Copyright

by

Serene Banerjee

2004

The Dissertation Committee for Serene Banerjee
certifies that this is the approved version of the following dissertation:

Composition-Guided Image Acquisition

Committee:

Brian L. Evans, Supervisor

Ross Baldick

Alan C. Bovik

Wilson S. Geisler

Joydeep Ghosh

Robert W. Heath, Jr.

Composition-Guided Image Acquisition

by

Serene Banerjee, B.Tech.(Hons), M.S.E.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2004

UMI Number: 3139186

Copyright 2004 by
Banerjee, Serene
All rights reserved.

UMI®

UMI Microform 3139186

Copyright 2004 ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
PO Box 1346
Ann Arbor, MI 48106-1346

This thesis is dedicated to my
Mom, Dad, Sis, and Bips.

Acknowledgments

On completing this dissertation my special thanks goes to my Mother. With her constant inspiration, she always stood up to be a role model that I could follow. Her constant sacrifice to ensure each and every comfort for me and my sister, in spite of her busy schedule, makes her truly the best Mom, ever. I would then like to thank my Father, for his constant support and encouragement, and forever being the lifeguard of our family. Along with Mom and Dad, I would like to thank Durga Didi for her guidance.

I would like to thank my advisor, Prof. Brian Evans, for his all around support. Without his help it would be impossible for me to get a taste of the Western Universities. The inspiration that he imbibed in me, and the tirelessly working example that he has set, will help me move forward. I would also like to thank him for letting me borrow his digital camera to acquire the pictures needed for my dissertation, and for all the times that he went out of his ways to make this research experience a pleasant one.

I would like to extend my thanks to Prof. Ross Baldick, Prof. Alan C. Bovik, Prof. Wilson S. Geisler, Prof. Joydeep Ghosh and Prof. Robert W. Heath, Jr., for their timely feedback and productive discussion, pertaining to my research. This has helped me a lot in formulating my research problem and carrying it out in a focused manner.

I would also like to thank all the Professors at the University for providing me the help, whenever I needed it most. Without the help of all my teachers at IIT Kharagpur (Undergraduate), Kendriya Vidyalaya (Class XI-XII), Hijli High School (Class VI-X), and St. Agnes Primary School (Nursery - Class V) it would be impossible for me to pursue studies. I will forever remain indebted to them for their inspiration and help.

I am grateful to all the people that I have interacted with at Hughes Software Systems, India, Nokia R&D Center, Irving, Ricoh Research Center, California, and Sozotek Wireless, Inc., Austin. Their perspectives has helped me look at any research problem from an application oriented view.

My special prayers will remain for Bijoy Jethu for helping me to appreciate art with a scientific mind. I would also thank Prof. Dennis Darling for giving me more insights on the artistic aspects of my dissertation problem. I am also grateful to all my art and music teachers for helping me to look at the world through an artists' eyes.

I sincerely thank all the staff members at the University whom I was fortunate to interact with. UT's vast success depends on their efficiency, skills, and cheerful personalities.

In every facet of life, I am thankful that very cheerful friends always beautified my days. In order to avoid exceeding the page limit of this dissertation, I am hereby omitting all of their individual names. However, my special thoughts will remain for Biao, Brooke, Christina, Claire, Devendra, Guner, Hamid, Hitesh, Indu, Jenn, Kaushikdas (both of them), Kaustav, Kunal, Magesh, Milos, Niranjan, Parineetha, Sanatanda, Sarit, Tatiana, Tutul, Umesh, Vagdevi and Wade.

I would like to thank God, for blessing me with my sweet Sister, without whose cheerful company, I would not be able achieve anything. She is a constant reminder that life is a joy. I would take this opportunity to thank my hubby to be, Dr. Bipul Das, for his constant support in every respect, and his limitless endurance. With his constant care, I always have the pleasure of walking on a bed of roses. I would also like to thank all other members of my immediate and extended family for making this a good world.

Last but not the least, I would like to thank some very efficient organizations in Austin, who have truly helped making my stay in Austin a pleasant one. They include, Southwest Airlines, Hula Hut, La Tazza Fresca, UT Informal Classes, St. David's Hospital, Austin Emergency Medical Service, and Austin Police Department. I am an admirer of their modes of operation and (most importantly) the friendly smiles of their representatives.

SERENE BANERJEE

The University of Texas at Austin
August 2004

Composition-Guided Image Acquisition

Publication No. ____________________

Serene Banerjee, Ph.D.
The University of Texas at Austin, 2004

Supervisor: Brian L. Evans

To make a picture more appealing, professional photographers apply a wealth of photographic composition rules, of which amateur photographers are often unaware. This dissertation aims at providing in-camera feedback to the amateur photographer while taking pictures. The proposed algorithms do not depend on prior knowledge of the indoor/outdoor setting or scene, and are amenable to software implementation on fixed-point programmable digital signal processors available in digital still cameras.

The key enabling step in automating photographic composition rules is to locate the main subject. Digital still image acquisition maps the 3-D world onto a 2-D picture. By using the 2-D picture alone, segmenting the main subject without prior knowledge of the scene is ill-posed. Even with prior knowledge, segmentation is often computationally intensive and error prone.

This dissertation defends the idea that reliable main subject segmentation without prior knowledge of scene and setting may be achieved by acquiring a single picture, in which the optical system blurs objects not in the plane of focus. After segmentation, photographic composition rules may be automated. In this context, segmentation only needs to approximately and not precisely locate the main subject.

In this dissertation, I combine optical and digital image processing to perform the segmentation of the main subject without prior knowledge of the scene. In particular, I propose to acquire a picture in which the main subject is in focus, and the shutter aperture is fully open. The lens optics will blur any object not in the plane of focus. For the acquired picture, I develop a computationally simple one-pass algorithm to segment the main subject.

The post segmentation objective is to automate selected photographic composition rules. The algorithms can either be applied on the picture taken with the objects not in the plane of focus blurred, or on a user-intended picture with the same focal length settings. This way, in-camera feedback can be provided to the amateur photographer, in the form of alternate compositions of the same scene.

I automate three photographic composition rules: (1) placement of the main subject obeying the rule-of-thirds, (2) background blurring to simulate the main subject being in motion or decrease the depth-of-field of the picture, and (3) merger detection and mitigation when equally focused main subject and background objects merge as one object.

The primary contributions of the dissertation are in digital still image processing. The first is the automation of segmentation of the main subject in a single still picture assisted by optical pre-processing. The second is the automation of main subject placement, artistic background blur, and merger detection and mitigation to try to improve photographic composition.

Contents

Acknowledgments **v**

Abstract **viii**

List of Tables **xiv**

List of Figures **xv**

Chapter 1 Introduction **1**

1.1 Motivation 1

1.2 Proposed Problem 5

1.3 Thesis Statement 8

1.4 Contributions 8

1.5 Notation 10

1.6 Dissertation Outline 11

Chapter 2 Background **14**

2.1 Introduction 15

2.2 Main Subject Detection with Bayes Nets 18

2.2.1 Algorithm Description 18

2.2.2 Shortcomings 19

2.3 Wavelet Based Main Subject Detection 19

2.3.1 Algorithm Description 19

2.3.2 Shortcomings 20

2.4 Main Subject Detection with Variance Maps 21

2.4.1 Algorithm Description 21

2.4.2 Shortcomings 22

2.5 Need for In-camera Algorithms 22

2.6 Conclusion 23

Chapter 3 Main Subject Segmentation 24

3.1 Introduction 25

3.2 Segmentation in Natural Images 26

3.3 Optical Model for Cameras 29

3.4 Formulation of Main Subject Detection 31

3.5 Algorithm for Main Subject Segmentation 34

3.5.1 Sharp Region Identification 35

3.5.2 Edge Detection 37

3.5.3 Contour Detection 39

3.5.4 Active Contours and the Gradient Vector Flow Algorithm 42

3.5.5 Modification of Mask Based on Difference Between Original and Supplementary Picture 44

3.6 Implementation Complexity 45

3.7 Main Subject Segmentation Results 49

3.8 Comparison with Prevalent Segmentation Methods 53

3.8.1 Comparison of Multiresolution-based (Wavelets) and Pixel-based Main Subject Detection 54
3.9 Conclusion 57

Chapter 4 Automation of Photograph Composition Rules 73
4.1 Introduction 74
4.2 Rule-of-Thirds: Automated Placement of the Main Subject . . 76
4.2.1 Algorithm Formulation 78
4.2.2 Proposed Algorithm 78
4.2.3 Implementation Complexity 80
4.2.4 Results 82
4.3 Motion Effects Rule: Simulating Background Blurring 95
4.3.1 Algorithm Description 95
4.3.2 Implementation Complexity 96
4.3.3 Results 96
4.4 Conclusion 97

Chapter 5 Merger Detection and Mitigation 99
5.1 Introduction 100
5.2 Merger Detection and Mitigation: Formulation 102
5.2.1 Background Segmentation 102
5.2.2 Merger Detection 107
5.2.3 Selective Blurring 115
5.3 Implementation Complexity 116
5.4 Results 121
5.5 Conclusion 122

Chapter 6 Conclusions **124**

6.1 Conclusions 124

6.2 Future Work 129

6.2.1 Lower Complexity Image Registration 129

6.2.2 Automation of Other Photographic Composition Rules 129

6.2.3 Extension for Video Acquisition Applications 130

6.2.4 Developing Algorithms for Image Stabilization 131

6.2.5 Inclusion of Other Cues in the Proposed Framework for Better Image/Video Acquisition 131

Appendix A Auto-Focus Filter **133**

Appendix B Distance Transform **134**

Bibliography **135**

Vita **150**

List of Tables

1.1 Acronym list. 12

2.1 Attributes for image appeal measures. 17

3.1 Segmentation accuracy measures. 52

5.1 Implementation complexity for merger mitigation. 121

List of Figures

1.1 Sample cave painting. 2

2.1 Framework for Wang's *et al.* algorithm. 20

3.1 Greater depth of field. 28
3.2 Shallow depth of field. 28
3.3 Optical model for cameras. 29
3.4 Cartoon for high depth of field. 30
3.5 Cartoon for shallow depth of field. 31
3.6 Image sharpening filter. 37
3.7 Highpass filter frequency response. 38
3.8 Lowpass filter frequency response. 39
3.9 Limitation example. 40
3.10 Supplementary image and mask for registration. 46
3.11 Possible image taken by user (1). 47
3.12 Possible image taken by user (2). 48
3.13 Framework for main subject detection. 49
3.14 Image with no edge. 56
3.15 Image with a strong edge. 57

3.16 Image with weak edges. 58
3.17 Segmented circle with pixel-based approach. 59
3.18 Segmented ramp with pixel-based approach. 59
3.19 Laplacian pyramid decomposition of an image with strong edge. 60
3.20 Laplacian pyramid decomposition of an image with blunt edge. 61
3.21 Example of main subject detection. 62
3.22 Automated main subject detection (Example 2). 63
3.23 Automated main subject detection (Example 3). 63
3.24 Automated main subject detection (Example 4). 64
3.25 Automated main subject detection (Example 5). 64
3.26 Automated main subject detection (Example 6). 65
3.27 Automated main subject detection (Example 7). 65
3.28 Automated main subject detection (Example 8). 66
3.29 Automated main subject detection (Example 9). 66
3.30 Automated main subject detection (Example 10). 67
3.31 Automated main subject detection (Example 11). 67
3.32 Comparison of the proposed and previous methods for main subject detection (Example 1). 68
3.33 Comparison of the proposed and previous methods for main subject detection (Example 2). 69
3.34 Comparison of the proposed and previous methods for main subject detection (Example 3). 70
3.35 Comparison of the proposed and previous methods for main subject detection (Example 4). 71
3.36 Comparison of the proposed and previous methods for main subject detection (Example 5). 72

4.1 Main subject in the middle. 75
4.2 Main subject placed following rule-of-thirds. 76
4.3 Foreground and background are equally focused. 77
4.4 Picture with background blurred. 77
4.5 Rule-of-thirds illustration. 79
4.6 Automated replacement following rule-of-thirds and background blurring (Example 1). 84
4.7 Automated replacement following rule-of-thirds and background blurring (Example 2). 85
4.8 Automated replacement following rule-of-thirds and background blurring (Example 3). 86
4.9 Automated replacement following rule-of-thirds and background blurring (Example 4). 87
4.10 Automated replacement following rule-of-thirds and background blurring (Example 5). 88
4.11 Automated replacement following rule-of-thirds and background blurring (Example 6). 89
4.12 Automated replacement following rule-of-thirds and background blurring (Example 7). 90
4.13 Automated replacement following rule-of-thirds and background blurring (Example 8). 91
4.14 Automated replacement following rule-of-thirds and background blurring (Example 9). 92
4.15 Automated replacement following rule-of-thirds and background blurring (Example 10). 93

4.16 Automated replacement following rule-of-thirds and background blurring (Example 11). 94

5.1 Example of an image modeled as a Gaussian mixture of hue values. 105
5.2 Example of a merger. 108
5.3 Histogram of image background. 109
5.4 Segmented background objects. 110
5.5 Contribution from distance transform coefficient for a floating-point implementation. 112
5.6 Contribution from distance transform coefficient for a fixed-point implementation. 114
5.7 Distance transform and first level Gaussian pyramid coefficients. 118
5.8 Merger mitigated image. 119
5.9 Merger mitigation framework. 120
5.10 Example of merger mitigation. 122

6.1 Proposed framework for main subject detection. 126

Chapter 1

Introduction

"The best place to start is from the beginning."

L. Frank Baum, The Wizard of Oz.

1.1 Motivation

Images have been the most primitive and intuitively the most effective form of expression in human civilization. Through the ages, the human cognitive system has shown its proficiency in communicating and expressing thoughts through this medium. Even when human civilization was in its infancy, during the upper Palaeolithic age (40,000 - 10,000 BC), the medium of communication and documentation were found to be images on caves, as is evident from the paintings of Altamira caves which date back to nearly 11,000 BC. The famous image of the bison (Fig. 1.1) painted on cave walls document their livelihood, hunting passion and the danger they faced at every step. As these images convey the story of a long forgotten era, many other expressions and messages in the modern world can be communicated through this medium.

Figure 1.1: An example of cave paintings in 11,000 BC that documents lifestyles of the people in the upper Palaeolithic age (40,000 – 10,000 BC).

Modern day civilization has also conceived the importance of pictures to communicate more effectively, preserve memories and communicate in the absence of someone. However, the main challenge was to devise a portable technology for capturing and communicating images. The industrial revolution in United States and Europe opened the floodgates to an entirely new set of technology that has brought a paradigm shift in the concept of image acquisition and communication. Thanks to modern inventions such as photography (1824), fax (1843) [1], movies (1891), television systems (1924), computer (1945), videophones (1960), cell phones (1973), camcorders (1980), and finally the World Wide Web (1990), which have opened up huge potential for communicating images. The present era is blessed with immense wealth of image communication for both still images and video.

Some of the basic steps involved in image and video communication can be divided into

- Image/video acquisition

- Representation of the acquired data in a compressed form
- Storage and transmission of this representation
- Decompression to retrieve the original image/video
- Correction of possible errors occurring during storage and transmission and
- Display of the image/video content

These stages can be individually or jointly optimized so that the effective throughput can be maximized with limited resources. Since the scope of the work presented in this dissertation is restricted to image acquisition stage, so, any discussions about the subsequent stages have been avoided.

Image acquisition research has been specific to an application area. The form of image acquisition depends upon (a) the interest of the application, (b) the location of object to be captured in the picture, (c) resolution of the object, and (d) material property of the object and the surrounding objects and background. For example, different image acquisition techniques are used in optical photography, aerial photography, medical imaging, archeology, etc. Also the object of interest in the different modes of image acquisition varies widely.

Perhaps the most widely used form of image acquisition that conveys information about the world at large is optical photography. This mode of image acquisition can capture a situation, a mood and an instant to make it long lasting, even immortal. The work proposed in this dissertation aims at general optical photographic image acquisition.

The digital camera is becoming mainstream in photography and serves a diverse population. The scope of flexibility in terms of image quality assessment is higher in digital cameras than analog cameras. The user can immediately view the acquired picture. The user can take many pictures of a scene (e.g. by changing camera settings) and delete poor quality images. The automatic settings on digital still cameras can invoke an auto-focus filter, set shutter aperture size and speed, and so forth to aid the amateur photographer in taking better pictures. Professional photographers know how to choose the settings to achieve high quality pictures with artistic effects. When amateur photographers are not satisfied with the pictures obtained in automatic modes, they face confusion and uncertainty in how to change camera settings to improve photographic composition.

One of the aspects in improving photographic composition is positioning the main object based on the content of the whole frame. Secondly, the choice of highlighting the desired objects makes an immense difference in photographic appeal. And, finally prominence and misplacement of unwanted objects in the image frame deteriorates the image quality as a photograph. Although there is no proven objective measure for assessing photographic appeal, studies [2] provide some rules for composition to make a photograph more appealing.

The proposed research concentrates on the image acquisition stage. This dissertation addresses the aforementioned challenges of object placement, by emphasizing the desired object(s) and removing and blurring unwanted object(s) to produce more appealing pictures. Before acquiring the image, if the perceived distortions or degradations can be corrected, then the performance of the subsequent stages in the image communication chain could automatically improve. At the same time, every user wants flexibility to exercise freedom in acquiring pictures. With these considerations, the proposed framework gives users freedom to make their own selection from a list of suggested alternative pictures.

I propose algorithms that are amenable to real-time implementation in digital cameras and that aid the amateur photographers in taking pictures by automating selected photographic composition rules. Other potential applications of this work include security and compression. The proposed still image framework could be extended to optical video acquisition.

Section 1.2 proposes the problem that this dissertation will address. Section 1.3 states the thesis statement. Section 1.4 discusses the contributions of this dissertation. Section 1.5 describes in short the notation used in this dissertation. Section 1.6 discusses the outline of the subsequent chapters.

1.2 Proposed Problem

By applying creative instincts and rules of photography [2], professional photographers can generate outstanding pictures. As mentioned previously, the notion of good photography is subjective. However, several underlying principles underlie professional photography. Guiding amateur photographers by some of these principles could improve one's photography skills immensely. A summary of these rules include

- Placing the main subject in the picture based on the content of the image. In general positioning the main subject(s) at one of the four one-third positions of the frame (i.e. the rule-of-thirds) provides appealing images. This way it captures the main subject and the context and content effectively.

- Removing or blurring unwanted prominent objects near or connected to the main subject(s) to reduce degradation in picture quality.
- Distinguishing foreground from background effectively.
- Photographing action pictures by blurring the background.
- Taking a close up action picture with telephoto lenses.
- Using lines in the picture for interest and unity, e.g. the main subject could be placed where the lines in the scene intersect. This automatically draws attention of the viewer to the main subject.
- Taking picture through frames available on the scene.
- Using the "best" camera angle, e.g. by using a low–angle to photograph active people or by using a low–angle to make the background be uncluttered sky, and
- Taking a picture that is well balanced on the eye, e.g. both the wheels of a photographed cart needs to be in the image frame to create a sense of balance.

Provided that the main subject(s) is(are) defined, it is most likely that the prominence of the main subject(s) is(are) preferred when compared to the background.

This research focuses on devising a framework for automating selected photographic composition rules to improve the quality of pictures taken by amateur photographers. Within the framework, I automate three photographic composition rules: placing the main subject(s) in a more suitable position on the frame, providing prominence to the main subject(s), and removing unwanted prominent objects. The image acquisition subsystem would provide alternative pictures that follow photographic composition rules and that are computed while acquiring pictures. For implementation in a digital still camera, I develop low-complexity algorithms that automatically segment the main subject(s), apply selected photographic composition rules, and provide flexibility to the photographer for manual intervention. Beyond personal use, such a smart camera system might be useful to professionals who need to take pictures for documentation, such as realtors and architects.

The main tasks involved in solving the proposed problem are

- Segmenting the main subject(s) in the photograph, and
- Postprocessing the image to place the main subject(s), blur the background and remove merged objects.

Segmentation of objects, without prior knowledge of the scene setting or content from a 2-dimensional picture, is an ill-posed problem. Even with *à priori* knowledge about the scene setting, segmentation is difficult and error prone. Instead, *à priori* knowledge of the optical settings makes the segmentation tractable. Postprocessing after the main subject(s) segmentation provides content-based background blurring and placement of the main subject(s). A region and inhomogeneity based segmentation is again performed for identifying the unwanted background objects in the image frame.

1.3 Thesis Statement

This dissertation defends the following idea:

> *Reliable main subject segmentation without prior knowledge of scene content and setting may be achieved by acquiring a picture, where the optical system blurs objects that are not in the plane of focus. After segmentation, selected photographic composition rules may be automated.*

1.4 Contributions

The main contributions of this dissertation are in the development of a smart image acquisition framework [3, 4, 5, 6]. Within this framework, we have developed the following algorithms:

- *Algorithm to segment the main subject(s) from the image frame based on frequency information:* I propose to use the camera optics to obtain a supplementary picture that has the main subject in focus, and objects that are not in the plane of focus are blurred. Depending on where the user points the camera, the auto-focus filter has the main subject in focus. The shutter aperture is then fully opened to blur out the background by using diffused light. This supplementary picture taken by the camera has a frequency content difference between the main subject and the background objects. I propose to utilize this frequency content information in segmenting the main subject. The proposed algorithm works independently of assumptions on scene setting or content.

- *Algorithm for automatic placement of the segmented main subject(s) in an advantageous position following the rule-of-thirds:* After segmentation, I have developed and implemented an algorithm to automate the rule-of-thirds, which is a guideline that professional photographers use to place the main subject on the canvas. Here, the image frame is defined as an imaginary grid defined in Euclidean coordinates that divide the image into three equal parts in horizontal and vertical directions. The objective is to place the main subject in one of the four places where these imaginary lines intersect.

- *Algorithm for adding simulated blur on the background depending on the content to make the image more appealing:* Artistic effects can be added on the image background, after segmenting the main subject. In this work I have proposed to add simulated blurs to the image background. These blurs can either simulate motion blurs when the main subject or the camera is in motion, or reduce the depth of field.

- *Algorithm for segmentation of unwanted mergers with the main subject(s) by using region and frequency based information:* A merger occurs in an image where the main subject is in focus, and an equally sharp background object appears to be a part of the main subject. Professional photographers avoid such a merger for a more appealing picture.

- *Algorithm that automatically identifies the background object that tends to merge with the main subject:* The proposed algorithm is a sub-optimal solution to detect a merger. Upon detection of the merger, the unwanted background object is artificially blurred so that is appears to be at a further distance away from the camera.

I have developed the proposed algorithms with low implementation complexity in mind so that the algorithms could be implemented in a digital still camera. In particular, the proposed algorithms take one pass over the image, have low computational complexity, and are amenable to fixed-point arithmetic. While the user is taking a picture, the camera would provide to the user alternative pictures that follow photographic composition rules. The user could then pick the picture that he/she likes the best. The software and color images for this paper are available at

http://www.ece.utexas.edu/~bevans/projects/dsc/

1.5 Notation

In this section, I briefly introduce the mathematical notation followed in this dissertation. The notations are also discussed in detail in the later chapters as they are used.

I denote a scene, $\mathcal{C} = (C, F)$, in an n-dimensional Euclidean space, $\Re^n$, where C is the scene domain and F gives the intensity function. The derived expressions are described in the n-dimensional Euclidean space, and later applied to the 2-dimensional Euclidean space. So, in the 2-dimensional space, the scene domain, C, is given as $C = \{\mathbf{v} | \mathbf{v} \in \text{Image frame}\}$ and $\mathbf{v} = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$ are the set of pixel positions.

When the scene is divided into object and background classes, the object and background class features are denoted by $(.)_o$ and $(.)_b$, respectively. For example, F_o and F_b are the intensity functions representing the object and background classes, respectively, in the scene $\mathcal{C}$ and $F_o \subset F$ and $F_b \subset F$.

I define a transformed scene $\mathcal{G} = (C, G)$ that has been obtained by using a transform over the intensity function on the scene domain. In the context of this dissertation, the transform is a gradient operator ∇. The intensity function F undergoes an onto transformation into the transform scene, $\mathcal{G} = (C, G)$, so that $G = \nabla F$. Thus, for any location $c \in C$, the mapping of $\{F(c) \in F | c \in C\}$ to $\{G(c) \in G | c \in C\}$ can be expressed as $G(c) = \nabla F(c)$. The gradient function, G can be further partitioned as G_H and G_L, where $G_H(c)$ and $G_L(c)$ represent the high and low frequencies, respectively.

An image in the 2-dimensional space is denoted by $I(x, y)$ of dimension $N \times M$ pixels. I define $I_{smooth}(x, y)$ as the image comprising of the intensity function derived from low frequency components, i.e., the smoothed image derived from the original image $I(x, y)$. Similarly, $I_{sharp}(x, y)$ denotes a sharpened version of the original image $I(x, y)$.

I denote a probability distribution by $P(.)$. The mean and the variance of the distribution are defined by μ and σ, respectively.

The acronyms used in this dissertation are listed in an alphabetical order in Table 1.1.

1.6 Dissertation Outline

The thesis is organized as follows:

Chapter 2 briefly overviews previous research in main subject detection and evaluation of image appeal. It also discusses the major challenges and shortcomings of the prevalent approaches, and illustrates the need for low-complexity algorithms for main subject segmentation.

The theory for main subject detection by using a gradient based ap-

CIELab	Commission Internationale d'Eclairage (Luminance, Position between red and green, Position between yellow and blue)
DT	Distance Transform
FIDT	Frequency Inverse Distance Transform
GVF	Gradient Vector Flow
HSV	Hue, Saturation, Value
JPEG	Joint Photographic Experts Group
MAP	Maximum A Posteriori
MSD	Main Subject Detection
RGB	Red, Green, Blue

Table 1.1: Alphabetically ordered list of acronyms used in this dissertation.

proach is presented in Chapter 3. The proposed theory uses the strength of inhomogeneity in the image frame to suppress the background and eventually segment the main subject(s). The algorithm derived from the proposed theory is also detailed in Chapter 3. The algorithm has been tested on a number of images and results are compared with other algorithms to show the effectiveness of this algorithm in an online environment. The developed algorithm has low implementation complexity and is amenable to fixed-point implementation on digital signal processors, which are commonly available in digital still cameras.

Chapter 4 describes two proposed algorithms. The first automatically places the main subject by following the rule-of-thirds. The second simulates background blurring depending on the content of the image. These increase the appeal of the acquired picture.

Chapter 5 presents merger detection and mitigation of the main subject with the background objects. An algorithm is proposed for segmentation of unwanted object near or spatially connected to main subject(s) based on region and gradient information. Implementation results are illustrated to show the strength of the proposed algorithm. The proposed algorithm mitigates the visibility of unwanted background objects.

Chapter 6 concludes the dissertation by highlighting the major contributions in the proposed research and discussing avenues for future work.

Chapter 2

Background

"The seeds of great discoveries are constantly floating around ..."

Joseph Henry

A photograph can be thought of as a medium with which the photographer communicates with a viewer. Photographers take pictures of people, objects or events to preserve memories and share the experience with others. Usually the photographer tends to convey the message of the photograph by having one or more main subjects in the picture. A system for automatically identifying the main subject(s) in the picture helps in capturing the main theme of the picture. This chapter summarizes previous research in main subject detection. Also, the shortcomings of the previous approaches are highlighted and the need for an in-camera main subject detecting algorithm is discussed. The ability to reliably detect the main subject provides a measure of saliency of relative importance of the different objects in a picture. So, this information may also be helpful for image/video display, enhancement, and content-based retrieval.

2.1 Introduction

Digital still cameras continue to include more and more advanced features to help the user take better pictures. As image quality assessment is quite subjective, two main streams of research are to (a) find measures to evaluate image [7, 8, 9, 10, 11, 12, 13, 14] and video [15, 16, 17] quality such that it corresponds with human judgement, and (b) develop methods so that more visually appealing pictures can be produced. This section will summarize research in the second category, in the context of photography. This research could potentially be extended for better image/video display on resource constrained mobile devices [18], automatic detection of main subject for sign language communication [19] and image and video retrieval [20, 21, 22].

Corey, Clayton, and Cuprey [23] demonstrated that perceived image quality is scene–dependent. Their experimental results show that the distance of the main subject from the camera has a significant effect on perceived image quality. Their work quantifies the bias in image quality perception of the main subject that occurs with the change in camera–to–subject distance (or magnification).

Biederman [24] promoted the idea that patterns and scenes are recognized with individual basic shapes and entities, and their spatial inter–relations. Studies showed that for very high objective quality portraiture, subjective quality decreased, as objective quality increased. For this reason photographers often use diffusing screens over lenses to soften the image. These results show that photographic appeal lies beyond objective quality metrics, and relates to image understanding.

Savakis, Etz, and Loui [25] conducted a subjective test by third–party evaluators to evaluate image appeal in consumer photography. As expected, their results show that perceived image quality depends on *people*, *composition*, and *subject of the photograph*, as well as *objective measures*. Table 2.1 gives their 38 criteria, which can be classified into the four aforementioned groups. The positive or negative effect defining the contribution of each factor is indicated in the rightmost column.

The above work explains the need for development of a new set of measures of image appeal. The first step to the development of image appeal measures lies in scene classification [26, 27, 28, 29, 30, 31, 32] and automated main subject detection [33, 34, 35]. After that, different regions of the image can be assigned their relative importance. This is closely related to region–of–interest [36] detection, and a discriminative treatment can be applied for image understanding, image enhancement, and constrained transmission.

The subsequent parts in this chapter discuss previous research for detecting the main subject in offline settings. Section 2.2 describes a neural network approach to detect the main subject for auto-album layout. Section 2.3 discusses a wavelet based method to compute the focused regions in an image from low depth-of-field pictures. Section 2.4 is an iterative solution for the same that is based on the variance map of the image. Based on the discussions of previous research, Section 2.5 describes the need for low-complexity algorithms to detect the main subject in-camera. This dissertation develops an algorithm to meet this need.

Categories	**Attributes**	**Index**
People/expression	People	44
	Whole group in photo	43
	Close–up	31
	Facial expression	28
	Personality/unusual	17
	Pose	16
	Baby	9
	Can see faces	9
	Kids	9
	Seniors	4
	Bride	3
	No irrelevant people	1
	Unflattering pose	-7
	No one facing camera	-9
Composition/subject	Composition	50
	Shows location	24
	Representative of event	22
	Shows action/fun	18
	Specific inanimate subject	18
	Panoramic	9
	Whole landscape	7
	Balance	3
	Historical subject	3
	Subject too far away	-12
	Occlusion	-13
	Boring	-20
	Poor composition	-23
Objective measures	Colorfulness	16
	Lighting	6
	Sharpness	4
	Good contrast/brightness	4
	Blurry	-7
	Low quality	-10
	Poor contrast/brightness	-20
Redundancy	Duplicates of lower quality	-10
	Duplicates	-21
	Same subject elsewhere	-31
	Redundant	-36

Table 2.1: Attributes that can positively or negatively influence image appeal, being categorized into four main classes, namely, People/expression, Composition/subject, Objective measures and Redundancy.

2.2 Main Subject Detection with Bayes Nets

Luo, Etz, Singhal, and Gray [34] developed a computational approach to main–subject detection by using Bayes neural network. Their algorithm is described and its shortcomings are highlighted in the subsequent subsections.

