TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Çağlar Aytekin

Quantum Cuts

A Quantum Mechanical Spectral Graph Partitioning Method for Salient Object Detection

Julkaisu 1440 • Publication 1440

Tampere 2016

Tampereen teknillinen yliopisto. Julkaisu 1440
Tampere University of Technology. Publication 1440

Çağlar Aytekin

Quantum Cuts

A Quantum Mechanical Spectral Graph Partitioning Method for Salient Object Detection

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Konetalo Building, Auditorium K1702, at Tampere University of Technology, on the 24th of November 2016, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2016

ISBN 978-952-15-3859-9 (printed)
ISBN 978-952-15-3872-8 (PDF)
ISSN 1459-2045

Abstract

The increasing number of cameras, their availability to the end user and the social media platforms gave rise to the massive repositories of today's Big Data. The largest portion of this data corresponds to unstructured image and video collections. This fact motivates the development of algorithms that would help efficient management and organization of the Big Data. This processing usually involves high level Computer Vision tasks such as object detection and recognition whose accuracy and complexity are therefore crucial. Salient object detection, which can be defined as highlighting the regions that visually stand out from the rest of the environment, can both reduce the complexity and improve the accuracy of object detection and recognition. Thus, recently there has been a growing interest in this topic. This interest is also due to many other applications of salient object detection such as media compression and summarization.

This thesis focuses on this crucial problem and presents novel approaches and methods for salient object detection in digital media, using the principles of Quantum Mechanics. The contributions of this thesis can be categorized chronologically into three parts. First part is constituted of a direct application of ideas originally proposed for describing the wave nature of particles in Quantum Mechanics and expressed through Schrödinger's Equation, to salient object detection in images. The significance of this contribution is the fact that, to the best of our knowledge, this is the first study that proposes a realizable quantum mechanical system for salient object proposals yielding an instantaneous speed in a possible physical implementation in the quantum scale.

The second and main contribution of this thesis, is a spectral graph based salient object detection method, namely Quantum-Cuts. Despite the success of spectral graph based methods in many Computer Vision tasks, traditional approaches on applications of spectral graph partitioning methods offer little for the salient object detection problem which can be mapped as a foreground segmentation problem using graphs. Thus, Quantum-Cuts adopts a novel approach to spectral graph partitioning by integrating quantum mechanical concepts to Spectral Graph Theory. In particular, the probabilistic interpretation of quantum mechanical wave-functions and the unary potential fields in Quantum Mechanics when combined with the pairwise graph affinities that are widely used in Spectral Graph Theory, results into a unique optimization problem that formulates salient object detection. The optimal solution of a relaxed version of this problem is obtained via Quantum-Cuts and is proven to efficiently represent salient object regions in images.

The third part of the contributions cover improvements on Quantum-Cuts by analyzing the main factors that affect its performance in salient object detection. Particularly, both unsupervised and supervised approaches are adopted in improving the exploited graph representation. The extensions on Quantum-Cuts led to computationally efficient algorithms that perform superior to the state-of-the-art in salient object detection.

Preface

The work presented in this thesis has been carried out at Tampere University of Technology (TUT), Finland and Middle East Technical University (METU), Turkey during the years 2011-2016.

First and foremost, I would like to express my deepest gratitude to Professor Moncef Gabbouj for his endless support and help in every possible issue that I faced during all these years. Second, I would like to thank Professor Serkan Kiranyaz for his support, as well as his technical supervision for the many studies that we have conducted. I would also like to thank Dr. Alexandros Iosifidis for his friendship and his technical supervision.

I am thankful to other co-authors of the studies presented in this study, especially my friend Ezgi Can Ozan. I would also like to thank all members of the Multimedia Research Group for their friendship.

Last, but definitely not least, I would like to express my deepest gratitude, esteem and love for my father, Professor Hüseyin Aytekin, and my mother, Hatice Aytekin. I would like to express here that this study would have never been possible without their endless support, love and belief in me.

Tampere 1.6.2016

Çağlar Aytekin

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

E: Energy

h: Planck's Constant

$\hbar$: Reduced Planck's Constant

ν: Frequency

λ: Wavelength

p: Momentum

k: Wavenumber

r: Radius

x: Location

$\hat{A}$: Quantum mechanical operator

A: Quantum mechanical observable

a: Quantum mechanical observable value

V: Potential field

m: Particle mass

$\hat{H}$: Hamiltonian, Quantum Mechanical Energy Operator

ψ: Wavefunction

$\mathcal{V}$: Nodes of a graph

$\mathcal{E}$: Edges of a graph

$\mathcal{G}$: Graph

A: Adjacency Matrix

W: Weighted Adjacency Matrix

D: Degree Matrix

L: Laplacian Matrix

$\overline{\mathrm{P}}$: Complement of P

φ: Eigenvector

γ: Eigenvalue

y^{*}: Conjugate transpose of y

y^{T}: Transpose of y

μ_{y}: Mean value of y

$\det(.)$: Determinant

$\mathrm{R}(.,.)$: Rayleigh Quotient

$\mathrm{R}_{\mathrm{g}}(.,.,.)$: Generalized Rayleigh Quotient

$\mathrm{K}(.,.)$: Kernel

$\mathrm{G}(.,.)$: Gaussian Kernel

PRE: Precision

REC: Recall

FPR: False Positive Rate

TPR: True Positive Rate

MSE: Mean Squared Error

MAE: Mean Absolute Error

RCUT: Ratio Cut

NCUT: Normalized Cut

QCUT: Quantum Cut

EQCut: Extended Quantum Cut

CKN: Convolutional Kernel Network

CNN: Convolutional Neural Network

UCKN: Unsupervised Convolutional Kernel Network

SCKN: Supervised Convolutional Kernel Network

List of Publications

[P1] C. Aytekin, S. Kiranyaz and M. Gabbouj, “Quantum Mechanics in Computer Vision: Automatic Object Extraction,” IEEE International Conference on Image Processing (ICIP), pp. 2489-2493, September 2013.

[P2] C. Aytekin, S. Kiranyaz and M. Gabbouj, “Automatic Object Segmentation by Quantum-Cuts,” International Conference on Pattern Recognition (ICPR), pp. 112-117, August 2014.

[P3] C. Aytekin, S. Kiranyaz and M. Gabbouj, “Learning to Rank Salient Segments Extracted by Multispectral Quantum-Cuts,” Pattern Recognition Letters, vol. 72, pp. 91-99, March 2016.

[P4] C. Aytekin, E. C. Ozan, S. Kiranyaz and M. Gabbouj, “Extended Quantum-Cuts for Unsupervised Salient Object Extraction,” Multimedia Tools and Applications, D.O.I.: 10.1007/s11042-016-3431-1.

[P5] C. Aytekin, A. Iosifidis, S. Kiranyaz and M. Gabbouj, “Salient Object Segmentation based on Linearly Combined Affinity Graphs,”, Accepted, International Conference on Pattern Recognition (ICPR), 2016.

[P6] C. Aytekin, A. Iosifidis, S. Kiranyaz and M. Gabbouj, “Learning Graph Affinities for Spectral-based Salient Object Detection,” Accepted, Pattern Recognition.

1 Introduction

The wide spread usage of cameras, a vast variety of mobile technologies, numerous social networks, along with the emerging context-based services have exponentially increased today's massive media collections and started the era of Big Data. Computer Vision constitutes an important field of research due to its ability to efficiently address the challenges and problems that arise for the organization and management of such massive media collections. One of the most important research domain lies in salient object detection, which provides regions of utter interest from the media. This process is important in terms of handling Big Data, as it both lowers the complexity and improves the accuracy of the subsequent higher level and usually computationally heavier tasks such as object recognition, by reducing the required search space to the regions that visually stand out. The fact that a similar task is performed prior to higher level ones as a part of the attention mechanisms in the human brain, further motivates the research in salient object detection. Besides assisting object detection and recognition, salient object detection also offers direct solutions to various problems such as auto-focusing, advertisement assessment, image summarization and compression, video surveillance, etc.

The computational complexity of salient object detection is required to be much less than that of the object detection and recognition, considering that one purpose of using salient object detection is to reduce the complexity of these subsequent tasks. Besides, its performance is crucial as it is required to provide accurate regions. Therefore, salient object detection needs to be formulated in such a way that will provide both computational efficiency and highly accurate regions of interest. In this thesis, a graph-based formulation is adopted, motivated by the success of such methods in many Computer Vision tasks. Especially, the success and efficiency of spectral graph based methods in similar tasks such as image segmentation, had led us to investigate the effectiveness of these methods in salient object detection. However, the spectral graph based methods that are widely used in Computer Vision, do not offer a direct solution to salient object detection, since they are mostly devised for problems involving partitioning the media data into many clusters that are equally important. However, salient object detection requires partitioning the media into two, where the important partition, the foreground, is the region that contains salient object or objects. There-

fore, a new spectral graph based method is needed in order to efficiently formulate and offer solutions for this problem.

In this thesis, such a spectral graph based method is proposed as a result of investigating the link between spectral graph theory and principles of Quantum Mechanics. First, a direct application of Quantum Mechanics to salient object detection is utilized via solving the Schrödinger's Equation with supplying raw image data as input. The solutions have proven both experimentally and theoretically, to form a salient object proposal pool of high accuracy. Next, with this motivation, an investigation of the Schrödinger's Equation is conducted in a spectral graph theory point of view. This analysis revealed that the ground state solution of the Schrödinger's Equation, exploiting a quantum mechanical operator that is obtained by integrating concepts from spectral graph theory and quantum mechanics, optimizes a problem that efficiently represents salient object detection. The improvements and modifications of the method devised by following this approach, i.e. Quantum-Cuts, led to state-of-the-art performances for salient object detection.

1.1 Objectives and Outline of the Thesis

In this thesis, the focus is on analyzing the exploitation of quantum mechanical principles in salient object detection problem. Accordingly, the objectives of the thesis can be listed as follows:

- To investigate a direct application of Quantum Mechanics to Salient Object Detection and to propose a realizable quantum mechanical system that would produce salient object proposals at an instantaneous speed.
- To reveal links between Quantum Mechanics and Spectral Graph Theory that would lead to an effective and efficient salient object detection method.
- To come up with salient object detection methods that perform superior to state-of-the-art, by applying several extensions and modifications on the proposed method.

The outline of the thesis is as follows. In Chapter 2, an overview of the major fields that are subjects of interest to this thesis is provided. This background section includes a brief overview of Quantum Mechanics in Section 2.1, Spectral Graph Theory in Section 2.2, graph affinity learning in Section 2.3 and salient object detection in Section 2.4. Next, the three contributions of this thesis are discussed in Chapter 3. First, a novel approach for segment proposals in quantum scale is described in Section 3.1. The major outcome of this thesis is the Quantum-Cuts technique, which is presented in Section 3.2. The improvements and extensions on Quantum-Cuts, namely Learning-to-Rank Salient Segments by Multispectral Quantum-Cuts, Extended Quantum-Cuts and Learning Graph Affinities for Quantum-Cuts, are described in the subsequent sections. Finally, in Chapter 4, major findings of this thesis will be highlighted along with the conclusive remarks, and some topics for future research will be discussed.

1.2 Publications and Author's Contribution

In [P1], an automatic object extraction method in quantum scale was proposed. The major contribution of the paper is the discovery that the solutions of time independent Schrödinger's Equation propose meaningful image segments, on a potential field formed by the gray-level image values. Being a direct application of Quantum Mechanics, an experimental setup for this method is feasible in quantum scale. The segment proposals obtained in this way are then ranked according to an objectness measure and the highest ranked segments are considered as the most interesting object(s) in the images. The candidate is the first author of this publication and is responsible for developing the whole method, performing all experiments and writing the manuscript.

In [P2], a spectral graph based foreground segmentation method, Quantum-Cuts (QCut), was proposed based on a link between Quantum Mechanics and Spectral Graph Theory. The method was developed by integrating the pairwise affinities that are widely used in spectral graph based Computer Vision tasks to the Schrödinger's Equation. The main discovery of the study is that the probabilistic interpretation of the ground state wavefunctions obtained in this way corresponds to approximate solutions of a unique optimization problem that can effectively address the salient object detection problem. The candidate is the first author of this publication and is responsible for developing the whole method, performing all experiments and writing the manuscript.

In [P3], a pool of salient object segments was formed by exploiting a multispectral version of QCut. These segments were ranked according to a quality measure that is learnt by a random forest regressor. Specifically, the regressor predicts scores of segments according to low and mid-level features extracted from them. The top ranked segments in the pool are then merged to obtain a final segmentation mask. This learning-to-rank approach provides improvement over the baseline QCut and achieves the performance level of the state-of-the-art in salient object detection. The candidate is the first author of this publication and is responsible for developing the whole method, performing all experiments and writing the manuscript.

In [P4], an in depth analysis of QCut was conducted and its theoretical advantages over other spectral graph based methods are explained. The main contribution of this study is proposal of extensions of QCut, which include a superpixel abstraction, a multi-resolution approach, increased graph connectivity and novel affinity assignments. These extensions result into a significant improvement over the baseline QCut and achieve a much superior performance than the state-of-the-art methods in automatic salient object detection. The candidate is the first author of this publication and is responsible for developing most of the method, performing most of the experiments and writing most of the manuscript.

In [P5], an affinity learning scheme for salient object detection was proposed based on learning a linear combination of widely used affinity calculation functions. A gradient-descent based method

was adopted for learning the linear combination coefficients via backpropagating the error of output of Extended Quantum Cuts method. It has been shown that even such a simple linear combination of affinity calculation functions leads to a decent improvement in salient object detection performance of EQCut. The candidate is the first author of this publication and is responsible for developing most of the method, performing all experiments and writing the manuscript.

In [P6], an end-to-end learning scheme for salient object detection was proposed based on graph affinity learning for Extended Quantum-Cuts (EQCut). A Convolutional Kernel Network (CKN) was exploited for prediction of affinity matrix related to a graph that represents an image. The CKN was concatenated by EQCut which takes the affinities as input and outputs a saliency map. This system produces a differentiable error with respect to CKN parameters, hence one can backpropagate the error for parameter learning. Such a supervised system can generalize well in test datasets for the error measures that are used during training. The candidate is the first author of this publication and is responsible for developing most of the method, performing all experiments and writing the manuscript.

2 Background

In this chapter, preliminary material will be introduced, which constitute the theoretical background of the contributions that are proposed in this thesis. Since the main focus of this thesis is analyzing the advantages of quantum mechanical principles in salient object detection via an interpretation of quantum mechanics from a spectral graph theory point of view, in this chapter the introduced topics include Quantum Mechanics, Spectral Graph Theory and salient object detection. Moreover, a graph affinity learning section is also included since this topic constitutes a background for the supervised extension of the spectral graph based methods proposed in this thesis.

2.1 Quantum Mechanics

Quantum Mechanics is a fundamental branch of physics that investigates the behavior of matter in very small scale, such as the behavior of electrons in an atom's orbit. The impact of Quantum Mechanics has been huge in the scientific community as it defied the rules of classical physics which was accepted as a complete description of the physical world until 20th century [1]. The unorthodox postulates of Quantum Mechanics are going to be explained in Section 2.1.2, however for an easier understanding of these postulates, a set of events that led to their discovery is first discussed in the next section.

2.1.1 A Historical Review

The birth of Quantum Mechanics followed from a pursuit to explain a set of experiments. The observations from these experiments could not be expressed by the classical mechanics and therefore led to a new branch in physics where objective determinism was replaced by subjective probability [1].

2.1.1.1 Blackbody Radiation

A blackbody is an object having the properties of an ideal absorber and an ideal reflector. When the body reaches the equilibrium temperature with the environment, it radiates light. At a particular equilibrium temperature, if the distribution of the intensity (power) of this light with respect to its wavelength is to be measured, one observes a Gaussian-like distribution with a peak at a certain light wavelength. However, according to Rayleigh-Jeans formula in classical physics, the intensity

should have decreased with increasing wavelength, which would have led to an ultraviolet catastrophe. Examples of intensity distributions for several temperatures and a classical prediction for 2000 K (Kelvin) are illustrated in Figure 1.

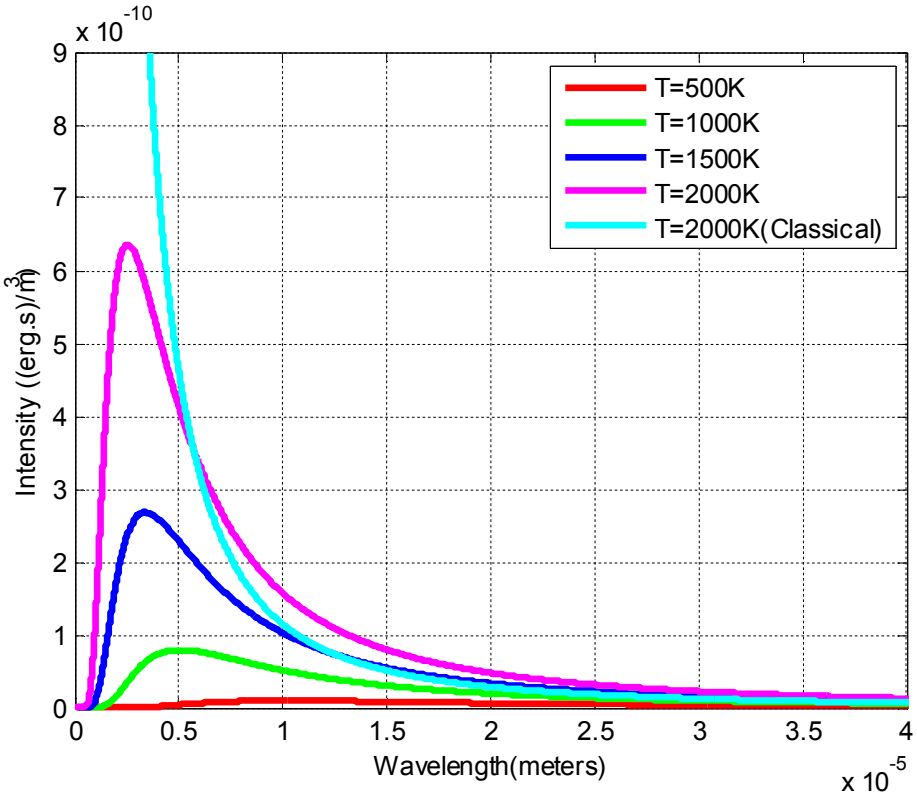


Figure 1 Intensity distributions of blackbodies at T=500K, 1000K, 1500K and 2000K and a prediction of the distribution with the classical Rayleigh-Jeans formulation for T=2000K are illustrated in red, green, blue, magenta and cyan colors, respectively.

The Gaussian-like distribution of the light intensity could only be explained by a theory proposed by Planck in 1901 [2], which states that the light comes in packets of energy and this energy E is expressed as multiples of the light's frequency ν ,i.e.:

$$E = h\nu, \tag{2.1}$$

where h, Planck's constant is a constant of nature. Therefore, it is less probable to have light packets with high frequency (low wavelength) since these packets have large energies according to (2.1). Hence, the intensity at higher frequencies is low. On the other hand, it is more probable to have light packets with low frequency, but due to their low energy, the intensity at this frequency is also low. The peak intensity occurs at a frequency that is mostly compatible with the current tem-

perature of the object. This explanation of blackbody radiation was made possible by (2.1) and each photon was called a quantum of radiation.

2.1.1.2 Photoelectric Effect

In 1905, Einstein constructed an experiment that measures the maximum energy of electrons emitted from a copper plate when it is irradiated with light [3]. Einstein observed that at a certain frequency of light, the maximum electron energy is independent of the intensity of the light. On the other hand, as the frequency of light increases, the maximum electron energy is increased linearly. This experiment also validates (2.1) in a different way. The maximum energy of an electron is given in (2.2) where ν_{th} is the threshold frequency indicating the energy E of the electrons which are easiest to move from the copper plate.

$$E = h\nu - h\nu_{th}. \tag{2.2}$$

In Figure 2, an illustration of the photoelectric effect is provided. The threshold frequency of copper is $\nu_{th} = 1.137\ PHz$. The maximum energy of the electrons emitted from the copper plate when the plate is irradiated with light of different frequencies is shown in Figure 2. Note that, no electrons are emitted before the threshold frequency.

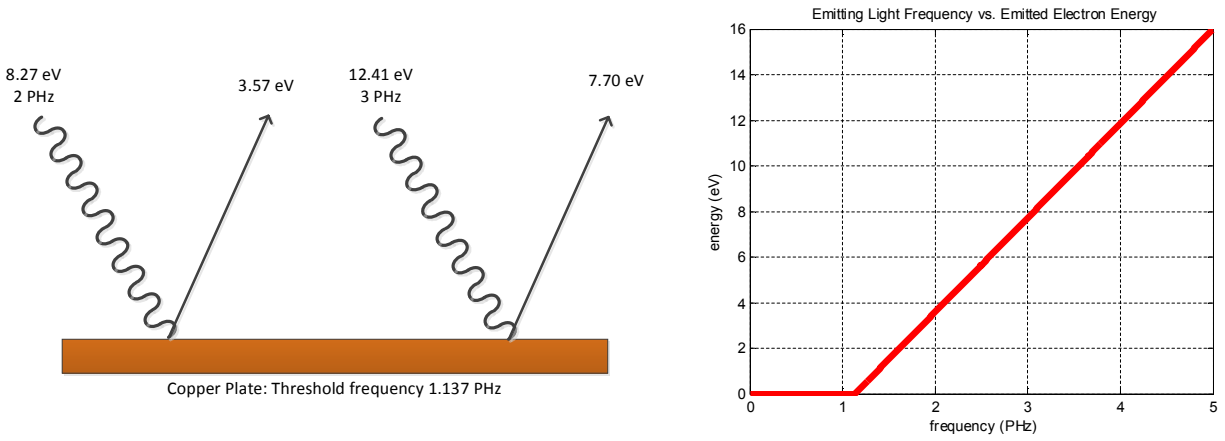


Figure 2 The Photoelectric Effect. (a) Energies of some irradiating lights on a copper plate and the maximum energy of the electrons emitted from the plate. (b) A plot of irradiating light frequency vs. maximum emitted electron energy from the copper plate.

