Regularizing variational methods for robust object boundary detection

Fang, Wen

2008

Fang, W. (2008). Regularizing variational methods for robust object boundary detection.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/14582>

<https://doi.org/10.32657/10356/14582>

Regularizing Variational Methods for Robust Object Boundary Detection

Fang Wen

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2008

*To my dear parents and husband,
who always give me love and support,
and made all this possible ...*

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Associate Professor Kap Luk Chan for his valuable guidance and encouragement on my research work throughout my Ph.D study in Nanyang Technological University (NTU). He helps a lot in problem solving and research with his broad and profound knowledge in image processing, pattern recognition, machine learning, and mathematics. The atmosphere in his group and the wide range of research topics provide me an inspiring environment for my work. He is a tremendous mentor, collaborator, and friend. This thesis would not have been possible without his supervision.

I would like to thank to the School of EEE, NTU, Singapore, for supporting me with research scholarship for three years, and the Biomedical Engineering Research Centre (BMERC) for providing me with excellent research facilities.

I would like to thank a number of fellow researchers, Jiayin Zhou, Peng Li, Lei Wang, Yan Gao, and Juan Xu for their hospitality and stimulating discussions on my work. I appreciated very much the fruitful discussions with them. I would also like to thank many of my friends in NTU for their great help in my research and life in Singapore.

Finally, I would like to thank my dear parents who always give me encouragement and support whenever and wherever. Specially, I would like to thank my dear husband. I dedicate this thesis to my family.

Contents

Acknowledgments	i
Summary	vii
List of Figures	x
List of Tables	xix
1 Introduction	1
1.1 Object Boundary Detection	1
1.2 Motivation and Objectives	4
1.3 Major Contributions	6
1.4 Organization of the Thesis	7
2 Review of Existing Variational Methods	9
2.1 Introduction	9
2.2 Boundary-based Variational Methods	10
2.2.1 Explicit Contour Representations	10
2.2.2 Implicit Contour Representations	15

2.2.3	Explicit vs. Implicit Contour Representations	18
2.3	Region-based Variational Methods	19
2.4	Boundary-based vs. Region-based Variational Methods	22
2.5	Problems of Using Boundary-based Variational Methods	23
2.6	Concluding Remarks	25
3	Regularizing Variational Methods by Adding Temporal Information	26
3.1	Background	27
3.1.1	The Dropout Problem in Echocardiography	28
3.1.2	Review of Related Work	30
3.1.3	Motivations and Objectives	31
3.2	Incorporation of Temporal Information into Variational Methods as a Regularizer	32
3.2.1	Block Matching	33
3.2.2	Identifying Strong / Weak Segments of Inner Heart Wall Bound- ary Using GMM	34
3.2.3	Regularizing Variational Methods by Incorporating Temporal In- formation	38
3.2.3.1	Incorporating temporal information into active contour framework	39
3.2.3.2	Incorporating temporal information into level set frame- work	41
3.3	Experimental Evaluations	42
3.3.1	Detection Performance	43
3.3.1.1	Active contour method using temporal information	43

3.3.1.2	Level set method using temporal information	47
3.3.2	Quantitatively Analysis	53
3.3.2.1	Active contour method using temporal information . .	53
3.3.2.2	Level set method using temporal information	54
3.3.3	Computational Burden	58
3.3.3.1	Active contour method using temporal information . .	59
3.3.3.2	Level set method using temporal information	60
3.3.4	Comparison of the Two Proposed Methods	60
3.4	Discussions and Conclusions	62
4	Regularizing Variational Methods by Adding a Shape Prior Model	64
4.1	Introduction	64
4.2	Learning a Shape Prior Model Using PCA and a Single Gaussian Function	67
4.2.1	Motivations and Objectives	67
4.2.2	Constructing the Shape Prior Model	68
4.2.2.1	Shape alignment	68
4.2.2.2	Creating the training data	69
4.2.2.3	Constructing the shape prior model	71
4.2.3	Combining the Shape Prior Model with GAC Equation	74
4.2.4	Experimental Results	78
4.2.4.1	Synthetic images	78
4.2.4.2	Real human shape images	81
4.2.4.3	Computational time	83
4.2.4.4	The influence of parameter γ	84

4.3	Learning a Shape Prior Model Using PCA and GMM	85
4.3.1	Motivations and Objectives	85
4.3.2	Constructing the Shape Prior Model	87
4.3.2.1	Bayesian classifier based on Gaussian Mixture Model (GMM)	87
4.3.2.2	Density estimation of the training data using GMM . .	89
4.3.2.3	Constructing the Bayesian classifier and the shape prior model	90
4.3.3	Experimental Results	91
4.3.3.1	Synthetic dataset	91
4.3.3.2	Real human shape dataset	95
4.3.3.3	Sign language dataset	97
4.3.3.4	Tolerance to the extent of occlusions	103
4.4	Learning a Shape Prior Model Using Kernel PCA and GMM	104
4.4.1	Motivations and Objectives	104
4.4.2	Constructing the Shape Prior Model	105
4.4.2.1	Kernel PCA	106
4.4.2.2	Density estimation in subspace	108
4.4.2.3	Combining the shape prior model with GAC equation	108
4.4.3	Experimental Results	112
4.4.3.1	Synthetic images	112
4.4.3.2	Real human shape from surveillance video sequence . .	114
4.5	Comparison of the three proposed shape prior models	119
4.5.1	Construction process	119

4.5.2	Application area	121
4.5.3	Computational time	121
4.5.4	Cognition level	123
4.6	Discussions and Conclusions	123
5	Conclusions and Future Work	125
5.1	Conclusions	125
5.1.1	Regularizing Variational Methods by Adding Temporal Information	126
5.1.2	Regularizing Variational Methods by Adding Shape Prior Model	127
5.2	Future Work	129
	Author's Publications	131
	Bibliography	133

Summary

From biomedical images such as those from ultrasound scans, computed tomography (CT) and Magnetic Resonance Imaging (MRI), to the surveillance videos in security systems, variational methods have played an important role in tasks such as image segmentation, object boundary detection, shape recovery, object localization, and visual target tracking. Variational methods can be divided into two categories: boundary-based and region-based. This thesis focuses on the boundary-based variational methods for object boundary detection.

Since the boundary-based variational methods utilize only gradient based information to locate the object boundary, they are very sensitive to the presence of imaging artifacts. If the object boundary is corrupted by image noise, or partially occluded by other objects, or interfered by surrounding local gradient magnitudes from the background, the original boundary may be lost and the gradient magnitudes in such vicinity will be too weak to constrain the curve evolution driven by the variational methods. This is regarded as the “**missing boundary**” problem in this thesis. The “missing boundary” problem is a major obstacle to successful application of boundary-based variational methods for robust object boundary detection.

To solve such a problem encountered in the application of boundary-based variational method for detecting the inner heart wall boundary of left ventricle from the echocardiographic image sequence, a new method is proposed which incorporates temporal information to regularize the variational methods. The object boundary is quantitatively partitioned and classified into strong and weak segments. For the weak segments, they are considered to be weakened by image degradations and thus the gradients in these places are too weak to constrain the curve evolution. Hence, temporal information from neighboring frames is incorporated into the variational framework to recover the original object boundary by strengthening the image gradient force in the weak segments. The main contribution in this work is that, the temporal information is incorporated in a novel way so that it can remedy the disruptions to object boundary.

Comparisons with similar existing methods have been made. The proposed methods show better detection performance for the intended task.

Object boundary detection also suffers from the problem of occlusion where the object of interest may be partially occluded or surrounded by cluttered background, and no temporal information can be utilized at hand. Hence, the shape prior information is proposed to regularize the variational methods. A set of object shapes are collected as training shapes. Their distribution is estimated and modeled by a probability density function. A shape prior model is constructed and is incorporated into the variational methods as a regularizer to constrain the curve evolution. Two methods are proposed in this work. The first proposed shape prior model approximates the distribution of training data by a single Gaussian function in the subspace projected by the principal component analysis (PCA), while the second method extends the first one by approximating the distribution with a Gaussian mixture model (GMM). Moreover, an explicit alignment method is used to speed up the convergence under the influence of shape prior knowledge, and a weighting scheme is developed to balance the influence between the image gradient and shape prior information. By evaluating the proposed methods on synthetic dataset, real human shape dataset and sign language dataset, and comparing with existing methods, it is observed that the proposed method can successfully detect and recover the original object shapes in the presence of partial occlusion and cluttered background. For the second proposed shape prior model, it is also able to generalize to multiple shape patterns either from multiple objects or from multiple viewpoints of a single object.

The linear shape prior models proposed above are based on principal component analysis and Gaussian density estimation of the training shapes. It assumes that the variations of training shapes correspond to the linear combination of a set of eigenvectors. However, very often the object shapes may undergo nonlinear deformation. Its shapes may follow a nonlinear change due to the change of viewing angles or silhouette deformations. Hence, a nonlinear shape prior model is proposed and is used as a regu-

larizer for curve evolution. It is constructed based on the kernel principal component analysis (KPCA) and Gaussian mixture model. The obtained probability function is treated as a shape energy term and is incorporated into a variational method as a regularizer. Experimental results demonstrate that the proposed shape prior model can successfully handle the object occlusion problem in images and show good detection performance.

List of Figures

2.1	The active contour method represents the contour by a set of discrete marker points [1].	11
2.2	Comparison of the external force field of traditional snake and the GVF snake [2]. It can be seen that the GVF field has a much larger capture range.	14
2.3	Level set formulation of front motion: the curve (darkened line) and the corresponding surface $\Phi(x, y, t)$ at different time t [1].	16
2.4	Failure of boundary-based variational method in the presence of missing boundary information. (a) Failure of level set method when encountering weak boundary containing “gaps”. The curve leaked out from the gaps instead of converging to it (the curve iteration process was manually stopped after the leakage problem happened). (b) Failure of level set method when encountering object occlusion. The curve converged to the wrong location.	24
3.1	The dropouts phenomenon in consecutive 4 frames in a echocardiographic sequence (signal dropouts in the two middle frames).	28

3.2	The detection problem of variational methods in the presence of signal dropout in echocardiographic images. (a) Ground truth. (b) The active contour method could not detect the real boundary and converged to a local extremum of gradient within the heart chamber. (c) The level set curve could not be stopped at the weak boundary but leaked out to the outside of the heart chamber.	29
3.3	Illustration of the block matching process.	33
3.4	Select the training data. The points around the predicted position of ventricular boundary are selected as training data.	36
3.5	Using GMM to approximate the distribution of training data. g_1 , g_2 and g_3 are three components of GMM. g_1 corresponds to the distribution of contour points, g_2 corresponds to the distribution of background points, and g_3 is a duplicate of g_1 in the negative x axis.	36
3.6	Position of weak segments on the contour that need to be strengthened by temporal information.	38
3.7	Parameters δ_1 and δ_2 that are used to control the incorporation process of temporal information into the variational frameworks.	40
3.8	The stopping criterion $k_I(= g(I))$ before adding the temporal information. The black region in the image corresponds to the maximum strength of gradient, while the white region corresponds to the minimum strength of gradient (where the strength of gradient value approximate 0).	42
3.9	The stopping criterion $k_I(= g(I))$ after adding the temporal information. The black region in the image corresponds to the maximum strength of gradient, while the white region corresponds to the minimum strength of gradient (where the strength of gradient value approximate 0).	42

3.10	Comparison of the detection results of the proposed active contour method using temporal information (row 1), with the ground truth (row 2), the traditional active contours (row 3), and Pardas’s method (row 4) and Lin’s method (row 5) in four of the image frames during the heart contraction period in the first echocardiographic sequence.	45
3.11	Comparison of the detection results of the proposed active contour method using temporal information (row 1), with the ground truth (row 2), the traditional active contours (row 3), and Pardas’s method (row 4) and Lin’s method (row 5) in four of the image frames during the heart relaxation period in the first echocardiographic sequence.	46
3.12	Detailed illustration of the improvements of our proposed active contour method using temporal information, comparing with the ground truth, the traditional active contour method, Pardas’ method and Lin’s method.	47
3.13	Comparison of the detection results of our proposed method (row 1) with the ground truth (row 2), the traditional active contour (row 3), Pardas’s method (row 4) and Lin’s method (row 5) in four image frames during heart contraction period in the second echocardiographic sequence. . .	48
3.14	Comparison of the detection results of our proposed method (row 1) with the ground truth (row 2), the traditional active contour (row 3), Pardas’s method (row 4) and Lin’s method (row 5) in four image frames during heart relaxation period in the second echocardiographic sequence.	49
3.15	Comparison of the detection results of the proposed level set method using temporal information (row 1), with the ground truth (row 2), and the traditional level set method (row 3) in four of the image frames during the heart contraction period in the first echocardiographic sequence. As the boundary information is strong enough to be detected, all algorithm perform well.	50

3.16	Comparison of the detection results of the proposed level set method using temporal information (row 1), with the ground truth (row 2), and the traditional level set method (row 3) in four of the image frames during the heart relaxation period in the first echocardiographic sequence. Since the ventricular boundaries suffer from signal dropout in some images, the traditional level set method cannot detect the weak segments of the boundary and leaked out. The proposed method can correctly detect the weak heart boundary without any leakage.	51
3.17	Comparison of the detection performance of our proposed level set method using temporal information with the active contour without edges in the third testing sequence. Row 1 shows the results of our proposed method, row 2 is the ground truth and row 3 is the results of active contour without edges. It can be seen that the proposed method shows more accurate detection performance than the active contour without edges method. .	52
3.18	Comparison of the MAD distance and the Jaccard similarity coefficient $\mathcal{J}$ of the proposed active contour method using temporal information with the traditional active contour method, Pardas's method and Lin's method in the first image sequence.	55
3.19	Comparison of the MAD distance of the three algorithms in the first testing sequence.	56
3.20	Comparison of the Jaccard similarity coefficient $\mathcal{J}$ of the three algorithms in the first testing sequence.	57
3.21	Comparison of the initial contour of the proposed method and the traditional active contour method.	59
4.1	Examples of the training shapes before and after alignment. The shapes generally share the same center, size and orientation after alignment. .	70
4.2	Comparison of the extent of shape overlap before and after alignment. .	70

4.3	Examples of level set functions of the synthetic data.	71
4.4	The trace of curve evolution process (corresponding to the shape “cross”) in the subspace under the influence of the shape prior knowledge. . . .	74
4.5	Illustration of the selection of γ	76
4.6	Parameter d_2 is obtained as the standard deviation of the distance among the training data.	77
4.7	Intermediate stages during the curve evolution process of the proposed method.	78
4.8	Comparison of the detection results of traditional geodesic active contours without using shape prior model and the proposed method using shape prior model for partially occluded objects.	79
4.9	Comparison of the detection results of traditional geodesic active contours without using shape prior model and out proposed method using shape prior model for partially cut objects.	80
4.10	Comparison of the detection results of traditional geodesic active contours without using shape prior model and out proposed method using shape prior model for objects in cluttered background.	81
4.11	Examples of selected human shapes.	82
4.12	Examples of the level set functions of training shapes.	82
4.13	Comparison of the detection results of the proposed method with the traditional GAC method.	83
4.14	Evaluation of the influence of thresholds d_1 and d_2 in parameter γ on the detection results.	84
4.15	The dimensionality reduction and density estimation using PCA and GMM.	92

4.16	Decision boundary of the Bayesian classifier for synthetic data. It separates the entire region into three subregions according to the posterior probability.	92
4.17	Comparison of the detection results of the proposed method with the traditional GAC method in the presence of object occlusions. Column (a) is the partially occluded objects, column (b) is the detection result of traditional GAC, column (c) is the Bayesian classification result (marked by $\star$) of the test shape, column (d) is the reconstructed shape prior model, and column (e) is the detection result of the proposed method.	93
4.18	Examples of training shapes from real human shape dataset.	94
4.19	Examples of the level set functions of human shapes.	94
4.20	Dimensionality reduction and density estimation on real human shape dataset using PCA and GMM.	95
4.21	Decision boundary of the Bayesian classifier for real human shape dataset. It separates the entire region into four subregions according to posterior probability.	95
4.22	Comparison of the detection results of our proposed method with the traditional GAC method in presence of cluttered background. Column (a) shows the partially occluded human shapes, column (b) shows the detection results of traditional GAC method, and column (c) shows the detection results of the proposed method. It can be seen that our proposed method can successfully recover the original human shapes in the presence of cluttered background.	97

4.23	Bayesian classification results of the partially occluded test shapes. Point 1 corresponds to the human shape of standing with open-armed, point 2 corresponds to the human shape of standing (view from the front), and point 3 corresponds to the human shape of standing (view from silhouette). It can be seen that all the testing data have been correctly classified to the most similar shape cluster by Bayesian classifier. A shape prior model can be constructed using the statistical properties of the corresponding cluster. The missing boundary of human shapes caused by object occlusion can thus be recovered using the shape prior knowledge.	98
4.24	Examples of hand shapes in the international sign language dataset. Each hand gesture corresponds to a letter.	99
4.25	Dimensionality reduction and density estimation using PCA and GMM on sign language dataset.	99
4.26	Decision boundary of the Bayesian classifier for sign language dataset. The entire region has been separated into five subregions according to posterior probability.	100
4.27	Comparison of the detection result of our proposed method and Cremers' method. (a) Partially occluded sign gesture "Y". (b) Detection result of our proposed method. (c) Detection result of Cremers' method.	101
4.28	Trace of the test shape during the energy minimization process of Cremers' method in the projected 2D space. The curve finally converged to the local minimum of the energy function.	101
4.29	Letter "W" in sign language. The detection result of Cremers' method is quite similar to the letter "W" in sign language although its original letter should be "Y". This detection error may directly cause misunderstanding in communications.	102

4.30	Bayesian classification result of the test shape of the proposed shape prior model. The test shape is assigned to a shape cluster according to the posterior probability.	102
4.31	Testing the tolerance of the shape model to the extent of shape distortions.	103
4.32	Classification results of the shape of cutting a circle with the amount of 10%, 20%, 30%, 40%, 50% and 60% respectively. It can be seen that when the shape distortion reaches 50%, the remaining shape information is not sufficient to identify the object, leading to a wrong classification result.	103
4.33	The basic idea of kernel PCA. In some high dimensional feature space F (bottom right), linear PCA is performed, just as a PCA in input space (figure (a)). Since F is nonlinearly related to input space (via Φ), the contour lines of constant projections onto the principal eigenvector (drawn as an arrow) become nonlinear in input space. Crucial to kernel PCA is the fact that we do not actually perform the mapping into F , but instead perform all necessary computations by the use of a kernel trick k [3].	106
4.34	Results of performing dimensionality reduction using KPCA and density estimation using Gaussian mixture model on the synthetic data.	113
4.35	Detection results of the proposed nonlinear shape prior model and the traditional geodesic active contour method on partially occluded objects. Column (a) is the examples of partially occluded objects. Column (b) is the detection results of the traditional geodesic active contours. Column (c) is the detection results of the proposed method.	114

4.36	Energy minimization process of the proposed method. Upper line corresponds to the minimization of partially occluded circle. Lower line corresponds to the minimization of partially occluded cross. Finally, they both converged to the center of the most similar shape cluster respectively.	115
4.37	Examples of alignment training shapes.	115
4.38	Examples of training data. Each image is the level set representation of a corresponding shape.	116
4.39	Training data in feature subspace after KPCA. Different color corresponds to different shape state (data are viewed in a 3D space). . . .	116
4.40	Fitting GMM to our training data in subspace. Different colors correspond to the different shape clusters (Data are viewed in a 2D space). .	117
4.41	Detection results of proposed method on different human shape patterns. Column (a) is the partially occluded human shapes, column (b) is the detection results of traditional GAC, column (c) is the detection results of our proposed method.	118
4.42	Energy minimization process of our proposed method. The two traces in the figure represent the convergence process of two human shape states. They start from an obscured shape and finally converged to the center of each shape cluster.	119

List of Tables

2.1	Comparison of advantages and disadvantages of explicit and implicit contour representations.	19
3.1	Comparison of the MAD distance between our proposed active contour method using temporal information with the other methods that incorporate temporal information into the external energy of active contour function	54
3.2	Comparison of the Jaccard similarity coefficient $\mathcal{J}$ between our proposed active contour method using temporal information with the other methods that incorporate temporal information into the external energy of active contour function	56
3.3	Comparison of the MAD distance of the three algorithms in all the five image sequences.	57
3.4	Comparison of the Jaccard similarity coefficient $\mathcal{J}$ of the three algorithms in all the five image sequences.	58
3.5	Comparison of the number of iterations and computation time of the proposed active contour method using temporal information with the traditional active contour method	60
3.6	Comparison of the MAD distance between the proposed method using 5000 points and 1000 points in all the five image sequences.	61

3.7	Comparison of the Jaccard similarity coefficient $\mathcal{J}$ between our proposed method using 5000 points and 1000 points in all the five image sequences.	61
3.8	Comparison of the average computation time per frame of the proposed level set method using temporal information, the traditional active contour method, and the traditional level set method.	61
4.1	Computational time of the proposed method on the partially occluded synthetic images.	113
4.2	Computational time of the proposed method on the partially occluded real human shapes.	117
4.3	Comparison of the construction process of the three proposed shape prior models.	120
4.4	Comparison of the application area of the three proposed shape prior models.	121
4.5	Comparison of the computational time of the three proposed shape prior models.	122
4.6	Comparison of the cognition level of the three proposed shape prior models.	122

Chapter 1

Introduction

1.1 Object Boundary Detection

Object boundary detection and tracking are often necessary in many computer vision applications such as video surveillance, object-based coding, and robotics etc. [4, 5, 6]. They are also found necessary in many applications of biomedical image processing and analysis including organ detection, outline delineation and shape recovery [7, 8, 9]. Many algorithms have been developed to accomplish these tasks.

