An Enhanced Markov Random Field (MRF) based Approach for Image Segregation

by

AKBAH A. KALIFA
(0940210438)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

School of Computer and Communication Engineering
UNIVERSITI MALAYSIA PERLIS

2017

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name: **AKBAH A. KALIFA**

Date of birth: **01 / 01 / 1971**

Title: **AN ENHANCED MARKOV RANDOM FIELD (MRF) BASED APPROACH FOR IMAGE SEGREGATION**

Academic Session: **2016 / 2017**

I hereby declare that the thesis becomes the property of Universiti Malaysia Perlis (UniMAP) and to be placed at the library of UniMAP. This thesis is classified as :

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
☑	**OPEN ACCESS**	I agree that my thesis is to be made immediately available as hard copy or on-line open access (full text)

I, the author, give permission to the UniMAP to reproduce this thesis in whole or in part for the purpose of research or academic exchange only (except during a period of _____ years, if so requested above).

Certified by:

______________________	______________________
SIGNATURE	**SIGNATURE OF SUPERVISOR**
A10834388	**Dr. Shahrul Nizam bin Yaakob**
(NEW IC NO. / PASSPORT NO.)	**NAME OF SUPERVISOR**
Date :________________	Date : ________________

NOTES : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentially or restriction.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my enthusiastic supervisor Dr. Shahrul Nizam bin Yaakob for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I would like to thank Ir. Prof. Dr. R. Badlishah Bin Ahmad the co-supervisor of this research. My sincere thanks to all members of the School of Computer and Communication Engineering, both staff and students, whose continuous support have made this thesis possible. I would like to thank Center for Graduate Studies (CGS) for correcting formatting and suggested possible improvements of this thesis. Also I would like to thank to Universiti Malaysia Perlis and Research, Management & Innovation center for the fellowship support.

I would like to thank all of the participants in the research: students, lecturer, friends and Local Education Authorities, for the time and help given throughout. Without their participation, this research would not have been possible.

Finally, but by no means least, I would also like to thank my brothers and sisters, without whom this thesis would not have been started or completed! Your encouragement and support has never faltered; thank you.

Dedicated to my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	One Dimension
2D	Two Dimension
3D	Three Dimension
BL	Background Label
CDAP	Contrast Depth Asymmetry Principle
CIE	Comission Internationale de lÉclairage (Color Space)
DBN	Dynamic Bayesian Network
EMD	Earth Moverś Distance
GRF	Gibbs Random Field
ICM	Iterated Conditional Modes
MAP	Maximum a Posteriori
MLL	Multi Level Logistic
MRF	Markov Random Field
OL	Object Label
RGB	Red-Green-Blue (Color Space)
RL	Relaxation Labeling
RP	Reward, Punishment
SA	Simulated Annealing
USB	Universal Serial Bus
TD	Tangent Discontinuities