2.1.1.3 Bohr's Model of Hydrogen Atom

In 1913, Bohr proposed a quantum theory of atoms [4]. This new theory was needed since a similar observation to photoelectric effect was made. In other words, when the atom was irradiated with light, the atom was observed to emit electrons for only certain frequencies of the irradiating light.

The main findings of Bohr's theory can be summarized in two postulates: (1) Electrons in hydrogen atom are in discrete energy levels. Accordingly, every discrete energy level is associated with a discrete orbit radius. At these levels the atom does not radiate. The discrete states are called bound states. (2) When an electron transition from one state n to another m occurs, the electron emits light with frequency defined by:

$$h\nu = E_n - E_m. \tag{2.3}$$

In Figure 3, a transition from the second electron orbit to the first one and the resulting emitted light and corresponding energies according to (2.3) are illustrated.

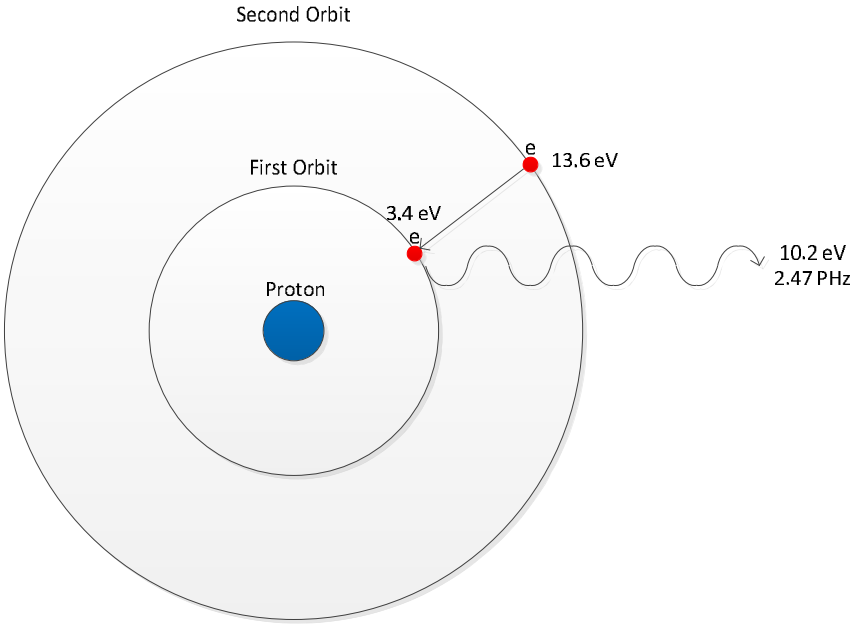


Figure 3 Bohr's Hydrogen Atom Model. A transition of an electron from second orbit to the first orbit and the resulting emitted light according to Bohr's model is illustrated. The real proportional sizes of particles and orbits are distorted in order to fit them in this illustration.

2.1.1.4 Particle-Wave Duality

Although Bohr's theory had a high impact, the reason of his first postulate was unknown. In order to understand why electrons in hydrogen atoms enjoy a discrete energy spectrum, one needs to review classical wave and particle properties. Traditionally, waves and particles were distinguished by the ability of interference. Basically, when two propagating disturbances combine in the same space, if they are particles, they add in intensities and if they are waves, they add in amplitudes. Adding in amplitude gives rise to interference in waves, whereas in particles interference was be-

lieved not to occur. In 1925, De Broglie proposed a counter theory to the common belief that particles also exhibit wave properties, hence they also interfere [5]. The wave nature of particles is described by the relations in (2.4), where p is the momentum, λ is the wavelength, k is the wave number and $\hbar$ is the reduced Planck's constant.

$$\lambda = \frac{h}{p} \; or \; p = \hbar k. \tag{2.4}$$

The relations in (2.4) when combined with Bohr's first postulate, gives the following relation for an electron at a bounded state n and at the corresponding radius r.

$$2\pi r = n\lambda. \tag{2.5}$$

This is the classical formulation for a standing wave on a circle. Interestingly, this relation expresses the fact that, if the electrons exhibit a wave-like behavior, the bounded states in an atom occur only on radii that a standing wave can occur. This explains the first postulate of Bohr's theory. Later in 1927, by Davisson-Germer experiment, it was verified that electrons indeed exhibit interference supporting De Broglie relations and the wave nature of particles.

2.1.1.5 Heisenberg's Uncertainty Principle

Following the wave-particle duality, in 1927, Heisenberg proposed the famous uncertainty principle [6]. This theory is particularly interesting because it defies the deterministic philosophy of classical mechanics which states that according to Newton's rules, once the forces on a particle and its initial location and velocity are known, it is possible to exactly know the position and velocity of the particle at any given instance of time. However, Heisenberg proposed that if the velocity (or momentum) of a particle is precisely known, the position of the particle is completely unknown. This inference is deducted from the following uncertainty principle that relates the standard deviation (uncertainty) of location Δx and of momentum Δp_x.

$$\Delta x \Delta p_x \geq \hbar. \tag{2.6}$$

The relation (2.6) also relates to the wave nature of particles. For example, in an extreme case, if the momentum of an electron is precisely known, one has no clue about its location; hence it acts like a wave in this case. On the other hand, if its location is precisely known, it acts purely as a particle. Uncertainty principle can explain important observations, such as the fact that for double-slit experiment, if the electrons are observed throughout their trajectories, they will lose their wave-

properties. This gives rise to philosophical inferences such as whether the observer has a role in the nature of the observable.

The above works bring the need for a new theory –Quantum Mechanics- in physics, as their conclusions cannot be explained with classical mechanics. For example, if particles have wave-nature, this wave behavior should be formalized in terms of wavefunctions. Another question to be answered is how to find the discrete states, for example in Bohr's hydrogen atom theory. Above, we have mentioned the role of the observer in measurement and uncertainty in particles' properties. How can one formalize the role of the observer and calculate statistical characteristics, e.g. average of particle properties? All these questions can be answered with the following postulates, which serve to formalize the rules of Quantum Mechanics.

2.1.2 Postulates of Quantum Mechanics

2.1.2.1 Postulate 1

An observable A is defined as a self-consistent and well-defined variable corresponding to a property of a particle, such as energy, momentum or location. The first postulate of Quantum Mechanics states that: for any observable in physics, there exists an operator $\hat{A}$, such that the eigenvalues of the operator a yield the measurement values of the observable [1]. This postulate can be formulated with the following eigenvalue equation:

$$\hat{A}\boldsymbol{\psi}(\boldsymbol{x}) = a\boldsymbol{\psi}(\boldsymbol{x}). \tag{2.7}$$

The function $\boldsymbol{\psi}(\boldsymbol{x})$ in (2.7) is called the eigenfunction of $\hat{A}$ corresponding to the eigenvalue a.

For example, the operator corresponding to energy observable is the Hamiltonian $\widehat{H}$ and for a particle of mass m in a potential field $\boldsymbol{V}(\boldsymbol{x})$, it is given as follows. Note that $\boldsymbol{x}$ is the 3-dimensional space coordinates. In the rest of the text, $\boldsymbol{\psi}$ and V will be used instead of $\boldsymbol{\psi}(\boldsymbol{x})$ and $\boldsymbol{V}(\boldsymbol{x})$, for simplicity.

$$\widehat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \boldsymbol{V}. \tag{2.8}$$

The corresponding eigenvalue equation for the energy operator $\widehat{H}$ is given as follows:

$$-\frac{\hbar^2}{2m}\nabla^2\boldsymbol{\psi} + \boldsymbol{V}\boldsymbol{\psi} = E\boldsymbol{\psi}. \tag{2.9}$$

Equation (2.9) is the *time-independent Schrödinger's Equation* and it gives the possible energies of a particle of mass m in a potential field $\boldsymbol{V}$. Following the discussion at the end of the previous subsection, the discrete energy levels of electrons in Hydrogen atom can be determined by (2.9) where m is the electron mass and $\boldsymbol{V}$ is the Coulomb potential caused by the Hydrogen nucleus. It can be noticed that the resulting eigenvalues will form a discrete energy spectrum which corresponds to energy levels of the electrons. The eigenfunction $\boldsymbol{\psi}$ in (2.9) is a wavefunction of the electron, defining its wave nature.

In 1927, Born interpreted the wavefunction $\boldsymbol{\psi}$ as a tool for calculating the probability distribution of a particle's whereabouts in space [7]. As $\boldsymbol{\psi}^* \circ \boldsymbol{\psi}$ is set to the intensity in waves, according to Born, it is more appropriate to define $\boldsymbol{\psi}^* \circ \boldsymbol{\psi}$ as a probability distribution. Following the Hydrogen atom example, for each energy level there is a corresponding wavefunction and the square of the modulus of this wavefunction gives the probability distribution of an electron in space. This interpretation particularly plays a crucial role in the algorithms that will be presented in Chapter 3 of the thesis.

2.1.2.2 Postulate 2

Postulate 1 together with Born's interpretation covers some of the questions that have been raised at the end of subsection 2.1.1. However, one of the questions that remains unanswered is the measurement (role of the observer) in Quantum Mechanics. The second postulate is particularly interested in the measurement of the observables and it states that a measurement on the observable A leaves the particle in a state $\boldsymbol{\varphi}_a$ with value a, hence any subsequent measurement will yield exactly the same value [1]. This means that until measurement, the state of the particle is undetermined, however once a measurement is conducted, the state is determined and any other measurement will find the particle in the same state. A philosophical explanation of this postulate is that until the observer interferes with the particle, the particle can be in many states at the same time and the observer actually plays a role in the latter state of the particle. This inference has been a striking one in physics community and has been the subject of many thought experiments such as the famous Schrödinger's Cat.

2.1.2.3 Postulate 3

This postulate introduces the state of a system that can be represented by the associated continuous and differentiable wavefunction. According to Postulate 3, all information about the system is contained in the wavefunction [1]. Moreover, statistical properties of observables can be extracted from the wavefunction. For example, the average value of an observable A at a state represented by $\boldsymbol{\psi}$ can be obtained as follows.

$$\langle A \rangle = \int \boldsymbol{\psi}^* \hat{A} \, \boldsymbol{\psi} d\boldsymbol{x}. \tag{2.10}$$

Note that a state can be either the eigenfunctions obtained by (2.7) or a linear combination of them as they also satisfy (2.7). For pure states, i.e. states represented by eigenfunctions only, $\hat{A}\,\boldsymbol{\psi} = a\boldsymbol{\psi}$ and $\int \boldsymbol{\psi}^*\boldsymbol{\psi} = 1$ due to Born's probability interpretation of wavefunction. Hence, (2.10) reduces to $\langle A \rangle = a$. In general, the expected value or the average of the value of the observable A has the following meaning. Consider a very large number of identical particles in the same state and we measure the value of an observable for each particle. If one averages over the values of all these measurements, the result is the same as given by (2.10).

2.1.2.4 Postulate 4

Postulates 1-3 cover properties for states at a given time. However, a complete wavefunction should also evolve in time. Postulate 4 describes how this evolution takes place [1]. Specifically, a wavefunction evolves in time according to the *time-dependent Schrödinger's Equation* i.e.:

$$i\hbar \frac{\partial}{\partial t}\boldsymbol{\psi}(\boldsymbol{x}, t) = \hat{H}\boldsymbol{\psi}(\boldsymbol{x}, t). \tag{2.11}$$

With the time evolution defined, postulates of Quantum Mechanics described briefly in this section constitute the basis of Quantum Mechanics.

The contributions mentioned in this thesis are inspired from the Quantum Mechanics postulates explained above and especially Postulate 1, the time independent Schrödinger's Equation and Born's probabilistic interpretation of the wavefunction will play the key role in these contributions. Together with these quantum mechanical concepts, definitions and theorems related to Spectral Graph Theory will give a complete background for the derivation of the proposed methods.

2.2 Spectral Graph Theory

Spectral Graph Theory is particularly interested in eigenproblems related to the characteristic matrices in the graph theory [8]. In order to understand the spectral graph theoretical methods that will be discussed in this section, graph theory preliminaries will be reviewed next.

2.2.1 Graph Theory Preliminaries

Graphs are structures that involve pairwise relations. A typical graph is formed by a set of nodes $\mathcal{V}$ and edges forming an edge set $\mathcal{E}$, and expressing links between the nodes. If an edge is connecting a node to itself, it is called a *loop*. Nodes generally represent objects/points in the related problem and edges represent the relationships between these points. For example, a circuit system can be modelled as a graph where nodes correspond to points that are at either ends of the electrical

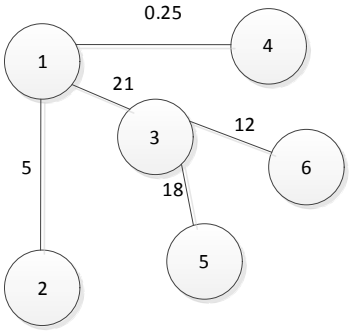
components and the edges of the representative graph reflect the properties of the electrical component. Another example is a model of cities and roads connecting them. In this case nodes correspond to cities and edges might be weighted by the distance of the roads between cities. The weight values and the objects represented by vertices are dependent on the problem at hand.

2.2.1.1 Graph Types

Graphs can be categorized according to their characteristics. For example, an *undirected* graph is a graph where the edges connecting nodes have no orientation. A *directed* graph includes this orientation information on the edges. A *path* in undirected graphs is defined as the set of edges connecting two nodes in a graph. For an undirected graph, a *connected* graph is a graph where each pair of nodes has a path connecting them. In directed graphs, a *directed path* is a set of directed edges with the same directions, which connects two nodes. For a directed graph, *a strongly connected* graph contains a directed path between each pair of nodes in the graph. A *weakly connected* directed graph is a graph where there is a path connecting every pair of nodes, which might include edges with different directions. Two nodes are *neighbors* if there is an edge connecting them. A *regular* graph is a graph where each node has the same number of neighbors. In a weighted graph, there is a number assigned to each edge, which represents a relation between the nodes that the edge connects. For example, in a graph representation of a number of cities and the roads that connects them, an edge values between two nodes may be assigned to the distances of the roads between cities that the nodes represent. There are other graph types and categorizations, however only the related ones to the problems addressed in this thesis are presented here.

2.2.1.2 Graph Matrices

It is possible to represent graphs by using matrices. For example, for an undirected graph, *adjacency matrix* $\mathbf{A}$ represents the connections in a graph. The entry a_{ij} is set to the number of edges connecting nodes i and j. Loops are counted twice and the adjacency matrix is symmetric as the graph is undirected. A *weighted adjacency matrix* $\mathbf{W}$ is designed for weighted graphs. For an undirected graph, the entry w_{ij} is set to the weight of the edge that connects nodes i and j. The *degree* matrix $\mathbf{D}$ related to the weighted adjacency matrix is a diagonal matrix, entries of which are obtained by row-wise summation of the adjacency matrix. The *Laplacian matrix* $\mathbf{L}$ is obtained as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. An exemplary undirected and connected graph and the corresponding graph matrices are illustrated in Figure 4.



$$\mathbf{W} = \begin{bmatrix} 0 & 5 & 21 & 0.25 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 \\ 21 & 0 & 0 & 0 & 18 & 12 \\ 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 18 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 26.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 61 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 \end{bmatrix}$$

$$\mathbf{L} = \begin{bmatrix} 26.25 & -5 & -21 & -0.25 & 0 & 0 \\ -5 & 5 & 0 & 0 & 0 & 0 \\ -21 & 0 & 61 & 0 & -18 & -12 \\ -0.25 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & -18 & 0 & 18 & 0 \\ 0 & 0 & -12 & 0 & 0 & 12 \end{bmatrix}$$

Figure 4 A graph and the corresponding weighted adjacency, degree and Laplacian matrices.

2.2.1.3 Graph Partitioning Measures

Consider a connected graph $\mathcal{G}$ is to be partitioned into two connected subgraphs, namely P and $\bar{P}$. The following definitions give important information about the partitioning.

Definition 2-1 *The cut value related to the partitioning of a graph is defined as the sum of weights of the edges to be cut in order to partition the graph as formulated in* (2.12)*.*

$$cut(P, \bar{P}) = \sum_{i \in P, j \in \bar{P}} w_{i,j}. \tag{2.12}$$

The cut value is extremely useful in many applications of graph theory, such as data clustering [10]. A common practice is to construct a graph with the data samples as graph nodes and assigning the pairwise graph weights to a pairwise similarity (affinity) measure. Then, partitioning the graph into two corresponds to clustering the data into two clusters. The cut value would give a clustering cost, as one would require low affinities in clustering borders.

Definition 2-2 *The volume of a subgraph P is defined as the sum of edge weights connecting the subgraph P to the graph $\mathcal{G}$, which is equal to the summation of the degree value of nodes contained in P,* i.e.:

$$vol(P) = \sum_{i \in P, j \in \mathcal{G}} w_{i,j} = \sum_{i \in P} d_i. \tag{2.13}$$

Definition 2-3 *The association of a subgraph P to another subgraph G is defined as the summation of weights of edges connecting the subgraph P to the subgraph G,* i.e.:

$$assoc(P, G) = \sum_{i \in P, j \in G} w_{i,j}. \tag{2.14}$$

Considering the data clustering example mentioned earlier, the association of a cluster to itself is particularly important, as it gives a notion of the similarity of the cluster within itself.

Definition 2-4 *The area of a subgraph P is defined as the number of nodes contained in P*, i.e.:

$$area(P) = \sum_{i \in P} 1. \tag{2.15}$$

A large number of graph-based methods in data clustering or Computer Vision related tasks, such as image segmentation, are based on optimizing objective functions based on the above measures. Most of these objective functions are intractable; however, approximations can be solved thanks to Spectral Graph Theory.

2.2.2 Spectral Graph Theory

Spectral Graph Theory is particularly interested in the eigen-decomposition of graph matrices discussed in 2.2.1.2. An eigenvalue problem is defined as follows.

$$\mathbf{M}\boldsymbol{\varphi} = \gamma\boldsymbol{\varphi}. \tag{2.16}$$

Specifically $\boldsymbol{\varphi}$ is called an eigenvector of $\mathbf{M}$ with eigenvalue γ if it satisfies (2.16). Hence, γ is an eigenvalue of matrix $\mathbf{M}$ only if $\gamma\mathbf{I} - \mathbf{M}$ is singular. Therefore, the eigenvalues are the roots of the characteristic polynomial of $\mathbf{M}$, which is given by (2.17).

$$det(\gamma \boldsymbol{I} - \boldsymbol{M}). \tag{2.17}$$

Since most of the proposed methods in this thesis involve undirected and connected graphs, Spectral Graph Theory review in this chapter will focus on symmetric graph matrices only. Next, a set of theorems and definitions that are crucial for real symmetric matrices are reviewed.

Theorem 2-1 *If a matrix M is a real symmetric nxn matrix, then there exist n real eigenvalues and n mutually orthogonal eigenvectors.*

This theorem explains a very useful property of symmetric matrices. For an asymmetric matrix, the existence of n eigenvalues is not guaranteed; furthermore if there exist n eigenvalues, the corresponding eigenvectors are not necessarily mutually orthogonal.

Definition 2-5 *The Rayleigh Quotient of a vector x with respect to a matrix M is defined as follows.*

$$R(\boldsymbol{M}, \boldsymbol{x}) = \frac{\boldsymbol{x^T M x}}{\boldsymbol{x^T x}}. \tag{2.18}$$

It can be noticed that if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$, then $R(\boldsymbol{M}, \boldsymbol{x})$ is the corresponding eigenvalue of $\boldsymbol{M}$.