2.2.1 Algorithm Description

Their algorithm is performance–scalable so that it need not be reconfigured for different sets of images, and involves (a) region segmentation, (b) perceptual grouping, (c) feature extraction, and (d) probabilistic reasoning and training. An initial segmentation is obtained based on the homogeneous properties of the image such as color and texture. False boundaries are removed with perceptual grouping of identifiable regions such as flesh tones, sky, and tree. Then, geometric features are extracted, including centrality, borderness, shape, and symmetry. The probability density function for the main subject location is estimated from the training data. Say there are $n \in \{1, ..., N\}$ observers and $k \in \{1, ..., K\}$ training images. Then a pixel (i, j) will be identified as the main subject with probability $p_k^n(i, j) = 1$, if it is in the main subject, and 0 otherwise. The probability that a pixel (i, j) is identified as a main subject can be determined as

$$P_k(i, j) = \frac{1}{N} \sum_{n=1}^{N} p_k^n(i, j) \tag{2.1}$$

The probability density function estimate can be applied to the unknown test set to guess what and where the main subject is.

2.2.2 Shortcomings

The Bayes Nets based method requires training time, and is not a low–complexity solution for detecting the main subject on the fly. Also, as this is a Bayes net based approach, the system performance will be poor if the test set is very different from the training examples. With the vast number of possibilities of scene content, scene settings, and user preferences, developing a good set of training examples to guarantee that the neural network would perform well for a varied number of circumstances is difficult.

2.3 Wavelet Based Main Subject Detection

Wang, Li, Gray, and Wiederhold [37, 38] proposed a wavelet based approach to detect the focused regions in an image from low depth-of-field pictures. Their algorithm is described and the shortcomings are listed in the subsequent parts.

2.3.1 Algorithm Description

In their proposed wavelet-domain approach, Wang, Li, Gray, and Wiederhold [37, 38] analyzed the statistics of the high-frequency wavelet coefficients to segment the focused regions in an image, thereby detecting the object-of-interest. Initially, the image is coarsely classified into object-of-interest and background regions by using the average intensity of each image block, and the variance of wavelet coefficients in the high frequency bands. The variance is higher for the focused regions in the image. Blocks are clustered by using k-means algorithm [39] by noting that blocks from a homogeneous image region will have similar average intensities. Each block is further subdivided into

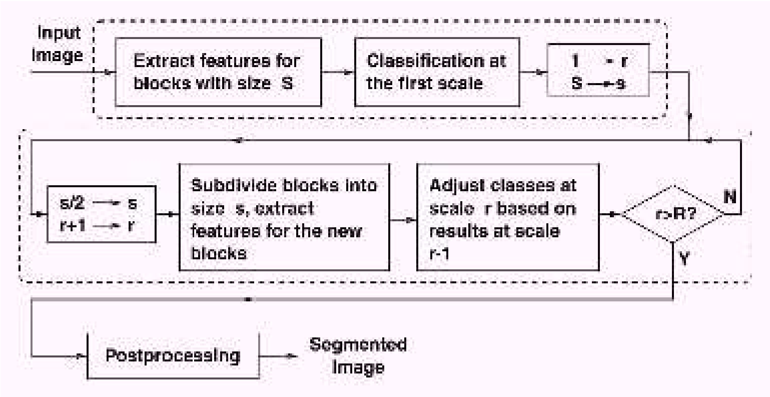

Input Image
Extract features for blocks with size S
Classification at the first scale
1 → r
S → s
s/2 → s
r+1 → r
Subdivide blocks into size s, extract features for the new blocks
Adjust classes at scale r based on results at scale r-1
r>R?
N
Y
Postprocessing
Segmented Image

Figure 2.1: Framework of detecting the main subject by using Wang's *et al.* wavelet based algorithm.

child blocks, and a multiscale context-dependent classification is performed for further refinement. Finally, a post-processing step removes small isolated regions and smoothes the boundaries. The framework for their algorithm is shown in Fig. 2.1.

2.3.2 Shortcomings

The segmentation accuracy of the wavelet-based segmentation algorithm is acceptable with the segmentation error varying between 4 to 7%. However, the method uses Haar wavelets, which have transfer functions that are scaled versions of $1+z^{-1}$ and $1-z^{-1}$, for the lowpass and highpass filters, respectively. The Haar wavelets and feature extraction can be implemented in fixed-point arithmetic. Nonetheless, the k-means clustering step, and further refinement through context-dependent classification in the multiscale wavelet-domain, is computationally intensive. For a straightforward implementation of the k-means clustering step, the computationally intensive part is in computing the nearest neighbors [39].

2.4 Main Subject Detection with Variance Maps

Won, Pyan, and Gray [40] developed a spatial domain approach to detect the focused region from low depth-of-field pictures. The algorithm description and its shortcomings follow.

2.4.1 Algorithm Description

In a spatial-domain approach, Won, Pyan, and Gray [40] developed an iterative algorithm based on variance maps. A local variance map is used to measure the pixel-by-pixel high frequency distribution in the image. This variance map has blob like errors both in the foreground (where the image is relatively smooth) and the background (where the background is highly textured) regions. To eliminate these errors, the authors employ a block-wise maximum *à posteriori* image segmentation. The local variance image is first divided into non-overlapping $B \times B$ image blocks. For each block, the average of the local variance is denoted by $\bar{y}_t$, and assigned a class label $\bar{x}_t$. The blocks are classified according to the following criterion

$$\bar{x}_t^{(n+1)} = \mathrm{argmax}_{\bar{x}_t} P(\bar{y}_t|\bar{x}_t^{(n)}) P(\bar{x}_t|\bar{x}_{\eta t}^{(n)}) \tag{2.2}$$

where n represents the number of iterations and η_t is the neighborhood of block t. After an initial blockwise segmentation the results are further refined to obtain a pixel-wise segmentation, by using the watershed algorithm [41, 42]. Their method yields more accurate segmentation when compared to the wavelet based approach in Section 2.3 [37, 38].

2.4.2 Shortcomings

The block-wise maximum *à posteriori* segmentation produces more accurate results compared to the wavelet-based segmentation. However, it requires recursion over image blocks and is computationally demanding. Further refinement of the segmentation by using the watershed algorithm [41, 42] adds to the implementation complexity.

2.5 Need for In-camera Algorithms

Previous research to detect the main subject with a computationally intensive algorithm may be appropriate for offline applications, such as image indexing for content-based retrieval [21, 22, 20], object-based image compression for image servers [36], and for content grouping for auto-album layout [31, 30]. However, one of the major challenges is to provide online feedback to the photographer, while a picture is being acquired. The previous methods do not interact with the image acquisition process and instead are applied after the image has been acquired. My proposed method manipulates the optical subsystem in the image acquisition system to blur out the objects not in the plane-of-focus. The optical filtering makes the problem of main subject segmentation well posed and simplifies the digital processing to detect the main subject.

The blurred picture is taken right after or right before the user-intended picture. The processing delay should be very short so as to give feedback to the user quickly enough so that the user can decide whether or not to try the picture again. The feedback is in the form postprocessed alternate pictures in addition to the picture the user intended to take. For a cost-effective implementation of main subject segmentation, on which the computations of the alternate pictures are based, the proposed algorithms must be of low complexity. For implementation in a digital still camera, the algorithms must be implementable in fixed-point arithmetic and with a low memory footprint.

The research reported in this dissertation provides a low-complexity solution to detect the main subject. I propose a low-implementation complexity one-pass segmentation algorithm based on gradient information in Chapter 3. The proposed algorithm can be implemented in fixed-point data arithmetic on digital signal processors, available in digital still cameras.

2.6 Conclusion

This chapter summarizes the previous research in main subject detection and explains the need for an in-camera algorithm. Luo, Etz, Singhal, and Gray [34] propose a neural network for main subject detection for auto-album layout. Wang, Li, Gray, and Wiederhold [37, 38] analyze statistics of wavelet coefficients to detect the main subject. Won, Pyan, and Gray [40] process the variance map of the image with their proposed iterative algorithm for main subject detection. The previous research have high implementation complexity and may be more suitable to detect the main subject in offline settings.

The subsequent chapters present a new low-complexity and one-pass algorithm for main subject detecting in photographs. After main subject detection, selected photographic composition rules are automated for presenting alternative well composed pictures.

Chapter 3

Main Subject Segmentation

"Imagination is more important than knowledge..."

Albert Einstein

When taking pictures, professional photographers employ a variety of composition rules. In automating these rules, it is often first necessary to detect and segment the main subject. I propose a detection and segmentation algorithm that leverages the optics in a digital still camera. Based on where the user points the camera, an auto-focus filter first puts the main subject in focus and takes a picture. Before or after the user takes the picture, I take a second picture. In this supplementary picture, I open the shutter aperture to diffuse light from objects that are out-of-focus, which blurs the background. The resulting difference in the frequency content of the main subject and the background in the supplementary picture is then used by the proposed algorithm to detect and segment the main subject. The algorithm does not depend on prior knowledge of the indoor/outdoor setting or scene content. Algorithm complexity is similar to that of a 5×5 filter.

3.1 Introduction

Segmentation is considered as one of the most salient tasks in image processing. With the advancement of imaging techniques, the necessity for segmentation is growing. Many segmentation techniques have been reported over the last few decades [43, 44, 45, 46, 47, 48]. However, the major challenge associated with image segmentation in any application is the *ill-posed* property of the problem itself. The ill-posed property comes from the projection of the 3-dimensional world onto a 2-dimensional image. Successful segmentation approaches are generally application-dependent. Since the domain of interest dealt with in this dissertation is digital photographic images, the primary thrust area is natural image segmentation. Thus, I will restrict all the discussion on this domain only and the research also has been carried on with natural images in focus.

Section 3.2 discusses the possible choices and selection of features to be used for main subject segmentation. Section 3.3 discusses the optical model for the camera. Section 3.4 formulates the theory. Section 3.5 describes the proposed algorithm. Section 3.6 measures implementation complexity of the proposed algorithm. Section 3.7 presents the segmentation results and discusses quantitative measures for evaluating segmentation accuracy. Section 3.8 compares the segmentation results of the proposed algorithm with the previous research described in Chapter 2. Section 3.9 concludes the chapter.

3.2 Segmentation in Natural Images

Natural images in general are composed of a wide variety of objects and background. This poses a major challenge to natural image segmentation and renders it yet more ill-posed, as it involves effective reduction of unnecessary regions while keeping the user defined important ones [44, 45, 46, 49]. Segmentation in general is usually accomplished from the knowledge of the various image features. The set of image features can be very broadly classified into *region*, *boundary features* and *shape information.* For natural image segmentation, human beings use both region and boundary cues as suggested by psychophysics experiments [50].

Region-based segmentation focuses on integrating features such as intensity, texture, color and similar parameters, which distinguishes a region from another [43, 49]. These approaches are motivated by Gestalt's notion that similar regions can be grouped together. However, these region–based approaches suffer from problems when two regions belonging to the same object in the image have quite different properties based on shading or perspective changes. In such cases, gradient–based approaches to segment the image based on local edges perform well [46, 51, 52]. Another main area of research is the use of shape-based features, which uses a mean model to represent a shape and allow some deviation from the mean model to accomplish the best fit to the shape that is to be segmented [53, 54, 55, 56]. A modified version of that approach uses region-based information along with shape information for segmentation [57, 58, 59, 60].

Photographic images are formed by the interaction of light upon the object surface. Two entirely different objects can have the same set of region features, while differing from one another in other properties. Thus, segmentation based only on region features is not viable for natural image segmentation. Addition of other features could aid the process of segmentation. A shape or morphological information along with the region–based features can be effective. However, introduction of shape information restricts the domain of application. Moreover, optimization of these features is dependent on training.

On the other hand, gradient information may be a more reliable feature for segmentation in certain applications. In this dissertation's application, the use of gradient information also removes the ambiguity that is prevalent in the region–based approaches. Another advantage with the gradient–based segmentation is that it does not depend on any *à priori* knowledge. Moreover, gradient–based segmentation is amenable to one-pass, fixed-point implementations.

In the problem of main subject detection, the image has to be divided into two separate classes: main subject and background. I propose to change the optical settings in the camera to take a supplementary picture, either before or after the amateur photographer acquires a picture. In this supplementary picture, the properties of the local edges are different between the main subject and the background. The main subject region (in focus) has crisp gradients whereas the background (blurred) has low-intensity gradients. Thus, I propose to use this property for unsupervised segmentation of the main subject. For example, Fig. 3.1 shows a picture taken in which the main subject (the central flower) and the background are equally focused. Fig. 3.2 is the same picture with a wider shutter aperture. Here the main subject has distinguishable image edge features compared to the background.

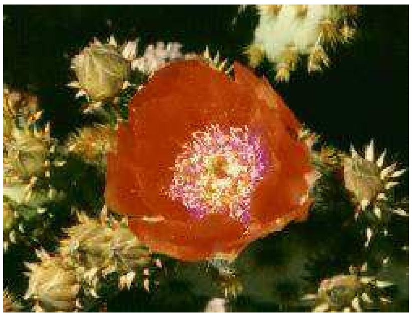


Figure 3.1: Greater depth of field obtained from a wider shutter aperture. In this picture, the shutter aperture is F22, and shutter speed is $\frac{1}{60}$. The picture was obtained from the World Wide Web.

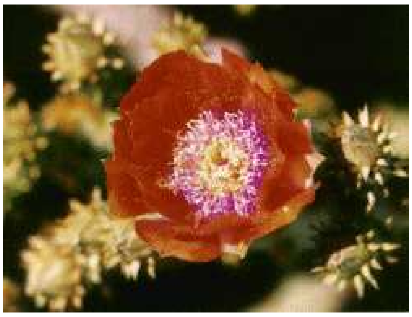


Figure 3.2: Shallow depth of field obtained from a smaller shutter aperture. In this picture, the shutter aperture is F5.6, and shutter speed is $\frac{1}{1000}$. The picture was obtained from the World Wide Web.

Given where the user is pointing the camera, the auto-focus filter (see Appendix A) puts the main subject in focus [61, 62]. Next, the shutter aperture is widened to blur the background. The blurring occurs because the light from out-of-focus objects does not converge as sharply as from objects in focus. By utilizing the significant difference in frequency content of the in-focus and background regions, the proposed algorithm detects the main subject by using filtering, edge detection, and contour smoothing.

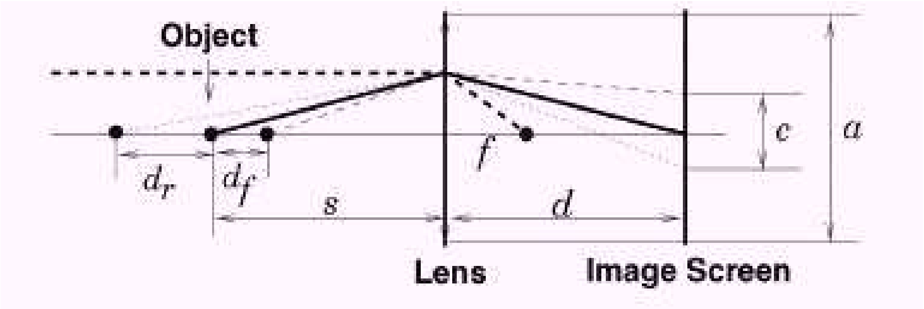


Figure 3.3: Optical model for a thin lens camera.

3.3 Optical Model for Cameras

Assuming a thin lens, the optical model for a typical camera is illustrated in Fig. 3.3 [63, 64]. Let the focal length of the lens be f, its diameter a, and the aperture f-stop number be p, so that $f = ap$. From the Gaussian thin lens formula, we have

$$\frac{1}{s} + \frac{1}{d} = \frac{1}{f} \tag{3.1}$$

where s is the distance of the object from the screen and d is the distance of the image screen from the lens. A point at a distance s will be in focus on the image screen, and will appear as a point. However, depending on the optics, a point closer or further away from the distance s will appear to be as a circle on the image screen. The largest circle that a human being would still tolerate to be a point is called the circle of confusion. Say the diameter of this circle of confusion be c. So, images of points that have a diameter greater than c on the image screen will be perceived to be blurred.

Let d_f and d_r be the limits on front and rear distances for which a point will be perceived to be sharp. Then by using geometry it can be shown that

$$d_f = \frac{scp(s-f)}{f^2 + cp(s-f)} \tag{3.2}$$

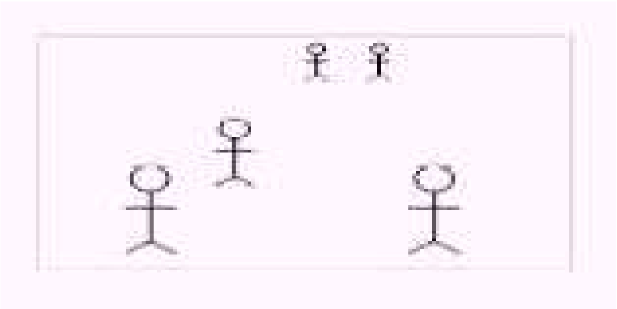


Figure 3.4: Depth perception from an image with high depth of field requires additional human knowledge.

and

$$d_r = \frac{scp(s-f)}{f^2 - cp(s-f)} \tag{3.3}$$

Figs. 3.4 and 3.5 illustrates the depth perception in an image with high and low depth of field, respectively. Based on the above discussion, it can be seen that depth of field variations could be obtained by tilting and shifting the camera, changing the camera aperture, moving the camera closer to the object or by a larger aperture. In this dissertation, I choose to take a supplementary low depth of field picture by changing the shutter aperture. The shutter speed could be automatically changed by the camera, so that the too much light from the larger shutter aperture does not wash out the supplementary picture.

Also, given a particular camera, so that the lens optics are known, the amount of blur at the defocused regions could be estimated by using deconvolution algorithms. However, this dissertation presents a more general framework where image processing would be used to identify the blurred regions from a supplementary shallow depth of field picture. Section 3.4 discusses the formulation for the proposed general framework. However, provided the lens optics are known precisely, the main subject detection stage in the proposed

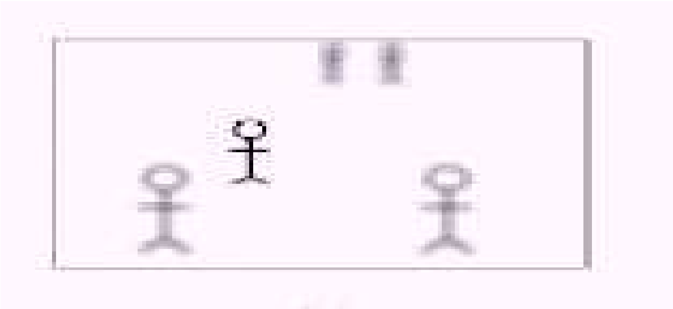


Figure 3.5: Depth perception is easier in an image with low depth of field.

framework could be modified to use deconvolution based algorithms.

3.4 Formulation of Main Subject Detection

As previously mentioned, a supplementary image would be taken that has the main subject in focus and objects not in the plane of focus blurred out. Thus, it can be assumed that the salient feature of the in–focus main subject is the presence of crisp boundaries and edges in it. On the other hand, the background has blurred edges. Now, the segmentation of the object from the background is accomplished based on gradient features. In this algorithm, I propose that *detecting regions with higher gradient values in contrast to the regions having more low frequency components provides a salient classification of pixels in either the main object class or the background class.* Thus, the aim of the proposed algorithm is to detect regions with high gradient in contrast to the blurred background, as the main subject.

Let a scene $\mathcal{C} = (C, F)$ be defined in an n-dimensional Euclidean space, $\Re^n$, where C is the scene domain and F gives the intensity function. Let F_o and F_b be the intensity functions representing the main subject (object) and background classes, respectively, in the scene $\mathcal{C}$ and $F_o \subset F$ and $F_b \subset F$. The objective is to define a spatially connected scene domain C_O, such that for any location $c \in C_O$, the intensity function $F(c) \in F_o$. I define a transformed scene $\mathcal{G} = (C, G)$, obtained by using a transform over the intensity function on the scene domain. In the present context the transform is a gradient operator ∇. The intensity function F undergoes a transformation onto the transform scene, $\mathcal{G} = (C, G)$, where $G = \nabla F$. Thus, for any location $c \in C$, the mapping of $\{F(c) \in F | c \in C\}$ to $\{G(c) \in G | c \in C\}$ can be expressed as $G(c) = \nabla F(c)$. In still image acquisition, I am only working with 2-dimensional photographic images, so I only consider an $\Re^2$ Euclidean space. Henceforth, the *scene domain* will in general be referred to as the *image domain*, and the transform function G as the gradient function.

The gradient function, G can be further partitioned as G_H and G_L, such that

$$G(c) = G_H(c) \text{ if } G(c) \geq \delta \tag{3.4}$$

and

$$G(c) = G_L(c) \text{ if } 0 \leq G(c) < \delta \ , \tag{3.5}$$

where δ is a predefined threshold. $G_H(c)$ and $G_L(c)$ represent the high and low frequency functions, respectively.

Let the inverse transform over G_H and G_L functions map them onto the intensity functions F_H and F_L, respectively. Next, I evaluate the desired relation of F_H and F_L to F. F_H is the intensity function contained in the image domain, which gives rise to the high frequency function $G_H \in G$ under the gradient operation. Similarly, F_L is the intensity function that generates the low frequency function $G_L \in G$. Since G is derived from the intensity function

F and also $G_H \in G$ and $G_L \in G$, the intensity functions F_H and F_L should be contained in F. By the choice of appropriate lowpass and highpass filters to generate F_L and F_H, respectively, F_L and F_H can be recombined to generate the original image F. Depending on the chosen filters, I define a constant θ so that the intensity function $F(c)$ at any point $c \in C$ can be expressed as

$$F(c) = \theta F_H(c) + (1 - \theta) F_L(c). \tag{3.6}$$

To define the contribution of each of the lowpass and highpass components, I study the effect of the constant θ. If the intensity function, $F(c)$, has a greater contribution from high frequency components, then the image essentially will be sharp with well defined edges. On the other hand, if there is more contribution from the low frequency components, the resultant intensity function will have a blurring effect. If I set $\theta > 0.5$, then the contribution of F_H into F becomes higher. In general, to avoid any bias towards the high or low frequency components, the probability can be equally shared; i.e., $\theta = 0.5$ can be taken. For a generic situation, I intend to allow choice of θ depending on the requirements of the image. So, let θ be defined as $\frac{1}{k+1}$, where the real non-negative parameter k can be adjusted to change θ. The unbiased situation can be obtained when $k = 1$.

The goal is to obtain the intensity distribution pertaining to the high frequency component in image in contrast to the blur background component, i.e., extract the sharp features of the image. So, if at any point $c \in C$ the intensity function $F(c)$ is subtracted from the intensity function pertaining to the high frequency component, $F_H(c)$, the generated image has sharper edges around the main subject. Thus from (3.6), and $\theta = \frac{1}{k+1}$, this difference can be expressed as

$$F_H(c) - F(c) = \frac{k}{k+1}\ (F_H(c) - F_L(c))\,. \tag{3.7}$$

In the proposed image acquisition framework, the main subject class, F_o, is in focus, and the background class, F_b, is blurred by widening the shutter aperture. Based on the first assumption, that the main subject in focus will have prominent gradient features and the background that is out of focus will have blurred features, $F_H(c) - F(c)$ will have sharper gradients around F_o and smoother gradients around F_b. Thus, the segmentation of F_o is induced by this difference of gradient information as postulated.

To generate $F_H(c)$ and $F_L(c)$ in the $\Re^2$ domain, highpass and lowpass filters can be designed, respectively. For the highpass filter, the criterion will be to select the frequencies so that $G(c) > \delta$. Similarly, the lowpass filter will have frequencies so that $G(c) < \delta$. The choice of filter coefficients will determine its characteristics, and the filter can be designed adaptively.

3.5 Algorithm for Main Subject Segmentation

Based on the basic assumption that the object in focus has higher gradient components compared to those not in the plane of focus, the proposed algorithm attempts to sharpen these gradients more in contrast to the blurred regions. Edges are detected over this sharpened image and subsequently a continuous smooth contour is defined by using deformable active contour model. The algorithm development, the design of the filters to accomplish this task and their desired properties are discussed in the subsequent parts. Details of the edge detection and active contour model are also presented.

Subsection 3.5.1 discusses detecting the sharper regions in the image.

Subsection 3.5.2 detects the strong edges from the processed image. Subsections 3.5.3 and 3.5.4 describe algorithms for closing the contour and generating the main subject mask. Subsection 3.5.5 finally modifies the generated mask based on the difference between the original picture and the supplementary picture.

3.5.1 Sharp Region Identification

For the 2-dimensional case, the conditions of (3.7) are satisfied with an image sharpening filter as modeled in Fig. 3.6. Let $I_{smooth}(x, y)$ define the image comprising of the intensity function derived from low frequency components, i.e., the smoothed image derived from the original image $I(x, y)$. To reduce the effect of the blurred components in the image $I(x, y)$, $I_{smooth}(x, y)$ is subtracted from $I(x, y)$. Let the resultant image be denoted as

$$g(x, y) = I(x, y) - I_{smooth}(x, y) \tag{3.8}$$

A sharpened image can be generated by adding $g(x, y)$ with the original image $I(x, y)$, as follows:

$$I_{sharp}(x, y) = I(x, y) + k\, g(x, y) \tag{3.9}$$

Here the factor k, as described in Section 3.4, gives the proportion of high gradient image into the resultant $I_{sharp}(x, y)$. From the relation in (3.6), $I(x, y)$ is composed of a weighted combination of $I_{sharp}(x, y)$ and $I_{smooth}(x, y)$:

$$I(x, y) = \frac{1}{k+1}\, I_{sharp}(x, y) + \frac{k}{k+1}\, I_{smooth}(x, y) \tag{3.10}$$

As the value of the non-negative real parameter k increases, the contribution of the smoothed image into the original image increases, thus blurring the image. The effect of this on the final output will be evident from the following step. To realize (3.7), the original image $I(x, y)$ needs to be subtracted from $I_{sharp}(x, y)$. Thus,

$$I_{sharp}(x, y) - I(x, y) = \frac{k}{k+1}\left(I_{sharp}(x, y) - I_{smooth}(x, y)\right) \tag{3.11}$$

The intention is to remove as much low frequency induced intensity as possible from the original image. Now, if $I(x, y)$ were mostly composed of $I_{smooth}(x, y)$, then $I_{sharp}(x, y) - I(x, y)$ would be mostly composed of high frequency induced intensity. Instead of computing $I_{sharp}(x, y) - I(x, y)$, the right hand side of (3.11) has been computed. Subtracting a smoothed version of the user-intended image from the sharpened image generates an edge map in which the edges around the main subject are sharper than the background edges. Hence, the problem of segmenting the main subject reduces to separating the regions with the sharper edges from the regions with smeared edges.

For the above tasks, I design linear time invariant filters both for lowpass and highpass filtering. The 3×3 sharpening filter used in this work is designed as follows:

$$\frac{1}{1+\alpha}\begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + \beta & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix} \tag{3.12}$$

Parameters α and β define the shape of the frequency response. I chose $\alpha = 0.2$ and $\beta = 5$. An integer implementation could choose $\alpha = 0.2$ and $\beta = 5$, remove the $\frac{1}{1+\alpha}$ factor, and scale the coefficients by 5. The frequency response of the above highpass filter is shown in Fig. 3.7. For the lowpass filter, a 3×3 Gaussian blur filter, that has frequency response shown in Fig. 3.8, is used. However, the filter characteristics could be adapted according to the strength

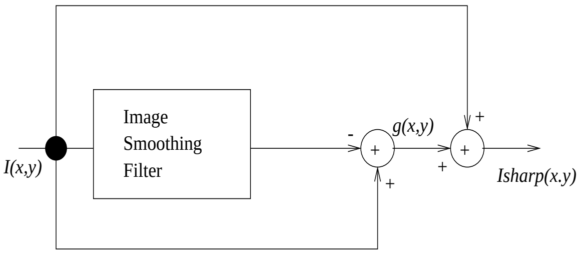


Figure 3.6: Model for an image sharpening filter.

of the image features. For example, an image having relatively weak edge features could be processed by a filter having a lower cut-off and greater span in the spatial domain, e.g. a 7 × 7 filter.

For a system in which the user is allowed to change parameters, this stage could be avoided. Edge detection could be directly performed on the supplementary image to identify strong edges. However, this requires user intervention in selecting the proper threshold for the edge detector for different images. So, this region identification stage removes such biases for a fully automated system. A particular threshold at the next edge detection stage has worked well for around 30 test images, for which the system was tested.

3.5.2 Edge Detection

The resulting image obtained by using (3.11) is passed through an edge detector. The Canny edge detector [65] first smoothes the difference image $g(x, y)$ in Fig. 3.6, then computes the gradient, and finally thresholds the gradient to preserve the strong edges and suppress the weak edges. The Canny edge detector preserves the directions of the edges, which is vital information for closing the boundary of the main subject by using gradient vector flow (Section 3.5.4).

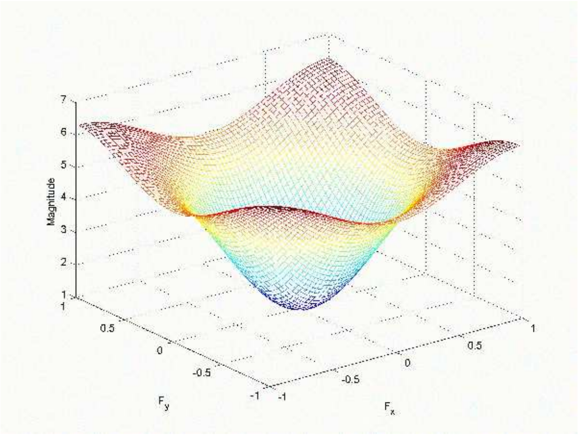


Figure 3.7: Frequency response of the 3×3 highpass filter in (3.12) for $\alpha = 0.2$ and $\beta = 5$ used for image sharpening.

Another popular edge detector, the Laplacian of Gaussian edge detector [66] be tuned to preserve strong edges and suppress weak edges, but it did not perform as well as the Canny edge detector. The non-directional derivatives used in the Laplacian of Gaussian edge detector produces responses both parallel and perpendicular directions to a given edge. The drawback could have been improved by using directional first and second derivatives. Nonetheless, the Laplacian of Gaussian edge detector would still not preserve the edge direction. The Canny edge detector also performs better than Roberts, Sobel, and Prewitt edge detectors [67].

To separate the strong edges in the focused parts from the weak edges in the out-of-focus parts, the hysteresis threshold of 0.3 for the Canny edge detector worked well for the test images shown in this dissertation. This selected hysteresis threshold depends on the value of k in (3.10). The value of k depends on the amount of blurring in the acquired image, i.e. the amount of background blur obtained from the lens in the camera. So, for each camera,

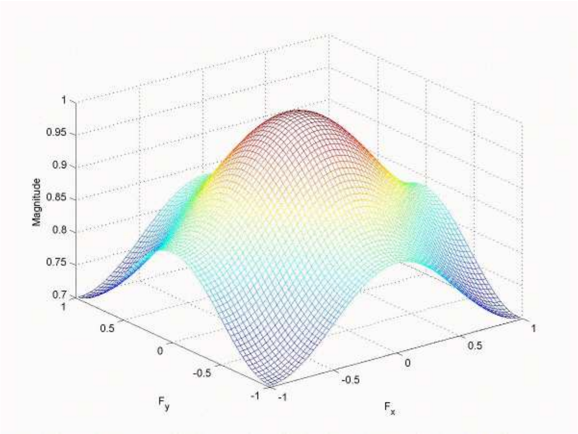


Figure 3.8: Frequency response of the 3 × 3 Gaussian lowpass filter used for image blurring.

the hysteresis threshold could be set to work for a range of natural images. However, with any preselected threshold, the strong edge detection step would still pick background edges for some acquired images where there is not enough background blur or strong edges in the main subject. For example, in Fig. 3.9, the main subject is the white flower, which does not have enough strong edges due to its monochromatic nature. Also, the background is more textured and not blurred enough. In this case, the proposed algorithm picks background edges with a hysteresis threshold set to 0.3 for the Canny edge detection step.