In the applications in computer vision, a variational method is proposed for detecting and tracking of boundaries of multiple moving objects in image sequences in [10]. This method estimates the distribution of the inter-frame difference using a mixture model. Thus, it is able to detect the motion of the object boundaries. The entire detection and tracking problem is formulated in a common framework that employs a geodesic active contour functional. In [11], a Bayesian framework is developed which can simultaneously detect and track multiple targets from image sequences. The method is implemented by computing the joint multitarget probability density (JMPD). In another research work, a method for the detection of human walking in video and the extraction of gait features is reported [12]. The method is based on

detecting the moving leg pairs and developing the symmetry based approach for more robust cases. It enables the method to be independent of noise and the varying frame rates. A new method for detecting object classes that exhibit variable shape structure in heavily cluttered images is proposed in [13]. The hidden State Shape Models (HSSMs), a reformulation of the Hidden Markov Models (HMMs), are introduced to model object classes of variable shape structure using a probabilistic framework.

In the biomedical applications, a Bayesian framework is proposed to detect the contour from sequences of echocardiographic images [14]. The contour sequences are assumed to be first-order Markov random processes and the image pixels are modeled as Rayleigh distributed random variables. In [15], a mathematical morphology method is employed to detect the ventricular boundary. Friedland and Adam reported an automatic boundary detection algorithm for sequential echocardiograms using a stochastic relaxation method [16]. Temporal information is integrated with this model. A multiple active contour algorithm based on a temporal domain extension of the active contour model was proposed in [17] to detect the epicardial and endocardial borders in cardiac sequences.

Variational methods have played an important role in both computer vision and biomedical applications [24, 25, 26, 27]. Variational methods usually rely on propagating or deforming an initial curve towards the lowest potential of a curve-based energy [28]. A number of variational methods [29, 30, 31, 32] have been developed to-date. They can be roughly divided into two categories: boundary-based approach and region-based approach. The boundary-based approach depends on the information of local image gradients which may correspond to the boundary of an object. Due to their relatively stable performance and comparatively fast computation than the region-based variational methods, they have been widely adopted [2]. However, as only gradient information is utilized, their detection performance is determined by the image gradient. The presence of object occlusion and other scene clutter can lead to erroneous detection as their presence will introduce undesired local image gradients.

Region-based approach depends on the information of the entire object region such as color or texture [33, 34, 35]. It is more robust than the boundary-based method since the information of the whole region is explored. However, it needs more computations and the region of interest has to be uniform and occlusion free. Hence, region-based method may not be suitable for more complex scenes. In this thesis, the focus is on the boundary-based variational methods due to many attractive properties for a wider range of applications. The objective of this thesis is to explore mechanisms that make the boundary-based approach more robust for object boundary detection.

There are several well known boundary-based variational methods: the active contour method [29], the level set method [36] and the geodesic active contour method [32]. The active contour (or snake) method is a parametric deformable model which pulls the curve towards the objects by minimizing an energy function composed of internal and external energies [4, 2]. The main problem of this method is that it cannot handle topological changes such as splitting or merging. Therefore, in recent years, the level set method has been used more favorably [37, 38, 39]. The level set method implicitly represents the contour as the zero level set of a higher dimensional surface function [1]. Then the ordinary differential equations evolving the explicit marker points are replaced by a partial differential equation which models the evolution of the hypersurface function [40]. This enables the level set method to handle topological changes such as cusps, and sharp corners. Later, Caselles [32] improved the level set formulation and proposed a geodesic active contour (GAC) method. It turns boundary detection into a problem of finding the path of minimal distance measured in the Riemannian space. This method achieves a stable detection when the gradients on the boundary exhibit variations or small gaps [39, 41].

In the existing work, most of the application of the variational methods focus on the functional for detecting the boundary of target region. In many practical situations, obtaining a complete boundary of a target is not always guaranteed. In this thesis, the level set method for the detection of target boundary when there is missing parts

on the boundary is closely studied. Hence, regularization is proposed to solve what is considered as "the missing boundary problem" in this thesis. In different applications, the causes of missing boundary parts are different. The thesis studied two situations, one is heart wall boundary detection from the echocardiograms in medical application, and the other is human detection in surveillance application. The cause of the missing boundary problem considered in the thesis is due to noise in ultrasound imaging. The cause of missing boundary problem considered in the surveillance data is due to object occlusion. Different approaches and methods are investigated in the thesis to solve the problem of different origins. Variational methods are regularized to handle the problem in both by considering the temporal information in the medical application and shape prior modelling in the surveillance application. The variation methods provide a natural framework to allow regularizing terms to be added to their functionals. The thesis explore such regularized functionals and evaluate their performance for the intended tasks.

1.2 Motivation and Objectives

Since the boundary-based variational methods only depend on local image gradient information to locate the object boundary, they are very sensitive to the interference of boundary due to the presence of unwanted image structures or artifacts from the imaging process. If the object boundary is contaminated by image noise, or partially occluded by other objects, or interfered by the spurious edges from the background, or loss of sensed signals, the original boundary information is either missing or interfered, and the local gradient may not be strong enough to constrain the curve evolution driven by boundary-based variational methods. The evolving curve can be pulled away from the real boundary and converge to wrong locations. In this thesis, this is regarded as the "**missing boundary**" problem. The "missing boundary" problem is a major obstacle to successful applications of boundary-based variational methods for robust object boundary detection. In order to tackle this problem, additional boundary

constraints need to be incorporated into the variational methods.

It is noticed that for object detection from an image sequence, temporal correlation and continuity of object boundaries can be very useful to improve the detection performance [42, 43, 44]. The same object in adjacent frames are usually similar to each other and are located within close vicinity of each other in the image sequence. They are highly correlated [45]. If the object boundary is corrupted by image degradations in the current frame, it is possible to reconstruct the missing parts by using the temporal information from its neighboring frames. In this way, object boundary detection can be greatly improved. However, the missing parts should reappear within a few frames so that the temporal continuity can be valid. If the missing parts are consistently absent for a prolonged period over a substantially large number of frames, e.g. 15 frames, the temporal continuity constraint is difficult to enforce.

In cases where there is no temporal information available at hand to be utilized (such as a targeted object is partially occluded or is in a cluttered background in a single snap shot and the temporal continuity constraint cannot be enforced), it is preferred to employ the shape prior information to constrain the curve evolution [46, 47]. A shape prior model can be constructed based on the statistical learning of the distribution of a set of training shapes. It contains the general shape characteristics and provides a global shape consistency of the object. If the object boundary is partially occluded or is disrupted by background noise, the evolving curve can recover the original boundary information by referring to the shape information in the shape prior model. Hence, the variational methods can cope with partial occlusions and cluttered background.

In order to solve the “missing boundary” problem, new methods are proposed to regularize the variational methods for robust object boundary detection in this thesis. The first part of this thesis reports on a method to utilize temporal information to regularize the active contour method and level set method. It is developed for a biomedical application. The second part of this thesis focuses on methods that utilize the shape prior information to regularize the geodesic active contour method. It is

developed for a computer vision application. The proposed methods are compared with similar existing methods, and are evaluated on different datasets to assess their detection abilities.

1.3 Major Contributions

The major contributions of this thesis can be summarized as follows:

- A new method is proposed which incorporates temporal information as a curve evolution regularizer into the variational methods. The object boundary is quantitatively partitioned and classified into strong and weak segments. The weak segments are considered to be disrupted by the image degradations and thus are not strong enough to constrain the curve evolution. Hence, temporal information from neighboring frames is incorporated into the variational framework to reconstruct the original boundary information in weak segments to strengthen their image gradient magnitudes. The temporal information is proposed to be incorporated into the external energy of active contour method, and the speed function of level set method. A novel formulation of the gradient restoration mechanism from temporal information is developed. Comparisons with the existing methods have been made. The experimental results show that the proposed methods perform better for the intended task than the compared methods in terms of the detection performance and computational burden.
- Two linear shape prior models are proposed to incorporate shape prior information for regularizing the curve evolution in the geodesic active contour (GAC) framework. A set of object shapes are collected as training shapes. They are projected onto a low dimensional subspace using principal component analysis (PCA). Their distribution is estimated and modeled by a probability density function. Then a shape prior model is constructed and is incorporated into the GAC method. The first proposed method models the distribution of training

data by a single Gaussian function while the second proposed method extends the first one by modeling the distribution with a Gaussian mixture model (GMM). The first method is able to recover the original object shapes in the presence of partial occlusions and cluttered background, while the second method is able to encode multiple shape patterns either from multiple different objects or from multiple different shape states of a single object. Synthetic dataset, real human shape dataset and sign language dataset are used to evaluate the detection performance. The experimental results demonstrate the detection abilities of the proposed method.

- Finally, a nonlinear shape prior model is proposed and is used to regularize the curve evolution. The training shapes are collected from a series of object shapes seen under different viewing angles or silhouette deformations. They are considered to form clusters in a submanifolds after a nonlinear dimensionality reduction using kernel principal component analysis (KPCA). A Gaussian Mixture Model (GMM) is used to estimate the distribution of multiple clusters in the subspace. The obtained probability density function is used as a shape energy term and is incorporated into the geodesic active contour function as a regularizer. During the curve evolution process, the energy function is reformulated to contain a classification scheme which will designate the curve to a most similar shape cluster based on the posterior probability. Experimental results show good detection performance of the proposed method.

1.4 Organization of the Thesis

This thesis is organized as follows:

- A literature review of the state-of-the-art variational methods is given in Chapter 2. The variational methods are considered to form two categories: boundary-based and region-based. The boundary-based variational methods are further

divided into the explicit and implicit contour representation methods. These two kinds of boundary based methods are reviewed respectively. The problems of using boundary-based variational methods are discussed.

- In Chapter 3, a new method is proposed which incorporates temporal information into the variational method as a curve evolution regularizer for the “missing boundary” problem encountered in a biomedical application. Temporal information is proposed to be added into the external energy of active contour framework and the speed function of level set framework. A Gaussian mixture model is used to quantitatively identify the boundary into strong and weak segments. A weight and a strengthening factor are developed to control the boundary restoration process. The proposed method is compared with the other methods in the experimental evaluations. Experimental results and discussions are given therein.
- In Chapter 4, the second method is proposed which employs the shape prior information as a regularizer to constrain the curve evolution. Three new shape prior models are presented. The first two methods are linear shape prior models which employ the principal component analysis (PCA) to project the training data onto a low dimensional subspace. Then the data distribution is approximated using a probability density function. The first proposed method approximates the distribution using a single Gaussian function while the second proposed method extends the first one by approximating the distribution using a Gaussian mixture model (GMM). The third proposed method is a nonlinear shape prior model which employs the kernel principal component analysis (KPCA) to project the training data onto a subspace. The data distribution is approximated by GMM. These three proposed methods are tested on synthetic dataset, real human shape dataset, sign language dataset, and real surveillance videos to evaluate their detection performance. Comparisons and quantitative analysis are given in the chapter.
- Conclusions and recommended future work are given in Chapter 5.

Chapter 2

Review of Existing Variational Methods

2.1 Introduction

In the last twenty years, there is a significant increase in the applications of variational methods in image processing and computer vision [48, 49]. From biomedical images such as those from ultrasound scans, computed tomography (CT) and Magnetic Resonance Imaging (MRI), to the surveillance videos in security systems, variational methods have been widely used for image segmentation [33, 50, 51, 52, 53], boundary detection [30, 31, 32], shape recovery [54, 55], object localization [56, 17], and visual tracking tasks [10, 22, 57]. Generally, the variational methods can be roughly divided into two categories [10]:

- The **boundary-based** (also referred to as edge-based) methods. They depend on the gradient information of object boundaries to perform the detection. Typical algorithms include the active contour method [29], the balloons [58], the level sets [31], and the geodesic active contours [32].
- The **region-based** methods. They depend on the information of the entire region

such as color, texture and motion properties to perform the detection. Typical algorithms include the active contour without edges [30], and the geodesic active region method [59].

In the following, a review of recent developments in both boundary-based and region-based variational methods is given in sections 2.2 and 2.3. The boundary-based and the region-based variational methods are compared with each other in section 2.4. As this thesis focuses on the boundary-based approaches, the problems encountered when using the boundary-based variational methods to detect the object boundary are discussed in section 2.5. Finally, this chapter is summarized and concluded in section 2.6.

2.2 Boundary-based Variational Methods

The boundary-based variational methods can provide efficient and reliable information of the object boundary which does not depend on the object position, color or shape information. Hence, they have been widely adopted for image segmentation and object detection. The boundary-based variational methods can be further divided into two types: the explicit (parametric) contour representation methods and the implicit (geometric) contour representation methods.

2.2.1 Explicit Contour Representations

The active contour (also known as the snake) method [29] is a well-known explicit contour representation method (with over 3000 citations to date [60]). It represents the contour by a set of discrete marker points (as shown in Figure 2.1) and evolves them over time. The curve is driven towards the desired object boundary by minimizing an energy function.

Let $\mathcal{C}(q) : [0, 1] \rightarrow \mathbf{R}^2$ be a parametrized planar curve and let $I : [0, a] \times [0, b] \rightarrow \mathbf{R}^2$

Figure 2.1: The active contour method represents the contour by a set of discrete marker points [1].

be a given image. The energy function of the snake method is defined as

$$E(\mathcal{C}) = \alpha \int_0^1 E_{int}(\mathcal{C}(q))dq + \beta \int_0^1 E_{img}(\mathcal{C}(q))dq + \gamma \int_0^1 E_{con}(\mathcal{C}(q))dq \quad (2.1)$$

where the internal energy term, E_{int} , constrains the curve to be regular and smooth, the image term, E_{img} , attracts the contour to the desired image features, and the constraint term, E_{con} , constrains the solution space. Each term creates its own potential field and the contour actively adjusts its position and shape until it reaches a minimum.

- The **internal** energy term, E_{int} , stands for the curve regularity and imposes a smoothness constraint [28] given by

$$E_{int}(\mathcal{C}(q)) = w_{tension}(\mathcal{C}(q)) \left| \frac{\partial \mathcal{C}}{\partial q}(q) \right|^2 + W_{stiffness}(\mathcal{C}(q)) \left| \frac{\partial^2 \mathcal{C}}{\partial q^2}(q) \right|^2 \quad (2.2)$$

where the first order term makes the snake behaving like a membrane (i.e. resisting stretching), and the second order term makes the snake acting like a thin plate (i.e. resisting bending). The symbol $|\cdot|$ calculates the absolute value of the variable. The weights $w_{tension}$ and $w_{stiffness}$ control the relative importance of the two terms, where $w_{tension}$ regulates the tension and $w_{stiffness}$ regulates the snake rigidity [28].

- The **image** energy term, E_{img} , is derived from the observed data, where the snake may be attracted to *lines*, *edges* or *terminations* given by

$$E_{img}(\mathcal{C}(q)) = w_{line}E_{line}(\mathcal{C}(q)) + w_{edge}E_{edge}(\mathcal{C}(q)) + w_{term}E_{term}(\mathcal{C}(q)) \quad (2.3)$$

where $\{w_{line}, w_{edge}, w_{term}\}$ are weight constants [28]. In most cases, the *line* term is defined by the image function

$$E_{line}(\mathcal{C}(q)) = I(\mathcal{C}(q)) \quad (2.4)$$

The *edge* term is defined by

$$E_{edge}(\mathcal{C}(q)) = |\nabla I(\mathcal{C}(q))|^2 \quad (2.5)$$

which attracts the snake to large image gradients (usually w_{edge} is negative). The symbol ∇ is the grad operator which calculates the gradient of an image. Finally, the *termination* term allows terminations (i.e. free ends of lines) or corners to be developed by the snake. It can be defined as

$$E_{term}(\mathcal{C}(q)) = \frac{\partial \theta(\mathcal{C}(q))}{\partial n_t(\mathcal{C}(q))} \quad (2.6)$$

where $\theta(\mathcal{C}(q))$ is the gradient direction along the snake and $n_t(\mathcal{C}(q))$ is the unit vector perpendicular to the gradient direction [28].

- Finally, the **constraint** term, E_{con} , is usually derived from user-defined constraints that are imposed by the nature of the application, and this may come from high-level knowledge.

Many efforts have been made to improve the robustness and stability of the active contour method [61]. Towards this end, **balloons** are proposed by Cohen [58] which

introduces an additional “balloon force”, E_{bal} , into the snake energy function

$$E(\mathcal{C}) = \int_0^1 [\alpha E_{int}(\mathcal{C}(q)) + \beta E_{img}(\mathcal{C}(q)) + \gamma E_{con}(\mathcal{C}(q))] + \delta E_{bal}(\mathcal{C}(q)) dq \quad (2.7)$$

where

$$E_{bal}(\mathcal{C}(q)) = \frac{\partial \mathcal{C}}{\partial q}(q) \mathcal{C}(q) \quad (2.8)$$

This new term E_{bal} is an anisotropic pressure potential that controls the evolution of the area enclosed by the model and it can either inflate or deflate the contour due to the sign of δ . This force helps the active contour to pass over spurious isolated weak image edges, and counters its tendency to shrink. The resulting snake is more robust with respect to initial curve position and image noise. However, this additional term reduces the generality of the snake. In a practical application, one needs to know beforehand whether the initial contour is located inside the object of interest or whether it encloses the object. Moreover, the strength of the balloon force may be difficult to set, since it must be large enough to overcome weak edges and noise, but small enough so that it will not overwhelm legitimate edge forces [28].

A significant advancement in the active contour method is the **GVF snake** [2] which employs a new class of external forces called the gradient vector flow (GVF) field for the active contour method. Let us define an *edge* map $f(x, y)$ derived from the image $I(x, y)$ having the property that is larger near the image edges

$$f(x, y) = -E_{ext}(x, y) \quad (2.9)$$

where the external energy term can be of the following two forms

$$E_{ext}^1(x, y) = -|\nabla I(x, y)|^2 \quad (2.10)$$

or

$$E_{ext}^2(x, y) = -|\nabla(G_\sigma(x, y) * I(x, y))|^2 \quad (2.11)$$

(a) External force field of traditional snake. (b) External force field of GVF snake.

Figure 2.2: Comparison of the external force field of traditional snake and the GVF snake [2]. It can be seen that the GVF field has a much larger capture range.

where G_σ is a Gaussian filter with standard deviation σ , and symbol $*$ represents a convolution.

The gradient vector flow (GVF) field is the vector field $\mathbf{v}(x, y) = (u(x, y), v(x, y))$ that minimizes the energy functional

$$\mathcal{E} = \int \int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 dx dy \quad (2.12)$$

where u_x and v_x are the first order derivative with respect to x , u_y and v_y are the first order derivative with respect to y .

When $|\nabla f|$ is small, the energy is dominated by partial derivatives of the vector field, yielding a smooth field. On the other hand, when $|\nabla f|$ is large, the second term dominates the integral, and is minimized by setting $\mathbf{v} = \nabla f$. The parameter μ is a regularization parameter governing the tradeoff between the first and the second terms. The advantages are that it can capture a snake from a large spatial range - from either side of the object boundary, and it can force the contour curve into the concave regions. Figure 2.2 compares the the external force field of the snake method with the GVF external force field of GVF snake. It can be seen that the GVF field has a much larger capture range than the traditional snake. A large capture range means that the curve

can be evolved from a relative faraway position to the object boundary. Thus the active contour method can be initialized more freely in the image instead of restricting it to be initialized near the object boundary.