Pendekatan Berasaskan Penambahbaikan Bidang Rawak Markov untuk Imej Pengasingan

ABSTRAK

Objektif utama tesis ini adalah untuk meningkatkan teknik alternatif bagi menambahbaik pengasingan kedalaman $2D$ monokular imej menggunakan Bidang Rawak Markov. Kedalaman pengasingan adalah tugas yang mencabar disebabkan oleh segmen yang luas bukan ketegaran dan mempunyai perbezaan tekstur antara objek. Penampilan objek dalam berbeza bentuk dan lokasi di tempat kejadian, kedalaman pengasingan adalah juga dibuat sukar kerana komponen tambahan, objek tersumbat, yang boleh sama ada boleh dilihat atau sama sekali tidak kelihatan dari tempat kejadian atau yang sedia ada imej tidak menentu menampung dalam persekitaran yang tidak dikekang, dimana ianya mungkin meningkatkan kesukaran proses pengasingan. Satu perubahan dalam pengagihan cahaya dalam imej boleh menimbulkan perubahan yang ketara dalam aspek objek dalam imej. Tesis ini menunjukkan bahawa kedalaman pengasingan monokular boleh berjaya digunakan dalam segmentasi imej kawasan berasaskan pinggir. Kedalaman teknik imej pengasingan telah dicadangkan untuk mencari isyarat kedalaman di kawasan imej, kawasan imej yang sepadan dengan tahap yang ditetapkan kedalaman petunjuk boleh dikenal pasti dan dikelaskan. Teknik pemisahan imej mendalam telah dicadangkan untuk mencari isyarat kedalaman di wilayah imej, wilayah imej yang sepadan dengan tahap petunjuk mendalam tertentu dalam objek bertindih seperti persimpangan-T dan persimpangan-L dapat dikenal pasti dan diklasifikasikan. Teknik yang dicadangkan pada mulanya dengan melaksanakan segmentasi imej; untuk mengenal pasti kawasan dalam imej, operasi morfologi; untuk mengurangkan apa-apa kesilapan piksel, maka kelebihan pengesanan; untuk mengenal pasti sempadan kawasan, dan menggunakan maklumat ini untuk melaksanakan proses pengasingan kedalaman untuk mengenal pasti dan melabelkan objek secara terperinci untuk dilabel. Hasil eksperimen menunjukan bahawa teknik ini telah berjaya mengasingkan kawasan bagi kedalaman dari $2D$ monocular imej. Teknik imej bersegmen telah digabungkan dengan imej pengesanan kelebihan untuk melaksanakan proses pengasingan kedalaman dan hasilnya adalah baik dari segi susunan objek pelabelan. teknik yang dicadangkan telah diuji dengan berbeza bilangan lelaran, dengan $(800,000)$ lelaran ketepatan persembahan teknik adalah (85.85%), manakala ketepatan sebelum gabungan itu adalah (74.74%) dengan jumlah yang sama lelaran. Eksperimen menunjukkan keupayaan teknik dapat diintegrasikan dengan sistem sejagat seperti yang ditunjukkan dalam keputusan yang dibentangkan kedalaman diasingkan dalam lapisan $2.1D$. Keputusan perbandingan telah menunjukkan bahawa, pendekatan model berdasarkan yang dicadangkan adalah lebih fleksibel kerana ia boleh merawat kes-kes stalemate itu dan rantau latar depan tidak hanya disambungkan. Oleh kerana pendekatan yang mantap dengan model pengasingan tersebut, ia boleh berurusan dengan imej termasuk kawasan bertekstur dengan variasi intensiti yang tinggi dan berbeza.

An Enhanced Markov Random Field (MRF) based Approach for Image Segregation

ABSTRACT

The main objective of this thesis is to present an alternative technique to enhance the depth segregation of a $2D$ monocular image using Markov Random Field (MRF). Depth segregation considered challenging task due to an extensive segment of non-rigidity and textural contrasts among objects. Object appearance in various shape and location in the scene, depth segregation is likewise made difficult due to extra components, occluded objects, which can be either visible or totally invisible from the scene, variation in light distribution in image can give rise to a significant change in the aspect of the objects in the image most likely to increase the difficulty of the process. Here is what gives the rise to the problem of developing image segregation tools that can deal with this variation in images, this tool required to be flexible and robust to successfully segregate objects in to layers from a $2D$ monocular image. This thesis shows that monocular depth segregation can be successfully used in edge region based image segmentation. Depth image segregation enhanced technique has been proposed to search for depth cues in the image regions, the image regions corresponding to a specified level of depth cue in occluding and occluded objects such as T-junction and L-junction can be identified and classified. The proposed technique initially executes image segmentation; to identify region in the image, morphological operation; to eliminate any error pixel in the region, then edge detection; to identify the boundaries of the regions, and use these information to perform depth segregation processes to identify and label the objects in depth labeled order. The experimental result shows that the technique has successfully segregate the regions in depth order from a $2D$ monocular image. The segmented image has been combined with the edge detection image to perform the depth segregation process and the result was efficient in terms of object labeling order. The proposed technique was tested with varies number of iterations, with the $(800,000)$ iteration the enhanced technique show accuracy of (85.85%), while the accuracy before the combination was (74.74%) with the same number of iterations. The result shows the ability of the enhanced technique to be integrated with a universal system as shown in the presented result of the segregated depth in $2.1D$ layers. Comparative results have shown that, the proposed technique based approach is more flexible since it can treat cases such occlusion and not simply connected foreground region. Due to the robust approach by segregation enhanced technique, since it can deal with images including textured regions with high intensity variations.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, the world of 3*D* and multi-dimensions system well expected that it will make a powerful realistic and immersive way of presenting images and videos (H. Su, Huang, Mitra, Li, & Guibas, 2014). Also with respect to the aspect of sophisticated 3*D* hardware and software system, the consumer demands is rapidly increasing for 3*D* contents (Moustakas et al., 2006). Thus, 3*D* conversion technique of existing 2*D* images and video frames has begun to be a serious issue. In order for the consumer market to grow, the availability of having an efficient conversion tool from 2*D* to 3*D* is required.

