1-28-2016

Semantic Segmentation and Object Detection Based On Active Contour Model and Fuzzy Clustering

Sara Memar Kouchehbagh
University of Windsor

Semantic Segmentation and Object Detection Based On Active Contour Model and Fuzzy Clustering

By

Sara Memar Kouchehbagh

A Dissertation
Submitted to the Faculty of Graduate Studies
through the School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

2016

Semantic Segmentation and Object Detection Based On Active Contour Model and Fuzzy Clustering

By

Sara Memar Kouchehbagh

APPROVED BY:

_____
Maria A. Amer, External Examiner
Department of Electrical and Computer Engineering, Concordia University

_____
A. Y. Alfakih,
Department of Mathematics and Statistics

_____
Imran Ahmad
School of Computer Science

_____
Robin Gras
School of Computer Science

_____
Riadh Ksantini, Co-Advisor
School of Computer Science

_____
Boubakeur Boufama, Advisor
School of Computer Science

Nov 23, 2015

DECLARATION OF CO-AUTHORSHIP AND PREVIOUS PUBLICATION

I.   Co-Authorship Declaration:

I hereby declare that this Dissertation incorporates the outcome of a joint research undertaken in collaboration with Dr. Boufama, Dr. Ksantini, and Dr. Jin. The collaboration is covered in Chapter 3, Chapter 4, and Chapter 5 of the Dissertation.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my Dissertation, and have obtained written permission from each of the co-author(s) to include the materials in my Dissertation. I certify that, with the above qualification, this Dissertation, and the research to which it refers, is the product of my own work.

II.   Declaration of Previous Publications:

This Dissertation includes 3 original papers that have been previously published/submitted for publication in conferences and peer reviewed journals, as follows:

- Sara Memar, Karen Jin, Boubakeur Boufama: Object Detection Using Active Contour Model with Depth Clue. ICIAR 2013: 640-647.
- Sara Memar, Riadh Ksantini, Boubakeur Boufama: Feature-based Active Contour Model for Object detection. Submitted to *Journal of the Optical Society of America.* Date of submission: Sept. 16, 2015.
- Sara Memar, Riadh Ksantini, Boubakeur Boufama: Multiple Object Detection with Occlusion Using Active Contour Model and Fuzzy C-Mean. ICIAR (1) 2014: 224-233.

I certify that I have obtained a written permission from the copyright owners to include the above published materials in my Dissertation. I certify that the above materials describe work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my Dissertation does not infringe upon anyone copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my Dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include the above materials in my Dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

We propose a novel method for image segmentation and object detection. The proposed strategy is based on two major steps. The first step corresponds to image segmentation which is based on Active Contour Model (ACM) algorithm. The gradient stopping function has been widely used in most ACMs as an edge indicator. Because of the gradient high sensitivity to texture and noise, other stopping functions, such as polarity, have been proposed with some limited success. Unfortunately, most of these proposed stopping functions, including gradient and polarity, fail to detect objects effectively in many circumstances. On the other hand, depth information, if available, could provide the better clue for object detection. The proposed method takes the advantage of the existing contour models by using the depth clue, from either Kinect sensor or stereo vision algorithm, instead of two-dimensional clues, in the model stopping function. However, even with depth clue, it is still difficult to accurately detect a salient object when it is located at similar depths of others. Indeed, based on specific image data or genre of the image, the best candidate for a stopping term could be either a single feature such as gradient, polarity or depth, or a combination of them. So the proposed ACM is based an automatic selection of best candidate features among gradient, polarity and depth, coupled with a combination of them by Kernel Support Vector Machine (KSVM). Although existing techniques, such as the ones based on ACM perform quite well in the single-object case and non-noisy environment, these techniques fail when the scene consists of multiple occluding objects, with possibly similar colors. Thus, the next step corresponds to the identification of salient and occluded objects based on Fuzzy C-Mean

(FCM) algorithm. In this latter step, the depth is included as an important clue that allows us to estimate the cluster number and to make the clustering process more robust. In particular, occlusions are easily handled this way, and the objects are properly segmented and identified. The experiments, carried out on real images, have shown the success and effectiveness of our proposed method to detect the salient objects.

DEDICATION

To My Husband, My Parents and My Brother for Their Love

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Boubakeur Boufama, who shared with me his expertise and research insight. I also want to express my gratitude to my co-advisor, Dr. Riadh Ksantini, who gave me many useful suggestions during my research.

I also wish to express my appreciation to Dr. Maria A. Amer, Concordia University, Dr. Alfakih, Department of Mathematics and Statistics, Dr. Imran Ahmad and Dr. Robin Gras, School of Computer Science for being in the committee and spending their valuable time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACM	Active Contour Model
SVM	Support Vector Machine
KSVM	Kernel Support Vector Machine
FCM	Fuzzy C-Mean
PDE	Partial Differential Equation
FAB	Forward-And-Backward
RGB	Red-Green-Blue
GAC	Geodesic Active Contour
CV	Chan-Vase
SDK	Software Development Kit
RD	Reaction Diffusion
LSF	Level Set Function
SPF	Signed Pressure Function
LIF	Local Image Fitting
SBGFRLS	Selective Binary and Gaussian Filtering Regularized Level Set
DRLSE	Distance Regularized Level Set Evolution
FMLE	Fuzzy Maximum Likelihood Estimation
B3DO	Berkley 3-D Object
w.r.t.	With Respect To

CHAPTER 1

INTRODUCTION

1.1. Background

Computer vision is a discipline aiming to automatically extract, analyze and understand meaningful information from visual data such as images and videos. It has a variety of applications in our modern world. To mention a few, autonomous vehicles, medical image analysis, robotics and face recognition are some famous applications of computer vision. Image processing and pattern recognition are two fundamental techniques computer vision utilizes to reach its goals. The main objective in image processing is to pre-process the image and convert it to a suitable form for further analysis, while pattern recognition employs machine learning and statistical techniques to classify various patterns. Image segmentation is a central goal in computer vision, and it pertains to extracting the salient and coherent regions that semantically belong to different objects. The more accurate the segmentation is accomplished, the more effectively semantic objects can be detected. Segmentation and object detection play the essential role toward the image understanding, and they are still considered as the open challenges due to the complexity of images and object classes.

In segmentation methods, region is considered as the most significant feature, and based on this feature, segmentation methods were categorized into four groups, namely pixel-

1

based, area-based, edge-based and physics-based. In the edge-based method, which is our main concern in this study, region is defined as a connected group of pixels surrounded by edge pixels creating a contour (Skarbek & Koschan, 1994). In the last two decades, active contour model (ACM) has attracted great attention from the computer vision research community. It has been widely applied as a framework for extracting the contour of salient objects from images (Ge *et al*., 2015; Wang *et al*., 2014; Zhang *et al*., 2013).

Current ACMs can be grouped into parametric ACMs and geometric ACMs, w.r.t. (with respect to) their implementation and representation (Li *et al*., 2005). Parametric active contours (Xu & Prince, 1998) are presented as parameterized curves in a Lagrangian framework, while geometric active contours are represented via level set functions in an Eulerian framework (Malladi *et al*., 1995). The important advantage of geometric active contours is that they can break or split automatically in the evolution process, thanks to the level set function (LSF). The fundamental concept is to evolve the LSF w.r.t. a partial differential equation (PDE). Therefore, a new evolution of PDE for a LSF can be derived by minimizing a particular energy function defined on a LSF. These methods (Osher & Fedkiw, 2002; Zhao *et al*., 1996), known as variational level set methods, have shown more robust results than the methods using PDE for parameterized curves.

Geometric active contours (snakes) are implemented via the level set methods w.r.t. minimizing an energy function. Snake is presented by the zero LSF ϕ_0 which is the result of the intersection of the plane with a surface. Generally, the energy of snake is expressed

2

in terms of the internal energy and external energy. While internal energy controls the shape and size of the snake, the external energy acts as forces that aim at moving the curve toward the object boundaries, based on some image features. The zero LSF evolves w.r.t. time using an evolution equation, allowing it to be re-initialized and to keep the LSF close to an approximation of the signed distance function. However, in traditional level set methods, when the zero LSF is very steep, it may move away from its main pose or LSF. This issue causes the LSF to be far away from a signed distance function. Hence, the re-initialization of the LSF to the signed distance function becomes impossible. Furthermore, this re-initialization is usually expensive and complicated (Li *et al*., 2005).

To address the above problems, a variational LSF was proposed by (Li *et al*., 2005) where the distance regularization term was defined with a potential function as a penalty term in the internal energy function. Such a penalty term makes the LSF have a unique forward-and-backward (FAB) diffusion effect, which retains the LSF to the signed distance function. However, the FAB diffusion with the defined penalty term for energy minimization is a backward diffusion with a large rate. Thus, such a strong backward diffusion leads to the periodic "peaks" and "valleys" in ϕ. Although such fluctuations happen at a specific distance to ϕ, they may deform the zero LSF a little. To overcome this problem, (Li *et al*., 2010) proposed a new potential function, called double-well, which shows a different diffusion rate and diffusion behavior. We follow the approach in (Li *et al*., 2010) for the variational level set method to avoid doing the expensive process of re-initialization.

3

Object detection is a challenging vision task that is an important step in numerous applications such as scene understanding, video surveillance, image search and medical applications. This task aims to find and locate the salient regions of the image which semantically belong to objects, assuming that the type of the object is unknown. In object detection, promising results have been recently obtained for various object classes including pedestrian (Liu *et al*., 2014; Negri *et al*., 2014), car (Zheng & Liang, 2009; Hota *et al*., 2010), and face (Ban *et al*., 2014; Jun & Kim, 2012). Such detection methods typically learn from object classifiers using a labelled training set. Given a testing set, the classifier is used on the sub-windows with different sizes at all positions.

 Although some studies (Cheng *et al*., 2014; Rafiee *et al*., 2013) addressed the problem of segmentation and object detection without relying on object class, only one salient object has been the main concern. However, the presence of multiple occluding objects makes the detection even harder. The problem of partially occluded objects was tackled in the work of (Wu *et al*., 2008; Shet *et al*., 2007) using part-based representations and joint analysis of multiple objects. However, such techniques were employed on some specific moving objects, and their performance for very crowded scenes is far from perfect.

4

1.2. Motivation and Objectives

In this study, we followed the level set method in (Li *et al*., 2010) as it is free of the costly re-initialization process in level set evolution. However, in the external energy function of (Li *et al*., 2010), only image gradient was applied as an edge indicator function. Hence, only objects whose edges are well defined by gradient can be correctly detected. Moreover, as the gradient is sensitive to textured or noisy regions, the curve may stop before reaching the actual edge of the salient object. In some cases, it may pass through the edge if the gradient has small local maxima on the edges of the object. To overcome such difficulties, image polarity was applied in the external energy function as a stopping term in (Allili *et al*., 2007; Ksantini *et al*., 2009; Ksantini *et al*., 2013), and it was indicated that polarity can detect region or boundaries of the object of interest better in the presence of texture. However, both gradient and polarity are sensitive to noise whether texture is present or not, and their results are not always optimal, especially when the initial contour is set inside of the objects contour.

Besides these two features, depth information, if available, could provide the better clue for object detection as the pixels of the depth image correspond to a depth or a distance. Disparity map, obtained from two images, can be used as a depth clue and is a good alternative feature in an ACM stopping term. Depth information does not have to be actual depth. In our case, we have used disparities as depth clue since they are easier to obtain and do not have any limitations. There are also newly available and affordable

5

depth sensors that can be used to obtain depth clue. Depth disparity does not suffer from ambiguities that are inherent to two-dimensional images. So one of our objectives in this study is to detect the salient object more effectively by applying depth clue as a stopping term in ACM (Memar *et al*., 2013).

However, it is still challenging to clearly detect or differentiate objects using depth clue when they are located at similar depths. For example, the depth clue might not be useful for detecting a dish plate (flat object) located on a table, as both the table and the plate have very similar depth information. Furthermore, depth sensors have some limitations; for example, estimated depth for small and/or distant objects is not accurate. Indeed, based on specific image data or genre of the image, the best feature candidate for a stopping term could be either a single feature such as gradient, polarity or depth, or any combinations of them. So another main objective followed in this study is to automatically and correctly select and combine the best candidate feature(s) as a stopping term w.r.t. image data.

When a 3D scene is projected to a 2D image, some information will be lost. For example, parallel lines will intersect, and separate objects may become connected and occluded. Therefore, occluding objects are segmented within the same contour using ACM-based segmentation techniques. So a fuzzy clustering approach (FCM) might be employed to help in better segmenting the occluding objects in still images. However, the C-value in the FCM algorithm should be known in advance. Many validity indices have been

6

proposed to determine the best number of clusters (Bouguessa *et al*., 2006; Mishra *et al*., 2012) with mixed results. In particular, different validity indices yield different cluster numbers, depending on the image data. This is because these methods rely on image statistics, and they are affected by pixel noise, which is always present in real images. Finally, we aim to properly segment and identify the salient objects particularly in the challenging case of occluded multi-object scenes by automatically choosing the cluster number in fuzzy clustering algorithm.

1.3. Research Scope

This section presents the summary of the direction followed in this thesis to achieve the objectives w.r.t. problem statements in segmentation and object detection.

Computer vision has numerous applications like scene reconstruction, video tracking, object recognition, object detection, motion estimation, and image restoration. Detecting the object of interest in the complex image is our main concern in this study. Image segmentation, a prerequisite for many computer vision applications, can be used to help with object detection.

Generally, segmentation techniques were categorized into five groups, namely threshold-based techniques, edge-based methods, region-based methods, clustering techniques, and

7

matching (Lucchese & Mitra, 2001). In particular, ACM, an edge-based segmentation technique and FCM, a clustering technique, are our main focus in this study that aim at segmenting and detecting objects in still images. Note that both ACM and FCM have been used in a variety of applications.

ACMs are categorized into three main groups, namely, edge-based, region-based, and shape-based models. Since region-based models are sensitive to intensity homogeneity, and shape-based ACM needs a large amount of training samples, we focus on edge-based models. As mentioned earlier, ACM is presented by internal and external energy functions. Many outstanding techniques have been proposed to address the costly re-initialization problem in the internal energy function of ACMs (Li *et al*., 2010; Zhang *et al*., 2013). However, the external energy function in most studies (Li *et al*., 2010; Tian *et al*., 2011; Zhang *et al*., 2013; Mumford *et al*., 1989) is based on image gradient which is so sensitive to texture. So we aim to address this main problem of the edge-based ACM.

Among clustering techniques, FCM is utilized for better segmentation and object detection in this study. FCM algorithm has shown promising results in segmentation by improving the compactness of the regions. It is a pixel clustering technique where, similar pixels should be in the same group while, the dissimilar ones are assigned to different clusters. Although FCM algorithm benefits from simplicity of its implementation, the number of clusters, the C-value in FCM, should be determined

8

manually and known in advance. Thus, we aim to improve object detection by automatically determining C-value or the number of the objects in the still images.

1.4. Research Contributions

To address the problem of segmentation and object detection in occluded multi-object scenes using ACM and fuzzy clustering algorithm, we proposed different methods w.r.t. research motivation and research objectives. The main contributions of this thesis are summarized as follow:

- Proposing a new method which takes the advantage of the existing contour models by using the depth clue, from Kinect sensor or stereo vision algorithm, instead of two-dimensional clues, in the model stopping function. The depth disparity is applied in the external energy function of the ACM for detecting the object of interest in the image (Chapter 3).

- We proposed a new segmentation method that automatically selects the best candidate feature(s) among gradient, polarity, and depth to be used in the stopping function of the well-known ACM. The proposed method has full flexibility w.r.t. the above-mentioned features, as the latter can be employed either individually or collectively in a coherent framework, depending on each individual image. We have also proposed a method to combine the selected

9

features in a semi-supervised manner using the kernel-based Support Vector Machine (KSVM) that will be the most effective stopping term in the ACM (Chapter 4).

- The final contribution of this thesis corresponds to the identification of salient and occluded objects based on Fuzzy C-Mean algorithm. In this step, the depth is included as an important clue that allows us to estimate the cluster number and to make the clustering process more robust. In particular, occlusions are easily handled, and the objects are properly segmented and identified (Chapter 5).

1.5. Thesis Organization

This thesis is outlined in 6 chapters.

Chapter 2: This chapter consists of the literature review on the segmentation and object detection approaches; different ACMs and in particular state-of-the-art techniques based on ACM are also highlighted.

Chapter 3: Sara Memar, Karen Jin, Boubakeur Boufama: Object Detection Using Active Contour Model with Depth Clue. ICIAR 2013: 640-647.

This chapter presents the details of the proposed method in which depth clue is employed and incorporated in the external energy function of the ACM for segmenting and detecting the salient object in the image. The experiments carried out on real images, comparison and discussion are also provided.

Chapter 4: Sara Memar, Riadh Ksantini, Boubakeur Boufama: Feature-based Active Contour Model for Object detection. Submitted to *Journal of the Optical Society of America A*.

It describes different steps of the proposed feature-based ACM. The details of an automatic selection of best candidate features among gradient, polarity and depth, coupled with a combination of them by KSVM will be discussed. This chapter also includes the results, discussion and comparison.

Chapter 5: Sara Memar, Riadh Ksantini, Boubakeur Boufama: Multiple Object Detection with Occlusion Using Active Contour Model and Fuzzy C-Mean. ICIAR (1) 2014: 224-233.

This chapter corresponds to detecting multiple occluded objects in images with complex background based on FCM algorithm which is followed by experimental results and discussion.

11

Finally, chapter 6 concludes and summarizes the most important aspects of the research. This chapter ends with the suggested future works.

CHAPTER 2

LITERATURE REVIEW

2.1. Segmentation

Image segmentation can be defined as the process of dividing an image into homogeneous and meaningful regions (Lucchese & Mitra, 2001) which is an important step in image analysis and image understanding. Over-segmentation and under-segmentation are two major problems with image segmentation. As Figure 2.1 indicates, in over-segmentation, the pixels of the same object are segmented into different regions, while pixels of the different objects are identified as the same object in under-segmentation. Many segmentation techniques have been proposed, and they are commonly divided into the following categories:
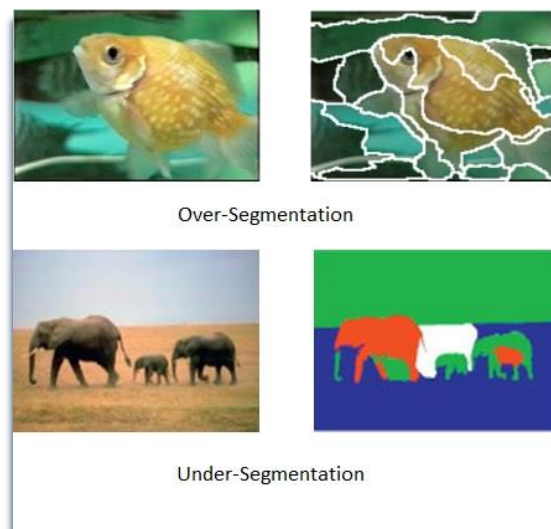
Figure 2-1: Over Segmentation and Under Segmentation (Yuan *et al*., 2005).

- **Threshold-based:** This technique is commonly used for segmenting the image. It is useful when there is a bright object on a dark background or vice versa. The threshold is often tuned interactively until a suitable segmentation result has been achieved. Histogram is a useful tool for setting the threshold value. For example, in (Siang Tan & Mat Isa, 2011), the histogram thresholding method was proposed to tackle the problem of image segmentation. Threshold techniques are usually utilized in combination to other segmentation techniques.

- **Region-based:** Region-based techniques are usually based on region growing which aims at grouping pixels or sub-regions into larger regions considering pre-defined criteria. Split and merge and texture analysis are also common strategies in region-based image segmentation (Shih and Cheng, 2005; Preetha, 2012). Unlike region-growing, the image is split into four sub-images so that $P(R_i)$ = False. Then, any adjacent regions are merged if $P(R_i \cup R_j)$ = True. This process will terminate when no more merge and split is feasible w.r.t. a specific criterion. The main drawback with this technique is that it is not possible to merge two segments belonging to two different levels of pyramid (Gonzalez & Woods, 2006).