2.2.2 Implicit Contour Representations

The main drawbacks of explicit contour representation methods are that, they cannot cope with the topological changes of object shapes, and they depend on an initial position which must be close to the target object [62]. Hence, in order to solve these problems, an alternative approach which can implicitly represent the contour is proposed [63, 64]. It embeds the initial contour in terms of a surface, i.e. a signed distance function. Instead of evolving the contour, the embedding surface is moved to track the motion of the contour. The corresponding contour at a given time is obtained by determining the zero level set of the evolving surface. In this way, this approach can handle cusps, sharp corners, and automatic topological changes of the evolving curves [65, 66]. The level set method [36, 31] is one of the famous implicit contour representation method.

Let $\mathcal{C}(\mathbf{p}, t)$, defined as $\{x(\mathbf{p}, t), y(\mathbf{p}, t)\}$, denote a family of closed contours generated by moving an initial contour $\mathcal{C}_0(\mathbf{p})$ in the direction of its Euclidean normal inward vector $\mathcal{N}$ as shown in Figure 2.3, and let $I: [0, a] \times [0, b] \rightarrow \mathbb{R}^+$ denote a given image [10, 67]. Assuming the speed of the curve movement is a scalar function F of the curvature $\mathcal{K}$, the curve evolution equation can be written as

$$\begin{cases} \mathcal{C}_t(\mathbf{p}, t) = F(\mathcal{K})\mathcal{N} \\ \mathcal{C}(\mathbf{p}, 0) = \mathcal{C}_0(\mathbf{p}) \end{cases} \quad (2.13)$$

To solve the partial differential equation Eqn. (2.13), the **level set method**, developed by Osher and Sethian [36, 1], represents the contour $\mathcal{C}(\mathbf{p}, t)$ implicitly as the zero level set of a smooth, continuous scalar function $\Phi(x, y, t)$, known as the

Figure 2.3: Level set formulation of front motion: the curve (darkened line) and the corresponding surface $\Phi(x, y, t)$ at different time t [1].

level set function, where $(x, y) \in \mathcal{R}^2$. The implicit contour at any time t is given by $\mathcal{C}(\mathbf{p}, t) = \{x, y | \Phi(x, y, t) = 0\}$. Taking derivative of $\Phi(x, y, t) = 0$ with respect to time and space, Eqn. (2.14) can be derived

$$\begin{cases} \frac{\partial \Phi}{\partial t} = -F(\mathcal{K})|\nabla \Phi| \\ \Phi(\mathcal{C}_0(\mathbf{p}), 0) = 0 \end{cases} \quad (2.14)$$

where ∇ is a gradient operator and $|\nabla \Phi|$ denotes the gradient norm. Thus, a connection is established between the family of the moving curves $\mathcal{C}(\mathbf{p}, t)$ and the family of evolving level set surfaces $\Phi(x, y, t)$.

To apply the above level set model for object boundary detection, the speed function $F(\mathcal{K})$ should be multiplied with a stopping criterion. This stopping criterion is determined by image gradient and is often defined in the following form

$$g(I) = \frac{1}{1 + |\nabla G_\sigma * I|^p} \quad p = 1, 2 \quad (2.15)$$

where G_σ is a Gaussian filter with standard deviation being σ . Parameter p can be set to 1 or 2 which gives two different options for $g(I)$.

Caselles *et. al.* [68] and Malladi *et. al.* [31] proposed the **geometric active contour model**, which can also be considered as **application of the level set method on images**. For a speed function of the form $F = c + \kappa$, where c is a constant force that pushes the curve towards the object, and κ is the curvature of the level set function ϕ ,

$$\kappa = \mathbf{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \quad (2.16)$$

and the geometric active contour model is defined as

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| g(I) \left(c + \mathbf{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right) = g(I)c|\nabla \phi| + g(I)\kappa|\nabla \phi| \quad (2.17)$$

Later, Caselles *et. al.* [32] improved the geometric active contour model and proposed a **geodesic active contour model** based on the relation of classical snake and the geometric active contours. They proved that for a particular case (based on the finding from [32], that is when parameter $\beta = 0$ in Eqn. (2.1)), minimizing the classical snake energy function is equivalent to finding a geodesic curve in a Riemannian space with a metric derived from the image.

Let $\mathcal{C}(q) : [0, 1] \rightarrow \mathbf{R}^2$ be a parameterized planar curve and let $I : [0, a] \times [0, b] \rightarrow \mathbf{R}^+$ be a given image, in a particular case (based on the finding from [32], when parameter $\beta = 0$ in Eqn. (2.1)), minimizing the snake energy function of Eqn. (2.1) is equivalent to minimizing of Eqn. (2.18)

$$\min \int_0^1 g(|\nabla I(\mathcal{C}(q))|) |\mathcal{C}'(q)| dq \quad (2.18)$$

As stated in [32], computing the Euler-Lagrange of Eqn. (2.18) and searching in

the gradient descent direction to find its minimum value, and hence we have

$$\frac{\partial \mathcal{C}(t)}{\partial t} = g(I)\kappa \vec{N} - (\nabla g \cdot \vec{N})\vec{N} \quad (2.19)$$

where κ is the Euclidean curvature, $\vec{N}$ is the unit inward normal to the curve. By introducing the level set formulation, the curve $\mathcal{C}$ can be implicitly represented by the zero level set of a function $\phi : [0, a] \times [0, b] \rightarrow \mathbf{R}$. In this way, the curve evolution process of the snake method can be rewritten as in Eqn. (2.20)

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \operatorname{div} \left(g(I) \frac{\nabla \phi}{|\nabla \phi|} \right) = g(I)|\nabla \phi| \kappa + \nabla g(I) \cdot \nabla \phi \quad (2.20)$$

Comparing Eqns. (2.17) and (2.20), it can be seen that a $g(I)c|\nabla \phi|$ term which acts as an “area constraint” is in Eqn. (2.17) but not in Eqn. (2.20). Whereas a $\nabla g(I) \cdot \nabla \phi$ term which attracts the curve to the boundary with large gradient variations or small gaps, is in Eqn. (2.20) but not in Eqn. (2.17). Hence, Caselles added the $\nabla g(I) \cdot \nabla \phi$ term into Eqn. (2.17) and proposed the **geodesic active contour model**

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \operatorname{div} \left(g(I) \frac{\nabla \phi}{|\nabla \phi|} \right) + c g(I)|\nabla \phi| = g(I)(c + \kappa)|\nabla \phi| + \nabla g(I) \cdot \nabla \phi \quad (2.21)$$

The advantages of geodesic active contours are that, firstly, as it inherits the properties of geometric active contour model, so it can handle topology changes and yet still easy to implement. Secondly, it can detect boundary reliably when the gradients exhibit small variations or small gaps.

2.2.3 Explicit vs. Implicit Contour Representations

Both explicit and implicit contour representations are widely used for object boundary detection. However, due to their different numerical implementations, they have their own advantages and disadvantages. They are compared with each other from 5 aspects in table 2.1 to show their characteristics.

Table 2.1: Comparison of advantages and disadvantages of explicit and implicit contour representations.

	Explicit representation	Implicit representation
Topological ability	Does not allow the evolving contour to undergo topological changes such as contour curve merging or splitting.	Can successfully handle topological changes and segmentation of several objects or multiple-connected objects is straightforward.
Numerical implementation	Requires a delicate regriding (or reparameterization) process to avoid self-intersection and overlap of marker points.	Does not need regriding process. Is capable to extract contour in cusps, sharp corners and protrusions.
Computational cost	Much faster than implicit representation method as it only needs to update a fairly small number of points.	Slower due to updating the full 2D signed distance function (or at least a narrow band around the current contour).
Initialization	Depends on the initialization. Can only find the boundary from an initial position close to the object boundary.	Less sensitive to the initialization problem and can converge to the desired boundary from an initialization far-away from the object boundary.
Shape representation	Represents a contour with marker points. The training shapes need to be carefully aligned to the same orientation, translation and scale when performing the shape analysis or statistical modeling.	Embeds the curve as the zero level set of a high dimensional surface (there is much information redundancy of object shape). Tolerant to slight misalignment of object shapes.

2.3 Region-based Variational Methods

In the previous section, we reviewed the boundary-based variational methods which essentially depend on the object boundary information (i.e. gradients) to drive the evolving curve. Yet, sometimes the properties of enclosed area of interested objects, such as color, texture or region of coherent motion, can be very useful [69, 70]. Hence, in this section, the region-based variational methods are reviewed.

Let $P(\mathcal{R}) = \{\mathcal{R}_A, \mathcal{R}_B\}$ denote an image partitioned into two regions. Two regions are defined by a curve on the image domain as two non-overlapping regions where $\mathcal{R}_A$ is the interior and $\mathcal{R}_B$ is the exterior of the curve region. Let $\{\partial\mathcal{R}\}$ be the boundaries between the different regions. Chakraborty *et. al.* [56] proposed a deformable object model for medical image segmentation that integrates edge and region information.

Their objective function is

$$E(\partial\mathcal{R}) = \alpha \underbrace{E_{prior}(\partial\mathcal{R})}_{PriorTerm} + \beta \underbrace{\int_0^1 I_G(\partial\mathcal{R}(f))df}_{BoundaryTerm} + \gamma \underbrace{\iint_{\mathcal{R}_A} I_R(x, y)dxdy}_{RegionTerm} \quad (2.22)$$

where I_G is an image that contains the gradient values, and I_R is an image that accounts for the region information. $\{\partial\mathcal{R}(f)\}$ is a parameterization of the boundaries using Fourier descriptors. In Eqn. (2.22), the prior term introduces the prior knowledge about the shape, the boundary term propagates the curve towards the high gradient values points (edges), and the region term incorporates some region-based information like the intensity homogeneity. Although this model suffers from drawbacks such as being sensitive to the initial position and not being able to handle topological changes, it is the first attempt to combine the boundary and region information together within an active contour framework [28].

Another well-known region-based variational method is the active region model proposed by Ivins *et. al.* [71]. They introduce a pressure force that is a function of the statistical characteristics of image data and shrinks/expands the curve until it encounters pixels that lie outside the user-defined limits relative to the statistical characteristics of a seed region [28]. When these limits are violated, the pressure force is reversed to make the model contract. The corresponding objective function is

$$E(\partial\mathcal{R}(c)) = \alpha \int_0^1 |\partial\dot{\mathcal{R}}(c)|^2 dc + \beta \int_0^1 |\partial\ddot{\mathcal{R}}(c)|^2 dc + \gamma \iint_{\mathcal{R}} G(I(x, y))dxdy \quad (2.23)$$

where the function $G(\cdot)$ measures the total energy of the interior curve energy, $\{\partial\mathcal{R}(c)\}$ is a parameterization of the boundaries using a set of control points and $\{\partial\dot{\mathcal{R}}, \partial\ddot{\mathcal{R}}\}$ denote the first and second order partial derivative of the curve with respect to its parameter c . The weakness of this active region model is that, the final result will show a bias towards the interior of the enclosed area of the segmented object.

Mumford and Shah [72] proposed to approximate a given input image I with a

piecewise smooth function u by minimizing the functional

$$E(u, \mathcal{C}) = \frac{1}{2} \int_{\Omega} (I - u)^2 dx + \frac{\lambda^2}{2} \int_{\Omega - \mathcal{C}} |\nabla I|^2 dx + \nu |\mathcal{C}| \quad (2.24)$$

simultaneously with respect to the image u and the contour $\mathcal{C}$ [73]. Parameter Ω represents the image domain, and parameter ν is a weight. The first term is a fidelity term which enforces the function u be similar to the input image I . The second term enforces smoothness of the segmented image but permits discontinuities of u across a boundary denoted by $\mathcal{C}$. The last term gives the one-dimensional Hausdorff measure of the length of the boundary [73]. Similar models can be found for the discrete case in a Markov random field by Geman [74] and the weak membrane model by Blake [75, 76].

In the case that $\lambda \rightarrow \infty$, Eqn. (2.24) is reduced to the *cartoon limit* [72].

$$E(u, \mathcal{C}) = \frac{1}{2} \sum_i \int_{R_i} (I - u_i)^2 dx + \nu |\mathcal{C}| \quad (2.25)$$

Minimization of Eqn. (2.25) leads to an approximation of the input image I by a function u which is piecewise constant on a set of regions R_i separated by the boundary set $\mathcal{C}$, where the constants u_i are the mean grey intensity value in each region R_i [73]:

$$u_i = \frac{1}{|R_i|} \int_{R_i} I dx \quad (2.26)$$

This simplified model in Eqn. (2.25) provides a segmentation for which the grey intensity value in each region R_i is approximated by a constant u_i . Instead of approximating by a constant, one can more generally approximate the intensity in each region R_i by a probabilistic model $P(I(x)|\theta_i)$ with a parameter θ_i [73]. The energy density for region R_i in Eqn. (2.25) is then replaced by the negative log-likelihood of a grey value $I(x)$ at point x ,

$$E(\{\theta_i\}, \mathcal{C}) = -\frac{1}{2} \sum_i \int_{R_i} \log P(I(x)|\theta_i) dx + \nu |\mathcal{C}| \quad (2.27)$$

This variational model in Eqn. (2.27) is proposed by Zhu [77] in their work on *region competition*. They employed a Gaussian probability for each region R_i . Thus, not only a different mean u_i but also a different standard deviation σ_i are fitted to each region

$$P(I(x)|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(I - u_i)^2}{2\sigma_i^2}\right), \quad \text{where } \theta_i = \{u_i, \sigma_i\} \quad (2.28)$$

The resulting segmentation process is then able to separate regions which have the same mean but different variances.

There are a number of region-based variational methods that are based on the Mumford-Shah functional [78]. Chan and Vese proposed the active contours without edges method [30]. It can detect objects whose boundaries are not necessarily defined by gradient nor being smooth. Paragios proposed a geodesic active region model which aimed at finding a partition where the interior and the exterior region preserved the desired region properties [33, 59, 79]. Cremers proposed a diffusion snake which was a hybrid model combining the Mumford-Shah functional and the snake functional [35]. The other methods are the binary level set method [80] and the region-based perceptual features method [81].

2.4 Boundary-based vs. Region-based Variational Methods

By comparing the region-based and the boundary-based variational methods, some fundamental differences between them can be identified:

- Since the properties (e.g. the probabilistic model of color, texture or motion) of the entire region of interested object are utilized, region-based variational methods are usually more robust than the boundary-based variational methods.

- For region-based variational methods, the enclosed region should be homogeneous in grey intensity, color or texture, otherwise they cannot work properly. However, the boundary-based methods are more flexible to complex object regions and backgrounds.
- Region-based variational methods are a little slower than the boundary-based ones due to more involved computation of region properties.

In practice, it is often difficult to find an object having a homogeneous region in real images. For example, when performing image segmentation or object detection in surveillance video sequences, the target object (such as human) may be composed of different subregions due to different color and texture of skin and clothing. Moreover, the properties of these regions may vary due to the changes of viewing angle and lighting. Hence, it is hard to find a unique model to represent the object region, and the region-based variational methods are not suitable for detecting object with complex region interior. Instead, boundary-based methods are more competent for this task. As long as the object maintains a clear boundary from the background, it is able to extract the object contour regardless how complex its interior may be. Moreover, in the case of biomedical applications, since the medical images such as ultrasound often suffer from noise problem, it is difficult to find a homogeneous region in these images. They are hard to be processed by using region-based variational methods. Hence in this thesis, the boundary-based variational methods for object detection are studied.

2.5 Problems of Using Boundary-based Variational Methods

Since the boundary-based variational methods only utilize the boundary information to locate the object, they are very sensitive to the loss of boundary information. Any weak or spurious edge may cause the curve to move away from the real boundary

Figure 2.4: Failure of boundary-based variational method in the presence of missing boundary information. (a) Failure of level set method when encountering weak boundary containing “gaps”. The curve leaked out from the gaps instead of converging to it (the curve iteration process was manually stopped after the leakage problem happened). (b) Failure of level set method when encountering object occlusion. The curve converged to the wrong location.

and converge to the wrong place.

Figure 2.4 shows two examples of the above mentioned problems of boundary-based variational methods. In Figure 2.4 (a), part of the object boundary is destroyed, leaving small “gaps” on the boundary. For a contour curve starting within the object region, the expanding curve of the boundary-based variational method could not find the correct object contour and leaked out from the gaps. In Figure 2.4 (b), the target object is partially occluded by another object. It can be seen that part of the original object boundary is missing due to the occlusion. Hence, the shrinking curve started by enclosing the object was “deceived” by the edges of the occluding object and converged to a wrong location.

The above two examples belong to the “**missing boundary**” problem as the original object boundary information is missing due to either boundary blur, background noise or object occlusion. Under these circumstances, the object boundary cannot provide reliable information and the evolving curve often converges to a wrong position. Hence, the “missing boundary” problem is the main obstacle to boundary-based variational methods for robust object detection. In order to improve their detection performance, additional boundary constraints are incorporated into the variational

methods to force the curve to stop at the desired boundary of interests. In the following chapters, new algorithms which incorporate additional boundary constraints, such as the temporal information and the shape prior information, as regularizers into the variational methods to constrain the curve evolution process are developed and evaluated. These shape regularizers can help to reconstruct the missing boundary information and recover the original object shapes. The overall detection performance of boundary-based variational methods can be therefore greatly improved.

2.6 Concluding Remarks

In this chapter, we reviewed recent developments in variational methods for image segmentation and object detection in computer vision.

The variational methods can be divided into two categories: the boundary-based and the region-based. The boundary-based methods can be further divided into the explicit and implicit contour representation methods. Typical methods of the explicit representations include the active contour method, the balloons and the gradient vector flow (GVF) method. Typical methods of the implicit representations are the level set method, the geodesic active contour method. Typical algorithms of the region-based variational methods are the active contour without edges, the geodesic active region method.

The region-based variational methods usually are more robust than the boundary-based ones. However, they require the enclosed area of an object to be homogeneous. Besides, they are slower than the boundary-based variational methods. Hence, in this thesis, the boundary-based variational methods are chosen for object detection. However, the boundary-based variational methods usually performs poorly when encountering the “missing boundary” problem. Additional boundary constraints are needed to improve their detection performance. The ways these additional constraints are imposed into the variational framework are new contributions of this thesis.

Chapter 3

Regularizing Variational Methods by Adding Temporal Information

In this chapter, a new method is presented which utilizes temporal information as a regularizer into the variational methods. This method is proposed to deal with the “missing boundary” problem when detecting the inner heart wall boundary from echocardiographic image sequence.

The echocardiographic (heart ultrasound) image sequence suffers from signal dropouts which weakens or totally destroys the registration of the ventricular boundary in the ultrasound images. When a variational method is applied to detect the inner heart wall boundary from echocardiographic data, it may fail to find the original boundary and converge to the wrong location. In order to solve this problem caused by “missing boundary” problem, a new method to integrate the temporal information into the variational frameworks is proposed to regularize the curve evolution process.

The novelty of this method is as follows. The ventricular boundary is quantitatively partitioned and classified into strong and weak segments. The weak segments are considered to be disrupted by signal dropout and thus there is low confidence on the presence of heart wall boundaries. Hence, temporal information from neighboring

frames is exploited and is used as a regularizer into variational methods to reconstruct the missing boundary and to strengthen the image gradient force of the weak segments. Two variational methods are studied here and they are the active contour method and the level set method. Temporal information is proposed to be incorporated into the external energy of the active contour method, and the speed function of the level set method.

This chapter is arranged as follows. In section 3.1, the background that motivates the research is given. The signal dropout problem in echocardiography which causes the missing boundary problem is discussed, and related work is reviewed. Section 3.2 presents the incorporation of temporal information into the variational framework. Block matching and Gaussian Mixture Model are used in the proposed method to identify the weak boundary segments. A weight and a strengthening factor are developed to control the process of boundary recovery from neighboring frames. Experimental results of the proposed methods are presented in section 3.3. Finally, the work in this chapter is concluded in section 3.4.

3.1 Background

In modern medicine, many medical imaging techniques such as Positron Emission Tomography (PET), Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and ultrasound have played an important role in diagnosis of diseases. Echocardiography, a sonic method utilizing ultrasound, is one of the widely used imaging utilities for clinical assessment of the heart function [82] due to its many attractive properties such as low cost, portability, minimal patient discomfort and real-time capture and display [16, 83]. An echocardiographic image is produced by the ultrasound scanning of the heart. The inner boundary of the left ventricle (LV) in echocardiographic images is an important part to be examined [84, 18, 85]. By extracting the ventricular contour, the shape, volume and the movements of the heart chamber can be

Figure 3.1: The dropouts phenomenon in consecutive 4 frames in a echocardiographic sequence (signal dropouts in the two middle frames).

estimated, and certain heart diseases, such as ischaemia, can be diagnosed. However, the traditional way to extract LV boundary is by a clinician manually delineating it on every image frame of the echocardiographic sequence and computing relevant information from it [86, 83]. This manual process is tedious and time-consuming and is subject to clinician variability. It is desirable to use an automatic or semi-automatic computer-aided method to extract the inner heart wall boundary during the clinical examination so as to assist the doctor in the diagnosis of heart diseases.

