(i,j)\in P}$ is called a *labeling*, where $w'_{i,j} = d$ if $x'_{i,j,d} = 1$.

Lemma 1. *Given a feasible* $\mathbf{x}'$ *and its corresponding labeling* $\mathbf{w}'$*, the equality* $H(\mathbf{x}') = F(\mathbf{w}')$ *holds, where* F *is the global Stereo Matching energy function in* (3).

Proof. Considering $H(\mathbf{x}')$ in (5),

- Since $\mathbf{x}'$ is feasible, $\sum_{d\in D} x'_{i,j,d} = 1$ for all pixels $(i,j) \in P$ by Definition 1. Therefore, we have

$$\alpha \sum_{(i,j)\in P} \left(1 - \sum_{d\in D} x'_{i,j,d}\right)^2 = \alpha \sum_{(i,j)\in P} (1-1)^2 = 0.$$

- Given a pixel $(i,j) \in P$, only one variable in the vector $(x'_{i,j,d_{min}}, \dots, x'_{i,j,d_{max}})$ is one, and all the others are zero. This non-zero variable is $x'_{i,j,w'_{i,j}}$ by Definition 2. Therefore, we have

$$\sum_{(i,j)\in P} \sum_{d\in D} \theta_{\{i,j\}}(d) x'_{i,j,d} = \sum_{(i,j)\in P} \theta_{\{i,j\}}(w'_{i,j}) x'_{i,j,w'_{i,j}}$$

$$= \sum_{(i,j)\in P} \theta_{\{i,j\}}(w'_{i,j}).$$

- Given $\{(i,j),(i',j')\} \in N$, the two corresponding vectors of binary variables are
  - $(i,j) : (x_{i,j,d_{min}}, \dots, x_{i,j,d_{max}})$
  - $(i',j') : (x'_{i',j',d_{min}}, \dots, x'_{i',j',d_{max}})$

  Since $\mathbf{x}'$ is feasible, only one of the variables in each vector is one, and the others are zero. These variables are $x'_{i,j,w'_{i,j}}$ and $x'_{i',j',w'_{i',j'}}$, respectively, by Definition 2. Thus, we can write

$$\begin{aligned}
&\lambda \sum_{\{(i,j),(i',j')\}\in N} \sum_{d_1\in D} \sum_{d_2\in D} \delta(d_1,d_2) x'_{i,j,d_1} x'_{i',j',d_2} \\
&= \lambda \sum_{\{(i,j),(i',j')\}\in N} \delta(w'_{i,j}, w'_{i'j'}) x'_{i,j,w'_{i,j}} x'_{i',j',w'_{i'j'}} \\
&= \lambda \sum_{\{(i,j),(i',j')\}\in N} \delta(w'_{i,j}, w'_{i'j'}).
\end{aligned}$$

Therefore, we can rewrite $H(\mathbf{x}')$ as follows.

$$H(\mathbf{x}') = \sum_{(i,j)\in P} \theta_{\{i,j\}}(w'_{i,j}) + \lambda \sum_{\{(i,j),(i',j')\}\in N} \delta(w'_{i,j}, w'_{i'j'}) = F(\mathbf{w}'). \quad \square$$

Lemma 2. *Let* $\mathbf{x}^* = \arg\min_{\mathbf{x}} H(\mathbf{x})$. $\mathbf{x}^*$ *is feasible.*

Proof. For ease of reference, we rewrite $H(\mathbf{x})$ as follows:

$$H(\mathbf{x}) = \alpha \mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{x}),$$

where

$$\begin{aligned}
\mathcal{A}(\mathbf{x}) &= \sum_{(i,j)\in P} \left(1 - \sum_{d\in D} x_{i,j,d}\right)^2, \\
\mathcal{B}(\mathbf{x}) &= \sum_{(i,j)\in P} \sum_{d\in D} \theta_{\{i,j\}}(d) x_{i,j,d} \\
&\quad + \lambda \sum_{\{(i,j),(i',j')\}\in N} \sum_{d_1\in D} \sum_{d_2\in D} \delta(d_1,d_2) x_{i,j,d_1} x_{i',j',d_2}.
\end{aligned}$$

Towards a contradiction, suppose that $\mathbf{x}^*$ is not feasible. In this case, $\mathcal{A}(\mathbf{x}^*) \neq 0$ and it is non-negative. Therefore,

$$H(\mathbf{x}^*) = \alpha \mathcal{A}(\mathbf{x}^*) + \mathcal{B}(\mathbf{x}^*). \tag{6}$$

Given a feasible $\mathbf{x}'$, $\mathcal{A}(\mathbf{x}') = 0$, and we have

$$H(\mathbf{x}') = \mathcal{B}(\mathbf{x}'). \tag{7}$$

Since $\mathbf{x}'$ is feasible, $\mathcal{B}(\mathbf{x}')$ adds penalty values up to a maximum of $\left(\sum_{(i,j)\in P} \max\{\theta_{i,j}(d) \mid d \in D\}\right) + \lambda|N|$. Considering (7), we have

$$H(\mathbf{x}') \leq \left(\sum_{(i,j)\in P} \max\{\theta_{i,j}(d) \mid d \in D\}\right) + \lambda|N|. \tag{8}$$

We know that $\alpha > \left(\sum_{(i,j)\in P} \max\{\theta_{i,j}(d) \mid d \in D\}\right) + \lambda|N|$, $\mathcal{A}(\mathbf{x}^*)$ is non-zero and non-negative, and $\mathcal{B}(\mathbf{x}^*)$ is non-negative. Considering (6), we can write

$$H(\mathbf{x}^*) = \alpha \mathcal{A}(\mathbf{x}^*) + \mathcal{B}(\mathbf{x}^*) \tag{9}$$

$$> \left(\sum_{(i,j)\in P} \max\{\theta_{i,j}(d) \mid d \in D\}\right) + \lambda|N|. \tag{10}$$

The following statement is true by (8) and (9): $H(\mathbf{x}') < H(\mathbf{x}^*)$, which is a contradiction. Therefore, $\mathbf{x}^*$ is feasible. $\square$

Theorem 1. *Given* $\mathbf{w}^*$ *as the corresponding labeling of* $\mathbf{x}^*$, $\mathbf{w}^*$ *minimizes the global Stereo Matching energy function* F *defined in* (3).

Proof. Towards a contradiction, we suppose that $\mathbf{w}^*$ does not minimize F. In this case, there must be a feasible $\mathbf{x}'$ for which its corresponding labeling $\mathbf{w}'$ minimizes F. Therefore, we have $F(\mathbf{w}') < F(\mathbf{w}^*)$. Since $\mathbf{x}'$ and $\mathbf{x}^*$ are both feasible (see Lemma 2), we have $H(\mathbf{x}') < H(\mathbf{x}^*)$ by Lemma 1. This is a contradiction because in this case $\mathbf{x}^* \neq \arg\min_{\mathbf{x}} H(\mathbf{x})$. $\square$

Eq. (5) is versatile and can be adapted for a variety of CV labeling problems by replacing P with any desired set of image features and replacing D with an appropriate set of labels depending on the application. Then, the first and second terms in Eq. (3) can be defined accordingly. The modified QUBO remains consistent with the QUBO model described in Eq. (5). This adaptability showcases the broader applicability of the model, making it a flexible tool for addressing a range of CV labeling challenges.

Example 1. We provide a simple example to show how our quantum model (5) can be modeled and minimized via QA. Fig. 2(a) and Fig. 2(b) show a pair of (3×4)-sized stereo images with $D = \{0,1\}$. The intensity values for the left and right images are shown on the pixels. The corresponding pixel coordinates are illustrated in Fig. 2(c). Without loss of generality, we ignore the first column of pixels in the left image since d_{max} is 1, and we would obtain negative coordinates to match this column in the right image. The main goal is to compute the disparity map allocated to the shown red square in Fig. 2(a). Fig. 2(d) shows the ground truth disparity map.