The tool developed by researcher is focusing on the depth maps in image generation, mainly motivated be the market industry. Most of the researcher make use of a reliable cue for depth estimation; moreover, the cues like T-junctions and L-junctions cues can be detected in specific image areas and provide depth order information that can be present in 2.1*D* layer presentation. The 2.1*D* model considered to be an intermediate state between 2*D* images and full/absolute 3*D* maps, 2.1*D* represent the image as a partition with its object ordered by its relative depth. State of the art depth ordering systems on 2*D* monocular images focus on the extraction of foreground regions from the background. Although this may be appropriate for some applications, more information can be extracted from a 2*D* monocular image (Palou & Salembier, 2012).

Depth segregation is a process in automated image analysis and objects extraction. It is during the depth segregation that object of interest will be extracted from an image in subsequent processing like, segmentation, edge detection, region surface and object identifier. In this low-level operation, that concern in partitioning images by determine the dis-joint and homogeneous objects region, or determined by finding boundaries of an object in image. Where the homogeneous region or the joint pixel grouped to present the genuine objects or part of the object inside the image. Therefore, in a most applications in 3*D* image processing, computer vision and virtual reality, depth estimation starts with the process of partitioning the objects in a monocular image to help provide better depth information. At this point, the $2.1D$ sketch method is applied; such sketch divides the regions into layers based on local occlusion cues of the overlapping objects in an image, are consider to be a fundamental role and the first step of applying technique in 3*D* images in order to achieve higher-level operations such as object recognition, distance measuring, region interpretation and color representation.

The need of new development of automatic techniques to find the depth and related information in a 2*D* monocular image by combining several techniques to obtain the region information from an existing monocular image to generate 3*D* content, over all for 2*D* to 3*D* depth information and conversion technique. Algorithms that apply this technique and collect depth information are yet to exist due to a several uncontrolled environment such as noise images, camera resolution and illuminations.

1.2 Related Work

It is obvious at this point that most of digital images are merely a projection of a 3*D* scene (Gonzalez et al., 2002). Thus, as a consequence of the projection, objects spatially separated in the 3*D* world might interfere with each other in the projected 2*D* plane and each of them occludes part of the ground (Moustakas et al., 2006). With that said, decomposing 2*D* image data into different objects and determining how objects and surfaces interact in the scene from their 2*D* projection is usually an effortless task for human vision, but it still represents one of the major challenges that both neuroscience and computer vision are facing nowadays.

A possible explanation of the importance of binocular disparity as source of information about the 3D structure of the world emerged from the work of (Garcia-Diaz et al., 2012), when the invention of the stereoscope allowed demonstrating that viewers could achieve a vivid 3D perception of an object depicted by a pair of appropriately hand-drafted drawings. Given that we can achieve this ideal, binocular disparity has been for more than one century considered the undisputed source of 3D surface structure. When evaluating perception, it is important to consider the foundation of the Gestalt School of Psychology, where the role of monocular factors has been evidenced. Recognizing this, in many chapters of his book in (M. Singh & Anderson, 2002) dedicated to depth perception, also demonstrates that depth information can also be observed in absence a close similarity, connection.