Theorem 2-2 *If a matrix **M** is a real symmetric matrix, then the vector $\boldsymbol{x}$ maximizing/minimizing $R(\boldsymbol{M}, \boldsymbol{x})$ is an eigenvector of **M** and the corresponding eigenvalue is equal to the Rayleigh Quotient and is the maximum/minimum eigenvalue of **M**.*

Proof. Since **M** is an nxn real and symmetric matrix, due to Theorem 2-1, its eigenvectors form an orthonormal basis. Hence any vector $\boldsymbol{x}$ can be written as follows.

$$\boldsymbol{x} = \sum_{i} (\boldsymbol{\varphi}_i{}^T \boldsymbol{x})\, \boldsymbol{\varphi}_i. \tag{2.19}$$

Note that eigenvalues are sorted in an ascending order so that first eigenvalue is the minimum and the n^{th} is the maximum. Next, notice that the scaling of $\boldsymbol{x}$ is irrelevant while calculating the minimum or maximum of $R(\boldsymbol{M}, \boldsymbol{x})$. Hence, without the loss of generality $\boldsymbol{x}$ can be assumed to be a unit vector, i.e. $\boldsymbol{x^T x} = 1$. Then $R(\boldsymbol{M}, \boldsymbol{x})$ can be written as follows.

$$R(\boldsymbol{M},\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{M}\boldsymbol{x} = \left(\sum_{i}(\boldsymbol{\varphi}_i{}^T\boldsymbol{x})\,\boldsymbol{\varphi}_i\right)^T\left(\sum_{j}\left(\boldsymbol{\varphi}_j{}^T\boldsymbol{x}\right)\gamma_j\boldsymbol{\varphi}_j\right) = \sum_{ij}(\boldsymbol{\varphi}_i{}^T\boldsymbol{x})\left(\boldsymbol{\varphi}_j{}^T\boldsymbol{x}\right)\gamma_j\boldsymbol{\varphi}_i{}^T\boldsymbol{\varphi}_j$$
$$= \sum_{i}(\boldsymbol{\varphi}_i{}^T\boldsymbol{x})^2\,\gamma_i \le \gamma_n\sum_{i}(\boldsymbol{\varphi}_i{}^T\boldsymbol{x})^2 = \gamma_n. \tag{2.20}$$

In (2.20), we used of the following facts: (1) $\boldsymbol{\varphi}_i{}^T\boldsymbol{\varphi}_j = 0$ if $i \neq j$ as a result of Theorem 2-1 and (2) previously assumed $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x} = 1$. According to (2.20), γ_n is the upper bound for the Rayleigh Quotient and it can be observed that it can only be achieved by setting $\boldsymbol{x} = \boldsymbol{\varphi}_n$. Remember that $\boldsymbol{\varphi}_n$ is the eigenvector corresponding to the largest eigenvalue γ_n of $\boldsymbol{M}$. Hence, the proof is completed.

Definition 2-6 *The Generalized Rayleigh Quotient of a vector $\boldsymbol{x}$ with respect to matrices **M** and **N** is defined as follows.*

$$R_g(\boldsymbol{M},\boldsymbol{N},\boldsymbol{x}) = \frac{\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{N}\boldsymbol{x}}. \tag{2.21}$$

Theorem 2-3 *If the matrices **M** and **N** are real symmetric matrices, then the vector $\boldsymbol{x}$ maximizing/minimizing $R_g(\boldsymbol{M},\boldsymbol{N},\boldsymbol{x})$ is a generalized eigenvector of **M** and **N** and the corresponding eigenvalue is equal to the Rayleigh Quotient and is the maximum/minimum eigenvalue of **M** and **N**.*

$$\underset{\boldsymbol{x}}{\operatorname{argmin}}\left(R_g(\boldsymbol{M},\boldsymbol{N},\boldsymbol{x})\right) = \boldsymbol{\varphi}_1, \tag{2.22}$$
$$\boldsymbol{M}\boldsymbol{\varphi}_1 = \gamma_1\boldsymbol{N}\boldsymbol{\varphi}_1.$$

The proof can be made by following a similar procedure to the proof of Theorem 2-2. As mentioned in Section 2.2.1.3, Spectral Graph Theory is particularly useful in approximating the intractable optimization problems related to graph partitioning. Theorem 2-2 and Theorem 2-3 are in fact the key theorems making these approximations possible. In the next section, some spectral graph partitioning methods will be discussed making use of these theorems.

2.2.2.1 Spectral Graph Partitioning Methods

Graph partitioning measures defined in Section 2.2.1.3, can be represented in terms of matrix algebra. Consider a graph $\mathcal{G}$ is to be partitioned into two, P and $\bar{P}$. For example, one can represent this partitioning with a labeling vector $\boldsymbol{x}$ with the following property.

$$x_i = \begin{cases} 0\,, i \in P \\ 1\,, i \in \bar{P} \end{cases}. \tag{2.23}$$

Then, the cut measure can be formulated as follows.

$$cut(P, \bar{P}) = \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \frac{1}{2} \sum_{ij} \left(x_i - x_j\right)^2 w_{ij}. \tag{2.24}$$

(2.24) sums the graph weights that are between nodes that are labelled differently in $\boldsymbol{x}$. Since the summation takes both w_{ij} and w_{ji}, it is multiplied by $\frac{1}{2}$. Note that, we assume a real symmetric weighted adjacency matrix, hence $w_{ij} = w_{ji}$.

Each spectral graph partitioning method has a unique objective function; therefore, the labeling vector is specifically defined.

Ratio Cut

The Ratio Cut [9] measure related to partitioning a graph $\mathcal{G}$ into P and $\bar{P}$ is given as follows.

$$Rcut(P, \bar{P}) = \frac{cut(P, \bar{P})}{area(P) area(\bar{P})}. \tag{2.25}$$

This measure is widely used for data clustering. Specifically, in a graph where the nodes correspond to samples and the graph weights correspond to affinities between data pairs, the minimization of Ratio Cut measure is assumed to give partitions that correspond to good clusters. A smaller Ratio Cut results into clusters that are separated from the graph edges that have weaker affinities. This is desirable since it means that the nodes corresponding to different clusters are highly different from each other at cut edges. Furthermore, area normalization is useful if one expects clusters of similar area. Ratio Cut can be formalized by matrix multiplication in a similar way as explained in Section 2.2.2.1. However, one needs to define a labelling vector as follows:

$$x_i = \begin{cases} p\,, i \in P \\ -q\,, i \in \bar{P} \end{cases}. \tag{2.26}$$

In (2.26), $p = \frac{area(\bar{P})}{n}$ and $q = \frac{area(P)}{n}$, where $n = area(V)$. Hence according to this labeling vector construction, the following relation between the cut measure and $\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}$ can be observed as follows.

$$\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} = \frac{1}{2}\sum_{ij}\left(x_i - x_j\right)^2 w_{ij} = \frac{1}{2}\sum_{i\neq j}(p+q)^2\, w_{ij} = \frac{1}{2}\sum_{i\neq j} w_{ij} = cut(P,\bar{P}). \tag{2.27}$$

Finally, from (2.26) one can infer the following.

$$\boldsymbol{x}^T\boldsymbol{x} = q^2pn + p^2qn = pq(qn+pn) = pqn = \frac{1}{area(V)}area(P)area(\bar{P}). \tag{2.28}$$

By combining (2.27) and (2.28), $Rcut(P,\bar{P})$ can be written as follows.

$$Rcut(P,\bar{P}) = n\frac{\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}. \tag{2.29}$$

With the definition of labeling vector $\boldsymbol{x}$ in its current form (according to (2.26)) the minimization of (2.29) is intractable. However, if the values of $\boldsymbol{x}$ are relaxed to take other values than p and q, the solution can be obtained by Theorem 2-2 and $\boldsymbol{x} = \boldsymbol{\varphi}_1$, where $\boldsymbol{\varphi}_1$ is the eigenvector of $\boldsymbol{L}$ corresponding to the minimum eigenvalue. Notice that the constant n does not have an effect on the minimization. However, since $\boldsymbol{L}$ has the minimum eigenvalue of 0, $\boldsymbol{\varphi}_1$ is constant and meaningless. Therefore, it is a common practice to take the eigenvector $\boldsymbol{\varphi}_2$ corresponding to the second minimum eigenvalue as the approximation of the labeling vector as formulated in (2.30).

$$\begin{aligned} &\underset{\boldsymbol{x}}{\operatorname{argmin}}\left(Rcut(P,\bar{P})\right) = \boldsymbol{\varphi}_2, \\ &\boldsymbol{L}\varphi_2 = \gamma_2\boldsymbol{\varphi}_2. \end{aligned} \tag{2.30}$$

The partitioning can then be made by a simple thresholding on $\boldsymbol{\varphi}_2$, as in (2.31). The threshold τ can either be selected as zero or as the median value of $\boldsymbol{\varphi}_2$.

$$\begin{cases} i \in P\,, \varphi_2(i) > \tau \\ i \in \bar{P}, \varphi_2(i) \leq \tau \end{cases}. \tag{2.31}$$

The spectral approximation for Ratio Cut minimization is useful in the sense that it gives fair clustering results, though it should be noted that the assumption in (2.26) is a strong one and therefore the approximation is actually quite coarse. However, one cannot measure how coarse it is since the actual optimization is intractable.

Normalized Cut

As explained in the previous subsection, minimization of Ratio Cut objective function would ideally result into similar sized subgraphs that are separated from a weak boundary. However, the assumption of similar sized subgraphs is not necessarily a valid one and the objective function in (2.25) does not include some intra characteristics of partitions such as their volume. Often, in applications such as data clustering, one might expect data in a particular cluster to be similar within. Normalized Cut [10] takes this criterion into account, as well as the inter-cluster dissimilarity, by the following objective function:

$$Ncut(P,\bar{P}) = \frac{cut(P,\bar{P})}{vol(P)vol(\bar{P})}. \tag{2.32}$$

A minimization of the criterion in (2.32) would result in subgraphs that are separated from a boundary involving weak graph weights, with high within-subgraph similarity. $Ncut(P,\bar{P})$ can be formulated using Matrix Algebra using the following labeling vector.

$$x_i = \begin{cases} p\,, i \in P \\ -q\,, i \in \bar{P} \end{cases}. \tag{2.33}$$

In (2.33), $p = \frac{vol(\bar{P})}{v}$ and $q = \frac{vol(P)}{v}$, where $v = vol(\mathcal{G})$. Then, the following relation can be written between $\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x}$ and $cut(P,\bar{P})$:

$$\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} = \frac{1}{2}\sum_{ij}\left(x_i - x_j\right)^2 w_{ij} = \frac{1}{2}\sum_{i\neq j}(p+q)^2\, w_{ij} = \frac{1}{2}\sum_{i\neq j} w_{ij} = cut(P,\bar{P}). \tag{2.34}$$

Furthermore, using (2.33) one can infer the following:

$$\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{x} = q^2pv + p^2qv = pq(qv+pv) = pqv = \frac{1}{vol(\mathcal{G})}vol(P)vol(\bar{P}). \tag{2.35}$$

By combining (2.34) and (2.35), one can obtain:

$$\mathrm{Ncut}(\mathrm{P},\overline{\mathrm{P}}) = v\frac{\boldsymbol{x}^{\mathrm{T}}\mathbf{L}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\mathbf{D}\boldsymbol{x}}. \tag{2.36}$$

An approximation of minimization of $\mathrm{Ncut}(\mathrm{P},\overline{\mathrm{P}})$ is possible by relaxing the values of the labeling vector $\mathbf{x}$ to take values other than the ones defined in (2.33). Then, according to Theorem 2-3, the minimization is given by the generalized eigenvector of $\mathbf{L}$ and $\mathbf{D}$ corresponding to the minimum generalized eigenvalue γ_1. Notice that the constant v does not have an effect on the minimization. However, since $\mathbf{L}$ has the minimum eigenvalue of 0, $\boldsymbol{\varphi}_1$ is constant and meaningless. Therefore, it is a common practice to take the generalized eigenvector $\boldsymbol{\varphi}_2$ corresponding to the second minimum generalized eigenvalue γ_2 as the approximation of the labeling vector as formulated in (2.37).

$$\begin{aligned} &\underset{\mathbf{x}}{\operatorname{argmin}}\left(\mathrm{Ncut}(\mathrm{P},\overline{\mathrm{P}})\right) = \boldsymbol{\varphi}_2, \\ &\mathbf{L}\boldsymbol{\varphi}_2 = \gamma_2\mathbf{D}\boldsymbol{\varphi}_2. \end{aligned} \tag{2.37}$$

The partitioning can then be made by a simple thresholding on $\boldsymbol{\varphi}_2$ as in (2.38). The threshold τ can either be selected as zero or as the median value of $\boldsymbol{\varphi}_2$.

$$\begin{cases} i \in P, \boldsymbol{\varphi}_2(i) > \tau \\ i \in \bar{P}, \boldsymbol{\varphi}_2(i) \leq \tau \end{cases}. \tag{2.38}$$

Average Association

Normalized Cuts and Ratio Cuts try to separate a graph into meaningful subgraphs while treating each subgraph equally in the objective functions. In data clustering applications this might correspond to two clusters that successfully represent two separate classes. However, in some applications, one might need a 1-class clustering, i.e. there is a cluster representing a class and the rest of the data is basically noise. In such a case, it is meaningless to use Ratio Cut or Normalized Cut as their objective functions are not designed to handle these particular problems of this nature.

Average association [11] method is designed to partition the data according to the 1-class clustering paradigm. The objective function is defined as follows.

$$Aassoc(P) = \frac{assoc(P,P)}{area(P)}. \tag{2.39}$$

Maximization of (2.39) would ideally partition the graph such that the subgraph P representing the class is a compact one with high within-subgraph similarity.

As in the case of Ratio Cut and Normalized Cut, the maximization of $Aassoc(P)$ is intractable. Hence, one needs a spectral approximation. For this purpose, one can start with defining a labeling vector as follows.

$$x_i = \begin{cases} 1\,, i \in P \\ 0\,, i \in \bar{P} \end{cases}. \tag{2.40}$$

By this labeling vector definition, $Aassoc(P,P)$ can be represented as follows.

$$Aassoc(P,P) = \frac{1}{2}\sum_{ij} x_i x_j\, w_{ij} = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{x}. \tag{2.41}$$

Furthermore, following (2.40), $\boldsymbol{x}^T\boldsymbol{x}$ directly gives the area of region P. Thus, one can formulate $Aassoc(P,\bar{P})$ by using Matrix Algebra as follows.

$$Aassoc(P) = \frac{1}{2}\frac{\boldsymbol{x}^T\boldsymbol{W}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}. \tag{2.42}$$

The maximization of (2.42) can be approximated by relaxing the values of $\boldsymbol{x}$ to have real values. Then, by Theorem 2-2, the solution is given by the eigenvector $\boldsymbol{\varphi}_m$, corresponding to the largest eigenvalue of $\boldsymbol{W}$.

$$\begin{aligned} &\underset{\boldsymbol{x}}{\operatorname{argmax}}\big(Aassoc(P)\big) = \boldsymbol{\varphi}_{\mathrm{m}}, \\ &\boldsymbol{W}\boldsymbol{\varphi}_{\mathrm{m}} = \gamma_{\mathrm{m}}\boldsymbol{\varphi}_{\mathrm{m}}. \end{aligned} \tag{2.43}$$

In order to match the labeling vector given in (2.40), one can make use of non-negative or non-positive eigenvectors keeping in mind that non-positive eigenvectors can be easily converted into non-negative ones as changing the sign of the eigenvector does not change the fact that it is still an eigenvector. However, eigenvectors having positive and negative values were not taken into account in the original work. Once the eigenvector is converted into a non-negative one, then the partitioning can be utilized simply by thresholding the eigenvector with a small value τ as follows.

$$\begin{cases} i \in P\,, \varphi_m(i) > \tau \\ i \in \bar{P}, \varphi_m(i) \le \tau \end{cases}.$$

Each spectral graph partitioning method has its advantages in different types of problems. Below, an exemplary graph, where the assumptions raised by Average Association method hold, is illustrated.

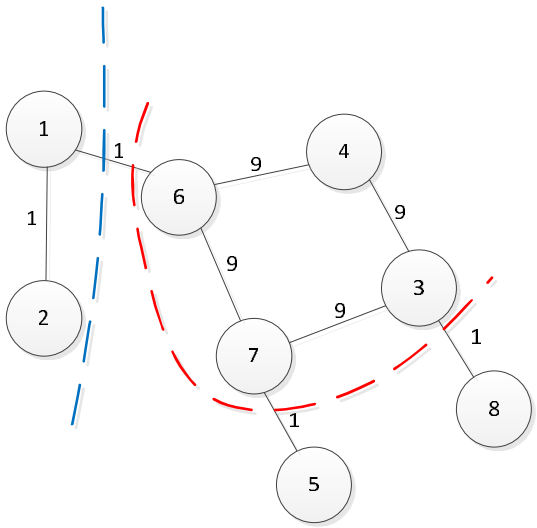


Figure 5 A graph and the partitions obtained by graph partitioning methods. (Blue: Ratio Cut and Normalized Cut, Red: Average Association)

By observing Figure 5, one can see that there is a group of nodes that are strongly connected to each other ($\{3,4,6,7\}$) and other nodes ($\{1,2,5,8\}$) are loosely connected to this group. Therefore, one can infer a partitioning $P = \{3,4,6,7\}$, $\bar{P} = \{1,2,5,8\}$. Such partitioning is indeed achieved only by Average Association method, since it is designed to find a strongly connected compact group, while Ratio Cut and Normalized Cut methods are not.

In Figure 6, another example is provided. In this particular case, if the weakest edges are cut, the graph can be divided into two subgraphs. This intuitive partitioning can only be achieved by Normalized Cuts. Average Association method focuses on the subgraph ($\{1,2,3\}$) with highest average association, whereas Ratio Cut partitions the graph from the middle. The cut value here is low, but not as low as the cut achieved by Normalized Cuts. The reason for this is the fact that Ratio Cut favors partitions with similar size.

Despite their differences, due to their strengths in representing specific cases, all of the above graph-partitioning methods are widely exploited in many applications in Computer Vision and Pattern Recognition tasks, such as image segmentation, data clustering, etc. The performances of the methods discussed in this section depend on the graph representation of the corresponding problem. In fact, from the objective functions of all methods, it might be seen that the performance of each method is only determined by the affinity matrix $\mathbf{W}$. Therefore, in some cases it might be useful to adopt supervised techniques for construction of appropriate affinity matrices. In the next section, these methods are briefly reviewed.

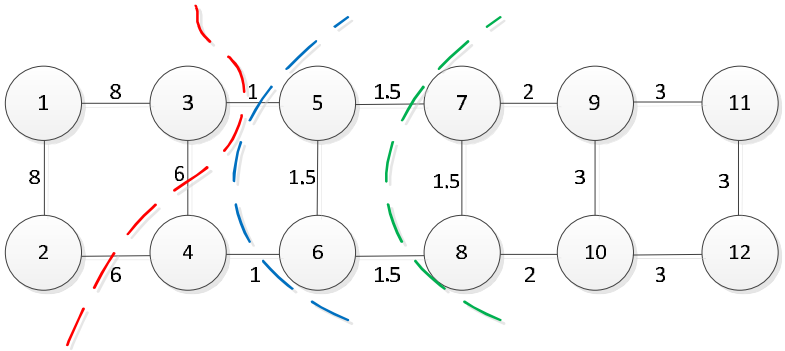


Figure 6 A graph and the partitions obtained by graph partitioning methods. (Red: Average Association, Blue: Normalized Cut and Green: Ratio Cut)

2.3 Graph Affinity Learning

As previously mentioned, the graph representation of the problem at hand is a factor for the performance of graph-based methods and contains three main steps.

a. Nodes: The graph nodes are designed in order to represent the variables/objects in the problem. For example, in data clustering, nodes usually correspond to data samples and in image segmentation to image pixels or regions.
b. Graph Connectivity: This step defines which pairs of the nodes are connected in the given graph and which are not. The connectivity is usually rule-based. For example, if the data is distributed in a regular grid, such as the case in images, the graph can be defined to be 4-connected, i.e. each pixel is connected to neighboring up, down, right and left pixels. For irregular mesh structures, one needs other connectivity rules. An example is the k-nearest neighbor graphs, where each node is connected to the closest –in terms of a given distance metric- k nodes.
c. Edge Weights: Most of the graph based methods make use of weighted graphs. In fact, edge weights contain the most crucial information in the graph representation of the problem at hand since they contain the pairwise relations of graph nodes. Each edge weight is usually selected as a measure of similarity (affinity) between a data pair. The affinity is often calculated as a nonlinear function of distance of feature vectors representing graph nodes.

A graph construction example for images is given in Figure 7. The graph is constructed in such a way that the nodes will correspond to regions (superpixels extracted with [12]) from the image to be represented. Superpixels are regions that provide an over-segmentation on the image while preserving the object and region boundaries. As shown in Figure 7-b, a node is connected to an-

other if the corresponding superpixels share a boundary. The features representing the nodes are selected as mean colors in superpixels. These colors are illustrated in Figure 7-b. Moreover, superpixel boundaries were indicated in red color and the connections of a particular superpixel to its neighbors are illustrated by coding the edges in yellow color. The affinities were calculated according to an exponentially decaying function as the feature distances increase. The resulting graph is shown in Figure 7-c. The strong connections are represented by a darker edge color. It can be observed from Figure 7-c that, such a graph representation assigns very weak edges around the objects which is desired if the goal is to partition the graph such that the objects in the image (bears) will be recovered.