Considering (1) and (2), we first define P and N follows:

$$\begin{aligned}
P = &\{(1,0),(2,0),(3,0),(1,1),(2,1),(3,1),(1,2),\\
&(2,2),(3,2)\}
\end{aligned}$$

$$\begin{aligned}
N = &\{\{(1,0),(2,0)\},\{(1,0),(1,1)\},\{(2,0),(3,0)\},\\
&\{(2,0),(2,1)\},\{(3,0),(3,1)\},\{(1,1),(2,1)\},\\
&\{(1,1),(1,2)\},\{(2,1),(3,1)\},\{(2,1),(2,2)\},\\
&\{(3,1),(3,2)\},\{(1,2),(2,2)\},\{(2,2),(3,2)\}\}.
\end{aligned}$$

The numbers of pixels and disparities are 9 and 2, respectively. Therefore, We define a vector of 18 binary variables as $\mathbf{x} = \{0,1\}^{18}$:

$$\begin{aligned}
\mathbf{x} = (&x_{1,0,0}, x_{1,0,1}, x_{2,0,0}, x_{2,0,1}, x_{3,0,0}, x_{3,0,1}, x_{1,1,0},\\
&x_{1,1,1}, x_{2,1,0}, x_{2,1,1}, x_{3,1,0}, x_{3,1,1}, x_{1,2,0}, x_{1,2,1},\\
&x_{2,2,0}, x_{2,2,1}, x_{3,2,0}, x_{3,2,1}).
\end{aligned}$$

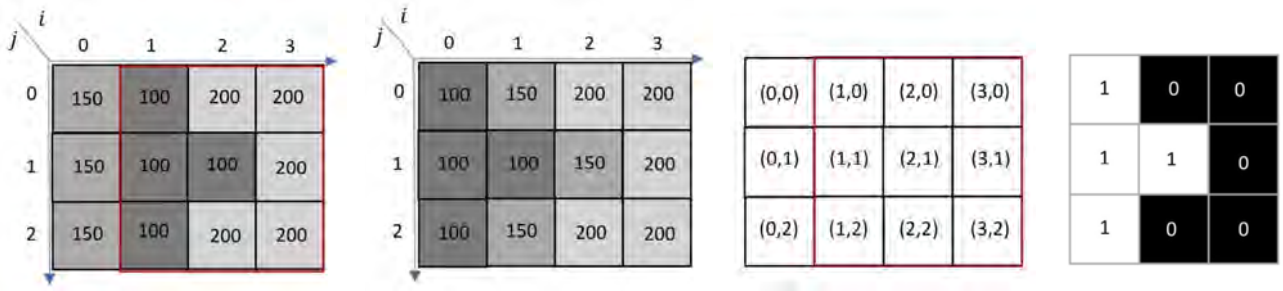
We set $\lambda = 10$ and $\alpha = 200$ by which we have $\alpha > \left(\sum_{(i,j)\in P} \max\{\theta_{i,j}(d) | d \in D\}\right) + \lambda|N|$. The QUBO model (5) is formulated as follows:

$$\begin{aligned}
H(\mathbf{x}) = 200 &\sum_{(i,j)\in P} \left(1 - \sum_{d\in D} x_{i,j,d}\right)^2 + \sum_{(i,j)\in P} \sum_{d\in D} \theta_{i,j}(d) x_{i,j,d} \\
&+ 10 \sum_{\{(i,j),(i',j')\}\in N} \sum_{d_1\in D} \sum_{d_2\in D} \delta(d_1,d_2) x_{i,j,d_1} x_{i',j',d_2}.
\end{aligned}$$

The QUBO model $H(\mathbf{x})$ has three terms denoted by H_1, H_2, and H_3 from left to right, respectively. The following shows each term's expansion separately. We then add them all at the end. We start with the first term denoted by H_1.

$$\begin{aligned}
H_1(\mathbf{x}) = 200(&- x_{1,0,0} - x_{1,0,1} + 2x_{1,0,0}x_{1,0,1} + 1 \\
&- x_{2,0,0} - x_{2,0,1} + 2x_{2,0,0}x_{2,0,1} + 1 \\
&- x_{3,0,0} - x_{3,0,1} + 2x_{3,0,0}x_{3,0,1} + 1 \\
&- x_{1,1,0} - x_{1,1,1} + 2x_{1,1,0}x_{1,1,1} + 1 \\
&- x_{2,1,0} - x_{2,1,1} + 2x_{2,1,0}x_{2,1,1} + 1 \\
&- x_{3,1,0} - x_{3,1,1} + 2x_{3,1,0}x_{3,1,1} + 1 \\
&- x_{1,2,0} - x_{1,2,1} + 2x_{1,2,0}x_{1,2,1} + 1 \\
&- x_{2,2,0} - x_{2,2,1} + 2x_{2,2,0}x_{2,2,1} + 1 \\
&- x_{3,2,0} - x_{3,2,1} + 2x_{3,2,0}x_{3,2,1} + 1)
\end{aligned}$$

Next, we expand the second term as H_2.



(a) (b) (c) (d)

Fig. 2. (a) the left stereo image, (b) the right stereo image, (c) the pixel coordinates, (d) the corresponding disparity map.

$$
\begin{aligned}
&H_2(\mathbf{x}) = \\
&= |I_l(1,0) - I_r(1,0)|x_{1,0,0} + |I_l(1,0) - I_r(0,0)|x_{1,0,1} \\
&+ |I_l(2,0) - I_r(2,0)|x_{2,0,0} + |I_l(2,0) - I_r(1,0)|x_{2,0,1} \\
&+ |I_l(3,0) - I_r(3,0)|x_{3,0,0} + |I_l(3,0) - I_r(2,0)|x_{3,0,1} \\
&+ |I_l(1,1) - I_r(1,1)|x_{1,1,0} + |I_l(1,1) - I_r(0,1)|x_{1,1,1} \\
&+ |I_l(2,1) - I_r(2,1)|x_{2,1,0} + |I_l(2,1) - I_r(1,1)|x_{2,1,1} \\
&+ |I_l(3,1) - I_r(3,1)|x_{3,1,0} + |I_l(3,1) - I_r(2,1)|x_{3,1,1} \\
&+ |I_l(1,2) - I_r(1,2)|x_{1,2,0} + |I_l(1,2) - I_r(0,2)|x_{1,2,1} \\
&+ |I_l(2,2) - I_r(2,2)|x_{2,2,0} + |I_l(2,2) - I_r(1,2)|x_{2,2,1} \\
&+ |I_l(3,2) - I_r(3,2)|x_{3,2,0} + |I_l(3,2) - I_r(2,2)|x_{3,2,1}
\end{aligned}
$$

$$
\begin{aligned}
H_2(\mathbf{x}) &= 50x_{1,0,0} + 50x_{2,0,1} + 50x_{2,1,0} \\
&+ 50x_{3,1,1} + 50x_{1,2,0} + 50x_{2,2,1}.
\end{aligned}
$$