3.5.3 Contour Detection

To close the boundary of the detected strong edges and to generate the main subject mask, I choose to feed the edge detection output into a contour detection framework. I prefer to use this approach over by using morphological operators or snake algorithms to close the boundary of the detected strong edges.

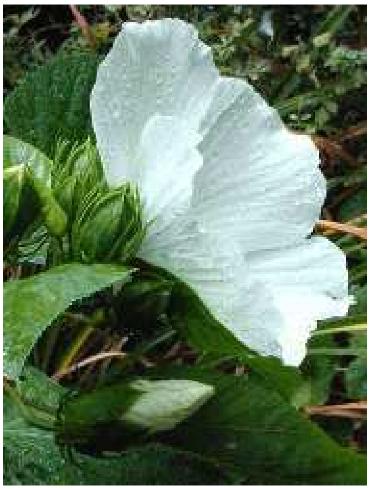


Figure 3.9: An example of an image where the proposed main subject detection algorithm picks both the main subject and background edges with a preselected hysteresis threshold of 0.3 for the Canny edge detector. Due to the monochromatic nature of the main subject, the white flower, and more textured background, there is not enough strong edges on the main subject or enough background blur.

Tsai and Wang [68] experimented with using morphological operators for the edge linking procedure. Their proposed approach consisted of dilation, thinning, and line linking. Simple dilation [67] was carried out to close the gaps in the initial edge map. Then thinning [67] was performed to ensure that the edges are 1-pixel wide. Finally the edges were linked based on finding the nearest neighbor and the direction of a particular edge. As searching for the nearest neighbor is time consuming, I did not choose this approach to connect the edges. Also, this approach is not suitable for finding edges in objects with blunt boundaries [38].

The snake [69] is described as an energy minimizing spline guided by internal and external forces towards the desired image features. Internal forces are determined by the curve characteristics and are generally defined in terms of elasticity and rigidity of the curve. The user constraints and image features, e.g., image intensity and edge functional, define the external forces, which guide the simulated elastic material to conform to the local image features. Although this approach is very effective for blob-like structures, some of the major challenges of the deformable spline is making it able to conform to the image concavities and to segment objects having sharp corners or elongated structures. Hence, the snake [69] algorithm and its direct descendants fail to track the concavities in the contour or require the initial control points to be placed near the actual contour. This limits its automated application for natural images.

Many improvements on the basic snake algorithm exists. Cohen and Cohen [70] proposed a balloon force to make the curve move towards the desired features, which reduces its sensitive to initial conditions. Berger [71] introduced the snake-growing algorithm, which allows snake to grow along features and also break by using local features. Neuenschwander, Fua, Iverson, Szekely, and Kubler [72] later utilized a similar idea in the Ziplock Snakes. Gunn and Nixon [73] developed a dual active contour model, where two curves approach true boundaries from both inside the object and from the background. Yuen [74] designed an enhanced snake algorithm, which uses a split and merge technique to make the snake track the concave boundaries. Research also has been motivated to the introduction of more image and task specific information along with filter orientations into the snake framework [48, 75, 76, 77, 78, 79]. One of the major challenges in the above approaches is the requirement of *à priori* knowledge about the image space and the intensity distribution within the object. This may not be possible in many situations. Moreover the task specific approaches cannot be generalized in many situations.

3.5.4 Active Contours and the Gradient Vector Flow Algorithm

In this research, main subject detection relies only on gradient information of the image. So, the appropriate active contour model to fit into this framework would primarily depend on the gradient information rather than region-based or *à priori* information. Thus, the gradient vector flow [80, 81] (GVF) algorithm, which is guided by the diffusion of the gradient vectors from the edge map of the image, is a good choice because it requires no initialization in terms of control points and has a higher capture range in its ability to track image contour concavities. This subsection describes the active contour principle and presents the GVF theory.

The theory of snakes has been motivated by the idea of deforming a spline based on various levels of information incorporated in a meaningful unified representation for segmentation of the object class from the image. With this motivation Kass, Witkin and Terzopoulos [69] defined an energy minimizing parametric curve $v(s) = (x(s), y(s))$, having the walk along the arc, s ($s \in [0, 1]$), as parameter. The energy function of this deformable curve is defined as

$$E_{snake}(v(s)) = E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)) \tag{3.13}$$

where $E_{int}(v(s))$ represents the internal energy of the contour due to bending or discontinuities, $E_{image}(v(s))$ is the image energy and $E_{con}(v(s))$ represents the external constraints. Internal energy of the spline is represented as

$$E_{int}(v(s)) = \alpha(s) \left| \frac{\partial v}{\partial s} \right|^2 + \beta(s) \left| \frac{\partial^2 v}{\partial s^2} \right|^2 \tag{3.14}$$

The first order term tends to shrink the contour, while the second order term restricts bending of the spline. The values of $\alpha(s)$ and $\beta(s)$ control the shrinking and rigidity strength at the concerned point of the spline. The purpose of image-based energy $E_{image}(v(s))$ is to incorporate image features in guiding the contour deformation.

GVF [80, 81] is essentially a force balance equation defined by the internal force attributed by the geometric properties of the spline and the external force derived from image features. The static external force field defined in GVF aims at having non-rotational (curl-free) and solenoidal (divergence free) components. In the Euler formulation of force field from the energy minimization definition of (3.13), the derivative of E_{image} has been replaced by a vector field $\mathbf{f}(x, y) = [a(x, y), b(x, y)]$ that minimizes the energy functional

$$\mathcal{E} = \int \int \mu(a_x^2 + a_y^2 + b_x^2 + b_y^2) + |\nabla I|^2 \, |\mathbf{f} - \nabla I|^2 \, dxdy. \tag{3.15}$$

The first term $(a_x^2 + a_y^2 + b_x^2 + b_y^2)$ is basically the optical flow vector. From (3.15) the first term dominates the energy equation when the gradient, ∇I, is small, i.e., in the homogeneous regions in the image. The effect of the gradient, which is high in inhomogeneous regions, is minimized by setting the factor $|\mathbf{f} - \nabla I|$ to zero. This is the motivating spirit in using the GVF into the proposed framework. Since the contour detection is performed on the edge map of the image, the situation here is more tricky in the sense that it has perfectly homogeneous regions and high gradients are given as the Dirac delta function. The GVF, however, takes care of this situation by the strength of the aforementioned property of the potential field. Also, GVF is a better choice since it has a higher capture range with the ability to track image contour concavities.

Since one of the main aims of the proposed work is reduction of complexity, initialization of the contour plays an important role. The initialization is generated automatically from the edge map. The outer most edge is considered, and a contour is inflated by using a balloon force along the direction normal to the edge at each point for initialization. Then this initialized contour undergoes deformation under the action of the force field defined by GVF.

3.5.5 Modification of Mask Based on Difference Between Original and Supplementary Picture

One of the drawbacks of taking a supplementary picture with a shallow depth of field is that the subject or the camera could have moved while the supplementary picture is taken. This drawback could be reduced by mounting the camera on a tripod. However, for small ranges of motion, I propose to modify the generated picture based on the difference between the original and the supplementary picture. I employ a simple image registration method [82, 83] between the original and the supplementary image to reduce implementation complexity.

Once the original mask has been generated, a difference is computed between the original and the supplementary image. Now, the difference image will contain pixels where the main subject has moved and pixels of the background that are in different focus compared to the supplementary picture. However, any change in main subject motion shows up more significantly compared to the change in focus. So, by using a threshold on the difference image, one can identify if the main subject has moved. A second mask is thus generated that identifies the pixels in the difference image lying above this threshold. For my application, this threshold is chosen to be 70 for an 8-bit image. This second mask is added to the generated main subject mask to create the mask that will be used on the original image.

For example, Fig. 3.10(a) shows the supplementary image and Fig. 3.10(b) shows the generated main subject mask. Now let Fig. 3.11(a) be the original image. In this case, the main subject has not moved significantly. Fig. 3.11(b) shows the difference between Fig. 3.10(a) and Fig. 3.11(a). Fig. 3.11(c) is the mask generated from the difference image in Fig. 3.11(b). Fig. 3.11(d) is the modified main subject mask, depending on the original main subject mask in Fig. 3.10(b) and the difference image in Fig. 3.11(c). Similarly, there can be another case where the main subject has moved significantly as in Fig. 3.12(a). In this case, the difference image is shown in Fig. 3.12(b). The thresholded difference image is shown in Fig. 3.12(c). Depending on the thresholded difference image, the main subject mask is modified as in Fig. 3.12(d).

3.6 Implementation Complexity

The basic flow for the proposed algorithm is shown in Fig. 3.13. Initially the auto-focus filter has the main subject in focus and shutter aperture is widened to blur the background. This image then undergoes the subsequent stages of image sharpening, edge detection and contour tracking as has been detailed in the previous section.

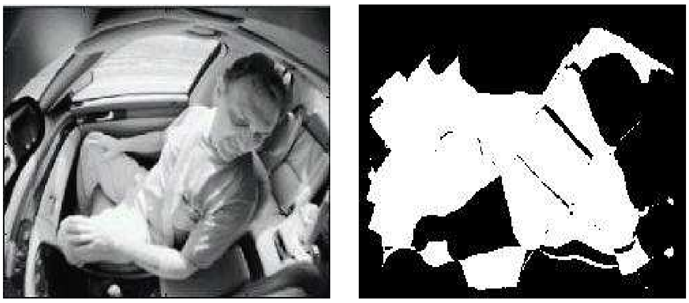

(a) (b)

Figure 3.10: (a) Supplementary image acquired by the digital still camera to detect the main subject. (b) The detected main subject using image (a).

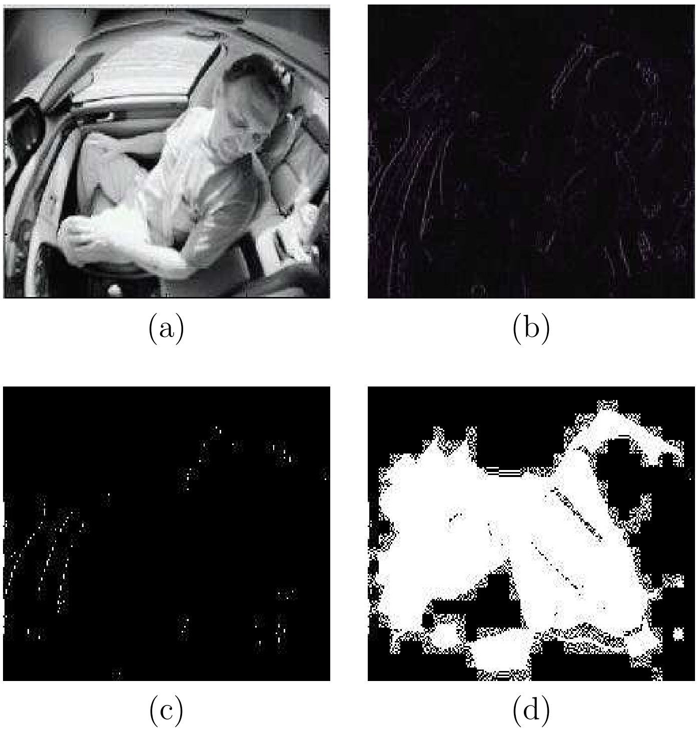
The RGB color image is converted to intensity by either

$$I = (R + G + B)/3 \text{ or } I = (R + 2G + B)/4 \tag{3.16}$$

The former step requires 2 multiply-accumulates, which matches a programmable digital signal processor well. The later, which requires 2 adds, a left shift by one bit (multiplication by 2) and a right shift by two bits (division by 4), reduces the digital hardware overhead. Shifts can be used here to implement division be a power-of-two positive integer because RGB values are non-negative.

The sharpening and smoothing operations convolve the image with a 3× 3 filters, respectively. However, the sharpening, smoothing and the difference calculation can be combined so that 9 multiply-accumulates are required per pixel. Moreover, the sharpening filter coefficients in (3.12) can be scaled to be integer values as described in Section 3.5.1.

Canny edge detection first smoothes the image in order to lower the noise sensitivity, then computes a gradient, and finally suppresses the non-maximum pixels by using two thresholds. The smoothing and the gradient



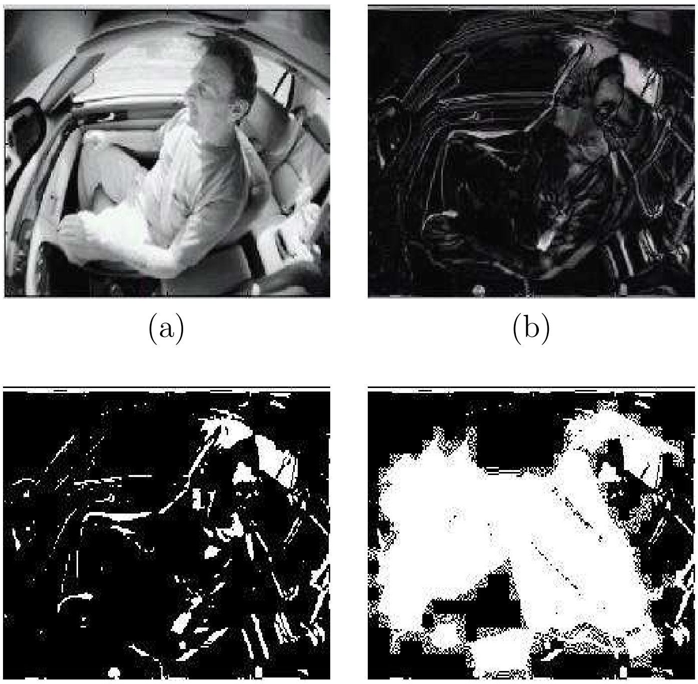
(a) (b)

(c) (d)

Figure 3.11: (a) Possible original image taken by the user. (b) Difference between the original image (a) and the supplementary image. (c) Mask generated by thresholding the difference image in (b). (d) Modified main subject mask depending on the mask in (c).

computation takes 9 multiply-accumulates, assuming a 3×3 pre-computed filter kernel that is the derivative of a Gaussian mask. The nonmaximum suppression step requires 2 comparisons per pixel. The two 3×3 filters can be cascaded to a 5×5 filter to reduce the number of memory accesses per pixel. This requires 5 memory reads per pixel.

As the exact implementation of the gradient vector flow algorithm to close the contour is computationally intensive, I propose to use an approxima-



(a) (b)

(c) (d)

Figure 3.12: (a) Possible original image taken by the user. (b) Difference between the original image (a) and the supplementary image. (c) Mask generated by thresholding the difference image in (b). (d) Modified main subject mask depending on the mask in (c).

tion. From the map of the detected sharper edges, the pixel position of the first “ON” pixel from the left and the right boundaries of the image is calculated. Every pixel between these two pixels is turned “ON”. This approximation detects the convex parts correctly, but fails at the concavities in the shape of the main subject. The approximate procedure requires 2 comparisons per pixel. The generated mask is written back with 1 memory access operation per pixel.

Depending on the difference between the original and supplementary

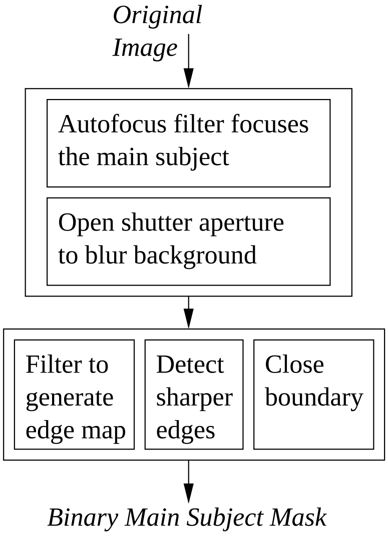

Original Image
Autofocus filter focuses the main subject
Open shutter aperture to blur background
Filter to generate edge map
Detect sharper edges
Close boundary
Binary Main Subject Mask

Figure 3.13: Proposed automated main subject detection algorithm for digital still cameras.

image, the generated main subject mask is modified. To compute the difference 1 subtraction per pixel and to modify the mask 1 addition per pixel is required. 2 additional memory accesses are required to accesses the image and write back the mask

The main subject mask can be generated with 20 multiply-accumulates, 4 comparisons and 8 memory accesses per pixel. As digital still cameras use approximately 160 digital signal processor instruction cycles per pixel, the main subject can be detected with relatively low implementation complexity.

3.7 Main Subject Segmentation Results

The proposed algorithm has been tested on several natural images. A quantitative analysis has been performed to evaluate the segmentation performance of this algorithm. Extensive experiments have also been conducted to make a comparative study of the proposed algorithm with the prevalent techniques for main subject detection.

Fig. 3.21 extensively illustrates the performance of the proposed algorithm for main subject detection at various levels. Fig. 3.21(a) shows the supplementary image obtained with the main subject in focus and the background blurred. Fig. 3.21(b) is the difference image obtained from a sharpened and smoothed versions of Fig. 3.21(a). The strong edge detection results from Fig. 3.21(b) is shown in Fig. 3.21(c). The gradient of the edge map is illustrated in Fig. 3.21(d). The gradient vector flow field is shown in Fig. 3.21(e). Fig. 3.21(f) shows the initial contour. Figs. 3.21(g) and 3.21(h) show the contours at iterations 5 and 10, respectively, that are generated by using the GVF field shown in Fig. 3.21(e). This iterative step is not mandatory, and depending on the computational resources available and the allowable implementation complexity, the iteration can be terminated at any point. Fig. 3.21(i) shows the binary mask generated from the detected contour.

Similar studies were conducted for around 30 images. I either downloaded these images from the World Wide Web in the year 2001 or acquired them with a Cannon Powershot G3 camera. The shutter aperture was varied from F2 through F2.8 to make sure that the acquired images are low depth-of-field photographs. The test set consisted of variety of pictures having human or inanimate main subjects are were taken under different light conditions and scene settings (indoor/outdoor). The original pictures are available at

http://www.ece.utexas.edu/˜bevans/projects/dsc/

Figs. 3.22(a) through 3.31(a) show background blur achieved by a wider shutter aperture, while the main subjects are in focus. The results of locating the main subjects before contour closing are shown in Figs. 3.22(b) through 3.31(b). Figs. 3.22(c) through 3.31(c) show the detected main subject mask. Qualitative visual inspection shows that the generated mask closely represents the main subject in focus. As can be seen, the developed algorithms are independent of scene settings or content.

Figs. 3.22 and 3.21 have human main subjects in outdoor settings. In Figs. 3.23, 3.26, 3.27 and 3.28 the main subjects are inanimate objects in indoor settings. Figs. 3.24, 3.25 and 3.30 are close up shots of house plants in indoor settings. For Fig. 3.29 a cognitive model would recognize the stuffed bear to be the main subject. However, due to depth of focus, the beaded curtains are sharper and hence proposed algorithm chooses that to be the main subject. In Fig. 3.31 the bush is the main subject outdoor settings.

For better evaluation a quantitative study has also been performed over the first three images. Three measures – sensitivity, specificity, and the error rate as suggested by [47] – have been used for evaluating the performance of the proposed segmentation algorithm. The sensitivity is defined as the ratio of the area of the detected main subject to the total area of the main subject in the image. The specificity is the ratio of the area of the detected background to the total area of the background in the image. Here the total area of the main subject or the background are the number of pixels that actually represent the main subject or the background, respectively, as would have been observed by a human. The error rate is the ratio of the number of pixels that are misclassified to the total area of the image.

Let the scene domain C be classified into object class, C_o, and background class, C_b, by the proposed algorithm. Let C'_o and C'_b represent the

Image	Resolution	Sensitivity	Specificity	Error rate
Man & child	280 × 350	88.0%	97.2%	4.1%
Man	246 × 276	77.8%	90.3%	8.0%
Stuffed animal	316 × 422	82.2%	94.6%	6.3%

Table 3.1: Segmentation accuracy measures for the proposed main subject detection algorithm for images in Figs. 3.21, 3.22, and 3.23.

actual object and background classes, respectively. Then the measures are defined as

$$\text{Sensitivity} = \frac{C_o}{C_o'}, \tag{3.17}$$

$$\text{Specificity} = \frac{C_b}{C_b'}, \text{ and} \tag{3.18}$$

$$\text{Error rate} = \frac{(C_o \cup C_o') - (C_o \cap C_o')}{C}, \tag{3.19}$$

where $\cup$ and $\cap$ represent the union and intersection operations, respectively.

For the segmented images given in Figs. 3.21(c), 3.22(c), and 3.23(c) the sensitivity, specificity, and the error rate are given in Table 3.1. The accuracy in segmentation as seen in Table 3.1 is within the tolerable limit as a trade off for low-complexity in detecting the main subject for subsequent automation of the photographic composition rules. The results also are comparable with Wang's *et al.* [37, 38] reported values for the three quantifiable measures for low depth of field images. For their test images, sensitivity, specificity, and error rate varied from 73.7% to 97.5%, 80.1% to 97.5%, and 3.4% to 5.5%, respectively.

3.8 Comparison with Prevalent Segmentation Methods

Figs. 3.32 through 3.36 compare the proposed main subject segmentation algorithm with Wang's *et al.* wavelet based method [37, 38], and Won's *et al.* iterative method [40]. As described in Section 3.6, the proposed method takes 20 multiply-accumulates, 4 comparisons, and 8 memory accesses per pixel, and does not require any *à priori* training.

The multiscale wavelet based method [37, 38] generates the wavelet coefficients for each stage and classifies the image based on the variance of the wavelet coefficients by using the k-means clustering algorithm. The process is repeated for multiple wavelet levels. Generating the wavelet coefficients involves filtering the image with lowpass and highpass filters, respectively. Also, the computationally intensive part of the k-means clustering lies in computing the Euclidean distance of each point from the neighboring clusters. Taking into account all these factors, the wavelet based method will at least be $2 \times n \times k$ more complex than the proposed method, where n is the number of wavelet levels computed and k is the number of clusters.

The iterative approach by Won *et al.* [40] starts by dividing the image into non-overlapping blocks. A few probability measurements are computed from the image variance to classify each block as foreground or background. The block classification is further refined into pixel-level classification by using recursion and the watershed algorithm. So, if the original image in divided into $B \times B$ blocks, this method would be at least B times as complex than the proposed method. For the results in Figs. 3.32(d) through 3.36(d), Won *et al.* substitute the grey level values from the original image onto the generated mask for visual inspection.

The proposed method generates a reasonable mask of the main subject with much lower complexity than the aforementioned methods. Also, the proposed algorithm can be implemented in fixed-point arithmetic. As the proposed pixel-based approach to detect the main subject produces comparable results with the more complex wavelet-based method [37, 38], the following subsection compares the two paradigms.

3.8.1 Comparison of Multiresolution-based (Wavelets) and Pixel-based Main Subject Detection

Section 3.8 shows that the proposed pixel-based approach to detect the main subject shows comparable accuracy in detecting the main subject compared to the multiresolution wavelet-based approach at a much lower computational complexity. Comparing the multiresolution wavelet-based [37, 38] and pixel-based [3, 4, 5, 40] approaches to segment the main subject, it can be seen that any wavelet or filter-based multiresolution approach to segment an image would be better at representing regional features of the image. Depending on the filter length and the resolution which is being used for analysis, the regional properties of the image would show up in the frequency transformed domain. So, any analysis based on regional properties will have estimation errors depending on the length of the used filter and the resolution at which it is being viewed at. The pixel-based approaches however analyze the image on a pixel by pixel basis, and the errors will depend on how well each pixel is classified. Thus, in this dissertation, I present a pixel-based approach that is fast and classifies the pixels with tolerable accuracy required for this application.

So, for images with substantially large smooth regions that are separated by well defined edges, both the wavelet-based or pixel-based algorithms would provide similar results. However, in images with many edges, texture and smaller smooth regions, the pixel-based approach would be more accurate. Also, in Wang, Li, Gray and Wiederhold's [37, 38] wavelet-based approach, the segmentation accuracy is further reduced when the authors use a block-based approach, in which in the subsequent iterations, the class of a subblock is switched, depending on the subblock neighborhood.

For simplicity, I choose three model images to compare the proposed pixel-based and a multiresolution approach. For the time being I use a Laplacian pyramid for the multiresolution analysis. The first image is a plain image with no edge and is generated as $f(x, y) = 0$, as shown in Fig. 3.14. The second is an image with a white circle on a black background, and it has a defined strong edge. It is generated as $f(x, y) = 1$ if $x^2 + y^2 \leq r^2$ where r is the radius of the circle, as shown in Fig. 3.15. The third image is a ramp modeling an image with a very blunt edge, as shown in Fig. 3.16. This image is generated as $f(x, y) = \sqrt{(x - x_{mid})^2 + (y - y_{mid})^2} / \sqrt{(x_{max} - x_{mid})^2 + (y_{max} - y_{mid})^2}$, where (x_{mid}, y_{mid}) are the mid points and (x_{max}, y_{max}) are the dimensions of the image.

Both the pixel-based and multiresolution Laplacian pyramid based approaches identify that there is no edge in Fig. 3.14. The results of the proposed pixel-based approach to segment Fig. 3.15 is shown in Fig. 3.17. The 6 levels of the Laplacian pyramid decomposition are shown in Fig. 3.19, where Fig. 3.19(a) represents the highest frequency content and Fig. 3.19(f) represents the lowest frequency content. Now, as this image has a sharp edge the highest frequency octave identifies the circle correctly. However, as more and

Figure 3.14: A plain image without edges for comparing pixel and multiresolution based approaches to detect the main subject.

more lower resolutions will be considered to segment the image, the accuracy of segmentation would reduce. But, the regional properties of the image is present across all the octaves. Similarly, for Fig. 3.16 depending on the chosen thresholds the proposed pixel-based approach either chooses none of the image or almost the whole of the image as shown in Fig. 3.18. The 6 octaves of the Laplacian pyramid decomposition for this image is shown in Figs. 3.20(a) through 3.20(f). Here also the segmentation would depend of which levels are being considered.

Now in a natural image, the strength of the edges cannot be predetermined, and the strength of all the edges would not likely be the same. So, a multiresolution approach would be better at representing the regional properties of the image but the segmentation accuracy would depend on which frequency level is being considered for segmentation. The accuracy of the pixel-based approach on the other hand will depend on how well each pixel is classified.

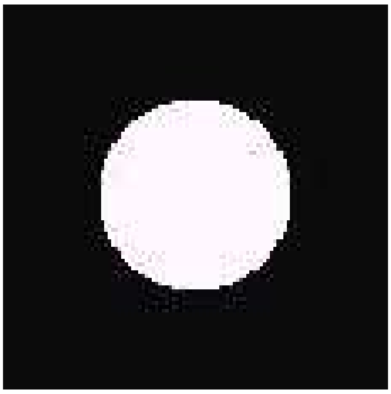


Figure 3.15: An image with a well defined strong edge for comparing pixel and multiresolution based approaches to detect the main subject.

3.9 Conclusion

This chapter proposes an algorithm for using frequency information difference between the main subject and the background, to segment the main subject in a photograph. The camera takes a supplementary picture with the main subject in focus, and background object blurred from diffused light through a larger shutter aperture in the camera. In this supplementary picture, there is an initial region-based segmentation, as the main subject has distinctive gradient-content when compared to the out-of-focus objects.

I segment the main subject from this supplementary picture, by using the difference in local gradient (edge) information between the main subject and the background objects. The algorithm is independent of scene setting or content, as long as there is(are) an identifiable main subject(s) in the picture. The implementation complexity of the proposed algorithm is similar to a 5×5 filter. The developed algorithm has lower implementation complexity compared to the existing methods and produces visually comparable main subject

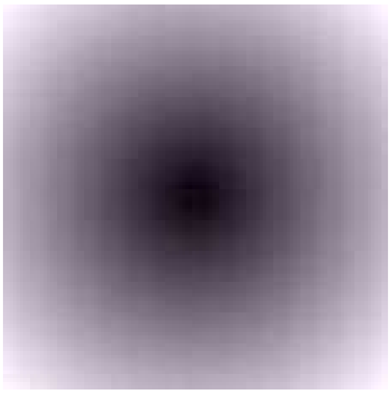


Figure 3.16: An image with a blunt edge for comparing pixel and multiresolution based approaches to detect the main subject.

masks.

After detecting the main subject, the subsequent chapters describe proposed algorithms for automating selected photographic composition rules. Photographic composition rules could be broadly divided into two categories. One category of rules could be applied to the image just knowing the main subject mask itself. The other requires information of the main subject as well as the image background. Chapter 4 automates two rules, namely, rule-of-thirds and background blurring, from the first category. Chapter 5 automates one rule, merger detecting and mitigation, from the second category. Other photographic composition rules could be automated by using the frameworks presented in the next two chapters.

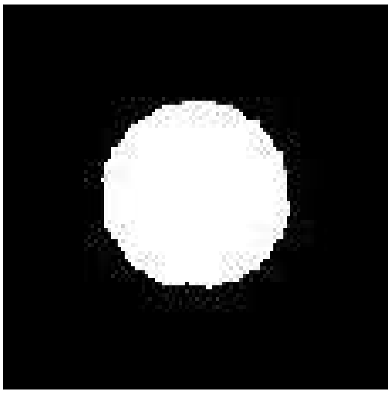


Figure 3.17: Segmented circle with the proposed pixel-based approach.

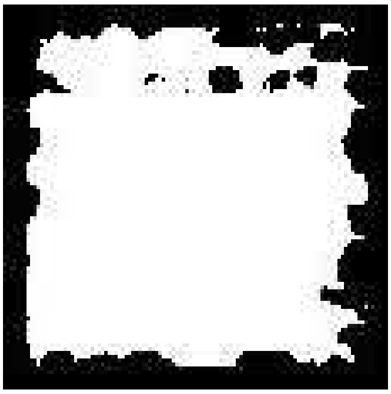


Figure 3.18: Segmented ramp with the proposed pixel-based approach.