- **Edge-based:** The object is identified by its boundaries, and the main goal of this segmentation technique is to find the edges of the object. Edge linking, watershed, and active contour are famous edge-based

14

segmentation models. Hough transform (Hart, 2009; Duda and Hart, 1972) has been commonly used as an edge linking technique aiming to find the ideal edges presenting a certain class of parametric shapes like line, circles, or ellipses. Hough transform is time-consuming for finding objects with complex shapes. It also demands post-processing to connect the correct line segments, and it is required to know the nature of the shape we are looking for in the image. Watershed segmentation (Beare, 2006; Vincent & Soille, 1991) is based on image gradient, and some barriers (dams) are built on image gradient. Then the built dams are formed to identify the semantic region. However, it may lead to over-segmentation when the gradient effect is high.

- **Clustering techniques:** The main goal of clustering techniques is to group the patterns which are similar to each other in a set of features. K-mean algorithm (Hartigan & Wong, 1979; Bishop, 2006), FCM (Bezdek, 1981), and Fuzzy Maximum Likelihood Estimation (FMLE) (Gath & Geva, 1989) are some common clustering algorithms that have been used in image segmentation. FMLE is the extension of the FCM by considering the covariance for each cluster rather than the mean value. The main disadvantages of such techniques are center initialization and determining the number of cluster.

15

- **Matching:** This technique is based on this assumption that the object we wish to locate and identify in the image should be known in advance (Dey *et al*., 2003). But, it is not always feasible to have this knowledge. So its application is useful when the object of interest is known such as detecting the hand contour (Park *et al*., 2012).

Edge-based and clustering techniques are taken into account among the above-mentioned techniques. The main problem with the edge-based techniques could be the presence of noise and artifacts which could diminish the performance of segmentation result. There are some classical edge detection techniques like Canny (Canny, 1986), Sobel (Pujar and Shambhavi, 2010), Ridge (Lindeberg, 1996), and Gabor (Lee, 1996) which perform by getting the convolution of the image with filters. However, such techniques fail to detect non-rectilinear objects, i.e. objects with the curved boundaries, and only rectilinear objects can be detected. This drawback has been addressed by some techniques like Hough transform and matching, but they demand the prior knowledge about the shape of the object. So we focus on edge-based active contour technique for image segmentation in this study.

2.2. Active Contour Model

Active Contour Model (ACM) was initially identified in 1988 by Kass *et al*. It is also called *Snake* because of the way the contour moves during the evolution. Image

16

segmentation based on ACMs is useful as the shape of the object is not required to be known, making this technique applicable in many applications. It is also robust to weak and incomplete edges. The main motivation for ACM is that it is able to effectively detect the object with the curved boundaries. In classical ACMs, the object boundary is presented by a parametric curve aiming to deform an ideal initial shape towards the shape of the salient object. This goal is accomplished by minimizing an energy function.

2.2.1. *Parametric Active Contour Model*

The contour is defined as a two-dimensional parametric curve (v(s)), where v(s) = (x(s), y(s)). Total energy of snake (E_{Snake}) is expressed in terms of internal energy (E_{int}), external energy (E_{ext}), and constraint energy (E_{con}) as follows:

$$E_{Snake} = E_{int} + E_{ext} + E_{con}. \tag{2.1}$$

1. **Internal Energy Function:** This function controls the length and smoothness of the curve. So this function consists of two terms, namely, *Elastic* energy and *Bending* energy ($E_{int} = E_{elastic} + E_{bending}$). Elastic energy ($E_{elastic}$) is the first order derivative of the contour that aims at adjusting the length of the curve by applying tension, and it is defined as below:

$$E_{elastic} = \frac{1}{2}\int_{s} \alpha(s)|v_s|^2 ds, \tag{2.2}$$

17

where $v_s = \dfrac{dv(s)}{ds}$, and the coefficient $\alpha(s)$ helps to adjust the elastic term along contours' points. Thus, as Figure 2.2 illustrates, this term tries to remove the curve concavity and shrink the curve. However, in the absence of any forcing term, the initial contours collapse to a point.
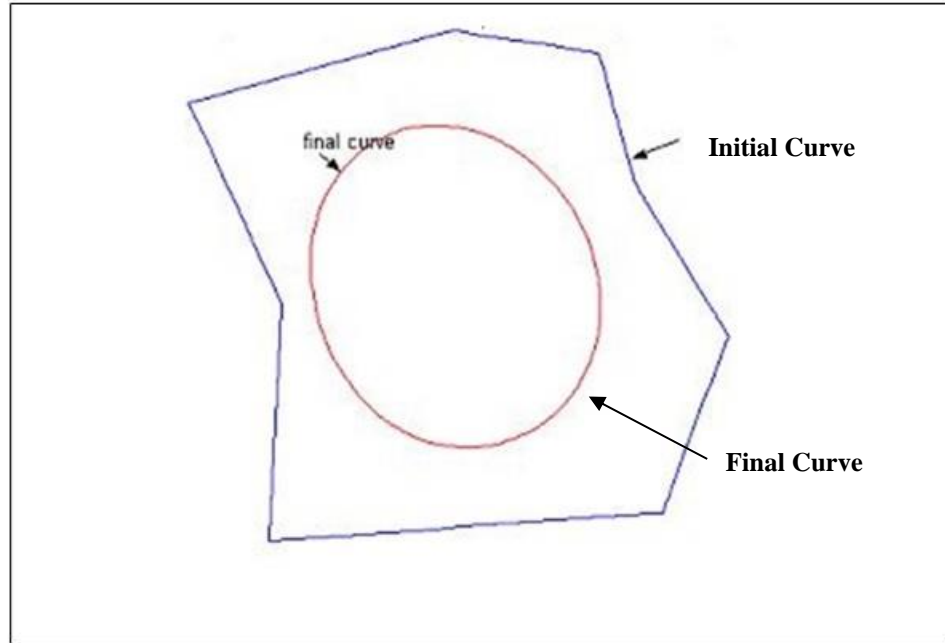
Figure 2-2: The Effect of Elastic Term on Contour (Poonawala & Milanfar, 2002).

The second term is bending force ($E_{bending}$) that aims at eliminating very sharp corners and making the final contour to be a smooth curve. This term is the second derivative of the contour and defined as follow:

$$E_{bending} = \frac{1}{2}\int_s \beta(s)|v_{ss}|^2 ds, \qquad (2.3)$$

18

where $v_{ss} = \dfrac{d^2 v(s)}{ds^2}$, and $\beta(s)$ is its coefficient which controls the level of the

smoothness. As Figure 2.3 demonstrates, using bending force, the initial

curve with high bending energy and sharp corners are smoothed and

deformed to a final curve with a lower bending energy. Bending term can

deform the initial curve to a circle if it is applied on the contour for the

infinite number of iteration as a circle has the minimum bending energy.
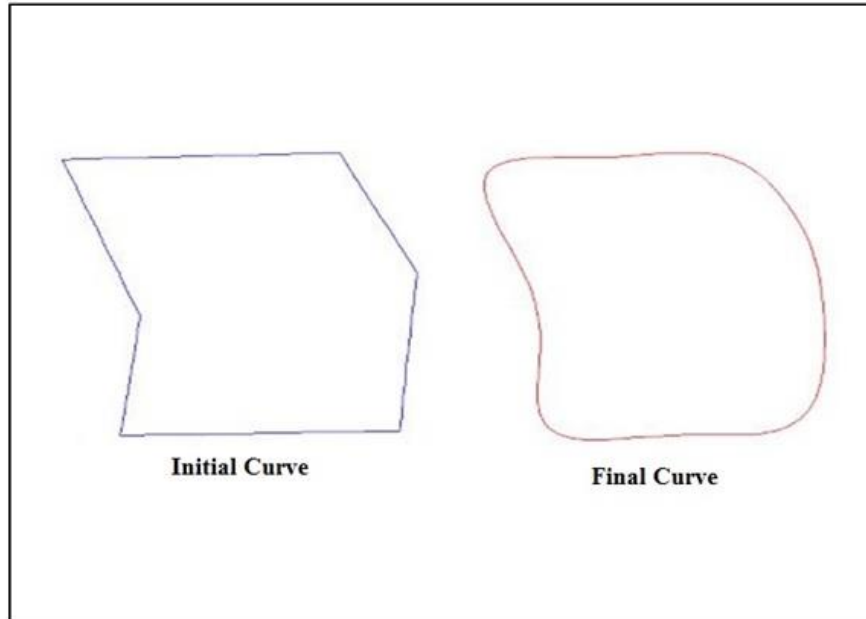
Figure 2-3: The Effect of Bending Term on the Contour (Poonawala & Milanfar, 2002).

2. **External Energy Function:** This function is defined based on image features.

Its definition is formulated as below:

$$E_{ext} = \int_s E_{image}(v(s))ds. \qquad (2.4)$$

19

www.manaraa.com

An image can be presented by some features like line or edge, and E_{image} should be minimized and get small values on such features. In many edge-based ACMs, the external energy function is defined based on image gradient which is commonly used for the edge detection. Some of these models will be discussed in details, later.

3. **Constraint Energy Function:** This energy allows the user to interactively manipulate the snake in the evolution process. The constraint term is optional, and it can be set by the user on the contour.

The parametric ACMs are very sensitive to parameters like $\alpha(s)$, $\beta(s)$ and standard deviation of Gaussian function (σ) which is normally used in the external energy function. Moreover, external force is not able to execute on points which are far away from the boundary. It also fails to extract the objects with boundary concavities.

2.2.2. Geometric Active Contour Model

Geometric ACMs were introduced by Caselles *et al*. (1993) and Malladi *et al*. (1995). The geometric models are implemented via LSF by minimizing an energy function. The initial curve in geometric ACM is presented by a zero LSF (ϕ_0) which is the result of the intersection of the plane with a surface. The main advantage with such models is that the

LSF can split or merge during the evolution, and the topological changes are automatically controlled (Li *et al*., 2005).

The active contour ($\mathbb{C}$) is presented by the zero LSF as follow:

$$\mathbb{C}(t) = \{(x, y) \mid \phi(t, x, y) = 0\}, \tag{2.5}$$

where $\phi(t, x, y)$ is the LSF which is evolved w.r.t. time using the following evolution equation:

$$\frac{\partial \phi}{\partial t} + F\left|\nabla \phi\right| = 0, \tag{2.6}$$

where F is the speed function, and it is defined based on LSF (ϕ) and image data.

LSF (ϕ) can develop very sharp and move away from its origin during the curve evolution. So a numerical scheme was used in traditional level set methods (Caselles *et al*., 1993; Malladi *et al*., 1995; Osher & Fedkiw, 2002) to keep the LSF close to the sign distance function using the following re-initialization equation:

$$\frac{\partial \phi}{\partial t} = sign(\phi_0)(1 - \left|\nabla \phi\right|), \tag{2.7}$$

where ϕ_0 is the re-initialized function, and $Sign(\phi)$ is the sign function. Various re-initialization techniques have been proposed (Peng *et al*., 1995; Sussman & Fatemi, 1999). However, such techniques are computationally costly, and it is often challenging

21

to find out how and when the LSF should be re-initialized to a signed distance function. Thus, a variety of studies have addressed this problem by proposing reinitializing free LSF. More details about such techniques will be provided, later.

2.3. Types of Active Contour Model

The various types of existing ACMs include region-based, edge-based, shape-based, and any combination of them. More details for each category is provided as follows:

2.3.1. Region-based (C-V model)

Region-based ACMs have been recently proposed and employed for image segmentation. The main goal of such region-based models is to incorporate the statistical information of the inside and outside the initial contour for the curve evolution and identifying the region of interest. Thus, they have performed more effectively for images with weak object boundaries (Tian *et al*., 2011; Wang *et al*., 2010). The main advantage with such models is that they are not very sensitive to the location of the initial contour. A review of literature has shown that the main application of such models is in medical image analysis where the images are presented with non-uniform daylight and artificial illumination (Crandall, 2009).

22

Chan-Vase (C-V) is the well-known geometric region-based model (Chan & Vase, 2001). The C-V model is the generalization of the Mumford-Shad model where the optimal image decomposition into several regions is accomplished by minimizing a smooth function proposed by Mumford and Shad (1989). The C-V model solved this minimization problem by incorporating LSF where the global statistical information was used for evolving the curve. In spite of getting larger convergence range and effectively handling topological changes using C-V model, it fails to perform well for images with intensity inhomogeneity. The main reason is that C-V algorithm is based on this assumption that image intensities in each region kept constant, and thus, this issue leads to the incorrect movement of the contour. Another limitation of this famous model is its dependence to numerical scheme like the re-initialization process, making this technique so costly and time-consuming. Some studies aimed to address the limitations of C-V model; however, their performance is still very expensive (Vese & Chan, 2002; Solem *et al*., 2006; Pi *et al*., 2007).

2.3.2. *Edge-based (GAC Model)*

In edge-based ACMs, the external energy function is defined based on an edge-based stopping term or an edge indicator. Such models mainly use image gradient to identify the desired object boundaries. Geodesics Active Contour (GAC) model is one of the most popular geometric edge-based models (Caselles *et al*., 1997).

23

The GAC model is defined based on geodesic or minimal path computation for identifying the boundaries. It also takes into account the relation between the classical active contours and geometric models. So the geometric GAC is based on the gradient-based stopping term for identifying the object boundaries and a balloon force term aiming to adjust the motion of the curve such as shrinking or expanding the contour during the evolution. However, most edge-based approaches including GAC model depend on image gradient for stopping the curve evolution. Hence, only objects whose edges are well defined by gradient can be correctly detected. Furthermore, gradient is sensitive to texture, and the curve may stop before capturing the desired boundaries. The curve may also pass the actual edge as gradient is prone to small local maxima on the edges of the object. The initial contour should be set properly in such edge-based models, and inappropriate location of the initial contour may lead to poor segmentation result.

2.3.3. Shape-based

The shape-based models demand the prior knowledge about the shape of the object we aim to detect in the target image. Although such approaches have shown good performance in case of texture and occlusion, they need a substantial amount of training samples (Malladi *et al*., 1995; Chen *et al*., 2002). Besides, since a shape-based model is based on this assumption that the prior shape knowledge should be known, its application is limited in practice.

There are also some ACMs based on the combination of region, edge, and shape information. For example, the model in (Bresson *et al*., 2006) is the combination of edge-based GAC model, Mumford-Shah region based model and shape information. Tian (*et al*., 2011) proposed an ACM based on both region and edge information. However, several parameters should be tuned for their models when tested on different images. Moreover, some of these techniques fail to perform effectively with images presented with local intensity inhomogeneity.

2.4. Review of Baseline Active Contour Models

The details of some baseline ACMs are discussed in this section. Such models are either region-based or edge-based, and have demonstrated good performance in segmenting the image.

Li *et al*. (2010) proposed an edge-based ACM called distance regularized level set evolution (DRLSE) where the regularity of the LSF is kept during the curve evolution, making this model free from expensive re-initialization procedure. The proposed regularization term is based on a Forward-and-Backward (FAB) diffusion effect which is able to effectively control the desired shape of the LSF during the evolution. The external energy function in this model is based on gradient information aiming to move the LSF toward the desired object boundaries. The proposed DRLSE method allows the use of large time steps, making this model significantly efficient as it needs less computation

time and number of iterations. So we follow this model for the variational LSF which is re-initialization free and efficient in practice.

The proposed ACM in (Allili & Ziou, 2007) is based on region and edge information. This unsupervised color/texture segmentation approach relies on polarity-based stopping function. This model is able to segment images with high texture due to the use of polarity information. However, homogeneous seeds are initialized at different regions of the image, and its ACM does not avoid the costly re-initialization procedure. Ksantini *et al*. (2013) also utilized polarity information as an edge indicator in their proposed ACM for salient object detection in grey-level images and color images. Hence, images with high level of texture can be effectively segmented. This model is also practically efficient by following the re-initialization free LSF proposed in (Li *et al*. 2010).

In (Zhang *et al*., 2010a), a region-based ACM method using local image fitting (LIF) energy was proposed. It extracts the local image information, w.r.t. intensity inhomogeneity for segmenting the target image. It also applies a Gaussian Filter to regularize LSF and to avoid the costly re-initialization process. It achieved satisfying segmentation results when applied on the synthesis images.

The method proposed in (Zhang *et al*., 2010b) is another region-based ACM that is based on the Selective Binary and Gaussian Filtering Regularized Level Set (SBGFRLS). This

26

model considers the advantages of the edge-based GAC model (Caselles *et al*., 1997) and C-V region-based ACM (Chan & Vase, 2001). A new region-based signed pressure function (SPF) was introduced using statistical information inside and outside the contour. The SPF behaves like the edge stopping function in GAC model for stopping the contour evolution on the object boundaries, and it does the global segmentation like C-V model for detecting all objects, regardless of where the initial contour is set in the image. When using the model in (Zhang *et al*., 2010b), the initial contour can be set anywhere.

More recently, Zhang *et al*. (2013) proposed the new method called RD-LSF where a reaction-diffusion (RD) term is identified, making the level set evolution absolutely free from the costly re-initialization procedure. The main advantages of the RD-LSF method include its satisfying performance on weak boundary, and simple implementation of the RD equation without the requirement of initialization scheme. The proposed RD-LSF method has been tested on the famous region-based C-V model and edge-based GAC model, and it outperforms other classical region-based and edge-based ACMs, especially on noisy/textured images.

2.5. Image Features

In this section, image features such as gradient, polarity and depth will be discussed in more details. Gradient has been widely applied in many ACMs. However, it fails to work properly in case of high level of texture/noise. Polarity is another feature which has been

27

recently employed, and has shown promising results when there are a lot of textures. On the other hand, depth provides structure information of the scene and can be also used as a stopping function to detect the salient object. Overall, each feature has its own pros and cons, and one may outperform the others in a particular condition. Here, more details for each feature are provided as follow:

- **Gradient:**

Generally, the image gradient measures the intensity changes for each pixel. It has been applied in many applications, especially in edge detection. In most previous works (Zhang *et al*., 2013; Li *et al*., 2010; Tian *et al*., 2011; Caselles *et al*., 1997), stopping function (F_g) in ACM is defined based on the gradient information as follows:

$$F_g = \frac{1}{1 + |\nabla G_\sigma * I|^2},$$ (2.8)

where $G_\sigma = \frac{1}{2\pi\sigma^2} \exp^{-\frac{x^2+y^2}{2\sigma^2}}$ is the Gaussian kernel with σ^2 standard deviation. The value of F_g is in the range of 0 to 1, where on main edges, this value is close to 0 while on non-edge regions, it tends to be close to 1. The main drawback with the gradient is its poor performance when the salient object is located on a textured background. This is because some areas, not belonging to the boundaries of the object, will be identified as edges and hence, making the curve stop before reaching the actual contour.

28

29

- **Polarity:**

Polarity was introduced as an alternative edge stopping function in ACMs (Allili & Ziou, 2007; Ksantini *et al*., 2013) with some limited success when the image background is clouted with texture. Polarity is considered to be a local image feature that shows a measurement of the extent to which the gradient vectors, in a specific pixel neighborhood, are oriented in a dominant direction η. Polarity is defined as:

$$P(x,y) = \frac{|E_+ - E_-|}{E_+ - E_-}. \tag{2.9}$$

$$E_+ = \sum_{x,y} G_\sigma(x,y) * [\nabla I . n]_+ . \tag{2.10}$$

$$E_- = \sum_{x,y} G_\sigma(x,y) * [\nabla I . n]_- . \tag{2.11}$$

In the above equations, ∇I is the gradient of image intensity which can be calculated using the convolution of the image with the first derivative of a Gaussian filter. G_σ is a smoothing function with variance σ^2. The rectified positive and negative are denoted as $[\ ]_+$ and $[\ ]_-$, respectively. n is the unit vector which is perpendicular to the dominant vector (η) in a neighborhood of a pixel (x, y). While E_+ indicates how many gradient vectors in the window $G_\sigma(x,y)$ are in the positive side of the dominant vector (η), E_- shows such information in the negative side of the dominant vector.

30

The dominant vector is identified as the principal eigenvector of the matrix that is defined by the following expression:

$$M\sigma = \sum\nolimits_{x,y} G_\sigma(x,y) * (\nabla I) \times (\nabla I)^T.$$ (2.12)

The values of polarity for the image (I) are between 0 and 1. The polarity $P(I)$ gets very close to 1 on edges of the object of interest and gets very close to 0 when it falls inside a region. So polarity information as a stopping term in the external energy function of ACM can be formulated as:

$$F_p = 1 - P(I).$$ (2.13)

- **Depth:**

Disparity map (depth) in stereo images introduces additional information for improving the performance of the segmentation. Depth clue, obtained from disparity map, is not sensitive to texture as it reflects the depth information of the objects in the scene. There are several well-known stereo matching algorithms that effectively estimate the position of the objects (Yang, 2012; Heo *et al*., 2013; Banks *et al*., 1999).