Finally, we compute the third term as H_3:

$$
\begin{aligned}
&H_3(\mathbf{x}) = \\
&= 10(x_{1,0,0}x_{2,0,1} + x_{1,0,1}x_{2,0,0} \\
&+ x_{1,0,0}x_{1,1,1} + x_{1,0,1}x_{1,1,0} + x_{2,0,0}x_{3,0,1} \\
&+ x_{2,0,1}x_{3,0,0} + x_{2,0,0}x_{2,1,1} + x_{2,0,1}x_{2,1,0} \\
&+ x_{3,0,0}x_{3,1,1} + x_{3,0,1}x_{3,1,0} + x_{1,1,0}x_{2,1,1} \\
&+ x_{1,1,1}x_{2,1,0} + x_{1,1,0}x_{1,2,1} + x_{1,1,1}x_{1,2,0} \\
&+ x_{2,1,0}x_{3,1,1} + x_{2,1,1}x_{3,1,0} + x_{2,1,0}x_{2,2,1} \\
&+ x_{2,1,1}x_{2,2,0} + x_{3,1,0}x_{3,2,1} + x_{3,1,1}x_{3,2,0} \\
&+ x_{1,2,0}x_{2,2,1} + x_{1,2,1}x_{2,2,0} + x_{2,2,0}x_{3,2,1} \\
&+ x_{2,2,1}x_{3,2,0}).
\end{aligned}
$$

Adding the three terms together, we have the main QUBO model as follows:

$$
\begin{aligned}
H(\mathbf{x}) = &- 150x_{1,0,0} - 200x_{1,0,1} - 200x_{2,0,0} - 150x_{2,0,1} \\
&- 200x_{3,0,0} - 200x_{3,0,1} - 200x_{1,1,0} - 200x_{1,1,1} \\
&- 150x_{2,1,0} - 200x_{2,1,1} - 200x_{3,1,0} - 150x_{3,1,1} \\
&- 150x_{1,2,0} - 200x_{1,2,1} - 200x_{2,2,0} - 150x_{2,2,1} \\
&- 200x_{3,2,0} - 200x_{3,2,1} \\
&+ 400x_{1,0,0}x_{1,0,1} + 400x_{2,0,0}x_{2,0,1} \\
&+ 400x_{3,0,0}x_{3,0,1} + 400x_{1,1,0}x_{1,1,1} \\
&+ 400x_{2,1,0}x_{2,1,1} + 400x_{3,1,0}x_{3,1,1} \\
&+ 400x_{1,2,0}x_{1,2,1} + 400x_{2,2,0}x_{2,2,1} \\
&+ 400x_{3,2,0}x_{3,2,1} + 1800 \\
&+ 10x_{1,0,0}x_{2,0,1} + 10x_{1,0,1}x_{2,0,0} + 10x_{1,0,0}x_{1,1,1} \\
&+ 10x_{1,0,1}x_{1,1,0} + 10x_{2,0,0}x_{3,0,1} + 10x_{2,0,1}x_{3,0,0} \\
&+ 10x_{2,0,0}x_{2,1,1} + 10x_{2,0,1}x_{2,1,0} + 10x_{3,0,0}x_{3,1,1} \\
&+ 10x_{3,0,1}x_{3,1,0} + 10x_{1,1,0}x_{2,1,1} + 10x_{1,1,1}x_{2,1,0} \\
&+ 10x_{1,1,0}x_{1,2,1} + 10x_{1,1,1}x_{1,2,0} + 10x_{2,1,0}x_{3,1,1} \\
&+ 10x_{2,1,1}x_{3,1,0} + 10x_{2,1,0}x_{2,2,1} + 10x_{2,1,1}x_{2,2,0} \\
&+ 10x_{3,1,0}x_{3,2,1} + 10x_{3,1,1}x_{3,2,0} + 10x_{1,2,0}x_{2,2,1} \\
&+ 10x_{1,2,1}x_{2,2,0} + 10x_{2,2,0}x_{3,2,1} + 10x_{2,2,1}x_{3,2,0}.
\end{aligned}
$$

Giving $H(\mathbf{x})$ to the D-Wave Ocean SDK for the QPU minimization, we obtain the optimal solution $\mathbf{x}^* = \arg\min_{\mathbf{x}} H(\mathbf{x})$ as follows:

- $x^*_{1,0,0} = 0$,
- $x^*_{1,0,1} = 1$,
- $x^*_{2,0,0} = 1$,
- $x^*_{2,0,1} = 0$,
- $x^*_{3,0,0} = 1$,
- $x^*_{3,0,1} = 0$,
- $x^*_{1,1,0} = 0$,
- $x^*_{1,1,1} = 1$,
- $x^*_{2,1,0} = 0$,
- $x^*_{2,1,1} = 1$,
- $x^*_{3,1,0} = 1$,
- $x^*_{3,1,1} = 0$,
- $x^*_{1,2,0} = 0$,
- $x^*_{1,2,1} = 1$,
- $x^*_{2,2,0} = 1$,
- $x^*_{2,2,1} = 0$,
- $x^*_{3,2,0} = 1$,
- $x^*_{3,2,1} = 0$,

We used the D-Wave default parameter settings for the hardware properties and initialized the number of sample-reads as 1000. Given a pixel $(i, j) \in P$, if $x^*_{i,j,d} = 1$ for $d \in D$, then d is the allocated disparity to the pixel (i, j). Therefore, we have the following disparities for the pixels:

- $(1,0) \leftarrow 1$
- $(2,0) \leftarrow 0$
- $(3,0) \leftarrow 0$
- $(1,1) \leftarrow 1$
- $(2,1) \leftarrow 1$
- $(3,1) \leftarrow 0$
- $(1,2) \leftarrow 1$
- $(2,2) \leftarrow 0$
- $(3,2) \leftarrow 0$

which match the corresponding ground-truth disparities shown in Fig. 2(d). Fig. 3 illustrates the corresponding D-Wave minor embedding for the defined QUBO, obtained by the D-Wave Inspector tool.

5. Evaluation and experimental results on stereo image patches

5.1. Qubit complexity

D-Wave quantum computers have showcased remarkable potential in solving optimization problems. However, one significant challenge they face is the limited availability of qubits. D-Wave QPUs employ QA to find/estimate the global minimum of a QUBO/Ising model. While effective for specific problem types, this approach often requires a large number of qubits, and the current generation of D-Wave QPUs have constraints on the number of qubits that can be utilized. Consequently, proposing a QUBO model with fewer variables is paramount as it addresses the current limitations in qubit availability, enables the solution of larger and more complex problems, widens access to QA, and enhances the robustness and practicality of QA technology

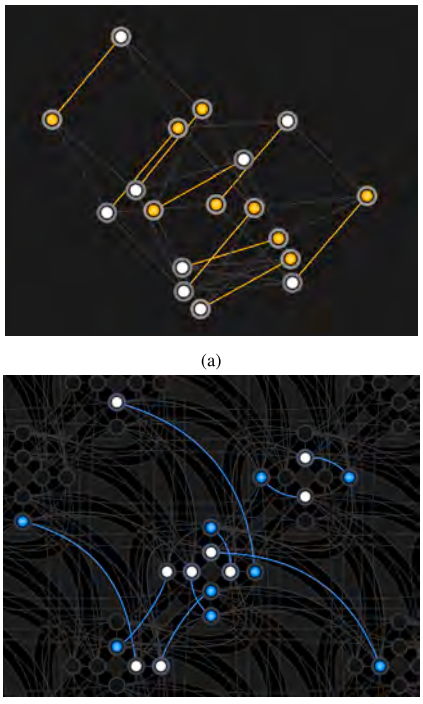


Fig. 3. D-Wave minor embedding for the given QUBO example. (a) the QUBO graph, and (b) the QPU graph.

in solving real-world optimization challenges. Recall P as the set of pixels for a pair of stereo images with size $n \times m$, and D as the set of possible disparities values, where $|P| = nm$ and $|D| = k$ denotes the number of elements in P and D, respectively. Given the defined vector of binary variables in (5), the number of QUBO variables in our general-purpose quantum model is nmk. Table 1 compares our quantum model with the existing labeling-based quantum solutions that can be utilized for Stereo Matching. Cruz-Santos et al. [39] and Heidari et al. [40] models are based on the minimum cut problem, and Heidari et al. [51] approach reduces a CV labeling problem to the minimum multi-way cut problem. Specifically,