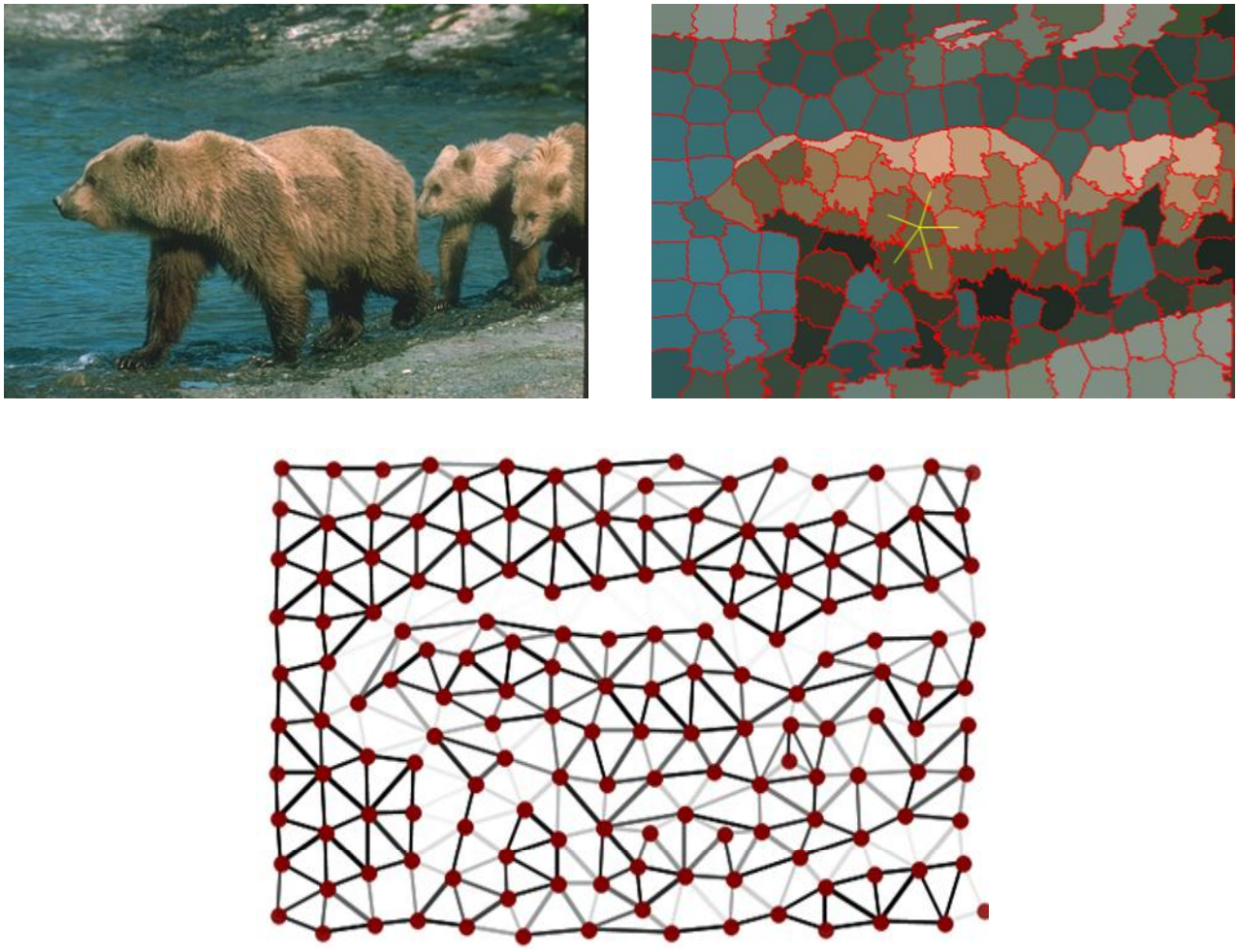


Figure 7 (a) An image from BSD dataset, (b) superpixel boundaries indicated by red colors, superpixel regions illustrated by their average color within, the graph connectivity rule for a particular superpixel (edges in yellow), (c) nodes and edges for the graph representation, weights of edges are coded in color where darker color means stronger connection.

As in the above example, the node assignment step is somewhat straightforward for most of the problems. In contrast, graph connectivity and affinity assignment are not as straightforward. The

variety of graph connectivity and affinity assignment rules are immense and these rules can affect the performance of the method greatly. For example, a more sophisticated affinity calculation might result in even weaker edges around and stronger ones within the salient object in Figure 7. However, it might be very difficult to come up with such a function. Thus, instead of relying on some hand crafted functions for this step, there have been efforts to learn them. A brief review of these graph affinity learning methods is provided next.

In [13], a convolutional neural network (CNN) based approach for affinity learning is proposed for 3D biomedical image segmentation. The graph connectivity was fixed to a 6-neighbourhood rule, i.e. every pixel is connected to immediate neighboring pixels in three dimensions. The affinities of such a graph can be represented in three images, each corresponding to the affinities in three dimensions. Therefore, a convolutional network can be designed which takes a 3D image as input and outputs three images representing affinities. Such a network can be trained if affinity ground truths are present. Therefore, affinity ground truths in each dimension were extracted for each 3D image manually. The image segmentation was performed by an algorithm which is independent from the affinity learning process. This approach is suboptimal since there is no guarantee that the learned affinities will work well with the image segmentation algorithm.

Another approach tackles this problem by proposing an end-to-end learning scheme for image segmentation. For this task, one needs a continuous and differentiable error measure for image segmentation so that one can make use of machine learning approaches exploiting gradient descent like approaches. Convolutional neural networks are such type of machine learning approaches. In [14],a modification of the work in [13] was presented by concatenating an image processing module to CNN model in [13]. The image segmentation module was designed by exploiting a continuous and differentiable approximation of the Rand index which is a simple measure of quality of the image segmentation. Rand index is based on a segmentation procedure by connected component labeling on a graph by cutting the graph edges that have affinities lower than a threshold. Such a simple image segmentation technique does not make use of global inference and relies heavily on the affinity assignment and therefore is incapable of solving many complex image segmentation tasks.

Image segmentation algorithms that have a superior capability for solving complex tasks such as Normalized-Cuts can be used to overcome the segmentation limitations of [14]. In [15] a system for segmenting out foreground objects was proposed by backpropagating the error from Normalized-Cuts output to the parameters of the affinity calculating function. However, the affinity calculating function has a very low capacity of predicting the affinities that will work well with Normalized Cuts, especially when compared to the high capacity CNN models of [14] and [15]. Moreover, although Normalized Cuts algorithm has been proven to work well in over-segmentation, it offers little for foreground segmentation.

A more sophisticated model in [16] has combined both a high capacity affinity learning module and a powerful segmentation algorithm for foreground segmentation task. The affinity learning module is selected as a 5-layer CNN and the segmentation method that is exploited is a convex variational relaxation of the minimum s-t graph cut algorithm in [17]. The connectivity rule on the graph representation of the 2D images is defined as 4-connected neighborhood so that the CNN can be trained to give three matrix outputs, affinities in two dimensions and a unary potential matrix which are the inputs for graph cut algorithm. The limitation of this model is the fact that it is only capable of learning graphs that are 4-connected. However, graphs with higher connectivity gives a much better representation of many problems and can help the performance in complex cases.

Convolutional Kernel Networks, capable of calculating graph affinities related to higher order connectivity have been recently proposed as a family of convolutional neural networks. In the following section, a brief overview will be given.

2.3.1 Convolutional Kernel Networks

Convolutional Neural Networks (CNNs) have proven to be extremely effective in many computer vision tasks such as object recognition [18]-[20], semantic segmentation [21], [22], etc. Usually, the CNNs are exploited in order to predict attributes of a single input, e.g. the classes of the objects that are present in an image. In order to make use of the powerful abilities of CNNs for learning pairwise attributes, Convolutional Kernel Networks (CKNs) were proposed as a hierarchical kernel producing Convolutional Neural Networks [23]. CKNs are designed to estimate the similarity of two inputs $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}'$ as formulated by a modification of the RBF kernel as follows:

$$K(\boldsymbol{\varphi}, \boldsymbol{\varphi}'): = \sum_{\boldsymbol{z}\in\Omega}\sum_{\boldsymbol{z}'\in\Omega} \|\boldsymbol{\varphi}(\boldsymbol{z})\|_2 \|\boldsymbol{\varphi}'(\boldsymbol{z}')\|_2 e^{-\frac{1}{2\beta^2}\|\boldsymbol{z}-\boldsymbol{z}'\|_2^2} e^{-\frac{1}{2\sigma^2}\|\tilde{\boldsymbol{\varphi}}(\boldsymbol{z})-\tilde{\boldsymbol{\varphi}}'(\boldsymbol{z}')\|_2^2}. \tag{2.44}$$

In (2.44), Ω represents image coordinates from which patches $\boldsymbol{z}$ and $\boldsymbol{z}'$ are sampled, and $\tilde{\boldsymbol{\varphi}}(\boldsymbol{z})$ is the l_2 normalized version of $\boldsymbol{\varphi}(\boldsymbol{z})$. The above kernel measures the similarity of two inputs $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}'$ of the same size, based on the spatial and feature similarity of sampled patches from them. These similarity measures are encoded in Gaussian kernels with smoothing parameters β and α for spatial and feature similarity, respectively.

The Gaussian kernel of two input vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ of dimension m can be expanded as follows.

$$G(\boldsymbol{x}, \boldsymbol{x}') = e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x}-\boldsymbol{x}'\|_2^2} = \left(\frac{2}{\pi\sigma^2}\right)^{\frac{m}{2}} \int e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}\|_2^2} e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}'-\boldsymbol{w}\|_2^2} d\boldsymbol{w}. \tag{2.45}$$

Instead of the exact expansion with infinite number of *filters* denoted by $\boldsymbol{w}$, one can approximate the Gaussian kernel as a finite sum which corresponds to the following dot product.

$$G(\boldsymbol{x}, \boldsymbol{x}') \cong \langle f(\boldsymbol{x}), f(\boldsymbol{x}') \rangle. \tag{2.46}$$

The kernel mapping function in (2.46) is defined as $f(\boldsymbol{x}) = \left[\sqrt{c}e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}_1\|_2^2} \ldots \sqrt{c}e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}_p\|_2^2}\right]$ where $\boldsymbol{w}_i$ represents the i^{th} filter and c is the constant in front of the integral in (2.45). With the approximation with finite number of filters, one can represent both Gaussians in (2.44) and train an unsupervised network by learning the filters for each Gaussian approximation.

For high dimensions, one can make a sparser, thus computationally more efficient approximation of the Gaussian kernel by introducing filters weights $\boldsymbol{\eta}$ as follows.

$$G(\boldsymbol{x}, \boldsymbol{x}') \cong \sum_{l=1}^{p} \eta_l e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}_l\|_2^2} e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}'-\boldsymbol{w}_l\|_2^2} = \langle f(\boldsymbol{x}), f(\boldsymbol{x}') \rangle, \tag{2.47}$$

where $f(\boldsymbol{x}) = \left[\sqrt{\eta_1}e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}_1\|_2^2} \ldots \sqrt{\eta_p}e^{-\frac{1}{\sigma^2}\|\boldsymbol{x}-\boldsymbol{w}_p\|_2^2}\right]$. Based on the above definitions, one can approximate the kernel in (2.44) as follows.

$$K(\boldsymbol{\varphi}, \boldsymbol{\varphi}') \cong \sum_{\boldsymbol{z},\boldsymbol{z}' \in \Omega} \langle \boldsymbol{\varsigma}(\boldsymbol{z}), \boldsymbol{\varsigma}(\boldsymbol{z}') \rangle e^{-\frac{1}{2\beta^2}\|\boldsymbol{z}-\boldsymbol{z}'\|_2^2}. \tag{2.48}$$

where $\boldsymbol{\varsigma}(\mathrm{z}) = \|\boldsymbol{\varphi}(\mathrm{z})\|_2 \left[\sqrt{\eta_1}\mathrm{e}^{-\frac{1}{\sigma^2}\|\tilde{\boldsymbol{\varphi}}(\mathrm{z})-\boldsymbol{w}_1\|_2^2} \ldots \sqrt{\eta_\mathrm{p}}\mathrm{e}^{-\frac{1}{\sigma^2}\|\tilde{\boldsymbol{\varphi}}(\mathrm{z})-\mathbf{w}_\mathrm{p}\|_2^2}\right]$. The Spatial Gaussian kernel $\mathrm{e}^{-\frac{1}{2\beta^2}\|\mathbf{z}-\mathbf{z}'\|_2^2}$ can be also approximated by uniform sampling on Ω and, finally, the following kernel approximation can be expressed.

$$K(\boldsymbol{\varphi}, \boldsymbol{\varphi}') \cong \sum_{\boldsymbol{u} \in \Omega} \langle \boldsymbol{\xi}(\mathbf{u}), \boldsymbol{\xi}'(\mathbf{u}) \rangle, \tag{2.49}$$

where $\boldsymbol{\xi}(\mathbf{u}) = \sum_{\boldsymbol{z} \in \Omega} e^{-\frac{1}{\beta^2}\|\boldsymbol{u}-\boldsymbol{z}\|_2^2} \varsigma(\boldsymbol{z})$.

The kernel approximation given by (2.48) can be represented by a 1 layer convolutional neural network where ς can be calculated by convolutions followed by Gaussian nonlinearities and ξ can be calculated by pooling on ς. This can be generalized to a higher number of layers to come up with a reproducing kernel. More detail on reproducing kernels can be found in [23].

If the inputs are considered as features rather than images, CKNs provide a high capacity network to calculate pairwise affinities of these features. This can be very useful for graph based models if one appends a powerful method for the task at hand. As will be described in Section 3.5, the related contribution of this thesis lies in the exploitation of such a machine for both unsupervised and supervised affinity learning.

The graph affinity learning topic discussed in this section gives a complementary background to the previously discussed concepts of Quantum Mechanics and Spectral Graph Theory in terms of constituting a theoretical basis for the contributions of this thesis. Next, the application topic of the thesis, i.e. salient object detection, will be reviewed.

2.4 Salient Object Detection

Salient objects in a scene can be defined as the objects that are more appealing to the eye when compared to the rest of the scene. Salient object detection from images and videos has been a very active research topic in the last decade. The reason for this interest is due to the wide application areas of salient object detection as a pre-processing block for higher level tasks such as object tracking [24], object detection [25], object region proposals [26], object recognition [26], image cropping [27], image manipulation [28], etc.

Early works on saliency were mostly inspired by neuroscience [29], particularly the attention mechanisms of the brain [30]. Thus, these works were mostly concentrated on estimating the eye fixations of humans on scenes with the aim of mimicking the brain's salient region selection function. Such sparse eye fixations can be useful in applications such as video surveillance [31], compression [33] and coding [32]. However, for tasks including object recognition and tracking, one needs a well-defined object region with closed boundaries. Thus, salient object detection emerged as a research track to address these needs. The difference between the eye fixation prediction and salient object detection is basically on the ground truths of the images. Salient object detection datasets include ground truths of object regions as opposed to sparse point clouds of eye fixations. An example is illustrated in Figure 8.

Figure 8 An image (left) from DUT-OMRON dataset showing the eye fixations (middle) and the pixel-wise salient object ground truth (right).

2.4.1 State-of-the-art in Salient Object Detection

Most of the works on salient object detection are computational models based on saliency cues either highlighted by earlier findings in neuroscience or by hand-crafted measures that represent the saliency problem. These cues can be listed as follows:

a. Local Contrast Cue: This cue is based on the assumption that a salient object is expected to be dissimilar from, hence in contrast with its immediate surrounding.
b. Global Contrast Cue: A salient object may also have characteristics that make it globally distinct in the whole scene.
c. Center Prior: Due to the humans' inclination to take photos where the object of interest is in the center, one might assume apriori information that the salient object will be close to the center of the image.
d. Boundary Connectivity Cue: Center prior is a strong assumption. A related, yet a more general assumption is that the salient object is less likely to be present on the boundaries of the image.

Most of the successful salient object detection methods make use of at least one of the above saliency cues above. In fact, the top performing methods in a recent salient object benchmark [45], exploit a combination of above cues. Besides these cues, a number of supporting techniques are also adopted by salient object detection methods. These techniques are listed as follows.

a. Image Abstraction: Working on pixel-level is computationally costly. Moreover, pixel-level features such as colors can be noisy. Both of these problems can be overcome by representing the image with superpixels which are regions that contain pixels similar in color and preserve the boundaries existing on the image.

b. Multi Resolution: Image abstraction can be made in several resolutions which are measured by the average area of the superpixels, each level includes different scale information. It is a common practice to run a saliency detection method in several resolutions and then merge the results.
c. Smoothness Enforcement: The smoothness of a saliency map is defined as the consistency in saliency labels of regions that are similar to each other in a feature space such as color and spatial coordinate features. Enforcing smoothness either in saliency evaluation or as a post-processing stage results in more consistent and connected regions.

Next, a brief description of the state-of-the-art methods in salient object detection according to the survey in [44] and benchmark in [45] is given. It should be noted that, among these methods, Quantum-Cuts [P2] is not discussed in this section, since it is a contribution of this thesis and will be described in detail in Section 3.2. Besides the ones mentioned in [44], [45] a brief description of two recent salient object detection methods that make use of deep learning is also provided.

2.4.1.1 Saliency Detection via Dense and Sparse Reconstruction

In this method [46], first a set of background templates are extracted from the boundaries of an image. These templates are represented by a feature vector of color and coordinate values. Top principal components of the background feature vectors are evaluated by performing singular value decomposition on the covariance matrix of these features. Next, each region in the image is densely reconstructed from the top principal components and the reconstruction error is taken as a measure of saliency. This makes sense in boundary connectivity cue perspective, since the samples taken from the image boundaries are not expected to represent the salient object. The dense reconstruction is susceptible to noise; hence another reconstruction procedure is proposed that enforces the sparsity of the representation achieved by the templates for each image region. Sparse reconstruction is complementary to dense reconstructions, since it is less susceptible to noise yet finds a smaller number of regions. Therefore, both measures were combined following a Bayesian approach in order to find a good saliency measure.

2.4.1.2 Saliency Optimization from Robust Background Detection

The boundary connectivity cue is a valid assumption, however although *most* of the regions in the image boundaries are background, still there might be a portion of salient objects touching image boundaries. This problem was challenged in [47] in order to reduce misclassification of such regions. The handcrafted measure in (2.50) was proposed in order to measure the boundary connectivity of a segment.

$$BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}}. \tag{2.50}$$

$Len_{bnd}(p)$ is the length of the segment p across the image boundary and $Area(p)$ is the number of elements contained in p. $BndCon(p)$ is assumed to be low if the segment is touching the boundary. Since segmentation itself is a challenging task, length and area of each segment was approximated by an approximate method based on the geodesic distance of the regions. Final saliency was calculated by exploiting this measure together with local region contrast and smoothness enforcement within regions.

2.4.1.3 Saliency Detection via Absorbing Markov Chain

A graph-based model was proposed in [48] where image regions are considered as nodes of a graph in Markov chain model. The nodes corresponding to image regions on image boundaries are considered as absorbing nodes in the chain. The saliency measure is calculated as the absorption time of each node by the absorbing nodes. The more salient the region is, the later it will be absorbed due to its dissimilarity to the image boundary regions. The similarities between regions are encoded in the related transition matrix $P = D^{-1}W$ of the Markov chain. If the transition matrix is organized in such a way that it will be a block matrix as in (2.51) where Q and R are transition probabilities of transient nodes to each other and to absorbing nodes, respectively.

$$P \to \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}. \tag{2.51}$$

Then, the fundamental matrix of P can be defined as $N = (I - Q)^{-1}$ and the saliency can be estimated by the absorption time Nu where u is a seed vector of prior belief of saliency.

2.4.1.4 Saliency Tree

A salient object detection technique based on a tree model was proposed in [49]. First, the image of interest is over-segmented into regions, then by a combination of global contrast, spatial sparsity and objectness cues, initial estimations for saliency of regions are calculated. Next, a saliency tree is generated by region merging. Finally, salient regions in this tree are selected according to a local contrast based regional center-surround measure.

2.4.1.5 Discriminative Regional Feature Integration

A supervised salient object detection method was proposed in [50]. First, for each region in the image of interest, a regional contrast descriptor which is calculated by taking the distance between feature vectors of the region and its immediate surrounding is constructed. The feature vector corresponding to a region includes color properties in several color spaces and texture filter responses. Besides the region contrast descriptor, a region property descriptor was also proposed which includes appearance features related to the region. Finally, a backgroundness descriptor was proposed which is obtained by taking the distance between the region property descriptor with the

image boundary property descriptor. A random forest based regressor is trained, which maps the mentioned descriptors to saliency scores. Finally, a multi-level fuser is trained in order to combine saliency scores from multiple resolutions.

2.4.1.6 Visual Saliency Based on Multiscale Deep Features

A supervised salient object detection method based on deep learning was proposed in [51]. First, a hierarchical image segmentation method is applied to the image. Each of the image segments at multiple scales are then fed to a feature extraction module exploiting a pre-trained convolutional neural network. Next, two fully connected layers are learned in order to classify the obtained features as salient or not. In order to ensure spatial coherence, a post processing algorithm is applied on saliency maps from each scale. The saliency maps obtained from each level in hierarchical segmentation are then merged to obtain the final saliency map.

2.4.1.7 Saliency Detection by Multi-Context Deep Learning

A multi-context approach was adopted in [52] for salient object detection using deep learning. First, a superpixel extraction method is applied on the input image. Then, two separate convolutional neural networks that aim to extract local and global context were applied on the input images. Each superpixel is treated as a new sample and the input image was centered according to a superpixel for the global context network and another superpixel-centered, but close-up image was used for the local context network. The high-level features extracted from each network are then combined and fed to a fully-connected layer in order to predict the saliency score of the superpixel.