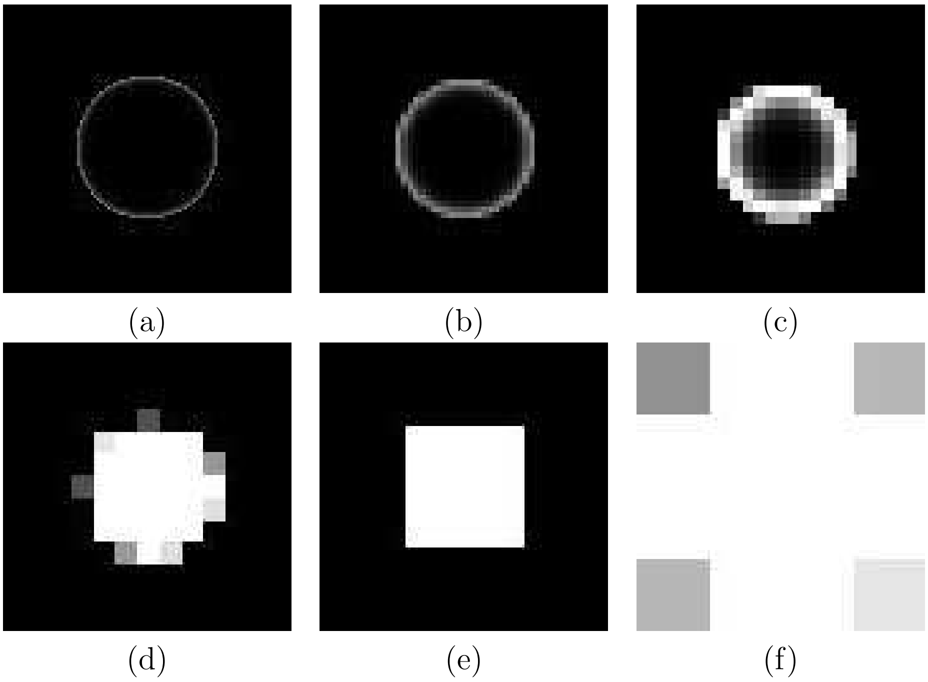
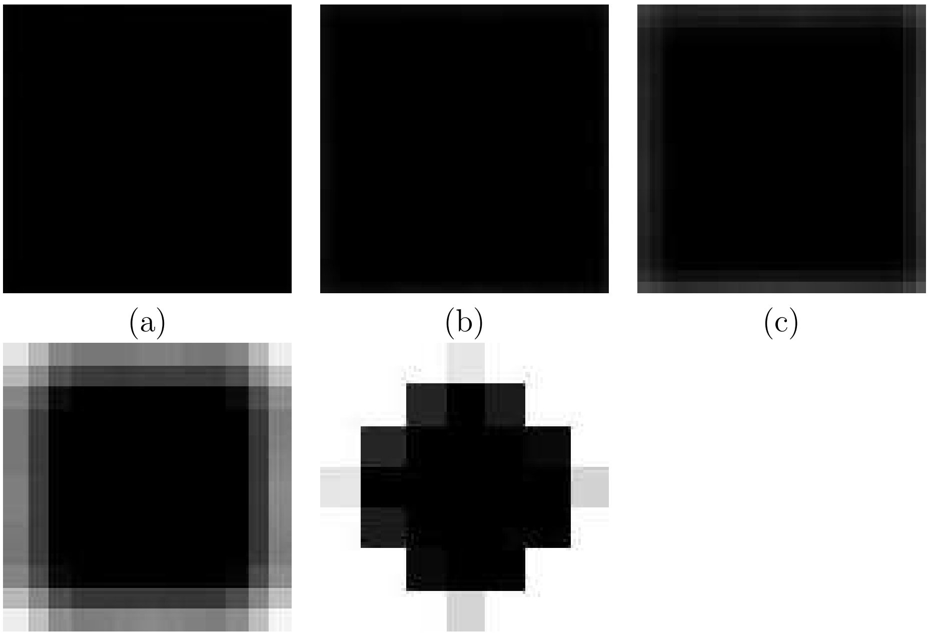

(a) (b) (c) (d) (e) (f)

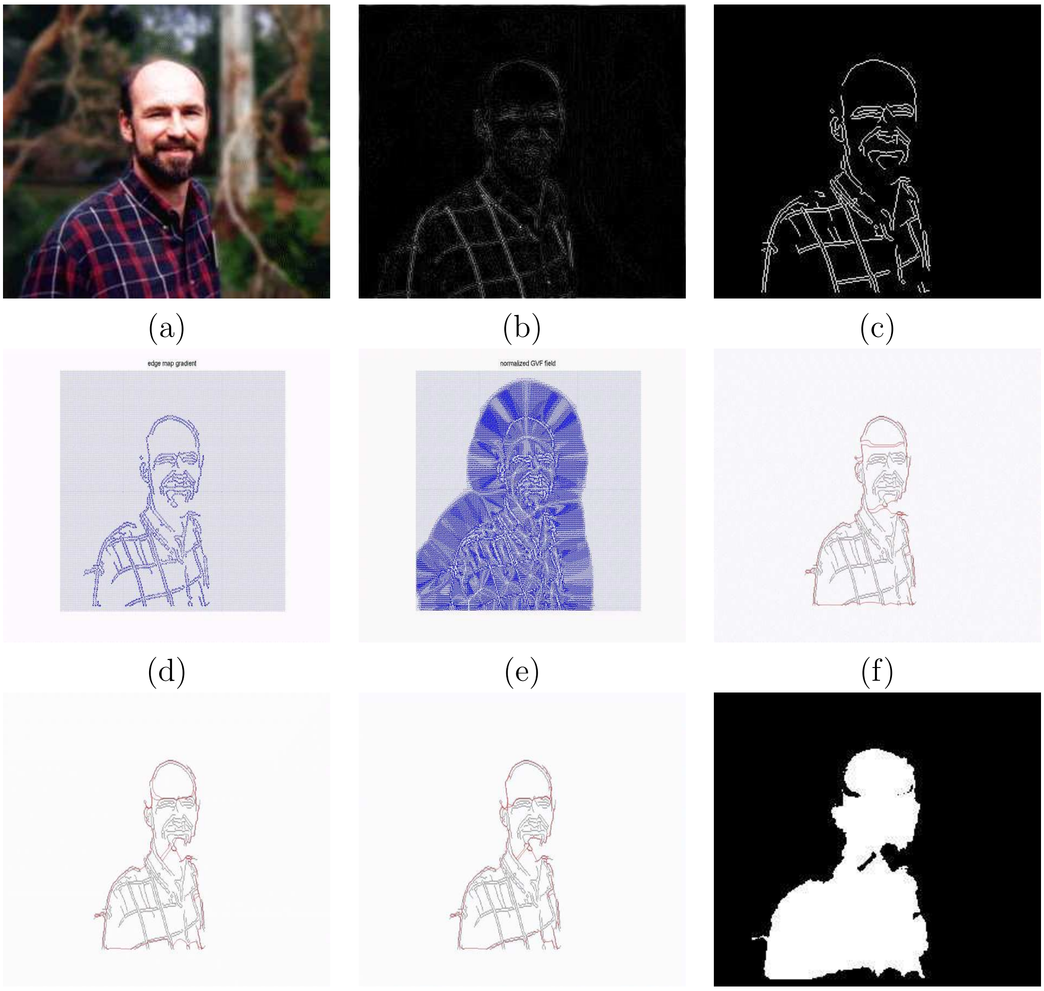
Figure 3.19: The six levels of the Laplacian pyramid for the image with a strong circular edge from highest (a) through the lowest (f) frequency octaves.



(a) (b) (c)

(d) (e) (f)

Figure 3.20: The six levels of the Laplacian pyramid for the ramp image from highest (a) through the lowest (f) frequency octaves.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3.21: Detecting the main subject, the man, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Difference image; (c) Detected strong edges; (d) Gradients of the edge image; (e) Gradient vector flow field; (f) Initial contour of the main subject; (g) Contour at iteration 5 (if required); (h) Contour at iteration 10 (if required); and (i) Detected main subject mask.

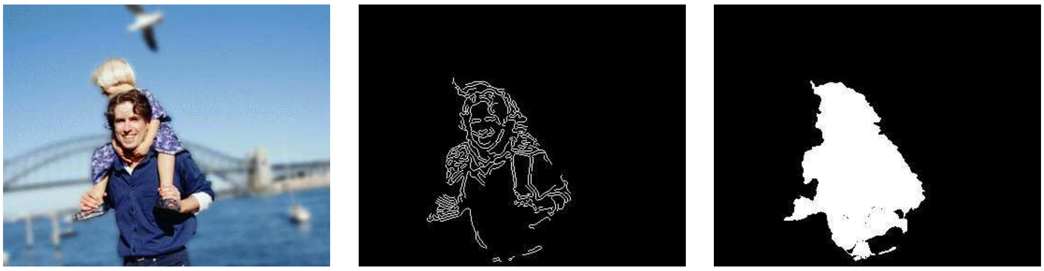
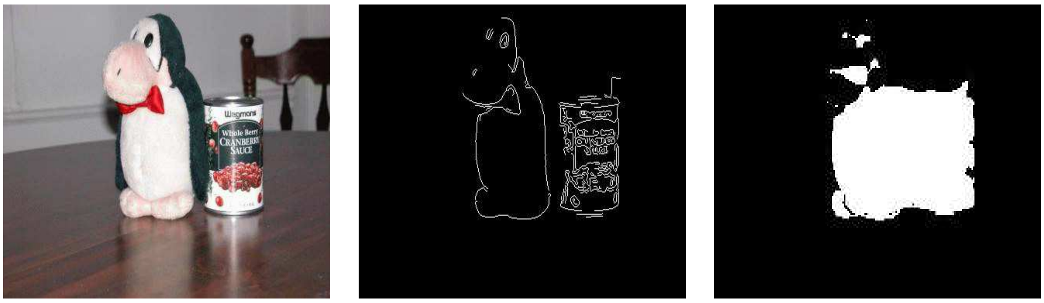
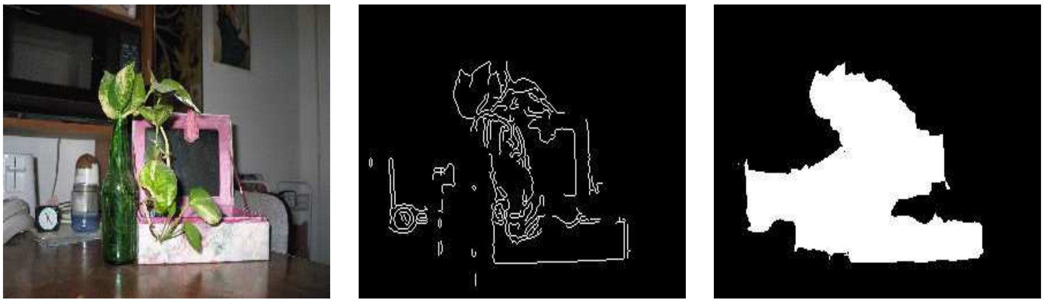

(a) (b) (c)

Figure 3.22: Detecting the main subject, the man and the child, which are in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask.



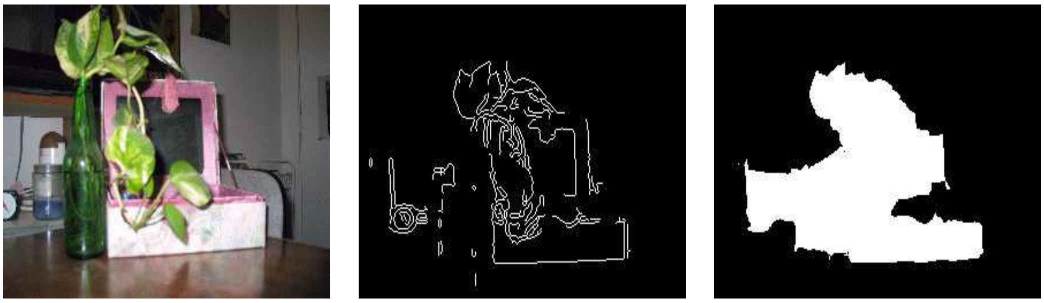
(a) (b) (c)

Figure 3.23: Detecting the main subject, the stuffed animal, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



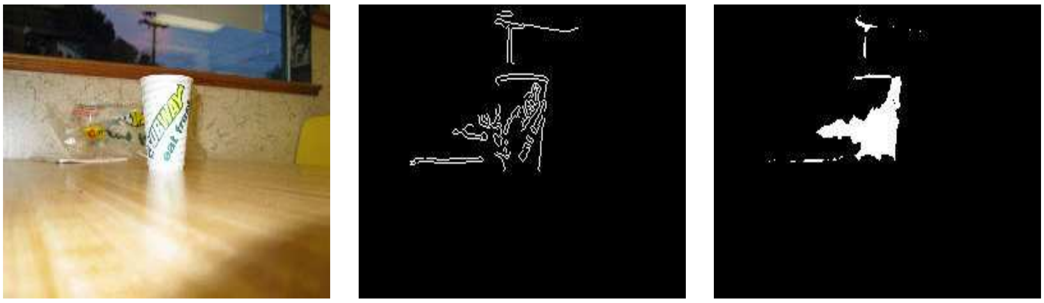
(a) (b) (c)

Figure 3.24: Detecting the main subject, the plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



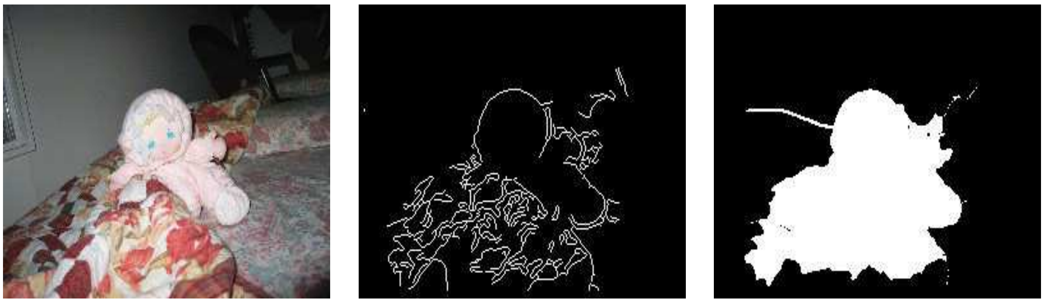
(a) (b) (c)

Figure 3.25: Detecting the main subject, the plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



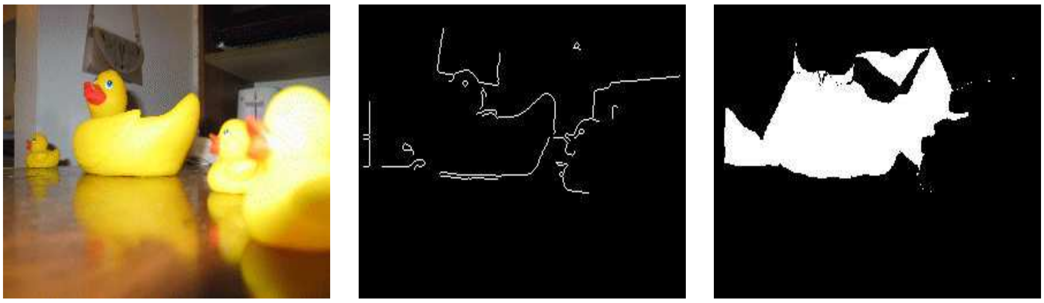
(a) (b) (c)

Figure 3.26: Detecting the main subject, the water cup, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



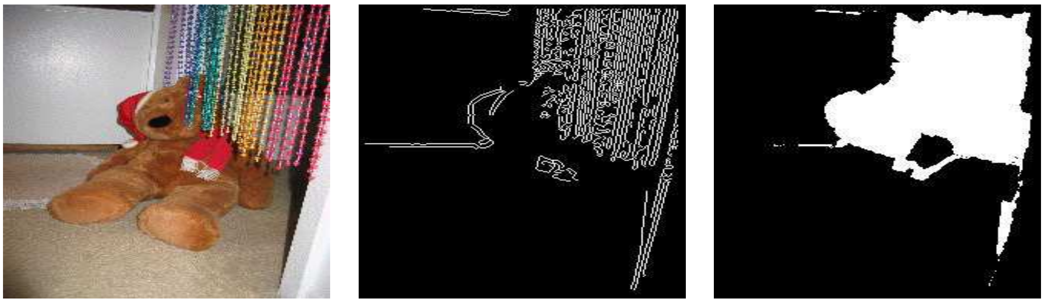
(a) (b) (c)

Figure 3.27: Detecting the main subject, the stuffed doll, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



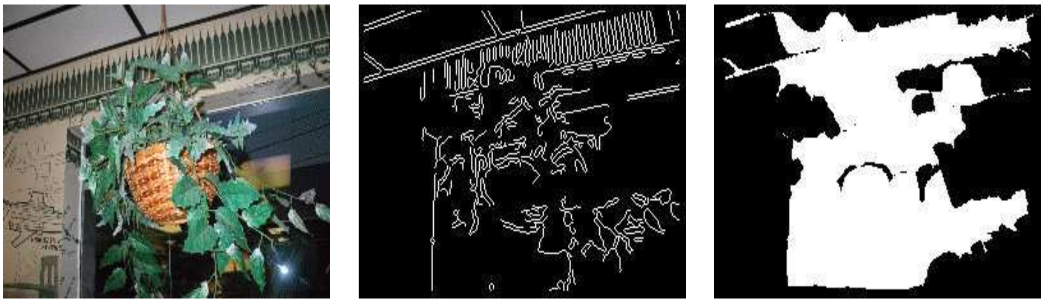
(a) (b) (c)

Figure 3.28: Detecting the main subject, the duck, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



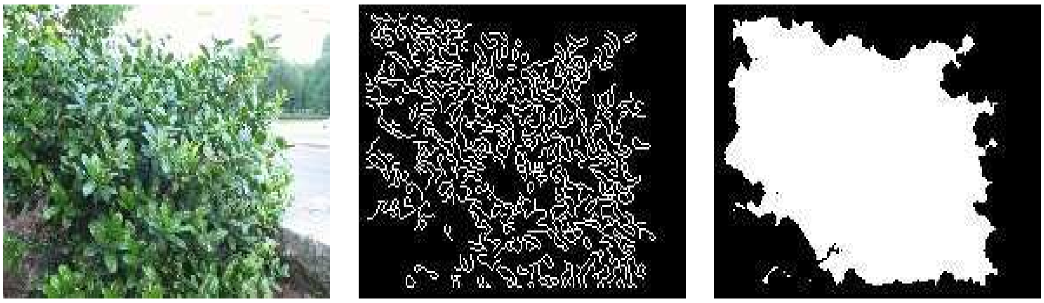
(a) (b) (c)

Figure 3.29: Detecting the main subject, the beaded curtain, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



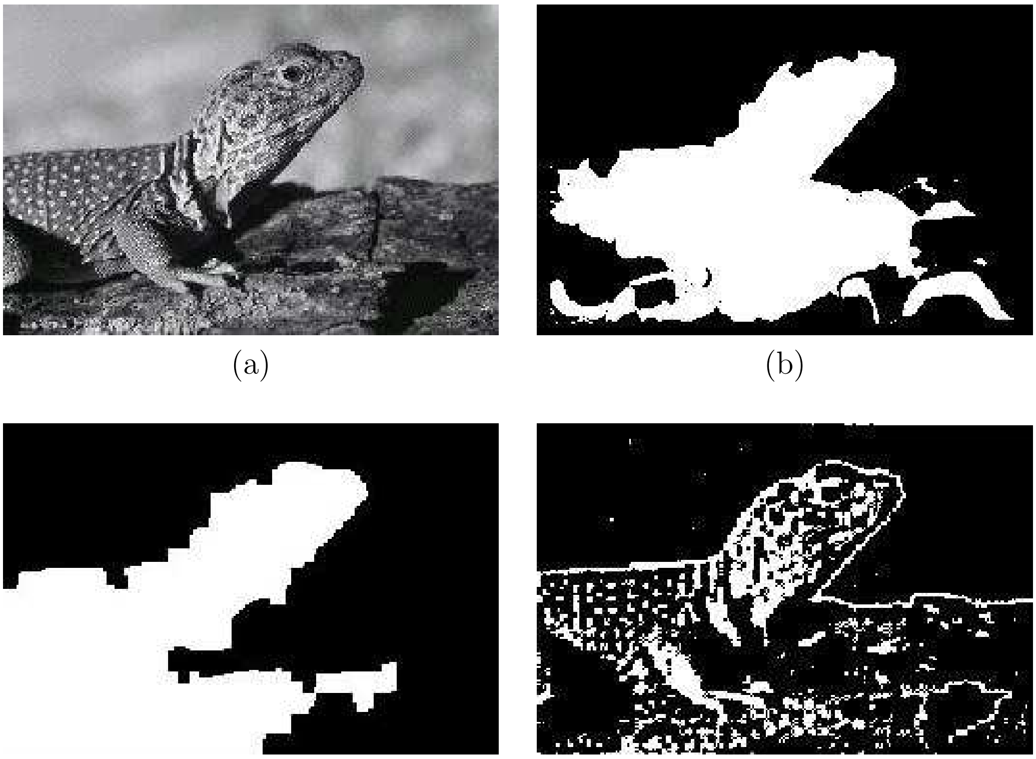
(a) (b) (c)

Figure 3.30: Detecting the main subject, the house plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



(a) (b) (c)

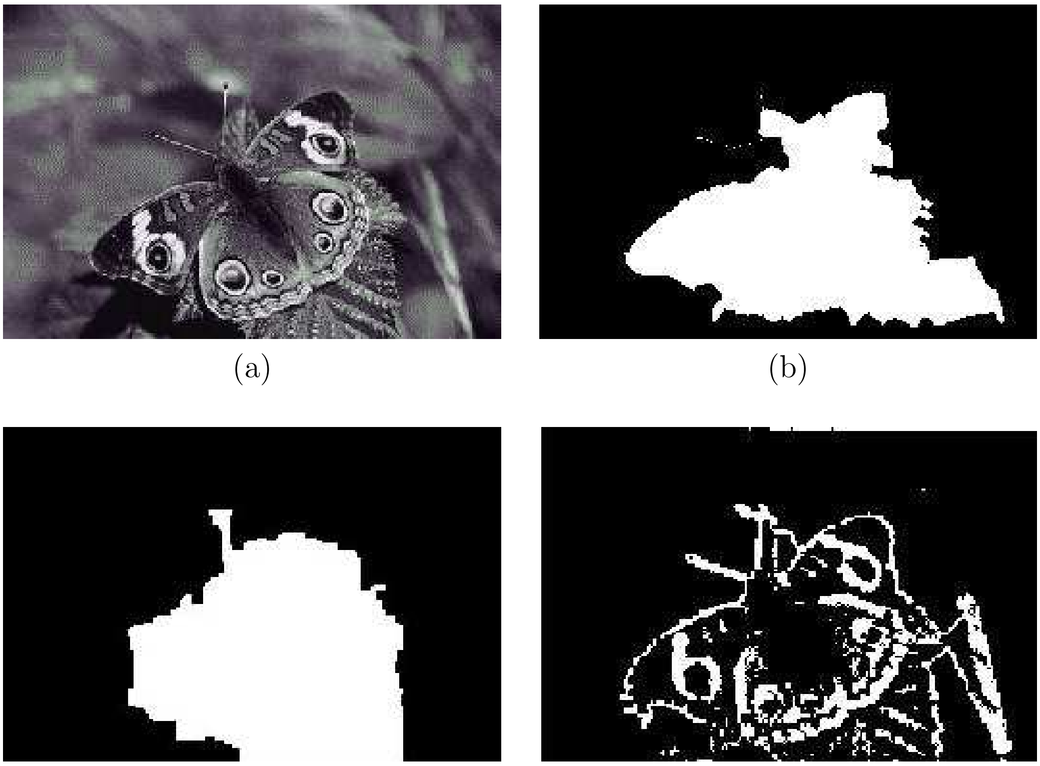
Figure 3.31: Detecting the main subject, the bush, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Rough outline of main subject; and (c) Detected main subject mask



(a) (b)

(c) (d)

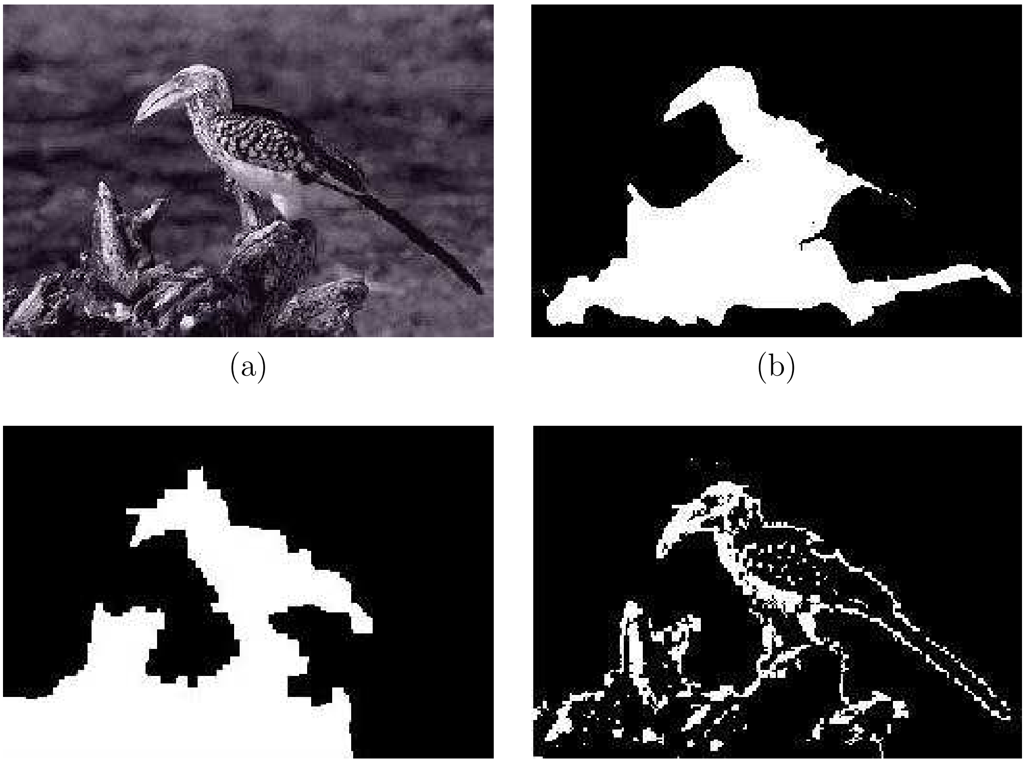
Figure 3.32: Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subject (the alligator) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et al.* multiscale Wavelet based approach; and (d) Detected main subject by Won's *et al.* maximum *à posteriori* probability estimation approach (the authors fill the segmented region with original gray levels for visual inspection).



(a) (b)

(c) (d)

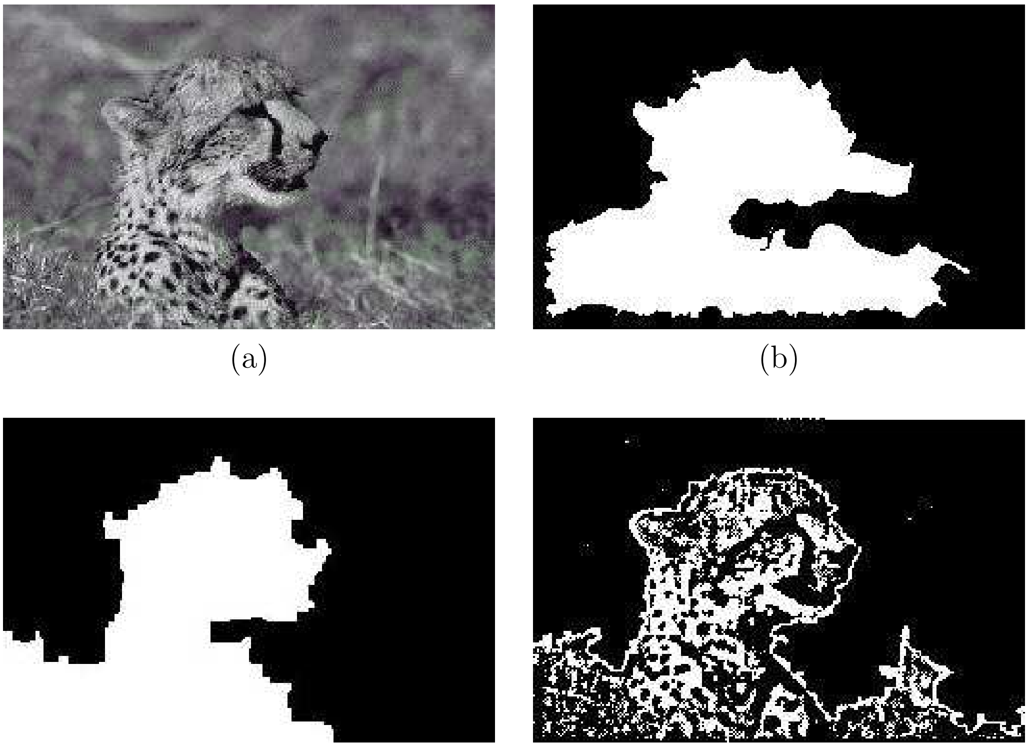
Figure 3.33: Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subject (the butterfly) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et al.* multiscale Wavelet based approach; and (d) Detected main subject by Won's *et al.* maximum *à posteriori* probability estimation approach (the authors fill the segmented region with original gray levels for visual inspection).



(a) (b)

(c) (d)

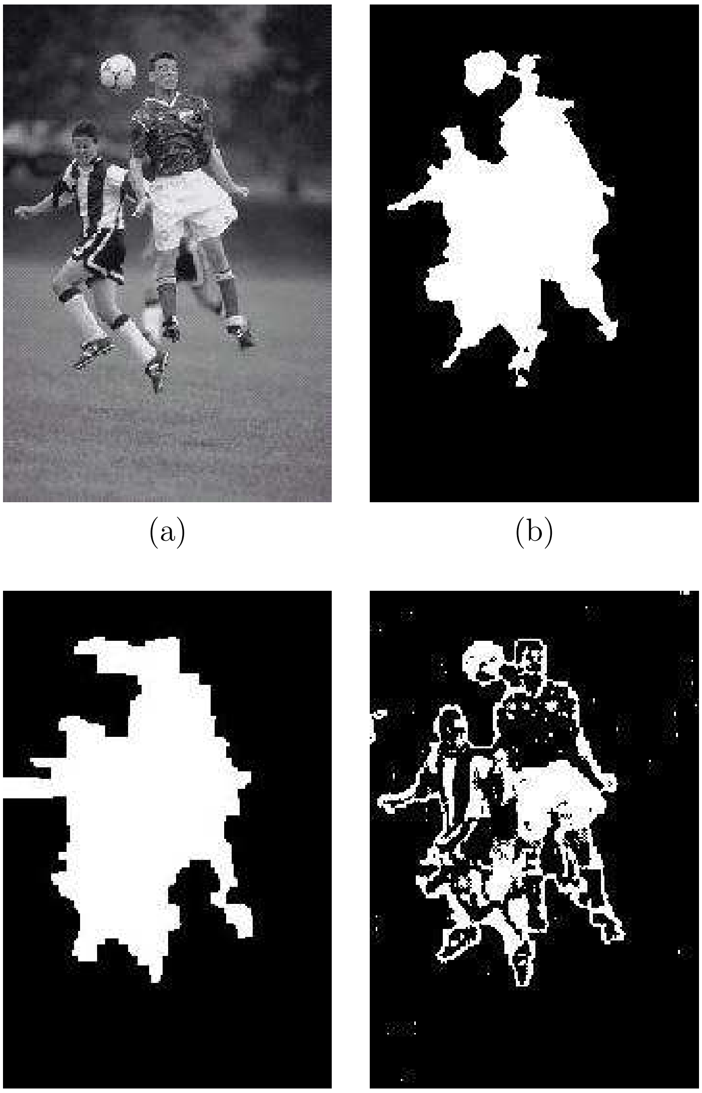
Figure 3.34: Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subject (the bird) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et al.* multiscale Wavelet based approach; and (d) Detected main subject by Won's *et al.* maximum *à posteriori* probability estimation approach (the authors fill the segmented region with original gray levels for visual inspection).



(a) (b)

(c) (d)

Figure 3.35: Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subject (the tiger) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et al.* multiscale Wavelet based approach; and (d) Detected main subject with Won's *et al.* maximum *à posteriori* probability estimation approach (the authors fill the segmented region with original gray levels for visual inspection).



(a) (b)

(c) (d)

Figure 3.36: Comparison of the proposed method with prevalent methods for main subject detection: (a) Original image, with the main subjects (the players) in focus; (b) Detected mask of the main subject with the proposed low–implementation complexity one–pass algorithm; (c) Detected mask by Wang's *et al.* multiscale Wavelet based approach; and (d) Detected main subject by Won's *et al.* maximum *à posteriori* probability estimation approach (the authors fill the segmented region with original gray levels for visual inspection).

Chapter 4

Automation of Photograph Composition Rules

"No great discovery was ever made without a bold guess."

Isaac Newton

When taking pictures, professional photographers apply photographic composition rules, e.g. rule-of-thirds and background blurring. The rule-of-thirds says to place the main subject's center at one of four places: at 1/3 or 2/3 of the picture width from left edge, and 1/3 or 2/3 of the picture height from the top edge. Background blurring is used to induce a sense of motion of the main subject or to create an effective distinction between the main subject and the background. This chapter develops unsupervised methods for digital still cameras to (1) realize the rule-of-thirds and (2) add artistic blurs to the image background. I also develop low-complexity algorithms for these methods.