Besides stereo-vision techniques, there are other ways to obtain this information. In particular, recent hardware using time-of-flight (sonic, infrared, laser) is capable to produce good quality depth information. For example, Microsoft Kinect is available for a low cost and allows the estimation of depth in addition to an RGB image. Microsoft Kinect sensor has attracted great attention both from the entertainment and

31

scientific communities. It has been employed in many computer vision applications ranging from skeletal tracking to speech recognition and 3D reconstruction (Barnachon *et al*., 2013; Shotton *et al*., 2011; Henry *et al*., 2012; Burgin *et al*., 2011). However, the main drawback with these techniques is that depth is only accurate for a predefined distance range and indoor scenes.

In this study, in addition to stereo-vision technique (Yang, 2012), we have used Microsoft Kinect sensor for obtaining the depth information of the scene. Kinect sensor consists of three devices, the center one is an RGB camera which takes about 30 frames per second. On the left side of the RGB camera, an infrared device projects the infrared rays towards the scene. On the right side of the RGB camera, an infrared sensor measures the time-of-flight of the returning infrared ray for each pixel, and hence getting the point-sensor distance. Kinect generates a 640x480 RGB image and a matching depth image as outputs. Due to different positions of the RGB and depth sensors in the Microsoft Kinect, depth image is not aligned with the RGB image. Hence, an RGB-to-depth calibration should be performed to ensure that the depth matrix is aligned with the RGB matrix. This was achieved using library functions from OpenNI. While pixels in the RGB image represent a measure of color or intensity, pixels of the depth image correspond to a depth or a distance.

In order to incorporate the depth clue in our method, the gradient of the disparity map is calculated and is used the same way as the classical image gradient in the ACM.

32

Note that when the actual depth is available, for example when using a Microsoft Kinect sensor, we will be using the gradient of the depth image itself.

More details about the use of disparity map and 3D information in ACM for the salient object detection will be explained in Chapter 3.

33

CHAPTER 3

OBJECT DETECTION USING ACTIVE CONTOUR MODEL WITH DEPTH CLUE

3.1 Background

Active contour models (ACM) have been widely employed for delineating an object of interest in images. In an edge-based ACM, the external energy function makes zero LSF move toward the edge of the objects. In other words, curve moves toward the object boundaries dynamically to produce the shape of the salient object. Therefore, stopping function should be defined in an external energy in order to make the curve stop on object boundaries. Although the gradient, as a stopping term, has shown promising results in ACMs (Zhang *et al*., 2013; Li *et al*., 2010; Tian *et al*., 2011; Caselles *et al*., 1997), its sensitivity to local minima makes it susceptible to stop before reaching the actual edge. Polarity is another feature used as a stopping function with some limited success, particularly, when the image background is clouted with texture (Allili & Ziou, 2007; Ksantini *et al*., 2013). In particular, the stopping functions defined in terms of gradient and polarity failed to detect salient objects effectively in many cases.

 On the other hand, depth information, if available, could provide the better clue for object detection. Disparity map in stereo images introduces additional information for improving the performance of the segmentation. In particular, depth disparity does not suffer from ambiguities that are inherent to two-dimensional images. The disparity map is

34

typically calculated from two images through the pixel matching process. The disparity is usually used to get the depth information of the entire object regardless of noise or texture inside or outside of that object. Thanks to the newly available low cost depth sensors, such as Microsoft Kinect, the use of depth disparity for solving object detection becomes reality. So we proposed a new method that uses depth clue with the classical contour models for detecting salient objects in noisy images.

The proposed method takes the advantage of the existing contour models by using the depth disparity clue, from Kinect sensor, instead of two-dimensional clues, in the model stopping function. The depth disparity is applied in the external energy function of the ACM for detecting the object of interest in the image. Experiments on several images collected by Microsoft Kinect indicated the superiority of depth clue over the polarity and gradient in detecting the object of interest using the ACM.

3.2 Method: Active Contour Model Based on Depth Information

In our study, depth disparity is defined as the stopping function in the external energy function. While pixels in the RGB image represent a measure of color or intensity, pixels of the depth image correspond to the calibrated depth of the scene. There are several well-known stereo matching algorithms that effectively estimate the position of the objects (Yang, 2013; Heo *et al*., 2013; Banks *et al*., 1999). Steps of stereo algorithms generally include matching cost computation, cost aggregation, disparity

35

computation/optimization, and disparity refinement. Traditional matching cost aggregation techniques were accomplished by summing/averaging matching cost over windows with constant disparity. Thus, such techniques are based on the local nature of traditional window-based cost aggregation algorithms and are vulnerable to the lack of texture. Yang (2012) addressed the issue of cost aggregation in stereo algorithm by proposing a non-local solution.

In (Yang, 2012), the matching cost values are aggregated adaptively based on pixel similarity on a tree structure derived from the stereo image pair to preserve depth edges. The nodes of this tree are all the image pixels, and the edges are all the edges between the nearest neighboring pixels. The similarity between any two pixels is decided by their shortest distance on the tree. The proposed method is non-local as every node receives supports from all other nodes on the tree. As can be expected, the proposed non-local solution outperformed all local cost aggregation methods. So in this study, we follow the method in (Yang, 2012) to get the disparity map.

In order to incorporate the depth clue in our method, the gradient of the disparity map, called hereafter F_D, is calculated and used the same way as the classical image gradient in the ACM. Note that when the actual depth is available, for example when using a Microsoft Kinect sensor, we will be using the gradient of the depth image itself. So the

gradient of depth information is defined and utilized in an external energy function of the ACM as follow:

$$F_D = \frac{1}{1+\left|\nabla G_\sigma * I_D\right|^2}, \tag{3.1}$$

where G_σ is the Gaussian kernel with σ standard deviation, and I_D is the depth image can be computed using either Kinect or stereo vision algorithm. The values for gradient of disparity are between 0 and 1, where this value is very close to 0 on edges of the salient object, and it is close to 1 on non-edge areas. So the stopping function is defined in terms of disparity information or F_D, and it is combined in the external energy function of the variational level set formulation proposed by Li *et al*. (2010). The reasons of employing this efficient and re-initialization free technique was discussed in chapter 2. So the proposed ACM based on the depth clue is defined as below:

$$\varepsilon_{F_D,\lambda,v}(\phi) = \lambda L_{F_D}(\phi) + v A_{F_D}(\phi), \tag{3.2}$$

where

$$L_{F_D}(\phi) = \int_\Omega F_D \delta(\phi)\left|\nabla\phi\right|dxdy. \tag{3.3}$$

$$A_{F_D}(\phi) = \int_\Omega F_D H(-\phi)dxdy. \tag{3.4}$$

In equation (3.3) and equation (3.4), δ and H are the Dirac delta and Heaviside functions, respectively. So in the external energy function (equation (3.2)), L_{F_D} controls

37

the line integral of the function F_D in equation (3.1) along the zero LSF, and the energy of L_{F_D} minimizes in a case that zero LSF of ϕ reaches the boundaries of the salient object. Besides that, A_{F_D} is applied on the energy function as well in order to control the motion of the zero LSF during the evolution, and its coefficient v can get either positive or negative values, where the positive value makes the initial curve shrink and the negative one makes it expand.

As mentioned before, we follow the approach of (Li *et al.*, 2010). The total energy function in this study is defined as below:

$$\varepsilon(\phi) = \mu R_P(\phi) + \varepsilon_{F_D, \lambda, v}, \tag{3.5}$$

where $\varepsilon_{F_D, \lambda, v}$ is the external energy function which was previously discussed and modified based on our needs, and it was defined in the equation (3.2). $R_P(\phi)$ is the internal energy function with the constant $\mu > 0$ which is considered as the level set regularization term in order to control the deviation of the LSF from the signed distance function, and it is described as below:

$$R_p(\phi) = \int_\Omega p(|\nabla \phi|) dx, \tag{3.6}$$

where P is regarded as a potential or energy density function $P : [0, \infty) \to \Re$ and it is defined as follow:

38

$$P(\phi) = \frac{1}{2} \int_{\Omega} (|\nabla \phi| - 1)^2 dx. \tag{3.7}$$

In (Li *et al*., 2005), the internal energy function was described only over $P(\phi)$, as a penalty term to retain the LSF close to the signed distance function and panelizes the deviation of the LSF from the signed distance function. However, due to unpleasant side effects of $P(\phi)$ on the LSF in some circumstances, a new potential function ($R_p(\phi)$) was added in the distance regularization term to remedy the situation (Li *et al*., 2010). In overall, the total energy function should be minimized using the following equation:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial \varepsilon}{\partial \phi}. \tag{3.8}$$

In equation (3.8) which is also known as gradient flow, $\dfrac{\partial \varepsilon}{\partial \phi}$ is the Gateaux derivative (first derivative) (Evans, 1998). Therefore, the Gateaux derivative of the total energy function which is expressed in terms of the function $\varepsilon(\phi)$ can be extended as:

$$\frac{\partial \varepsilon}{\partial \phi} = -\mu div(d_p(|\nabla \phi|)\nabla \phi) - \lambda \delta(\phi) div(F_D \frac{\nabla \phi}{|\phi|}) - v F_D \delta(\phi). \tag{3.9}$$

In equation (3.9), ∇ denotes the Laplacian operator, and d_p is a function which is described as below:

$$d_p(s) = \frac{P'(S)}{(S)}. \tag{3.10}$$

Then, the gradient flow or the proposed evolution equation of the LSF in (Li *et al*., 2010) is defined as below:

39

$$\frac{\partial \varepsilon}{\partial t} = +\mu div(d_p(|\nabla \phi|)\nabla \phi) + \lambda \delta(\phi) div(F_D \frac{\nabla \phi}{|\phi|}) + vF_D\delta(\phi). \qquad (3.11)$$

All the spatial partial derivatives $\frac{\partial \phi}{\partial x}$ and $\frac{\partial \phi}{\partial y}$ are approximated by the central difference,

and the temporal partial derivative $\frac{\partial \phi}{\partial t}$ is approximated by the forward difference. The

approximation of equation (3.11) by the above difference scheme can be simply written as:

$$\frac{\phi^{k+1}_{i,j} - \phi^{k}_{i,j}}{\Delta t} = L(\phi^{k}_{i,j}), \qquad (3.12)$$

where $L(\phi_{i,j})$ is the approximation of the right hand side by the above spatial difference

scheme. The difference equation (3.12) can be expressed as the following iteration:

$$\phi^{k+1}_{i,j} = \phi^{k}_{i,j} + \Delta t L(\phi^{k}_{i,j}). \qquad (3.13)$$

3.3 Results & Discussion

A set of experiments were conducted on several images from RGB-D Washington dataset (Lai *et al.*, 2014), synthetic stereo dataset (Scharstein & Szeliski, 2003; Scharstein & Szeliski, 2002), and our own real world dataset to show the effectiveness of the proposed ACM using depth clue. Our own real world images were captured by Microsoft Kinect. Microsoft Kinect for windows SDK version 1.5 was used for extracting the depth

40

information of the scene. The proposed ACM is compared to the gradient-based and polarity-based ACMs in (Li *et al*., 2010) and (Ksantini *et al*., 2013), respectively.

The results of each active-contour model on our real world RGB-D images are shown and compared together in Figure 3.1. The initial contour is set outside of the salient object for most images. In other words, the coefficient of the A_{F_D} i.e., v is set to the positive value which makes the initial contour shrink, whereas this coefficient (v) is set to the negative value making the initial contour expands during the curve evolution for image 5 and image 6 in Figure 3.1. Figure 3.2 indicates the comparison results of the proposed method and other baseline methods on images of several standard datasets.

As experimental results show, the proposed ACM using depth clue could detect the salient object more effectively in comparison to gradient-based and polarity-based ACMs presented in (Li *et al*., 2010) and (Ksantini *et al*., 2013), respectively. The main reason is that depth disparity provides the depth information of the salient object regardless of the noise or texture inside or outside of it. Considering depth information, objects are recognized from each other based on their distance to the camera. However, gradient and polarity present objects with a lot of noise. Thus, the curve stops before reaching the actual object's boundary.

41

a) Image 1:Initial Contour   b) (Li *et al*., 2010)   c) (Ksantini *et al*., 2013)   d) Proposed ACM with Depth Clue

e) Image 2:Initial Contour   f) (Li *et al*., 2010)   g) (Ksantini *et al*., 2013)   i) Proposed ACM with Depth Clue

j) Image 3:Initial Contour   k) (Li *et al*., 2010)   l) (Ksantini *et al*., 2013)   m) Proposed ACM with Depth Clue

n) Image 4:Initial Contour   o) (Li *et al*., 2010)   p) (Ksantini *et al*., 2013)   q) Proposed ACM with Depth Clue

r) Image 5:Initial Contour   s) (Li *et al*., 2010)   t) (Ksantini *et al*., 2013)   u) Proposed ACM with Depth Clue

v) Image 6:Initial Contour   w) (Li *et al*., 2010)   x) (Ksantini *et al*., 2013)   y) Proposed ACM with Depth Clue

Figure 3-1: Salient Object Detection Using ACM Based on Gradient, Polarity, and Depth Clue on Our Own Images.

42

a) Image 1:Initial Contour

b) (Li *et al*., 2010)

c) (Ksantini *et al*., 2013)

d) Proposed ACM with Depth Clue

e) Image 2:Initial Contour

f) (Li *et al*., 2010)

g) (Ksantini *et al*., 2013)

h) Proposed ACM with Depth Clue

i) Image 3:Initial Contour

j) (Li *et al*., 2010)

k) (Ksantini *et al*., 2013)

l) Proposed ACM with Depth Clue

m) Image 4:Initial Contour

n) (Li *et al*., 2010)

o) (Ksantini *et al*., 2013)

p) Proposed ACM with Depth Clue

q) Image 1:Initial Contour

r) (Li *et al*., 2010)

s) (Ksantini *et al*., 2013)

t) Proposed ACM with Depth Clue

Figure 3-2: Salient Object Detection Using ACM Based on Gradient, Polarity, and Depth Clue on Images of Several Standard Datasets.

3.4 Conclusion

In this study, an ACM based on depth information extracted by either Microsoft Kinect or stereo matching algorithm is proposed for detecting objects of interest in the image. The main significance of depth disparity is that it gets the depth information of the entire object regardless of noise or texture inside or outside of that object. Both gradient and polarity, which were previously applied as stopping terms in the external energy function, are sensitive to the noise, and the curve may stop before reaching the edge of objects. Therefore, depth clue is applied as a stopping function in the ACM for detecting the salient objects. Experiment on some standard datasets and several real world images indicated the superiority of depth disparity over polarity and gradient in detecting the object of interest using the ACM. Although, the use of depth clue has outperformed other clues for most objects, the gradient or polarity may perform better when the depth information becomes unreliable. In particular, Microsoft Kinect has its own limitation as it is unable to provide accurate depth information for very long distances. Hence, because of the wide variety of image data, our future work will be devoted to develop an ACM that will automatically select the best clue among the depth, gradient, polarity, or any combination of them. The selected clue will be used as the stopping term in the external energy function of the ACM.

44

CHAPTER 4

FEATURE BASED ACTIVE CONTOUR MODEL FOR OBJECT DETECTION

4.1 Background

Among the most recent ACMs based on the level set evolution, edge-based information like image gradient is utilized as an edge indicator in their external energy term to stop the curve on the object boundaries. In spite of obtaining promising results in some images using gradient-based ACMs, its sensitivity to texture makes it difficult stop before reaching the actual edge. Polarity has performed more effectively in the presence of texture, but it has failed to detect salient objects effectively in images with high level of texture. Depth clue can provide the scene information based on the distance of image objects to the camera, and hence, it is not vulnerable to the texture. However, depth clue is not always available for flat objects, and some objects might be located at the same distance or have the same depth information.

In fact, the best candidate for a stopping term could be either a single feature such as gradient, polarity and depth, or a combination of them, w.r.t. image characteristics. Therefore, we proposed a feature-based ACM that is able to automatically decide which subset of the features among gradient, polarity, and depth leads to the most effective result in detecting the salient object.

45

Figure 4.1 indicates the overview diagram of the proposed feature-based ACM. First, given an input image, the three features, namely gradient (G), polarity (P) and depth (D), are computed. Second, the texture ($I_{texture}$) is estimated based on image gradient, and if it is below a threshold, P is kept and G is ignored; otherwise, we ignore P and consider G as a more suitable feature. Then, the histogram of depth on the initial region is extracted and analyzed. If the number of peaks (non-zero bin) is greater than 1, it means that depth information is available and can be combined with another feature using KSVM. More details about feature selection and combination are provided in section 4.2.
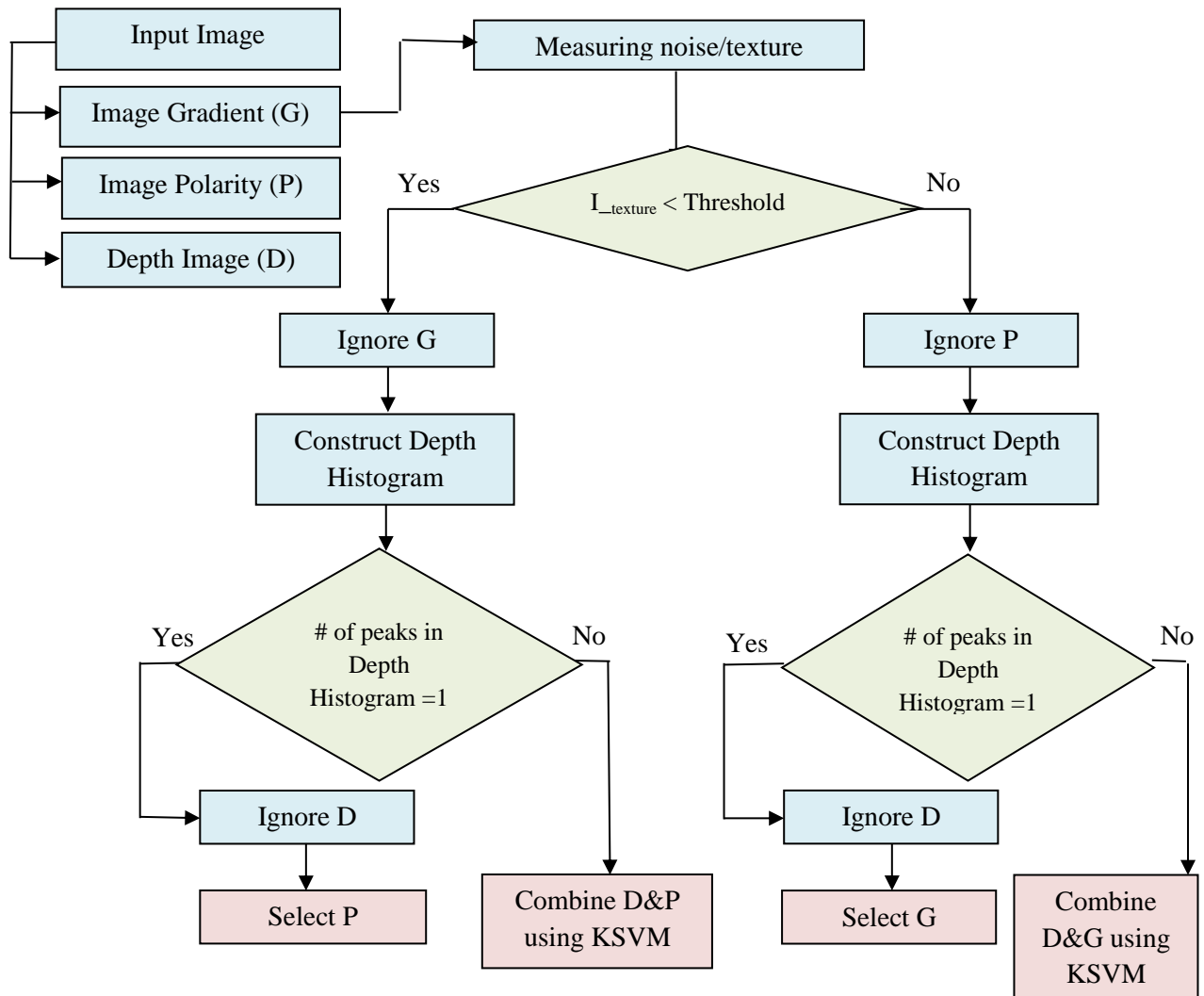
Figure 4-1: The Overview Diagram of The Proposed Feature-based ACM.