- Cruz-Santos et al. [39] formulated a specific type of Stereo Matching problem as a quantum model. They first modeled the Stereo Matching problem as a labeling problem. Inspired by classical Stereo Matching approaches, the authors constructed a graph for which the minimum cut provided the optimal solution for minimizing the defined objective function. Classically, finding the minimum cut on an arbitrary undirected weighted graph is trivial, and the cut with the minimum cost can be precisely calculated in polynomial time.
- Heidari et al. [40] improved Cruz-Santos et al. quantum model [39] in terms of number of required variables. They used an existing quantum model in the literature to find the minimum cut on an arbitrary weighted graph and then incorporated the model to solve a Stereo Matching problem. Likewise, the minimization was to solve a trivial problem, finding the minimum cut on an undirected weighted graph.
- Inspired by solving a more complex CV problem using QA, Heidari et al. [51] introduced the first quantum model to find the minimum multi-way cut on an arbitrary weighted graph, which is known to an NP-Hard problem, and at least by current algorithmic means, cannot be solved in polynomial time on a Turing Machine. They then defined a CV labeling problem and minimized the corresponding quantum model using QA. Their method is capable of handling Stereo Matching problems. Therefore, we include their method in our qubit complexity comparisons.

Table 1
Qubit-complexity comparison of the proposed quantum Stereo Matching models.

Model	Qubit complexity
[39]	$7nmk + 9nm - 2nk - 2mk - 2n - 2m + 2$
[40]	$nmk + nm + 2$
[51]	$nmk + k^2$
Ours	nmk

The aim of comparing qubit complexities was to illustrate the superiority of our approach over existing potential quantum solutions. Given the same Stereo Matching energy function, minimizing these quantum models is expected to yield identical results due to the same minimization approach (QA), with differences only arising from the number of required variables. Thus, we include a set of classical minimization methods in our comparison in the next section.

5.2. Experimental results

Once a QUBO model is prepared, it needs to be embedded within the QPU hardware architecture for the minimization process. Embedding is the crucial step of mapping QUBO variables onto the available qubits on the hardware. Embedding can be challenging due to the relatively limited qubits and the restricted hardware connectivity. Consequently, it is common to chain two or more qubits together on the QPU to represent a single QUBO variable. While many real-world applications can successfully run on the D-Wave QPUs, there are cases where the input data is too large to be directly solved by QA, primarily because of the qubit scarcity. To overcome this size limitation, hybrid solvers combine classical and quantum approaches for problem-solving. D-Wave hybrid solvers can handle problems with a significantly higher number of variables than those directly solvable by a D-Wave QPU, offering a reliable estimate of the future accuracy of D-Wave QPUs once more qubits become available on the hardware. As a proof of concept, we utilize the Constrained Quadratic Model (CQM) D-Wave hybrid solver to minimize the proposed quantum Stereo Matching model. This solver has the capability of handling up to 500,000 QUBO variables, but it still poses restrictions on the size of the input stereo images and the number of disparities that can be processed. Therefore, we had to use cropped pairs of stereo images to analyze the performance of the Stereo Matching quantum model. We chose four pairs of stereo images from 2001-Middlebury image datasets [52], namely *Venus* and *Bull*, *Sawtooth*, and *Barn*. We could not use the latest stereo datasets because of their high disparity range. Given a pair of cropped regions from recent Middlebury stereo datasets (see Fig. 4), the majority of regions in both cropped stereo images would be occluded due to a large disparity range, resulting in only a small portion of the scene being visible in both images. This makes them not suitable to evaluate our quantum model due to the simplicity of the defined global Stereo Matching energy function in (3).

To identify a more common region of interest in both stereo images, we selected our pairs of stereo images from the 2001 Middlebury image dataset [52], as well as two "natural" images (*Tree* and *Castle*) from the real world with low disparity ranges to incorporate complex scene structures into our prepared dataset. Fig. 5 illustrates our prepared stereo dataset with the corresponding "ground truths". We did not have ground truths for the natural images since they were not created in controlled laboratory settings like 2001-Middlebury image datasets [52]. Therefore, we incorporated a deep-learning-based model [54] to get fairly accurate disparity maps to be used as the corresponding ground truths. Note that we used the gray-scale versions of the shown stereo images.

We also selected the best-performing and state-of-the-art classical minimization algorithms commonly used in CV, so that we can

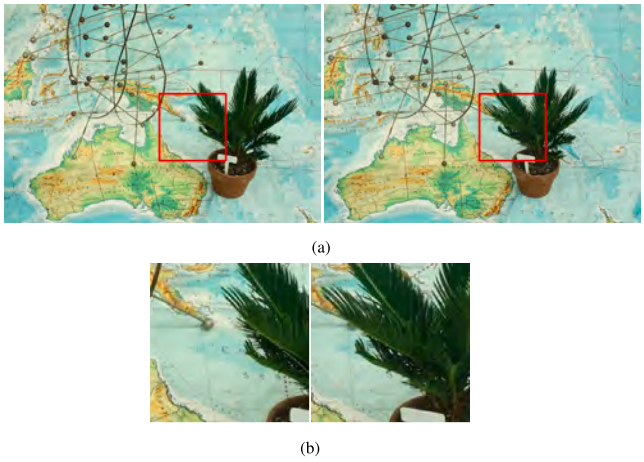


Fig. 4. A pair of stereo images from 2014-Middlebury image datasets: (a) the quarter-resolution version of the *Australia* stereo dataset from 2014-Middlebury image datasets [53], and (b) the cropped regions of 150 × 150 pixels from the left and right stereo images, respectively. Most regions in both cropped stereo images are occluded due to a high disparity range, making them unsuitable for evaluating our minimization models with the classical counterparts.

compare our quantum model with the classical counterparts. The selected classical algorithms include two move-making methods and two message-passing methods.

- Move-making methods break down the minimization problem into sub-problems over subsets of the label space. A move is defined as a change from one labeling solution to another. These methods typically start with a random labeling solution and then iteratively improve the initial labeling solution until no improvement can be achieved. We selected two well-performing move-making methods, Swap move [8] and Expansion move [8].
  - Swap move algorithm begins with an arbitrary labeling solution, where each pixel has a label from a set of possible disparity values. Given a pair of disparities $\alpha, \beta \in \mathbb{N}$, a move is called a swap move if it takes some subsets of pixels currently given the disparity α and allocates them to the disparity β (and vice versa). The swap move algorithm terminates at a local minimum where no swap move can produce a lower energy labeling solution. For each pair of disparities, minimization is performed by solving the minimum-cut problem on a specially structured graph.
  - Likewise, the Expansion move algorithm starts with an arbitrary labeling. Given a disparity $\alpha \in \mathbb{N}$, a move is called an expansion move if it extends the set of pixels with the disparity α. This algorithm finds a local minimum such that there is no expansion move for the disparity α, by which a labeling solution with lower energy can be produced. Similar to the Swap move, minimization is performed based on finding the minimum cut over a specially structured graph.
- Message-passing methods are based on graphical models where messages are defined on the edges of the graph and updated based on other messages. These messages effectively re-parameterize the model, enabling local optimizations to converge towards a globally consistent solution. We selected two well-performing message-passing methods, Max product Loopy Belief Propagation (LBP) [13], and Improved Tree Re-weighted Message Passing (TRW-S) [55].
  - Belief Propagation (BP) was initially proposed to solve the inference problem in a probability distribution, which can be defined over different graphical models (such as MRF, Bayesian networks, or factor graphs) without cycles. BP equations do not depend on the way that a graphical model is structured; therefore, nothing can stop us from defining it for graphical models with cycles. Standard LBP methods can be categorized into two types based on their message update rules: sum–product and max-product LBP. The max-product LBP is more prevalent in Stereo Matching problems because it can be transformed into a min-sum problem in negative-log space, consistent with most Stereo Matching energy functions.
  - An alternative message-passing approach is TRW, which is comparable to LBP techniques. The TRW algorithm requires a coefficient calculated based on a set of trees from a neighborhood graph of each pixel. When this coefficient is equal to one, the TRW method becomes equivalent to the standard LBP. The TRW method uniquely provides a lower bound estimate for the energy function. This lower bound can be utilized to assess the proximity of the energy of an optimal solution. Similar to LBP techniques, TRW can be utilized for any CV/Stereo-Matching energy function. The initial TRW method underwent significant improvement [55] and now is referred to as TRW-S.