The rule-of-thirds method moves the centroid of the main subject to the closest of the four rule-of-thirds locations. I first define an objective function that measures how close the main subject placement obeys the rule-of-thirds, and I then reposition the main subject. For multiple main subjects, the proposed algorithm could be extended by using rule-of-triangles by adding an appropriate constraint.

4.1 Introduction

This chapter describes automation of two photographic composition rules, namely the rule-of-thirds and background blurring. In the *rule-of-thirds* imaginary lines are drawn on the image canvas to divide the canvas into three equal parts in both horizontal and vertical dimensions. The main subject would need to be placed at one of four places where these imaginary lines intersect. By using this rule, photographers can produce pictures that are nicely balanced for the human eye in which each object in the frame tends to have good interaction with the others [2].

It may seem that the most harmonious and appealing image would be symmetrical — the main subject in the center of a picture or other elements distributed evenly. But such photographs strike us as static, because they lack two qualities that often elicit a viewer's interest — tension and movement. Placement of the main subject at a one-third point in the image frame otherwise lends a contextual importance to the scene. For example, a photographer wants to take a picture of a garden, where a yellow crocus has bloomed in front of purple crocuses. Fig. 4.1 shows an example, in which the yellow crocus (the main subject) is placed in the middle. In this picture, the background does not provide much support to the main subject. Fig. 4.2 is a more dynamic

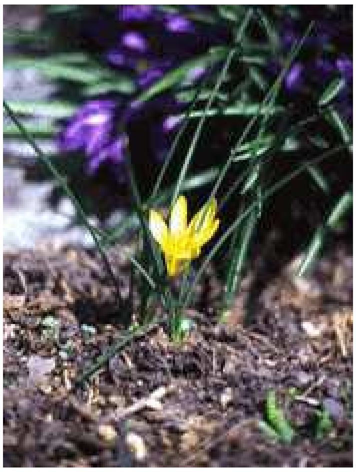


Figure 4.1: Picture with the main subject, the yellow crocus, in the middle.

alternative of the scene, and more pleasing to look at [2], as it accentuates the interaction between the yellow crocus and the surrounding violet crocuses.

Background blurring physically occurs when either the main subject(s) or the camera is in motion. Possible blurs include linear motion, circular motion or zoom blurs. These help in simulating the viewer's experience of motion, which is as much a physical sensation as it is an observed thing. Professional photographers also blur the background to reduce the depth of field in the picture. This in turn gives more attention and importance to the main subject(s) in the picture [2].

Pleasing addition of background blurs are also good for constrained transmission of images [84]. For example, Fig. 4.3 shows the original image that would be good for printing. Fig. 4.4 shows the same image with the background blurred. The blurring produces a new appeal in the image, by giving more importance to the main subject, the shell. Also, Fig. 4.4 produces more savings in file size compared to Fig. 4.3, when compressed with image com-

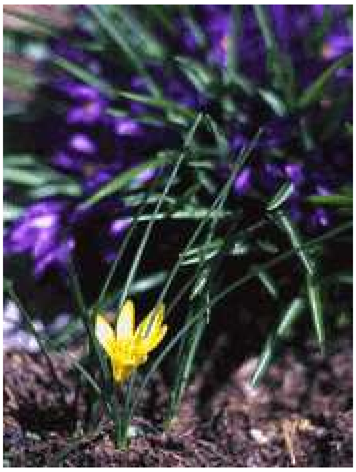


Figure 4.2: The same picture where the main subject, the yellow crocus is placed by following the rule-of-thirds. This picture is more dynamic and pleasing to look at as it shows interaction of the yellow crocus with the surrounding violet crocuses.

pression softwares. For example, by using JPEG compression software, there is around a 32% savings in file size of the blurred image (Fig. 4.4) verses the original one (Fig. 4.3) for the same JPEG quality factor due to the background blurring [84]. The images are of dimension 400 × 416 pixels.

4.2 Rule-of-Thirds: Automated Placement of the Main Subject

For this rule, the post-segmentation objective is to automatically place the main subject by following the rule-of-thirds. The rule-of-thirds says to place the main subject in one of four places: at 1/3 or 2/3 of the picture width from left edge, and 1/3 or 2/3 of the picture height from the top edge. A mathematical measure is defined to check how close the picture follows the

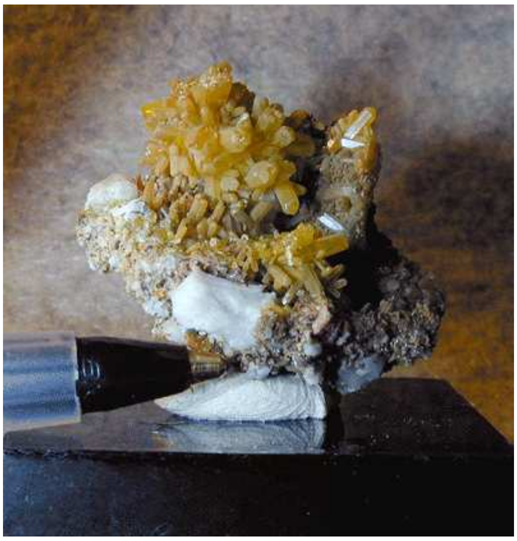


Figure 4.3: Original image where the main subject, the shell, and the image background are equally focused. This picture is good for printing.

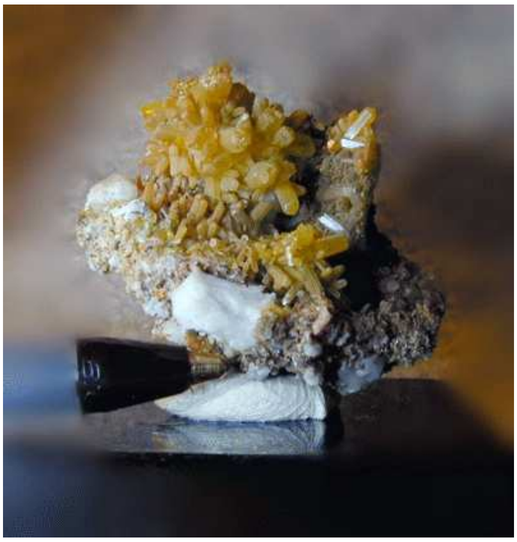


Figure 4.4: The same picture where the background is blurred. This adds an added appeal to the main subject, the shell. The processed picture is better for constrained image transmissions (for example, over the World Wide Web) than Fig. 4.3.

rule-of-thirds, and this measure is optimized to reposition the main subject.

4.2.1 Algorithm Formulation

Let C be the scene domain of the main subject where $C = \{\mathbf{v}|\mathbf{v} \in \text{Main subject}\}$ such that $\mathbf{v} = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$ is the set of pixel positions. Then, the center of mass is defined as the weighted sum of the components and cardinality of the scene domain. Consider that there are n main subjects. The center of mass for each of them is computed independently. A 2-dimensional function $f(x, y)$ is defined such that it reaches a minimum when a center of mass is at the one-third position in the canvas both along the x and y axis. The objective will be to minimize the summation of the value of the function generated by the center of mass positions (x'_n, y'_n) of the n main subjects.

4.2.2 Proposed Algorithm

For the current implementation, I assume that there is one main subject (i.e. $n = 1$), and the function $\chi(x, y)$ is a product of the Euclidean distance from the four one-third corners on the canvas. Let $\mathbf{v_1} = (x_1, y_1)$, $\mathbf{v_2} = (x_2, y_2)$, $\mathbf{v_3} = (x_3, y_3)$ and $\mathbf{v_4} = (x_4, y_4)$ be the four one-third corners. And, $\mathbf{v} = (x, y)$ is the position of the center of the mass of the main subject. Then,

$$\chi(x, y) = \prod_{i=1}^{4} (\mathbf{v} - \mathbf{v_i})^2 \tag{4.1}$$

So, $\chi(x, y) \geq 0$ with $\chi_{min}(x, y) = 0$, and the minimum is attained when the center of mass is at one of the one-third corners. Fig. 4.5 illustrates the function $\chi(x, y)$ on a 300×150 pixel canvas. Thus, after computation of the center of mass, the image pixels are shifted so that the center of mass falls at a one-third corner.

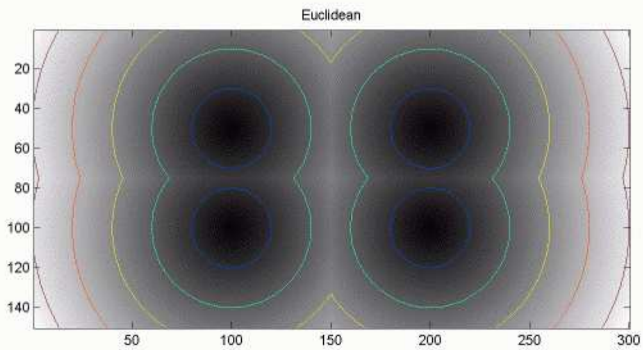


Figure 4.5: The defined function, $\chi(x, y)$, that illustrates automation of the rule-of-thirds. This function attains a minimum value of zero when the center of mass of the image falls at a one-third corner. It is positive otherwise.

The center of mass is computed along the rows and columns respectively. For each row (or column) if w_n is the number of "ON" pixels in the detected main subject mask, then the center of mass is defined as

$$center = \frac{w_n * \text{row (or column) location}}{\Sigma w_n} \tag{4.2}$$

After computing the center of mass, a comparison is made as to which of the four one-third corners is closest to the current position of the center of mass. The picture is then shifted so that the center of mass falls at the closest one-third corner.

Preliminary subjective observations show that if the main subject is cropped, during the shifting process, the generated shifted picture looses some appeal. So, an additional constraint is added, to avoid the main subject from being cropped after it has been shifted. In this case, the algorithm automatically chooses to move the main subject to the next closest one-third corner. Whether the main subject would be cropped during shifting or not is determined by comparing the distance of the center of mass of the main subject to the desired one third corner point, and the distance of the edge of the main subject mask to the image boundary.

4.2.3 Implementation Complexity

By using the main subject mask, the rule-of-thirds algorithm requires 2 multiply-accumulates, 1 comparison, and 1 or 3 memory access per pixel, plus eight comparisons and one division (explained below) for the entire image. One memory access per pixel is needed to calculate the center of mass. An additional two memory accesses per pixel is needed only if the picture is shifted instead of cropped.

For automated placement of the main subject by following the rule-of-thirds, the center of mass for the detected main subject mask is computed with 2 multiply-accumulates and 1 memory read per pixel, and one division per image. The closest one-third corner is computed with eight comparisons (four comparisons of each of the $x-$ and $y-$coordinates). The next step is to alter the picture so that the center of mass of the main subject is at one of the four one-third corners.

One approach is to crop the picture so that the center of mass of the main subject falls on one of the one-third corners. This is computationally very simple. During cropping the two competing criteria to optimize are (1) moving the center of mass as close to one of the four one-third points as possible, and (2) minimizing the number of rows and columns cropped in the picture to retain the most picture content possible, subject to the constraint that no pixels of the main subject are cropped.

In this dissertation instead of cropping the picture, every pixel in the entire image is shifted by the same amount so that the center of mass of the main subject occurs at one of the one-third corners. However, if shifting the picture to the closest one-third corner crops the main subject, the picture is shifted to the next closest one-third corner. After shifting the image, many pixel values along two of the edges of the image would be undefined. For simulation purposes, depending on the picture, these pixels could be given values through pixel replication or boundary pixel duplication along the boundary of known pixel values. As a practical implementation viewpoint, it might be possible to use a wide angle lens camera, and capture a picture with a broader view. This way, when the pixels are shifted to follow the rule-of-thirds, there would not be any undefined pixels along the image edges. The shifting approach requires one memory read and one memory write per pixel.

In the best case, the center of mass falls at one of the one-third corners so that the image does not have to be altered. In the worst case, the center of mass is at one of the corners of the picture so that one-third of the rows and one-third of columns would be cropped or need to be given new values. In the average case, e.g. if the main subject were originally in the middle of the picture, one-sixth of the rows and one-sixth of the columns would be cropped or be given new values.

4.2.4 Results

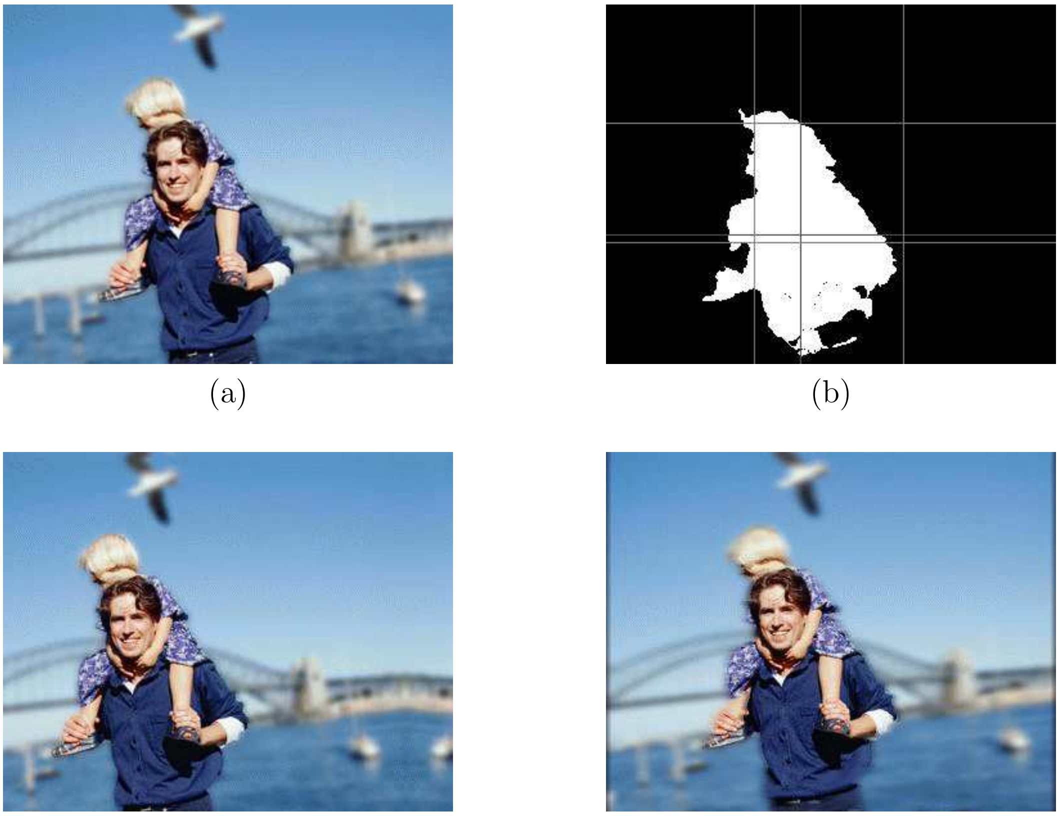
The algorithms were tested on about 30 low depth of field pictures. Some pictures were obtained from the World Wide Web, and I acquired the others with a Cannon Powershot G3 camera. I kept the shutter aperture of the camera at F2. For some pictures the aperture was varied from F2 through F2.8. I varied the shutter speed from $\frac{1}{60}$ seconds through $\frac{1}{250}$ seconds depending on wether I was using a tripod or not. I used faster shutter speeds when I was hand holding the camera while taking the picture. I acquired the pictures under different conditions, such as different settings (indoor or outdoors), lighting (day light, incandescent light and filament bulb light), main subjects (human beings or inanimate objects), camera orientations (landscape or portrait modes), and distances to the main subject. The test set was very diverse to show that the algorithms are independent of scene setting or content. While acquiring the pictures, I generally placed the main subject in the middle of the camera as most amateurs do. However, for some pictures the main subject was closer to following rule-of-thirds.

The original pictures are shown in Figs. 4.6(a) through 4.16(a). Figs. 4.6(b) through 4.16(b) show the detected main subject masks, the 1/3 and 2/3 lines on the canvas along the height and width, respectively, and the position of the center of mass of the detected main subject. Figs. 4.6(c) through 4.16(c) show the main subjects repositioned by following the rule-of-thirds.

In Figs. 4.6 and 4.7 the main subjects are humans in outdoor settings. In Fig. 4.8 (a) the main subject, the stuffed animal, is shifted to the second nearest one-third corner, so that the main subject does not get cropped in Fig. 4.8 (c). Figs. 4.9 and 4.10 are close up shots of a house plant in indoor settings. Figs. 4.11, 4.12 and 4.13 are indoor pictures with inanimate main subjects. In Fig. 4.14 a cognitive model would pick the stuffed bear to be the main subject. However, due to depth of focus, the beaded curtains are sharper and hence chosen by the algorithm to be the main subject. Also, in this picture, the detected main subject almost follows the rule-of-thirds. So, the picture could have been left unprocessed by using a little leeway in the algorithm. In Figs. 4.15 and 4.16 the plants are the main subjects in indoor and outdoor settings, respectively.

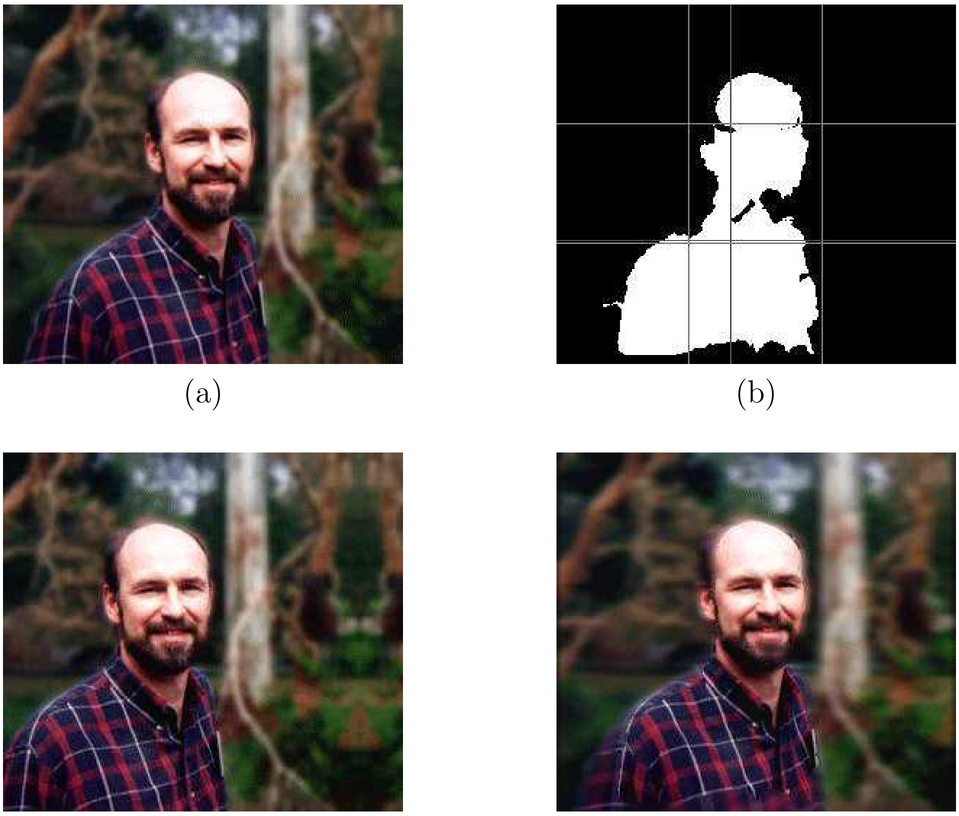
For some pictures I use mirror reflection and for others I extend the boundary pixels for the undefined pixels are shifting the main subject. Usually if the borders are smoother, I use boundary extension, and textured or high frequency borders, I use mirror reflection. It can be seen that for some results visible artifacts can be seen after either of these techniques. But, these artifacts could be reduced by capturing an image by using a wide-angled lens camera. In that case, even after shifting the picture to follow the rule-of-thirds, the boundary pixels would still be valid.

For multiple main subjects in the photograph, the proposed algorithm could be extended for automation of the *rule-of-triangles* [2]. The rule-of-triangles states that if there is more than one main subject in the picture, then their centers of mass should not lie on the same line in the canvas, but should form triangles on the canvas. This can be automated by adding a constraint during minimization so that no two center of masses lie on the same row in the canvas.



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

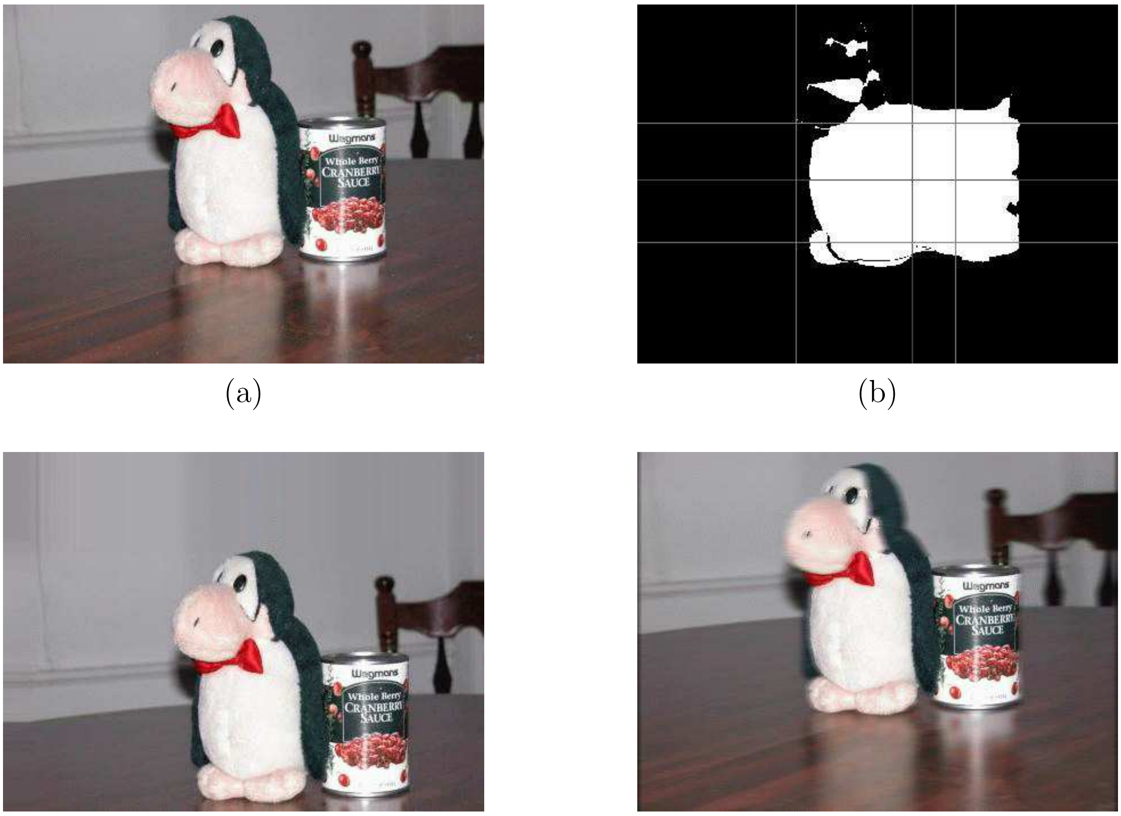
Figure 4.6: Automation of photographic composition rules by detecting the main subject, the man and the child, which are in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



(a) (b)

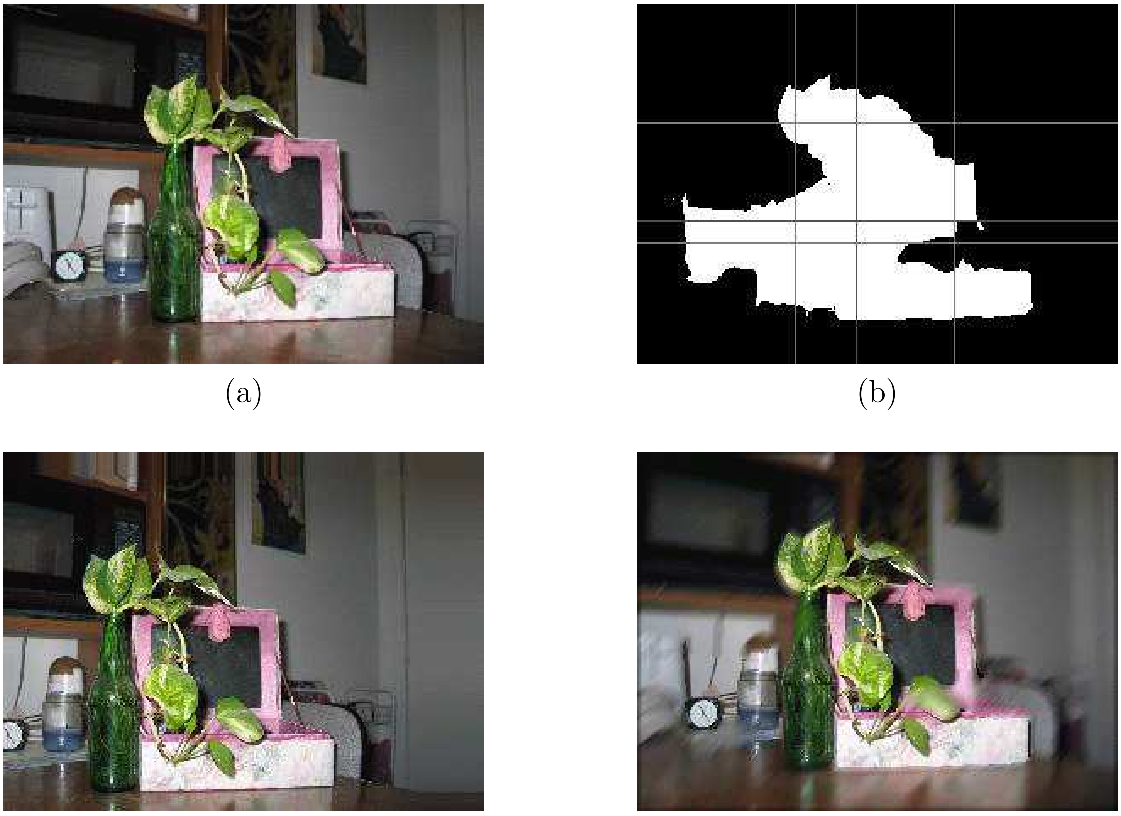
(c) (d)

Figure 4.7: Automation of photographic composition rules by detecting the main subject, the man, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



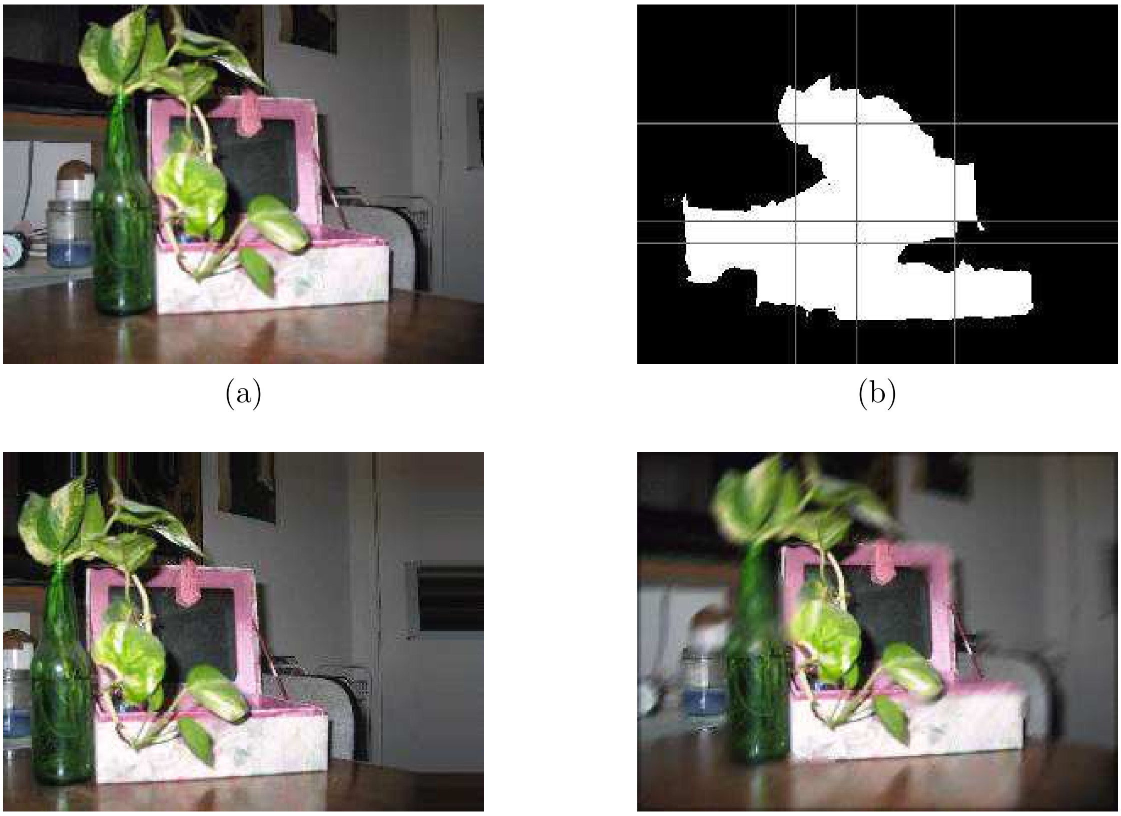
(a) (b) (c) (d)

Figure 4.8: Automation of photographic composition rules by detecting the main subject, the stuffed animal, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds (the picture does not crop the main subject, the stuffed animal); and (d) Simulated background blur which could result from camera panning.