As noted earlier, in spite of the fact that this results were known widely at the time when computer vision come into view as a new regulation, this issue remains alive, as signaled by great work of effort that has been invested by the community in the attempt of designing a new algorithms to recover depth information from stereo (Szeliski, 2010) and create other new cues and roll that can process multiple images, such as pattern or organization from motion (Helmholtz, 2000. Original publication 1866) or depth information from defocus (Amer et al., 2015). The controversy of interaction between Gestalt Psychology and computer vision is mainly due to the qualitative nature of Gestalt theory. The distinction use of mathematical simplification of digital image did not to take notice of or acknowledge by Gestaltists and the related issues of image information like blur and noise were not well-considered qualitatively. These prevalent practical realities have somehow mitigated the evident strength of monocular factors as source of depth information in computer-based systems (Dimiccoli, Morel, & Salembier, 2008). The problems of segmentation with depth and monocular depth cue detection are intimately bound together; a good segmentation mask gives extremely clean contours, useful for monocular depth cue detection, while monocular depth cues constitute an excellent source of information to be exploited in low-level segregation.

A notable finding in recent years, motivated by the number of applications such as object removal, image understanding, 3*D* scene reconstruction and synthesis, which benefits the advances in the field and encouraged by the enormous progresses in machine learning over the last decade, the computer vision community has focused its interest on recovering the spatial layout from single images. Additionally, the state-of-the-art techniques aim to learn the structure of the visual world from a set of training images, in order to attempt depth recovery in unseen test images. The proliferation of such approaches allow to incorporate prior experience about the structure of the environment as for instance that blue patches are more likely to be the sky and green patches are more likely to be grass on the ground and therefore green patches should be closer to the viewpoint than blue patches.

While learning is arguably the most crucial factor for recognition, the issue raised in this section of whether it really influences basic depth segregation and grouping remains controversial. After a series of studies (Anderson, 2011) proposed a theory that emphasizes the importance of learning but only in the high stages of the visual perception process, while most visual information is provided by inherent cues, which are a direct response to the retinal stimulation. In a more formidable visual development, (Shipley & Kellman, 2001) have proposed a framework for object perception that includes depth cues into the grouping process and does not incorporate any feedback from high stage processes to the low ones. A note of caution from them is that object perception can undoubtedly proceed without such feedback, and likely does so in cases where there is no obvious involvement of learned information. Thus, it is more appropriate to suggest that the lack of reliable methods for computing inherent monocular depth cues still represents one of the major limitations not only for a more effective exploration of this fundamental issue but also for a potential exploitation of these cues in computer-based applications.

Standing in full contrast to state-of-the-art approaches and having image segregation as main objectives, the proposes of methods for monocular depth cue detection in single images rely neither on previously learned information about the structure of the world nor on any assumption on the image structure. Substantially, these depth cue detectors open the door to the introduction of an important intermediate layer in applications such as image segmentation and segregation. As such, the integration of depth ordering information provided by depth cues into low-level processing allows not only a more accurate segregation, which preserves low level structures, but also gives a new image representation, closer to the real world, in which the scene is considered as a set of independent objects with an associated relative depth. In a strict sense, this spatial understanding of images can be used as a foundation for other visual tasks, enabling a wide variety of applications such as for example automatic object removal, image indexing and $3D$ reconstruction.

1.3 Problem Statements

The applications and the challenging of the depth segregation problem arousing curiosity or interest . In the field of image processing, depth segregation is consider a first step for $3D$ systems. Hence, it becomes the most important step in designing $3D$ systems. So far, researchers pay particular attention to the depth segregation problem in term of technique and applying role, and the prior data in a monocular image is usually separated by either segregation or depth information. Moreover depth segregation having the potential to become or develop into a human computer interface and surveillance systems in the future.