4.2 Method: The Proposed ACM based on Feature Selection and Feature Combination

In this section, the details of the proposed method for selecting the best candidate feature

as a stopping term in ACM is discussed. Then, in the case when more than one feature is

selected, we describe how two features can be combined, using KSVM, and be used in ACM.

4.2.1   Feature Selection

The proposed feature selection approach aims to find the best candidate features among gradient, polarity and depth. The number of edge pixels in a specified region helps determine a characteristic of texture complexity (Stockman & Shapiro, 2001; Haralick, 1973). Hence, the edge indicator F_g calculated using equation (2.8), can be used to measure the intensity of texture (or noise). For detecting a salient object using ACM, an initial contour surrounding the object is first defined. Then, the mean of $F_g(i)$ values in this initial region R is calculated as a measurement for texture and is given by:

$$I_{texture} = \frac{1}{N}\sum_{i\in R}(F_g(i)). \tag{4.1}$$

As mentioned earlier, the value of $F_g(i)$ is very close to 0 on edge areas. So when $I_{texture}$ is closer to 0, it means that the texture intensity is high. A threshold value should be defined to determine whether the gradient can be a good candidate for object detection, w.r.t. the texture intensity $I_{texture}$. For this purpose, we have defined the threshold value based on Otsu method (Otsu, 1979). The Otsu threshold value is computed on the values of $F_g(i)$ in the initial region R. Generally, Otsu threshold method aims to find the value which minimizes the variance of samples within the class and maximizes the variance

48

between two classes. Thus, when $I_{texture}$ is below Otsu threshold value (close to 0), then the background is too textured, making the gradient not suitable for detecting the object effectively. In this case, the gradient should be ignored, while polarity would be a better feature candidate.

As discussed earlier, the depth has its own limitations. For example, when objects are located at similar depths, it would be impossible to distinguish between them using the depth clue. In order to evaluate the usefulness of the depth information for a given image, we have analyzed the depth histogram of I_D within the region of interest. Such histogram provides significant information about the geometry of scene, and it was already shown that histograms are very useful. For example, in (Park *et al.*, 2012), the detection of the hand contour was achieved based on the depth histogram. In our case, the analysis of the histogram of depth in the initial region of interest gives us a good indication on whether the depth information of the salient object is beneficial. When the number of peaks (non-zero bins) in the depth histogram is equal to 1, the depth of the salient object is confused with the background, making the depth feature useless. On the other hand, when the number of peaks is greater than 1, depth is considered as a good candidate feature. It might be combined with either gradient or polarity using KSVM. This process of feature selection is described in the following algorithm:

49

Algorithm 1 Feature Selection
___
1: $Img \leftarrow$ Input Image
2: $F_g \leftarrow$ Edge indicator based on Gradient information of Img
3: $F_p \leftarrow$ Edge indicator based on Polarity information of Img
4: $F_D \leftarrow$ Edge indicator based on Depth information of Img
5: $I_{texture} \leftarrow$ Texture of Initial Region in Img
6: $Threshold \leftarrow$ Otsu Threshold of Img
7: $S_F \leftarrow$ Selected Feature(s)
8: **if** $I_{texture} < Threshold$ **then**
9:     $S_F \leftarrow \{F_p\}$
10: **else**
11:     $S_F \leftarrow \{F_g\}$
12: **end if**
13: Construct Depth Histogram
14: **if** Count of peaks in Depth Histogram > 1 **then**
    $S_F \leftarrow \{F_D\} \cup \{S_F\}$
15: **end if**
16: **return** S_F
___

4.2.2 Feature Combination based on KSVM

If more than one feature is selected as a stopping term, they should be combined using
KSVM, and the result of this combination is applied as an optimal edge indicator in the
external energy function of ACM.

Support Vector Machines (SVM) are a set of related supervised learning methods which
are applied for prediction and classification. The main aim of SVM is to build a set of
hyperplanes and select the hyperplane that maximizes the margin from the training
samples. Once such a hyperplane is discovered, SVM can effectively classify the testing
samples. The corresponding data samples are called Support Vectors (SVs). These points

50

are the crucial samples for classification. Therefore, since SVM considers only those data points which are close to the decision hyperplane and are critical to find the decision boundary, it is robust to noisy data and can properly handle the high-dimensional data. Note that SVM has been successfully applied in different domains and applications (Cheng *et al*., 2014; Ektefa *et al*., 2010; Guerbai *et al*., 2015).

Let $X_1 = \{x_i\}_{i=1}^{N_1}$ and $X_2 = \{x_i\}_{i=N_1+1}^{N_1+N_2}$ are two different classes of data samples constituting an input space of $N = N_1 + N_2$ samples, and the associated tags are represented by $\tau = \{t_i\}_{i=1}^{N}$, where $t_i \in \{0,1\}$, $\forall i = 1,2,...,N$. Since real-life data has inherent non-linearity, KSVM tries to map the data samples to a higher dimensional feature space F, where linear classification might be achieved. Let the two classes are mapped to higher dimensional feature classes $F_1 = \{\Phi(x_i)\}_{i=1}^{N_1}$ and $F_2 = \{\Phi(x_i)\}_{i=N_1+1}^{N}$ by the function Φ. Our target is to learn the weight vector $\boldsymbol{w} = (w_0, w_1, ..., w_N)$ for functions of the form:

$$y(x;w) = \sum_{i=1}^{N} f_i^x w_i + w_0 = \Phi^T(x_i)\boldsymbol{w} + w_0 \quad , \quad \text{where} \quad \Phi(x) = (f_1^x, f_2^x, ..., f_N^x) : \chi \to F$$

describes the non-linear mapping from the input space to the feature space for the input variable x (Cristianini & Shawe-Taylor, 2000). The *kernel trick* (Vapnik, 1998) is used in case when the dimension of F is very high, where a kernel function κ calculates the inner products of the higher dimensional data samples: $\kappa(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle, \forall_{i,j} \in \{1, 2, ..., N\}.$

51

In the feature space, KSVM tries to find the optimal decision hyperplane. The optimal hyperplane is the one with the largest margin; in other words, the plane which has the largest minimal distance from any of the samples. Maximizing the distance of samples to the optimal decision hyperplane is equivalent to minimizing the norm of w. As a result, this becomes part of the objective function. However, it might be the case that the problem is non-linear even in the higher dimensional space. To solve this, the margin constraint is relaxed or slacked. A penalty factor is also introduced in the objective function to control the amount of slack. This penalty factor is the form of a loss function, usually a hinge loss function (Vapnik, 1998). Incorporating all these, the KSVM optimization problem can be written as:

$$\min_{w \neq 0, w_0} \left\{ \frac{1}{2} w^T w + C \sum_{i=1}^{N} \max(0, 1 - t_i (\Phi^T(x_i) w + w_0)) \right\}. \qquad (4.2)$$

Here, $\max(0, 1 - t_i (\Phi^T(x_i) w + w_0))$ is the hinge loss function (Rosasco *et al.*, 2004). For correctly classified training samples, this function does not incur any loss. For misclassification, the loss factor is controlled by C. Since the weight vector w resides in the feature space, it cannot be calculated directly. Instead, the Lagrangian dual problem is solved (Scholkopf & Smola, 2001). The optimal weight vector for this problem is a linear combination of the data points and is of the form $w^* = \sum_{i=1}^{N} t_i \alpha_i^* \Phi(x_i)$, where $\{\alpha_i\}_{i=1}^{N}$ are the Lagrangian variables. The decision function for any test sample is obtained by:

$$g(x) = \sum_{i=1}^{N} t_i \alpha_i^* \kappa(x, x_i) + w_0^*. \qquad (4.3)$$

52

Where w_0^* is computed using the primal-dual relationship (Scholkopf & Smola, 2001), and where only samples with non-zero Lagrange multipliers α_i contribute to the solution.

Using KSVM, we aim to effectively predict the class label $t_i \in \{0,1\}$ for pixels of the target image where $t_i \in \{0\}$ corresponds to the class edge and $t_i \in \{1\}$ is for the class non-edge. Each pixel from set $X = \{x_i\}_{i=1}^N$ is a vector presented by features of S_F, where S_F includes the selected features or edge indicators determined by Algorithm 1. So the external energy function in our proposed ACM is denoted as F_{ext}, where $F_{ext} = g(x)$, computed based on the kernel-based prediction model using equation (4.3) and incorporated into the ACM (Li *et al.*, 2010) as below:

$$\varepsilon_{F_{ext},\lambda,v} = \mu R_p(\phi) + \lambda L_{F_{ext}}(\phi) + \alpha A_{F_{ext}}, \qquad (4.4)$$

where F_{ext} is the class labels assigned to the pixels of the image using the method described above. $R_p(\phi)$, $L_{F_{ext}}(\phi)$, and $A_{F_{ext}}(\phi)$ were defined in details in chapter 3.

4.2.3 *Providing Training Samples for KSVM*

Since KSVM is a supervised method, some training samples (pixels are samples in our case) should be provided to train the classifier. However, in our case the class labels of the pixel vectors are unknown. In other words, it is not known in advance if the pixel

vector belongs to the edge class ($t_i \in 0$) or to the non-edge class ($t_i \in 1$). In the absence of training samples, we proposed a method, based on the nearest-neighbor, to provide the training samples from the (unknown) input image itself.

Let $X = \{x_i\}_{i=1}^{N}$ includes all pixel vectors within the initial region of interest, with N being the total number of these pixels. We aim at selecting training samples from the set X and insert them into the edge training samples set $X_{edge} = \{x_i\}_{i=1}^{M_1}$ and the non-edge training samples set $X_{non-edge} = \{x_i\}_{i=M_1+1}^{M}$, respectively, where $X_{edge} \cap X_{non-edge} = \varnothing$.

For any pixel, the value of its gradient F_g, polarity F_p or depth F_D ranges from 0 to 1, with 0 representing a perfect edge and 1 a perfect non-edge, respectively. This information is used to provide the training samples as follows. First, for every pixel we calculate the Euclidean distance between its features values and the features values of the two ideal vectors, namely, the true edge (T_{edge}) (features values are all 0), and the true non-edge ($T_{non-edge}$) (features values are all 1).

Then, all pixel vectors are sorted based on their distance from T_{edge} and $T_{non-edge}$ vectors, and only 1% of nearest pixel vectors to these two ideal vectors are kept in the edge training set (X_{edge}) and in the non-edge training set ($X_{non-edge}$), respectively. For instance, Table 4.1 shows the values of the features for two pixels, and the last column shows the

calculated Euclidean scores. One can see that the first pixel (first row) is presented by gradient and depth features. This pixel is close to a true edge vector and might be a part of the training sample for the edge class, while another pixel (i.e. 20th pixel) is very close to the true non-edge vector and can be considered as a training sample for the non-edge class. Once these training samples are constructed, it becomes possible to classify the remaining pixels into either edge or non-edge classes, using KSVM.

Table 4-1: An Example of Providing Training Samples.

Pixels of The Image	Feature Value	True *edge* and *non-edge* Vectors	Euclidean Distance Score
$P_1 = \{F_g, F_D\}$	$P_1 = [0.1, 0.2]$	True edge [0.0]	**0.22**
		True non-edge [1,1]	1.2
$P_{20} = \{F_g, F_D\}$	$P_{20} = [0.75, 0.9]$	True edge [0.0]	1.17
		True non-edge [1,1]	**0.32**

4.3 Result and Discussion

A set of experiments was conducted to evaluate the effectiveness and accuracy of our object detection method. The proposed ACM-based method is tested on several datasets, consisting of Berkeley 3-D objects (B3DO) dataset (Janoch *et al.*, 2011), RGB-D Washington dataset (Lai *et al.*, 2014), synthetic stereo dataset (Scharstein & Szeliski, 2003; Scharstein & Szeliski, 2002), and our own outdoor stereo dataset. Then, we have

55

compared our results to some recent baseline ACMs (Zhang *et al.*, 2013; Zhang *et al.*, 2010a; Zhang *et al.*, 2010b).

There are several parameters for the above methods that are empirically tuned to get the best result for each image in our experiment. For example, in the method (Zhang *et al.*, 2010a), the initial contour is a circle whose center and radius should be set properly based on the size and location of the object in the test images. The number of iterations and the parameter σ, the standard deviation of the kernel function, should be properly chosen according to the image. This parameter is a scale parameter that controls the region-scalability from small neighborhood to the whole image domain (Li *et al.*, 2008). A too small σ may cause undesirable result, while a too large σ will cause high computational cost. In (Zhang *et al.*, 2010b), the number of iterations and the parameter α, which controls the shrinking or expanding the capacity of curve evolution, are tuned based on the test images. For the method in (Zhang *et al.*, 2013), with both GAC model and CV model, the initial region, number of iterations, Δt_1 and Δt_2 should be set. The RD-based level set evolution considers two time steps Δt_1 and Δt_2 to keep numerical stability, and they are set according to the noise level in the image. The values of all parameters for each method and dataset will be shown.

When implementing the ACM using our proposed method, some parameters like λ, μ, α and Δt (time-step) are set. Since the model is not sensitive to λ and μ, they are kept

56

fixed and set to $\lambda = 5$, $\mu = 0{:}04$ and $\Delta t = 5$ for all images in our experiments. The parameter α, the coefficient of $A_{F_{ext}}$ in equation (4.4), should be set depending on the image, and it is set to a positive value which makes the initial contour shrink. Note that a large value for α causes the contour to pass the objects with weak boundaries. If the proposed feature selection method chooses more than one feature, these features should be combined using KSVM. Hence, we have tested different kernel functions, like RBF and polynomial, and have found that the polynomial kernel $K(x, y) = (x^T y + c)^d$ is able to separate edge from non-edge pixels best. Gradient (G), Polarity (P), Depth (D), and their combination result in seven feature subsets: {G, P, D, GP, GD, PD, GPD}. Hence, all seven feature subsets are applied and experimented on the external energy function of the ACM for all input images to evaluate the validity of the proposed method on selecting the suitable feature subset.

The effectiveness of the proposed method and baseline methods are measured objectively in terms of the evaluation metric such as Dice similarity. Such a similarity measure considers the overlap between the segmented area and the ground truth. So it is computed as the intersection between the segmented area and the ground truth divided by the area of their union. Perfect segmentation results in a similarity score of 1, while such score is about zero if the segmented region and the ground truth are disjoint.

4.3.1   *Experiment with RGB-D Datasets*

Berkeley 3-D object (B3DO) dataset includes images taken in crowded domestic and office settings with Microsoft Kinect sensor. The sensor produces the RGB image in addition to per-pixel depth information. Although some areas of depth image may lack depth information, the B3DO dataset provides smoothed images with estimate of depth, when missing. RGB-D Washington is another dataset used in the experiment. This dataset (Lai *et al*., 2014) consists of 14 scenes containing furniture and a subset of the objects, with both RGB and depth images provided for each view. The chosen objects are commonly found in home and office environments.

Table 4.2 and Table 4.3 show how a particular feature subset is selected for each target image using our proposed feature selection method in both Berkeley and Washington RGB-D datasets, respectively. Moreover, the results of applying different feature subsets on the external energy function of the ACM using KSVM are indicated in Figure 4.2 and Figure 4.3 for each dataset. Ten images of Berkeley dataset and 6 scenes of Washington dataset are employed for the experiment (Figure 4.2 and Figure 4.3). The Dice similarity score is computed for each feature subset and shown below each image for each dataset in both Figure 4.2 and Figure 4.3.

58

Table 4-2: Feature Selection Process Based on Image Data for Berkeley Dataset.

Image ID	$I_{texture}$	Threshold	Number of Peaks in Depth Histogram	Selected Feature(s)
Image 43	0.42	0.44	3	{PD}
Image 170	0.71	0.54	2	{GD}
Image 205	0.59	0.46	4	{GD}
Image 294	0.20	0.37	4	{PD}
Image 344	0.70	0.45	1	{G}
Image 455	0.41	0.45	1	{P}
Image 800	0.18	0.38	1	{P}
Image 835	0.59	0.51	1	{G}
Image 841	0.33	0.40	2	{PD}
Image 705	0.46	0.43	2	{GD}

Table 4-3: Feature Selection Process Based on Image Data for Washington Dataset.

Image ID	$I_{texture}$	Threshold	Number of Peaks in Depth Histogram	Selected Feature(s)
Scene-1	0.27	0.45	2	{PD}
Scene-4	0.29	0.46	2	{PD}
Scene-7	0.45	0.42	3	{GD}
Scene-9	0.43	0.38	2	{GD}
Scene-12	0.47	0.38	2	{GD}
Scene-14	0.54	0.47	2	{GD}

As Table 4.2 indicates, the proposed method considers {PD} as the best feature subset, w.r.t. texture value $I_{texture}$ and depth histogram for image 43 in Berkeley dataset. Having experimented different feature subsets in the external energy function of the ACM (Figure. 4.2), the combination of polarity and depth ({PD}) using KSVM, with dice similarity score of 0.913, outperforms other feature subsets for image 43. This performance is also achieved for other images of Berkeley dataset.