We utilized the Middlebury software framework for our classical implementations [56]. Considering the global Stereo Matching energy function defined in (3), we established an initial $\lambda = 20$ for all the benchmark minimization algorithms to ensure a fair comparison. Fig. 6 shows the computed disparity maps by the benchmark minimization algorithms. Next, we define two widely-used metrics [57] to evaluate the accuracy of our results in comparison to the corresponding ground truths: the root-mean-squared error (*rms*) and the percentage of mismatched/bad pixels (*bad-β*). Given a disparity map D and a ground truth $\mathcal{T}$ defined by $n \times m$ matrices of integers, *rms* and *bad-β* are defined as (11).

$$rms = \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(D_{i,j} - \mathcal{T}_{i,j}\right)^2}, \tag{11}$$

$$bad\text{-}\beta = \left(\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(|D_{i,j} - \mathcal{T}_{i,j}| > \beta\right)\right) \times 100 \tag{12}$$

where $\beta \in \mathbb{R}^+$ is the disparity error tolerance. In the following evaluation, we set β to 0.5 and 1.0, named *bad-0.5* an *bad-1.0*, respectively. Table 2 compares the performance of each minimization algorithm based on the defined metrics. Given the cropped stereo images, the results suggest that the quantum model outperformed the classical counterparts on the *Bull*, *Sawtooth*. and *Tree* datasets, and performed competitively on the *Venus*, *Barn*, and *Castle* datasets. Our findings show that QA can offer promising results in CV applications compared to the state-of-the-art CV minimization inference algorithms. Due to the scarcity of available qubits on the current D-Wave QPUs, we were not able to use a pure QA minimization, and we used a D-Wave hybrid solver, which offers a reliable estimate of the future accuracy of D-Wave QPUs once more qubits become available on the hardware. Since our model (5) is a direct equivalent to the global Stereo Matching energy function (3), its energy solution can be compared with that of the iterative classical minimization algorithms. Fig. 7 shows the energies of the solutions obtained by each minimization model over the provided stereo datasets. According to the findings, our approach demonstrated a capacity to obtain solutions of lower energy in comparison to the iterative classical minimization methods for each provided stereo dataset. This observation underscores the effectiveness of QUBOs when solved by D-Wave hybrid solvers. We do not provide a comparison in terms of running time, as the classical iterative minimization algorithms were significantly faster than the D-Wave hybrid solver when minimizing the corresponding QUBO models. The reason is because of the way that a

Table 2
Numerical evaluation for the prepared stereo dataset.

Dataset	Method	*rms*	*bad-0.5 (%)*	*bad-1.0* (%)
Venus	Ours	2.25	40.44	10.45
	Swap	2.09	47.57	10.23
	Expansion	1.94	43.81	9.76
	BP-M	1.96	47.09	9.30
	TRW-S	1.92	44.07	9.54
Bull	Ours	2.33	36.07	7.08
	Swap	2.38	37.43	7.40
	Expansion	2.38	37.28	7.36
	BP-M	2.39	37.42	7.25
	TRW-S	2.38	37.06	7.32
Sawtooth	Ours	2.27	22.54	10.26
	Swap	2.44	22.76	10.27
	Expansion	2.41	22.85	10.44
	BP-M	2.41	23.59	10.30
	TRW-S	2.36	22.67	10.36
Barn	Ours	2.27	14.37	7.41
	Swap	2.23	16.11	7.51
	Expansion	2.21	16.09	7.55
	BP-M	2.38	20.21	8.25
	TRW-S	2.23	15.54	7.33
Tree	Ours	2.99	24.99	13.22
	Swap	3.32	33.85	14.39
	Expansion	3.27	31.73	13.23
	BP-M	3.16	33.45	13.91
	TRW-S	3.12	31.81	13.22
Castle	Ours	2.74	34.52	17.62
	Swap	2.83	32.85	17.25
	Expansion	2.76	33.68	16.94
	BP-M	2.99	41.12	21.21
	TRW-S	2.66	33.36	16.62

D-Wave hybrid solver works. A D-Wave hybrid solver is based on the D-Wave Hybrid Solver Service (HSS). Once a QUBO is provided to the HSS, it activates one or more heuristic solvers that run in parallel, either on a CPU or a GPU platform, to identify high-quality solutions. Each heuristic solver comprises a classical heuristic module that navigates the search space, and a quantum module is responsible for formulating quantum queries directed to the D-Wave Advantage QPU. Solutions retrieved from the QPU assist the heuristic modules in pinpointing more viable search space regions or refining the current solutions. Each heuristic solver forwards its top solution to the HSS solver. The HSS solver then determines the best solution from the collective outputs of the heuristic solvers and relays this optimal solution back to the user [58]. Therefore, the running time is not derived from a direct QPU minimization to be compared with the classical minimization methods. We utilized a D-Wave hybrid solver due to the limited availability of qubits on the hardware. If we were to directly employ the actual QPU, we would be constrained to testing the minimization process on very small patches. Stereo Matching conducted on small patches would not encompass enough complex disparity estimation scenarios, such as occluded regions, to effectively compare against classical approaches. Moreover, classical methods typically perform well on small patches, with their drawbacks becoming apparent as input images grow larger, resulting in a larger search space within the underlying MRF graph structure (the defined 4-neighborhood system). Hence, we opted to work with input images of an acceptable size, ensuring they encompassed enough complexities representative of various stereo-vision scenarios. In addition to mathematically demonstrating the correctness of the QUBO model, we employed the hybrid D-Wave solver to showcase a proof of concept that the QUBO model is viable in practice. However, the actual minimization process on the QPU should be evaluated once sufficient qubits are available on the QPU. Our primary objective in utilizing the D-Wave hybrid solver was to demonstrate that minimizing the QUBO model yields results consistent with minimizing the original objective function, as expected.

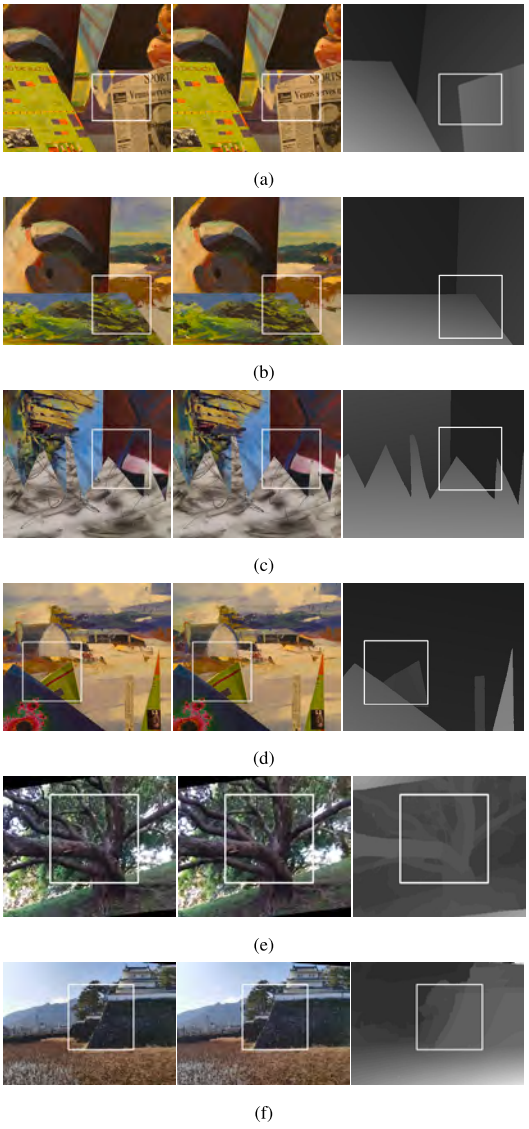


Fig. 5. The prepared stereo datasets, (a) *Venus*, (b) *Bull*, (c) *Sawtooth*, (d) *Barn*, (c) *Tree*, and (d) *Castle*. In each row of images, we have the left stereo image, the right stereo image, and the corresponding ground truth for the left stereo image. The white squares show the cropped regions for our experiments.