3.1.1 The Dropout Problem in Echocardiography

As echocardiography employs a sonic method which generates the images by a probe scanning from a constant angle towards the heart [87], this imaging modality produces relatively poor image quality [44]. A typical image degradations is the signal dropout problem [82, 84]. Dropouts occur at places where the endocardium is nearly parallel to the direction of the travel of acoustic wavefront [15]. Therefore, little ultrasonic signal is reflected back towards the probe, resulting in a registration of zero or low echo amplitude. Figure 3.1 shows a dropout phenomenon in four consecutive video frames. In the first and fourth frames, the heart boundary indicated by the arrow registers well, but in the two middle frames, it is lost due to dropout. Since the signal dropout phenomenon usually does not last when imaging a moving target, it is reasonable for us to assume that it will typically occur at a particular location in at most two consecutive image frames.

Figure 3.2: The detection problem of variational methods in the presence of signal dropout in echocardiographic images. (a) Ground truth. (b) The active contour method could not detect the real boundary and converged to a local extremum of gradient within the heart chamber. (c) The level set curve could not be stopped at the weak boundary but leaked out to the outside of the heart chamber.

The noise effects plaguing echocardiographic data make conventional techniques based on edge enhancement such as Sobel or Canny operators inappropriate for ventricular boundary detection [88, 14]. More sophisticated methods are needed. Many methods have been proposed such as simulated annealing [16, 89], Markov random field [14, 90], artificial neural networks [91, 92, 93], multiple active contours [17], optical flow [86, 94], and morphological filters [15]. More recent research work includes the shape-space-based approach [18], the active appearance motion model [19], and the methods based on information fusion [22, 23]. Among all the algorithms reported in literature, variational methods have been used favorably for such a task. Particularly, the active contour method [29] and the level set method [20, 21, 95] have been attempted several times due to their superiority for such a task [96]. Many of their advantages are highlighted in [97, 86]. Hence, in this chapter, these two variational methods are explored for detecting the ventricular contours.

Since signal dropout in echocardiography weakens or totally destroys the registration of ventricular boundary in the ultrasound images, it creates “gaps” on the contour where the original boundary information is lost. Hence, the dropouts in echocardiographic images can be regarded as the “**missing boundary**” problem. The broken ventricular boundary may not be strong enough to constrain the curve evolution, and the curve can be led to the wrong position. Figure 3.2 (b) and (c) show the detec-

tion results by the active contour method and the level set method in the presence of dropouts. Compared with the ground truth in Figure 3.2 (a), it can be seen that the active contour curve could not detect the real boundary and converged to a local extremum of the gradient within the heart chamber. Similarly, the level set curve could not be stopped at the weak boundary and leaked out to the outside of the heart chamber. Both of these methods failed to find the correct ventricular boundary.

3.1.2 Review of Related Work

Since the original boundary information is missing due to dropout, the gradient force of the contour cannot effectively constrain the curve evolution. Research that address this problem has been reported in which information other than the gradient is used to define a special function to constrain the curve [98, 99]. Lin [100] employs a Generalized Fuzzy Operator-based speed term to evolve the curve, Ho [101] replaces the constant propagation term by a signed local statistical force generated from region competition, and Jin [102] uses a constraint force computed from a local homogeneity measurement. However, the image features used in the above methods to enforce the constraints have to be homogeneous and distinctive. Thus, these methods can only work under certain conditions. They are hard to be applied to complicated cases.

Hence, methods that fuse the gradient-based constraint with a regularizer have been reported. One type of regularizers is based on region information [103]. In [10], the level set function is combined with a motion detection term which constrains the evolving curve to detect objects only in the moving area. In [104], a region term which is obtained from fuzzy clustering is added to the speed function. Chan and Vese [30] utilize the region information and propose a new stopping term which is based on the Mumford-Shah functional [72]. In [105], the area information is used by combining it as a weighted area minimizing flow. A combination of methods in [105] and [30] is reported in [106]. However, a few special requirements such as the grey level uniformity within the region of interest have to be met. Therefore, region regularizer is not suitable to

handle more complicated cases such as images containing high level of noise or textured non-uniform regions. Another type of regularizer is based on shape prior information [107, 108]. Leventon [46] constructs a shape prior model from a set of training shapes using principle component analysis, which is then combined with the level set function as a global shape constraint. A prior shape knowledge generated by modeling the distribution of level set representation of object shapes with a Gaussian function is used in [21]. Although these prior models can provide reliable shape constraints, they need a training process beforehand to learn the specific characteristics of object shapes. Although the authors claim in [109] that their shape prior model does not need a training process, they still need to pre-assume the object as having a specified shape, such as an ellipse. Therefore, their method can only regulate the evolving curve to the assumed shapes and cannot accommodate large shape variations.

As the ventricular boundary may undergo large deformation in size and shape throughout an echocardiographic image sequence, and its internal region may vary in gray level owing to noise in ultrasound, it is difficult to be processed using either region or shape regularizers. Therefore, what is needed is a better variational method which can effectively solve the “missing boundary” problem in echocardiographic data without using elaborated shape priors.

3.1.3 Motivations and Objectives

It is noticed that for object detection from image sequences, temporal continuity of object boundary can be extremely useful for improving the detection performance [42, 43, 44]. The objects in adjacent frames usually are similar to each other, making them highly correlated [45]. If the ventricular boundary is disrupted by signal dropout in the current frame, it is possible to reconstruct the missing part by using temporal information from its neighboring frames. In this way, the failure of variational methods can be prevented and no off-line training process is needed. The above considerations lead us to explore the abundant information available in temporal domain [110].

Among the currently reported work for combining temporal information with the variational methods, Pardas [116] proposed to embed a compensation error term generated from motion estimation of object position in adjacent frames into the external energy of active contour function. Lin [117] employed global motion prediction and local motion estimation to estimate the new location of curves. However, the way to incorporate temporal information into the active contour framework in Pardas's and Lin's work is not robust enough to handle the missing boundary information and the image noise. When dealing with poor quality echocardiographic data containing lots of dropouts and speckle noise, these methods cannot work well.

In this chapter, a new method to incorporate temporal information into variational methods as a curve evolution regularizer is proposed. It aims at solving the missing boundary problem caused by signal dropout when detecting the ventricular boundary from echocardiographic image sequence. The ventricular boundary is quantitatively partitioned and classified into strong and weak segments. The weak segments are considered to be caused by signal dropout in ultrasound and there is low confidence on the presence of boundary. Temporal information from neighboring frames is utilized and is embedded in the variational framework as a regularizer to reconstruct the missing boundary in the weak segments. Hence, the evolving curve will be constrained to stop at the weak segments, and the overall boundary detection can be improved.

3.2 Incorporation of Temporal Information into Variational Methods as a Regularizer

Before incorporating the temporal information into variational methods, the inner heart wall boundary needs to be quantitatively identified into strong and weak segments first. This is done in two steps: a *block matching* step to roughly locate the position of moving ventricular boundary, and an *identification* step to identify the strong / weak segments using Gaussian Mixture Model (GMM).

Figure 3.3: Illustration of the block matching process.

3.2.1 Block Matching

Block matching is an efficient motion estimation method for finding the approximate location of object in motion. It is employed to predict the location of ventricular boundary in the current image frame based on its detection from the previous frame (except for the first frame of a sequence). As long as the detection from block matching is roughly close to the real boundary, the subsequent steps can be successfully performed. Minor degree of erroneous matching can be tolerated in the proposed method due to the narrow band used in the second step.

Each point on the detected contour from the previous frame is used as the center of a rectangular block, a search window is defined centered around the block. All blocks in the search window are checked to find the best match with the rectangular block. The center of the matched block is the predicted new location of the ventricular boundary in the current frame. This block matching process is depicted in Figure 3.3. The block matching algorithm employed is a *Three-Step Search* method [111], and the matching

criterion is the *mean squared error* (MSE) defined by Eqn.(3.1).

$$\text{MSE}(k, l; u, v) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I_{r-1}(k+i, l+j) - I_r(u+i, v+j)]^2 \quad (3.1)$$

where M and N define the size of the matching block. Parameters k and l are the coordinates of the first pixel of the matching block in the previous frame I_{r-1} . u and v are the coordinates of the first pixel of the matching block in the current frame I_r .

3.2.2 Identifying Strong / Weak Segments of Inner Heart Wall Boundary Using GMM

Once the predicted contour from block matching is obtained, it needs to be quantitatively identified into strong and weak segments. We propose to employ the *Gaussian Mixture Model* (GMM) [112] to compute the data dependent thresholds that separate them.

The Gaussian mixture model (GMM) is a popular method for representing an arbitrarily complex probability density function of the data. It comprises a number of component Gaussian functions which are then combined with different weights to approximate a multi-modal density. The advantages of using GMM in that, it can help us to automatically find the thresholds partitioning the contour into strong and weak segments. Thus the reconstruction of boundary information in weak segments does not depend on ad-hoc thresholds determination, but rather depends on data dependent thresholds determined by a more sophisticated density estimation. Furthermore, the GMM [112] used is an unsupervised method. We do not need to give the “label” of each training sample. All the data will be clustered automatically based on their actual distribution.

The points within a narrow band around the predicted contour are chosen to be the training data $\mathcal{X}$ for GMM (see Figure 3.4). Their image energy, which is the

absolute value of gradient intensity as defined in Eqn. (3.2), is chosen to be the feature representing the properties of points.

$$E_{image}(x, y) = |\nabla G_\sigma * I(x, y)| \quad (3.2)$$

where G_σ is a Gaussian filter with standard deviation of σ . These points come from two categories: the points belonging to the contour and the points belonging to the background. Hence, they can be assumed to follow two different types of distributions. The distribution of contour points can be considered to follow a full Gaussian. However, since E_{image} is always above zero, the distribution of background points can only be assumed as a half Gaussian. In order to obtain a full Gaussian-like distribution of the background points, the entire training data are reflected to the negative x axis. Hence, the distribution of contour points will be duplicated in the negative x axis. As a result, there will be totally three Gaussian components in our training data. Figure 3.5 shows the result of using GMM to model the distribution of the training data. It can be seen that three Gaussian components are obtained. Hence, the probability density function of the training data can be written as

$$p(\mathcal{X}'|\Theta) = \sum_{m=1}^3 \alpha_m p(\mathcal{X}' | \theta_m) \quad (3.3)$$

where $\mathcal{X}'$ is the original training data selected from the narrow band plus the duplicate of this data to the negative axis. Parameter $\alpha_m, m = 1, 2, 3$ are the mixing probabilities which satisfy

$$\alpha_m \geq 0, \quad m = 1, 2, 3, \quad \sum_{m=1}^3 \alpha_m = 1 \quad (3.4)$$

and each θ_m is the set of parameters defining the m -th Gaussian component $\theta_m = (\mu_m, \sigma_m)$, and $\Theta = (\alpha_m, \mu_m, \sigma_m), m = 1, 2, 3$ is the complete set of parameters specifying the mixture model.

To find the thresholds separating the strong and weak segments, let us take a look

Figure 3.4: Select the training data. The points around the predicted position of ventricular boundary are selected as training data.

Figure 3.5: Using GMM to approximate the distribution of training data. g_1 , g_2 and g_3 are three components of GMM. g_1 corresponds to the distribution of contour points, g_2 corresponds to the distribution of background points, and g_3 is a duplicate of g_1 in the negative x axis.

at Figure 3.5 (b). Component g_1 has a larger mean and variance, it is considered to correspond to the distribution of contour points. Component g_2 has a smaller variance and centered at the origin, it is considered to correspond to the background points. Component g_3 is a duplicate of g_1 in the negative x axis. Thus, if the image energy of any point on the predicted contour is above the mean value of component g_1 , it can be considered to belong to the “strong” segments of the contour, which means the gradients on the boundary are strong enough to constrain the curve evolution

without any temporal information. Therefore, the mean value μ_1 of g_1 is chosen to be the first threshold T_1 . The intersection of component g_1 and g_2 is a key point: if the image energy of a point on the predicted contour is below this level, then it is regarded to be more likely from the background as it has a larger probability of being within component g_2 . This also implies that the point could have been mixed up with the background even though it actually belongs to the contour. The gradient of this point on the boundary is too weak to be distinguished from the background. In this case, temporal information is fully utilized to recover the original boundary information. Hence, this intersecting point is selected to be the second threshold T_2 . In this way, each point on the predicted contour can be classified into “strong” and “weak” segments by thresholds T_1 and T_2 .

The parameters Θ of GMM is computed using the *Expectation Maximization* (EM) algorithm [113]. There are two main steps in EM: (i) the *expectation step* which calculates the expectation of the complete log-likelihood expression under the current estimate $\hat{\Theta}(t)$. The result is the so-called Q-function

$$\mathcal{Q}(\Theta, \hat{\Theta}(t)) = E[\log p(\mathcal{X}', \mathcal{Z}|\Theta)|\mathcal{X}', \hat{\Theta}(t)] \quad (3.5)$$

where $\mathcal{Z}$ is a set of binary labels associated with the training samples indicating which component each sample belongs to; (ii) the *maximization step* which updates the parameters of GMM at iteration $(t + 1)$ according to

$$\hat{\Theta}(t + 1) = \arg \max_{\Theta} \mathcal{Q}(\Theta, \hat{\Theta}(t)). \quad (3.6)$$

By alternately applying these two steps in each iteration until convergence, the final value of parameters can be found and the two thresholds can be calculated by Eqns. (3.7) and (3.8)

$$T_1 = \mu_1 \quad (3.7)$$

Figure 3.6: Position of weak segments on the contour that need to be strengthened by temporal information.

$$\frac{\alpha_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(T_2-\mu_1)^2}{2\sigma_1^2}} = \frac{\alpha_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(T_2-\mu_2)^2}{2\sigma_2^2}} \quad (3.8)$$

where T_2 is the positive solution of Eqn. (3.8).

Figure 3.6 shows the position of identified weak segments on the contour that need to be strengthened by using temporal information. These segments are obtained by the thresholds calculated from GMM. It can be seen that the positions match very well with the failed detections by active contour method and level set method.

3.2.3 Regularizing Variational Methods by Incorporating Temporal Information

To recover the missing boundary information in weak segments, we propose a novel way to incorporate the temporal information as defined in Eqn. (3.9) into the variational methods

$$E_{time}(x, y) = \left| \nabla \left[\frac{f_{r-1}(x, y) + f_{r+1}(x, y)}{2} \right] \right| \quad (3.9)$$

where f_{r-1} is the previous image frame, f_{r+1} is the next image frame. In fact, it is the temporally interpolated gradient between neighboring frames. E_{time} works as a regularizer that regularizes the deformable curve to stop at the weak segments on the contour instead of converging to wrong position.

As aforementioned, the dropout is assumed to only occur in at most two consecutive image frames. Therefore, the missing boundary in the current frame can mostly be recovered by its neighboring frames, as at least one of its neighboring frames contains the reliable boundary information. However, since the information in the next image frame is used in computing E_{time} , the detection process will lead to one frame delay. To increase the robustness of the algorithm, it is preferable to use more consecutive frames (e.g. previous two and next two frames), and the algorithm will be able to handle dropouts in more consecutive frames. However, the detection process will be delayed a few more frames as a result.

3.2.3.1 Incorporating temporal information into active contour framework

Since the external energy of active contour method is the force that drives the moving curve, we propose to incorporate E_{time} into the external energy of the active contour equation Eqn. (2.1) in the following way

$$E'_{ext}(x, y) = -\delta_2 \left[\delta_1 E_{time}(x, y)^2 + (1 - \delta_1) E_{image}(x, y)^2 \right] \quad (3.10)$$

Parameters δ_1 and δ_2 in Eqn. (3.10) are weight and strengthening factors respectively. Parameter δ_1 is a weight which balances the strength between the image energy E_{image} and the temporal information E_{time} and it controls how much E_{time} should be emphasized. If the image energy E_{image} of a point on the predicted contour is above threshold T_1 , then it will be classified as strong segments. Hence, no extra information is needed, as the point itself is strong enough to constrain the level set curve. Therefore, δ_1 is chosen to be 0. Only the E_{image} term contributes to the speed function, the E_{time} term vanishes. However, if a point on the predicted contour is classified as weak segments (image energy E_{image} is less than threshold T_2), its boundary information may be weakened or destroyed by dropout. Temporal information is used to reconstruct its original boundary structure. In this case, the E_{image} term will be totally replaced by the E_{time} term, and δ_1 is thus set to 1. If E_{image} of a point is between T_1 and T_2 ,

Figure 3.7: Parameters δ_1 and δ_2 that are used to control the incorporation process of temporal information into the variational frameworks.

E_{image} is still useful, and temporal information is added selectively according to the magnitude of E_{image} . For simplicity, we define it to be a linear function of E_{image} . The final expression of δ_1 is defined in Eqn. (3.11) and is depicted in Figure 3.7 (a).

$$\delta_1 = \begin{cases} 0 & \text{if } E_{image} > T_1 \\ 1 & \text{if } E_{image} < T_2 \\ \frac{1}{T_2 - T_1}(E_{image} - T_1) & \text{if } T_2 < E_{image} < T_1 \end{cases} \quad (3.11)$$

Although the missing boundary information can be recovered by temporal information, the contour may still be weaker than what it should be. So δ_2 function is used to boost the total energy $\delta_1 E_{time} + (1 - \delta_1) E_{image}$ close to the average level. In this way, boundary can be detected much easier. Consider that if the image energy E_{image} is above threshold T_1 , then the strength of the total energy is strong enough to constrain the level set curve evolution. It is not necessary to boost it and δ_2 is set to 1. This value is hence chosen to be the average level of the total energy we want to boost. Thus, if the image energy of a point is below threshold T_2 , it is classified as weak segments. Then δ_2 is set to be T_1/T_2 in order to boost the energy to the average level. Finally, if the image energy is between the two thresholds, then δ_2 is defined to

be a reciprocal function of E_{image} so that the energy can be boosted to the average level T_1 . The final expression of δ_2 function is defined in Eqn. (3.12) and is depicted in Figure 3.7 (b).

$$\delta_2 = \begin{cases} 1 & \text{if } E_{image} > T_1 \\ T_1/T_2 & \text{if } E_{image} < T_2 \\ T_1/E_{image} & \text{if } T_2 < E_{image} < T_1 \end{cases} \quad (3.12)$$

3.2.3.2 Incorporating temporal information into level set framework

Since the speed function of level set method is the force that drives the moving curve, we propose to incorporate E_{time} into the stopping criterion $k_I(= g(I))$ of the level set equation (2.17), and hence we have

$$k'_I(x, y) = \exp\{-\delta_2[\delta_1 E_{time}(x, y) + (1 - \delta_1)E_{image}(x, y)]\} \quad (3.13)$$

For a speed function F with the form of $F = c - \epsilon\mathcal{K}$, where c is a balloon force like constant velocity which pushes the curve inwards (or outwards), and $\mathcal{K}$ is the curvature which controls the “smoothness” of the curve, the final expression of the speed function of level set method becomes

$$F'(x, y) = (c - \epsilon\mathcal{K}) \cdot \exp\{-\delta_2[\delta_1 E_{time}(x, y) + (1 - \delta_1)E_{image}(x, y)]\} \quad (3.14)$$

Parameter δ_1 and δ_2 are defined in the same way as in the previous sub-section 3.2.3.1.

Figures 3.8 and 3.9 illustrate an example of stopping criterion $k_I(= g(I))$ in the level set method before and after adding the temporal information. The subfigures on the right are the magnified images of k_I in the circled region. We can see that before adding the temporal information, the contour indicated by arrow is missing due to dropouts. There is a “gap” on the heart boundary which causes the leakage of level set curve. However, by adding the temporal information, the missing part is reconstructed

Figure 3.8: The stopping criterion $k_I(= g(I))$ before adding the temporal information. The black region in the image corresponds to the maximum strength of gradient, while the white region corresponds to the minimum strength of gradient (where the strength of gradient value approximate 0).

Figure 3.9: The stopping criterion $k_I(= g(I))$ after adding the temporal information. The black region in the image corresponds to the maximum strength of gradient, while the white region corresponds to the minimum strength of gradient (where the strength of gradient value approximate 0).

and a clear segment is obtained.