As observed in Table 4.3, polarity and depth clue {PD} are selected as the best features for Scene-1 and Scene-4, while gradient and depth {GD} are selected for the other scenes in Washington dataset. The dice similarity scores shown under each image of Figure 4.3, for different feature subsets, also identifies the {PD} for Scene-1 and Scene-4, and {GD} for other scenes. This confirms one more time that our feature selection method is capable to automatically detect the right candidate features.

| **a) Initial 43** | **a)** D=0.735 | **a)** G=0.780 | **a)** P=0.885 |
| **a)** GD=0.849 | **a)** GP=0.841 | **a) PD=0.913** | **a)** GPD=0.863 |

www.manaraa.com

b) Initial 170 | **b)** D=0.312 | **b)** G=0.556 | **b)** P=0.271

b) GD=0.747 | **b)** GP=0.461 | **b)** PD=0.297 | **b)** GPD=0.421

c) Initial 205 | **c)** D=0.361 | **c)** G=0.702 | **c)** P=0.468

c) GD=0.797 | **c)** GP=0.646 | **c)** PD=0.517 | **c)** GPD=0.682

d) Initial 294 | **d)** D=0.821 | **d)** G=0.528 | **d)** P=0.854

d) GD=0.551 | **d)** GP=0.593 | **d)** PD=0.911 | **d)** GPD=0.755

61

e) Initial 344 | **e) D=0.205** | **e) G=0.924** | e) P=0.395

e) GD=0.622 | e) GP=0.458 | e) PD=0.582 | e) GPD=0.578

f) Initial 455 | **f) D=0.523** | **f) G=0.148** | **f) P=0.940**

f) GD=0.165 | f) GP=0.303 | f) PD=0.645 | f) GPD=0.196

g) Initial 800 | **g) D=0.683** | **g) G=0.456** | **g) P=0.885**

g) GD=0.458 | g) GP=0.459 | g) PD=0.755 | g) GPD=0.773

Figure 4-2: Applying Different Feature Subset as a Stopping Term in Berkeley Dataset.

a) Initial Scene-1   a) D=0.385   a) G=0.570   a) P=0.598

a) GD=0.675   a) GP=0.681   **a) PD=0.699**   a) GPD=0.582

b) Initial Scene-4   b) D=0.489   b) G=0.558   b) P=0.359
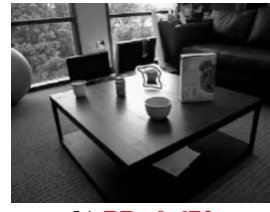
b) GD=0.652   b) GP=0.440   **b) PD=0.672**   b) GPD=0.670

c) Initial Scene-7   c) D=0.844   c) G=0.524   c) P=0.576
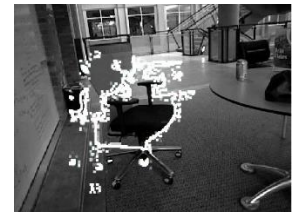
c) GD=0.854   c) GP=0.666   c) PD=0.838   c) GPD=0.782
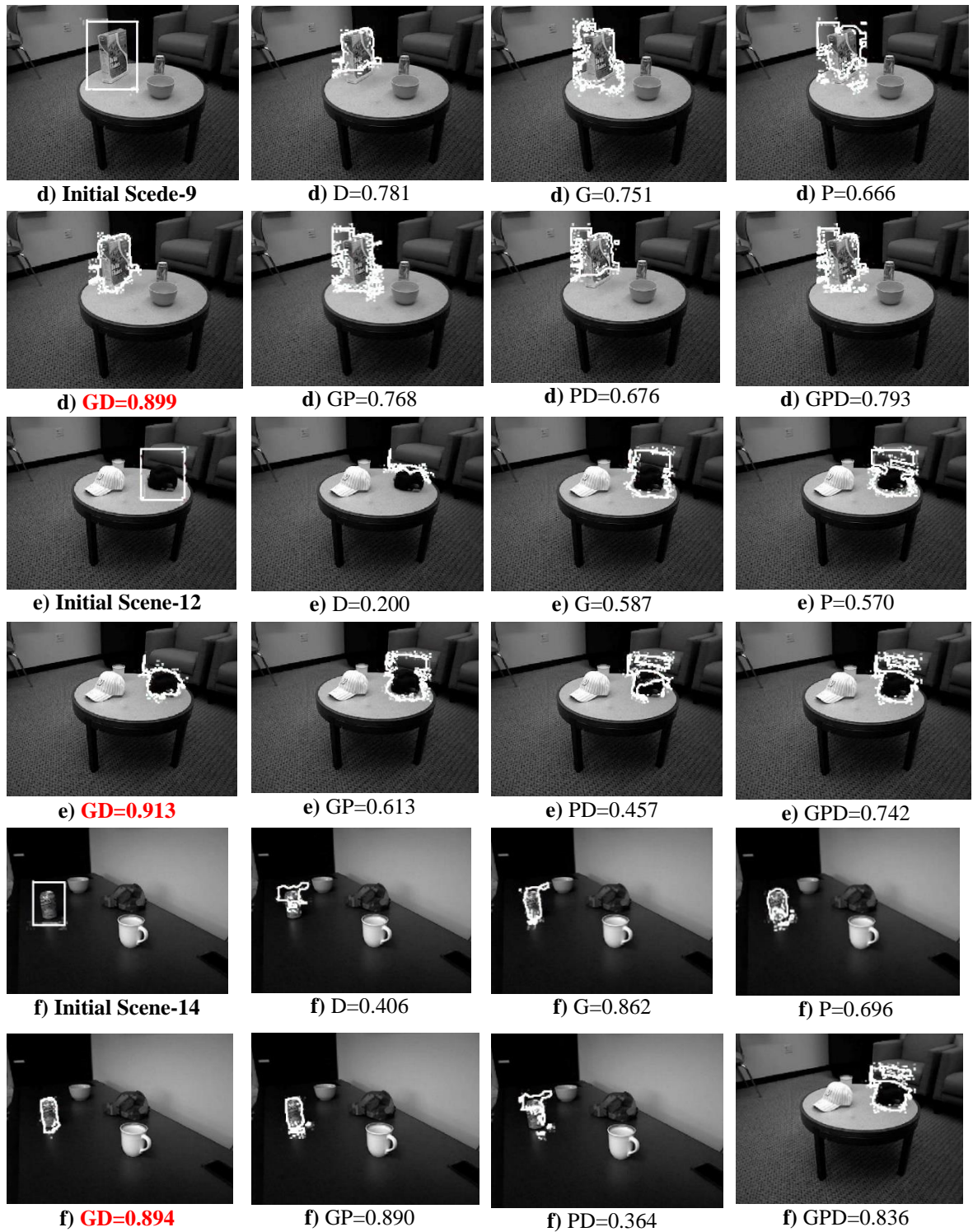
Figure 4-3: Applying Different Feature Subset as a Stopping Term in Washington Dataset.

The images in the figure are labeled as follows:

d) Initial Scede-9 | d) D=0.781 | d) G=0.751 | d) P=0.666

d) **GD=0.899** | d) GP=0.768 | d) PD=0.676 | d) GPD=0.793

e) Initial Scene-12 | e) D=0.200 | e) G=0.587 | e) P=0.570

e) **GD=0.913** | e) GP=0.613 | e) PD=0.457 | e) GPD=0.742

f) Initial Scene-14 | f) D=0.406 | f) G=0.862 | f) P=0.696

f) **GD=0.894** | f) GP=0.890 | f) PD=0.364 | f) GPD=0.836

Table 4.4 and Table 4.5 demonstrate the Dice similarity scores using the proposed feature-based ACM and other baseline methods (Zhang *et al.*, 2013; Zhang *et al.*, 2010a; Zhang *et al.*, 2010b) for Berkeley and Washington datasets, respectively. The last row in these tables shows the average similarity score over all the test images. As Table 4.4 and Table 4.5 suggest, the proposed ACM outperforms, and for some images, performs very close to other baseline methods in detecting the salient object. The method in (Zhang *et al.*, 2010b), with average scores of 0.247 and 0.429, performs very poorly with the Berkeley and Washington datasets. This region-based baseline method fails to detect the salient object properly because test images are clouted with texture. On the other hand, the main drawback with most region-based ACMs is that they are based on statistical information of the image and cannot achieve good performance in segmenting and detecting objects in images with high intensity inhomogeneity. Moreover, the initial contour is not set, and the method is supposed to do the global segmentation regardless of where the initial contour is. So the performance is quite low for images with high texture. Another region-based method tested in this study is (Zhang *et al.*, 2010a), which has shown better performance than the method in (Zhang *et al.*, 2010b), with similarity score of 0.778 for Berkeley dataset and 0.761 for Washington dataset. One of the main reasons is that the initial contour should be set. As mentioned earlier, the initial contour is a circle, and its center and radius is set considering the size and location of the object in the image. So this local segmentation leads to the better object detection than the global segmentation in (Zhang *et al.*, 2010b).

The RD-based ACM in (Zhang *et al.*, 2013) experimented with edge-based GAC model and region-based CV model. The RD-based with GAC model method yields better results than the region-based RD-CV model. As discussed earlier, region-based methods are sensitive to intensity inhomogeneity. The RD-based method with GAC has performed very close to our proposed method with average similarity score of 0.830 in Berkeley dataset. For example, RD-GAC method (Zhang *et al.*, 2013) on image 455, image 835, and image 841 of Berkeley dataset has shown very good performance by getting similarity score close to 1. However, this method fails to perform effectively for images with high texture. For instance, objects in image 294 and image 800 of Figure 4.2 were not detected effectively using RD-GAC method (Zhang *et al.*, 2013). The reason is that these images are clouted with texture, and RD-GAC, which has the gradient-based edge indicator, is not able to perform effectively. As mentioned before, polarity is a better feature in the presence of texture. Hence, as Figure 4.2 demonstrates, our proposed method is able to successfully select {PD} as an edge indicator in image 294 and image 800 considering texture level in the images, making object detection accomplished more effectively. Overall, our proposed ACM with average similarity score of 0.885 and 0.821 outperforms other baseline methods in Berkeley and Washington datasets, respectively.

Table 4-4: Comparing the Proposed Method with Baseline Methods Using Dice
Similarity Score in Berkeley Dataset.

Image ID	(Zhang et al., 2010b)	(Zhang et al., 2010a)	(Zhang et al., 2013) RD-GAC	(Zhang et al., 2013) RD-CV	Proposed Method
Image 43	0.00	0.888	**0.962**	0.963	0.913
Image 170	0.032	0.682	0.734	0.00	**0.747**
Image 205	0.092	0.756	0.766	0.044	**0.797**
Image 294	0.615	**0.938**	0.607	0.490	0.911
Image 344	0.882	0.913	0.869	0.763	**0.924**
Image 455	0.00	0.928	**0.977**	0.843	0.940
Image 800	0.215	0.794	0.505	0.689	**0.885**
Image 835	0.195	0.834	**1**	0.946	0.985
Image 841	0.00	0.277	**0.962**	0.072	0.807
Image 705	0.444	0.771	0.923	0.342	**0.941**
Average	0.247	0.778	0.830	0.515	**0.885**

Table 4-5: Comparing the Proposed Method with Baseline Methods Using Dice
Similarity Score in Washington Dataset.

Image ID	(Zhang et al., 2010b)	(Zhang et al., 2010a)	(Zhang et al., 2013) RD-GAC	(Zhang et al., 2013) RD-CV	Proposed Method
Scene-1	0.00	**0.794**	0.693	0.579	0.699
Scene-4	0.00	0.587	0.571	**0.697**	0.672
Scene-7	0.804	0.671	0.608	0.401	**0.854**
Scene-9	0.00	0.758	0.895	0.142	**0.899**
Scene-12	0.900	**0.957**	0.737	0.00	0.913
Scene-14	0.873	0.803	0.884	0.565	**0.894**
Average	0.429	0.761	0.731	0.397	**0.821**

68

Based on the above findings and depending on the input image, a single feature, like the gradient, polarity or depth, can be used as a stopping function in the ACM to effectively detect the salient object. Moreover, in some cases, a combination of such features may improve the result. Therefore, the proposed feature-based ACM in this study, that automatically selects the best suitable feature(s), outperforms previous ACMs that were limited to a single feature in their stopping functions.

Values of different parameters, tuned for each experimented method, are indicated in Table 4.6 and Table 4.7 for both Berkeley and Washington datasets. As these tables suggest, the proposed method is able to detect the salient objects more efficiently by going through less number of iterations. Although RD-GAC method (Zhang *et al.*, 2013) performs close to our proposed method, it has gone through about 300 iterations, while curves evolutions take maximum 260 iterations using our proposed method.

Table 4-6: Parameter Values for Each Image in Implementing the Proposed ACM and Other Baseline Methods in Berkeley Dataset.

Image ID	(Zhang et al., 2010b)		(Zhang et al., 2010a)		(Zhang et al., 2013) RD-GAC		(Zhang et al., 2013) RD-CV		Proposed Method	
	α	itr #	σ	itr #	Δt_1	itr #	Δt_1	itr #	α	itr #
Image 43	36	100	10	400	0.9	300	0.1	150	1	260
Image 170	80	200	10	600	0.8	300	0.3	150	1.5	260
Image 205	80	200	10	300	0.9	300	0.3	150	0.8	285
Image 294	80	200	10	600	0.7	300	0.4	300	1.3	235
Image 344	60	150	10	200	0.9	300	0.4	70	1.3	160
Image 455	100	200	10	200	0.9	300	0.1	450	1.3	235
Image 800	100	150	10	400	0.9	300	0.1	600	1	155
Image 835	100	150	10	600	0.9	300	0.1	100	1.5	235
Image 841	100	250	10	400	0.7	300	0.4	200	2.5	260
Image 705	60	100	10	600	0.9	300	0.4	200	6	260

Table 4-7: Parameter Values for Each Image in Implementing the Proposed ACM and Other Baseline Methods in Washington Dataset.

Image ID	(Zhang et al., 2010b)		(Zhang et al., 2010a)		(Zhang et al., 2013) RD-GAC		(Zhang et al., 2013) RD-CV		Proposed Method	
	α	itr #	σ	itr #	Δt_1	itr #	Δt_1	itr #	α	itr #
Scene-1	60	200	20	100	0.9	300	0.3	500	30	85
Scene-4	60	200	20	100	0.9	500	0.1	200	30	70
Scene-7	100	270	10	600	0.7	400	0.2	300	10	210
Scene-9	60	150	10	100	0.9	200	0.3	200	1.5	160
Scene-12	36	120	10	100	0.8	180	0.1	200	1.5	210
Scene-14	36	85	10	100	0.9	200	0.3	300	10	160

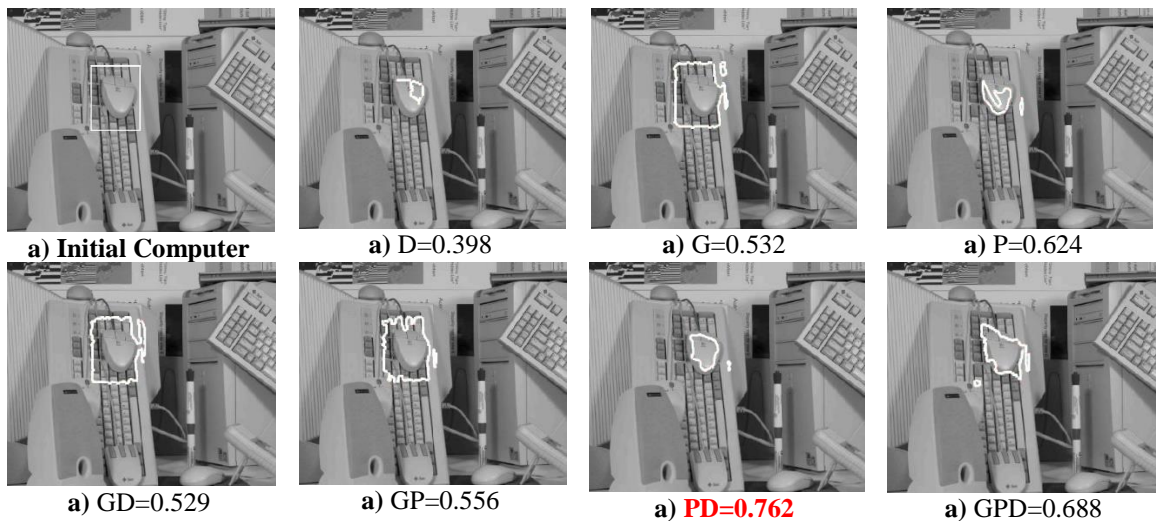### *4.3.2   Experiment with Stereo Indoor and Outdoor Images*

Table 4.8 presents the results when the proposed automatic feature selection method is applied on both indoor and outdoor stereo images, using disparity maps. Furthermore, different feature subsets are tested on the external energy function, and their results are shown in terms of dice similarity score under each image in Figure 4.4, to confirm the validity of the proposed method in selecting the best candidate feature(s) as the edge indicator.

The images with objects Car, Chair, and Flower are outdoor scenes while the other images are indoor scenes. For example, as Table 4.8 shows, because of the texture score of 0.22 for Image Computer (less than the threshold value), the gradient will be ignored. In addition, the depth histogram has 2 peaks, making the depth feature a reliable one. Hence, polarity and depth ({PD}) are selected as the good candidate features. They are combined, using KSVM, to be used in the external energy function. Figure 4.4 also confirms the accuracy of this automatic feature selection as it clearly indicates that {PD} outperforms the other feature subsets in detecting the salient object, with a dice similarity score of 0.762 in Image Computer. As Table 4.8 and Figure 4.4 suggest, applying our automatic feature selection method on other stereo images confirms once again that the best feature subsets can be selected for detecting the object of interest. $I_{texture}$ values for most stereo images and especially for outdoor scenes are close to zero, indicating the presence of high texture. So gradient cannot detect the object effectively, and polarity is

71

chosen as a better option. Depth clue, in case it is available, is not sensitive to texture and is considered as a helpful feature as it carries a different kind of information. Thus, it is combined with polarity using KSVM and applied in the ACM stopping term.

Table 4-8: Feature Selection Process Based on Image Data for Stereo Dataset.

Image ID	I$_{texture}$	Threshold	Number of Peaks in Depth Histogram	Selected Feature(s)
Computer	0.22	0.40	2	{PD}
Cone	0.40	0.29	2	{GD}
Laundry	0.24	0.40	2	{PD}
Monopoly	0.32	0.45	2	{PD}
Star	0.23	0.39	2	{PD}
Teddy	0.29	0.48	2	{PD}
Tsukuba	0.15	0.36	2	{PD}
Car	0.18	0.35	3	{PD}
Chair	0.20	0.35	3	{PD}
Flower	0.12	0.30	4	{PD}

a) Initial Computer   a) D=0.398   a) G=0.532   a) P=0.624

a) GD=0.529   a) GP=0.556   a) **PD=0.762**   a) GPD=0.688

72

b) Initial Cone   **b) D=0.457**   **b) G=0.663**   **b) P=0.191**

b) GD=0.879   **b) GP=0.699**   **b) PD=0.463**   **b) GPD=0.678**

c) Initial Laundry   **c) D=0.806**   **c) G=0.689**   **c) P=0.255**

c) GD=0.746   **c) GP=0.660**   **c) PD=0.923**   **c) GPD=0.837**
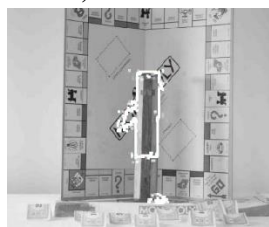
d) Initial Monopoly   **d) D=0.736**   **d) G=0.675**   **d) P=0.376**
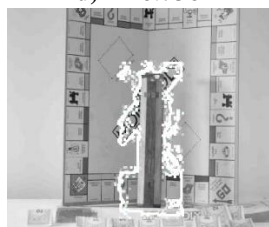
d) GD=0.868   **d) GP=0.671**   **d) PD=0.896**   **d) GPD=0.872**

73
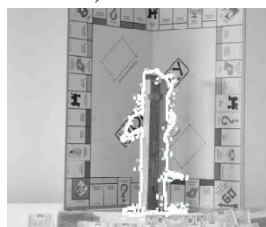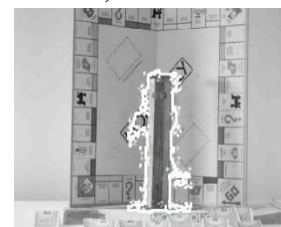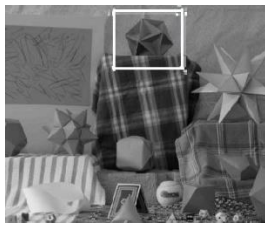
e) Initial Star | e) D=0.698 | e) G=0.622 | e) P=0.815

e) GD=0.927 | e) GP=0.727 | e) **PD=0.959** | e) GPD=0.857

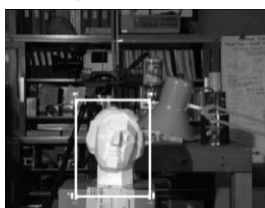**f) Initial Teddy** | f) D=0.920 | f) G=0.615 | f) P=0.396

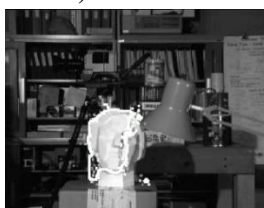f) GD=0.817 | f) GP=0.647 | f) **PD=0.924** | f) GPD=0.847

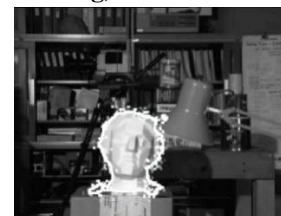**g) Initial Tsukuba** | g) D=0.688 | g) G=0.730 | g) P=0.322

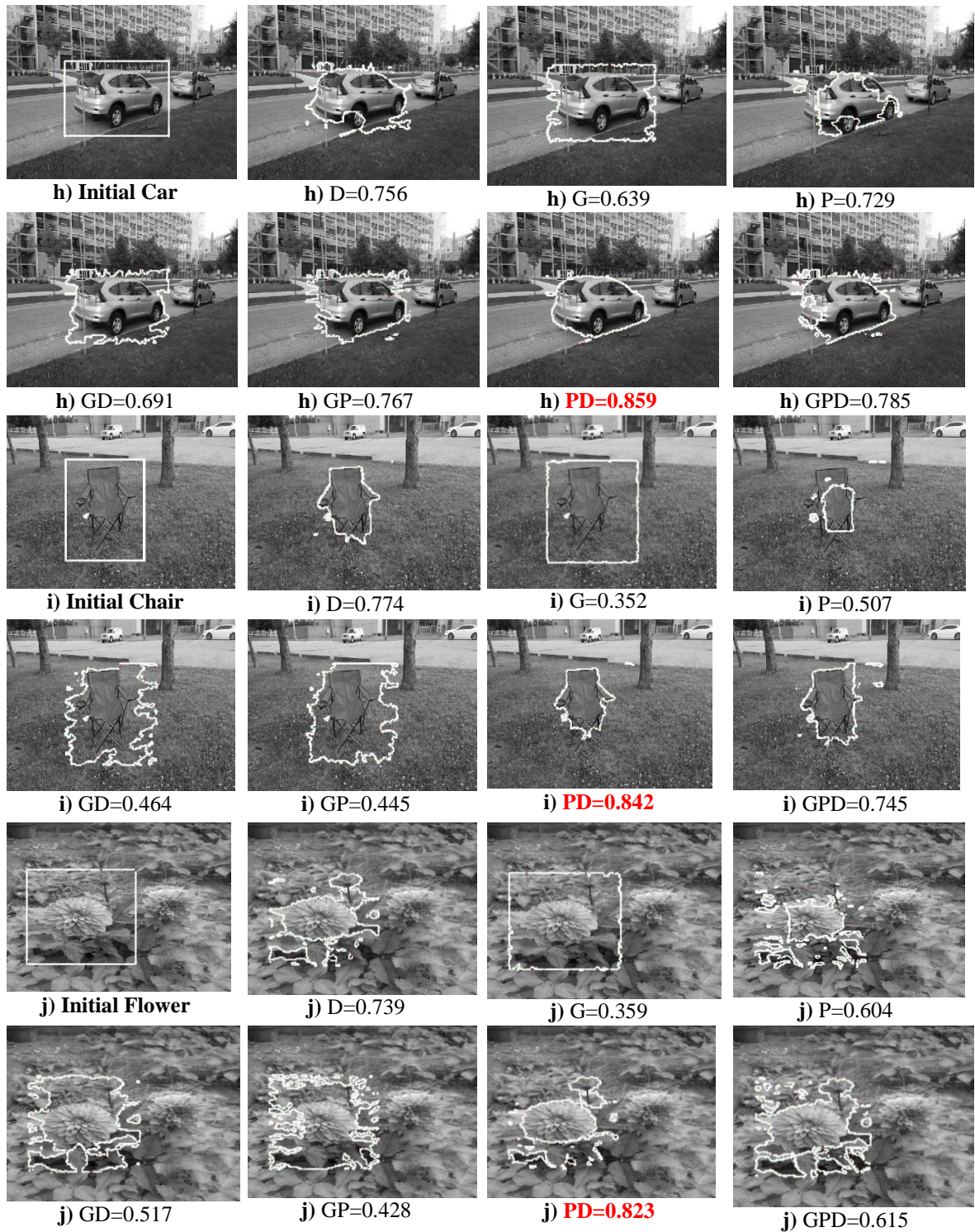g) GD=0.900 | g) GP=0.900 | g) **PD=0.949** | g) GPD=0.944

Figure 4-4: Applying Different Feature Subset as a Stopping Term in Stereo Images.