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

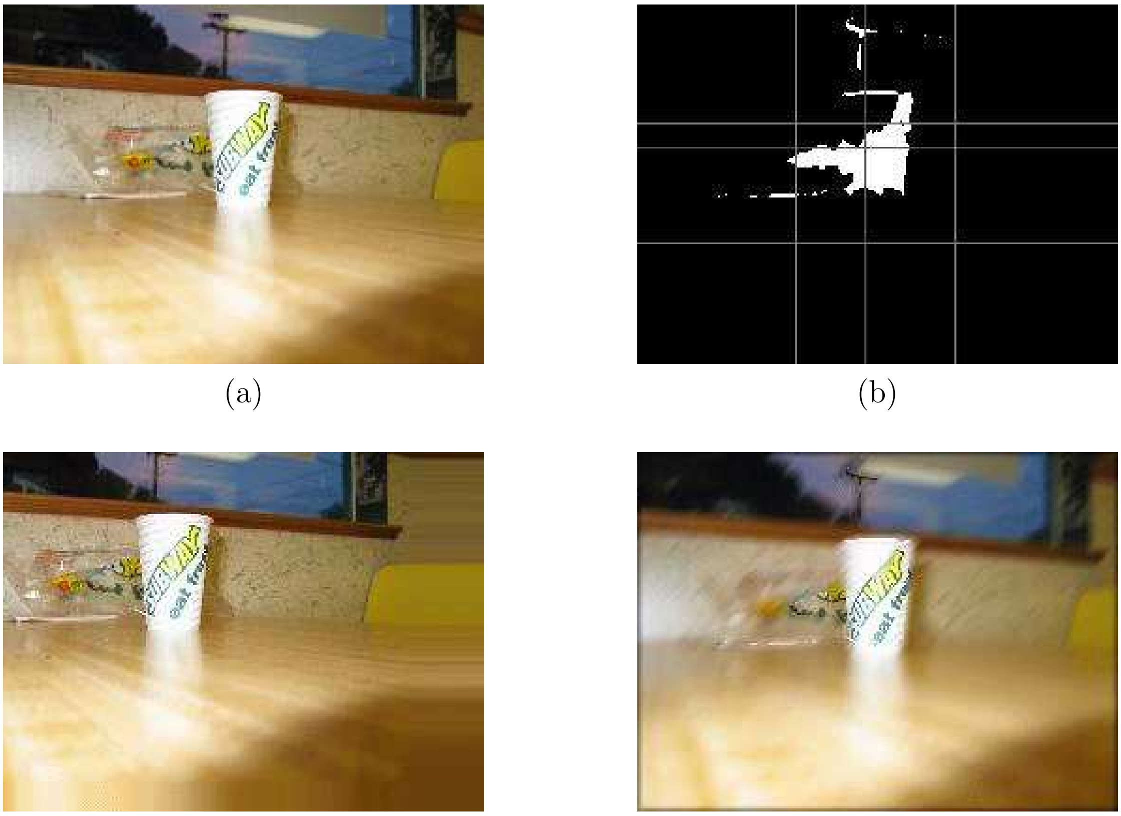
Figure 4.9: Automation of photographic composition rules by detecting the main subject, the plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



(a) (b)

(c) (d)

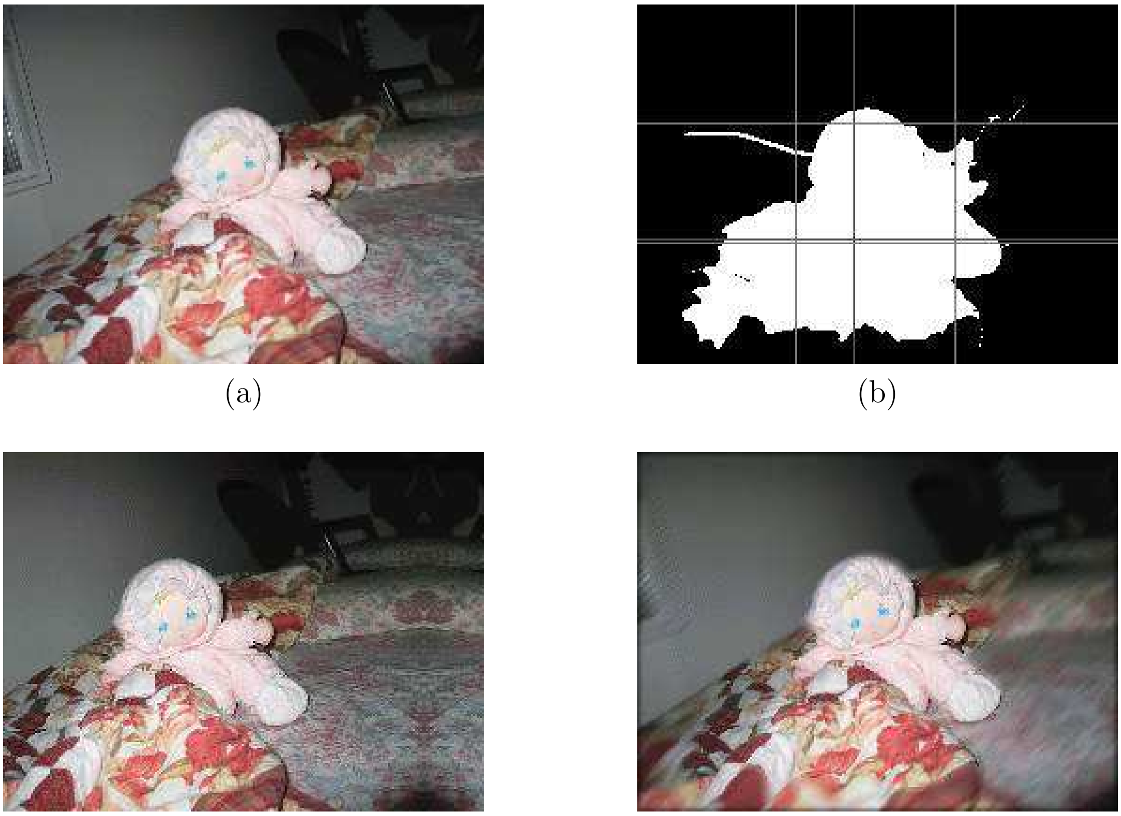
Figure 4.10: Automation of photographic composition rules by detecting the main subject, the plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



(a) (b)

(c) (d)

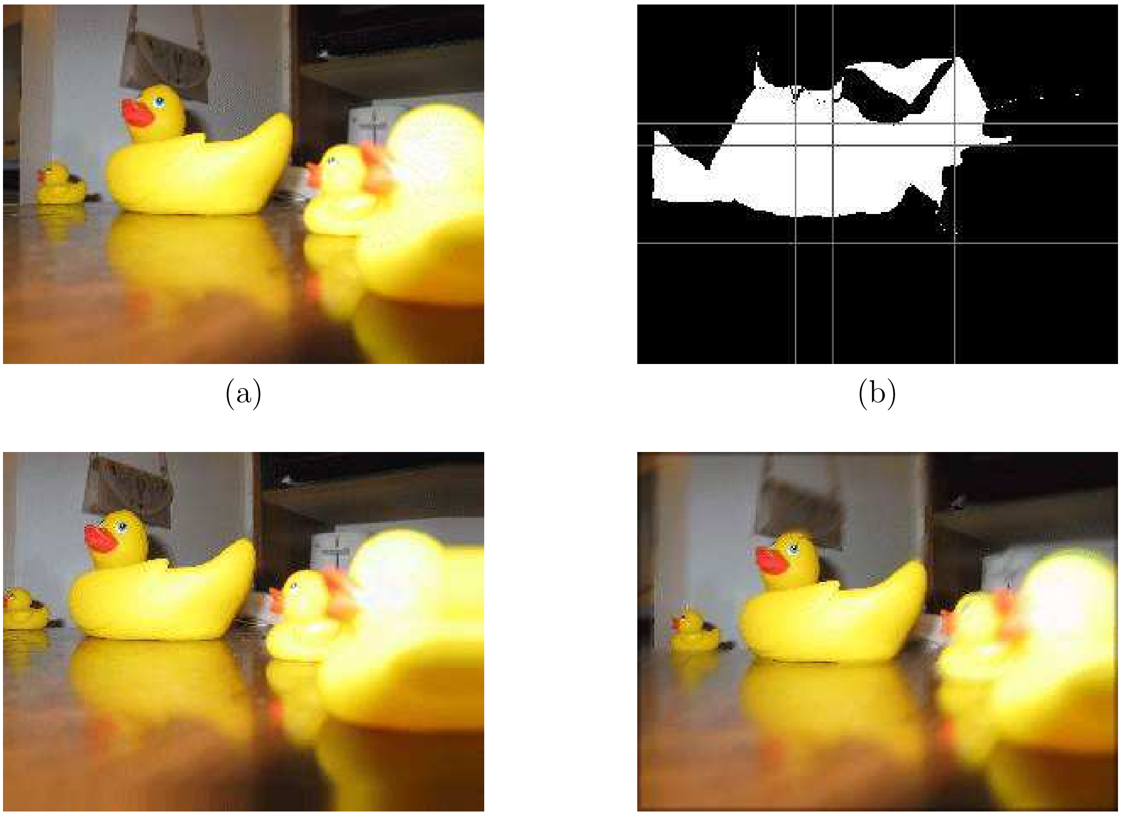
Figure 4.11: Automation of photographic composition rules by detecting the main subject, the water cup, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



(a) (b)

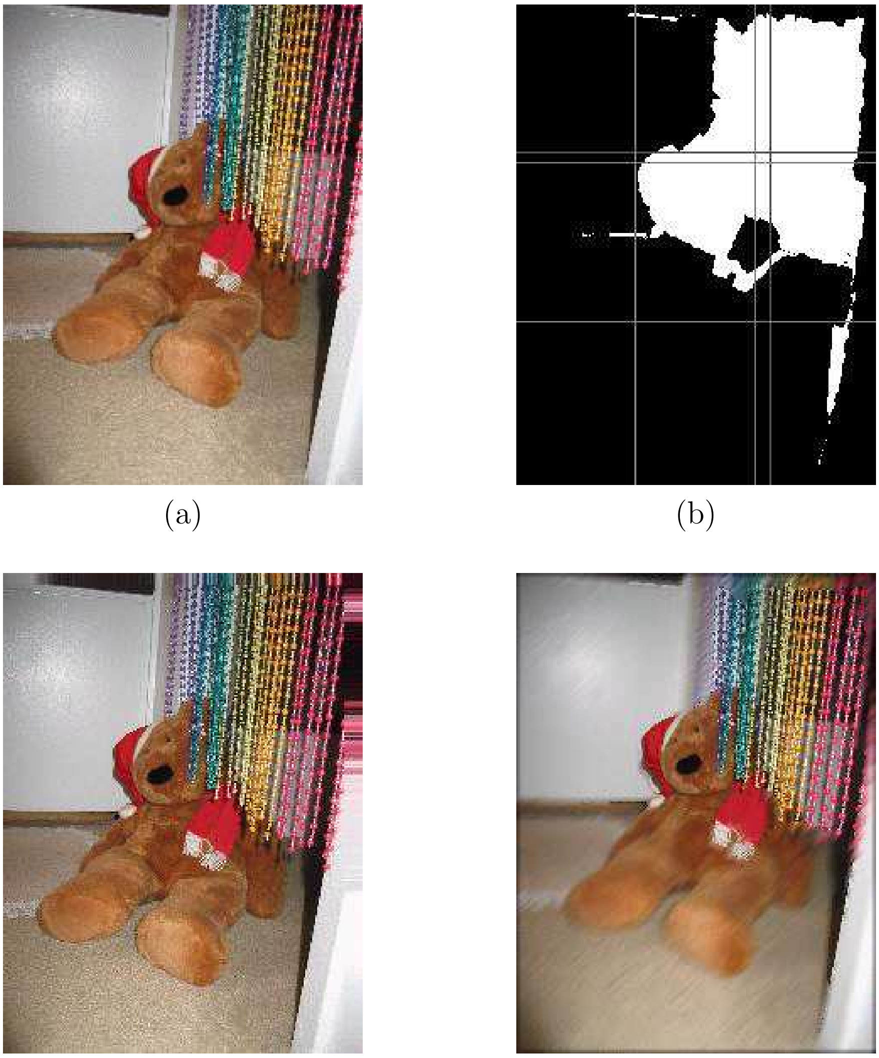
(c) (d)

Figure 4.12: Automation of photographic composition rules by detecting the main subject, the stuffed doll, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



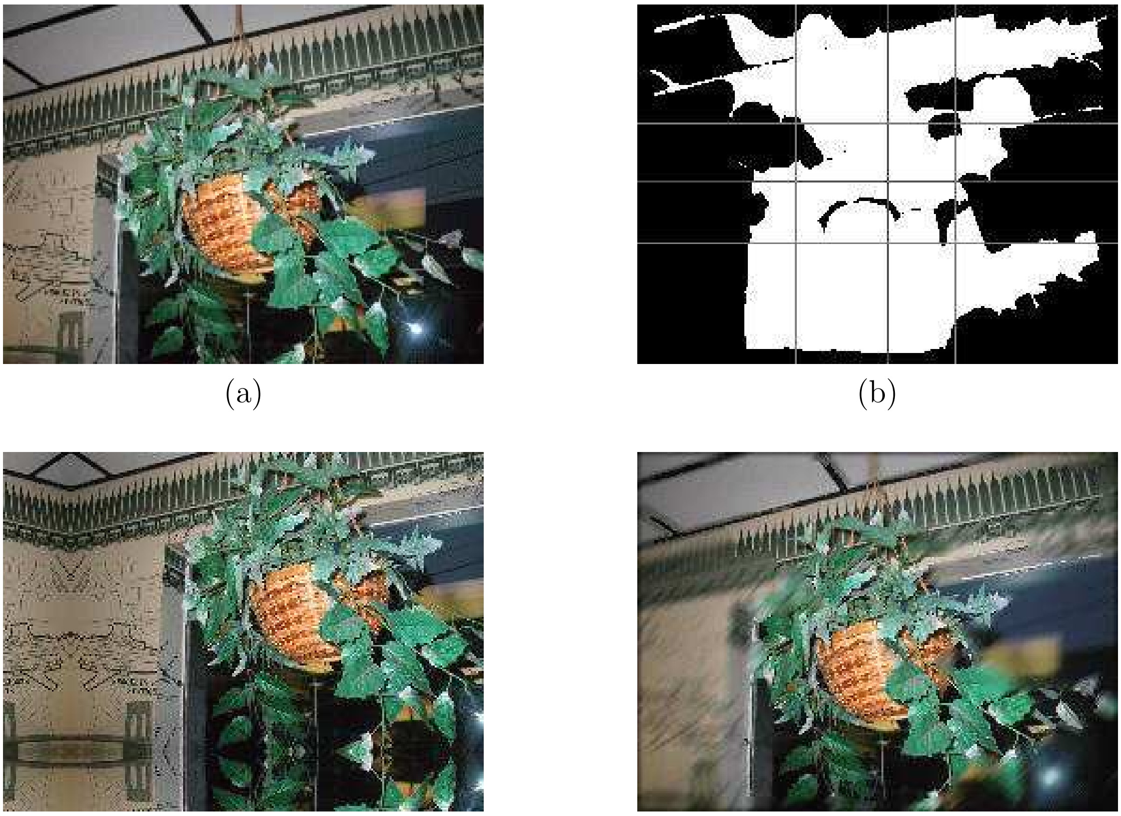
(a) (b) (c) (d)

Figure 4.13: Automation of photographic composition rules by detecting the main subject, the duck, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



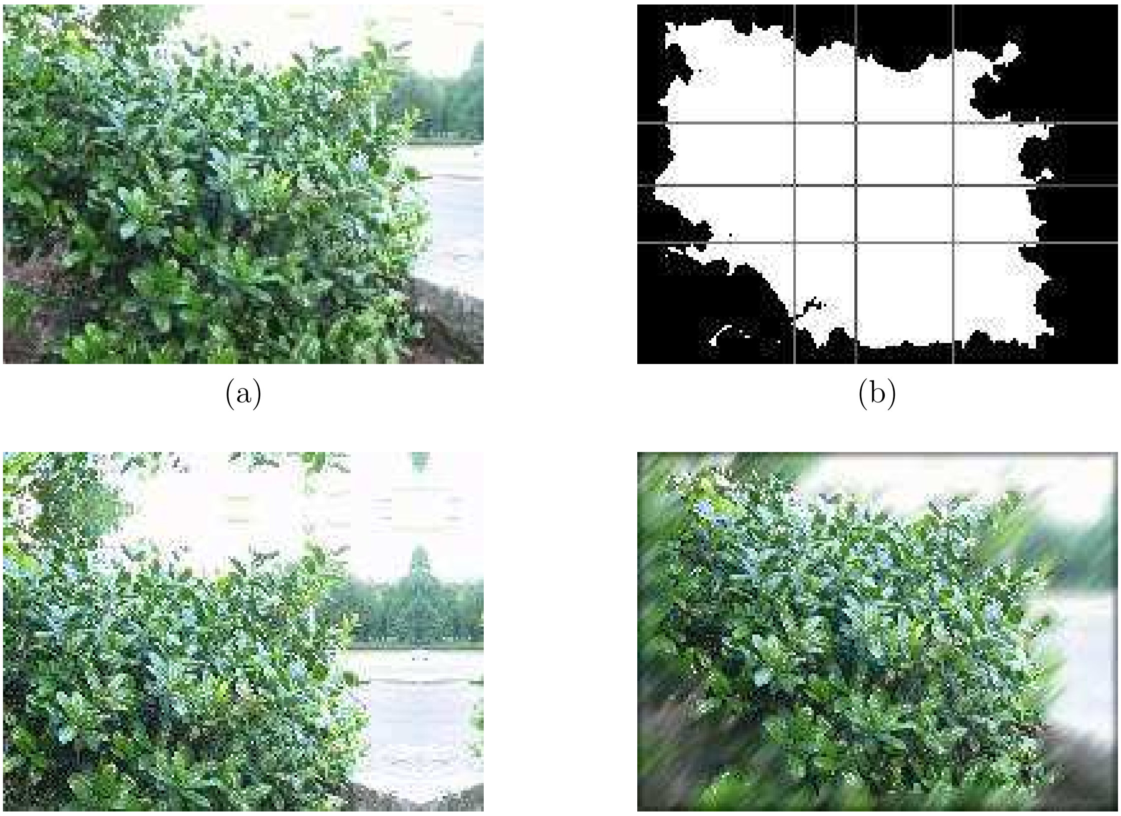
(a) (b) (c) (d)

Figure 4.14: Automation of photographic composition rules by detecting the main subject, the beaded curtain, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.


(a) (b)

(c) (d)

Figure 4.15: Automation of photographic composition rules by detecting the main subject, the house plant, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.



(a) (b)

(c) (d)

Figure 4.16: Automation of photographic composition rules by detecting the main subject, the bush, which is in focus: (a) Digital image with background blur from large shutter aperture; (b) Detected main subject mask, with center of mass not following the rule-of-thirds; (c) Generated picture obeying rule-of-thirds; and (d) Simulated background blur which could result from camera panning.

4.3 Motion Effects Rule: Simulating Background Blurring

For simulating background blur, the original image is first masked with the main subject mask detected by the algorithm proposed in Chapter 3. Then region of interest filtering is performed on the masked image so that the main subject pixels remain unaltered and the background pixels are altered to add artistic effects.

4.3.1 Algorithm Description

When the main subject or the camera is in motion, linear, radial, or zoom blurs could occur. A linear model for the imaging process is defined as a convolution of the original image, $I(x, y)$, with a space-invariant point-spread function, $h(x, y)$ [67]. So, the observed image, $\gamma(x, y)$, can be represented as

$$\gamma(x,y) = (I * h)(x,y) \tag{4.3}$$

In case of linear blur the point-spread function, $h(x, y)$ can be expressed as

$$h(x,y) = \begin{cases} \frac{1}{vt}\delta(y) & 0 \leq |x| \leq vt\cos(\alpha), y = \sin(\alpha) + vt \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

where v is the motion velocity, t is the exposure time, α is the angle of the blur and δ is the Dirac delta functional.

The discrete equivalent [85] of the point-spread function for a linear blurring distance of L, and an angle of α, is given by

$$h(x,y) = \begin{cases} \frac{1}{L+1} & 0 \leq |x| \leq (L+1)\cos(\alpha), y = \sin(\alpha) + (L+1) \\ 0 & \text{otherwise.} \end{cases} \tag{4.5}$$

Here, L is the number of additional points in the image resulting from a single point in the original scene.

The radial blur occurs when there is circular motion of the camera or the main subject, and a zoom blur occurs when there is a change of scale during the image acquisition process. In this dissertation, I simulate these possible motion blurs to create an appealing alternative of the acquired picture.

4.3.2 Implementation Complexity

Given the main subject mask, background blurring can be implemented in two ways. In the first case, the original image may be masked and the blur filter can be used to filter the non-masked regions. At the main subject edges, the higher frequencies would be cut off by the filter, leading to ringing artifacts. In the second approach, the entire image is first low-pass filtered, and then the pixels contained in the main subject mass are superimposed with the low-pass filtered image. I choose to use the second approach to avoid ringing artifacts. The image background can be blurred with a 3×3 Gaussian low-pass filter with $\sigma = 0.2$. Depending on the amount of blur required the filter dimension or the σ could be changed. Thus, background blurring requires 9 multiply-accumulates and 4 memory accesses per pixel.

4.3.3 Results

I convolve the images with a motion blur filter that simulates linear and radial blurs produced by horizontal and rotational movement of the camera. The filtering involves convolving the image with a series of filters and compositing the filtered images. Figs. 4.6(d) through 4.16(d) show simulated background blurring that could have resulted from camera panning. The current example simulates linear motion of the camera by 10 pixels. Other values of linear, radial, or zoomed motion blurs can also be simulated. As can be seen from the results the algorithms can be applied independent of the scene setting or content.

4.4 Conclusion

This chapter presents algorithms for automating two photographic composition rules. One of them guides the placement of the main subject on the canvas. The other could be used to add simulated artistic blurs to the image background, by knowing the placement of the main subject. A digital still camera uses approximately 200 digital signal processor instruction cycles per pixel. Automating background blurring or rule-of-thirds requires far fewer digital signal processor cycles. The proposed algorithms are amenable for implementation completely in fixed-point data types and arithmetic.

Other photographic composition rules that can be automated with the framework proposed in this chapter are

- *Automation of the best possible zoom:* Given the main subject mask, the amount of zoom could be selected based on the photograph content.
- *Taking a picture through frames available in the scene:* Sometimes objects in the scene can be grouped together and the picture of the main subject be taken so that it appears that the main subject is framed. Automating this effect would involve first detecting the available frames in the picture, and repositioning the main subject to fall within one of

the detected frames.

- *Placing the main subject where lines of the scene intersect:* After detecting straight lines in the picture, the main subject could be relocated to where these lines intersect, for an added appeal of the picture. This way the observers' eye is automatically drawn to the main subject.

All of the above rules can be applied irrespective of the background content. The subsequent chapter deals with automating another photographic composition rule, where background content also needs to be considered.

Chapter 5

Merger Detection and Mitigation

> *"When asked what single event was most helpful in developing the Theory of Relativity, Albert Einstein replied, "Figuring out how to think about the problem"."*
>
> W. Edwards Deming

When taking pictures, professional photographers apply photographic composition rules, e.g. avoidance of mergers. A merger occurs when equally focused foreground and background regions appear to merge as one object. This chapter presents an unsupervised algorithm that (a) detects background objects merging with the main subject, and (b) reduces the visibility of merging background objects. The main subject is detected by using the algorithm presented in Chapter 3 and outlined in Fig. **??**. Detection of the main subject requires automated adjustment of optical settings, optical blurring of objects not in focused, and digital image processing. The rest of the merger detection and mitigation algorithm does not adjust or use the camera settings. The algorithm does not make assumptions about the scene setting (indoor/outdoor) or content. The algorithm is amenable to implementation in fixed-point arithmetic.

5.1 Introduction

During image acquisition, the three-dimensional world is mapped to a two-dimensional picture. Professional photographers change camera settings so that the main subject is in focus, while the objects in the background that merge with the main subject are blurred [2]. This preserves the sense of distance between the objects in the photograph. However, amateur photographers often take pictures in which an object in the background remains in focus, seems to be a part of the main subject, and appears to be merged with the main subject frame. Fig. 5.2(a) shows an example of a merger in which the trees appear to grow out of the main subject, the man's head. Other examples include a horizontal line shooting through the subject's ears, and a knee or elbow extending from the frame edge.

The previous chapter describes automation of photographic composition rules that can be applied by knowing the position of the main subject alone. This chapter discusses automation of merger removal in photographs, and hence both the main subject and the image background will be considered. In-focus objects other than the main subject, which are adjacent to the main subject, often produce annoying effects in the photograph. These objects seem to merge with the outline of the main subject and confuse the shape of the main subject. The work presented in this chapter aims at identifying the background object that seems to merge with the main subject. The detected object is then blurred to make it identifiable as a part of the background rather than having the object appear to be part of the main subject.

This chapter presents a method that automatically identifies the background objects that merge with the main subject. The unsupervised method classifies merged objects without using *á priori* assumptions on scene content and indoor/outdoor setting. First, I generate the set of background objects by performing color segmentation on the background image. Then, each background object is classified as a merging object or a non-merging object according to the following features:

- Distance to the main subject: a merging object appears to adjoin the main subject
- Magnitude of gradient values: a merging object, because it is in focus, will have gradient values that are similar to those of the main subject, which is also in focus.

Finally, I blur the merging objects to make the merging objects appear farther behind the main subject.

Section 5.2 formulates the algorithm for merger detection and mitigation. Section 5.3 analyzes the complexity of the proposed algorithm. The results are presented in Section 5.4. Section 5.5 concludes this chapter by summarizing the algorithm and discussing possible extensions.

5.2 Merger Detection and Mitigation: Formulation

The method proposed in Chapter 3 generates a main subject mask that divides the picture into foreground and background regions. Fig. 5.2(b) is the generated main subject mask for Fig. 5.2(a). After the segmentation into foreground and background regions, the goals will be to segment the background, identify merging objects, and blur the merged part in the picture. The theoretical formulations follow.

5.2.1 Background Segmentation

The color information is used for segmentation of the background objects. The red, green, and blue (RGB) image provided by the camera is transformed to the hue channel found in the hue, saturation, value (HSV) space. In HSV space, hue corresponds to color perception, saturation provides a purity measure, and value provides the intensity [86]. A histogram in the hue space is then utilized for segmentation of the background region. Although hue does not model the color perception of the human visual system as accurately as CIELab [86], it is chosen because the transformation from RGB to hue has lower implementation complexity. RGB to CIELab requires calculation of cube roots.

Let the hue values be on the interval $[0, 255]$ and broken into m-bins. The discrete probability distribution for hue values belonging to each bin is

$$P(\mathrm{hue}_i) = \frac{c(\mathrm{hue}_i)}{T_c} \tag{5.1}$$

where $c(\mathrm{hue}_i)$ is the count corresponding to each bin and T_c is the total count of values in all bins. By modeling the background picture as a Gaussian mixture of hue values, the task is to further segment these m-bins into n-groups, where each group will identify a different object.

The term $\frac{T_c}{m}$ gives the average of the hue values. Any hue value above this average is marked as a dominant hue. Based on the available dominant hues, the n-groups are determined automatically so that each group contains only one dominant hue. Each group boundary lies halfway between two of the dominant hues. This ensures that the local maximums of the probability distribution, $P(\text{hue}_i)$, is captured in each group. Pixels with hue values falling in each of the identified n-groups form different background objects. For the proposed algorithm, m is chosen to be 64, as it is assumed that a difference in four hue levels (i.e., 256/64 levels) would correspond to approximately the same perceived color [87].

The implicit assumption in the above threshold selection is that the different background objects are of different colors, and one object is distinguishable from the other based on the color information only. Also, each object occupies a substantial amount of spatial area on the canvas. For example, the background objects could be green trees, blue sky and an orange house in Fig. 5.2(a), and all three objects occupy substantial spatial areas on the canvas. The theoretical aspects of the threshold selection process are described below.

Optimal Threshold Selection from the Image Hue Histogram

Equation (5.1) gives the probability of the gray level i in the hue image. For optimal bi-level threshold selection [88], the pixels are divided into two object classes, O_1 and O_2, where O_1 and O_2 contain gray levels $[1, ..., t]$ and $[t+1, t+$

$2, ..., m]$, respectively. Then the probability distribution for the two classes are

$$O_1 : P_1/\omega_1(t), ..., P_t/\omega_1(t) \tag{5.2}$$

and

$$O_2 : P_{t+1}/\omega_2(t), P_{t+2}/\omega_2(t), ..., P_m/\omega_2(t) \tag{5.3}$$

Here $\omega_1(t) = \sum_{i=1}^{t} P_i$, $\omega_2(t) = \sum_{i=t+1}^{m} P_i$ and $P_i = P(\text{hue}_i)$. Also, for classes O_1 and O_2, the means μ_1 and μ_2 are defined as

$$\mu_1 = \sum_{i=1}^{t} \frac{iP_i}{\omega_1(t)} \tag{5.4}$$

and

$$\mu_2 = \sum_{i=t+1}^{m} \frac{iP_i}{\omega_2(t)} \tag{5.5}$$

The mean hue intensity for the entire image is $\mu_T = \frac{\sum_i i \times c(\text{hue}_i)}{T_c}$. So,

$$\omega_1(t)\mu_1 + \omega_2(t)\mu_2 = \mu_T \tag{5.6}$$

and

$$\omega_1(t) + \omega_2(t) = 1 \tag{5.7}$$

Otsu [88] defined the between-class variance of the thresholded image, by using discriminant analysis, as

$$\sigma_B(t)^2 = \omega_1(\mu_1 - \mu_T)^2 + \omega_2(\mu_2 - \mu_T)^2 \tag{5.8}$$

For bi-level threshold selection, the optimal threshold, t^*, has to be chosen so that the between-class variance, $\sigma_B(t)$ is maximized [88]. So,

$$t^* = \arg_{1 \leq t \leq m} \max \, \{\sigma_B^2(t)\} \tag{5.9}$$

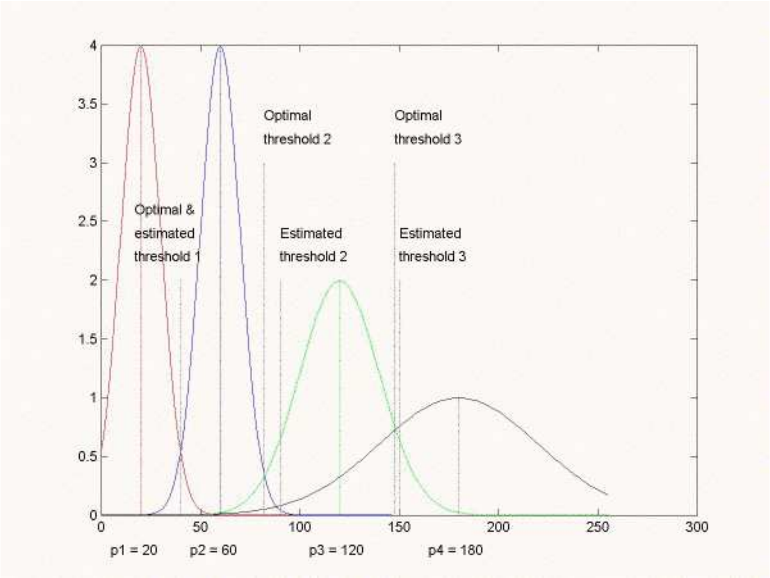


Figure 5.1: An example of an image being modeled as a Gaussian mixture of four hue values. The peaks of the Gaussian curves and shown as $p1$, $p2$, $p3$, and $p4$, respectively. The optimal and estimated thresholds for hue based color segmentation are also indicated.

For multilevel thresholding, the above equation can be extended as follows [89]. Suppose the image is divided into n object classes, $\{O_1, O_2, ..., O_n\}$ with $n-1$ thresholds, $\{t_1, t_2, ..., t_{n-1}\}$. The optimal thresholds, $\{t_1^*, t_2^*, ..., t_{n-1}^*\}$, are chosen by maximizing σ_B^2. For example, Fig. 5.1 shows a model where the hue histogram of an image can be modeled as a Gaussian mixture of four hue values. So, three optimal thresholds can be determined to segment the image. Thus,

$$\{t_1^*, t_2^*, ..., t_{n-1}^*\} = \arg_{1 \leq t_1 < ... < t_{n-1} < m} \max \, \{\sigma_B^2(t_1, t_2, ..., t_{n-1})\} \tag{5.10}$$

where $\sigma_B^2 = \sum_{k-1}^{n} \omega_k(\mu_k - \mu_T)^2$, $\omega_k = \sum_{i \in O_k} P_i$ and $\mu_k = \sum_{i \in O_k} \frac{iP_i}{\omega_k}$.

When it is assumed that the background image can be modeled as a Gaussian mixture of hue values, the optimal thresholds have the local maxima of the histogram in different classes [90]. In this research, I also assume that each of the background objects occupies a substantial region in the image. The method of finding the optimal threshold for this application would be computationally intensive [88] and would involve an exhaustive search. So, based on the above assumptions, the proposed method generates sub-optimal thresholds with lower implementation complexity. These detected thresholds ensure that local maxima fall in different classes. The assumption that the merged object covers a substantial percentage of the background emphasizes the major objects in the background. Thus, the smaller merged objects (if there were any) may not be noticeable. However, depending on the application, the algorithm could be modified for taking into account the smaller objects.