3.3 Experimental Evaluations

In this section, experimental evaluations of the proposed methods are presented. Five echocardiographic sequences obtained from “Atlas of Echocardiography” of Yale University [114] and “e-Echocardiography” of the online echocardiography CME resource [115] are used for testing the methods. They all contain the signal dropout phenomenon. The data are collected by using SIEMENS and Hewlett-Packard Sonos ultrasound machine with the imaging bandwidth varying from 2.5MHz to 3.5MHz. The

image resolution is 320×240 pixels. The data are all mid-papillary level ¹ short axis 1-chamber-view of the left ventricle, each covering one complete cardiac cycle. The length of each image sequence is around one second with 20 to 25 frames. As the data are collected by different ultrasound machines, with different settings and on different persons, they have different levels of noise. They were used to evaluate the robustness of the proposed methods.

The parameters in the active contour method are empirically set to $\alpha = 0.2$ and $\beta = 0.5$. The parameters in the speed function of the level set method are set to: $c = 1.0$ and $\epsilon = 0.02$. The standard deviation of the Gaussian filter is set to 2.5, and the radius of the searching window in block matching is set to 7.

3.3.1 Detection Performance

The proposed active contour method using temporal information is compared with the ground truth drawn by trained clinician, the traditional active contour method, and two other methods (Pardas’s method [116] and Lin’s method [117]) which also incorporate temporal information into the external energy of the active contour function to improve the detection performance. The proposed level set method using temporal information is compared with the ground truth, the conventional level set method, and the active contour without edges method [30] which employs region information as a regularizer to constrain the curve evolution.

3.3.1.1 Active contour method using temporal information

Figure 3.10 shows the detection results of the proposed active contour method using temporal information during the heart contraction period in the first echocardiographic image sequence. The results are compared with the other methods that also

¹The papillary muscle is the small muscles within the heart that anchor the heart valves. “mid-papillary” is a position that the echocardiography is taken around the papillary muscle in the middle of the heart chamber. This position is often used in clinical diagnosis of heart function [140, 141].

incorporate temporal information into the active contour method. Figure 3.11 shows the comparison results during the heart relaxation period in the first sequence. It can be seen from these figures that as the heart boundaries are generally clear during the contraction period, most of the algorithms perform well. However, as parts of the heart boundary are disrupted by dropouts and speckle noise during the relaxation period, the traditional active contour method, Pardas's method, and Lin's method all fail to detect the real boundary. Only the proposed method can correctly find the weak ventricular boundary.

Comparing the detection results, we can see that the contours extracted by the proposed method are much closer to the ground truth than others. The main improvements in the detection results are highlighted using circles as shown in Figure 3.12. It can be seen that in the left circle where the heart contour is disrupted by dropouts, the traditional active contour method cannot detect the weak inner heart wall boundary. However, by incorporating the temporal information to re-create the missing boundary information, the external energy of the active contour function is strengthened. Hence, we are able to correctly detect the boundary. Similarly, in the region that is highlighted by the circle in the bottom left of the images, the speckle noise appears to be very close to the real inner heart wall boundary. It degrades the detection accuracy of the traditional active contour method. However, by utilizing the predicted contour location from block matching, the active contour curve is pulled nearer to the real inner heart wall boundary and away from the local noise. It can converge to the correct boundary instead of stopping at the wrong position. Although Pardas's and Lin's methods also incorporate temporal information into the active contour function, their incorporation methods are sensitive to image degradations and thus show poor detection performance here.

Figures 3.13 and 3.14 show the detection results of the proposed active contour method using temporal information in the second echocardiographic sequence. The results are compared with the ground truth, the traditional active contour method,

Figure 3.10: Comparison of the detection results of the proposed active contour method using temporal information (row 1), with the ground truth (row 2), the traditional active contours (row 3), and Pardas's method (row 4) and Lin's method (row 5) in four of the image frames during the heart contraction period in the first echocardiographic sequence.

Pardas's method and Lin's method. It can be seen that the proposed method still performs better than others. The traditional active contour method and Pardas's

Figure 3.11: Comparison of the detection results of the proposed active contour method using temporal information (row 1), with the ground truth (row 2), the traditional active contours (row 3), and Pardas's method (row 4) and Lin's method (row 5) in four of the image frames during the heart relaxation period in the first echocardiographic sequence.

method fail to detect the ventricular boundaries during the heart relaxation period. Although Lin's method can detect the rough shape of ventricular wall, it cannot detect

Figure 3.12: Detailed illustration of the improvements of our proposed active contour method using temporal information, comparing with the ground truth, the traditional active contour method, Pardas' method and Lin's method.

the papillary muscles as accurately as the proposed method.

3.3.1.2 Level set method using temporal information

The proposed level set method using temporal information is compared with the conventional level set method and the ground truth drawn by trained clinician in the first echocardiographic sequence. Figure 3.15 and Figure 3.16 show the results during the heart contraction and relaxation period, respectively. It can be seen that, as the inner heart wall boundary is clear enough during the contraction period, both the proposed method and the conventional level set method work well. The contours extracted by them are all similar to the ground truth. However, during the heart relaxation period, parts of the ventricular boundary are weakened by dropout, hence the traditional level set method suffers a severe leakage problem and its detection performance is poor. By strengthening the weak segments using temporal information,

Figure 3.13: Comparison of the detection results of our proposed method (row 1) with the ground truth (row 2), the traditional active contour (row 3), Pardas's method (row 4) and Lin's method (row 5) in four image frames during heart contraction period in the second echocardiographic sequence.

we can see that the proposed method can correctly detect the inner heart wall boundary without any leakage. The boundary detected by the proposed method is more similar to the ground truth and it shows a considerable improvement of detection performance.

Figure 3.14: Comparison of the detection results of our proposed method (row 1) with the ground truth (row 2), the traditional active contour (row 3), Pardas's method (row 4) and Lin's method (row 5) in four image frames during heart relaxation period in the second echocardiographic sequence.

To show the efficiency of the proposed method, we then compare it with the active contour without edges method [30] which employs region information as a regularizer for curve evolution. Unlike the traditional active contour model which depends only on

Figure 3.15: Comparison of the detection results of the proposed level set method using temporal information (row 1), with the ground truth (row 2), and the traditional level set method (row 3) in four of the image frames during the heart contraction period in the first echocardiographic sequence. As the boundary information is strong enough to be detected, all algorithm perform well.

the image gradient to stop the curve evolution, this algorithm adopts a new stopping term based on the Mumford-Shah functional. The advantages of the active contour without edges method is that, it can detect objects whose boundaries are not necessarily defined by gradient nor being smooth.

Suppose C is the evolving curve, the energy function $F(C)$ of the active contour without edges method is defined as

$$\begin{aligned}
 F(C) = & \mu \cdot \text{Length}(C) \\
 & + \lambda_1 \int_{\text{inside}(C)} |\mu_0(x, y) - c_1|^2 dx dy \\
 & + \lambda_2 \int_{\text{outside}(C)} |\mu_0(x, y) - c_2|^2 dx dy
 \end{aligned} \tag{3.15}$$

Figure 3.16: Comparison of the detection results of the proposed level set method using temporal information (row 1), with the ground truth (row 2), and the traditional level set method (row 3) in four of the image frames during the heart relaxation period in the first echocardiographic sequence. Since the ventricular boundaries suffer from signal dropout in some images, the traditional level set method cannot detect the weak segments of the boundary and leaked out. The proposed method can correctly detect the weak heart boundary without any leakage.

where $\mu \geq 0$, $\lambda_1, \lambda_2 > 0$ are fixed parameters, and the constants c_1, c_2 , depending on C , are the averages of μ_0 inside C and respectively outside C [30].

Figure 3.17 compares the detection results of these two methods in three image frames containing dropouts in the third testing sequence. Although the region information is utilized, however, the active contour without edges method treats the weak or missing boundary caused by dropout as part of the interior region of the heart chamber, and thus it keeps on evolving to the outside of the inner heart wall boundary. Therefore, it still suffers from the boundary leakage problem. Moreover, as there is often noise in the ultrasound images, the heart chamber is not a homogenous region.

Figure 3.17: Comparison of the detection performance of our proposed level set method using temporal information with the active contour without edges in the third testing sequence. Row 1 shows the results of our proposed method, row 2 is the ground truth and row 3 is the results of active contour without edges. It can be seen that the proposed method shows more accurate detection performance than the active contour without edges method.

The evolving curve may converge to the local noise within the heart chamber instead of moving outward. Hence, the overall detection performance of the active contour without edges method is not satisfactory. By utilizing the temporal information to recover the weakened ventricular boundary, the proposed method can successfully find the correct inner heart wall boundary without any leakage problem.

3.3.2 Quantitatively Analysis

To quantitatively measure the detection accuracy of the proposed methods, two metrics are used. The first one is the mean absolute distance (MAD) between the extracted object boundary and the ground truth [17]. For contour A and ground truth B , their MAD distance is defined as

$$e(A, B) = \frac{1}{2} \left\{ \frac{1}{n} \sum_{i=1}^n d(\mathbf{a}_i, B) + \frac{1}{m} \sum_{j=1}^m d(\mathbf{b}_j, A) \right\} \quad (3.16)$$

where $A = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$, $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\}$ represent the coordinates of the points on the contours and $d(\mathbf{a}_i, B)$ is the distance of point $\mathbf{a}_i$ to the closet point on B

$$d(\mathbf{a}_i, B) = \min_j \|\mathbf{b}_j - \mathbf{a}_i\| \quad (3.17)$$

The second metric is the Jaccard similarity coefficient $\mathcal{J}$ which is defined as the intersected area bound by the extracted contour and the ground truth divided by their united area [118]. A Jaccard similarity coefficient of 1.0 represents perfect overlap, whereas an index of 0.0 represents no overlap.

$$\mathcal{J} = \frac{S_{\text{extracted_contour}} \cap S_{\text{ground_truth}}}{S_{\text{extracted_contour}} \cup S_{\text{ground_truth}}} \quad (3.18)$$

where S represents the area of the enclosed region of the curve.

3.3.2.1 Active contour method using temporal information

For the active contour method using temporal information, Figure 3.18 show its MAD distance and the Jaccard similarity coefficient $\mathcal{J}$ in the first echocardiographic sequence. Compared with the MAD distance and similarity coefficient of the traditional active contour method, Pardas's method and Lin's method, it can be seen that during the heart contraction period (1^{st} to 12^{th} frame), as the boundaries are strong

Table 3.1: Comparison of the MAD distance between our proposed active contour method using temporal information with the other methods that incorporate temporal information into the external energy of active contour function
(Mean $\pm$ Standard deviation (pixel))

	Proposed method	Traditional snake	Pardas's method	Lin's method
Seq. 1	1.62 ± 0.23	1.90 ± 0.86	1.72 ± 0.45	3.82 ± 1.27
Seq. 2	3.19 ± 0.75	4.97 ± 2.88	4.86 ± 2.43	3.20 ± 0.96
Seq. 3	7.88 ± 2.66	8.95 ± 3.26	5.73 ± 0.85	9.41 ± 3.51
Seq. 4	2.32 ± 0.55	6.15 ± 3.48	7.45 ± 4.69	2.82 ± 0.65
Seq. 5	2.53 ± 0.76	3.06 ± 1.14	5.21 ± 1.56	4.41 ± 0.97

enough to be detected, most of the algorithms achieve small MAD distance and high similarity coefficient. However, during the relaxation period (13th to 21st frame), as the boundaries suffer from high degree of image degradations, only the proposed method can still achieve a satisfactory similarity with the ground truth (less than 2 pixels MAD distance and around 0.9 $\mathcal{J}$). The other algorithms fail to detect the boundaries as seen from their large MAD distances and low similarity coefficients. The proposed method can extract the ventricular boundaries more accurately throughout the whole image sequence.

We then apply the proposed active contour method using temporal information on all five echocardiographic image sequences. Tables 3.1 and 3.2 show the mean and standard deviation of the mean absolute distance and the Jaccard similarity coefficient of the proposed method compared with the traditional active contour method, Pardas's method and Lin's method. The results indicate that overall, the proposed method obtains a lower MAD distance and a higher $\mathcal{J}$ with the ground truth, while achieving smaller standard deviations at the same time.

3.3.2.2 Level set method using temporal information

For the level set method using temporal information, Figures 3.19 and 3.20 compare its MAD distance and the Jaccard similarity coefficient $\mathcal{J}$ with the traditional level set

(a) Comparison of MAD distance

(b) Comparison of Jaccard similarity coefficient $\mathcal{J}$

Figure 3.18: Comparison of the MAD distance and the Jaccard similarity coefficient $\mathcal{J}$ of the proposed active contour method using temporal information with the traditional active contour method, Pardas's method and Lin's method in the first image sequence.

method and the active contour without edges method in the first echocardiographic sequence (21 frames). The figures indicate that during the heart contraction period (1^{st} to 12^{th} frame), most of the algorithms achieve a small MAD distance and a high similarity coefficient. However, during the relaxation period (13^{th} to 21^{st} frame), as the boundaries suffer from high degree of image degradations, only the proposed method can still achieve a satisfactory similarity with the ground truth (less than 1.2 pixels

Table 3.2: Comparison of the Jaccard similarity coefficient $\mathcal{J}$ between our proposed active contour method using temporal information with the other methods that incorporate temporal information into the external energy of active contour function
(Mean $\pm$ Standard deviation)

	Proposed method	Traditional snake	Pardas's method	Lin's method
Seq. 1	0.903 ± 0.016	0.878 ± 0.053	0.890 ± 0.032	0.759 ± 0.082
Seq. 2	0.827 ± 0.041	0.697 ± 0.142	0.696 ± 0.113	0.793 ± 0.146
Seq. 3	0.589 ± 0.138	0.588 ± 0.126	0.678 ± 0.089	0.575 ± 0.136
Seq. 4	0.874 ± 0.047	0.793 ± 0.076	0.618 ± 0.158	0.810 ± 0.062
Seq. 5	0.851 ± 0.039	0.763 ± 0.154	0.652 ± 0.164	0.853 ± 0.058

Figure 3.19: Comparison of the MAD distance of the three algorithms in the first testing sequence.

MAD distance and around 0.9 $\mathcal{J}$). The conventional level set method is disrupted by echocardiographic dropout, hence shows a low similarity to the ground truth. The active contour without edges method is also disrupted by the dropout and the image noise within the heart chamber, its results are not accurate either. The proposed method can extract the ventricular boundaries more accurately throughout the whole image sequence.

We then apply the proposed level set method using temporal information on the other four echocardiographic image sequences and compare it with the traditional level

Figure 3.20: Comparison of the Jaccard similarity coefficient $\mathcal{J}$ of the three algorithms in the first testing sequence.

Table 3.3: Comparison of the MAD distance of the three algorithms in all the five image sequences.

(mean $\pm$ standard deviation)

	Proposed method	Conventional level sets	Active contour without edges
Seq. 1	1.0925 ± 0.2588	1.3255 ± 0.6198	1.1668 ± 0.2196
Seq. 2	2.1035 ± 0.5876	2.9595 ± 0.5427	2.3869 ± 0.6876
Seq. 3	3.2215 ± 0.9620	3.6068 ± 0.9779	6.0652 ± 1.1669
Seq. 4	3.1477 ± 0.5500	5.5611 ± 1.7435	4.6852 ± 0.8883
Seq. 5	2.2212 ± 0.4762	3.2661 ± 0.6543	2.7106 ± 0.6039

set method and the active contour without edges method. Tables 3.3 and 3.4 show the mean and standard deviation of the MAD distance and the similarity coefficient of the detection results of these three algorithms in all five image sequences. It can be seen that overall, the proposed method obtains a lower mean absolute distance and a higher similarity coefficient $\mathcal{J}$ with the ground truth, while achieving smaller standard deviations at the same time.

Table 3.4: Comparison of the Jaccard similarity coefficient $\mathcal{J}$ of the three algorithms in all the five image sequences.

	(mean $\pm$ standard deviation)		
	Proposed method	Conventional level sets	Active contour without edges
Seq. 1	0.9171 ± 0.0301	0.8676 ± 0.0384	0.8825 ± 0.0336
Seq. 2	0.8488 ± 0.0416	0.8054 ± 0.0713	0.8230 ± 0.0380
Seq. 3	0.7601 ± 0.0546	0.7304 ± 0.0542	0.6495 ± 0.0783
Seq. 4	0.7679 ± 0.0437	0.7233 ± 0.0483	0.7789 ± 0.0510
Seq. 5	0.8348 ± 0.0445	0.7724 ± 0.0461	0.8187 ± 0.0585

3.3.3 Computational Burden

Apart from the detection performance, the proposed methods are evaluated in their computational burden. As there are extra computations resulting from using the EM algorithm (around 100 to 200 extra iterations) in GMM, the overall detection process of the proposed methods is slowed down. To compensate for the extra computation, the following two steps are taken:

- Firstly, by making use of the temporal consistency of object boundaries in adjacent image frames, the active contour curve and level set curve can be initialized with a contour similar to the result from block matching. Thus, the distance between the initial contour and the real inner heart wall boundary is reduced and less iterations are needed for convergence.
- Secondly, the number of training data for GMM can be reduced. We currently use 5000 points to train GMM in order to keep a reliable estimation. In fact, this number can be reduced to around 800 to 1000 points and we have proved that it will not affect the detection performance.

(a) Initial contour in proposed method (b) Initial contour in traditional active contours

Figure 3.21: Comparison of the initial contour of the proposed method and the traditional active contour method.

3.3.3.1 Active contour method using temporal information

For the active contour method using temporal information, we compare its initialization position and the computation time with the traditional active contour method which is well known for speedy computation. Figure 3.21 shows the comparison of the initialization position in one of the image frames. It can be seen that by using the results from block matching, the initial contour of our proposed method is much closer to the real inner heart wall boundary. Thus, less iterations are needed for the evolving curve to converge to the real inner heart wall boundary. However, in the traditional active contour method, the detected boundary in the previous frame is usually used to initialize the contour in the current frame. Its initial positions are some distance away from the real boundary. Hence, it will take more iterations for the curve to converge to the real boundary. Table 3.5 shows the number of iterations and the computation time of the active contour method using temporal information and the traditional active contour method. The experiments were conducted using MATLAB on the same single user computer (Pentium 4, 3.2 GHz CPU). The table shows that the active contour method using temporal information indeed needs less iterations (usually no more than 100 iterations). It is a little slower than the traditional method and yet considering the detection performance, this computational overload is acceptable.

Table 3.5: Comparison of the number of iterations and computation time of the proposed active contour method using temporal information with the traditional active contour method

(Number of iterations / Computation time (second))

	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5
Proposed method	40 / 3.53	100 / 4.56	80 / 4.65	55 / 5.15	70 / 4.74
Traditional snake	75 / 2.21	200 / 3.91	200 / 4.83	80 / 4.00	100 / 3.72

3.3.3.2 Level set method using temporal information

For the proposed level set method using temporal information, we assess the impact on the detection performance of our proposed method when the number of training points drops from 5,000 to 1,000 in the EM algorithm. Tables 3.6 and 3.7 show the comparison between the MAD distance and Jaccard similarity coefficient $\mathcal{J}$ of the method using 5,000 points and 1,000 points, respectively. It can be seen that although the number of training data decreases, it won't significantly affect the detection performance. Hence, 1,000 points are enough to train a reliable threshold. In this way, the computational burden of using EM algorithm can be reduced. We then compare the proposed level set method using temporal information with the traditional level set method and the traditional active contour method which is well known for speedy computation. It can be seen from Table 3.8 that generally, our method only needs around 0.5 second more than theirs. Considering the detection performance, this computational overload is acceptable.

3.3.4 Comparison of the Two Proposed Methods

Finally, the detection results of the proposed active contour method using temporal information, and the proposed level set method using temporal information are compared with each other. It can be seen from Figures. 3.11 and 3.16 that although temporal information has been incorporated into the active contour method, however, due to its intrinsic properties of numerical implementation, the active contour method

Table 3.6: Comparison of the MAD distance between the proposed method using 5000 points and 1000 points in all the five image sequences.

(mean $\pm$ standard deviation)

	5000 points	1000 points
Seq. 1	1.0925 ± 0.2588	1.1038 ± 0.2951
Seq. 2	2.1035 ± 0.5876	2.2224 ± 0.5723
Seq. 3	3.2215 ± 0.9620	3.2564 ± 0.9918
Seq. 4	3.1477 ± 0.5500	3.2059 ± 0.6213
Seq. 5	2.2212 ± 0.4762	2.2436 ± 0.4940

Table 3.7: Comparison of the Jaccard similarity coefficient $\mathcal{J}$ between our proposed method using 5000 points and 1000 points in all the five image sequences.