We have also tested other baseline methods on both indoor and outdoor stereo images and compared their detection results to our method. The parameters for each method were also tuned and shown in Table 4.10. As Table 4.9 shows, the similarity scores, obtained using the proposed ACM, are close to 1 for most images, outperforming all other methods. Thanks to disparity map, obtained from the stereo matching algorithm, the proposed method is able to detect the objects effectively in outdoor scenes. In spite of getting 0.783 average similarity score, the RD-GAC method in (Zhang *et al.*, 2013) failed to perform efficiently, as it takes a lot of time for curve evolution by going through 700 iterations.

Table 4-9: Comparing the Proposed Method with Baseline Methods Using Dice Similarity Score in Stereo Dataset.

Image ID	(Zhang *et al.*, 2010b)	(Zhang *et al.*, 2010a)	(Zhang *et al.*, 2013) RD-GAC	(Zhang *et al.*, 2013) RD-CV	Proposed Method
Computer	0.00	0.699	0.656	0.603	**0.688**
Cone	0.009	0.809	0.846	0.833	**0.879**
Laundry	0.00	0.703	0.779	0.068	**0.923**
Monopoly	0.776	0.404	0.743	0.207	**0.896**
Star	0.00	0.804	0.838	0.555	**0.959**
Teddy	0.052	0.789	0.810	0.808	**0.920**
Tsukuba	0.754	0.882	**0.966**	0.604	0.949
Car	0.391	0.661	0.825	0.513	**0.859**
Chair	0.519	0.669	0.744	0.405	**0.842**
Flower	0.310	0.631	0.627	0.328	**0.823**
Average	0.281	0.704	0.783	0.492	**0.873**

Table 4-10: Parameter Values for Each Image in Implementing the Proposed ACM and Other Baseline Methods in Stereo Dataset.

Image ID	(Zhang *et al.*, 2010b)		(Zhang *et al.*, 2010a)		(Zhang *et al.*, 2013) RD-GAC		(Zhang *et al.*, 2013) RD-CV		Proposed Method	
	α	itr #	σ	itr #	Δt_1	itr #	Δt_1	itr #	α	itr #
Computer	100	150	10	300	0.9	200	0.5	800	1.5	260
Cone	200	250	10	300	0.9	250	0.5	450	1.5	210
Laundry	60	200	7	600	0.9	500	0.5	200	1.5	210
Monopoly	80	250	6	1000	0.9	300	0.5	200	3	210
Star	80	250	10	100	0.9	200	0.5	200	1	260
Teddy	100	200	10	150	0.9	400	0.5	200	1.5	260
Tsukuba	100	85	10	300	0.9	500	0.5	200	20	210
Car	40	150	7	600	1	500	0.2	300	1	260
Chair	80	150	7	600	1	400	0.1	200	1.5	260
Flower	80	350	7	600	1.1	700	0.3	300	20	210

4.4 Conclusion

We have proposed a new method that automatically selects the best feature(s) to be used in the stopping function of the well-known ACM. Three features were considered, the gradient, the polarity and the depth. The proposed method has full flexibility w.r.t. the above-mentioned features, as the latter can be employed either individually or collectively in a coherent framework, depending on each individual image. We have also proposed a method to combine the selected features in a semi-supervised manner using the KSVM that will be the most effective stopping term in the ACM. To the best of our

77

knowledge, this work is the first to explore the selection and combination of several features to improve the detection of salient objects. We have successfully tested our method on Washington and Berkeley RGB-D datasets where depth is obtained using 3D sensors. More tests were also carried on indoor and outdoor stereo images, where the disparity map was used as depth clue. In particular, we were always able to select the right feature(s) to be used in the stopping term of the ACM depending on the image input, where the detection of the salient object was always correct. The proposed method has been also compared with several baseline existing ACM-based methods for object detection. Given the flexibility of our method to select and use from three features, the obtained results were superior for almost all the used input images.

CHAPTER 5

MULTIPLE OBJECT DETECTION WITH OCCLUSION USING ACTIVE CONTOUR

MODEL AND FUZZY C-MEAN

5.1 Background

Detection and segmentation of salient objects in static images is one of the fundamental research areas in computer vision. Occlusion in real world images brings a main challenge for object detection. When several objects overlap one another in the image, the problem of object detection becomes even more complicated. Some studies addressed the problem of detecting partially occluded object. However, such studies mainly focused on one class of object. For example, promising results have been obtained for various object classes including pedestrian (Liu *et al*., 2014; Negri *et al*., 2014), car (Zheng & Liang, 2009; Hota *et al*., 2010), and face (Ban *et al*., 2014; Jun & Kim, 2012). Such detection methods typically learn from object classifiers using a labelled training set. Given a testing set, the classifier is used on the sub-windows with different sizes at all positions. Although some studies (Rafiee *et al*., 2013; Hota *et al*., 2010) addressed the problem of segmentation and object detection without relying on object class, only one salient object has been the main concern. However, the presence of multiple occluding objects makes the detection even harder. The problem of partially occluded objects was tackled in the work of (Wu *et al*., 2008; Shet *et al*., 2007) using part-based representations and joint analysis of multiple objects. However, such techniques were employed on some specific moving objects, and their performance for very crowded

79

scenes is far from perfect. In this study, we aim to detect occluded unknown objects in complex static images.

Occluding objects are segmented within the same contour, when using an ACM-based segmentation technique explained earlier in Chapter 4. Hence, FCM algorithm is utilized to improve the object detection and handle the possible occlusions. However FCM algorithm has two major drawbacks.

First, cluster centers should be initialized randomly when using FCM. This issue may have negative effects on the outcome of the segmentation quality. To overcome this problem, some researchers proposed a number of efficient initialization algorithms (Bouguessa *et al.*, 2006; Capitaine, 2011). For example, the Fuzzy Maximum Likelihood Estimation (FMLE) was used as a clustering technique for classifying data points in (Bouguessa *et al.*, 2006) where, FCM algorithm was applied as an initialization first step for the FMLE algorithm. In (Capitaine, 2011), a centroid initialization method, called ordering split, was proposed, where all n-dimensional samples are converted to 1-dimension by getting the mean over the feature space. Then, all samples are uniformly split into C groups and C subsets of indices that are iteratively built. Finally, cluster centers are calculated. More details about this method can be found in (Capitaine, 2011). The second main issue about FCM is the number of clusters, the C value, which should be known in advance. Several validity indices have been proposed to determine the best number of clusters (Bezdek *et al.*, 1999; Bouguessa *et al.*, 2006). However, various

80

validity indices have yielded different cluster numbers that depend on the genre of the specific data. The reason is that most of these proposed methods are based on statistical information, and they are typically sensitive to noisy data, like real images, which might consist of texture and illumination.

Each image pixel is represented by a set of features in FCM clustering technique. The applied feature space can have a significant impact on the clustering quality. In (Siang Tan & Mat Isa, 2011), the histogram thresholding method was proposed to tackle the problem of cluster number and cluster centroid. However, color was the only feature used, making the method less robust. For instance, when multiple regions have a similar color but belong to different objects. It is therefore not sufficient to use color only. Other features, such as texture, shape and depth, should be explored as well. Another issue with (Siang Tan & Mat Isa, 2011) and other clustering methods is that they require high computational time to cluster pixels in high-resolution images. Authors in (Rafiee *et al*., 2013) clustered image blocks rather than image pixels to improve the efficiency of clustering techniques in segmentation. So the average feature values of each block are used in the clustering process. It was noted that the block size should not be too large to avoid decreasing the accuracy of the clustering. In our study, we followed the bock-wise FCM clustering to improve efficiency.

To address the above discussed issues, we proposed a new approach based on Fuzzy C-Mean (FCM) algorithm, which takes the advantage of the recently available Microsoft

Kinect sensor RGB-depth outputs, to solve the difficult problem of object segmentation in complex scenes. Although existing methods, such as the ones based on Active Contour Model (ACM), perform quite well in the single-object case and non-noisy environment, these methods fail when the scene consists of multiple occluding objects, with possibly similar colors. We aim to overcome this problem by using both the RGB and depth images of a scene, typically what Kinect sensor delivers as outputs.

5.2 Method: Salient Object Identification Using ACM and FCM

The proposed ACM based on automatic feature selection (described in details in Chapter 4) is able to effectively identify the salient contours when applied on the target image. However, when objects are occluded by each other, multiple objects might be segmented within a single contour. So we have used a FCM-based algorithm and disparity map to separate occluded objects.

Figure 5.1 illustrates the overview diagram of the proposed FCM-based method. It shows in particular how each stage tackles the challenges of detecting the multiple occluding objects in a complex scene. We first use the ACM to identify the main final contours, and then a FCM is used for clustering within the contours for possible further segmentation. In this later stage, the depth is included as an important clue that allows us to estimate the cluster number and to make the clustering process more robust. In

82

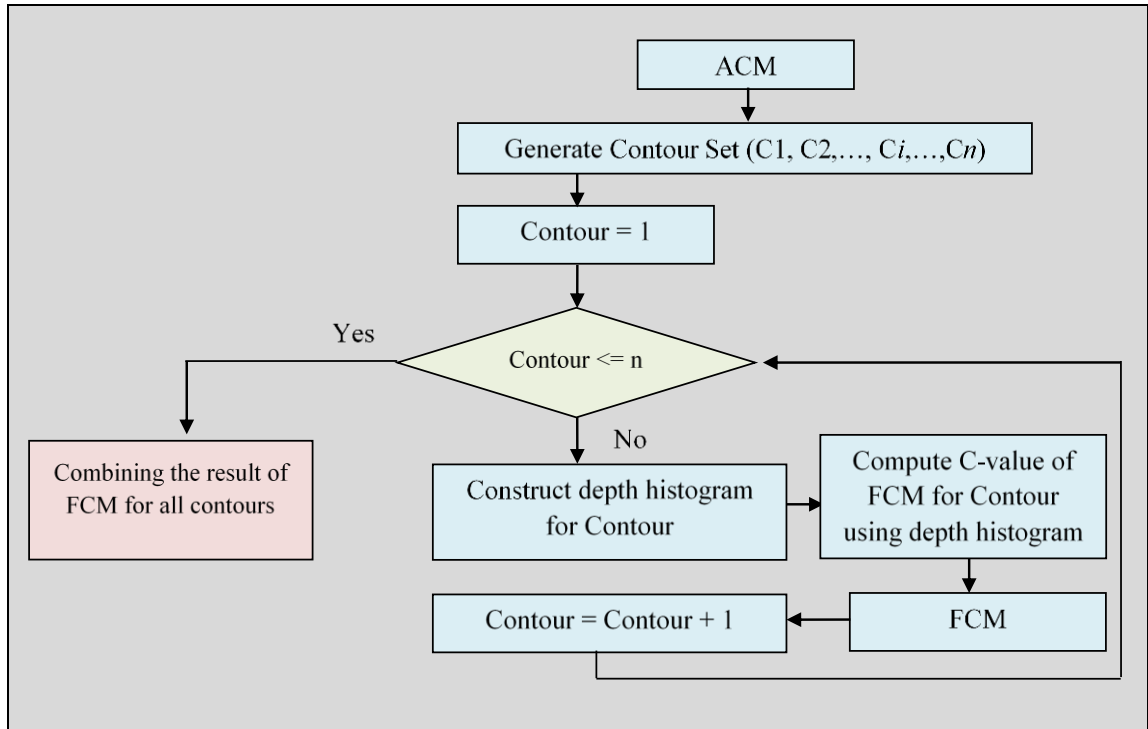particular, occlusions are easily handled this way, and the objects are properly segmented.

Figure 5-1: Overview Diagram of Proposed Method for Occluding Object Detection.

5.2.1   Determining Cluster Number

The proposed feature-based ACM yields a set of final contours. Given a contour, we should analyze the contents of this contour to detect possible occluded objects. Depth information can be used to help obtain the number of clusters for FCM algorithm. Both the color and depth images were obtained with the affordable Microsoft Kinect sensor. Note that such depth, even if it is not very accurate, it is beneficial for the clustering

method, as it provides approximate distances between objects and the sensor. In particular, depth histograms, well-known graphical tools for frequency distribution, were successfully experimented in (Park et al., 2012) for the detection of the hand contour. We have used depth histograms to determine the number of potential objects within the given contour.

Figure 5.2 depicts how a number of objects in each contour are extracted for an image. In this example, ACM results in three final contours. The area out of the final contours is the image background, to be discarded by setting its color to white. For each contour, the histogram of depth is constructed, and the number of peaks or non-zero bins is used as the number of clusters, or the C-value for the FCM algorithm. The C-value is at least two, as one peak belongs to the background and another peak corresponds to a contour, indicating the availability of one object in that contour. For example, the number of peaks for the middle contour in Figure 5.2 (h) is three, clearly showing the presence of two objects within this contour. The highest peak associates to the background (white color) in the histogram of each contour. Having found the cluster number for each contour, FCM is then employed on each contour to further segment and identify the objects. This process is done for all contours, and the different FCM results are combined to identify the total number of objects in the whole image (Figure 5.2 (c)).
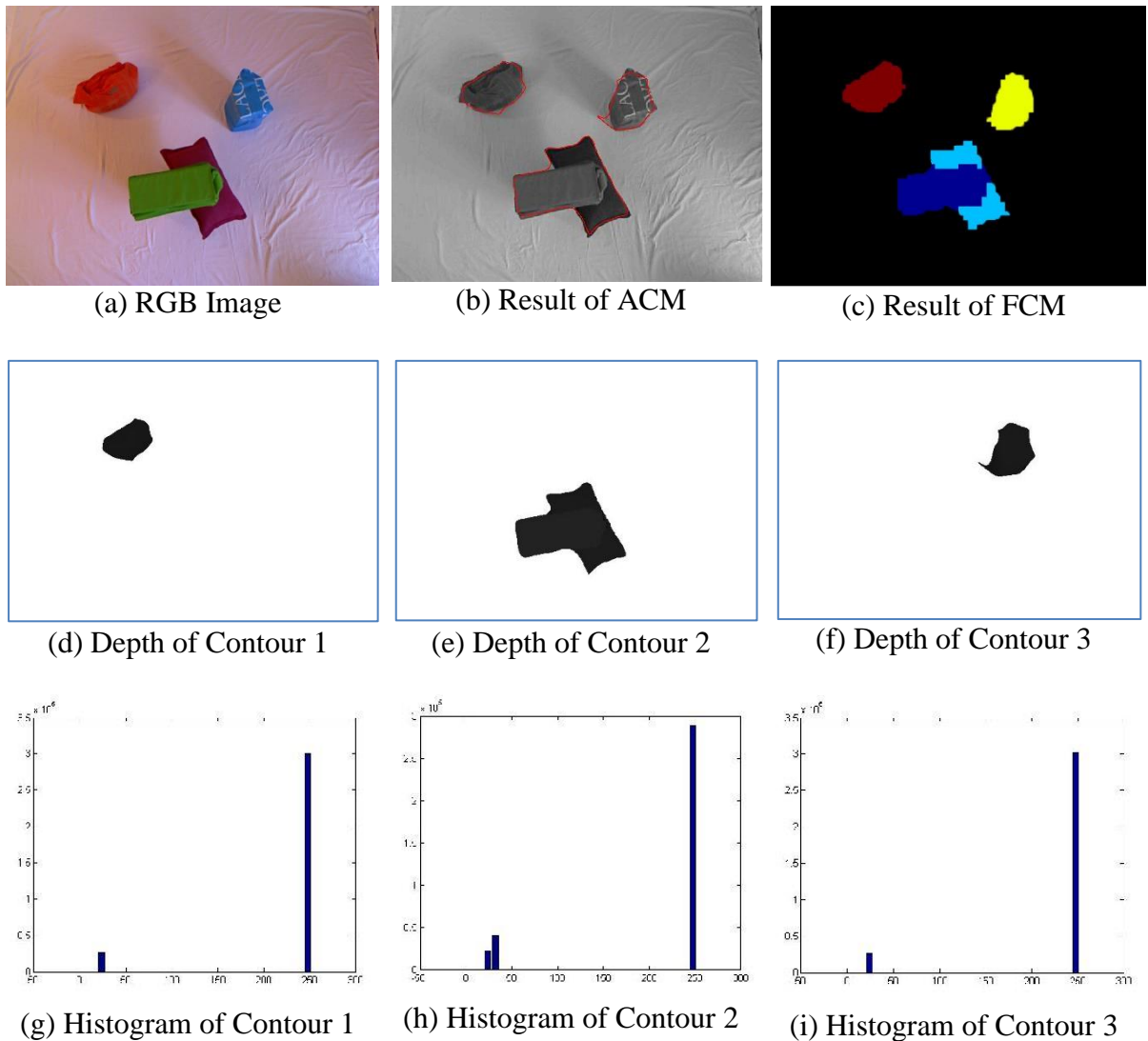
84

(a) RGB Image   (b) Result of ACM   (c) Result of FCM

(d) Depth of Contour 1   (e) Depth of Contour 2   (f) Depth of Contour 3

(g) Histogram of Contour 1   (h) Histogram of Contour 2   (i) Histogram of Contour 3

Figure 5-2: The Process of Determining the Cluster Number.

5.2.2   Occluding Object Detection

Once the cluster number for each contour is found, FCM can be used to explore the possible occlusions. Color (we have used the CIE-L*a*b* color space), texture and depth are the three features used in our FCM.