The above methods can be categorized according to the saliency cues and supporting techniques that they use while calculating saliency maps and according to their approach to the salient object detection problem. Such a categorization is made in Table 1. Note that center prior was not included in the categories since none of the state-of-the-art methods mentioned in this section make use of this cue. As can be observed from Table 1, all methods make use of the image abstraction technique, most methods exploit the boundary connectivity cue and almost half of the methods are graph-based approaches. It should be noted here that the categorization of the deep learning based methods [51], [52] are not as straightforward as the other methods. For example, it is obvious whether other methods make use of local and global contrast since their approaches are handcrafted, however deep learning based methods make use of these cues implicitly, which is not an easy task to directly show.

Table 1 Categorization of Salient Object Detection Techniques

	DSR[46]	RBD[47]	MC [48]	ST [49]	DRFI [50]	MDF [51]	MCD [52]
Local Contrast	No	Yes	No	Yes	Yes	Yes	Yes
Global Contrast	No	No	No	Yes	No	Yes	Yes
Boundary Connectivity	Yes	Yes	Yes	No	Yes	No	No
Image Abstraction	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multi Resolution	No	No	No	Yes	Yes	Yes	No
Smoothness	No	Yes	No	Yes	No	Yes	No
Graph Based	No	Yes	Yes	Yes	No	No	No
Probabilistic Approach	Yes	No	Yes	No	No	No	No
Supervised	No	No	No	No	Yes	Yes	Yes

2.4.2 Salient Object Detection Datasets

Many salient object detection datasets were proposed during the last decade. In this section a brief overview of the most popular datasets is given.

ASD [35] contains 1000 images from the MSRA [34] salient object dataset. The contribution of ASD is manual pixel-wise labeling of the ground truth, whereas in MSRA, the ground truth was shared as bounding boxes. ASD contains large salient objects with high local and global contrast. There is usually one salient object which tends to occur at image center. The dataset is relatively less challenging when compared to the rest of the datasets. MSRA10k [36] is also formed from the images in MSRA and extends the bounding box ground truths to pixel-wise ones. MSRA10k presents a slightly more challenging salient object detection task compared to ASD. Although the image characteristics and complexity of the salient objects are similar, it involves a higher number of classes. PASCAL1500 [37] contains 1500 images selected from the PASCAL-VOC 2012 [38] challenge. The images contained in this dataset contain usually more than one salient object per image. Moreover, the objects appear at a higher variety of size, shape and location. Hence, PASCAL1500 is more challenging compared to the MSRA datasets. SOD [39] is constructed from Berkeley seg-

mentation dataset (BSD300) [40] and involves 300 images. In BSD300, the boundaries related to the objects and other regions already exist. During the construction of SOD, 7 subjects were asked to choose the enclosed boundaries that correspond to the boundaries of salient objects. Then, closed filled regions were formed according to these boundaries, which form the salient object ground truths. The images include many object classes. Most of the images include high texture information and the salient object detection is not so straightforward as in some other datasets. DUT-OMRON [41] is one of the most challenging datasets involving 5168 images. The dataset involves images covering a high number of object classes, with cluttered background and usually more than one object per image. The dataset was collected from a higher number of images collected from the web and salient object ground truths were selected by contribution of 25 subjects in pixel-wise labeling. JUDD-A [42] is the most challenging dataset containing 900 real life images with much cluttered background and the highest spatial distribution and class variety of objects. The pixel wise salient object ground truths for this dataset were obtained by annotating the eye predictions obtained by [43]. The highest density in the eye predictions was assumed to fall on the most salient object. 2 subjects were responsible for pixel-wise ground truth labeling on the most salient object in the scene.

Exemplary images from each dataset are illustrated in Figure 9. For a more in-depth analysis of the datasets, one can refer to [45].

Figure 9 Exemplary images from ASD, MSRA10k, PASCAL1500, SOD, DUT-OMRON and JUDD-A datasets.

2.4.3 Performance Measures

A large number of measures have been defined for evaluating the performance of salient object detection methods. In this section, a brief overview of the popular metrics is given.

2.4.3.1 Precision-Recall (PR) Curve

In salient object detection, a sample image pixel is true positive if it is correctly classified as salient, false positive if it is incorrectly classified as salient, true negative if it is correctly classified as background and false negative if it is falsely classified as background. The precision value of a segment is the ratio of the number of true positives to the number of samples that are classified as positive. The recall value of a segment is the ratio of the number of true positives to the number of positives. Precision and recall are complementary evaluation metrics and are formalized as follows.

$$pre(\tau) = \frac{|g \cap s_\tau|}{|s_\tau|}, rec(\tau) = \frac{|g \cap s_\tau|}{|g|}. \tag{2.52}$$

It should be noted that salient object detection methods usually output a gray-level saliency confidence map s instead of a binary salient object segment. Hence, the saliency map is converted to a segment s_τ by thresholding it by τ. Then, precision-recall values are extracted for several thresholds by comparing s_τ with the ground truth g and a performance curve is extracted.

2.4.3.2 Receiver Operating Characteristics (ROC) Curve

Another popular salient object detection evaluation metric is Receiver Operating Characteristics (ROC) curve which investigates the true positive and false positive rates. True positive rate is yet another name for recall. False positive rate on the other hand is the ratio of the number of false positives to the total number of negatives and is formulated as follows.

$$fpr(\tau) = \frac{|\bar{g} \cap s_\tau|}{|\bar{g} \cap s_\tau| + |\bar{g} \cap \bar{s_\tau}|}. \tag{2.53}$$

ROC curve is extracted similarly to precision-recall curves by finding the true positive and false positive rates for different thresholds on the saliency maps.

2.4.3.3 F_β measure and Area under the ROC curve (AUC)

PR and ROC curves give performance evaluations in terms of complementary performance metrics. However, it is possible to represent these curves with single measures as well. These

measures, namely F_β measure for precision recall curves and AUC for ROC curves are defined as follows.

$$F_\beta(\tau) = (1+\beta^2)\frac{pre(\tau)rec(\tau)}{\beta^2 pre(\tau) + rec(\tau)}, \qquad AUC = \int rec(\tau)dfpr'(\tau). \tag{2.54}$$

F_β measure gives a threshold dependent evaluation where β is a parameter the choice of which gives more weight to either precision or recall. The maximum or the mean F_β measure throughout the PR curve is often used for evaluation of salient object methods with a single value. On the other hand, AUC gives a single threshold-free measure by integrating over the ROC curve.

2.4.3.4 Mean Squared Error (MSE)

Unlike PR and ROC curves, the Euclidean or Mean Squared Error (MSE) directly gives a measure of the accuracy of saliency map without thresholding it and is formulated as follows.

$$MSE = \frac{1}{|I|}\sum_{i \in I}(g_i - s_i)^2. \tag{2.55}$$

For the MSE measure both the ground truth and saliency map vectors are linearly normalized between 0 and 1. It can be observed from (2.55) that any saliency value that is not exactly equal to the ground truth is penalized. Thus, MSE differs from the thresholding based methods in this sense and acts as a complementary measure to them.

2.4.3.5 Mean Absolute Error (MAE)

Similar to the MSE, the Mean Absolute Error (MAE) is also a threshold-free evaluation metric and for min-max normalized ground truth and saliency map vectors, it is defined as follows.

$$MAE = \frac{1}{|I|}\sum_{i \in I}\|g_i - s_i\|. \tag{2.56}$$

MAE differs from MSE in the sense that the penalization is linear. This might be favorable especially in the regions that the ground truth is closer to zero. The error in these regions can be ignorable in terms of MSE due to the squaring, however for better precision they should not be.

2.4.3.6 Kullback-Leibler (KL) Divergence

The saliency map and the ground truth vectors can be treated as probability densities if they are l^2 normalized. Then, a measure of similarity between these vectors can be written as a measure of probability density divergence given by the Kullback-Leibler divergence as follows.

$$D_{KL}(\boldsymbol{g}||\boldsymbol{s}) = \sum_i g_i \log\left(\frac{g_i}{s_i}\right). \tag{2.57}$$

All the evaluation metrics defined above evaluates the saliency in a different point of view. Hence, for a complete comparison between salient object detection methods, a wide range of evaluation metrics should be used.

3 Contributions

In this chapter, the contributions related to this thesis will be briefly introduced. First, Schrödinger's Equation will be analyzed in detail and its direct applications to object segmentation problem will be discussed in Section 3.1. Next, the relation between Quantum Mechanics and Spectral Graph Theory will be explained by exploiting a graph-based analysis on Schrödinger's Equation in Section 3.2 and by exploiting this relation, a spectral foreground extraction method, Quantum-Cuts, (QCut) will be introduced. The optimization problem that QCut solves will be shown to contain saliency cues. Next, the improvements and extensions on QCut in this context will be explained briefly. Specifically, in Section 3.3, a method for learning to rank salient segments obtained by multispectral Quantum-Cuts will be introduced. In Section 3.4, several extensions on Quantum-Cuts related to the exploited graph structure and affinity matrix construction will be explained. Finally, in Section 3.5, a supervised approach for affinity matrix learning will be presented.

3.1 Salient Segment Proposals in Quantum Scale

Schrödinger's Equation given in (2.9) is formulated in the continuous domain. However, the problems of interest in Computer Vision are usually addressed in discrete domain due to the digital nature of the visual data. In order to draw links between both fields, one can write Schrödinger's equation in a discrete grid by using a discrete approximation of the Laplacian operator with finite difference method as follows.

$$\nabla^2 \boldsymbol{\psi}|_{\mathrm{i}} = \left(\sum_{\mathrm{j} \in \mathrm{N_i}} \psi_{\mathrm{j}} \right) - |\mathrm{N_i}| * \psi_{\mathrm{i}}. \tag{3.1}$$

In (3.1), $\mathrm{N_i}$ is the set of coordinates that are in an immediate neighborhood of the coordinate indicated by i. In such discrete representation, $\boldsymbol{\psi}$ is a finite length vector representing discrete wavefunction and the potential field can be represented with a vector $\boldsymbol{v}$ that is of same dimension with $\boldsymbol{\psi}$.

As an example, in the above representation, an image would correspond to a graph with a regular 2D grid structure. Due to the locality principle in physics [53], the neighborhood rule needs to be

defined as connectivity to immediate surrounding only. For images, 4-connected or 8-connected graphs might be utilized. In such a graph, a matrix representation of (3.1) is identical with the negative graph Laplacian matrix for an unweighted graph. Therefore, the discrete version of Schrödinger's Equation in matrix algebra can be written as the following eigenvalue problem.

$$\mathbf{H}\boldsymbol{\psi} = \mathrm{E}\boldsymbol{\psi}. \tag{3.2}$$

In (3.2), $\mathbf{H} = (\mathbf{L} + \mathbf{V})$, $\mathbf{L}$ is the graph Laplacian matrix of a weighted graph where all weights are equal to $\frac{\hbar^2}{2m}$ and $\mathbf{V}$ is a diagonal matrix which includes the values of the potential $\boldsymbol{v}$ on its diagonal. Therefore, solutions of the discrete Schrödinger's Equation can be simply obtained by applying eigen-analysis on $\mathbf{H}$.

From the discussion in Section 2.2, one can infer that the pure wavefunctions obtained by the eigenvalue problem in (3.2), correspond to critical points of the Rayleigh Quotient $(\boldsymbol{H}, \mathbf{z}) = \frac{\boldsymbol{z}^T \mathbf{H} \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}}$. Considering the probabilistic interpretation of Born, the intensity of the wavefunction is accepted to be a probability density function. Therefore, $\boldsymbol{z}^T \boldsymbol{z} = 1$. Hence the Rayleigh Quotient reduces to $R(\boldsymbol{H}, \mathbf{z}) = \boldsymbol{z}^T \mathbf{H} \boldsymbol{z}$. According to the third postulate of Quantum Mechanics, this gives the expected value of the observable that is represented by $\mathbf{H}$, which is the energy of the particle. Hence, one can conclude that pure wavefunctions occur at the critical points of expected particle energy. By Theorem 2-2, the minimum and maximum eigenvalues of $\mathbf{H}$ correspond to the minimum and maximum expected energies of a particle, and the other eigenvalues correspond to energies that are in a local optimality. The meaning of these critical points can be observed by rewriting $R(\boldsymbol{H}, \mathbf{z})$ as follows.

$$R(\boldsymbol{H}, \mathbf{z}) = \boldsymbol{z}^T \boldsymbol{L} \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{V} \boldsymbol{z} = \sum_i \sum_{\mathrm{j} \in \mathrm{N_i}} \frac{\hbar^2}{2m} \left(z_i - z_j\right)^2 + \sum_i {z_i}^2 v_i. \tag{3.3}$$

The first term in (3.3) is a weighted measure of spatial smoothness in elements of $\boldsymbol{z}$ and the second term gives the average potential value in the region that is represented by the wavefunction. The global minimum of $R(\boldsymbol{H}, \mathbf{z})$ is therefore expected to give a wavefunction ψ_1 which has a low rate of oscillation, which covers a region of low potential value. On the other hand, the global maximum of $R(\boldsymbol{H}, \mathbf{z})$ would give a wavefunction ψ_n that is of highest oscillation rate and that is present in a region of high potential value. Other local optima correspond to wavefunctions that span regions corresponding to a structured potential. For example, based on the discussion in Section 2.1, the bounded wavefunctions correspond to standing waves that are limited by the walls of a potential structure, such as the Coulomb potential of nucleus acting on electrons.

The discussion above has motivated our work in [P1] which investigates the wavefunctions corresponding to a potential that is defined on gray-level image intensities. Based on the above discussion, the wavefunctions are expected to span gray-level intensities of a certain structure which may correspond to an object or a region in an image. Thus, the thresholded probability densities corresponding to wavefunctions were considered as segments. In order to handle segmentation errors caused by the oscillations in wavefunctions, a closing operation was applied on the segments. Moreover, the regularization factor $\frac{\hbar^2}{2m}$ between the spatial smoothness and average intensity terms in (3.3) corresponds to a scaling measure. Thus, the above segment proposal procedure is utilized for values selected from an empirical range of $[0.2, 1.2]$ for this regularization factor. Smaller factors correspond to smaller scales. Some of the segments obtained in this way are given in Figure 10.

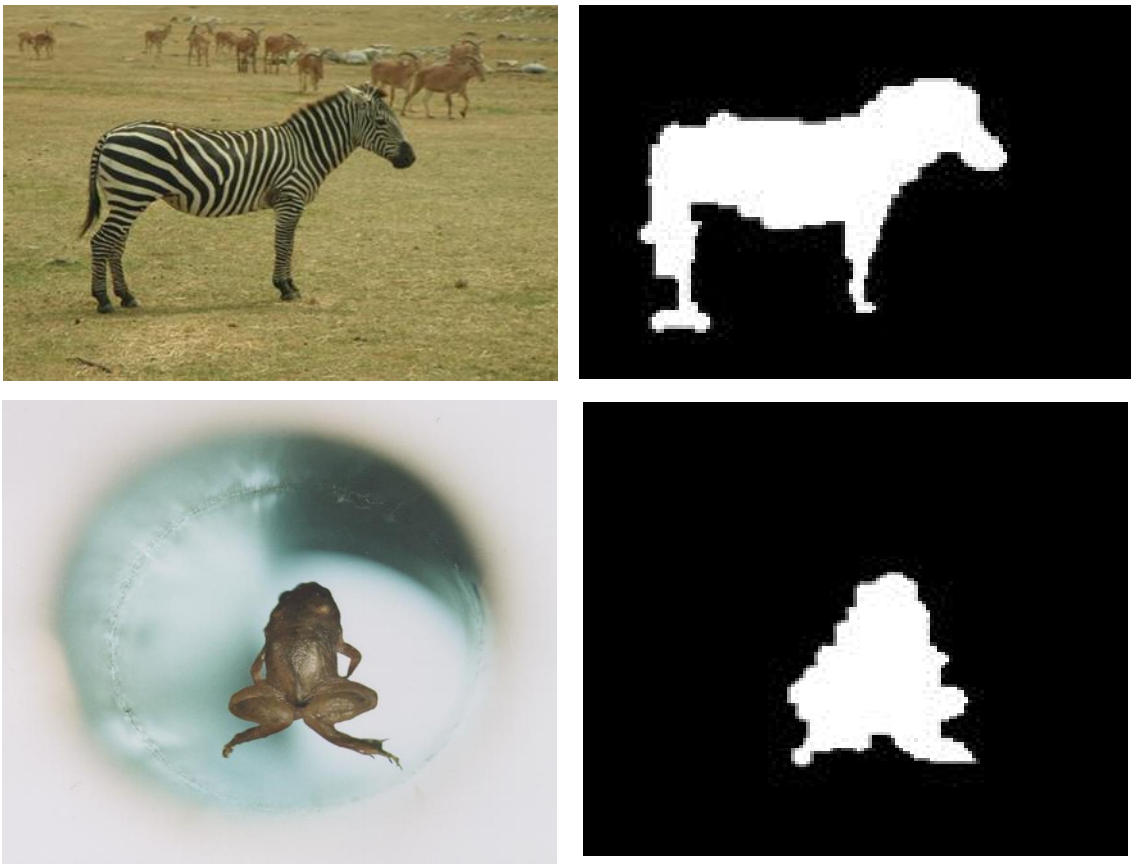


Figure 10 (Left) Images from Corel1000 dataset, (Right) a segment corresponding to a wavefunction.

The segments illustrated in Figure 10 show a potential to extract a salient object from images. However, how to select salient segments from the segment pool obtained by the above procedure is of question. For this purpose, the segments in the pool were sorted according to an objectness measure proposed in [P1] as follows.

$$O(S) = c|S \cap E| - (1 - c)\text{area}(S). \quad (3.4)$$

In (3.4), E is the output of an edge detection algorithm on the image, $|S \cap E|$ is the number of edge pixels covered by the segment S. The objectness measure $O(S)$ of a segment S is then defined as a tradeoff between the edge encapsulation and the area of the segment, where the tradeoff is controlled by the parameter c. Edge encapsulation measure follows from the assumption that an interesting object is expected to have a strong boundary and the area is discouraged in order to prevent bias to large segments including all boundaries in the image. The segments in the pool were ranked according to the measure in (3.4) and the top ranked ones were selected as salient segments.

To the best of our knowledge, [P1] is the first study that investigates object/region segment proposals in a realizable quantum mechanical setup. Such a realization in the quantum scale would be computationally instantaneous as the wavefunctions of a particle in a given potential are immediately formed upon a measurement. Thus, the computational complexity of such a physical system is only limited by the measurement process.

Computer simulations of such system on the other hand are computationally very heavy due to the eigenvector calculation on the Hamiltonian matrix which is of size $MKxMK$, where the corresponding image is MxK. A full eigendecomposition on H is infeasible, therefore only a limited number of eigenvectors could be calculated. Therefore, the optimality of the selection from the segment pool which relies on the primitive measure in (3.4), is limited by the segments in the pool that do not cover the entire eigenspectrum. Even with the limited number of proposals, a non-optimized MATLAB implementation of the algorithm takes around 2 minutes for a typical image of size 75x100 on a PC with a IntelCore i7-3632QM CPU @2.20 GHz with 8 GB RAM. Finally, the corresponding graph representation based on Schrödinger's Equation is only based on a restricted neighborhood rule and completely omits the pairwise relations between image pixels, but relies on a unary potential only. Such a graph results into segments that are noisy and cannot fully recover object edges. In fact, it is pointed out in [P1] that due to the above limitations, the best segments in the pool can only reach an average precision of 0.58 and recall of 0.75 on a dataset constructed from images in PASCAL VOC-2012 [38].

Such disadvantages of [P1] can be overcome by combining the unique concepts of Quantum Mechanics, such as the potential and the probabilistic interpretation of the wavefunction and the concepts in spectral graph partitioning methods.