(mean $\pm$ standard deviation)

	5000 points	1000 points
Seq. 1	0.9171 ± 0.0301	0.9049 ± 0.0522
Seq. 2	0.8488 ± 0.0416	0.8343 ± 0.0489
Seq. 3	0.7679 ± 0.0437	0.7587 ± 0.0693
Seq. 4	0.7601 ± 0.0546	0.7462 ± 0.0521
Seq. 5	0.8348 ± 0.0445	0.8394 ± 0.0532

Table 3.8: Comparison of the average computation time per frame of the proposed level set method using temporal information, the traditional active contour method, and the traditional level set method.

(Average Computation time (second))

	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5
Proposed method	3.25	3.68	4.84	3.95	2.79
Traditional snake	2.28	3.13	3.92	3.11	2.54
Traditional level set	2.77	3.27	4.76	3.24	2.88

has a difficulty of converging into the concave regions. Thus, detected contours of the active contour method using temporal information are not closely following the inner heart wall boundary as those achieved by the level set method using temporal information. By referring to the MAD value in Tables 3.1 and 3.3, and the Jaccard similarity coefficient $\mathcal{J}$ in Tables 3.2 and 3.4, it can be seen that overall, the proposed

level set method using temporal information achieves a more accurate detection than the proposed active contour method using temporal information. However, as the active contour method are much easier to be implemented in comparison to the level set method, and the active contour method using temporal information is faster than the level set method using temporal information, the final choice depends on how accurate the detection should be and whether one is willing to achieve higher accuracy at cost of heavier computation. It also depends on the frame rate to be achieved in the particular application.

3.4 Discussions and Conclusions

A new method to incorporate temporal information into the variational frameworks to regularize the curve evolution process is proposed to solve the “missing boundary” problem caused by dropout in echocardiographic image sequences. The ventricular boundary is quantitatively identified and classified into strong and weak segments. As for the weak segment, it is considered that the boundary information is destroyed by the signal dropout and they are difficult to be detected. Temporal information from neighboring frames is incorporated into them to re-build the missing boundary information and strengthen the gradient force. To attain this goal, the following is proposed: (i) a Gaussian Mixture Model (GMM) is employed to calculate the thresholds that quantitatively partition the strong and weak segments; (ii) the temporal information is incorporated as a regularizer into the external energy of the active contour framework and the speed function of level set framework as shown in Eqn. (3.10) and Eqn. (3.14); (iii) a weight is developed to balance the strength between the image gradient force and temporal information, and a strengthening factor is introduced to boost the total energy (the image gradient force and temporal information) to the average level.

In the experimental evaluations, the proposed active contour method using temporal information was compared with the ground truth, the traditional active contour

method, and other two methods that also incorporate temporal information into active contour equation. The proposed level set method using temporal information was compared with the ground truth, the traditional level set method and the active contour without edges method which uses a region regularizer. The detection results are assessed according to their detection performance, the quantitative analysis of the results and the computational burden. The experiments demonstrate that the proposed methods both outperform the peer methods with which they are compared and that no heavy computation burden is involved. By comparing the proposed active contour method using temporal information with the proposed level set method using temporal information, it can be seen that overall, the proposed level set method can obtain a more accurate detection result than the proposed active contour method using temporal information. However, it will be a little slower than the proposed active contour method using temporal information in terms of the computational time.

Chapter 4

Regularizing Variational Methods by Adding a Shape Prior Model

4.1 Introduction

In Chapter 3, temporal information is utilized to regularize the variational methods for solving the “missing boundary” problem. However, there are many cases in which temporal information is either insufficient or unavailable. For example, when an object in a single snap shot is partially occluded or is put in a cluttered background, there is no temporal information available to be used to recover the original object boundary. Moreover, even in an image sequence where temporal information is readily available, it may also happen that an object is partially occluded for a few seconds and thus the original object boundary is missing over dozens of image frames. In this situation, it is not feasible to find the original object shape by using the temporal information only from the neighboring frames. Additional information or other alternatives, should be explored to solve the “missing boundary” problem.

In recent years, a common approach to solve this problem is to incorporate a shape prior model into the variational methods to regularize the detection process [119]. The

shape of the evolving curve is constrained within a set of pre-selected training shape patterns [20, 120]. Chen [121, 122] proposed to use the average of a training set of aligned curves as a shape model, while Zhang [123] defined their shape model as the level set function of a pre-selected reference shape. In [124, 125] and [54], only a single shape was considered as the shape prior information. The above methods focus more on the numerical computation of the shape priors, and they do not elaborate on the issue of representing the shape of objects in general. Hence, the shape priors they used are too simple to capture the diverse characteristics of object shapes.

There is recent research work that calls for the use of statistical learning methods to construct the shape prior model. Shape prior models are usually constructed on the basis of learning the shape of a set of training samples. Rousson [47] modeled the distribution of their training data with a Gaussian function. The curve is evolved towards a preferred shape pattern by seeking the maximum probability of the density function. Cremers [35] also modeled the distribution of the training data by a Gaussian function, then the density function is used as a shape energy term to constrain the curve evolution. However, statistical density estimation in the above algorithms is directly carried out in the original data space. As the training data are usually generated from a certain representation of shapes (e.g. level set function), they can be very high dimensional, whereas the number of training samples is also limited. The data are distributed sparsely in a high dimensional space, making the density estimation insufficient to capture the intrinsic data structure.

Hence, dimensionality reduction techniques are used during the construction of the prior model. Wang [126] computed the corresponding points across a set of training shapes and constructed a statistical model in the subspace determined by the principal component analysis (PCA) of these points. Leventon [46] proposed performing a principal component analysis (PCA) on the training data first before the density estimation is made. The data distribution is modeled in a low dimensional subspace. A similar technique was used in [127, 108] where the shape prior model is constructed

by the linearly weighted sum of k principal eigenshapes plus a mean shape. The same shape prior model was also used in [50, 128]. However, Leventon's method suffers from slow computation. At each step of its curve evolution, the *maximum a posterior* (MAP) position must be found to adjust the pose of the prior model. Moreover, many parameters need to be set some of which are determined empirically.

In order to constrain the shape of the curve under current evolution within a cluster of preselected similar shapes using the shape prior knowledge, a collection of training shape patterns are needed from which we can learn the general characteristics of object shapes. Furthermore, a more sophisticated method to represent these shape characteristics should be considered. Hence, in this chapter, we aim at using a statistical method to describe the structure of a set of preselected shapes and construct the shape prior model based on these shapes.

In this chapter, three proposed methods that incorporate the shape prior model to regularize the variational methods for robust object boundary detection are presented. The level set functions of the training shapes are used as training data. They are projected onto a low dimensional subspace using a dimensionality reduction technique. Then their data distribution is approximated by a density estimation function. Based on the density function, a shape prior model is constructed and is incorporated into the geodesic active contour formulation to constrain the contour evolution process. The curve converges with due consideration of both the local shape variations and the global shape consistency. In the first proposed method, the training data are projected by the principal component analysis (PCA) method, and the data distribution is approximated by a single Gaussian function. In the second proposed method, the training data is also projected by PCA method, but it improves the first method by approximating the data distribution with a Gaussian mixture model (GMM). Finally, in the third proposed method, the training data are projected by a nonlinear mapping using kernel principal component analysis (KPCA), and their distribution is approximated by GMM in the feature subspace. The experimental results of the proposed methods

are given respectively to demonstrate their advantages.

4.2 Learning a Shape Prior Model Using PCA and a Single Gaussian Function

4.2.1 Motivations and Objectives

As reviewed in section 4.1, a common approach to solve the missing boundary problem is to incorporate a shape prior model into the variational framework. An efficient shape prior model is usually constructed based on a collection of training shapes, and a sophisticated density estimation method which can represent the shape characteristics of the object should be used. Apart from the above issues, it is found that the following points should also be considered:

- Since the level set representations of object shapes are commonly used as training data, they are usually high dimensions. The data distribute sparse in the space, which makes the density estimation inappropriate to capture the intrinsic structure of data distribution. Hence, dimension reduction technique is needed. The data distribution should be more efficiently estimated in the lower dimensional subspace.
- If the shape prior model is created to improve the detection performance in surveillance videos, it should be incorporated into the variational methods without causing a substantial increase in computational complexity.

In this section, the first shape prior model is presented which is constructed on the basis of statistical learning of the distribution of training data using a single Gaussian function in the subspace. The level set function of object shapes is collected as training data and projected onto the first k principal directions using the principal component

analysis method (PCA). In the reduced subspace, the data distribution is approximated by a Gaussian function. A shape prior model is constructed and used as a regularizer to constrain the curve evolution process. To speed up the convergence under the influence of shape prior knowledge, we propose to explicitly align the shape model with the shape of current curve under evolution. Thus less iterations are needed in adjusting the pose compared to the Leventon's MAP approach for determining the pose parameter. Moreover, a weighting scheme is developed to balance the influence between the image gradient and shape prior information by quantifying their differences, unlike Lenvenon's work in which some of the parameters have to be set empirically.

4.2.2 Constructing the Shape Prior Model

Before constructing the shape prior model, the collected shapes need to be aligned to the same position, orientation and scale. Then their level set functions are computed as training data. The shape prior model is constructed based on statistical estimation of the distribution of training shapes.

4.2.2.1 Shape alignment

In this section, we only consider the objects whose shapes are either view-angle independent or less affected by viewing angles. Hence the training shapes are just subject to affine transformations such as translation, rotation and scale. Since matching geometric shapes is not the focus of our work, a simple alignment method, the Euclidean similarity transformation [129] is adopted. A Euclidean similarity transformation of a point x into a point x' is defined as

$$x' = \alpha Ax + T \quad (4.1)$$

where α is the isotropic scaling factor, A is the rotation matrix, and T is the translation vector. For each training shape, a set of (α, A, T) needs to be found to adjust it to

the same pose as a predefined (usually arbitrary selected) reference shape. These parameters are determined in the following ways:

- Translation T : defined as the distance between the centroid of target shape (x_i, y_i) and the centroid of reference shape (x_j, y_j)

$$T = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (4.2)$$

- Scaling α : defined as the ratio between enclosed area of the reference shape and the target shape

$$\alpha = \frac{S_{reference}}{S_{target}} \quad (4.3)$$

- Rotation matrix A : defined as

$$A(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (4.4)$$

Figure 4.1 shows some examples of training shapes before and after the alignment respectively. It can be seen that after alignment, the shapes share the same center, with approximately equal size, and align towards the same orientation. Figure 4.2 compares the amount of shape overlap before and after alignment. It's obvious that the characteristics of shape patterns are much clearer after the alignment.

This alignment method will be also used during the incorporation of shape prior model into the geodesic active contour equation. The prior model will be adjusted to the same pose as the shape of the curve under evolution at each iteration step.

4.2.2.2 Creating the training data

In some existing work, marker points are used to represent the contours [126, 35]. However, the number of points has to be determined and the correspondence between points is needed during the shape alignment. In this chapter, we adopt another method,

Figure 4.1: Examples of the training shapes before and after alignment. The shapes generally share the same center, size and orientation after alignment.

Figure 4.2: Comparison of the extent of shape overlap before and after alignment.

the level set function of contour, to represent the shape. It can avoid previous problems and tolerant to slight misalignment of shape features [46].

Each training shape is embedded as the zero level set of a higher dimensional surface Ψ_i . The points on the surface encode the distance to the nearest point on the boundary of object, with negative distance assigned to the inside and positive distance to the outside of object region.

$$\Psi_i(x, y) = \begin{cases} 0, & (x, y) \in S \\ -D((x, y), S), & (x, y) \in \mathbb{R}^2 \\ D((x, y), S), & (x, y) \in \overline{\mathbb{R}^2} \end{cases} \quad (4.5)$$

Figure 4.3: Examples of level set functions of the synthetic data.

where $D((x, y), S)$ computes the minimum Euclidean distance between the grid pixel (x, y) and the shape contour S , $\mathbb{R}$ is the region inside the object shape contour [130]. Function Ψ_i is a matrix (signed distance map) of the same size as the training shape template. Figure 4.3 shows the examples of the level set functions.

These level set functions are rearranged into column vectors and form a matrix M

$$M = \{\Psi_1, \Psi_2, \dots, \Psi_n\} \quad (4.6)$$

where M is a $N \times n$ matrix, with N being the dimension of feature vectors and n being the number of training data. Since N is the size of the shape template, it is usually large. The training data distribute sparse in a high dimensional space, making the density estimation inappropriate to capture the intrinsic structure of data distribution. A dimensionality reduction technique is needed to project the training data onto a low dimensional subspace for more efficient shape modeling.

4.2.2.3 Constructing the shape prior model

Following the construction of shape prior in Leventon's work [46], let $\bar{\Psi}$ be the mean of training vectors calculated from the the column vectors formed by rearranging rows of the level set function as columns and concatenate them into a single column vector.

$$\bar{\Psi} = \frac{1}{n} \sum_i \Psi_i \quad (4.7)$$

where Ψ_i is the rearranged discrete lattice of level set function from the i -th training sample as defined in Eqn. (4.6).

Each data is subtracted by the mean vector $\bar{\Psi}$ to create the zero-mean vector $\hat{\Psi}_i$

$$\hat{\Psi}_i = \Psi_i - \bar{\Psi} \quad (4.8)$$

Hence, matrix M becomes $\hat{M}$. Constructing the covariance matrix C

$$C = \frac{1}{n} \hat{M} \hat{M}^\top \quad (4.9)$$

Performing the eigen-decomposition on C

$$C = U \Sigma U^\top \quad (4.10)$$

where U is a unitary matrix containing the eigenvectors of C and Σ is the corresponding diagonal matrix of eigenvalues. Matrix U defines a coordinate transform which decorrelates the data and reveals the subspace the data resides in. By projecting the data onto the largest k principal components,

$$\tilde{\Psi}_i = U_k^\top (\Psi_i - \bar{\Psi}) \quad (4.11)$$

the dimension of the training data can thus be reduced to only k . The data are easier to be approximated in the low dimensional subspace.

In the reduced subspace, the feature vectors $\tilde{\Psi}$ are assumed to follow a Gaussian distribution.

$$p(\tilde{\Psi}) = \frac{1}{(2\pi)^{k/2} |\Sigma_k|^{1/2}} \cdot \exp \left\{ -\frac{1}{2} \tilde{\Psi}^\top \Sigma_k^{-1} \tilde{\Psi} \right\} \quad (4.12)$$

Thus, for a test shape with level set function Ψ_t , its probability in the subspace is $p(\tilde{\Psi}_t)$. Our task is to evolve the geodesic active contour towards a shape pattern so that its probability can be maximized. For computational convenience, we compute

the negative logarithm of the probability and define it to be an energy term (ignore the constant part in Eqn. (4.12) for its minimization)

$$\begin{aligned} E_{shape}(\tilde{\Psi}_t) &= -\log [p(\tilde{\Psi}_t)] = \log [(2\pi)^{k/2} |\Sigma_k|^{1/2}] + \left[\frac{1}{2} \tilde{\Psi}_t^\top \Sigma_k^{-1} \tilde{\Psi}_t \right] \\ &\propto \tilde{\Psi}_t^\top \Sigma_k^{-1} \tilde{\Psi}_t \end{aligned} \quad (4.13)$$

Minimizing E_{shape} is equivalent to maximizing the probability function. As E_{shape} is a convex function, we search along its gradient descent direction

$$[\tilde{\Psi}_t]_{n+1} = [\tilde{\Psi}_t]_n - \epsilon \nabla E_{shape} = [\tilde{\Psi}_t]_n - \epsilon \Sigma_k^{-1} [\tilde{\Psi}_t]_n \quad (4.14)$$

By iteratively updating $\tilde{\Psi}_t$ in Eqn. (4.14), the maximum value of $p(\tilde{\Psi}_t)$ can be found. During each step of the iteration, a shape prior model $[\Psi^*]_n$ is constructed using the largest k principal eigenvectors U_k and the mean of the entire training data $\bar{\Psi}$.

$$[\Psi^*]_n = U_k [\tilde{\Psi}_t]_n + \bar{\Psi} \quad (4.15)$$

$[\Psi^*]_n$ is a reverse mapping of the point $[\tilde{\Psi}_t]_n$ from the low dimensional feature subspace to the high dimensional data space. It is a reconstructed level set function of the same size as the training shape template. Although it may not be a true distance function due to the information lost during the reconstruction, it generally contains the properties of smoothness, local region correlation, and zero level set consistence with the original shape, it can be regarded as a level set function [46]. The shape prior term is updated at each step of the minimization of Eqn. (4.14). Since the training data have been mean shifted when performing PCA, they are centered at the origin in the subspace. Thus on convergence, the value of $\tilde{\Psi}_t$ in Eqn. (4.14) will be equal to 0, which means the shape prior model is just the mean of the training data $\bar{\Psi}$ at the last iteration step. Figure 4.4 shows the trace of a test vector in the subspace $\tilde{\Psi}_t$ during the optimization process. The point is moved from the tail region of the Gaussian

Figure 4.4: The trace of curve evolution process (corresponding to the shape “cross”) in the subspace under the influence of the shape prior knowledge.

function to the center of it, and the reconstructed level set function changed from a coarse shape pattern to finally a more accurate one.

4.2.3 Combining the Shape Prior Model with GAC Equation

To evolve the curve in the direction of the shape prior model, Eqn. (4.16) is proposed

$$\Psi_{n+1} = \Psi_n + \Delta t(\Psi^* - \Psi_n) \quad (4.16)$$

This is the shape template defined in terms of a level set function at iteration n .

Combining Eqn. (4.16) with geodesic active contour equation (2.21), and reformulating them into iterative equation, yields the final expression of the complete curve evolution process which has a different weighting scheme from [46]

$$\Psi_{n+1} = \Psi_n + \gamma \left[\Delta t_1 (g(c + \kappa) |\nabla \Psi_n| + \nabla \Psi_n \cdot \nabla g) \right] + (1 - \gamma) \left[\Delta t_2 (\Psi^* - \Psi_n) \right] \quad (4.17)$$

where Δt_1 and Δt_2 are the iteration step sizes which control the amount of increments of the second and third terms on the right hand side of Eqn. (4.17), and γ is a weighting factor which balances the strengths between them.

The rationale of using a different force balancing mechanism from the one in [46] is explained as follows. The second term on the right hand side of Eqn. (4.17) evolves the curve based on object boundary information, it provides an edge-driven force. While the third term evolves the curve based on the shape prior information, it provides a knowledge-driven force. The curve will strike a balance between these two forces during the evolution process, thus it can not only capture the local shape variations but also preserve a global shape consistency. Hence, the weighting factor γ has to be determined carefully. We propose to quantify the difference between the current evolving curve and the shape prior as shown in Figure 4.5 and choose γ by measuring their distance. Suppose there are two thresholds d_1 and d_2 which act as upper and lower limits of the dissimilarity between the curve under evolution and the shape prior model. If the distance is larger than d_1 pixels ($d > d_1$), the curve under evolution and the shape model are so dissimilar that the edge-driven force is not reliable and it should be ignored. Thus, γ is set to 0 to strengthen the influence of prior information to recover the object shape. If the distance is smaller than d_2 pixels ($d < d_2$), the distance between the curve and the shape model is so small that they can be regarded as equal to each other and no shape prior information is needed. Hence, γ is set to 1. Finally, if the distance is between d_2 and d_1 pixels, γ is set to 0.5 such that the image gradient and shape information are jointly used to constrain the curve evolution. In summary, parameter γ is defined as

Figure 4.5: Illustration of the selection of γ .

$$\gamma = \begin{cases} 0 & d > d_1 \\ 1 & d < d_2 \\ 0.5 & d_1 < d < d_2 \end{cases} \quad (4.18)$$

To determine the value of parameters d_1 and d_2 , we note that the distance below d_2 is considered as an acceptable level of the difference between the curve under evolution and the shape prior model. Hence, we propose to compute d_2 as the standard deviation σ of the distance among the training shapes as shown in Figure 4.6, and d_1 to be 4 times of d_2 which is 4σ .

For training shape A and B , their distance $e(A, B)$ can be computed using Eqn. 3.16. Hence, parameter d_2 can be computed as

$$d_2 = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (e_i - \mu_0)^2} \quad (4.19)$$

where μ_0 is the mean of the distance among the training shapes, and k is the number of distance among the training shapes. For example, if the number of training shapes is n , parameter k is $n(n-1)/2$.

Figure 4.6: Parameter d_2 is obtained as the standard deviation of the distance among the training data.

In this way, with a larger variance of the training shapes, larger d_1 and d_2 are used and the proposed algorithm is able to tolerate larger shape variations. On the other hand, with a smaller variance of the training shapes, smaller d_1 and d_2 are used and the curve is restricted to a small range of shape variations.