Based on the two assumptions — Gaussian mixture model of hue values and large merging objects relative to the area of main subject — the next task is to identify the thresholds from the histogram such that σ_B^2 is maximized. A level, λ, is determined which acts as a cut-off value for the dominant hues to be considered. For any given histogram, this cut-off value λ can be determined by using the histogram mean. Depending on how stringent the assumption is, i.e., whether the smaller objects are considered or not, λ can be designed as a factor (η) of the mean. Where the constraint is relaxed, η will be high. In this dissertation, η is 1. On the other hand, if the constraint is hard, i.e., smaller objects are to be taken into account, then the value of η needs to be smaller. Once λ is defined, the dominant hues, which have values greater than the cut-off λ, are selected as the valid regions to be segmented. Other regions with hue below the cut-off λ lie within the same region.

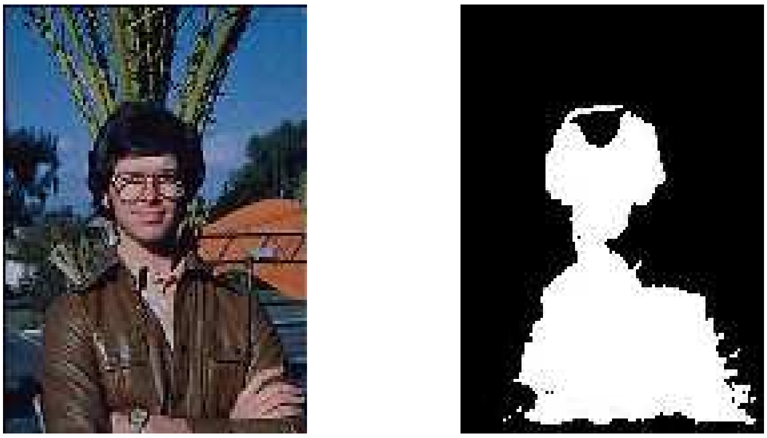
The distributions of the hue values are assumed to be normal. Since the theory of color segmentation of the background does not use an *á priori* model, the exact values of the standard deviation is not known. Also determination of the same is computationally intensive. Here it is assumed that each hue class has a similar distribution; i.e., they have the same standard deviation. This assumption leads to some error with respect to the choices of the optimal thresholds. For simplicity, the threshold is chosen to be the center of two consecutive dominant hues. Fig. 5.1 illustrates the segmentation error that would result from this approximation. Theoretically, an optimal threshold is where the distributions intersect. In simulation, more or less the Gaussian curves for the hue values intersect at a point close to the mid point of the two consecutive peaks. The main motivation of using this assumption is to reduce computational complexity, which otherwise is high for an exhaustive search procedure.

Fig. 5.3 shows the color histogram for the hue values with the average (i.e. $\lambda = \frac{T_c}{m}$ and $\eta = 1$) and the peaks for the background of Fig. 5.2(a). Based on the color histogram and the average value, $n = 10$ background objects are automatically identified for Fig. 5.2(a). Fig. 5.4 shows three of these identified background objects.

5.2.2 Merger Detection

Based on the aforementioned background segmentation, the background image can be modeled as a linear combination of the background objects. Thus,

$$S_b = \sum_{i=1}^{n} O_i \tag{5.11}$$



(a) (b)

Figure 5.2: Examples of (a) a merger of the main subject, the man, with the trees in the background (in color) and (b) the detected main subject mask in (a).

where S_b is the background image and O_i are the identified n background objects. Now, one or more of these background objects may merge with the main subject.

Features for Identifying Merged Background Object

A merged background object has sharp edges and spatially touches the main subject in the 2-dimensional photograph. Thus to identify the merged background object, I choose a measure that takes the above two features into consideration. I choose the background object that has the largest high frequency content and is touching the main subject mask to be the merged object. For each of the segmented background objects, (a) the high frequency content gives a measure of the sharp edges in the object, and (b) the distance of each

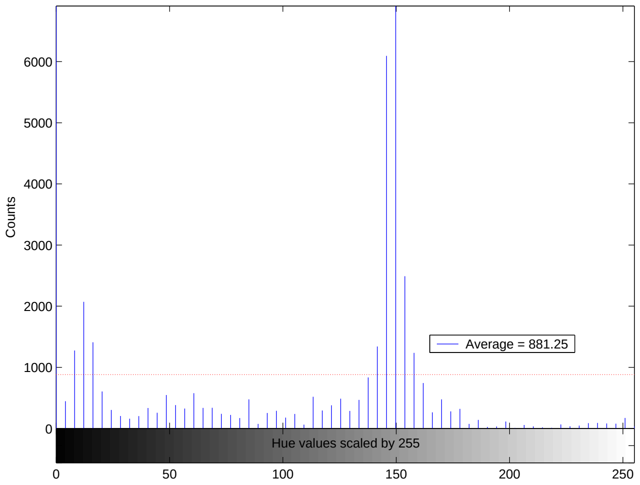


Figure 5.3: Histogram of the hue values for the background of Fig. 5.2(a), which shows the average and peaks.

of the points in the object from the main subject mask is used to determine whether the object is touching the main subject.

Measure for Merged Object Identification

To identify the merged object automatically, each object O_i is transformed into a feature space representation, Ω_i, where $\Omega_i \in \Gamma$. Γ is defined as a weighted sum of the high frequencies contained in the spatial region of each object. High frequency coefficients are obtained from the first level of the two-dimensional Gaussian pyramid [91] of the intensity image. In the Gaussian pyramid representation, the image is represented in several fine to coarse layers, where the next layer is generated by smoothing the previous layer with a symmetric Gaussian kernel and resampling it at one-half the size along each dimension. Gaussian pyramids are localized in space. The Gaussian pyramid could be replaced with a Laplacian pyramid, for an extra implementation cost

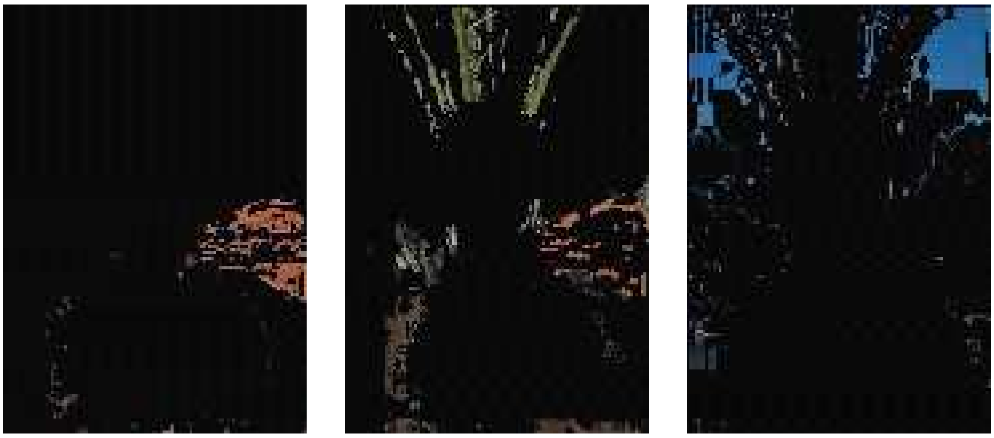

(a) Object 1 (b) Object 2 (c) Object 3

Figure 5.4: Some of the background objects (segmented by color content) for Fig. 5.2(a) identified by the color background segmentation.

of one subtraction per pixel.

The high frequency coefficients are weighted with the inverse of the distance in space from the main subject mask. To compute the inverse distance transform, the distance transform coefficients (see Appendix B) are stored as a grayscale image, and are subtracted from 255 before multiplication with the high frequency coefficients. This assigns more penalty to the higher frequencies closer to the main subject. Fig. 5.7(a) and (b) show the Euclidean distance transform [92, 93] coefficients and high frequency coefficients obtained from the first level of the Gaussian pyramid, respectively. In Section 5.3, I will reduce the implementation complexity of the computation of the inverse distance measure.

The measure is further normalized by the area occupied by each of the background objects, to remove biasses based on the size of the background object.

Frequency Inverse Distance Measure

For each background object, O_i, let ω_i^H be the high frequencies contained in the spatial area occupied by O_i. In this dissertation, I generate ω_i^H from the first level of the two-dimensional Gaussian pyramid [91] of the intensity image. Let d_i be the distance transform coefficient for every pixel of object O_i. I now define the feature space representation, Ω_i, for each background object, O_i, as a function of the frequency component and distance transform coefficient. I normalize the measure by the area of each of the segmented background objects, which is denoted by A_i.

As previously stated, the main motivation of merger detection and blurring is that the prominent background objects near the main subject mask needs to the detected and blurred. Thus, the nature of the function Ω_i should be such that it increases for high frequency and simultaneously decreases with distance. The possible forms of the functions have been tested and their properties have been detailed below. Finally, the choice of the function for this application has been described. The computation can be in fixed-point arithmetic or floating point arithmetic. For the floating-point arithmetic, the distance transform value lies within $[0, 1]$, whereas for fixed-point arithmetic the distance transform has been scaled to lie between $[0, 255]$. First, I discuss the approach by using floating-point arithmetic.

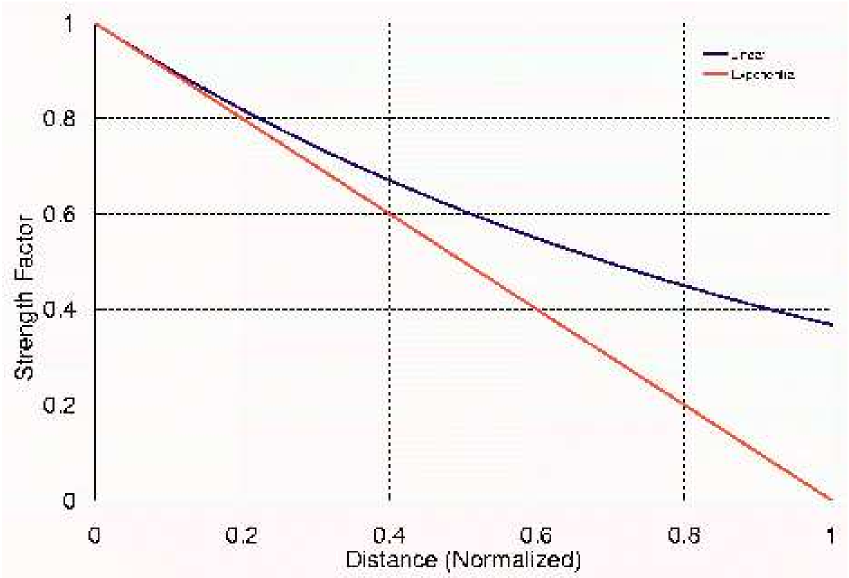


Figure 5.5: Illustration of the characteristics of the weighting factor, $d_i \in [0, 1]$ for the Frequency Inverse Distance Measure. The red and the black curves represent the exponential and linear forms of the distance, d_i, respectively.

The functions for floating-point arithmetic can be as follows:

$$\begin{aligned} \Omega_i &= (1 - d_i)\frac{\omega_i^H}{A_i} && \text{Linear Form} \\ \Omega_i &= \frac{\omega_i^H}{A_i d_i} && \text{Division Form} \\ \Omega_i &= \frac{\omega_i^H e^{-d_i}}{A_i} && \text{Exponential Form} \end{aligned} \tag{5.12}$$

Figure 5.5 shows the behavior of the contribution from the distance transform coefficient, d_i, for the linear and exponential forms. The division form is not a suitable choice in floating-point situation with $d_i \in [0, 1]$, since it has a high value as $d_i \to 0$.

The division form could be rewritten as

$$\Omega_i = \frac{\omega_i^H}{A_i(n + d_i)}$$

where n is an integer value. Although this restricts the upper limit of the division form within 1, the dynamic range gets reduced considerably as the lower limit goes up to 0.5 for $n = 1$. Thus, this option can also be discarded.

Now, from the other two forms, it seems that the linear operator performs better compared to the exponential for the same reason of reduced dynamic range. Also, the exponential form is computationally expensive.

The function characteristics change drastically for fixed-point arithmetic. The expressions for fixed-point arithmetic can be similarly written as

$$\begin{aligned} \Omega_i &= (255 - d_i)\frac{\omega_i^H}{A_i} && \text{Linear Form} \\ \Omega_i &= \frac{\omega_i^H}{A_i d_i} && \text{Division Form} \\ \Omega_i &= \frac{\omega_i^H e^{-d_i}}{A_i} && \text{Exponential Form} \end{aligned} \tag{5.13}$$

Figure 5.6 illustrates the behavior of the contribution from the distance transform coefficient, d_i, for the exponential and the division forms. Here the distance transform coefficient, d_i is scaled to be between $[0, 255]$.

One interesting observation of these curves is the trend of their change with the transform value. Both the exponential and division factor curves fall off rapidly with increase in distance. The exponential value drastically changes from 1 to near 0 within a very short span, while the division factor term initially drops very fast (although slower than exponential) and then falls slowly over

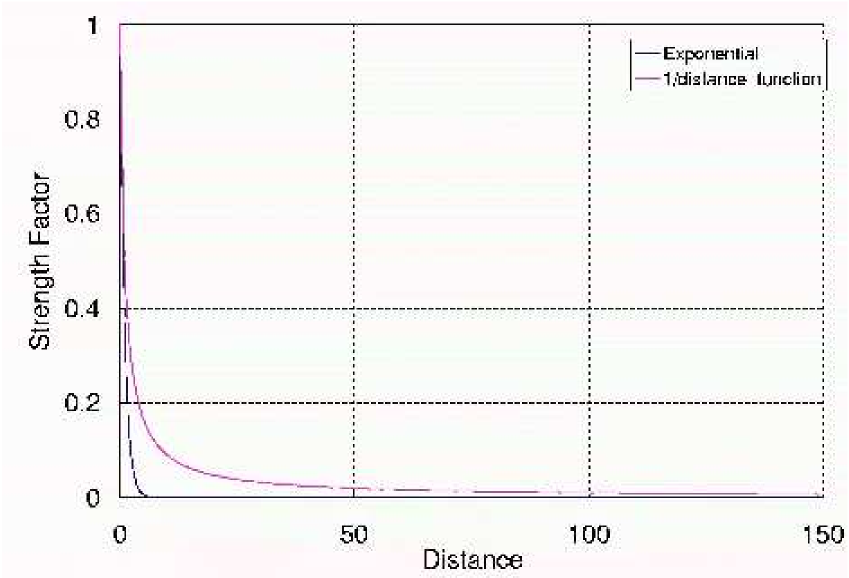


Figure 5.6: Illustration of the characteristics of the weighting factor, $d_i \in [0, 1]$ for the Frequency Inverse Distance Measure. The red and the black curves represent the exponential and division forms of the distance, d_i, respectively.

a quite wider range. This gives a wide range of possibility of exploiting these features according to the applications. For example, the images where the merger is quite small and the cut-off in the color thresholding is selected quite low, the exponential value might be good choice. In those images it is desired the Frequency Inverse Distance Measure should decrease quite fast so that only the small merged objects are detected and blurred, while the others are left untouched. In some other situations that require more blurring at the very near regions and gradual low blurring as distance increases; e.g. to provide some special effect, the division factor is quite appropriate.

However, these functions are not the appropriate choice for the images having considerable mergers, e.g. those in Fig. 5.2(a). The linear form performs perfectly. Moreover, the linear form has the least complexity when compared to the other transforms. Also it can be easily implemented in fixed-point arithmetic, in contrast to the other forms. However, in applications that require the other forms, exponential and division forms of Ω_i could be implemented in fixed-point by using lookup tables or iterative algorithms.

An object O_i is detected to be merged with the main subject if its feature space representation, Ω_i, is greater than a threshold. This threshold could be selected by the user. This chapter presents an unsupervised approach in which the object O_i yielding the maximum value of the feature space representation, Ω_i, is identified to be the merged object. This unsupervised approach detects the object that produces the strongest merger and blurs the produced artifact. For Fig. 5.2(a), the tree object shown in Fig. 5.4(b) produces the maximum of the weighted sum of high frequencies, which identifies that the tree merges with the main subject.

5.2.3 Selective Blurring

The detected merged object, O_i^*, has feature a space representation, Ω_i^*. To reduce the effect of the merger, Ω_i^* needs to be reduced. As Ω_i is the weighted sum of the high frequencies, the high frequency coefficients, ω_i^H, are masked when the image is reconstructed from the Gaussian pyramid representation. From the chosen definition of Ω_i in the linear form of (5.13), the factor $(255-d_i)$ cannot be changed. So, Ω_i^* is reduced by lowering ω_i^{*H}. In Fig. 5.2(a), the high frequency coefficients of the first level of the Gaussian pyramid are masked out by using the approximate shape of the detected tree object. The resulting image is shown in Fig. 5.8. In order to increase the amount of smoothing, masking could be extended to higher levels of the Gaussian pyramid decomposition.

5.3 Implementation Complexity

The proposed algorithm is shown in Fig. 5.9. The original RGB image of dimension $N \times M$ requires $3NM$ grayscale pixels (8 bits per grayscale pixel) of storage without compression. The main subject is detected with 18 multiply-accumulates, 4 comparisons and 6 memory accesses per pixel [3]. The output binary main subject mask requires NM bits.

Background segmentation starts with a conversion from RGB to hue. The hue value calculation uses an intermediate variable, H', which is in the interval $[-255, 1275]$ and can be represented by a 12-bit signed integer. The pseudocode for the conversion follows:

```
min = min(R, G, B);
max = max(R, G, B);
$\delta$ = max - min;
if (R == max) H' = G-B; (within yellow & magenta)
else if (G == max) H' = 2$\delta$+B-R; (within cyan & yellow)
else H' = 4$\delta$+R-G; (within magenta & cyan)
H = (H' + 255) >> 3;
```

In the worst case, the conversion to hue requires 2 shifts, 3 adds, 6 compares, and 4 byte memory accesses per pixel. Computing the histogram and threshold value requires 1 add and 1 compare per pixel. The hue values are stored in NM pixels (or $N \times M \times 8$ bits), and a buffer of $NM \log_2 n$ bits stores the information of the n segmented objects. Now, for many practical applications, the number of segmented objects, n, will be less than 2^8. So,

$$n \leq 2^8 \text{ or } \log_2 n \leq 8 \tag{5.14}$$

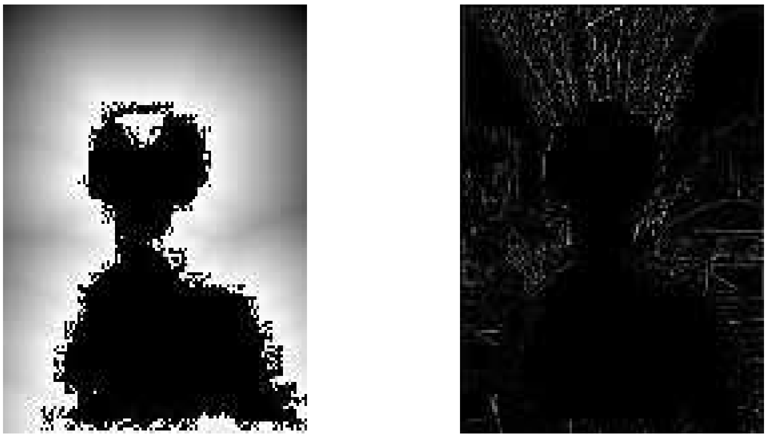
So, the information regarding the segmented objects can be stored in the buffer that originally contained the hue values.

The intensity Gaussian pyramid first converts the color image to an intensity image by either

$$I = (R + G + B)/3 \text{ or } I = (R + 2G + B)/4 \tag{5.15}$$

The former step, which requires 2 adds and 1 multiply, is suitable for programmable digital signal processors. For a hardware implementation, I could use the later, which requires 2 adds, a shift left by one bit (multiplication by 2) and a shift right by two bits (division by 4). Shifts can be used because the RGB values are non-negative. The intensity image is stored in NM pixels. Any level of the Gaussian pyramid can be computed by convolving the grayscale image with a 3 × 3 filter with power-of-two coefficients, which requires 9 shifts, 8 adds and 4 byte memory accesses per pixel. The 9 reads in image values to compute the convolution can be stored in registers in order to reduce the number of memory reads to 3 per pixel. The first level of coefficients are stored in NM pixels, and the intensity image may be overwritten in a sequential implementation of Fig. 5.9.

The inverse distance transform could be determined from the Euclidean distance transform [92, 93] by subtracting its value from 255. In this case, the inverse distance transform would be computationally intensive. I propose an approximate, lower complexity, inverse distance measure. Along each row (column) the distance of each "off" pixel from the nearest "on" one is computed and a ramp function is generated. The maximum of the horizontal (row) distance and the vertical (column) distance is taken as the distance from the nearest "on" pixel. In order to assign more penalty to the high frequency



(a) (b)

Figure 5.7: (a) The Euclidean distance transform coefficients and (b) the high frequency coefficients from the first level of the Gaussian pyramid for Fig. 5.2(a). The background object is detected to be merged if it yields the maximum of the weighted sum of (a) and (b).

coefficients close to the main subject, the pixels closer to the main subject mask have a higher weight. The weights are stored in NM pixels. The distance measure requires 2 adds, 1 compare, and 2 byte memory accesses per pixel.

Both the approximated and the Euclidean distance measures give similar weights to the objects oriented in the horizontal and vertical directions. However, for objects that are oriented diagonally, this approximation would assign less weights to them, compared to the Euclidean distance measure. So, if the merging objects are located horizontally or vertically then the accuracy in identifying the merged object would be the same, for both the approximated and the Euclidean inverse distance measures. However, if a merging object is diagonally aligned the weighting and hence the accuracy of merger detection would be better with the Euclidean distance measure.

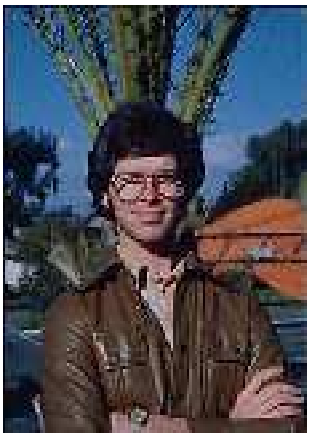


Figure 5.8:]
The detected merged region is processed in the frequency domain to reduce the effect of the merger. The blurred trees induce a sense of distance.

For each background object, the intensity Gaussian pyramid coefficients are weighted by the inverse distance transform coefficients and summed. The background object with the highest sum is chosen as the background merging object, and the corresponding background object mask is output. The background object mask can be stored in the main subject mask buffer so as to reuse memory. All totaled, 1 multiply, 1 add, and 1 compare are required per pixel.

In the final step, the color Gaussian pyramid and reconstruction only have to be applied to those pixels in the binary mask input. For each pixel in the binary mask input, the first level of the color Gaussian pyramid transformation is calculated separably for each RGB planes. For each color plane, 9 shifts, 8 adds, and 3 byte memory accesses are required for a 3×3 filter kernel. The high frequency coefficients for the merging background object are masked with 1 compare and 1 memory access per pixel. The output (merger reduced)

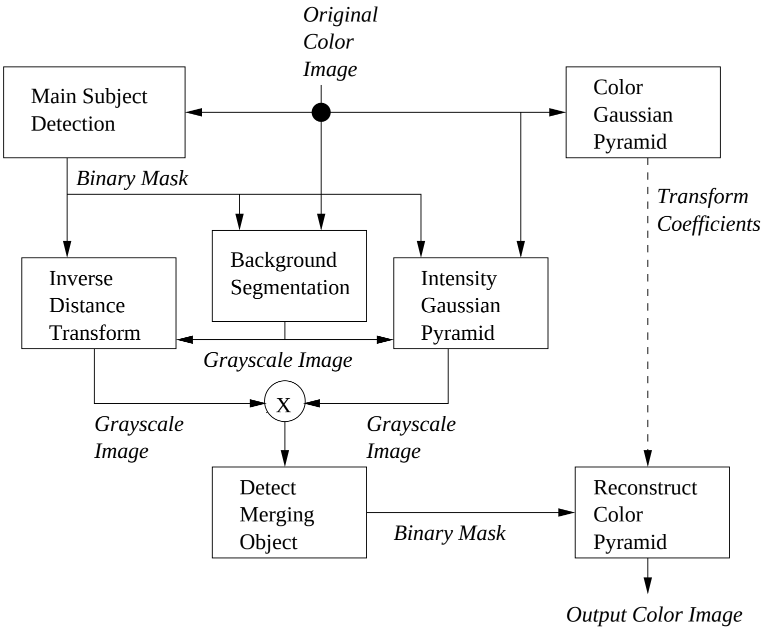


Figure 5.9: Proposed merger reduction algorithm for an original $N \times M$ color image. Storage is $3NM$ grayscale pixels (bytes) for the original, NM bits for a mask, and $3NM$ grayscale pixels (bytes) for the output (merger reduced) image. For a parallel implementation of the subsystems, an additional storage of $2NM$ pixels (bytes) is needed.

image takes 9 shifts, 8 adds, 1 compare, and 1 byte memory access per pixel, and would be stored in $3NM$ pixels.

The merger detection and mitigation algorithm is explained in Fig. 5.9. The computational requirements for the blocks in Fig. 5.9 are given in Table 5.1. All the blocks, except for the main subject detection and color Gaussian pyramid/reconstruction, work only on the background image. Hence, the complexity will depend on the number of background pixels in the image.

Block	$\times$	$<<$	$+$	$\geq$	m
Segment background		2	4	7	4
Intensity Gaussian pyramid	1	9	10		4
Inverse distance transform			2	1	2
Detect merging object	1		1	1	1
Color Gaussian pyramid		27	24		9
Reconstruct pyramid	1	27	24		3
Total	**3**	**65**	**65**	**9**	**23**

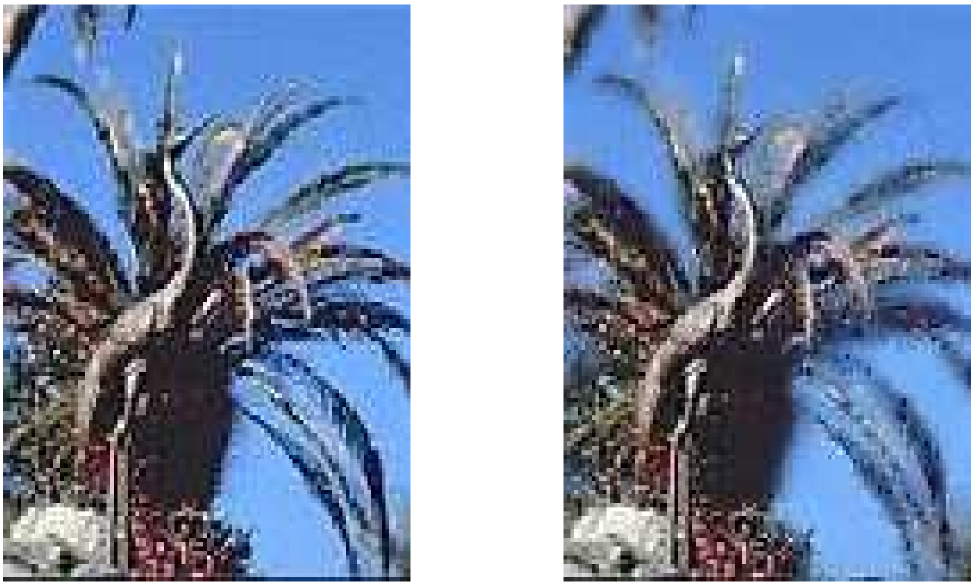
Table 5.1: Per pixel implementation complexity of the proposed algorithm in number of multiplications ($\times$), shifts ($<<$), additions ($+$), comparisons ($\geq$), and byte memory accesses (m). The last two steps are only applied to the merging background object. The other steps are applied only to the background.

5.4 Results

The proposed algorithm was tested on several pictures. The merger reduced image for Fig. 5.10(a) is shown in Fig. 5.10(b). The background trees merging with the bird are blurred out, thereby inducing a sense of distance.

The pictures were downloaded from the World Wide Web and comprise mainly of outdoor pictures. It was possible to segment the background of these pictures with color segmentation. However, for pictures with more complicated background, color and texture segmentation could be combined to detect the background objects.

Also, the assumption that the background objects occupied large areas on the canvas was met in the test pictures. However, the proposed algorithm could be modified to identify smaller merging objects.



(a) Original image (b) Merger reduced

Figure 5.10: The proposed algorithm reduces the effect of the merger of the tree with the bird. The blurred trees in the processed image are distinguishable as a separate object from the main subject.

5.5 Conclusion

This chapter presents an unsupervised algorithm for automatic merger detection and mitigation when taking photographs in digital still cameras. The performance of the color based segmentation will be limited for highly textured backgrounds, which may require texture segmentation instead. Alternately, merger detection could be used to warn the user of a possible merger.

With the framework presented in this chapter, two other photographic composition rules that may be automated are

- *Using the "best" camera angle:* For example, when photographing active people it is appealing to have an uncluttered background. The feedback from the background segmentation discussed in this chapter could be used to prompt the amateur photographer to acquire an uncluttered background by using a low–angle.

- *Taking a picture that is well balanced on the eye:* It is better to have both the wheels of a cart or all the legs of a table in the picture to create a sense of balance in the picture. If the background segmentation determines that the user is acquiring only parts of such objects, the camera zoom could be adjusted to capture a wider area of the scene.

Chapter 6

Conclusions

"When you aim for perfection, you discover it's a moving target."

Geoffrey F. Fisher

6.1 Conclusions

This dissertation proposes a framework for helping the amateur photographers take pictures with better photographic composition. The framework, which is shown in Fig. 6.1, acquires the image the user intended as well as a second image. This second image is taken immediately before or after the user-intended picture and uses the same autofocus settings. In the second image, however, the shutter aperture is fully opened and shutter speed is automatically adjusted so that objects not in the plane of focus are blurred by the optical subsystem. This second blurry image is then digital processed to locate the main subject. With the main subject identified, selected photographic composition rules may be automated to generate new alternate pictures with better photographic composition. Three photographic compositions rules are automated in a way that does not make assumptions on scene setting or content.

This dissertation also moves towards the goal of implementing the framework in a digital still camera. A digital still camera implements a variety of digital image processing algorithms on a fixed-point programmable processor with little on-chip memory and relatively slow clock speeds and off-chip data transfers. I present low-complexity, non-iterative algorithms for automatic main subject detection and for automating three photographic composition rules: rule-of-thirds, artistic background blurring, and blurring merging background objects. The algorithms are amenable to implementation in fixed-point arithmetic.

Chapter 2 summarizes previous research in main subject detection. The previous approaches either require *à priori* training or have higher implementational complexity. These approaches could be appropriate for offline applications, such as image indexing for content-based retrieval [20, 21, 22], object-based image compression for image servers [36], and for content grouping for auto-album layout [30, 31]. I assert that reliable main subject detection can be performed in a way that does not require *á priori* training and is not restricted to off-line computation by guiding the image acquisition process and offloading most of the computation to the optical subsystem. My proposed approach makes it possible to provide in-camera feedback to the amateur photographer for taking pictures with better photographic composition.