3.2 Quantum-Cuts

3.2.1 A Spectral Graph Based Analysis of Schrödinger's Equation

There have been already some connections drawn to Spectral Graph Theory in Section 3.1 by pointing out that the quantum wavefunctions are indeed critical points of the Rayleigh Quotient $R(\boldsymbol{H}, \mathbf{z})$. However, the meaning of the wavefunctions obtained by Schrödinger's Equation in a spectral graph partitioning point of view remains unclear. To reveal this point, we turn to spectral graph partitioning methods that use the Rayleigh Quotient and **Theorem 2-2** to approximate solutions for intractable graph partitioning problems. For such problems, one need a labeling vector as an indicator for the labels of the subgraphs to be partitioned. Examples were given in (2.23), (2.26) and (2.33). One can similarly define a labeling vector as follows.

$$x_i = \begin{cases} \frac{1}{|P|}, i \in P \\ 0, i \in \bar{P} \end{cases}. \tag{3.5}$$

This definition follows from the probabilistic interpretation of the wavefunction. If $\boldsymbol{x}$ would be thought as a probability density function, it would correspond to a uniform distribution on the space represented by the partition P. Furthermore, one can define another vector $\boldsymbol{z}$ as follows.

$$z_i = \begin{cases} \frac{\mp 1}{\sqrt{|P|}}, i \in P \\ 0, i \in \bar{P} \end{cases}. \tag{3.6}$$

In other words, $\boldsymbol{x} = \boldsymbol{z} \circ \boldsymbol{z}$ where $\circ$ denotes the Hadamard product. In the above definitions, we have defined vectors $\boldsymbol{z}$ and $\boldsymbol{x}$ analogous to $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^* \circ \boldsymbol{\psi}$ respectively in a graph partitioning context. Next, we can infer the meaning of the Rayleigh Quotient in terms of graph partitioning by investigating it with the new vector definitions.

$$R(\boldsymbol{H}, \mathbf{z}) = \boldsymbol{z}^T \mathbf{H} \boldsymbol{z} = \boldsymbol{z}^T \mathbf{L} \boldsymbol{z} + \boldsymbol{z}^T \mathbf{V} \boldsymbol{z}. \tag{3.7}$$

The second term in (3.7), $\boldsymbol{z}^T \mathbf{V} \boldsymbol{z}$**,** is the average potential value in the region P represented by the vector $\boldsymbol{z}$**.** The meaning of the first term on the other hand is not so straightforward and can be written as follows.

$$\boldsymbol{z}^T\boldsymbol{L}\boldsymbol{z} = \sum_{j,i} w_{j,i}\left({z_j}^2 - z_j z_i\right) = \sum_{j,i} w_{j,i}\left({z_j}^2 - |P|{z_i}^2{z_j}^2\right) + \sum_{j,i} w_{j,i}\left(|P|{z_i}^2{z_j}^2 - z_i z_j\right). \tag{3.8}$$

Remember that the graph weight $w_{i,j}$ between the node i and j is $\frac{\hbar^2}{2m}$ in a neighborhood rule defined in Section 3.1. Then, the first term in (3.8) can be easily seen to correspond to the $\frac{cut(P,\bar{P})}{area(P)}$. The second term is a weighted average of the oscillations within the foreground region P. An oscillation is defined as a sign change in the vector $\boldsymbol{z}$. Therefore, this term was named as phase term in [P2]. Based on the above discussion, one can write $R(\boldsymbol{H},\mathbf{z})$ as follows.

$$R(\boldsymbol{H},\mathbf{z}) = \frac{cut(P,\bar{P})}{area(P)} + phase(P) + \mu_v. \tag{3.9}$$

In (3.9), $phase(P) = \sum_{j,i} w_{j,i}\left(|P|{z_i}^2{z_j}^2 - z_i z_j\right)$ and $\mu_v = \boldsymbol{z}^T\mathbf{V}\mathbf{z}$**.** If the global minimum of the criterion in (3.9) would be considered, one can easily omit the phase term, since this term can easily be set to zero by changing the signs of $\boldsymbol{z}$ which does not affect the actual labeling vector $\boldsymbol{x}$ in any way [P2]. Furthermore, one can think of an artificial sink node S and formulate the unary potential values as weights of connections of each node to S. In such a model, the minimization of (3.9) reduces to the following.

$$\min_{\mathbf{z}} R(\boldsymbol{H},\mathbf{z}) = \min_{\mathbf{z}} \frac{cut(P,\bar{P}+S)}{area(P)}. \tag{3.10}$$

It should be noted that the cut in (3.10) corresponds to the cut in the new graph formed by the artificial sink node S. Moreover, since all the weights in the graph related to Schrödinger's Equation is defined as equal, this cut value merely represents a weighted value of the perimeter of the region that is cut from the graph.

Based on the discussion above, one can conclude from (3.10) that the minimization of $R(\boldsymbol{H},\mathbf{z})$ corresponds to large circular regions in a low potential and a spectral approximation of segmentation of such regions is given by the ground-state of Schrödinger's Equation.

3.2.2 Quantum-Cuts

From an object segmentation point of view, the conclusion of Section 3.2.1 does not make much sense, especially if the potential is selected to correspond to gray-level intensities as in [P1]. However, by combining this result with traditional graph construction approaches with different potential

definition, one can obtain a unique optimization that results into a very effective salient object detection method.

As discussed in Section 3.1, due to the structure of Schrödinger's Equation, one disadvantage of the method in [P1] is that it omits pairwise relations in the graph. To overcome this drawback, in [P2], a weighted graph was considered, where the weights correspond to affinities between nodes following the common approach in applications of spectral graph partitioning to Computer Vision [10]. Furthermore, the sink node S was thought as a representative node for the background. The graph weights of the connections to this node, i.e. the potential then represent a prior knowledge related to the background.

Then, the minimization of $R(\boldsymbol{H}, \mathbf{z})$ corresponds to regions which are in high local and global contrast with their surroundings, and have large area and low prior as background. Such regions are highly compatible with the definition of a salient object. Considering the saliency cues that are introduced in Section 2.4.1, high contrast cue is covered by the small cut value. Furthermore, if a high background prior is assigned to the image boundaries, the boundary connectivity cue could also be covered. Then, the approximation of the minimization of $R(\boldsymbol{H}, \mathbf{z})$ in such graph and potential definitions is expected to give good saliency maps.

The method explained above is called Quantum-Cuts [P2] and can be formulated as follows.

$$\boldsymbol{H}\boldsymbol{z}_* = E_m \boldsymbol{z}_*, \tag{3.11}$$

$$\boldsymbol{y}_* = \boldsymbol{z}_* \circ \boldsymbol{z}_*.$$

In (3.11), $\boldsymbol{H} = \boldsymbol{L} + \boldsymbol{V}$ where $\boldsymbol{L}$ is a graph Laplacian of a weighted graph constructed using image pixels as nodes and pairwise color affinities as graph weights. $\boldsymbol{V}$ is a diagonal matrix formed by high values on the image boundary pixels and zeros elsewhere. $\boldsymbol{z}_*$ is the eigenvector corresponding to the minimum eigenvalue E_m of $\boldsymbol{H}$. The saliency map is then obtained by $\boldsymbol{y}_*$.

QCut method, originally presented in [P2], forms a graph out of image pixels with a 4-connected neighborhood rule, then calculates the affinities according to the RGB color affinities between pixels. The potential is set very high on image boundaries and zero for the rest of the image. The saliency maps were obtained by saliency calculation in (3.11) on this graph. The graph partitioning was utilized by thresholding the saliency maps.

3.2.3 Quantum-Cuts vs. Spectral Graph Partitioning Methods

QCut method differentiates from the other spectral graph partitioning methods in the following points:

a. Unlike Ratio Cut and Normalized Cuts, QCut optimizes a measure related to the foreground only. Therefore, it is more suitable for applications such as salient object extraction.
b. Average Association also optimizes a measure related to the foreground only, however the optimization criterion is suboptimal in terms of salient object detection as it is inclined to find very small regions with high uniformity. On the other hand, QCut favors large areas with strong boundaries, which are much less likely to correspond to noise.
c. Unlike Ratio Cut and Normalized Cuts, QCut can assign prior on one of the partitions (the background). This is due to the labeling vectors that are designed accordingly to the probabilistic interpretation of wavefunctions.
d. Average Association method can also assign a prior on one of the partitions, however on the foreground region. QCut is more favorable in this sense, because it is easier to define background priors than foreground ones. For example, the boundary connectivity cue in salient object detection simply assumes background regions on the image boundary.

The above distinctions of QCut are mainly due to two key properties that are inspired from Quantum Mechanics:

a. The selection of labeling vectors which are inspired by the probabilistic interpretation of the wavefunctions in Quantum Mechanics.
b. The notion of potential field in Quantum Mechanics applied as a background prior.

In Figure 11, a visual comparison of the spectral graph partitioning algorithms and QCut is illustrated. The graph partitioning results in Figure 11 were obtained as follows. For every method, the same graph construction was utilized. For average association and Quantum-Cuts, the labeling vectors were thresholded by the Otsu method [54]. As for the Normalized Cuts and Ratio Cut, the labeling vectors were thresholded with 0. As one can observe, the maximum average association results into small noisy regions that have high association within. Normalized Cuts and Ratio Cut both cut the image from strong boundaries; however, resulting segments are not meaningful in terms of being a salient object. The partitioning obtained by QCut corresponds to salient objects with high contrast, large area and dissimilar from the image boundaries as it is expected from the optimization criterion in (3.10).

Computational Complexity and Performance

In [P2], it is pointed out that, for a typical image of size 200x150, a non-optimized simulation of the algorithm in MATLAB-R2013a, in an 8-GB RAM, Intel i7-2,20 GHz computer with 64-bit OS, takes about 1.5 seconds per image on average. The salient segment obtained by thresholding the QCut saliency map at the threshold value that corresponds to the equal error rate point of the precision-recall curve, yields an average performance of 0.8614 precision and 0.8642 recall on ASD dataset.

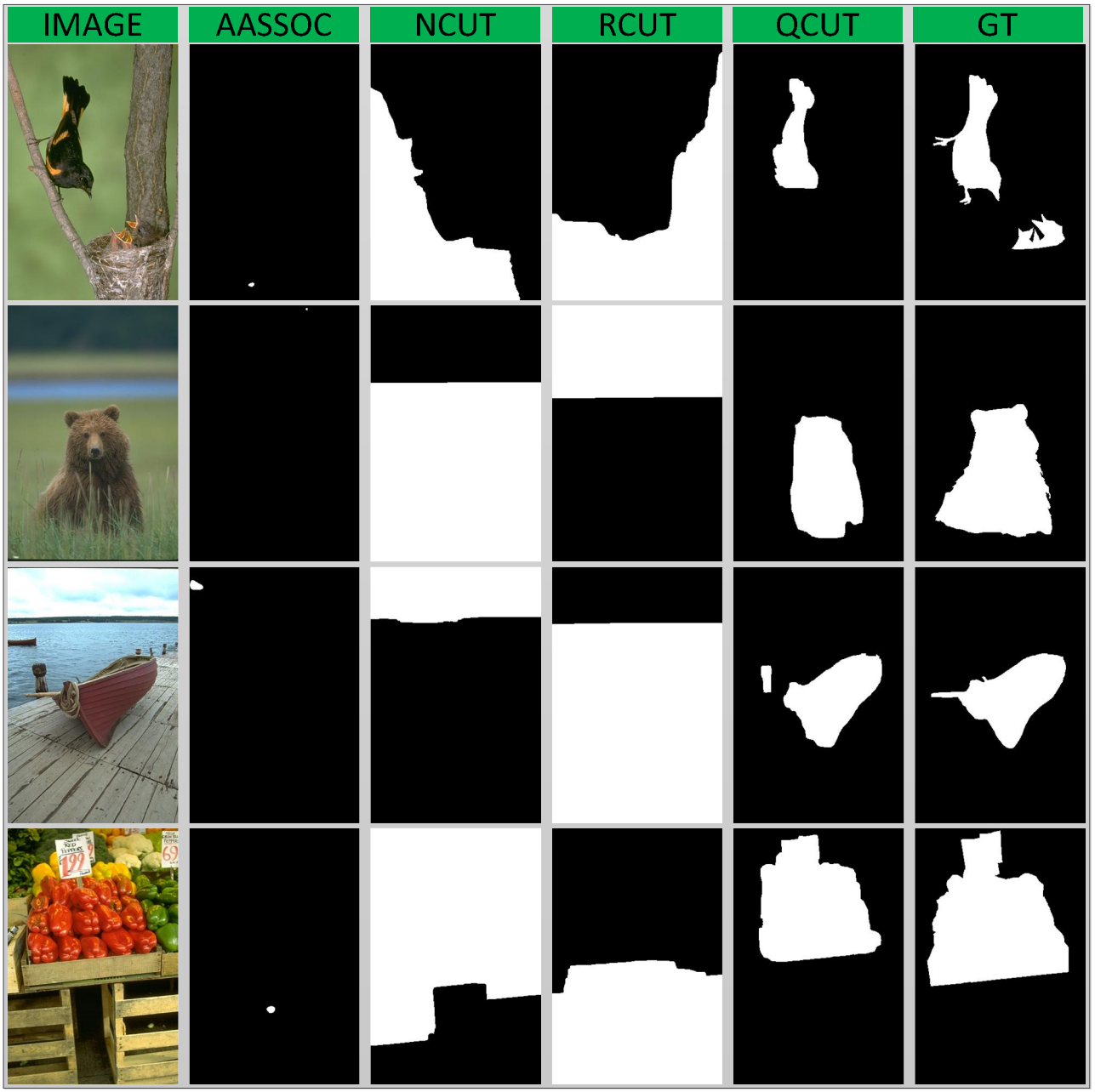


Figure 11 Images from SOD dataset, graph partitioning obtained by Average Association (AAS-SOC), Normalized Cut (NCUT), Ratio Cut (RCUT) and Quantum Cut (QCUT) and the related ground truth (From left to right).,

3.3 Multispectral Quantum-Cuts

In the previous section, we explained that the minimization of the objective function in (3.10) results into segments that are meaningful in terms of corresponding to salient objects in the image. However, besides the global minimum obtained by (3.11), other local minima are also observed to correspond to good segments. These local minima can be found by exploiting a larger spectrum of H. The segments are obtained in a similar fashion, i.e. each eigenvector is converted to a probability density function by taking the power of it and segments can be obtained by an adaptive thresholding method. In Figure 12, segments obtained by exploiting three eigenvectors corresponding to the smallest three eigenvalues for some images are shared.

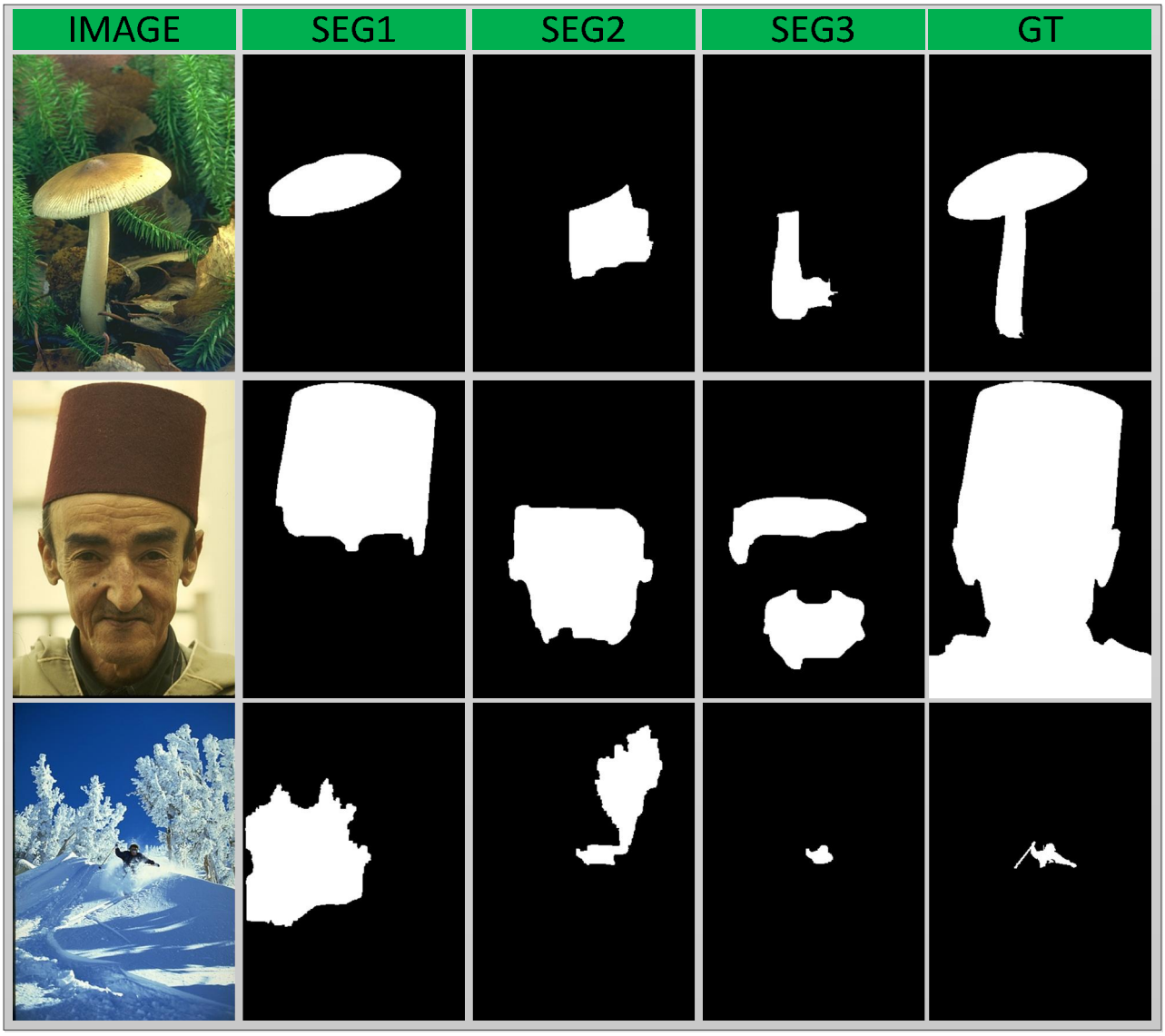


Figure 12 Images from SOD dataset, salient segments obtained by the first three eigenvectors of H matrix and the ground truth. (From left to right).

For the examples in Figure 12, the segment obtained by the first eigenvector may sometimes correspond to a portion of the salient object, or it can be a noisy segment. On the other hand, the objects or part of objects can be obtained by the segments related to the eigenvectors with higher eigenvalues. A selection from or combination of these segments would result into a better representation of the salient object. For example, in the first image the combination of first and third segments and in the second image the combination of first and second segment can give a good salient object segment. For the third image, one should discard the first and second segments and select the third segment.

A procedure for such selection/combination method was proposed in [P3] within a learning to rank segments strategy. For this purpose, first the saliency maps corresponding to the first 5 eigenvectors of H were taken. To obtain segments, the naïve thresholding methodology that is adopted in [P2] is not a robust approach; hence in [P3] the saliency maps obtained by multispectral QCut were used as prior seeds for an automatic foreground extraction method in [55]. Such a post-processing stage does not only provide an adaptive thresholding, but also takes the global contrast measure into account for foreground region extraction. The above segmentation procedure was utilized for the first 5 eigenvectors and the connected components in the segmentation masks were collected in a pool of segments.

From each segment in the pool a set of features related to low and mid-level appearance and shape cues were extracted. Then, a random forest regressor was trained that learns to predict the overlap value of each segment with the ground truth. The overlap value of a segment is defined as follows.

$$overlap(s, g) = \frac{|s \cap g|}{|s \cup g|}. \tag{3.12}$$

For a test image, the segments in the pool were ranked according to their predicted overlap values by the learnt random forest regressor. The segments having larger than 80% of the predicted overlap value of the highest rank segment were merged in order to obtain the final salient object segmentation mask. Such a learning to rank procedure results into a significant increase in the salient object detection performance.

In Figure 13, exemplary images from ECSD dataset, salient segments obtained by multispectral QCut and the segment obtained by the method in [P3] are illustrated. The approach in [P3] can successfully select and combine connected components that have the highest possibly to belong to a salient object.

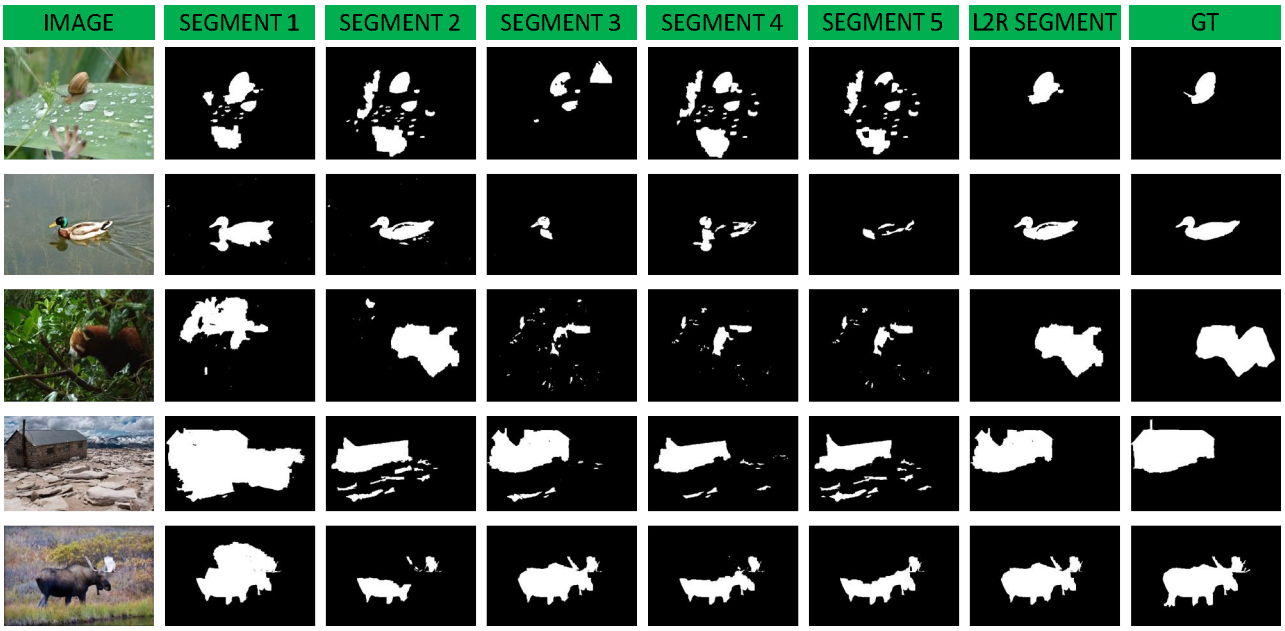


Figure 13 Images from ECSD dataset, 5 salient segments obtained by multispectral QCut, the segment obtained by [P3] and the Ground Truth (from left to right).