During the curve evolution process, the curve is first moved according to edge-driven force. After becoming stationary, then the shape prior force is introduced. This process allows an easier way to adjust the pose (translation, rotation and scale) of the shape model to be the same as the shape of curve under evolution. Unlike Leventon who used a maximize a posterior (MAP) method to find the pose parameter of best matching orientation [46], we propose to explicitly align the shape prior model using the Euclidean similarity transformation. The translation and scaling can be found by only one computation. As for the rotation, since the angle between successive iterations will not vary significantly after the first alignment, it is possible to find the matching rotation within a small search range. In this way, less iterations are needed than the MAP estimation. By updating together with the update of level set function using the narrow band approach, this alignment can be implemented faster than the MAP approach in [46].

Figure 4.7: Intermediate stages during the curve evolution process of the proposed method.

4.2.4 Experimental Results

In this section, the proposed shape prior model based on a single Gaussian function is tested on different dataset to evaluate its detection performance. We used both synthetic images and real human shape images.

4.2.4.1 Synthetic images

The synthetic shapes are circle, triangle and cross, each containing 20 samples drawn by hand as shown in Figure 4.1. These shapes were extracted and aligned to a template of size 64×64 pixels, and their level set functions were calculated to form the training data as shown in Figure 4.3. Then PCA was carried out, and the data were projected onto their k eigenvectors corresponding to k largest eigenvalues which can retain over 90% of the total variance. According to the variance of synthetic shapes (their standard deviations are between 1.4 to 2.0), the distance d_1 and d_2 in parameter γ are set to 8 and 2 pixels respectively.

In the first experiment, we partially occluded the target objects with another object. It is obvious that the original shapes of objects are destroyed. Hence, it is impossible for the traditional geodesic active contours to find the real object boundary by only using the edge information in the image. To recover the original object shape, the shape prior model should be incorporated into the detection algorithm. Figure 4.8 compares the detection results of traditional geodesic active contours without using a shape model

Figure 4.8: Comparison of the detection results of traditional geodesic active contours without using shape prior model and the proposed method using shape prior model for partially occluded objects.

and our proposed method which incorporates shape prior model into GAC. The curve of traditional GAC simply stopped at the local maximum of gradient and could not encompass the real shape of object. However, our proposed method can successfully find the object boundary by utilizing both the edge information and shape information. Figure 4.4 shows the trace of curve evolution process of the proposed method in the subspace under the influence of the shape prior knowledge. It can be seen that at the beginning, the curve was evolved under the influence of a slightly distorted shape pattern, but finally it converged to a more accurate one. The trace indicates that the curve is moved towards a most likely shape pattern under the influence of the shape prior. In this way, even though the original shape of the object has been destroyed, the curve can still detect the correct shape by referring to the shape information in the shape prior model. Figure 4.7 shows the intermediate curve evolution process of the proposed method. It can be seen that the evolving curve overpassed the spurious contours and recovered the desired shape of object.

In the second experiment, the objects are partially cut so that parts of the shapes

Figure 4.9: Comparison of the detection results of traditional geodesic active contours without using shape prior model and out proposed method using shape prior model for partially cut objects.

are missing. Moreover, the objects are with different translation, rotation and scale value. Figure 4.9 (a) shows the detection results of traditional geodesic active contours without using shape prior model and Figure 4.9 (b) shows the detection results of geodesic active contours using the proposed shape model. It can be seen that without a shape model, the curve can only find an incomplete shape of the object and converge to the wrong position. Whereas by using the proposed shape model, the curve is driven towards the desired boundary. The missing portion of the object shape is successfully recovered by the shape prior information.

In the third experiment, the objects are put in a cluttered background with different translation, rotation and scale. The undesired gradient from the background clutter severely interferes with the gradient of object boundary and makes it difficult to differentiate the object from background. Hence, in Figure 4.10 (a), the geodesic active contours without shape prior is disturbed by the spurious edge information and cannot obtain the original object shapes. However, by utilizing the shape model, the interference from the background clutter is eliminated and the original shapes of objects are

Figure 4.10: Comparison of the detection results of traditional geodesic active contours without using shape prior model and out proposed method using shape prior model for objects in cluttered background.

correctly extracted as shown in Figure 4.10 (b).

4.2.4.2 Real human shape images

Now, we apply the proposed shape prior model on human shapes in real images. Four human postures, standing (view from the front), standing (view in silhouette), standing with open-armed, and walking (view in silhouette), are selected for testing. Each of the posture contains 40 samples. Figure 4.11 shows some examples of the training shapes. The human shapes are extracted and aligned to a 80×80 pixels template to compute their level set functions. Figure 4.12 shows the examples of the level set functions. Then these training data are projected onto a low dimensional subspace by PCA. As the standard deviations of human shapes are around 1.8 to 2.4 pixels, distance d_1 and d_2 are set to 10 and 2 pixels in this case.

The proposed method was tested on three kinds of occlusions. Figure 4.13 row 1 shows the detection results of the proposed method. Columns (a) and (b) are the human objects occluded by a signage board, column (c) is the human object occluded

Figure 4.11: Examples of selected human shapes.

Figure 4.12: Examples of the level set functions of training shapes.

by handrails on stairs. It can be seen that due to occlusion, the original human shapes is disturbed, and the object edge is changed. Hence, the traditional geodesic active contours without using shape prior model could not find the real shape of human but converged to the wrong position as shown in Figure 4.13 row 2. However, by referring to the shape information in the shape prior model, the GAC curve can successfully recover the original human shape. Compared with the detection results of traditional GAC, the proposed method shows better detection performance.

Detection results of the proposed method.

Detection results of traditional geodesic active contour method.

(a)

(b)

(c)

Figure 4.13: Comparison of the detection results of the proposed method with the traditional GAC method.

4.2.4.3 Computational time

As the shape prior model is involved in the detection process, more computation time is needed to find the boundaries than the same method without using shape prior, and most of the computation will be spent on modulating the pose of the shape model with the current curve under detection. For efficiency, we update this alignment together with the update of level set function when using the narrow band approach [31], and employ the Euclidean similarity transformation to explicitly align the shape prior model. Thus, after the shape prior is imposed, normally no more than 100 extra iterations are needed and the whole detection process can be finished within 3 to 4 seconds using VC++ on a 3.2 GHz Pentium 4 CPU.

Figure 4.14: Evaluation of the influence of thresholds d_1 and d_2 in parameter γ on the detection results.

4.2.4.4 The influence of parameter γ

Finally, the influence of d_1 and d_2 in parameter γ on the detection results needs to be further discussed. We vary the value of d_1 with 4, 6, 8, 10 and 12 respectively, and d_2 with 1, 2, 4, 6 respectively. The experimental results are shown in Figure 4.14. It can be seen that with a small d_2 , the detection results become more accurate. But more computation time is needed as most of the iterations are under the influence of shape prior (need to adjust the pose of shape prior). Accordingly, with a larger d_2 ,

the detection accuracy decreased as the curve is still based on image gradient to detect boundary although the dissimilarity between the curve and shape model is d_2 already. As for the threshold d_1 , it can be seen that with a smaller d_1 , the shape prior may “overfit” the current curve, shaping the curve to be the same as shape model even though their difference is within the variance of shapes. But with a larger d_1 , more iterations are needed and the detection accuracy is also affected. Since the standard deviations of our training shapes are usually between 1 and 3 pixels, in experiments, we usually set d_1 to be 8 or 10, and d_2 to be 2 or 3.

4.3 Learning a Shape Prior Model Using PCA and GMM

4.3.1 Motivations and Objectives

Although our first shape prior model works well in the presence of partial occlusions and cluttered background, it is constructed from training samples of a single shape pattern. It cannot be applied to other shape patterns at the same time. If the training samples are collected from multiple different shape patterns that are either from multiple objects or multiple shape states of a single object, there can be multiple shape clusters in the feature subspace, with each cluster corresponding to a certain shape pattern. Hence, the shape prior model constructed on the basis of a single Gaussian function cannot accommodate all the characteristics of multiple shape patterns. Even though it can be solved by creating several shape prior models, each corresponds to a shape pattern to account for the multiple shape patterns, in practice, it is laborious and tedious in repeating the same construction process of the prior model. In addition, a user intervention is needed to artificially designate the GAC evolving curve to one of the prior models in the detection process. Therefore, a shape prior model only encodes one shape pattern is not suitable to deal with this problem.

Recently, Cremers [125] proposed to integrate nonlinear shape statistics with variational methods. They performed a nonlinear mapping using an extension of kernel PCA first, and then modeled the distribution of training data with a single Gaussian function in the feature subspace. The obtained probability density function is treated as a shape energy term to constrain the curve evolution. For a set of training data $z_1, z_2, \dots, z_m$, the shape energy is defined as

$$E_\phi(z) = \sum_{k=1}^r \left(\sum_{i=1}^m \alpha_i^k \tilde{k}(z_i, z) \right)^2 \cdot (\lambda_k^{-1} - \lambda_\perp^{-1}) + \lambda_\perp^{-1} \cdot \tilde{k}(z, z) \quad (4.20)$$

where λ_k is the nonzero eigenvalues of the regularized covariance matrix of training data, r is the number of the nonzero eigenvalue λ_k , α_i^k is the expansion coefficients, $\tilde{k}(x, y)$ is the centered kernel function, and $\lambda_\perp$ is a constant. This nonlinear shape prior can encode different shape patterns. However, the optimization process can stop at a local extremum at which a new shape not in the training samples may be created, and it may not be the interested shapes we need. In fact, when the Gaussian peak in the kernel space is reached, it corresponds to the centers of multiple shape clusters and there is no distinction among the shapes. Although this is claimed by Cremers as a superiority of his approach so that classification of shape patterns is avoided, knowing which shape cluster the current evolving curve may correspond to can be helpful. In another piece of work, Cremers proposed another shape prior model which uses a multiphase dynamic labelling method [124, 40]. This model can also generalize to multiple shape patterns. However, as a labelling function is used to indicate the region where a prior model needs to be enforced, extra parameters and iterations are involved and the curve evolution process becomes complicated.

In order to overcome the limitations of the aforementioned shape prior models, a new classification-based shape prior model is proposed and it is incorporated into the geodesic active contour method to regularize the curve evolution process. The training shapes contain multiple shape patterns either from multiple object shapes or multiple shape states of a single object. Their level set representations are computed

and projected onto the k largest eigenvectors using the Principal Component Analysis (PCA). In the feature subspace, the data form distinct clusters with each corresponding to a certain shape pattern. The data distribution is approximated by a Gaussian Mixture Model (GMM). Then a Bayesian classifier is created based on GMM which will automatically assign the current evolving curve to the most similar shape cluster by maximizing a posterior probability. In this way, a shape prior model can be constructed by using the statistical properties of that cluster and the curve is evolved under the influence of it. The advantage of this shape prior model lies in the fact that, by only creating one shape model, it can generalize to multiple shape patterns from either multiple different objects or multiple different shape states of a single object. Moreover, it can successfully handle the problems of object occlusion and cluttered background in the detection process.

4.3.2 Constructing the Shape Prior Model

Before constructing the classification-based shape prior model, the collected training shapes need to be aligned to the same translation, rotation and scale and their level set function should be computed and used as training data. These two steps are the same as described in sections 4.2.2.1 and 4.2.2.2. Thus, we will not describe them here and move directly to the construction process of the shape prior model.

4.3.2.1 Bayesian classifier based on Gaussian Mixture Model (GMM)

The Gaussian mixture model (GMM) is a ubiquitous statistical tool for density estimation [131]. Due to its analytical tractability, asymptotic properties, and universal approximation ability to any continuous density function, it has been widely used in pattern recognition for approximating the data distribution. Each component in GMM stands for a class that consists of patterns with some similarity, and it is usually assumed to follow a Gaussian distribution.

Let $\mathbf{x} = [x_1, \dots, x_d]^\top \in \mathbb{R}^d$ be a d -dimensional random variable. It is said that $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ follows a M -component finite mixture distribution if its probability density function can be written as

$$P(\mathbf{x}|\Theta) = \sum_{m=1}^M \alpha_m p(\mathbf{x}|\theta_m) \quad (4.21)$$

where $\alpha_1, \dots, \alpha_M$ are the *mixing probabilities* which satisfy $\alpha_m \geq 0, m = 1, \dots, M$, and $\sum_{m=1}^M \alpha_m = 1$, and each θ_m is the set of parameters defining the m th components, and $\Theta \equiv \{\theta_1, \dots, \theta_M, \alpha_1, \dots, \alpha_M\}$ is the complete set of parameters needed to specify the mixture. Each component density $p(\mathbf{x}|\theta_m)$ is a multivariate Gaussian probability distribution

$$p(\mathbf{x}|\theta_m) = \frac{(2\pi)^{-\frac{d}{2}}}{\sqrt{|\Sigma_m|}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_m)^\top \Sigma_m^{-1} (\mathbf{x} - \mu_m) \right] \quad (4.22)$$

with μ_m and Σ_m being the statistical parameters of a cluster.

Once the parameters of GMM have been determined, a Bayesian classifier can be constructed [132, 133]. For a given pattern $\mathbf{x}_t$, consider α_m to be a prior probability and $p(\mathbf{x}_t|\theta_m)$ to be a class conditional probability, the posteriori probability of Ω_k can be computed as

$$p(\Omega_k|\mathbf{x}_i) = \frac{P(\Omega_k)p(\mathbf{x}_i|\Omega_k)}{P(\mathbf{x}_i)} = \frac{\alpha_k p(\mathbf{x}_i|\theta_k)}{\sum_{m=1}^M \alpha_m p(\mathbf{x}_i|\theta_m)} \quad (4.23)$$

By maximizing the posterior probability, the shape pattern $\mathbf{x}_t$ can be classified to the desired cluster Ω_k^* using the Bayesian classifier.

$$\Omega_k^* := \arg \max_{\Omega_k} \{p(\Omega_k|\mathbf{x}_i)\} = \frac{\arg \max_{\Omega_k} \{\alpha_k p(\mathbf{x}_i|\theta_k)\}}{\sum_{m=1}^M \alpha_m p(\mathbf{x}_i|\theta_m)} \quad (4.24)$$

Since the value of term $\sum_{m=1}^M \alpha_m p(\mathbf{x}_i|\theta_m)$ is the same to each variable $\mathbf{x}_i$, the above

equation can be written as

$$\Omega_k^* = \arg \max_{\Omega_k} \{p(\Omega_k | \mathbf{x}_i)\} = \arg \max_{\Omega_k} \{\alpha_k p(\mathbf{x}_i | \theta_k)\} \quad (4.25)$$

4.3.2.2 Density estimation of the training data using GMM

As aforementioned, the training shapes may contain multiple and different shape patterns either from multiple objects or from multiple shape states of a single object. There can be multiple clusters in the subspace after PCA and thus a single Gaussian is inappropriate to model the data distribution. We propose to employ the Gaussian mixture model to estimate the data distribution.

Following Eqn. (4.21), the final probability density function of the training data is

$$P(\tilde{\Psi}_i) = \sum_{m=1}^M \alpha_m p(\tilde{\Psi}_i | \theta_m) \quad (4.26)$$

The parameters Θ of GMM can be computed by the Expectation-Maximization (EM) algorithm. To determine the number of Gaussian components M , we employ an unsupervised GMM method [112] which can find a “best” overall model in the entire set of available models by measuring the Minimum Message Length (MML).

The rationale behind minimum encoding length criteria is: if you can build a short code for your data, that means that you have a good data generation model [112]. Consider a dataset $\mathcal{Y}$, known to have been generated according to $p(\mathcal{Y} | \theta)$, following Shannon theory [136], the shortest code length $\mathcal{L}(\theta, \mathcal{Y})$ for $\mathcal{Y}$ is $-\log p(\mathcal{Y} | \theta)$.

In terms of Gaussian mixture model (GMM), by necessary calculation as shown in [112], the final expression of MML is defined as Eqn. (4.27)

$$\mathcal{L}(\Theta, \tilde{\Psi}_i) = \frac{N}{2} \sum_{m=1}^M \log \left(\frac{n \alpha_m}{12} \right) + \frac{M}{2} \log \frac{n}{12} + \frac{M(N+1)}{2} - \log p(\tilde{\Psi}_i | \Theta) \quad (4.27)$$

where N is the number of parameters specifying each component, i.e., the dimension-

ality of θ_m . The one with the minimum message length $\mathcal{L}$ is the desired number of Gaussian components.

4.3.2.3 Constructing the Bayesian classifier and the shape prior model

Now consider a curve that is evolving towards an object boundary which is partially occluded or is in a cluttered background. Our purpose is to incorporate the shape prior knowledge into the geodesic active contour (GAC) function so that the shape of the curve under evolution can be constrained within a cluster of similar shapes, and the original shape of the object can be recovered. With several shape clusters in the training data, we need to find out which cluster is the desired one for which the curve should be designated. Hence, a Bayesian classifier is created according to Eqn. (4.23). It finds the target cluster by maximizing the posterior probability. For the level set function $\tilde{\Psi}$ of the evolving curve, the desired shape cluster Ω_k^* is

$$\Omega_k^* = \arg \max_k \left\{ \alpha_m p(\tilde{\Psi} | \theta_m) \right\} \quad (4.28)$$

After the “label” of the shape of the current evolving curve is known, a shape prior model Ψ^* can be constructed by using the statistical properties of that cluster

$$\Psi^* = U_k \bar{\Psi}(\Omega_k^*) + \bar{\Psi} \quad (4.29)$$

where $\bar{\Psi}(\Omega_k^*)$ is the means of the target cluster, U_k is the largest k principal eigenvectors of the PCA projection, and $\bar{\Psi}$ is the means of the whole training data. Also, Eqn. (4.29) is a reverse mapping of the point $\bar{\Psi}(\Omega_k^*)$ from the feature subspace to the original data space, and similar to the one in the previous method, Ψ^* is a reconstructed level set function of the point $\bar{\Psi}(\Omega_k^*)$.

Since the proposed shape prior model includes a classification step during its construction process, we name it a classification-based shape prior model. The advantages of this shape model are as follows: firstly, it is capable of generalizing to different

multiple shape patterns either from multiple objects or from multiple shape states of a single object. As long as the shape pattern of the evolving curve is within the training shapes, the proposed shape prior model can recover its original shape. Secondly, it can handle occlusions and cluttered background. The shape prior model constrains the evolving curve to follow a global shape consistency in the detection process. If there are disruptions in the object boundary, it can recover the original shape of object by referring to the shape information in the prior knowledge.

By incorporating the classification-based shape prior model into the GAC equation in the same way as described in section 4.2.3 Eqn. (4.17), the curve evolves under the constraints of both the image edge information and the shape prior information.

4.3.3 Experimental Results

This proposed shape prior model using the Gaussian mixture model is applied on the synthetic dataset, the human shape dataset, and the sign language dataset to assess its detection performance.

4.3.3.1 Synthetic dataset

The synthetic images shown in section 4.2.4.1 are used to test the proposed shape prior model on two aspects: generalization to multiple shape patterns and detecting object boundary in the presence of partial occlusions.

The level set functions of the synthetic images are computed and used as training data. They are projected onto the largest k (in order to retain over 90% of the total variance, k is chosen to be 5) eigenvectors using PCA. As there are multiple shape patterns in the training data, they form clusters in the low dimensional subspace. GMM is then used to approximate the data distribution. Figure 4.15 (a) shows the result of the PCA projection (data are embedded in a three dimensional subspace) and Figure 4.15 (b) shows the result of fitting GMM to the data distribution (data are

Figure 4.15: The dimensionality reduction and density estimation using PCA and GMM.

Figure 4.16: Decision boundary of the Bayesian classifier for synthetic data. It separates the entire region into three subregions according to the posterior probability.

embedded in a two dimensional space). A Bayesian classifier is created on the basis of GMM, with Figure 4.16 showing its decision boundary. The entire region has been divided into three subregions according to the posterior probability. To incorporate the shape prior model, the parameter d_1 in γ is set to be 1.3 to 2.4 according to the variance of each shape cluster and d_2 is empirically set to be $4d_1$.