85

Energy and homogeneity are two texture features computed from the co-occurrence matrix. Co-occurrence matrix determines the frequency of a pixel with grey value i occurs adjacent to a pixel with grey value j in a specific orientation. We compute the co-occurrence matrix for four orientations $\theta \in \{0, 45, 90, 135\}$ and one distance value D $\in \{1\}$. Then the average of energy and homogeneity are calculated over the number of orientations and distances. The other feature we can use is the depth obtained from Microsoft Kinect or stereo matching algorithm. So the feature space is six-dimensional, including color (L, a, b), texture (Energy, Homogeneity) and depth. Because pixel clustering is computationally costly, especially for today high resolution images, we have followed the approach in (Rafiee *et al*., 2013), where block-wise clustering was used. In particular, the image is divided into several blocks, where the block size should cover approximately 17% of the maximum number of pixels in the height and width of the image. Then the average of all features values for each block is calculated.

Finally, FCM is applied to cluster the blocks instead of pixels. The occluded objects are hence identified by their corresponding contours. This process is done for all contours, and the different FCM results are combined to identify the number of objects in the whole image.

5.2.3   FCM Algorithm

The FCM algorithm has been a popular clustering technique in computer vision and pattern recognition. It was first presented by Dunn (1973) and developed by Bezdek (1981). FCM as a soft clustering technique measures the degree of membership of each data point to the particular cluster. Such membership values fall in the range of [0, 1] and exhibit the consistency of the relationship between the data point and the cluster. In this work, blocks are partitioned into C Fuzzy clusters. Let *N* be the total number of blocks in the image and *m* be the exponential weight of membership. The fuzzy data clustering is done w.r.t. the minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} U_{ij}^m . X_i, \tag{5.1}$$

where X_i is the distance of the *i*th block to the *j*th cluster center $S_j = \{s_1, s_2, \ldots, s_C\}$, and matrix $U = [u_{ij}]$ measures the membership degree between *i*th block to the *j*th cluster. To minimize the objective function, FCM algorithm should go through the following process:

1. Set the termination threshold $\varepsilon \in [0,1]$, iteration step $q = 0$

2. Initialize membership matrix $U = [u_{ij}], U^{(0)}$

3. At *q* iteration, calculate the cluster centroid $S^{(c)} = [S_j]$ with $U^{(q)}$

87

www.manaraa.com

$$S_j = \frac{\sum_{i=1}^{N} U_{ij}^m . X_i}{\sum_{i=1}^{N} U_{ij}^m}. \tag{5.2}$$

4. Update the membership function

$$U_{ij} = \frac{1}{\sum_{K=1}^{C} (\frac{\|x_i - S_j\|}{\|x_i - S_K\|})^{\frac{2}{m-1}}}, \tag{5.3}$$

where $1 \le j \le C$ and $1 \le i \le N$

5. If $\|U^{q+1} - U^q\| < \varepsilon$ then stop; otherwise, $q = q+1$ and go to step 3

5.3 Result and Discussion

The proposed FCM-based method has been tested on several real images, captured by Microsoft Kinect, on some images from the RGB-D Berkeley dataset, and on some real world stereo images where depth information is computed using the stereo vision algorithm (Yang, 2013). Note that because Kinect depth images are not registered to their corresponding RGB images, we have used a utility from OpenNI to approximately perform the depth-to-RGB registration.

In the first stage, our proposed ACM-based image segmentation method, using automatic feature selection, is applied to extract the salient and coherent regions from the target

88

image, yielding a set of contours. For all these experiments, the initial contour is set globally in the ACM. It goes through up to 500 iterations in Berkeley images and 430 iterations for our own images. Other parameters of ACM like λ, μ, and Δt (time step) are kept as default values and contours with the number of pixels less than a threshold are ignored. Objects occluded by each other within the same contour are discriminated and identified using the FCM algorithm. As discussed earlier, the depth histogram of each contour is taken into account to determine the C-value for the FCM algorithm.

Figure 5.3, Figure 5.4, and Figure 5.5 show the effectiveness and success of our proposed method in detecting the salient occluding objects in our real images, Berkeley dataset, and real world stereo images, respectively. The first column in these figures depicts the RGB images while the middle and third columns are the results of our ACM-based and FCM-based algorithms, respectively.

As Figure 5.3, Figure 5.4, and Figure 5.5 suggest, ACM was able to effectively identify the salient regions in the images. However, occluded objects were not segmented correctly. Hence, our FCM-based algorithm was applied on each contour to properly handle occlusions. The final results of FCM for all contours are shown in the last column of the figures. The number of objects for most scenes was correctly estimated, and the occluding objects were effectively segmented. However, in some cases, small objects with the same color and texture were grouped in the same cluster.

89

We have also tested the method in (Bouguessa *et al.*, 2006) on our images and compared it to our proposed method. Fuzzy Maximum Likelihood Estimation (FMLE) was used as a clustering technique for classifying data points in (Bouguessa *et al.*, 2006) where SC (compactness and separation) and PC (Partition Coefficient) are used to validate the cluster number, i.e., the number of objects in the image. However, as Figure 5.6 shows, such validity indices may get different values for the same image due to texture and outliers. The reason is that most of these proposed methods are based on statistical information, and they are typically sensitive to noisy data, like real images. Moreover, since the background is not subtracted, objects cannot be properly detected and separated from each other. Overall, our proposed method has outperformed (Bouguessa *et al.*, 2006) and was able to effectively and accurately detect all objects in a given image. The use of the depth information in the clustering process was very helpful to accurately obtain the C-value for the FCM algorithm. In particular, objects which are occluded by each other are properly detected using our method.
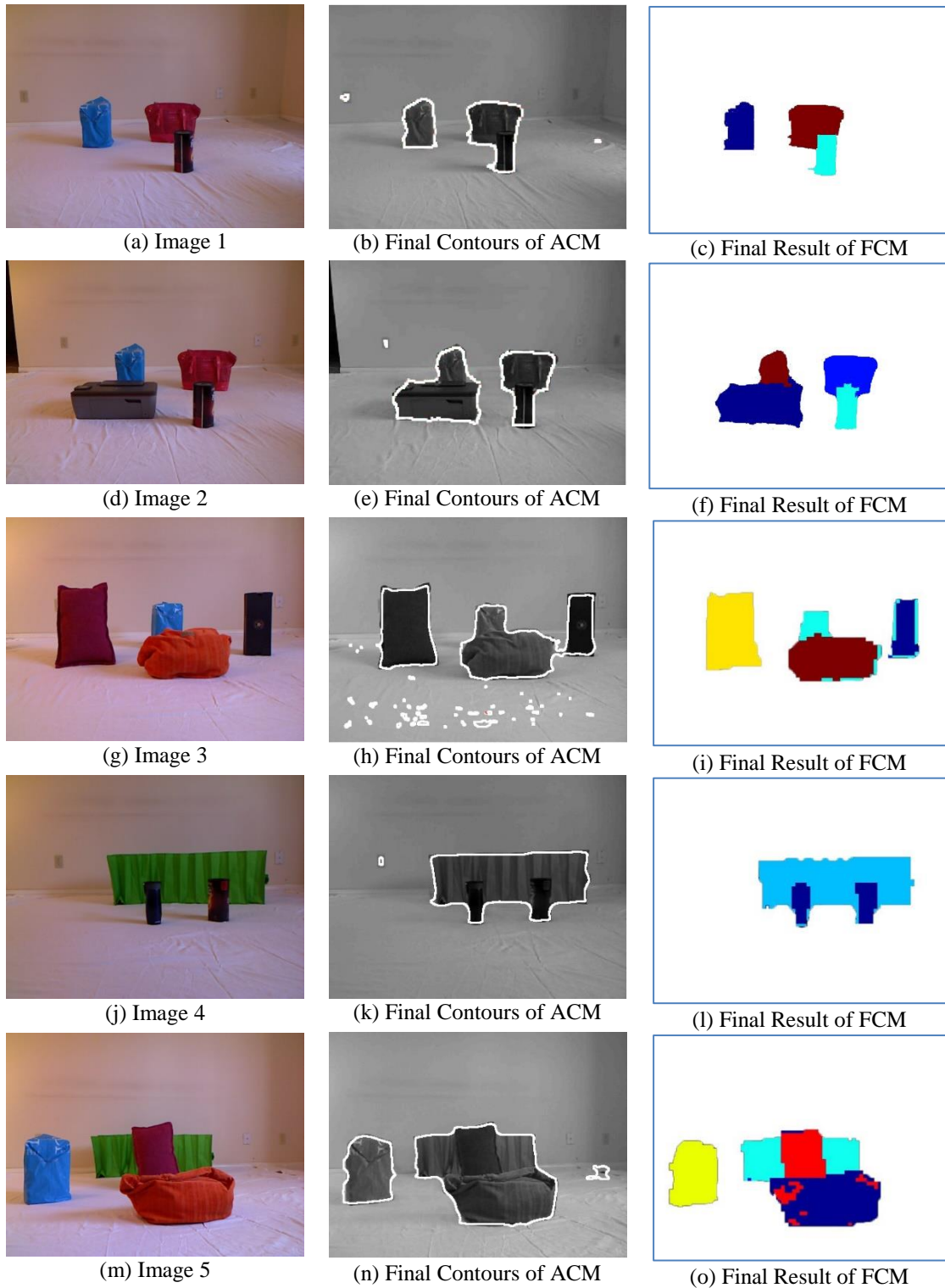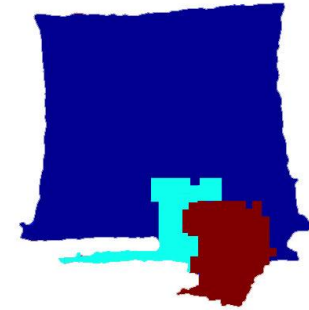
(a) Image 1   (b) Final Contours of ACM   (c) Final Result of FCM

(d) Image 2   (e) Final Contours of ACM   (f) Final Result of FCM

(g) Image 3   (h) Final Contours of ACM   (i) Final Result of FCM

(j) Image 4   (k) Final Contours of ACM   (l) Final Result of FCM

(m) Image 5   (n) Final Contours of ACM   (o) Final Result of FCM

Figure 5-3: Results of Applying FCM-based Method on Real RGB-D Images.

(a) Image 460   (b) Final Contours of ACM   (c) Final Result of FCM
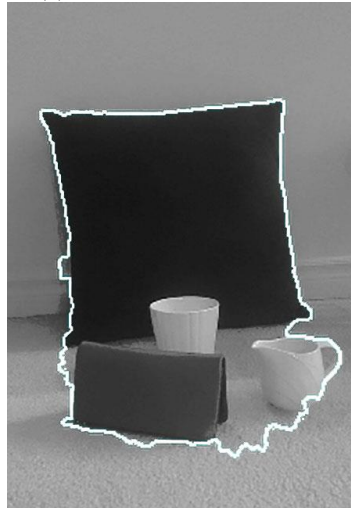
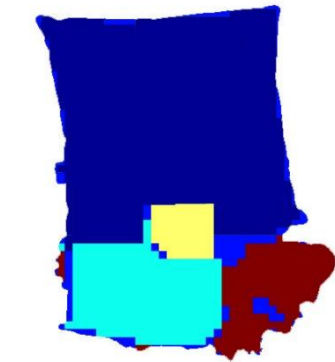(d) Image 50   (e) Final Contours of ACM   (f) Final Result of FCM

(g) Image 3   (h) Final Contours of ACM   (i) Final Result of FCM

(j) Image 145   (k) Final Contours of ACM   (l) Final Result of FCM

(m) Image 146   (n) Final Contours of ACM   (o) Final Result of FCM

Figure 5-4: Results of Applying FCM-based Method on Berkeley Dataset.

(a) Image 1   (b) Final Contours of ACM   (c) Final Result of FCM

(d) Image 2   (e) Final Contours of ACM   (f) Final Result of FCM
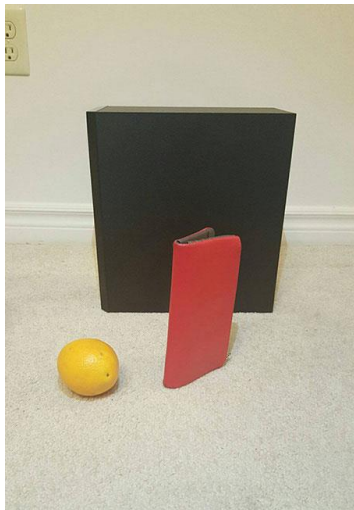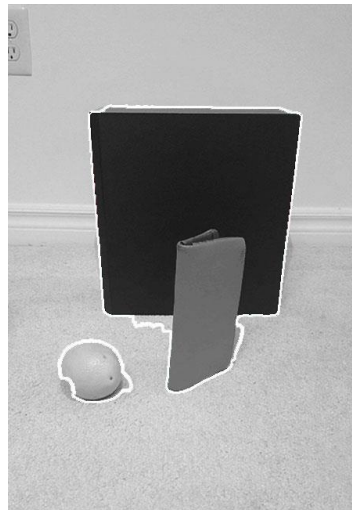
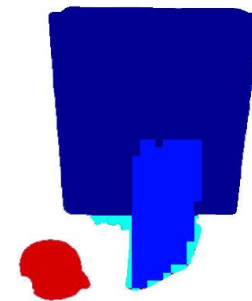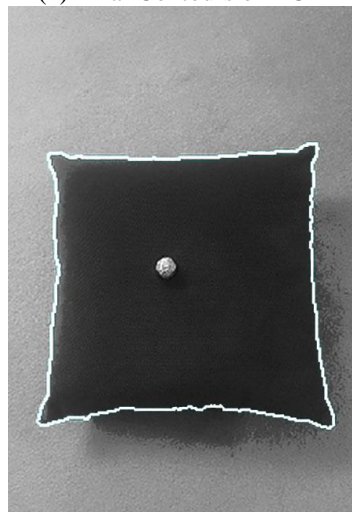(g) Image 3   (h) Final Contours of ACM   (i) Final Result of FCM
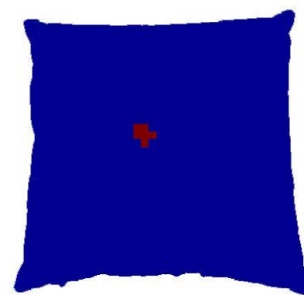
93

(j) Image 4   (k) Final Contours of ACM   (l) Final Result of FCM

(m) Image 5   (n) Final Contours of ACM   (o) Final Result of FCM
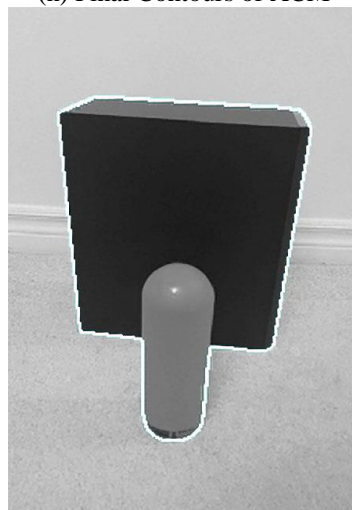
(p) Image 6   (q) Final Contours of ACM   (r) Final Result of FCM
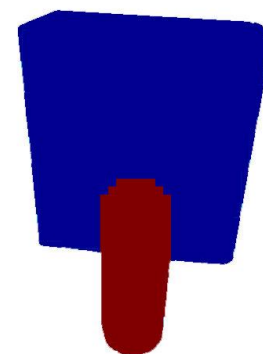
Figure 05-5: Results of Applying FCM-based Method on Stereo Images.

94

(a) Image 1   (b) PC=2   (c) SC=4

(d) Image 2   (e) PC=3   (f) SC=3

(g) Image 3   (h) PC=4   (i) SC=4

(j) Image 4   (k) PC=2   (l) SC=5

(m) Image 5   (n) PC=3   (o) SC=4

Figure 05-6: Results of Method (Bouguessa et al., 2006) Using SC and PC Validity Indices.

5.4 Conclusion

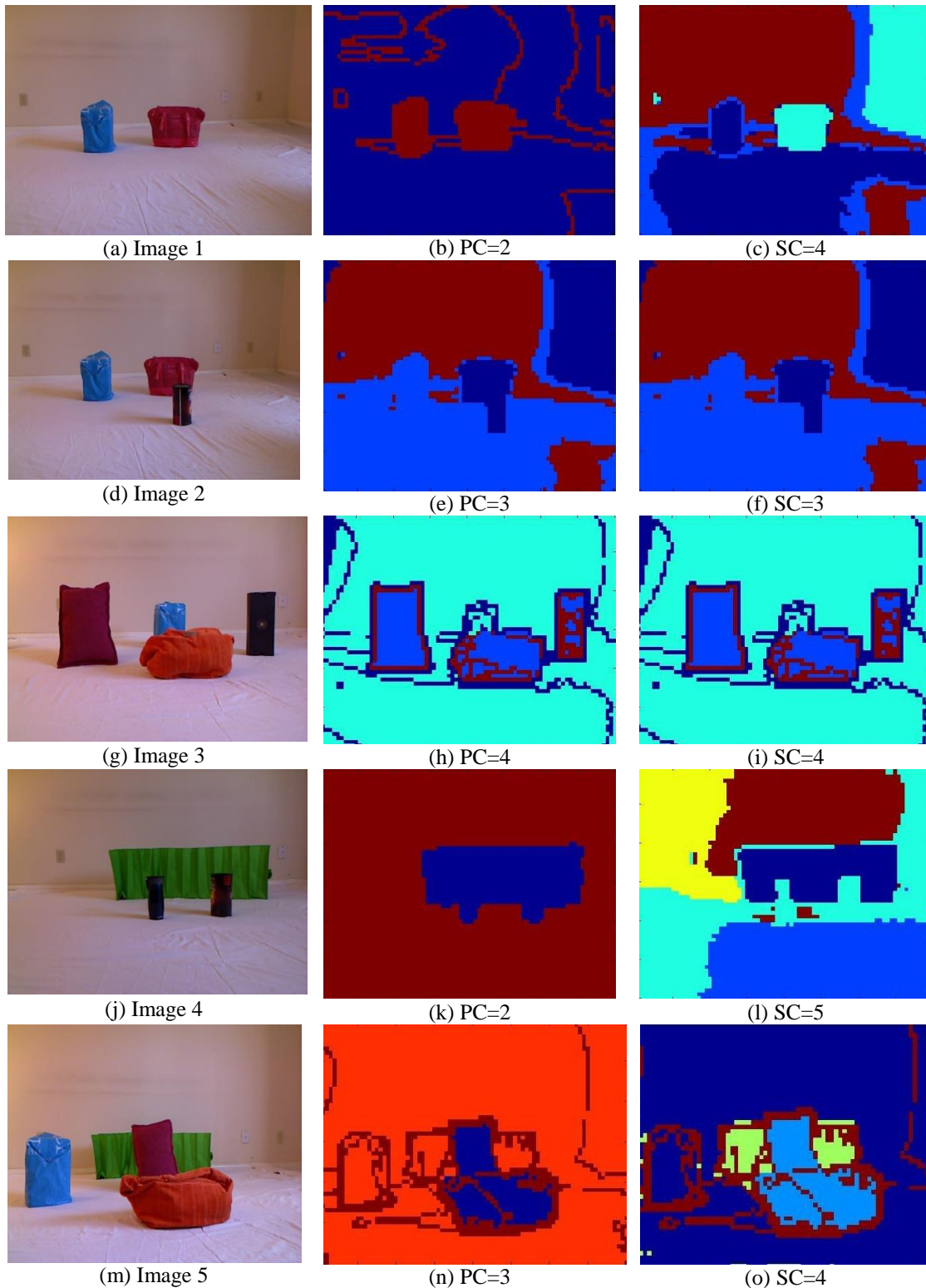We tackled the challenging task of detecting multiple occluding objects in complex scenes in this study. In addition to the improved performance of the classical ACM method through the proposed automatic feature selection for the stopping function, the use of depth histogram has made it possible to get an accurate cluster number. The latter is used in the final fuzzy clustering step, yielding accurate object detection. Our method outperformed some existing techniques that use validity indices for specifying the best cluster number. Experimental results obtained on the Berkeley B3DO dataset, our own RGB-D dataset, and real world stereo images have demonstrated that the proposed method performed very well for real complex images. However, in some challenging situations when different objects have the same color and texture, they might be assigned to the same cluster. This could be resolved by adding the spatial information as another feature to help clustering. Our future work will explore the use of pixel continuity for resolving this natural clustering ambiguity.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1. Conclusion

This thesis proposed novel segmentation and object detection methods for the real world image understanding. The proposed methods in this study are able to do the important tasks of segmentation and object detection in an unsupervised manner and perform effectively on our dataset and real world datasets from different sources. The main contribution of this study can be summarized as follow:

- Introducing and employing the depth clue, obtained from Microsoft Kinect or stereo vision algorithm, in the stopping function of the well-known edge-based ACM for detecting the salient object in the target image. The main significance of depth disparity is that it gets the depth information of the entire object regardless of noise or texture inside or outside of that object. Thus, this clue is able to perform more effectively than polarity and gradient in detecting the object of interest using the ACM. (Chapter 3)

- Proposing the novel feature-based ACM that is able to automatically select the best candidate features among gradient, polarity, and depth w.r.t. image characteristics. We also proposed the semi-supervised method based on KSVM for combining the selected features and classifying them to the class edge or non-

97

edge in the absence of class label. In this stage, the training samples are provided for the classifier based on nearest neighbor technique. Using the kernel function, pixels of the real world images, which have inherent non-linearity, can be mapped to higher dimensional feature space and hence accurately classified to their corresponding class. To the best of our knowledge, this work is the first to explore the selection and combination of several features to improve the detection of salient objects, and the obtained results were superior for almost all the used input images on well-known ACMs. (Chapter 4)

- Finally, the problem of detecting multiple occluding objects in the complex scenes was addressed in this study by proposing a method based on fuzzy clustering algorithm and disparity map. Thanks to disparity map, unlike existing techniques where the number of clusters is typically set manually, the proposed method is able to automatically estimate the cluster number. Thus, the occluded objects, segmented in the same contour using the proposed feature-based ACM, can be identified and separated from each other. Experimental results on Berkeley dataset and our own RGB-D dataset indicated the success and effectiveness of the proposed method in identifying the correct number of objects in the target images. (Chapter 5)

6.2. Future Work

This work can be improved from different perspectives. The possible future works are outlined as follows:

- In edge-based ACMs, the object is surrounded by the initial contour. The initial contour is determined manually in most studies. The possible future work can pertain to automatically determine the initial contour w.r.t. the size and location of the salient object in the image. So new internal energy function can be identified for estimating the coordinates of the initial contour in curve evolution.