In [P3], experiments were conducted on the effect of the dataset that is used in the training stage. It was found out that training in a more complex dataset such as DUT-OMRON results in a random forest regressor that is more capable of handling various saliency cases. This is also evident from the feature importance analysis that is conducted in [P3]. It is observed that training with the relatively easier dataset MSRA10k focuses on simple features such as area and minor axis length, whereas the model trained by DUT-OMRON results in a rather distributed importance among all features.

During the computation of saliency maps, the method in [P3] does not directly utilize the QCut method in [P2] but employs several improvements such as a superpixel abstraction and multi-resolution approach. An extended version of QCut, including these along with several other improvements was proposed in [P4] and will be briefly covered in the Section 3.4.

Computational Complexity and Performance

The C implementation on a 64-bit PC with Intel i7-3740QM processor at 2.70 GHz yields an average running time of 0.42 seconds for the extraction of 5 saliency maps, on a typical image size of 400x300. The post-processing of each saliency map takes around 1.17 seconds and the feature extraction per saliency map takes 7.74 seconds. In total the whole algorithm takes around 45 seconds. On PASCAL1500 dataset, the saliency map with the smallest eigenvalue achieves a maximum 0.7514 F1 score, the binary segment obtained by post-processing on this saliency map achieves a 0.7846 F1 score, finally the segment selected by the learning-to-rank scheme achieves an F1 score of 0.8014. More extensive experimental results are given in [P4].

3.4 Extended Quantum-Cuts

3.4.1 Superpixel Abstraction

The graph construction in [P2] was based on representing the image pixels by graph nodes. Such a graph construction is computationally not efficient, especially considering the eigenvector calculation on the related H matrix. For a M-by-N image, the corresponding H matrix is MN-by-MN. Performing eigenanalysis on such a matrix is computationally heavy even for images of average size. Furthermore, pixel-based representation is very susceptible to noise. In order to overcome such issues, a superpixel abstraction of the image was proposed in [P4]. By a simple linear iterative clustering method, the image is over-segmented into superpixels which provides a close to regular grid representation of the image while preserving the object and region boundaries. The graph representation of the image with superpixel regions as nodes results into a significantly fewer number of nodes when compared to the pixel-level graph. Furthermore, for each node, the representing feature was selected as the average color values in the corresponding superpixel. This reduces the pixel-level color noise hence provides more robust feature values per region.

3.4.2 A Multi-resolution Approach

The superpixel abstraction extension also makes a multi-resolution approach possible. The superpixel representation of the graph can be made at several granularities each of which corresponds to the abstraction of the image in a different resolution. In [P4], three different superpixel resolutions with average number of superpixels 300, 600 and 1200 were utilized and the resulting saliency maps from each resolution were merged.

3.4.3 Increased Graph Neighborhood

The graph connectivity rule in [P2] was based on a 4-connectivity rule for pixels. This can easily be extended to superpixels by defining a neighborhood rule that defines two superpixels that share a boundary as connected in the representative graph. However, a cut value in a graph obtained in this way would only represent the foreground segment in terms of its dissimilarity from its surroundings in the segment border. Therefore, in [P4], a higher order connectivity rule was introduced in order to represent a more sophisticated local and global contrast value of the cut segment to the rest of the image. Due to the varying spatial distance of the connected nodes, the connections should contribute to the cut value accordingly. Hence, neighbors of a node were categorized into sets. An illustration of the sets is given in Figure 14.

Figure 14 An image from MSRA1000 dataset, boundaries of a superpixel, its 1st, 2nd,3rd,4th and 5th neighbor sets shown in white, blue, yellow, cyan and magenta colors respectively. Other superpixel boundaries are shown in dark red color.

As observed from Figure 14, the neighboring set definitions made in [P4] encodes geometrical properties of superpixels whereas a spatial distance based categorization of neighborhoods would not include such information.

3.4.4 A Novel Affinity Assignment

The neighboring set definitions made in the previous subsection indicates the degree of the connections that are to be formed between nodes in the graph. For example, the 5th set of neighbors of a node should be relatively loosely connected when compared to the ones in 1st set of neighborhood of it. Such a connection procedure was utilized in [P4] by normalizing the color affinities between nodes with a measure related to the number of nodes in the set that defines the degree of neighborhood between nodes.

The affinity assignment in [P4] is asymmetric; hence it allows controlling the cost of a cut link according to the labeling regarding the nodes that are related with this link. In other words, a cut cost is defined as w_{ji} if the node j is to be labeled as foreground and i as background and it is defined as w_{ij} if the labeling is the other way around. This affinity assignment results into an asymmetric H matrix. The result of Theorem 2-2 is only valid for symmetric matrices; however it gives a good solution in our special case. The analysis is made in the Appendix.

Besides the extensions that are briefly introduced here, [P4] also makes use of the post processing and multispectral approaches proposed in [P3]. These extensions result in a significant perfor-

mance improvement on top of QCut. Some visual illustrations of the improvement of EQCut on QCut are provided in Figure 15. As can be observed, the saliency maps obtained by EQCut are more accurate than the ones obtained by QCut in terms of both highlighting more regions from the salient objects and reducing the false positives.

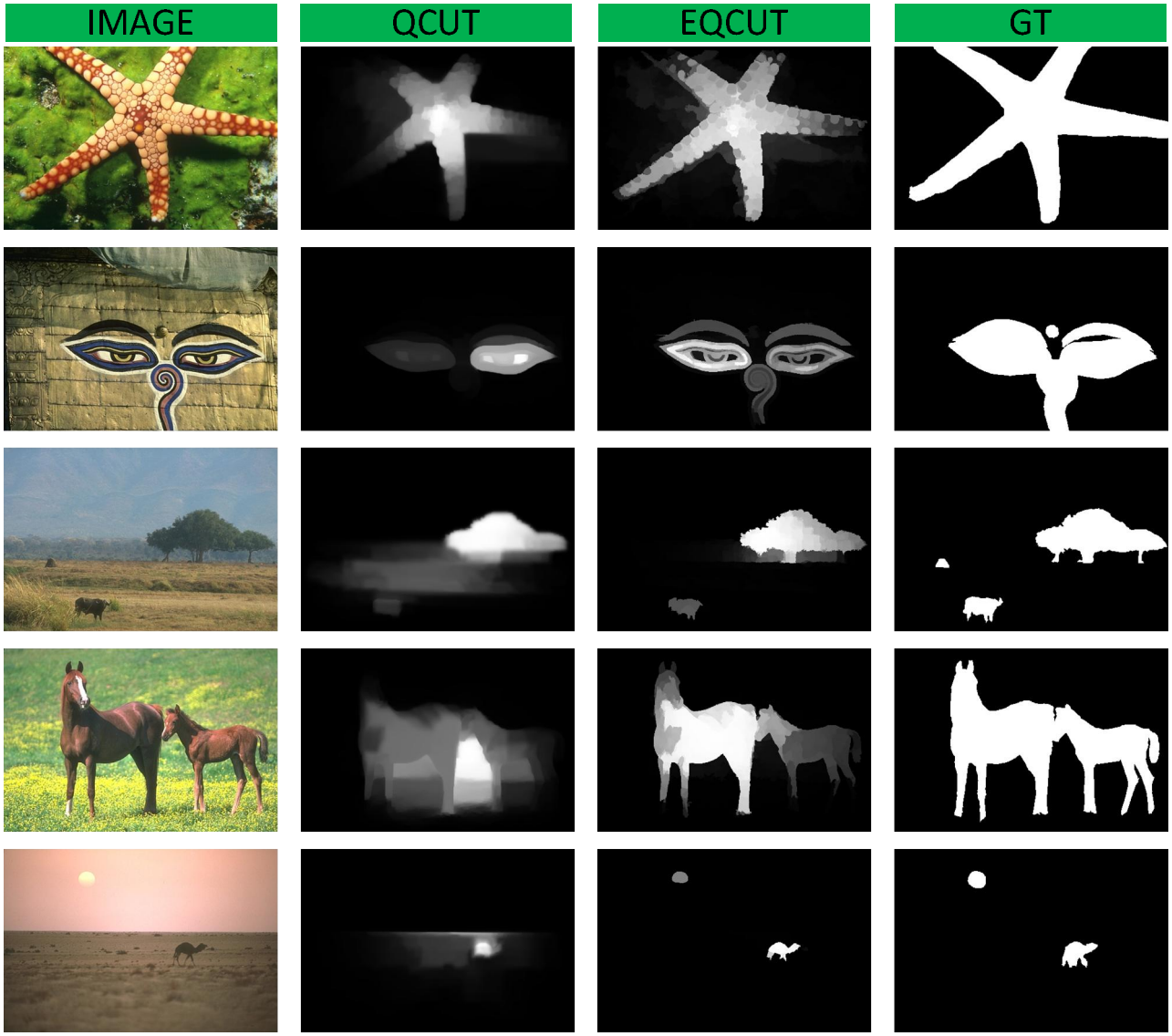


Figure 15 Images from SOD dataset, saliency maps obtained by QCut and EQCut and the related ground truths.

Computational Complexity and Performance

The C implementation of EQcut on a 64-bit PC with Intel i7-3740QM processor at 2.70 GHz yields an average running time of 0.852 seconds for the extraction of the EQCut saliency map.

As an unsupervised method, EQCut has been found to outperform many other salient object detection methods [P4]. Moreover, among 42 salient object detection methods, EQCut was enlisted within five best performing salient object detection methods in a comprehensive benchmark conducted recently [45]. As a representative example, EQCut achieves a maximum F1 measure of 0.683 on DUT-OMRON dataset whereas the closest competitor DRFI achieves a maximum F1 score of 0.665. More extensive experiments and comparisons are present in [P4] and [45].

3.5 Learning Graph Affinities for Quantum-Cuts

The main factor that provides the performance improvement achieved by EQCut over QCut is the selection of the hand-crafted affinity matrix, which is expected to encode the salient object detection task better than the basic one in [P2]. An affinity matrix possesses all the information related to the graph, i.e. the connection rules and the weighting strategy. As explained in Section 2.3, these are the most crucial steps in designing a graph based algorithm. Hence, in cases where it is possible, it might be favorable to use a supervised approach while constructing the affinity matrix. Moreover, such an approach provides automatic setting of some algorithm parameters. For this task, we have employed two different approaches. Next, these approaches will be described briefly.

3.5.1 Linearly Combined Affinity Graphs

Due to the fact that the saliency error obtained by (3.11) is differentiable, one can use a backpropagation based learning scheme to learn the affinity matrix. For this purpose, first, in [P5], we have proposed a system to learn to linearly combine several affinity calculation functions. In the original EQCut work, the affinities were calculated from raw color features by a specific affinity calculation function that is inversely proportional to the color distances. In [P5], we have instead exploited several affinity calculation functions by linearly combining them. The linear coefficients related to each affinity calculation function were learnt by a stochastic gradient descent based method via backpropagating the salient object detection error of EQCut. In this work, we have chosen Mean Average Error (MAE) as the loss function and maximum F1 measure as the objective function. As a major finding, it has been observed that the learnt affinity functions lead to exponentially increasing affinities for distances that are very small, to exponentially decreasing affinities for distances that are very high and to a linear-like relation in between. This learnt function is observed to lead to better saliency maps than using a steeply descending affinity function with increasing color distance as in [P4].

Computational Complexity and Performance

The learnt model was observed to improve the performance of baseline EQCut up to 3.1% on maximum F1 score. In particular, on JUDD dataset, the supervised method proposed in [P5] achieves

a maximum F1 score of 0.522 whereas EQCut in [P4] can achieve 0.5061. The MATLAB implementation of the method in [P5] on a 64-bit PC with Intel i7-3740QM processor at 2.70 GHz yields an average running time of 1.213 seconds for a typical image size of 400x300.

3.5.2 Feature Transformation

The study conducted in [P5] is limited to learning to convert color distances to affinities by linearly combining some affinity calculating functions. In [P6], we take another approach to learning the exploited affinity matrix via fixing the affinity calculating function to the one used in [P4] and learning the features to be used. To this end, we have made use of a modified version of Convolutional Kernel Networks (CKNs) explained in Section 2.3.1 in order to learn a feature transformation. The affinity matrix is then generated by applying the fixed affinity calculation function on these transformed features. The CKN block was concatenated with the global inference layer, EQCut which applies (3.11) on the affinity matrix obtained by the CKN block in order to predict saliency maps.

More specifically, the 1-layer CKN structure that is exploited in [P6] is designed to convert Lab color vectors to a new color space; such that the pairwise affinities calculated in this color space will encode salient object detection task better. This is ensured by updating the CKN parameters by backpropagating the error obtained by the subsequent EQCut block. The error is measured in terms of two loss functions: Mean Squared Error (MSE) and Kullback-Leibler Divergence (KLD) measures both of which are explained in Section 2.4.2. A different CKN was trained for each loss function.

The major findings of the work in [P6] can be listed as follows.

a. EQCut with the color space learnt with the unsupervised CKN (UCKN) performs better than EQCut in Lab color space in both performance metrics.
b. EQCut with the color space learnt with the supervised CKN (SCKN) proposed in [P6] performs better than UCKN for the learnt performance metric.
c. As the number of filters used in UCKN increases, the performance tends to increase.
d. The performance of SCKN is observed to be highest for a specific number of filters.

These findings can be explained as follows. The Gaussian kernel approximation performed in UCKN converts Lab color space to a better one in terms producing affinities that work better with EQCut in salient object detection task. As expected, an increase in the number of filters that are used in UCKN leads to better approximation of the Gaussian kernel, hence a performance improvement in salient object detection. As expected, the supervised approach SCKN produces better saliency maps than UCKN, as it is specifically trained to reduce the salient object detection error. Unlike in UCKN, a higher number of filters do not necessarily correspond to a better performance for SCKN since this also results into a model that is more difficult to learn. Hence a specific number of filters are observed to give the best performance.

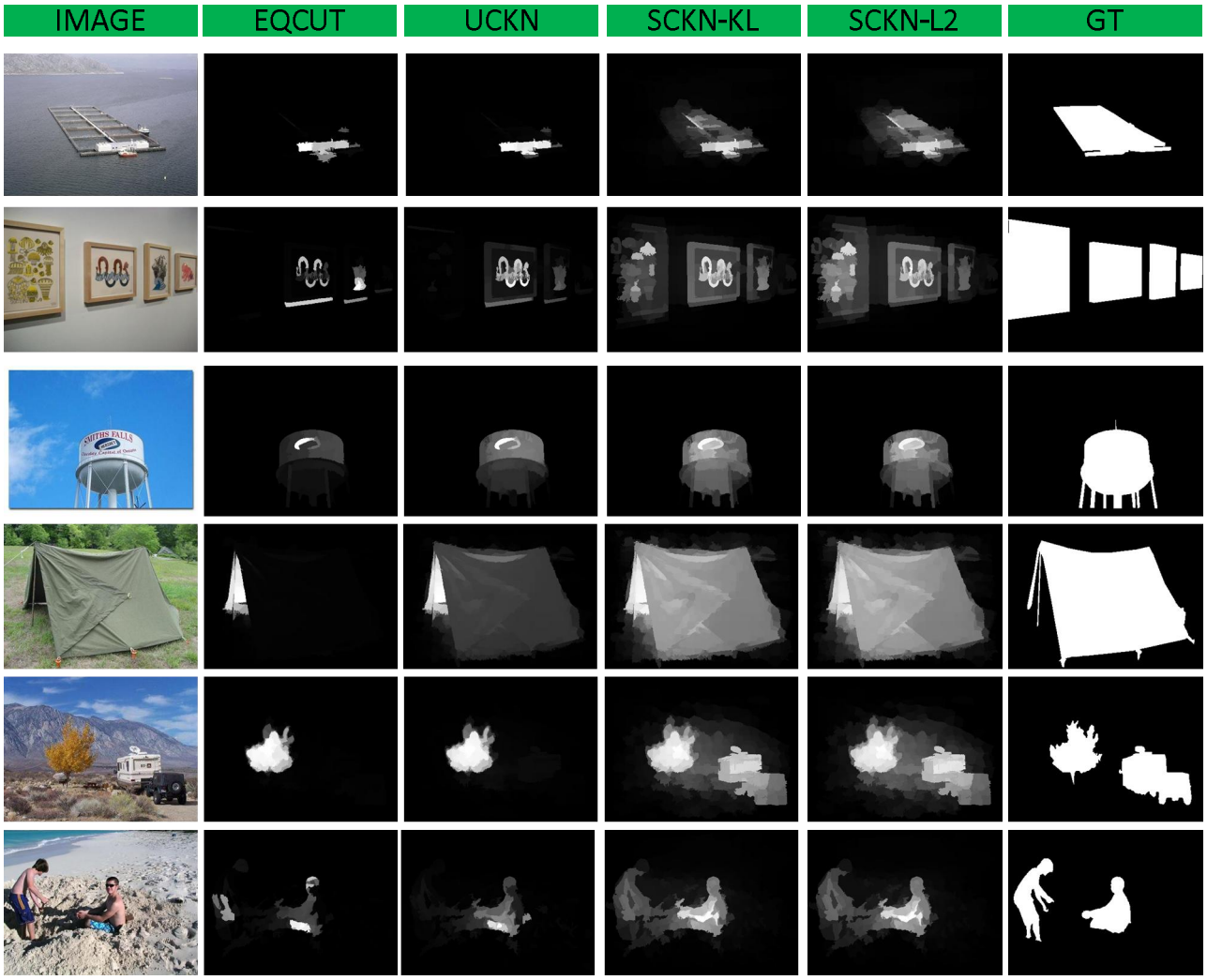


Figure 16 Images from DUTOMRON dataset, saliency maps obtained by EQCut, EQCut with unsupervised CKN (UCKN), EQCut with supervised CKN trained with Kullback-Leibler divergence loss (SCKN-KL) and Mean Squared Error (SCKN-MSE), and the related ground truths.

In Figure 16, the saliency maps obtained by EQCut, UCKN and SCKN for images from DUT-OMRON dataset are illustrated. UCKN is observed to highlight more salient regions with more uniformity within when compared to EQCut. The saliency maps obtained by SCKN trained on both loss functions are observed to result in both high recall and high uniformity within the salient object. This is expected since both KLD and MSE highly penalize the saliency maps that are not uniform within the salient object. SCKN trained with MSE is observed to result into regions that are slightly noisier than the one trained with KLD. This is due to the nonlinear penalization obtained by the MSE in the regions that have zero saliency which is explained in Section 2.4.3.5.

Computational Complexity and Performance

The MATLAB implementation of SCKN/UCKN on a 64-bit PC with Intel i7-3740QM processor at 2.70 GHz yields an average running time of 1.213 seconds for a typical image size of 400x300. The SCKN and UCKN approaches yields an average MSE error of $1.626\text{x}10^{-4}$ and $1.662\text{x}10^{-4}$ whereas EQCut obtains an MSE error of $2.496\text{x}10^{-4}$. More detailed experiments and comparisons are provided in [P6].

4 Conclusions

The main contributions in this thesis, present a unique interdisciplinary research in the sense that they link two major fields; Quantum Mechanics and Spectral Graph Theory. A deep investigation and due analysis of this link resulted into such novel algorithms that achieved the state-of-the-art performance in the salient object detection problem. In other words, the foundations of Quantum Mechanics pave the way to an automatic, spectral graph based foreground segmentation method - Quantum-Cuts (QCut) that can efficiently address the salient object detection problem.

The prior study of segment proposals in quantum scale in [P1] is not only inspired from Quantum Mechanics, but it applies the Quantum Mechanics principles directly. The contribution of this study is unique in the sense that such a system is physically realizable, i.e. image segment proposals can be produced in a physical system at quantum scale. Besides, the computational complexity is only limited by the measurement cost, since wavefunctions already naturally occur once the potential is defined, which motivates future studies for constructing such a system. However, due to the limitations of a computer simulation of this system, a practical simulation of the method in [P1] in a PC is computationally heavy. A typical image of size 75x100 takes around 2 minutes to process, which is much higher compared to many salient object detection methods. Moreover, the performance of this method is limited to the ability of Schrödinger's Equation applied on raw gray-level images, and is around 0.65 maximum F1 measure at best in PASCAL saliency dataset, which is lower than many state-of-the-art salient object detection methods. Hence, although the method in [P1] constitutes a basis for the important research problem of salient object detection in quantum scale, its simulation in a PC brings computational and performance drawbacks.