Depth segregation considered to be a complex task due to a three primary reasons. First, there is an extensive segment of non-rigidity and textural contrasts among objects. Object appearance varies in scene of shape and location. Second, depth segregation is likewise made difficult due to extra components, for example occluded objects, which can be either visible or totally invisible from scene. All these extra elements increment the inconstancy of the depth segregation that an object detection system must process. Third, A variation in light distribution in image can give rise to a significant change in the aspect of the object in the image, most likely to increase the difficulty of the process. All this reasons should be carefully thought and considered when developing a depth segregation system. Depth segregation is an interpretative first step in any 3*D* system. Even though research on depth segregation and 3*D* construction begun very early, there was no great attention or interest to the monocular image to solve the problem until recently, especially after the 3*D* images become more demanded in present for both industrial and social science. Over the last ten years, more attention has been given to the depth segregation and depth information tools and there is a growing number of methods contribute to depth segregation in image.

Among these tools and methods, there is an attention in MRF energy function modeling that is used in image segregation and segmentation. Modeling problems in this thesis are seen mainly from the computational viewpoint. The primary concerns are how to define an energy function for the optimal solution to a vision problem and how to find the optimal solution. The reason for defining the solution in an optimization sense is due to various variation in vision processes. It usually look for an optimal one in the sense that an objective in which constraints are encoded is optimized. The energy function theory provides a convenient and consistent way of modeling context dependent entities such as image pixels and other spatially correlated features. This is achieved through characterizing mutual influences among such entities using MRF probabilities. A particular MRF model favors its own class of patterns by associating them with larger probabilities than other pattern classes. MRF theory is often used in conjunction with statistical decision and estimation theories, so the use of formulate energy functions in terms of established optimality principles its relation with the Maximum a posteriori (MAP) probability which, is one of the most popular statistical criteria for optimality and in fact, has been the most popular choice in MRF vision modeling.

1.4 Research Objectives

The goal of this thesis is to present an alternative technique to enhance the efficiency of depth segregation with high accuracy and reliability from a $2D$ monocular image. There are three challenging objectives defined for the work been presented on this thesis;

1. Enhance the techniques of Markov Random Field (MRF) for image segregation.
2. Enhance the segregation process of overlapping objects in $2D$ monocular image.
3. To analyze the performance of the proposed technique in image segregation.

1.5 List of Contribution

This thesis is present an alternative enhanced technique to increase the efficiency of depth segregation in a $2D$ monocular image that is contributing to the image processing community.

1. An enhancement in the Markov Random Field (MRF) energy function computational, by adding the pixel feature texture τ combined with the edge ε of the object in the $2D$ image.
2. Enhancement in image segregation process, incorporate edge and segment image in the depth computation.
3. The enhanced segregation technique will be combined with four processes; image segmentation, morphological erosion, edge detection and depth segregation.
4. The related effect of the iterations number in the MRF energy function.
5. Test the performance of the enhanced technique with different type (Real and artificial) images.

1.6 Scope

This thesis is focused on the field of depth segregation from $2D$ monocular image technique and demonstrating the utility of image segmentation and depth segregation using Markov Random Field (MRF). For image segmentation, the knowledge of the objects or regions can systematically integrate various types for solving the depth segregation problem. With the design of the main theory of the depth segregation. Image segregation provide a systematic way to find the overlapped objects or regions by choosing direct or indirect links based on the natural relationship between the boundaries of the regions. This overlapped region than give another set of data to the depth segregation technique, so that the principle of the occlusion can be performed in the segregated image. Based on sparse occlusion cues on the boundary of the regions the depth segregation technique inference the overlapped regions. Finally the technique present the regions in a $2.1D$.