- Several parameters should be tuned empirically in implementing most ACMs w.r.t. image data. Most of such parameters are tuned depending on the noise or texture level in the test image. So such parameters can automatically fall in a specific range considering the statistical information of the image.

- Analyzing the depth information is beneficial as it can help to obtain the cluster number for FCM technique. However, in addition to depth, exploring more feature like color or texture may result in more accurate cluster number and improve the result of clustering.

- The outcome of the proposed ACM are salient regions in the target image, and some irrelevant information, i.e., background is subtracted. This issue could help

99

the FCM method to more effectively handle the occlusion in those salient regions. However, the cluster centers are initialized randomly which might diminish the performance of the FCM. So centroid initialization could be another suggested future work for improving the result of the clustering method.

- In some challenging situations when different objects have the same color and texture, they might be assigned to the same cluster. This could be resolved by adding the spatial information as another feature to help clustering. Another future work could be exploring the use of pixel continuity for resolving this natural clustering ambiguity.

- Another possible future work might be recognizing the class label of the detected object. Our proposed method is able to effectively detect the unknown objects in the static image. However, a certain class of the detected object (such as humans, buildings, or cars) has not been determined in the test images yet.

100

REFERENCES/BIBLIOGRAPHY

Allili, M. S., & Ziou, D. (2007). Globally adaptive region information for automatic color–texture image segmentation. *Pattern Recognition Letters*, *28*(15), 1946–1956. http://doi.org/10.1016/j.patrec.2007.05.002

Banks, J., Bennamoun, M., Kubik, K., & Corke, P. (1999). A hybrid stereo matching algorithm incorporating the rank constraint. In *Image Processing And Its Applications, 1999. Seventh International Conference on (Conf. Publ. No. 465)* (Vol. 1, pp. 33–37 vol.1). http://doi.org/10.1049/cp:19990276

Ban, Y., Kim, S.-K., Kim, S., Toh, K.-A., & Lee, S. (2014). Face detection based on skin color likelihood. *Pattern Recognition*, *47*(4), 1573–1585. http://doi.org/10.1016/j.patcog.2013.11.005

Barnachon, M., Bouakaz, S., Boufama, B., & Guillou, E. (2014). Ongoing human action recognition with motion capture. *Pattern Recognition*, *47*(1), 238–247. http://doi.org/10.1016/j.patcog.2013.06.020

Beare, R. (2006). A locally constrained watershed transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(7), 1063–1074. http://doi.org/10.1109/TPAMI.2006.132

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers.

Bezdek, J. C., Keller, J., Krisnapuram, R., & Pal, N. R. (1999). *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing* (Vol. 4). Boston, MA: Springer US. Retrieved from http://link.springer.com/10.1007/b106267

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Bresson, X., Vandergheynst, P., & Thiran, J.-P. (2006). A Variational Model for Object Segmentation Using Boundary Information and Shape Prior Driven by the Mumford-Shah Functional. *International Journal of Computer Vision*, *68*(2), 145–162. http://doi.org/10.1007/s11263-006-6658-x

Bouguessa, M., Wang, S., & Sun, H. (2006). An objective approach to cluster validation. *Pattern Recognition Letters*, *27*(13), 1419–1430. http://doi.org/10.1016/j.patrec.2006.01.015

Burgin, W., Pantofaru, C., & Smart, W. D. (2011). Using depth information to improve face detection. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 119–120).

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-8*(6), 679–698. http://doi.org/10.1109/TPAMI.1986.4767851

Caselles, V., Catté, F., Coll, T., & Dibos, F. (1993). A geometric model for active contours in image processing. *Numerische Mathematik*, *66*(1), 1–31. http://doi.org/10.1007/BF01385685

Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic Active Contours. *International Journal of Computer Vision*, *22*(1), 61–79. http://doi.org/10.1023/A:1007979827043

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, *10*(2), 266–277. http://doi.org/10.1109/83.902291

Cheng, M.-M., Mitra, N. J., Huang, X., & Hu, S.-M. (2014). SalientShape: Group Saliency in Image Collections. *Vis. Comput.*, *30*(4), 443–453. http://doi.org/10.1007/s00371-013-0867-4

Chen, Y., Tagare, H. D., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K. S., … Geiser, E. A. (2002). Using Prior Shapes in Geometric Active Contours in a Variational Framework. *International Journal of Computer Vision*, *50*(3), 315–328. http://doi.org/10.1023/A:1020878408985

Crandall, R. (2009). Image segmentation using the Chan-Vese algorithm, Project report from ECE 532

Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press.

Dey, T. K., Giesen, J., & Goswami, S. (2003). Shape Segmentation and Matching with Flow Discretization. In F. Dehne, J.-R. Sack, & M. Smid (Eds.), *Algorithms and Data Structures* (pp. 25–36). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-45078-8_3

Duda, R. O., & Hart, P. E. (1972). Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Commun. ACM*, *15*(1), 11–15. http://doi.org/10.1145/361237.361242

Ektefa, M., Memar, S., Sidi, F., & Affendey, L. S. (2010). Intrusion detection using data mining techniques. In *2010 International Conference on Information Retrieval Knowledge Management, (CAMP)* (pp. 200–203). http://doi.org/10.1109/INFRKM.2010.5466919

Evans, L. C. (2010). *Partial Differential Equations*. American Mathematical Soc.

Gath, I., Geva, A.B., (1989). Unsupervised Optimal Fuzzy Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7), 773-781.

Ge, Q., Li, C., Shao, W., & Li, H. (2015). A hybrid active contour model with structured feature for image segmentation. *Signal Processing*, *108*, 147–158. http://doi.org/10.1016/j.sigpro.2014.09.007

Gonzalez, R. C., & Woods, R. E. (2006). Digital Image Processing (3rd Edition). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Guerbai, Y., Chibani, Y., & Hadjadji, B. (2015). The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. *Pattern Recognition*, *48*(1), 103–113. http://doi.org/10.1016/j.patcog.2014.07.016

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-3*(6), 610–621. http://doi.org/10.1109/TSMC.1973.4309314

Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, *28*(1), 100–108. http://doi.org/10.2307/2346830

Hart, P. E. (2009). How the Hough transform was invented [DSP History]. *IEEE Signal Processing Magazine*, *26*(6), 18–22. http://doi.org/10.1109/MSP.2009.934181

Henry, P., Krainin, M., Herbst, E., Ren, X., & Fox, D. (2012). RGB-D Mapping: Using Kinect-style Depth Cameras for Dense 3D Modeling of Indoor Environments. *Int. J. Rob. Res.*, *31*(5), 647–663. http://doi.org/10.1177/0278364911434148

Heo, Y. S., Lee, K. M., & Lee, S. U. (2013). Joint Depth Map and Color Consistency Estimation for Stereo Images with Different Illuminations and Cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(5), 1094–1106. http://doi.org/10.1109/TPAMI.2012.167

Hoel Le Capitaine, C. F. (2011). A fast fuzzy c-means algorithm for color image segmentation. http://doi.org/10.2991/eusflat.2011.9

Hota, R. N., Jonna, K., & Krishna, P. R. (2010). On-road Vehicle Detection by Cascaded Classifiers. In *Proceedings of the Third Annual ACM Bangalore Conference* (pp. 27:1–27:5). New York, NY, USA: ACM. http://doi.org/10.1145/1754288.1754315

IEEE Xplore Abstract - Unsupervised optimal fuzzy clustering. (n.d.). Retrieved 18 August 2015, from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=192473

Janoch, A., Karayev, S., Jia, Y., Barron, J. T., Fritz, M., Saenko, K., & Darrell, T. (2011). A category-level 3-D object dataset: Putting the Kinect to work. In *2011 IEEE International*

Conference on Computer Vision Workshops (ICCV Workshops) (pp. 1168–1174). http://doi.org/10.1109/ICCVW.2011.6130382

Jun, B., & Kim, D. (2012). Robust face detection using local gradient patterns and evidence accumulation. *Pattern Recognition*, *45*(9), 3304–3316. http://doi.org/10.1016/j.patcog.2012.02.031

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, *1*(4), 321–331. http://doi.org/10.1007/BF00133570

Ksantini, R., Boufama, B., & Memar, S. (2013). A new efficient active contour model without local initializations for salient object detection. *EURASIP Journal on Image and Video Processing*, *2013*(1), 1–13. http://doi.org/10.1186/1687-5281-2013-40

Ksantini, R., Shariat, F., & Boufama, B. (2009). An Efficient and Fast Active Contour Model for Salient Object Detection. In *Canadian Conference on Computer and Robot Vision, 2009. CRV '09* (pp. 124–131). http://doi.org/10.1109/CRV.2009.19

Lai, K., Bo, L., & Fox, D. (2014). Unsupervised feature learning for 3D scene labeling. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3050–3057). http://doi.org/10.1109/ICRA.2014.6907298

Lee, T. S. (1996). Image Representation Using 2D Gabor Wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, *18*(10), 959–971. http://doi.org/10.1109/34.541406

Li, C., Kao, C.-Y., Gore, J. C., & Ding, Z. (2008). Minimization of Region-Scalable Fitting Energy for Image Segmentation. *IEEE Transactions on Image Processing*, *17*(10), 1940–1949. http://doi.org/10.1109/TIP.2008.2002304

Li, C., Xu, C., Gui, C., & Fox, M. D. (2005). Level set evolution without re-initialization: a new variational formulation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005* (Vol. 1, pp. 430–436 vol. 1). http://doi.org/10.1109/CVPR.2005.213

Li, C., Xu, C., Gui, C., & Fox, M. D. (2010). Distance Regularized Level Set Evolution and Its Application to Image Segmentation. *IEEE Transactions on Image Processing*, *19*(12), 3243–3254. http://doi.org/10.1109/TIP.2010.2069690

Lindeberg, T. (1996). Edge detection and ridge detection with automatic scale selection. In *Proceedings CVPR '96, 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996* (pp. 465–470). http://doi.org/10.1109/CVPR.1996.517113

Liu, Y.-F., Guo, J.-M., & Chang, C.-H. (2014). Low resolution pedestrian detection using light robust features and hierarchical system. *Pattern Recognition*, *47*(4), 1616–1625. http://doi.org/10.1016/j.patcog.2013.11.008

Lucchese, L., & Mitra, S. K. (n.d.). *Color Image Segmentation: A State-of-the-Art Survey*.

Malladi, R., Sethian, J. A., & Vemuri, B. C. (1995). Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, *17*(2), 158–175. http://doi.org/10.1109/34.368173

Memar, S., Jin, K., & Boufama, B. (2013). Object Detection Using Active Contour Model with Depth Clue. In M. Kamel & A. Campilho (Eds.), *Image Analysis and Recognition* (pp. 640–647). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39094-4_73

107

Mishra, B. K., Rath, A., Nayak, N. R., & Swain, S. (2012). Far Efficient K-means Clustering Algorithm. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics* (pp. 106–110). New York, NY, USA: ACM. http://doi.org/10.1145/2345396.2345414

Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, *42*(5), 577–685. http://doi.org/10.1002/cpa.3160420503

Negri, P., Goussies, N., & Lotito, P. (2014). Detecting pedestrians on a Movement Feature Space. *Pattern Recognition*, *47*(1), 56–71. http://doi.org/10.1016/j.patcog.2013.05.020

Osher, S., & Fedkiw, R. (2002). Level Set Methods and Dynamic Implicit Surfaces, Springer-Verlag, New York

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *N/A*, *9*(1), 62–66. http://doi.org/10.1109/tsmc.1979.4310076

Park, M., Hasan, M., Kim, J., & Chae, O. (2012). Hand Detection and Tracking Using Depth and Color Information. Las Vegas Nevada, USA

Peng, D., Merriman, B., Osher, S., Zhao, H., & Kang, M. (1999). A PDE-Based Fast Local Level Set Method. *Journal of Computational Physics*, *155*(2), 410–438. http://doi.org/10.1006/jcph.1999.6345

Pi, L., Shen, C., Li, F., & Fan, J. (2007). A variational formulation for segmenting desired objects in color images. *Image and Vision Computing*, *25*(9), 1414–1421. http://doi.org/10.1016/j.imavis.2006.12.013

Poonawala, A. & Milanfar, P. (2002). Active Contour Model (Snake). Lecture Notes.

Preetha, M. M. S. J., Suresh, L. P., & Bosco, M. J. (2012). Image segmentation using seeded region growing. In *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)* (pp. 576–583). http://doi.org/10.1109/ICCEET.2012.6203897

Pujar, J. H., & Shambhavi, D. S. (2010). A Novel Digital Algorithm for Sobel Edge Detection. In V. V. Das, R. Vijayakumar, N. C. Debnath, J. Stephen, N. Meghanathan, S. Sankaranarayanan, … N. Thankachan (Eds.), *Information Processing and Management* (pp. 91–95). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-12214-9_16

Rafiee, G., Dlay, S. S., & Woo, W. L. (2013). Region-of-interest extraction in low depth of field images using ensemble clustering and difference of Gaussian approaches. *Pattern Recognition*, *46*(10), 2685–2699. http://doi.org/10.1016/j.patcog.2013.03.006

Rosasco, L., De Vito, E., Caponnetto, A., Piana, M., & Verri, A. (2004). Are Loss Functions All the Same? *Neural Comput.*, *16*(5), 1063–1076. http://doi.org/10.1162/089976604773135104

Scharstein, D., & Szeliski, R. (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, *47*(1-3), 7–42. http://doi.org/10.1023/A:1014573219977

Scharstein, D., & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, 2003. Proceedings (Vol. 1, pp. I–195–I–202 vol.1). http://doi.org/10.1109/CVPR.2003.1211354

Scholkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

Shet, V. D., Neumann, J., Ramesh, V., & Davis, L. S. (2007). Bilattice-based Logical Reasoning for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07* (pp. 1–8). http://doi.org/10.1109/CVPR.2007.383133

Shih, F. Y., & Cheng, S. (2005). Automatic seeded region growing for color image segmentation. *Image and Vision Computing*, *23*(10), 877–886. http://doi.org/10.1016/j.imavis.2005.05.015

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., … Blake, A. (2011). Real-time Human Pose Recognition in Parts from Single Depth Images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1297–1304). Washington, DC, USA: IEEE Computer Society. http://doi.org/10.1109/CVPR.2011.5995316

Siang Tan, K., & Mat Isa, N. A. (2011). Color image segmentation using histogram thresholding Fuzzy C-means hybrid approach. *Pattern Recognition*, *44*(1), 1–15. http://doi.org/10.1016/j.patcog.2010.07.013

Skarbek, W., & Koschan, A. (1994). *Colour Image Segmentation: A Survey*.

Solem, J. E., Overgaard, N. C., & Heyden, A. (2006). Initialization Techniques for Segmentation with the Chan-Vese Model. In *18th International Conference on Pattern*

Recognition,   2006.   ICPR   2006   (Vol.   2,   pp.   171–174). http://doi.org/10.1109/ICPR.2006.713

Stockman, G., & Shapiro, L. G. (2001). *Computer Vision* (1st ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Sussman, M., & Fatemi, E. (1999). An Efficient, Interface-Preserving Level Set Redistancing Algorithm and Its Application to Interfacial Incompressible Fluid Flow. *SIAM J. Sci. Comput.*, *20*(4), 1165–1191. http://doi.org/10.1137/S1064827596298245

Tian, Y., Duan, F., Zhou, M., & Wu, Z. (2011). Active contour model combining region and edge   information.   *Machine   Vision   and   Applications*,   *24*(1),   47–61. http://doi.org/10.1007/s00138-011-0363-7

Vapnik, V. N. (1998). *Statistical Learning Theory* (1 edition). New York: Wiley-Interscience.

Vese, L. A., & Chan, T. F. (2002). A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, *50*(3), 271–293. http://doi.org/10.1023/A:1020874308076

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(6), 583–598. http://doi.org/10.1109/34.87344

Wang, H., Huang, T.-Z., Xu, Z., & Wang, Y. (2014). An active contour model and its algorithms with local and global Gaussian distribution fitting energies. *Information Sciences*, *263*, 43–59. http://doi.org/10.1016/j.ins.2013.10.033

111

Wang, X.-F., Huang, D.-S., & Xu, H. (2010). An efficient local Chan–Vese model for image segmentation. *Pattern Recognition*, *43*(3), 603–618. http://doi.org/10.1016/j.patcog.2009.08.002

Wu, B., Nevatia, R., & Li, Y. (2008). Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008* (pp. 1–8). http://doi.org/10.1109/CVPR.2008.4587750

Xu, C., & Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, *7*(3), 359–369. http://doi.org/10.1109/83.661186

Yang, Q. (2012). A non-local cost aggregation method for stereo matching. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1402–1409). http://doi.org/10.1109/CVPR.2012.6247827

Yuan, C., Ma, Y.-F., & Zhang, H.-J. (2003). Extracting video object's motion trajectory by velocity voting. In Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia (Vol. 3, pp. 1561–1565 vol.3). http://doi.org/10.1109/ICICS.2003.1292729

Zhang, K., Song, H., & Zhang, L. (2010). Active contours driven by local image fitting energy. *Pattern Recognition*, *43*(4), 1199–1206. http://doi.org/10.1016/j.patcog.2009.10.010

Zhang, K., Zhang, L., Song, H., & Zhang, D. (2013). Reinitialization-Free Level Set Evolution via Reaction Diffusion. *IEEE Transactions on Image Processing*, *22*(1), 258–271. http://doi.org/10.1109/TIP.2012.2214046

Zhang, K., Zhang, L., Song, H., & Zhou, W. (2010). Active contours with selective local or global segmentation: A new formulation and level set method. *Image and Vision Computing*, *28*(4), 668–676. http://doi.org/10.1016/j.imavis.2009.10.009

Zhao, H.-K., Chan, T., Merriman, B., & Osher, S. (1996). A Variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, *127*(1), 179–195. http://doi.org/10.1006/jcph.1996.0167

Zheng, W., & Liang, L. (2009). Fast car detection using image strip features. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009* (pp. 2703–2710). http://doi.org/10.1109/CVPR.2009.5206642

VITA AUCTORIS

NAME: Sara Memar Kouchehbagh

PLACE OF BIRTH: Tehran, Iran

YEAR OF BIRTH: 1986

EDUCATION: Karaj Azad University, B.Sc., Iran, 2008

                         University Putra Malaysia, M.Sc., Malaysia, 2011