These drawbacks are addressed in [P2]. First, a preliminary analysis of Schrödinger's Equation discovers some clues about why it would produce good image segments. However, an in-depth graph theoretical analysis concludes that it is limited as it only uses unary potentials. The well-studied concepts in Spectral Graph Theory when integrated into the Schrödinger's Equation, combine the strengths of both fields and result into the proposed method QCut, which approximates the global optimal solution of a unique graph optimization problem that formulates foreground segmentation from graphs. A striking discovery is that, this optimization problem inherently includes well-known saliency cues such as local contrast and boundary connectivity, hence QCut is discovered to work successfully in salient object detection problem in the sense of performing superior to many state-of-the-art salient object detection methods. In fact, on salient object datasets where the saliency problem is clear such as ASD dataset, it can achieve a performance of around

0.86 maximum F1 measure. Moreover, the average running time of QCut is around 1.5 seconds for a typical image of size 200x150, which is much less than the work proposed in [P1].

Although the ground-state of the system defined as a modified Hamiltonian provides an approximation of the global minimum of the related optimization problem which produces salient object segments, it has been observed in [P3] that other local minima may also efficiently represent parts of the salient object or another salient object. These local minima have been obtained by exploiting a large spectrum of the eigen-problem used in QCut. The challenge of efficiently combining segment proposals is efficiently addressed by the multispectral QCut, by adopting a learn-to-rank approach. Specifically, a machine learning method has been exploited to predict the accuracy of the segments. The ranked segments according to this measure are then combined using a rule-based approach to obtain the final salient object detection. The approach in [P3] provides a significant performance improvement on QCut. For example, on ASD dataset, it improves the performance of QCut from 0.86 to 0.92 maximum F1 measure. However, such improvement is brought at a high cost of around 45 seconds for a typical image of size 400x300.

The proposed method in [P4], both addresses the computational complexity of the method in [P3] and brings a high performance improvement on QCut. In particular, besides the exploitation of a larger spectrum, it was founded that the graph structure representing images plays a crucial role in the performance of QCut. In Extended Quantum-Cuts (EQCut) [P4], a better graph structure was investigated for salient object detection. Accordingly, the abstraction of images by superpixels reduces the noise when compared to the pixel-level representation. Moreover, it reduces the computational complexity of QCut and makes a multi-resolution approach possible. A multi-resolution approach is crucial as it covers a large variety of object sizes. Higher order connectivity within the superpixel graph provides a better representation of the pairwise relations in the image as long as the pairwise affinities are normalized according to the spatial relations. The novel affinity assignment procedure in EQCut can cover these aspects. In a graph obtained in this way, the optimization criterion that is solved takes more sophisticated local contrast and global contrast measures into account. The effect of this in salient object detection performance was proven by the significant performance improvement of EQCut over the QCut. For example, in SOD dataset, EQCut achieves a maximum F1 score of 0.682, whereas QCut performance in same measure is only 0.606. On the other hand the proposed method in [P3] can also offer a decent performance on 0.672, however the computational complexity of EQCut is much lower than this method, as it takes only around 0.85 seconds for a typical image of size 400x300.

The EQCut approach further reveals that the adopted affinity matrix defines the only determining factor for the performance of the salient object detection. Hence, in addition to handcrafted graph structures and affinity assignments strategies, the calculation of affinity matrix was further investigated by following the supervised approaches in [P5] and [P6]. In [P5], it has been observed that even learning a linear combination of well-known affinity calculation functions lead to a decent performance improvement on EQCut which employs only one of these functions. For example, in

JUDD dataset the learnt affinity calculation function leads to a performance improvement of 3% over EQCut, whereas the computational burden is increased by 11%. The main finding of the approach in [P6] is that an end-to-end salient object detection network can be trained by concatenating a convolutional kernel network (CKN) and the EQCut block. This network is able to learn affinities that work well with EQCut via learning a feature transform on color features by CKN. It was observed that this system achieves a significant performance improvement over EQCut in terms of the error metrics that are used during training. For example, in the challenging DUT-OMRON dataset, the average KLD error of EQCut is around 4.75, whereas the supervised approach in [P6] can achieve a very low error of 1.88. Besides, the average running time of this method is only around 5% higher than that of EQCut.

In conclusion, the contributions presented in this thesis offer fast and state-of-the-art solutions to the salient object detection problem from images. These methods follow from the unique research topic presented for the first time in this thesis: the link between Spectral Graph Theory and Quantum Mechanics. Such special relation is expected to offer future benefits for other Computer Vision tasks as well as salient object detection, via further exploitation of other quantum mechanical concepts such as superposition principle and the time evolution of wavefunctions. In particular, the superposition principle allows producing a segment pool of high quality which might act as object proposals. The time evolution of wavefunctions on the other hand might provide more accurate saliency maps as propagating waves might provide a better coverage of the entire salient object.

Appendix

Theorem 2-2 gives a solution for the minimization of the Rayleigh Quotient $R(\mathbf{M}, \boldsymbol{x}) = \frac{\boldsymbol{x}^*\boldsymbol{M}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}$, for real and symmetric matrices $\boldsymbol{M}$. Particularly, the solution is obtained by performing an eigenanalysis on matrix $\boldsymbol{M}$ and the minimizing vector is the eigenvector corresponding to the minimum eigenvalue. However, for asymmetric matrices, this is not the case. Similar to a procedure adopted in Section 2.2.2, without loss of generality, let us reformulate minimization problem by setting $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x} = 1$, as follows.

$$\boldsymbol{x}_* = \min_{\boldsymbol{x}} \boldsymbol{x}^*\mathbf{M}\boldsymbol{x}\ , \qquad s.t.\ \boldsymbol{x}^*\boldsymbol{x} = 1. \tag{4.1}$$

With use of Lagrangian Multiplier, one can reformulate the above problem as follows.

$$\boldsymbol{x}_* = \min_{\boldsymbol{x}}\big(\boldsymbol{x}^*\mathbf{M}\boldsymbol{x} - \gamma(\,\boldsymbol{x}^*\boldsymbol{x} - 1)\big). \tag{4.2}$$

The stationary points of this objective function are obtained by setting the partial derivative of $R(M, x)$ with respect to $\boldsymbol{x}$**,** to zero. Then, for real $\boldsymbol{x}$, the stationary points satisfy the following.

$$(\mathbf{M} + \mathbf{M}')\boldsymbol{x} = 2\gamma\boldsymbol{x}\,. \tag{4.3}$$

In other words, the stationary points are eigenvectors of $(\mathbf{M} + \mathbf{M}')$ and the minimizing vector $\boldsymbol{x}_*$ is the eigenvector of $(\mathbf{M} + \mathbf{M}')$ corresponding to the minimum eigenvalue.

On the other hand, if the $\boldsymbol{x}_*$ vector is generalized to have complex values, then the stationary points satisfy the following relations [56].

$$\mathbf{M}\boldsymbol{x} = \alpha\boldsymbol{x}, \quad \mathbf{M}'\boldsymbol{x} = \beta\boldsymbol{x}. \tag{4.4}$$

In other words, a complex stationary point $\boldsymbol{x}$ should be an eigenvector of both $\mathbf{M}$ and $\mathbf{M}'$. However, the existence of such an eigenvector is not guaranteed.

Next, we interpret these results in the special case of QCut optimization problem. It has been discussed in Section 3.2 that with the vector definition of $\boldsymbol{z}$ in (3.6), the phase term in (3.9) can be neglected. However, one can only get an approximation of the optimization problem in (3.9) with relaxing the values of $\boldsymbol{z}$ to real or even complex values. In this case, the phase term is not negligible and contributes to the approximation error. Let us formulate the above discussion as follows.

$$R(\mathbf{H}, \boldsymbol{z}) = \frac{\boldsymbol{z}^* \mathbf{H} \boldsymbol{z}}{\boldsymbol{z}^* \boldsymbol{z}} \cong \frac{cut(P, \bar{P})}{area(P)} + phase(P) + \mu_{\boldsymbol{v}}. \tag{4.5}$$

In (4.5), P is the foreground region represented by the vector with the definition of (3.6). In some cases the error from phase can dominate the $\frac{cut(P,\bar{P})}{area(P)}$ term which is desired to be minimized. It can be observed from (3.8) that $\frac{cut(P,\bar{P})}{area(P)}$ is guaranteed to be positive. However, no such guarantee exists for phase term. For $\mathbf{H}$ that is not positive-definite, $R(\mathbf{H}, \boldsymbol{z})$ can take negative values, and these negative values are an evidence that the phase term is dominating the cut term, which is not desired. Hence, one should look for a minimization of $R(\mathbf{H}, \boldsymbol{z})$ with the constraint that it will take positive values.

The asymmetric $\mathbf{H}$ matrix used in [P4] can be decomposed as $\mathbf{H} = \mathbf{L} + \mathbf{V}$, where $\mathbf{L}$ is a Laplacian matrix and $\mathbf{V}$ a diagonal real and positive definite matrix. Hence, the eigenvalue of $\mathbf{H}$ with minimum absolute value is guaranteed to be real and positive [57].

Thus, in [P4], we use the eigenvector of $\mathbf{H}$ corresponding to the eigenvalue with minimum absolute value. The result in (4.3) is avoided since it is not guaranteed to make $R(\mathbf{H}, \boldsymbol{z})$ positive. On the other hand, the stationary points offered by (4.4) are not guaranteed to exist, and even if they do, they have to be eigenvectors of $\mathbf{H}$. Hence, they would never offer a $R(\mathbf{H}, \boldsymbol{z})$ that is smaller than the one obtained by the eigenvector of $\mathbf{H}$ corresponding to the eigenvalue with minimum absolute value. Note that this eigenvector is not guaranteed to be a stationary point for $R(\mathbf{H}, \boldsymbol{z})$, however, in order to minimize $\frac{cut(P,\bar{P})}{area(P)}$ only and suppress the phase term, we do not exactly need a local minimum for $R(\mathbf{H}, \boldsymbol{z})$. Instead, we need a non-negative, small value so that $\frac{cut(P,\bar{P})}{area(P)}$ will be small and $R(\mathbf{H}, \boldsymbol{z})$ will not be dominated by the phase term.

References

[1] Liboff, R. L. “Introductory Quantum Mechanics”, Addison Wesley, 1987.

[2] M. Planck, “On the Law of Distribution of Energy in the Normal Spectrum,” Annalen der Physik, vol. 4 pp. 553-558, 1901.

[3] A. Einstein, “Uber einen die Erzeugung und Verwandlung des Lichtes betreffenden heurischen Gesichtpunkt,” Annalen der Physik, vol. 17, pp. 132-148, 1905.

[4] N. Bohr, “On the Constitution of Atoms and Molecules,” The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, vol. 26, no. 151, pp. 1-25.

[5] L. de Broglie, “Recherches sur la Theorie des Quanta,” Ann. De Physique, vol. 10, no.3, 1925.

[6] W. Heisenberg, “Über den Anschaulichen Inhalt der Quantentheoretischen Kinematik und Mechanik,” Zeitschrift für Physik, vol. 43, no 3-4, pp. 172-198, 1927.

[7] M. Born, “Mechanics of the Atom,” George Bell & Sons, 1927.

[8] F. RK Chung, “Spectral Graph Theory,” American Mathematical Society, vol. 92, 1997.

[9] L. Hagen, A. B. Kahng, “New Spectral Methods for Ratio Cut Partitioning and Clustering,” IEEE Transactions on Computer-Aided Design, vol. 11, no. 9, pp.1074-1085, September, 1992.

[10] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, August 2000.

[11] S. Sarkar and K. Boyer, “Quantitative Measures of Changes Based on Feature Organization: Eigenvalues and Eigenvectors,” Computer Vision and Image Understanding, vol. 71, no.1, pp. 110-136, July 1998.

[12] R. Achanta, A Shaji, A. Lucchi, P. Fua and S. Süsstrunk, “SLIC Superpixels Compared to the State-of-the-Art Superpixel Methods,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274-2281, Nov. 2012.

[13] S. C. Turaga, J. F. Murray, V. Jain, F. Roth, M. Helmstaedter, K. Briggman, D. Winfried and H. S. Seung, “Convolutional Neural Networks can Learn to Generate Affinity Graphs for Image Segmentation,” Neural Computation, vol. 22, no. 2, pp. 511-538, 2010.

[14] K. Briggman, W. Denk, S. Seung, M. N. Helmstaedter, and S. C. Turaga, “Maximin Affinity Learning for Image Segmentation,” Advances in Neural Information Processing Systems (NIPS), pp. 1865-1873, February, 2010.

[15] T. Cour, N. Gogin and J. Shi, “Learning Spectral Graph Segmentation,” Workshop on Artificial Intelligence and Statics (AISTATS), 2005.

[16] R. Ranftl and T. Pock, “A Deep Variational Model for Image Segmentation,” German Conference on Pattern Recognition, pp. 107-118, September 2014.

[17] Y. Boykov, O. Veksler and R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222-1239, November 2011.

[18] A. Krizhevsky, I. Sutskever and G. E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems (NIPS), pp. 1097-1105, December 2012.

[19] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” CoRR, abs/1409.1556, 2014.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, “Going Deeper with Convolutions”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9, June 2015.

[21] C. Farabet, C. Couprie, L. Najman and Y. LeCun, “Learning Hierarchical Features for Scene Labeling,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1915-1929, August 2013.

[22] J. Long, E. Shelhamer and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, June 2015.

[23] J. Mairal, P. Koniusz, Z. Harchaoui and C. Schmid, “Convolutional Kernel Networks,” Advances in Neural Information Processing Systems (NIPS), pp. 2627-2635, December 2014.

[24] C. Aytekin, E. Tunali and S. Oz, “Fast Semi-automatic Target Initialization based on Visual Saliency for Airborne Thermal Imagery,” International Joint Conference on Vision Theory and Applications (VISAPP), pp. 490-497, January 2014.

[25] M. Casares, S. Celipasalar and A. Pinto, “Light-weight Salient Foreground Detection for Embedded Smart Cameras,” Computer Vision and Image Understanding, vol. 114, no. 11, pp. 1223-1237, 2010.

[26] E. Horbert, G. M. Garcia, S. Frintop and B. Leibe, “Sequence-level Object Candidates Based on Saliency for Generic Object Recognition on Mobile Systems,” InternationalConverence on Robotics and Automation (ICRA), pp. 127-1234, 2015.

[27] X. Hou and L. Zhang, “Thumbnail Generation based on Global Saliency,” Advances in Cognitive Neurodynamics, pp. 999-1003, 2008.

[28] R. Margolin, L. Z. Manor and A. Tal, “Saliency for Image Manipulation,” The Visual Computer, vol. 29, no. 5, pp. 381-392, 2012.

[29] L. Itti, C. Koch and E. Niebur, “A Model of Saliency-based Visual Attention for Rapid Scene Analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, 1998.

[30] S. K. Ungerleider and G. Leslie, “Mechanisms of Visual Attention in the Human Cortex,” Annual Review of Neuroscience, vol. 23, no. 1, pp. 315-341, 2000.

[31] D. Rudoy, D. B. Goldman, E.Shechtman and L. Z. Manor, “Learning Video Saliency from Human Gaze Using Candidate Selection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1147-1154, 2013.

[32] Z. Li, S. Qin and L. Itti, “Visual Attention Guided Bit Allocation in Video Compression,” Image and Vision Computing, vol. 29, no. 1, pp. 1-14, 2011.

[33] H.Hadizadeh and I. V. Bajic, “Saliency-Aware Video Compression,” IEEE Transactions on Image Processing, vol. 23, no. 1, pp. 19-33, 2014.

[34] T. Liu, J. Sun, N. N. Zheng, X. Tang and H. Y. Shum, “Learning to Detect a Salient Object,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 353-367, December 2010.

[35] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk, “Frequency-tuned Salient Region Detection,” IEEE Conference Computer Vision and Pattern Recognition (CVPR), pp. 1597-1604, June 2009.

[36] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, S. Hu, “Global Contrast Based Salient Region Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, pp. 569-582, March 2015.

[37] W. Zou, K. Kpalma, Z. Liu and J. Ronsin, “Segmentation Driven Low-rank Matrix Recovery for Saliency Detection,” British Machine Vision Conference (BMVC), pp. 1-13,September 2013.

[38] M. Everingham, L. V. Gool, C. K. Williams, J. Winn and A. Zissermann, “The Pascal Visual Object Classes Challenge,” International Journal of Computer Vision, vol. 88, no. 2, pp. 303-338, 2010.

[39] V. Mohavedi and J. H. Elder, “Design and Perceptual Validation of Performance Measures for Salient Object Segmentation,” IEEE Computer Society Workshop on Perceptual Organization in Computer Vision (POCV), 2010.

[40] D. Martin, C. Fowlkes, D. Tal and J. Malik, “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics,” International Conference on Computer Vision (ICCV), pp. 416-423, 2001.

[41] C. Yang, L. Zhang, H. Lu and M. Yang, “Saliency Detection via Graph Based Manifold Ranking,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3166-3173, June 2013.

[42] A. Borji, “What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection,” IEEE Transactions on Image Processing, vol. 24, no. 12, February 2015.

[43] T. Judd, K. Ehinger and F. Durand, “Learning to Predict Where Humans Look,” IEEE Conference on Computer Vision (ICCV), pp. 2106-2113, September 2009.

[44] A. Borji, M. M. Cheng, H. Jiang and J. Li, “Salient Object Detection: A Survey,” arxiv preprint, arXiv: 1411.5878, 2014.

[45] A. Borji, M. M. Cheng, H. Jiang and J. Li, “Salient Object Detection: A Benchmark,” IEEE Transactions on Image Processing, vol.24, no. 12, pp. 5706-5722, vol. 24, no. 12, December 2015.

[46] X. Li, H. Lu, L. Zhang, X. Ruan, and M. H. Yang, “Saliency Detection via Dense and Sparse Reconstruction,” IEEE International Conference on Computer Vision (ICCV), pp. 2976-2983, December 2013.

[47] W. Zhu, S. Liang, Y. Wei and J. Sun, “Saliency Optimization from Robust Background Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2814-2821, June 2014.

[48] B. Jiang, L. Zhang, H. Lu, C. Yang and M. H. Yang, “Saliency Detection via Absorbing Markov Chain.” IEEE International Conference on Computer Vision (ICCV), pp. 1665-1672, December 2013.

[49] Z. Liu, W. Zou and O. L. Meur, “Saliency Tree: A Novel Saliency Detection Framework,” IEEE Transactions on Image Processing, vol. 25, no. 5, pp. 1937-1952, May 2014.

[50] J. Jiang, J. Wang, Z. Yuan, Y. Wu, N.Zheng and S. Li,” Salient Object Detection: A Discriminative Regional Feature Integration Approach,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2083-2090, June 2013.

[51] G. Li and Y. Yu, “Visual Saliency Based on Multiscale Deep Features,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5455-5463, June 2015.

[52] R. Zhao, W. Ouyang, H. Li and X. Wang, “Saliency Detection by Multi-context Deep Learning,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1265-1274, June 2015.

[53] A. Einstein, B. Podolsky and N. Rosen, “Can Quantum Mechanical Description of Physical Reality be Considered Complete?,” Physical Review vol. 47, no. 10, pp 777-780, March 1935.

[54] N. Otsu, “A Threshold Selection Method from Gray-level Hisograms,” IEEE Transactions on Systems, Man and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979.

[55] C. Rother, V. Kolmogorov and A. Blake, “GrabCut: Interactive Foreground Extraction using Iterative Graph Cuts,” ACM Transactions on Graphics (SIGGRAPH), pp. 309-314, 2004.

[56] B. N. Parlett, “The Rayleigh Quotient Iteration and Some Generalizations for Nonnormal Matrices,” Mathematics of Computation, vol. 28, no. 127, pp. 679-693, July 1974.

[57] R. Agaev and P. Chebotarev, “On the Spectra of Nonsymmetric Laplacian Matrices,” Linear Algebra and its Applications, vol. 399, pp. 157-168, 2005.

ORIGINAL PAPERS

I

QUANTUM MECHANICS IN COMPUTER VISION: AUTOMATIC OBJECT EXTRACTION

by

Ç. Aytekin, S. Kiranyaz & M. Gabbouj, September 2013

IEEE International Conference on Image Processing (ICIP), pp. 2489-2493.

II

AUTOMATIC OBJECT SEGMENTATION BY QUANTUM-CUTS

by

Ç. Aytekin, S. Kiranyaz & M. Gabbouj, August 2014

International Conference on Pattern Recognition (ICPR), pp. 112-117

III

LEARNING TO RANK SALIENT SEGMENTS EXTRACTED BY MULTISPECTRAL QUANTUM-CUTS

by

Ç. Aytekin, S. Kiranyaz & M. Gabbouj, March 2016

Pattern Recognition Letters, vol. 72, pp. 91-99

IV

EXTENDED QUANTUM-CUTS FOR UNSUPERVISED SALIENT OBJECT EXTRACTION

by

Ç. Aytekin, E. C. Ozan, S. Kiranyaz & M. Gabbouj, 2016

Multimedia Tools and Applications, DOI: 10.1007/s11042-016-3431-1

V

SALIENT OBJECT SEGMENTATION BASED ON LINEARLY COMBINED AFFINITY GRAPHS

by

C. Aytekin, A. Iosifidis, S. Kiranyaz and M. Gabbouj, 2016

International Conference on Pattern Recognition (ICPR), Accepted,

VI

LEARNING GRAPH AFFINITIES FOR SPECTRAL GRAPH-BASED SALIENT OBJECT DETECTION

by

C. Aytekin, A. Iosifidis, S. Kiranyaz and M. Gabbouj, 2016

Pattern Recognition, Accepted,

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3859-9
ISSN 1459-2045