Chapter 3 proposes a gradient induced algorithm for in-camera detection of the main subject in a picture. The main subject segmentation algorithm has been implemented and tested on various images. It is proposed to take a supplementary picture that has the main subject in focus and blurs out the objects not in the plane of focus, by leveraging the camera's optical subsys-

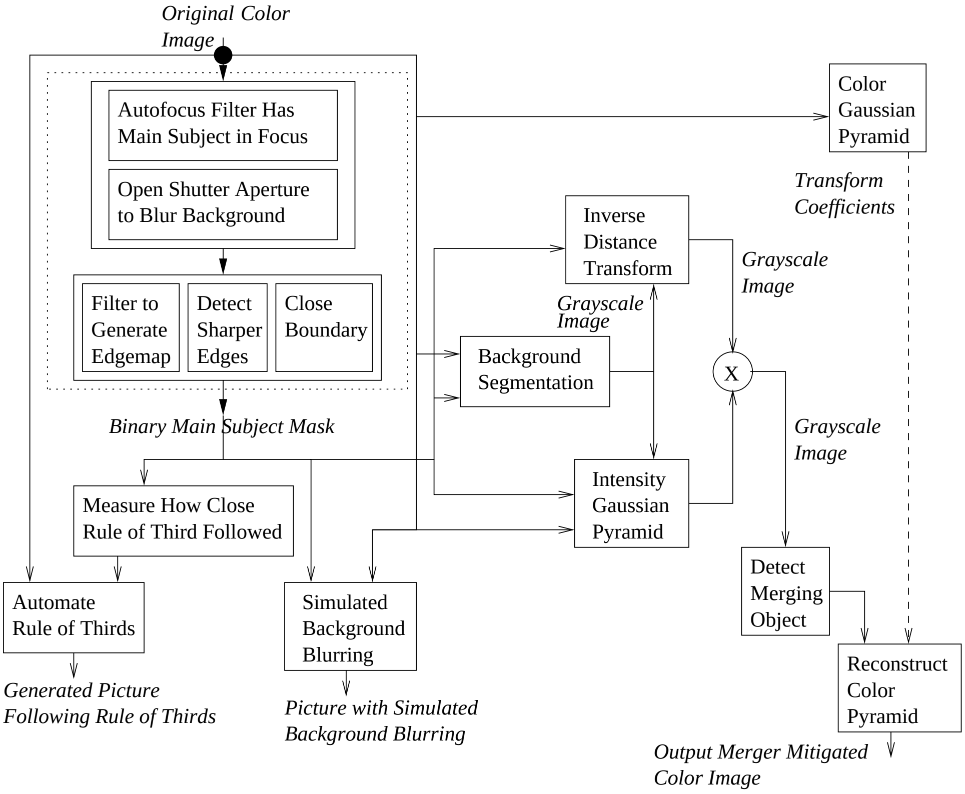

Original Color Image
Autofocus Filter Has Main Subject in Focus
Open Shutter Aperture to Blur Background
Filter to Generate Edgemap
Detect Sharper Edges
Close Boundary
Binary Main Subject Mask
Measure How Close Rule of Third Followed
Automate Rule of Thirds
Generated Picture Following Rule of Thirds
Simulated Background Blurring
Picture with Simulated Background Blurring
Color Gaussian Pyramid
Transform Coefficients
Inverse Distance Transform
Grayscale Image
Grayscale Image
Background Segmentation
X
Grayscale Image
Intensity Gaussian Pyramid
Detect Merging Object
Reconstruct Color Pyramid
Output Merger Mitigated Color Image

Figure 6.1: Proposed automation of selected photograph composition rules for digital still cameras.

tem. The resulting frequency content information difference between the main subject and the image background is used as a cue for segmentation. The proposed algorithm first prefilters the supplementary picture to isolate the strong edges around the main subject, then detects the strong edges with an edge detector, and finally generates the main subject mask by closing the contour with gradient vector flow based active contour algorithm. The implementational complexity of the proposed algorithm is similar that to a 5×5 filter. The previous wavelet based approach [37, 38] to detect the main subject would at least be $2 \times n \times k$ times more complex than the proposed method, where

n is the number of wavelet levels computed and k is the number of initial clusters. A previous block-based iterative approach [40] would be at least B times more complex, where the image is divided into $B \times B$ non-overlapping blocks. The proposed algorithm could be extended to segment more than one main subjects in the picture. The process would involve taking at least as many supplementary pictures as the number of main subjects.

Based on the generation of the main subject mask in Chapter 4 automates two photographic composition rules, that could be applied based on the main subject information only. The selected rules are (1) placing the main subject by following the *rule-of-thirds* and (2) *simulating background blurs* if the main subject or the camera is in motion or to reduce the depth of field of the picture. For the rule-of-thirds imaginary lines are drawn on the canvas, dividing the image into three equal parts horizontally and vertically. It is suggested [2] to have the center of mass at a point where these imaginary lines intersect. This research proposes an algorithm that automatically identifies how far apart the center of mass of the main subject is from the four one-third corners. The picture is then shifted so that the center of mass falls at the nearest desired corner. The proposed algorithm can be extended to automate the *rule-of-triangles* for placing multiple main subjects in the picture. For simulating background blurs, region-of-interest based filtering is performed on the image background, while the main subject is isolated with the generated mask. Possible linear, radial and zoom blurs can be simulated.

Chapter 5 proposes an algorithm for merger reduction in photographs. A merger occurs in a picture when equally focused foreground and background regions in the image tend to merge as one object. To avoid such mergers, professional photographers either change the camera angle, or blur the background (with a larger shutter aperture) to induce a sense of distance in the picture. This dissertation proposes an algorithm to automatically detect a merger, and blur the merging object. The possible future extensions of the proposed algorithm are to (1) detect and blur more than one background object in the picture, (2) experiment with different background segmentation methods for color and texture segmentation, and (3) substitute colors in the spatial domain from the non-merging background objects to the merging ones.

The implementation complexity of all the proposed algorithms are low enough to be implemented in real-time in fixed-point digital signal processors. In a nutshell, the research carried out in the scope of this dissertation proposes a theory and algorithms to

- Segment the main subject from a supplementary picture that has the main subject in focus and objects that are not in the plane of focus are blurred by using camera optics
- Reposition the main subject in the image frame for better context
- Blur the background based on image content
- Detect unwanted mergers in the photograph and reduce the visually unpleasant effect

Thus, this dissertation shows that it is possible to provide feedback to the amateur photographer to take pictures following photographic composition rules.

6.2 Future Work

The proposed framework for providing online feedback to the photographer presented in this dissertation, opens up a number of avenues for future research. Based on the presented algorithms and inclusion of other cues, the feedback system can be extended for video and a wide variety of applications. A brief discussion about possible directions for future research follows.

6.2.1 Lower Complexity Image Registration

Currently the proposed framework registers the supplementary picture to the user-intended picture by using the difference in the histogram of the two pictures. The complexity of this part is lower at the cost of accuracy. The complexity could be further reduced and the accuracy increased be developing a system where the two images are registered based on the main subject mask. The pixel values of the supplementary image lying within the main subject mask could be searched in a small search window in the user-intended picture. This way the accuracy could be improved.

6.2.2 Automation of Other Photographic Composition Rules

With the framework proposed in this dissertation a few other photographic composition rules could be automated. They are

- *Automation of the best possible zoom:* After detecting the main subject mask, the amount of zoom could be determined.

- *Taking a picture through frames available in the scene:* At first, the available frames in the scene need to be detected. The main subject could then be repositioned to fall within one of the detected frames.
- *Placing the main subject where lines of the scene intersect:* Here, the available lines in the scene need to be detected. For an added appeal, the main subject could be relocated to where these lines of interest intersect. This way the viewers' eye is automatically drawn to the main subject.
- *Using the "best" camera angle:* For example, when photographing active people it is appealing to have an uncluttered background. The feedback from the background segmentation discussed in Chapter 5 could be used to prompt the amateur photographer to acquire an uncluttered background by using a low–angle.
- *Taking a picture that is well balanced on the eye:* To create a sense of balance in the picture, it is better to have both the wheels of a cart or all the legs of a table in the picture. Background segmentation could determine if the user is acquiring only parts of such objects. In such cases, the camera zoom could be adjusted to capture a wider area of the scene.

6.2.3 Extension for Video Acquisition Applications

The algorithms presented in this dissertation could be directly extended for video acquisition. The added variable in the extension would be time. So, the algorithms could be applied either on a frame-to-frame basis, or in the motion compensated domain [94]. In a frame-to-frame basis implementation, the resulting video acquisition rate would be around 50% slower than the normal acquisition rate. For a faster implementation, the algorithms need to be applied in the motion compensated domain. For example, the main subject is detected from the first frame of video sequence. Later the motion vectors pertaining to the main subject, need to be modified depending on the possible relocation of the main subject by using the proposed algorithms.

6.2.4 Developing Algorithms for Image Stabilization

The proposed research could also be extended for a software-based image stabilization system that is capable of stabilizing acquired video with substantial displacements between frames [95, 96, 97, 98]. I suggest to detect the main subject from the first frame of the video sequence. The mask of the main subject could then be used as a feature that could detect possible camera displacements. Once the camera displacements are computed, the subsequent frames could be corrected for a more stable video sequence.

6.2.5 Inclusion of Other Cues in the Proposed Framework for Better Image/Video Acquisition

The main focus of this dissertation was to improve image acquisition for photography. The proposed framework could also be extended to image acquisition in other domain such as medical image acquisition [99, 100, 101], aerial photography [102, 103] or non-destructive testing [104]. However, in each of these domains, other cues need to be used in the proposed framework. For example, in medical image acquisition, pressure or velocity of blood flow at a region under investigation could be used to direct more appropriate image acquisition in image-guided operations for on-site or telemedicine applications [101, 105, 106]. For aerial photography, a pressure or temperature difference could guide camera angle during acquisition [102, 103]. Similarly a difference in elasticity could guide camera orientations, for acquiring images to detect leather or material defects [104].

On a closing note, this dissertation proposes and implements algorithms to help amateur photographer take pictures that follow photographic composition rules. This increases the aesthetic appeal of the acquired pictures. The presented research could be extended to other disciplines of image acquisition.

Appendix A

Auto-Focus Filter

An auto-focus filter [61, 62] automatically adjusts the optical settings in the camera to have the image in focus. The commonly used criterion to have a focused image is a measure of the sharpness of the acquired image. More high frequency components in an image make it sharper.

The lens of a camera system operates as a lowpass filter. So, the amount of blur in the acquired image depends on the lens settings. This blurring is modeled as convolution of the original image with a point spread function. The auto-focus filter is then designed to reduce the amount of the generated blur.

Auto-focus systems usually computes a feature of the filtered image and uses it as a criterion to design the auto-focus filter. Possible choices of this feature includes energy, gradient energy and Laplacian energy of the acquired image.

Appendix B

Distance Transform

The distance transform (DT) operator maps a gray level image from a binary image [92, 93]. In the binary image, a pixel is either a feature or background if its value is one or zero, respectively. The generated gray level image provides a measure of the distance of each background pixel from the nearest nonzero feature pixel in the binary image. This operator maps the distance based on four metrics: Euclidean, cityblock (4-connected), chessboard (8-connected), and quasi-Euclidean distances. The distance transform is closely related to the Voronoi diagram. In the construction of Voronoi diagrams, each pixel is assigned an identity of the nearest feature pixel.

Bibliography

[1] The Great Idea Finder, "History – Invention Facts and Myth: Fax Machine." http://www.ideafinder.com/history/inventions/story051.htm.

[2] Kodak, *How to Take Good Pictures: A Photo Guide by Kodak*. Ballantine, Sept. 1995.

[3] S. Banerjee and B. L. Evans, "A Novel Gradient Induced Main Subject Segmentation Algorithm for Digital Still Cameras," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2003.

[4] S. Banerjee and B. L. Evans, "Unsupervised Automation of Photographic Composition Rules in Digital Still Cameras," in *Proc. SPIE Conf. on Sensors, Color, Cameras, and Systems for Digital Photography VI*, Jan. 2004.

[5] S. Banerjee and B. L. Evans, "Unsupervised Merger Detection and Mitigation in Still Images Using Frequency and Color Content Analysis," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, May 2004.

[6] S. Banerjee and B. L. Evans, "Low-Complexity Unsupervised Automa-

tion of Photographic Composition Rules for In-Camera Image Acquisition," *IEEE Trans. on Image Processing*, submitted March 15, 2004.

[7] S. Daly, *Digital Images and Human Vision*, ch. The Visible Differences Predictor: An algorithm for the assessment of image fidelity, pp. 179–206. MIT Press, 1993.

[8] Z. Wang and A. C. Bovik, "A Universal Image Quality Index," *IEEE Signal Proc. Letters*, vol. 9, pp. 81–84, Mar. 2002.

[9] Z. Wang, A. C. Bovik, and L. Lu, "Why is Image Quality Assessment so Difficult?," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 4, pp. 3313–3316, May 2002.

[10] X. Li, "Blind Image Quality Assessment," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 1, pp. 449–452, Sept. 2002.

[11] Z. Liu and L. J. Karam, "Context Formation by Mutual Information Maximization," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 89–92, Sept. 2002.

[12] P. M. F. Dufaux, S. Winkler, and T. Ebrahimi, "A No-Reference Perceptual Blur Metric," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 57–60, Sept. 2002.

[13] D. S. Turaga, Y. Chen, and J. Caviedes, "No Reference PSNR Estimation for Compressed Pictures," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 61–64, Sept. 2002.

[14] P. G. Engeldrum, "Psychometric Scaling: Avoiding the pitfalls and hazards," in *Proc. IS&T's Conf. on Image Proc., Quality, Capture Sys.*, vol. 1, pp. 101–107, Apr. 2001.

[15] Z. Wang, L. Liu, and A. C. Bovik, "Video Quality Assessment Using Structural Distortion Measurement," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 65–68, Sept. 2002.

[16] M. Masry and S. S. Hemani, "Perceived Quality Metrics for Low Bit Rate Compressed Video," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 49–52, Sept. 2002.

[17] M. S. Moore, S. K. Mitra, and J. M. Foley, "Defect Visibility and Content Importance Implications for the Design of an Objective Video Fidelity Metric," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 45–48, Sept. 2002.

[18] J. Luo, A. Singhal, G. Braun, R. T. Gray, O. Seignol, and N. Touchard, "Displaying Images on Mobile Devices: Capabilities, issues, and solutions," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 1, pp. 13–16, Sept. 2002.

[19] U. M. Erdem and S. Sclaroff, "Automatic Detection of Relevant Head Gestures in American Sign Language Communication," in *Proc. IEEE Int. Conf. on Pattern Recognition*, vol. 1, Aug. 2002.

[20] J. Z. Wang, G. Weiderhold, O. Firschein, and X. W. Sha, "Content-based Image Indexing and Searching Using Daubechies' Wavelets," *Int.*

Journal of Digital Libraries, vol. 1, no. 4, pp. 311–328, 1998, Springer-Verlag.

[21] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and Effective Querying by Image Content," *Journal of Intelligent Information Systems*, vol. 3, pp. 231–262, 1994.

[22] A. Gupta and R. Jain, "Visual Information Retrieval," *Communication of the ACM*, vol. 40, pp. 69–79, 1997.

[23] G. P. Corey, M. J. Clayton, and K. N. Cupery, "Scene Dependance of Image Quality," *Photographic Science and Eng.*, vol. 27, pp. 9–13, 1983.

[24] I. Biederman, "Recognition by components: A theory of human image understanding," *Psychological Review*, vol. 94, no. 2, pp. 115–147, 1987.

[25] A. E. Savakis, S. P. Etz, and A. C. Loui, "Evaluation of Image Appeal in Consumer Photography," in *Proc. SPIE Conf. on Human Vision and Elec. Imag.*, pp. 111–120, Jan. 2000.

[26] J. Luo, A. E. Savakis, S. P. Etz, and A. Singhal, "On the Application of Bayes Networks to Semantic Understanding of Consumer Photographs," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 512–515, Sept. 2000.

[27] N. Serrano, A. Savakis, and J. Luo, "A Computationally Efficient Approach to Indoor/Outdoor Scene Classification," in *Proc. IEEE Int. Conf. on Pattern Recognition*, Aug. 2002.

[28] J. Luo and A. Savakis, "Indoor vs Outdoor Classification of Consumer Photographs using Low-Level and Semantic Features," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 2, pp. 745–748, Oct. 2001.

[29] J. Luo and S. P. Etz, "A Physical Model-Based Approach to Detecting Sky in Photographic Images," *IEEE Trans. on Image Proc.*, vol. 11, pp. 201–212, Mar. 2002.

[30] J. Luo, S. P. Etz, R. T. Gray, and A. Singhal, "Normalized Kemeny and Snell distance: A Novel Metric for Quantitative Evaluation of Rank-Order Similarity of Images," *IEEE Trans. on Pattern Anal. and Machine Intelligence*, vol. 24, pp. 1147–1151, Aug. 2002.

[31] S. P. Etz, J. Luo, R. T. Gray, and A. Singhal, "Quantitative Evaluation of Rank-Order Similarity of Images," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 1, pp. 485–488, Sept. 2000.

[32] A. Savakis, "Document Image Thresholding Using Foreground and Background Clustering." US Patent Filed, No. 6044179, Nov. 1997.

[33] S. P. Etz and J. Luo, "Ground Truth for Training and Evaluation of Automatic Main Subject Detection," in *Proc. SPIE Conf. on Human Vision and Electronic Imaging*, vol. 3959, pp. 434–442, Jan. 2000.

[34] J. Luo, S. P. Etz, A. Singhal, and R. T. Gray, "Performance-Scalable Computational Approach to Main Subject Detection in Photographs," in *Proc. SPIE Conf. on Human Vision and Electronic Imaging*, vol. 4299, pp. 494–505, Jan. 2001.

[35] J. Luo and C. Guo, "Non-Purposive Perceptual Region Grouping," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 2, pp. 749–752, Sept. 2002.

[36] Z. Wang, S. Banerjee, B. L. Evans, and A. C. Bovik, "Generalized

Bitplane-by-Bitplane Shift Method for JPEG2000 ROI Coding," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 81–84, Sept. 2002.

[37] J. Li, J. Z. Wang, R. M. Gray, and G. Wiederhold, "Multiresolution Object-of-Interest Detection for Images with Low Depth of Field," in *Proc. IEEE Int. Conf. on Image Analysis and Processing*, pp. 32–37, Sept. 1999.

[38] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold, "Unsupervised Multiresolution Segmentation for Images with Low Depth of Field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 85–90, Jan. 2001.

[39] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "A Efficient k-Means Clustering Algorithm: Analysis and Implementation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, July 2002.

[40] C. S. Won, K. Pyan, and R. M. Gray, "Automatic Object Segmentation in Images with Low Depth of Field," in *Proc. IEEE Int. Conf. on Image Proc.*, pp. 805–808, Sept. 2002.

[41] C. S. Won, "A Block-Based MAP Segmentation for Image Compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, pp. 592–601, Sept. 1998.

[42] L. Vincent and P. Soille, "Watersheds in Digital Space: An Efficient Algorithm Based on Imersion Simulations," *IEEE Trans. on Pattern Matching and Machine Intelligence*, vol. 13, pp. 583–598, June 1991.

[43] G. J. van Tonder and Y. Ejima, "Learning from Nature: Image Segmentation Based on Local Symmetry Detection," in *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*, vol. 6, pp. 804–809, Oct. 1999.

[44] M. Heiler and C. Schnorr, "Natural Image Statistics for Natural Image Segmentation," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1259–1266, Oct. 2003.

[45] C. Fowlkes, D. Martin, and J. Malik, "Learning Affinity Functions for Image Segmentation: Combining Patch-based and Gradient-based Approaches," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 54–61, June 2003.

[46] S. Novianto, L. Guimaraes, Y. Suzuki, J. Maeda, and V. V. Anh, "Multiwindowed Approach to the Optimum Estimation of the Local Fractal Dimension for Natural Image Segmentation," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 222–226, Oct. 1999.

[47] S. K. Warfield, K. H. Zou, and W. M. Wells, "Validation of Image Segmentation and Expert Quality with an Expectation- Maximization Algorithm," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 298–306, Sept. 2002, Springer-Verlag.

[48] T. McInerney and D. Terzopoulos, "Medical Image Segmentation Using Topologically Adaptable Snakes," in *Proc. Conf. on Computer Vision, Virtual Reality and Robotics in Medicine*, pp. 92–101, Apr. 1995.

[49] J. Maeda, S. Novianto, S. Saga, Y. Suzuki, and V. V. Anh, "Rough and Accurate Segmentation of Natural Images using Fuzzy Region-growing

Algorithm," in *Proc. IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 227–231, Oct. 1999.

[50] S. Wolfson and M. Landy, "Examining Edge- and Region-based Texture Analysis Mechanisms," *Vision Research*, vol. 38, no. 3, pp. 439–446, 1998.

[51] A. F. Limas-Serafim, "Natural Images Segmentation for Patterns Recognition Using Edges Pyramids and its Application to the Leather Defects," in *Proc. IEEE Int. Conf. on Industrial Electronics*, vol. 3, pp. 1357–1360, Nov. 1993.

[52] A. F. Limas-Serafim, "Segmentation of Natural Images Based on Multiresolution Pyramids Linking of the Parameters of an Autoregressive Rotation Invariant Model. Application to Leather Defects Detection," in *Proc. IEEE Int. Conf. on Pattern Recognition: Image, Speech and Signal Analysis*, pp. 41–44, Aug. 1992.

[53] A. M. Baumberg and D. C. Hogg, "An Efficient Method for Contour Tracking Using Active Shape Models," in *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 194–199, Nov. 1994.

[54] C. Davatzikos, T. Xiadong, and D. Shen, "Hierarchical Active Shape Models, Using the Wavelet Transform," *IEEE Trans. on Medical Imaging*, vol. 22, pp. 414–423, Mar. 2003.

[55] B. van Ginneken, A. F. Frangi, J. J. Stall, B. M. ter Haar Romeny, and M. A. Viergever, "Active Shape Model Segmentation with Optimal

Features," *IEEE Trans. on Medical Imaging*, vol. 21, pp. 924–933, Aug. 2002.

[56] M. M. Dickens, S. S. Gleason, and H. Sari-Sarraf, "Volumetric Segmentation via 3D Active Shape Models," in *Proc. IEEE Southwest Symp. on Image Analysis and Interpretation*, pp. 248–252, Apr. 2002.

[57] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models," *IEEE Trans. on Pattern Matching and Machine Intelligence*, vol. 23, pp. 681–685, June 2001.

[58] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Advances in Active Appearance Models," in *Proc. IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 137–142, Sept. 1999.

[59] T. F. Cootes and C. J. Taylor, "Constrained Active Appearance Models," in *Proc. IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 748–754, July 2001.

[60] F. Dornaila and J. Ahlberg, "Face Model Adaptation Using Robust Matching and Active Appearance Models," in *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 3–7, Dec. 2002.

[61] N. N. K. Chern, P. A. Neow, and J. M. H. Ang, "Practical Issues in Pixel-Based Autofocusing for Machine Vision," in *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 3, pp. 2791–2796, May 2001.

[62] C. H. Park, J. H. Paik, Y. H. You, H. K. Song, and Y. S. Cho, "Auto Focus Filter Design and Implementation Using Correlation between Fil-

ter and Auto Focus Criterion," in *Proc. IEEE Int. Conf. on Consumer Electronics*, pp. 250–251, June 2000.

[63] A. Adams, *The Camera.* Bulfinch, June 1995.

[64] L. Stroebel, R. D. Zakia, I. Current, and J. Compton, *Basic Photographic Materials and Processes.* Focal Press, Mar. 2000.

[65] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, Nov. 1986.

[66] D. Marr and E. Hildreth, "Theory of Edge Detection," in *Proc. Royal Society of London*, pp. 187–217, 1980.

[67] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (2nd Edition).* Addison-Wesley Pub. Co., Jan. 2002.

[68] D. M. Tsai and H. J. Wang, "Segmenting Focused Objects in Complex Visual Images," *Elsevier Patter Recognition Letters*, vol. 19, pp. 929–940, Aug. 1998.

[69] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int. Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[70] L. Cohen and I. Cohen, "Finite Element Methods for Active Contour Models and Balloons for 2D and 3D Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1131–1147, Nov. 1993.

[71] M. Berger, "Snake Growing," in *Proc. European Conf. on Computer Vision*, pp. 570–572, Apr. 1990.

[72] W. Neuenschwander, P. Fua, L. Iverson, G. Szekely, and O. Kubler, “Ziploc Snakes,” *Int. Journal of Computer Vision*, vol. 25, pp. 191–201, Dec. 1997.

[73] S. R. Gun and M. S. Nixon, “A Dual Active Contour for Improved Snake Performance,” *Research Journal on Image, Speech and Intelligent Systems*, vol. 6, 1995.

[74] P. C. Yuen, Y. Y. Wong, and C. S. Tong, “Contour Detection Using Enhanced Snake Algorithm,” *Electronic Letters*, vol. 32, pp. 202–204, Feb. 1999.

[75] A. Klein, T. K. Egglin, J. S. Pollak, F. Lee, and A. Amini, “Identifying Vascular Features with Orientation Specific Filters and B-spline Snakes,” in *Proc. Computers in Cardiology*, pp. 113–116, 1994.

[76] J. Wang and F. Cohen, “Part II: 3-D Object Recognition and Shape Estimation from Image Contours using B-splines, Shape-invariant Matching, and Neural Networks,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 13–23, Jan. 1994.

[77] J. Irvins and J. Porill, “Statistical Snakes: Active Region Models,” in *Proc. British Conf. on Machine Vision*, vol. 2, pp. 377–386, 1994.

[78] T. McInerney and D. Terzopoulos, “Topologically Adaptable Snakes,” in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 840–845, June 1995.

[79] M. Hoch and P. Litwinowicz, “A Semi-Automatic System for Edge Tracking with Snakes,” *Visual Computation*, vol. 12, pp. 75–83, 1996.

[80] C. Xu and J. L. Prince, “Snakes, Shapes, and Gradient Vector Flow,” *IEEE Trans. on Image Processing*, vol. 7, pp. 359–369, Mar. 1998.

[81] C. Xu, J. A. Yezzi, and J. L. Prince, “A Summary of Geometric Level-Set Analogues for a General Class of Parametric Active Contour and Surface Models,” in *Proc. IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pp. 104–111, July 2001.

[82] J. L. Moigne, X. Wei, P. Chalermwat, T. El-Ghazawi, M. Mareboyana, N. Netanyahu, J. C. Tilton, W. J. Campbell, and R. P. Cromp, “First Evaluation of Automatic Registration Methods,” in *Proc. IEEE Int. Symposium on Geoscience and Remote Sensing*, vol. 1, pp. 315–317, July 1998.

[83] B. Zitova and J. Flusser, “Image Registration Methods: A Survey,” *Elsevier Image and Vision Computing*, vol. 21, pp. 977–1000, June 2003.

[84] The Image, Inc., “Digital Photography – Image Work UP – Print vs. WEB.” http://www.theimage.com/photography/photopg15.htm, 1997.

[85] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, “Iterative Methods for Image Deblurring,” *Proc. of the IEEE*, vol. 78, no. 5, pp. 856–883, 1990.

[86] J. A. C. Yule and G. G. Field, *Principles of Color Reproduction.* GATF-Press, Jan. 2001.

[87] C. Zhang and P. Wang, “A New Method of Color Image Segmentation Based on Intensity and Hue Clustering,” in *Proc. IEEE Int. Conf. on Pattern Recognition*, vol. 3, pp. 613–616, Sept. 2000.

[88] N. Otsu, "A Threshold Selection Method from Gray-level Histogram," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, Jan. 1979.

[89] P. Liao, T. Chen, and P. Chung, "A Fast Algorithm for Multilevel Thresholding," *Journal of Information Science and Engineering*, vol. 17, pp. 713–727, 2001.

[90] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing: Analysis, and Machine Vision.* Brooks Cole, Sept. 1998.

[91] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multiscale Transforms," *IEEE Trans. on Information Theory*, vol. 38, pp. 587–607, Mar. 1992.

[92] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear Time Euclidean Distance Transform Algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 529–533, May 1995.

[93] O. Cuisenaire and B. Macq, "Fast and Exact Signed Euclidean Distance Transformation with Linear Complexity," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 6, pp. 3293–3296, Mar. 1999.

[94] M. A. Tekalp, *Digital Video Processing.* Pearson Education, Aug. 1995.

[95] C. Guestrin, F. Cozman, and E. Krotkov, "Fast Software Image Stabilization with Color Registration," in *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, vol. 1, pp. 19–24, Oct. 1998.

[96] P. Burt and P. Anandan, "Image Stabilization by Registration to a Reference Mosaic," in *DARPA Workshop on Image Understanding*, pp. 425–434, Nov. 1994.

[97] M. Irani, B. Rousso, and S. Peleg, "Recovery of Ego-motion Using Image Stabilization," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, pp. 454–460, June 1994.

[98] C. H. Morimoto and R. Chellappa, "Automatic Digital Image Stabilization," in *Proc. IEEE Int. Conf. on Pattern Recognition*, Aug. 1996.

[99] A. Matami, Y. Ban, O. Oshiro, and K. Chihara, "A System for Superimposing a Medical Image Within the Subject," in *Proc. IEEE Int. Conf. on Bridging Disciplines for Biomedicine*, vol. 2, pp. 637–638, Oct. 1996.

[100] N. Shareef, D. L. Wang, and R. Yagel, "Segmentation of Medical Images Using LEGION," *IEEE Trans. on Medical Imaging*, vol. 18, pp. 74–91, Jan. 1999.

[101] C. Weidong, D. Feng, and R. Fulton, "Web-based Digital Medical Images," *IEEE Trans. on Computer Graphics and Applications*, vol. 21, pp. 44–47, Jan. 2001.

[102] J. A. Franco, M. Moctezuma, and F. Parmiggiani, "A Fusion-based Segmentation Algorithm for High-resolution Panchromatic Aerial Photography," in *Proc. IEEE Symposium on Geoscience and Remote Sensing*, vol. 6, pp. 3396–3398, June 2002.

[103] S. Levitt and F. Aghdasi, "Texture Measures for Building Recognition

in Aerial Photographs," in *Proc. IEEE Symposium on Communications and Signal Processing*, pp. 75–80, Sept. 1997.

[104] F. P. Lovergine, A. Branca, G. Attolico, and A. Distante, "Leather Inspection by Oriented Texture Analysis with a Morphological Approach," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, pp. 669–671, Oct. 1997.

[105] D. L. Paul, K. E. Pearlson, and J. R. R. McDaniel, "Assessing Technological Barriers to Telemedicine: Technology-Management Implecations," *IEEE Trans. on Engineering Management*, vol. 46, pp. 279–288, Aug. 1999.

[106] C. Kugean, S. M. Krishnan, O. Chutatape, S. Swaminathan, N. Srinivasan, and P. Wang, "Design of a Mobile Telemedicine System with Wireless LAN," in *Proc. IEEE Conf. on Circuits and Systems*, vol. 1, pp. 313–316, Oct. 2002.

Vita

Serene Banerjee was born to Prof. J. N. Bandyopadhyay and Prof. Swapna Banerjee on November 20th, 1977. During the 1999–2000 academic year, she received Bachelors of Technology (with honors) in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur. In 2001, she graduated with Masters in Electrical Engineering from the University of Texas at Austin. Serene is currently a full–time Ph.D. student in Electrical Engineering.

In 1999, the Electrical Engineering Department employed her as a teaching assistant for Real–time Digital Signal Processing Laboratory and Senior Design Project under Prof. Brian L. Evans, and Prof. Baxter F. Womack, respectively. From 2000 through 2004, she was a research assistant under Prof. Brian L. Evans, where she developed real–time video compression techniques, and smart image acquisition software. In summers of 1998, 2001, and 2002, she was an intern at Hughes Software Systems, India; Nokia R&D Center, Irving, TX; and Ricoh Research Center, Menlo Park, CA, respectively. She has developed algorithms for fast image and video compression, and transmission of JPEG 2000 compressed images.

She is a student member of the Institute of Electrical and Electronics

Engineers (IEEE).

Permanent Address: B-165, IIT Kharagpur,
West Bengal 721302,
INDIA

This dissertation was typeset with LaTeX 2_ε[1] by the author.

---

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay and James A. Bednar.