6. Generalization

Preceding the study's conclusion, we show the general applicability of the proposed quantum model for a different CV minimization problem, Image Restoration. Image Restoration is a family of inverse problems to recover an original high-quality image from a corrupted input image. There are some reasons that corruption may occur, such as the image capture process (e.g., noise, lens blur), post-processing (e.g., JPEG compression), or photography in non-ideal conditions (e.g., haze, motion blur). Image Restoration can be modeled by a labeling problem where a set of pixels is labeled by some quantities. In Image Restoration, the main goal is to recover the image's original pixel intensities as much as possible. Therefore, the set of labels should contain the actual intensities (as opposed to Stereo Matching, where the set of labels is considered as disparity values). In the following, we first model the Image Restoration as a discrete minimization problem, and then we adapt our quantum model to it. In the following, we initially model Image Restoration as a discrete minimization problem and then integrate our quantum model. In the most general form, a digital image is a function $I : P \rightarrow H$ where is the set of two-dimensional spatial

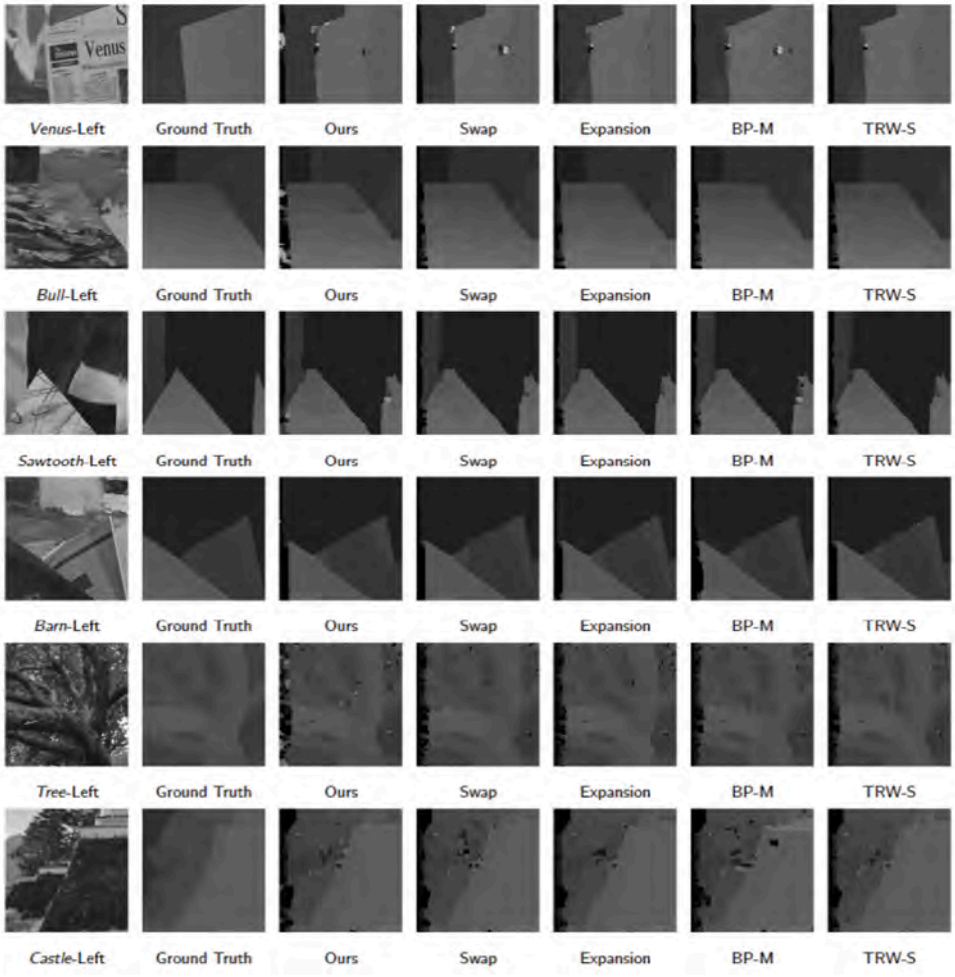


Fig. 6. Computed disparity maps by the benchmark minimization algorithms.

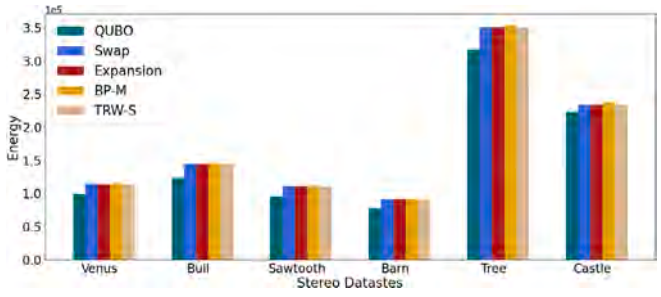


Fig. 7. A comparison between energies obtained by our model and the benchmarking classical minimization.

coordinates as defined in (1) and $H = \{0, \dots, h-1\}$ is a set of signal values. The coordinate $(i, j) \in P$ is referred to as a pixel, and $I(i, j)$ is called the intensity of the image at pixel location (i, j). Let N as defined in (2) be a 4–neighborhood system on a regular lattice by which each pixel in P has at most four neighboring pixels. We define $L = \{0, \dots, h-1\}$ as the set of labels. The main goal here is to label each pixel in P with a value in L. Let $\mathbf{w}$ be a vector of integer variables such that $\mathbf{w} = (w_{i,j})_{(i,j) \in P}$ where $w_{i,j} \in L$. Given I as the input noisy image, Image Restoration can be represented by the energy function $F' : L^{n \times m} \to \mathbb{R}$ as follows [8].

$$F'(\mathbf{w}) = \sum_{(i,j) \in P} (I(i,j) - w_{i,j})^2 + \lambda \sum_{\{(i,j),(i',j')\} \in N} \delta(w_{i,j}, w_{i',j'}), \tag{13}$$

$$\delta(w_{i,j}, w_{i',j'}) = \begin{cases} 0, & \text{if } w_{i,j} = w_{i',j'}; \\ 1, & \text{otherwise,} \end{cases}$$

where λ is a positive integer. The first term, defined as the Sum of Squared Differences (SSD), is to compute the cost of allocating a label $w_{i,j} \in L$ to a pixel $(i, j) \in P$. When $(I(i, j) - w_{i,j})^2$ is zero, the intensity $I(i, j)$ at pixel (i, j) is more likely to be $w_{i,j}$. The second term encodes a preference for the labels of the neighboring pixels, ensuring that the intensities of a neighborhood of pixels present some coherence and generally do not change abruptly. Given (i, j) and (i', j') as two neighboring pixels, $\delta(w_{i,j}, w_{i',j'})$ penalizes the solution if the allocated labels $w_{i,j}$ and $w_{i',j'}$ are different.