1.7 Overview of The Thesis

This thesis is divided into six chapters, the first of which is this introduction. Chapter 2 presents a summary of related work for depth segregation and other related fields. The context issues that been studied, investigated and addressed are extensively specified. Chapter 3 introduces the work related to the technique for $2D$ monocular image depth cue detection, focusing on Markov Random Field (MRF) and the cues of occlusion. Chapter 4 presented the problem of $2D$ monocular image depth cue integration; the cues depth occlusion based technique and the formulation of the technique. Chapter 5 presents the results and discussion to the segregated images based depth segregation and a comparative analysis of the enhanced technique performances. Finally, chapter 6 summarizes the major findings of this study and discusses limitations and possible future lines of research and point out the main contribution of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides a relevant historical context for the estimation of depth, ranging from early theory of vision to a present work in computer vision. Section 2.2 introduces the main issues related to depth perception in vision, focusing mainly on the factors that convey information about depth and on the theories attempting to explain their interactions. The connection between depth and object perception is established and the relevance of some monocular depth cues in segregation and grouping is justified from empirical studies. Section 2.3 overviews the work related to the depth segregation in single images in computer-based systems, providing a survey on computational models of cues as well as on algorithms that recover depth from single images. Finally, Section 2.4 gives an overview of the chapter with respect to the state-of-the-art.

2.2 Vision and Depth Perception

Despite the complexity of physical factors that act to generate 2*D* images on the retina, the visual system is remarkably adept at decomposing 2*D* image data into objects and recovering their depth relationships. Regardless of the approach, for centuries, scholars have pondered the mechanism underlying this spatial understanding (Dimiccoli & Salembier, 2009b). Though the debate continues on nearly every aspect of depth perception, some understanding is emerging and the current knowledge is broad. This section reviews only selected aspects of these topics, which are relevant for this thesis.

2.2.1 Image Depth Theory

During the first half of the $20th$ century, three main theoretical perspectives on vision have emerged. Helmholtz, often credited as the founder of the scientific study of visual perception, developed a theory known as unconscious inference (Helmholtz, 2000. Original publication 1866), following which retinal images do not provide direct access to objects because of their intrinsic ambiguity and therefore visual perception is a matter of inferring a probable interpretation for incomplete data. Generally, inference is based on an accumulation of evidences from a variety of cues as well as on a long history of visual experiences, which provides us with the context to interpret images. For example, human ability to interpret carefully crafted images as those in Fig. 2.1 relies on learned knowledge about the objects involved in the scene.

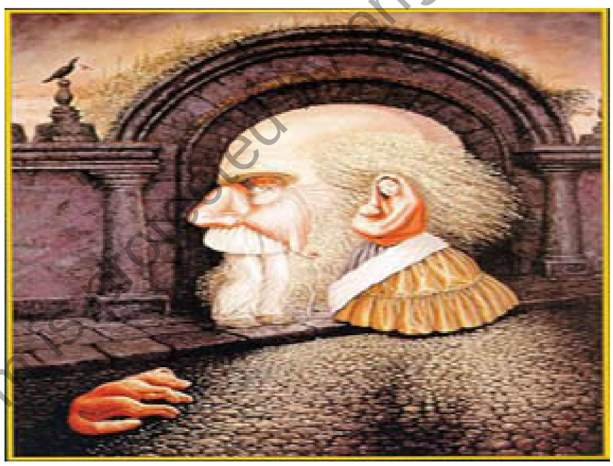

Figure 2.1: Ambiguous images, knowledge of the world learned over many years.

Further in line with interpreting images, (Helmholtz, 2000. Original publication 1866) published his research results, a new paradigm of vision which was developed within the Gestalt School of Psychology, officially initiated in 1912 by a book of (Schultz & Schultz, 2015). The core of it was the idea that the world is a sensible coherent whole, that reality is organized into meaningful parts, and that natural units have their own structure. It is obvious at this point that the human mind can discover these structures, by understanding the internal rules and principles of the phenomenon itself. Under this perspective, Gestaltists consider human perception as the result of a construction pro-