Considering the Image Restoration objective function in (13) and the Stereo Matching objective function in (3), they are both discrete energy functions derived from MRFs, and therefore, we can use a similar approach to represent the corresponding QUBO model to (13) for QA purposes. We first allocate $|L|$ binary variables to each pixel $(i, j) \in P$, where $|L|$ is the number of elements in L, $0 \le i \le n-1$, and $0 \le j \le m-1$ for an $n \times m$ input image. For such an allocation, we define $\mathbf{x} \in \{0, 1\}^{nm|L|}$ as a vector of $nm|L|$ binary variables such that $\mathbf{x} = (x_{i,j,l})$ for all $(i, j) \in P$ and $l \in L$. Let our QUBO model be defined as (14).

$$\begin{aligned} H'(\mathbf{x}) = \alpha \sum_{(i,j) \in P} \left(1 - \sum_{l \in L} x_{i,j,l}\right)^2 + \sum_{(i,j) \in P} \sum_{l \in L} (I(i,j) - l)^2 x_{i,j,l} \\ + \lambda \sum_{\{(i,j),(i',j')\} \in N} \sum_{l_1 \in L} \sum_{l_2 \in L} \delta(l_1, l_2) x_{i,j,l_1} x_{i',j',l_2}, \end{aligned} \tag{14}$$

where $\alpha > \left(\sum_{(i,j) \in P} \max\{(I(i,j) - l)^2 \mid l \in L\}\right) + \lambda |N|$, and $|N|$ is the number of elements in N. We set $\mathbf{x}^* = \arg\min_{\mathbf{x}} H'(\mathbf{x})$ and define a

vector of nm integer values as $\mathbf{w}^* = (w^*_{i,j})_{(i,j) \in P}$, where $w^*_{i,j} = l$ if $x^*_{i,j,l} = 1$. Then, $\mathbf{w}^*$ minimizes the Image Restoration energy function defined in (13). The proof of correctness follows the same approach discussed in Section 4. We refer interested readers to Heidari et al. solution [51] to Image Restoration on small image patches, which uses the same idea, but minimizes the same energy function based on a multi-way graph cut approach.

7. Conclusion

CV labeling algorithms play a pivotal role in the domain of low-level vision. For decades, it has been known that these problems can be elegantly formulated as discrete energy-minimization problems derived from probabilistic graphical models. Despite recent advances in inference algorithms, the resulting energy-minimization problems are generally viewed as intractable. In this study, we presented a QA-based method for solving CV discrete optimization problems, specifically for Stereo Matching. However, our proposed quantum model is not limited to Stereo Matching and can be applied to various CV labeling problems such as Image Segmentation, Image Restoration, Image Registration, Optical Flow, Object Detection, and Image Inpainting. We provided proof of correctness to demonstrate the equivalence of the proposed quantum model to the original discrete minimization energy function. Due to the limited availability of qubits on the quantum hardware, we were not able to minimize the Stereo Matching energy function directly on the QPU. Instead, we utilized a D-Wave hybrid solver to show the feasibility of our proposed quantum model. Our results showed promising solutions with lower energies compared to the best classical minimization algorithms in the literature. When there are enough qubits available, it may be a subject for future research to determine if a quantum-based CV inference offers any advantages over classical minimization methods in terms of accuracy and speed.

CRediT authorship contribution statement

Shahrokh Heidari: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michael J. Dinneen:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Patrice Delmas:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shahrokh Heidari reports article publishing charges was provided by University of Auckland. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank Prof. Cristian Calude for his helpful discussions on this paper.

References

[1] O. Veksler, Efficient Graph-Based Energy Minimization Methods in Computer Vision, Cornell University, 1999.
[2] P.F. Felzenszwalb, R. Zabih, Dynamic programming and graph algorithms in computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 33 (4) (2010) 721–740.
[3] V. Černỳ, Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, J. Optim. Theory Appl. 45 (1) (1985) 41–51.
[4] D. Geiger, F. Girosi, Parallel and deterministic algorithms from MRFs: Surface reconstruction, IEEE Trans. Pattern Anal. Mach. Intell. 13 (05) (1991) 401–412.
[5] J. Besag, On the statistical analysis of dirty pictures, J. R. Stat. Soc. Ser. B Stat. Methodol. 48 (3) (1986) 259–279.
[6] Y. Boykov, O. Veksler, R. Zabih, Markov random fields with efficient approximations, in: Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), IEEE, 1998, pp. 648–655.
[7] S. Birchfield, C. Tomasi, Multiway cut for stereo and motion with slanted surfaces, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 1, IEEE, 1999, pp. 489–495.
[8] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.
[9] N. Komodakis, G. Tziritas, Approximate labeling via graph cuts based on linear programming, IEEE Trans. Pattern Anal. Mach. Intell. 29 (8) (2007) 1436–1453.
[10] M. Wainwright, T. Jaakkola, A. Willsky, Tree consistency and bounds on the performance of the max-product algorithm and its generalizations, Stat. Comput. 14 (2) (2004) 143–166.
[11] M.J. Wainwright, T.S. Jaakkola, A.S. Willsky, MAP estimation via agreement on trees: message-passing and linear programming, IEEE Trans. Inf. Theory 51 (11) (2005) 3697–3717.
[12] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (7) (2003) 787–800.
[13] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient belief propagation for early vision, Int. J. Comput. Vis. 70 (1) (2006) 41–54.
[14] T. Yu, R.-S. Lin, B. Super, B. Tang, Efficient message representations for belief propagation, in: 2007 IEEE 11th International Conference on Computer Vision, IEEE, 2007, pp. 1–8.
[15] M.F. Tappen, W.T. Freeman, Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters, in: Computer Vision, IEEE International Conference on, Vol. 3, IEEE Computer Society, 2003, p. 900.
[16] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for markov random fields, in: European Conference on Computer Vision, Springer, 2006, pp. 16–29.
[17] V. Kolmogorov, C. Rother, Comparison of energy minimization algorithms for highly connected graphs, in: European Conference on Computer Vision, Springer, 2006, pp. 1–15.
[18] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for markov random fields with smoothness-based priors, IEEE Trans. Pattern Anal. Mach. Intell. 30 (6) (2008) 1068–1080.
[19] J.H. Kappes, B. Andres, F.A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B.X. Kausler, T. Kröger, J. Lellmann, et al., A comparative study of modern inference techniques for structured discrete energy minimization problems, Int. J. Comput. Vis. 115 (2) (2015) 155–184.
[20] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, et al., Deep learning for computer vision: A brief review, Comput. Intell. Neurosci. 2018 (2018).
[21] V.S. Denchev, S. Boixo, S.V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, H. Neven, What is the computational value of finite-range tunneling? Phys. Rev. X 6 (3) (2016) 031015.
[22] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A.D. King, M.M. Nevisi, J.P. Hilton, C.C. McGeoch, Quantum annealing amid local ruggedness and global frustration, J. Phys. Soc. Japan 88 (6) (2019) 061007.
[23] R. Yaacoby, N. Schaar, L. Kellerhals, O. Raz, D. Hermelin, R. Pugatch, A comparison between D-Wave and a classical approximation algorithm and a heuristic for computing the ground state of an Ising spin glass, 2021, arXiv preprint arXiv:2105.00537.
[24] C.C. McGeoch, Adiabatic quantum computation and quantum annealing: Theory and practice, Synth. Lect. Quant. Comput. 5 (2) (2014) 1–93.
[25] A. Lucas, Ising formulations of many NP problems, Front. Phys. 2 (2014) 5.
[26] C.S. Calude, M.J. Dinneen, Solving the broadcast time problem using a D-Wave quantum computer, in: Advances in Unconventional Computing, Springer, 2017, pp. 439–453.
[27] C.S. Calude, M.J. Dinneen, R. Hua, QUBO formulations for the graph isomorphism problem and related problems, Theoret. Comput. Sci. 701 (2017) 54–69.
[28] S.H. Adachi, M.P. Henderson, Application of quantum annealing to training of deep neural networks, 2015, arXiv preprint arXiv:1510.06356.
[29] V. Dixit, R. Selvarajan, M.A. Alam, T.S. Humble, S. Kais, Training and classification using a restricted Boltzmann machine on the D-wave 2000Q, 2020, arXiv:2005.03247, [cs, stat], URL http://arxiv.org/abs/2005.03247.

[30] Y. Koshka, D. Perera, S. Hall, M. Novotny, Empirical investigation of the low temperature energy function of the restricted Boltzmann machine using a 1000 qubit D-wave 2X, in: 2016 International Joint Conference on Neural Networks (IJCNN), (ISSN: 2161-4407) 2016, pp. 1948–1954.
[31] Y. Koshka, D. Perera, S. Hall, M.A. Novotny, Determination of the lowest-energy states for the model distribution of trained restricted Boltzmann machines using a 1000 qubit D-wave 2X quantum computer, Neural Comput. 29 (7) (2017) 1815–1837.
[32] Y. Koshka, M.A. Novotny, 2000 Qubit D-wave quantum computer replacing MCMC for RBM image reconstruction and classification, in: 2018 International Joint Conference on Neural Networks (IJCNN), (ISSN: 2161-4407) 2018, pp. 1–8.
[33] T. Birdal, V. Golyanik, C. Theobalt, L.J. Guibas, Quantum permutation synchronization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13122–13133.
[34] F. Arrigoni, W. Menapace, M.S. Benkner, E. Ricci, V. Golyanik, Quantum motion segmentation, in: European Conference on Computer Vision, Springer, 2022, pp. 506–523.
[35] J.-N. Zaech, A. Liniger, M. Danelljan, D. Dai, L. Van Gool, Adiabatic quantum computing for multi object tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8811–8822.
[36] A.-D. Doan, M. Sasdelli, D. Suter, T.-J. Chin, A hybrid quantum-classical algorithm for robust fitting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 417–427.
[37] M. Farina, L. Magri, W. Menapace, E. Ricci, V. Golyanik, F. Arrigoni, Quantum multi-model fitting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 13640–13649.
[38] J. Li, S. Ghosh, Quantum-soft QUBO suppression for accurate object detection, in: European Conference on Computer Vision, Springer, 2020, pp. 158–173.
[39] W. Cruz-Santos, S.E. Venegas-Andraca, M. Lanzagorta, A QUBO formulation of the stereo matching problem for D-Wave quantum annealers, Entropy 20 (10) (2018) 786.
[40] S. Heidari, M. Rogers, P. Delmas, An improved quantum solution for the stereo matching problem, in: 2021 36th International Conference on Image and Vision Computing New Zealand, IVCNZ, IEEE, 2021, pp. 1–6.
[41] M.S. Benkner, V. Golyanik, C. Theobalt, M. Moeller, Adiabatic quantum graph matching with permutation matrix constraints, in: 2020 International Conference on 3D Vision (3DV), IEEE, 2020, pp. 583–592.
[42] A. Yurtsever, T. Birdal, V. Golyanik, Q-FW: A hybrid classical-quantum frank-wolfe for quadratic binary optimization, in: European Conference on Computer Vision, Springer, 2022, pp. 352–369.
[43] M.S. Benkner, M. Krahn, E. Tretschk, Z. Lähner, M. Moeller, V. Golyanik, Quant: Quantum annealing with learnt couplings, 2022, arXiv preprint arXiv: 2210.08114.
[44] R.A. Hamzah, H. Ibrahim, Literature survey on stereo vision disparity map algorithms, J. Sens. 2016 (2016).
[45] A.J. Malekabadi, M. Khojastehpour, B. Emadi, Disparity map computation of tree using stereo vision system and effects of canopy shapes and foliage density, Comput. Electron. Agric. 156 (2019) 627–644.
[46] F. Remondino, C. Fraser, Digital camera calibration methods: considerations and comparisons, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 36 (5) (2006) 266–272.
[47] C. Wang, N. Komodakis, N. Paragios, Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey, Comput. Vis. Image Underst. 117 (11) (2013) 1610–1627.
[48] S.Z. Li, Markov Random Field Modeling in Computer Vision, Springer-Verlag, Berlin, Heidelberg, 1995.
[49] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, Quantum computation by adiabatic evolution, 2000, arXiv preprint quant-ph/0001106.
[50] D-Wave Systems, Discrete quadratic models, 2023, [Online]. Available from: https://docs.ocean.dwavesys.com/en/latest/concepts/dqm.html.
[51] S. Heidari, M.J. Dinneen, P. Delmas, An equivalent QUBO Model to the minimum multi-way cut problem, Tech. rep., Department of Computer Science, The University of Auckland, New Zealand, 2022.
[52] Middlebury Stereo Vision, 2001 Middleburry stereo datasets, 2022, [Online]. Available from: https://vision.middlebury.edu/stereo/data/scenes2001/.
[53] 2014 Middleburry stereo datasets, 2022, [Online]. Available from: https://vision.middlebury.edu/stereo/data/scenes2014/.
[54] J. Li, P. Wang, P. Xiong, T. Cai, Z. Yan, L. Yang, J. Liu, H. Fan, S. Liu, Practical stereo matching via cascaded recurrent network with adaptive correlation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16263–16272.
[55] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, in: International Workshop on Artificial Intelligence and Statistics, PMLR, 2005, pp. 182–189.
[56] Middlebury Stereo Vision, Middleburry MRF implementations, 2022, [Online]. Available from: https://vision.middlebury.edu/MRF/.
[57] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. 47 (1) (2002) 7–42.
[58] D-Wave Systems, D-Wave hybrid solver service + advantage: Technology update, 2022, [Online]. Available from: https://www.dwavesys.com/media/m2xbmlhs/14-1048a-a_d-wave_hybrid_solver_service_plus_advantage_technology_update.pdf.

Shahrokh Heidari is a Ph.D. candidate in the Computer Science department at the University of Auckland. He is studying the potential applications of quantum annealing for computer vision applications, such as stereo matching and image restoration. He has been research assistant in the Intelligent Vision Systems Lab (IVSLab) at the University of Auckland, working on several projects including 3D reconstruction and modeling from stereo vision systems, characterizing sediment topography and seagrass detection for Marine applications, and medical image segmentation. His research interests lie in stereo vision and 3D reconstruction, quantum annealing computations, machine learning, and neural networks.

Michael J. Dinneen received his Ph.D. from the University of Victoria (B.C. Canada) in 1996 and is a currently a senior lecturer at the University of Auckland. Prior to that he worked for several years at the Los Alamos National Laboratory (New Mexico, USA) working on grand-challenge combinatorial search and optimization problems using supercomputers. Michael's specialty area is of graph algorithms, network design and graph minors. He does research on unconventional models of computation such as (adiabatic) quantum computing and membrane computing, culminating in over 100 research papers.

Patrice Delmas was born, raised and educated in France (M.Eng., M.Sc. and Ph.D. degrees, in 1994, 1995 and 2000, respectively, at the National Polytechnic Institute, Grenoble, France). He has been with the Department of Computer Science at the University of Auckland since 2001 and currently holds an associate Professor position. Patrice is the founder and director of the Intelligent Vision Systems Lab - IVSLab, Department of Computer Science, The University of Auckland. Patrice has 25 years research experience in theoretical and practical computer vision with over 200 refereed publications on theoretical and applied computer vision.