Three test objects are artificially occluded by another object as shown in Figure 4.17

Figure 4.17: Comparison of the detection results of the proposed method with the traditional GAC method in the presence of object occlusions. Column (a) is the partially occluded objects, column (b) is the detection result of traditional GAC, column (c) is the Bayesian classification result (marked by $\star$) of the test shape, column (d) is the reconstructed shape prior model, and column (e) is the detection result of the proposed method.

column (a). It is obvious that the original shapes of the objects are destroyed. Hence, the traditional geodesic active contour method which depends on the image gradient force to constrain the curve evolution could not find the correct object boundary and converged to a wrong location (as shown in Figure 4.17 column (b)). In the proposed method, the shape of the curve under evolution is designated to a shape cluster which is most similar to the shape pattern of the curve by the Bayesian classifier. Figure 4.17 column (c) shows the classification results, each “ $\star$ ” in the figure represents the curve under evolution. It can be seen that the data all have been correctly classified to their corresponding shape clusters. In this way, a shape prior model is constructed by using the statistical properties of that cluster. It is a classification-based shape prior

Figure 4.18: Examples of training shapes from real human shape dataset.

Figure 4.19: Examples of the level set functions of human shapes.

model and thus is capable of generalizing to multiple shape patterns. For different shape patterns, it can provide different types of shape prior information as shown in Figure 4.17 column (d). The shape prior model will constrain further curve evolution to follow a global shape consistency learned from the training data. Figure 4.17 column (e) shows the detection results of the proposed method. The evolving curve avoided the spurious contours and successfully found the real boundary of objects.

Figure 4.20: Dimensionality reduction and density estimation on real human shape dataset using PCA and GMM.

Figure 4.21: Decision boundary of the Bayesian classifier for real human shape dataset. It separates the entire region into four subregions according to posterior probability.

4.3.3.2 Real human shape dataset

Now let us test our proposed method by using the real human shape dataset. Four human motion postures, such as standing (view from the front), standing (view in silhouette), standing with an open arm, and walking (view in silhouette), are selected as training shapes, with each posture containing 80 samples. They are aligned to a template of size 75×75 . Figure 4.18 shows the examples of training shapes, and Figure

4.19 shows their corresponding level set functions. These data are projected onto the first 8 eigenvectors in order to retain 99% of the total variance, then their distribution is approximated by GMM. Figure 4.20 (a) shows the result of the PCA projection and Figure 4.20 (b) shows the Gaussian components of GMM. Based on GMM, a Bayesian classifier is created which separates the entire region into 4 subregions. Figure 4.21 depicts its decision boundary. The parameters d_1 and d_2 in γ are set according to the variance of each shape cluster. The real human shape dataset is used to test the detection capability of the proposed shape prior model in the presence of real cluttered background.

The human object appears in front of bushes as shown in Figure 4.22 column (a). The spurious edges from the background severely interfered with the original shape of human objects. Thus, it is difficult to detect the real boundary by relying on the edge information only. The traditional geodesic active contour method performed poorly in the presence of this kind of “noise” as shown in Figure 4.22 column (b). However, by combining the shape prior model with GAC, the shape model will first find out the most similar shape cluster as shown in Figure 4.23, then constrain the evolving curve according to the shape characters of that cluster. Hence, the real contours of the human can be recovered successfully by the shape information captured in the prior model as shown in Figure 4.22 column (c). Moreover, with only one prior model, we can restore the original shapes of different human shape patterns. The proposed algorithm demonstrated more robust detection performance in the presence of cluttered background, and its ability to generalize to multiple shape patterns as well.

Another advantage of the proposed method is that, at the point of applying the shape prior model into the GAC framework, the shape of the evolving curve may be very different from the one on convergence. However, due to the redundant shape information in the signed distance map, the overall shape representative of the evolving curve shares the similarity with the shape of the object. Hence the curve still can be classified to the corresponding shape cluster.

Figure 4.22: Comparison of the detection results of our proposed method with the traditional GAC method in presence of cluttered background. Column (a) shows the partially occluded human shapes, column (b) shows the detection results of traditional GAC method, and column (c) shows the detection results of the proposed method. It can be seen that our proposed method can successfully recover the original human shapes in the presence of cluttered background.

4.3.3.3 Sign language dataset

The sign language is a kind of natural languages that has been widely used among hearing impaired people. It uses different hand gestures, each representing a certain meaning of words or sentences, for communication. In this experiment, it is employed to assess the ability of the proposed method to generalize to several different shape patterns.

Figure 4.23: Bayesian classification results of the partially occluded test shapes. Point 1 corresponds to the human shape of standing with open-armed, point 2 corresponds to the human shape of standing (view from the front), and point 3 corresponds to the human shape of standing (view from silhouette). It can be seen that all the testing data have been correctly classified to the most similar shape cluster by Bayesian classifier. A shape prior model can be constructed using the statistical properties of the corresponding cluster. The missing boundary of human shapes caused by object occlusion can thus be recovered using the shape prior knowledge.

The international sign language dataset, obtained from [137], is chosen for testing. Five hand gestures corresponding to the letters “A”, “B”, “L”, “R”, and “Y” are selected as training shapes. Each gesture contains 40 samples. Figure 4.24 shows the examples of these gestures. They are aligned to the same orientation, position and scale, with their level set functions computed. Then these data are projected onto the largest 7 eigenvectors by PCA and their distribution is approximated by GMM as shown in Figure 4.25. A Bayesian classifier is created using the GMM. Figure 4.26 depicts its decision boundary.

Now we test the proposed method on partially occluded hand gestures and compare

Figure 4.24: Examples of hand shapes in the international sign language dataset. Each hand gesture corresponds to a letter.

Figure 4.25: Dimensionality reduction and density estimation using PCA and GMM on sign language dataset.

it with the Cremers' method [125]. Figure 4.27 show the comparison results. Although it is claimed in Cremers' method that their shape prior model can also encode different multiple shape patterns, their method evolves the curve by minimizing a shape energy function using the gradient descent method. Hence, the optimization process may stop at a local extremum, and generates new shapes which are not contained in the training

Figure 4.26: Decision boundary of the Bayesian classifier for sign language dataset. The entire region has been separated into five subregions according to posterior probability.

shape patterns. Figure 4.28 plots the trace (white line) of the test shape in the energy minimization process in the projected 2D space by the Cremers' method. It can be seen from the figure that the curve evolved and finally converged to the local minimum of the shape energy instead of converging to the corresponding cluster of the most probable shape. Since the position of the local minimum is close to the shape cluster "Y", the recovered shape as shown in Figure 4.27 (c) is comparatively similar to the shape "Y". However, it is not an exact "Y" shape as it also includes part of the shape characteristics of the nearby cluster "L". Thus, the detection result of the Cremers' method is a new shape pattern between the shape of "Y" and that of "L". It is not a desired original shape for the partially occluded test data. For the sign language, a new shape pattern may directly mean a different letter or word, which may cause the misunderstanding in communications. Figure 4.29 shows the hand gesture of letter "W" in the sign language. Comparing the contour of the Cremers' method with the contour of letter "W", it can be seen that their contours are quite similar to each other and the detection result of the Cremers' method is probably suggesting the letter "W" although its original letter should be "Y". Hence, although the Cremers' method

Figure 4.27: Comparison of the detection result of our proposed method and Cremers' method. (a) Partially occluded sign gesture "Y". (b) Detection result of our proposed method. (c) Detection result of Cremers' method.

Figure 4.28: Trace of the test shape during the energy minimization process of Cremers' method in the projected 2D space. The curve finally converged to the local minimum of the energy function.

may be good for detecting objects with a large shape variation, it is not suitable for detecting the partially occluded hand gestures in the sign language.

On the contrary, as the proposed method performs a classification step first, the

Figure 4.29: Letter “W” in sign language. The detection result of Cremers’ method is quite similar to the letter “W” in sign language although its original letter should be “Y”. This detection error may directly cause misunderstanding in communications.

Figure 4.30: Bayesian classification result of the test shape of the proposed shape prior model. The test shape is assigned to a shape cluster according to the posterior probability.

test shape is assigned to a shape cluster (as shown in Figure 4.30) and then recovered by the shape information in that cluster. It ensures that the recovered shape must be a recognized shape pattern in the training data and that no unfamiliar shapes will be generated. In this way, our proposed method shows better detection performance in handling this kind of cases.

Figure 4.31: Testing the tolerance of the shape model to the extent of shape distortions.

Figure 4.32: Classification results of the shape of cutting a circle with the amount of 10%, 20%, 30%, 40%, 50% and 60% respectively. It can be seen that when the shape distortion reaches 50%, the remaining shape information is not sufficient to identify the object, leading to a wrong classification result.

4.3.3.4 Tolerance to the extent of occlusions

Although we have demonstrated that the proposed method performs well in presence of object occlusion and cluttered background, the tolerance to the degree of shape distortions is yet to be determined. Hence, in the following, we gradually increase the amount of shape distortions to test to what degree the proposed shape model can still correctly recover the object shapes. The experiments is tested on the synthetic dataset.

Figure 4.31 shows the results of cutting of a circle from the amount of 10% to 60% respectively. It can be seen that, when the distortion reaches 50%, the remaining

shape information of the circle is not sufficient to identify the original object, resulting in a wrong classification result as shown in Figure 4.32. As a consequence, the curve is constrained by a shape prior model constructed from a wrong shape cluster and it can only converge to a wrong shape pattern. Hence, in order to obtain a correct classification result, the remaining shape information should be above 50%.

4.4 Learning a Shape Prior Model Using Kernel PCA and GMM

4.4.1 Motivations and Objectives

As mentioned in sections 4.2 and 4.3, most of the existing shape prior models [46, 127, 128] employ a linear mapping (e.g. PCA) to project the training data onto a low dimensional subspace. Then the shape prior model is constructed by the linear weighted sum of k principal eigenshapes plus a mean shape. However, very often the object shapes may undergo nonlinear deformations. For example, when a human walking under a surveillance camera, its shapes may follow a nonlinear change due to the change of viewing angle and silhouette deformation. In this situation, a linear projection such as PCA is unable to capture the intrinsic structure of data distribution in subspace [134]. It is noticed that Cremers [125] proposed a recognition-based shape prior model which employs a kernel mapping on the training data first, then models the distribution with a single Gaussian function. The obtained probability density function is treated as a shape energy term to constrain the curve evolution. This shape prior model is able to generalize to multiple different shape patterns. Moreover, by minimizing the energy function towards a local extremum, new shape patterns which are not in the training samples may be created. These were considered by Cremers to be the superiority of his approach. However, as demonstrated in section 4.3, although Cremers' shape prior model is able to handle more shape variations, there are cases

where the method may not recovered the correct shapes..

Hence, in this section, a new shape prior model is proposed and is used as a curve evolution regularizer for variational methods. The training shapes are collected from a series of object shapes seen under different viewing angles or silhouette deformations. They are considered to form clusters in a submanifolds after a nonlinear dimensionality reduction using kernel principal component analysis (KPCA). Their level set functions are also assumed to lie in a low dimensional feature subspace and form clusters after KPCA. A Gaussian Mixture Model (GMM) is used to estimate the distribution of multiple clusters in the subspace. The obtained probability density function is used as a shape energy term and is incorporated into the geodesic active contour function as a regularizer. During the curve evolution process, the energy function is reformulated to contain a classification scheme which will designate the curve to a most similar shape cluster based on the posterior probability. The advantages of the proposed method are that, firstly, a nonlinear dimensionality reduction method, KPCA, is used to project the training data onto a subspace. The nonlinear shape variations in the training data are removed and the data distribution in subspace can be represented more efficiently. Secondly, the main difference of the proposed method with Cremers' method is that, unlike Cremers' recognition-based shape prior model, the proposed method is a classification-based shape prior model. The recovered shape pattern must be within the training shapes. Thus the proposed method can handle situations which cannot be solved by Cremers' method. Experiments demonstrate the advantages of the proposed method.

4.4.2 Constructing the Shape Prior Model

To construct the nonlinear shape prior model, the training shapes need to be aligned first. Then the level set functions of the shapes are computed and used as training data. These two steps are the same as described in sections 4.2.2.1 and 4.2.2.2. Hence, we will present the construction process directly.

Figure 4.33: The basic idea of kernel PCA. In some high dimensional feature space F (bottom right), linear PCA is performed, just as a PCA in input space (figure (a)). Since F is nonlinearly related to input space (via Φ), the contour lines of constant projections onto the principal eigenvector (drawn as an arrow) become nonlinear in input space. Crucial to kernel PCA is the fact that we do not actually perform the mapping into F , but instead perform all necessary computations by the use of a kernel trick k [3].

4.4.2.1 Kernel PCA

Kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the original linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping [138]. Figure 4.33 shows the basic idea of KPCA. In the high dimensional space (can be infinitely high), the mapped data will distribute in a way that they are more likely to be linearly separated [3, 135].

Given a set of variables $\mathbf{x}_i$, $i = 1, \dots, n$, $\mathbf{x}_i \in \mathbb{R}^N$, define a nonlinear mapping $\Phi : \mathbf{x} \in \mathbb{R}^N \rightarrow \mathbf{X} \in \mathbb{F}$, where $\mathbb{F}$ refers to the *feature space*. Here the data are assumed to be centered so that $\sum_{i=1}^n \Phi(\mathbf{x}_i) = 0$. The covariance matrix in $\mathbb{F}$ is

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^\top \quad (4.30)$$

Performing the eigen decomposition on $\bar{C}$ with $\lambda \mathbf{V} = \bar{C} \mathbf{V}$. Note that the solutions $\mathbf{V}$ lie in the span of $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)$, so the following equations

$$\lambda (\Phi(\mathbf{x}_i) \cdot \mathbf{V}) = (\Phi(\mathbf{x}_i) \cdot \bar{C} \mathbf{V}) \quad \forall i = 1, \dots, n \quad (4.31)$$

$$\text{and } \mathbf{V} = \sum_{i=1}^n \beta_i \Phi(\mathbf{x}_i) \quad (4.32)$$

can be generated, where β_i being the coefficients.

Define a $n \times n$ matrix K by $K_{ij} := (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$, this dot product matrix can be computed by kernel function $k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}))$. Thus, the eigenvalue problem is converted to

$$M \lambda \beta = K \beta \quad (4.33)$$

To extract the principal components, the test data should be projected onto the eigenvectors $\mathbf{V}_l$. For our case, let the level set function Ψ be a test data. The projected data in the feature subspace is

$$(kPC)_l(\Psi) = (\mathbf{V}_l \cdot \Phi(\Psi)) = \sum_{i=1}^n \beta_i^l k(\mathbf{x}_i, \Psi) \quad (4.34)$$

A more detailed discussion on the KPCA method can be found in [3, 135].

4.4.2.2 Density estimation in subspace

Since the training shapes are collected from a series of object shapes seen from different viewing angles or the object silhouette undergoes non-linear deformations, they may form clusters in the feature subspace. Hence, a Gaussian Mixture Model (GMM) [112] is employed to approximate the data distribution in the subspace. This GMM method is an unsupervised learning method which can automatically estimate the proper number of Gaussian components. This is desirable when the object shapes are extracted from video sequences since the data can thus be clustered according to their shape properties. We do not need to know the number of clusters in advance. Based on [112], the desired number of Gaussian components is determined by finding a minimum message length $\mathcal{L}$ as shown in Eqn. (4.27).

Thus, the final probability density function is

$$P(\mathbf{y}) = \sum_{m=1}^k \alpha_m p(\mathbf{y}|\theta_m) \quad (4.35)$$

with k being the number of components, $\alpha_1, \dots, \alpha_k$ being the mixing probabilities, and each $p(\mathbf{y}|\theta_m)$ being a multivariate Gaussian distribution

$$p(\mathbf{y}|\theta_m) = \frac{(2\pi)^{-\frac{d}{2}}}{\sqrt{|\Sigma_m|}} \exp \left[-\frac{1}{2}(\mathbf{y} - \mu_m)^\top \Sigma_m^{-1}(\mathbf{y} - \mu_m) \right] \quad (4.36)$$

4.4.2.3 Combining the shape prior model with GAC equation

For a GAC curve evolving towards the object boundary, the probability density function of its level set function Ψ can be computed as

$$P(\tilde{\Psi}) = \sum_{m=1}^k \alpha_m p(\tilde{\Psi}|\theta_m) \quad (4.37)$$

where $\tilde{\Psi}$ is the projected vector after using KPCA.

Our task is to regularize the GAC curve towards a preferred shape pattern so that its probability P can be maximized. For computational convenience, this problem is converted to minimize the negative log of the probability, and is defined to be a shape energy term

$$E_{shape} = -\log P(\tilde{\Psi}) \quad (4.38)$$

In order to minimize Eqn. (4.38), we search for the gradient descent direction with respect to the level set function Ψ

$$\frac{\Delta \Psi}{\Delta t} = \frac{\partial E_{shape}}{\partial \Psi} = \frac{\partial E_{shape}}{\partial \tilde{\Psi}} \cdot \frac{\partial \tilde{\Psi}}{\partial \Psi} \quad (4.39)$$

The first term on the right hand side of Eqn. (4.39) can be computed by referring to Eqn. (4.37)

$$\begin{aligned} \frac{\partial E_{shape}}{\partial \tilde{\Psi}} &= \frac{\partial \log P(\tilde{\Psi})}{\partial \tilde{\Psi}} = \frac{1}{P(\tilde{\Psi})} \cdot \frac{\partial P(\tilde{\Psi})}{\partial \tilde{\Psi}} \\ &= -\frac{1}{P(\tilde{\Psi})} \sum_{m=1}^k \alpha_m p(\tilde{\Psi}|\theta_m) \cdot \Sigma_m^{-1}(\tilde{\Psi} - \mu_m) \end{aligned} \quad (4.40)$$

To calculate the second term on the right hand side of Eqn. (4.39), it is noticed that $\tilde{\Psi}$ is the projected vector of Ψ after using KPCA, hence

$$\tilde{\Psi} = \sum_{i=1}^n \beta_i^l k(\Psi_i, \Psi) \quad (4.41)$$

Moreover, the kernel function $k(\mathbf{x}, \mathbf{y})$ employed in this method is a Gaussian radius basis function (RBF)

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (4.42)$$

So,

$$\frac{\partial \tilde{\Psi}}{\partial \Psi} = -\sum_{i=1}^n \beta_i^l k(\Psi_i, \Psi) \cdot \left(\frac{\Psi_i - \Psi}{\sigma^2}\right) \quad (4.43)$$

Finally, Eqns. (4.40) and (4.43) are put into Eqn. (4.39) to generate the final expression of the curve evolution function

$$\begin{aligned} \frac{\Delta \Psi}{\Delta t} = & \frac{1}{P(\tilde{\Psi})} \cdot \left\{ \sum_{m=1}^k \alpha_m p(\tilde{\Psi} | \theta_m) \cdot \Sigma_m^{-1}(\tilde{\Psi} - \mu_m) \right\} \\ & \cdot \left\{ \sum_{i=1}^n \beta_i^l k(\Psi_i, \Psi) \frac{(\Psi - \Psi_i)}{\sigma^2} \right\} \end{aligned} \quad (4.44)$$

The first term on the right hand side of Eqn. (4.44) acts as an adaptive increment at each iteration step. If the shape of evolving curve is far away from the preferred shape pattern, the probability $P(\tilde{\Psi})$ will be small, which results in a large increment step of the curve evolution. On the other hand, if the shape of evolving curve is similar to the shape prior model, then probability P will be large and the curve evolution process will change slowly until it finally converges. The second term in Eqn. (4.44) is derived from differencing shape energy with respect to the projected vector $\tilde{\Psi}$. During the energy minimization process, this term evolves the curve in the direction of maximum probability. Thus, it controls the convergence of the curve. The last term in Eqn. (4.39) explicitly links the shape information in the training data with the current test shape. It is the main force to recover the pattern of the object shape.

Now we need to analyze the second term in Eqn. (4.39) again. As aforementioned, it evolves the curve towards the maximum probability. However, as P is the probability function of mixture of Gaussian components, it has local maxima. But what we need is a shape energy term that can regularize the curve evolution to follow a globally shape consistency instead of a local one. Hence, the second term in Eqn. (4.39) is replaced with a classification term

$$\left[\max_{\tilde{\Psi}} \alpha_m p(\tilde{\Psi} | \theta_m) \right] \cdot \Sigma_m^{-1}(\tilde{\Psi} - \mu_m) \quad (4.45)$$

Equation (4.45) can be considered as a Bayesian classifier which will “classify” the current evolving curve to one of the shape clusters in the feature subspace according to

the posteriori probability. In this way, we can avoid the local extremum in the energy function and obtain a reliable result.

Rewrite Eqn. (4.44) in an iterative equation

$$\Psi_{n+1} = \Psi_n + \Delta t_1 \cdot A \quad (4.46)$$

where A represents

$$\begin{aligned} A = & \frac{1}{P(\tilde{\Psi}_n)} \cdot \left\{ \left[\max_{\tilde{\Psi}_n} p(\tilde{\Psi}_n | \theta_m) \right] \Sigma_m^{-1} (\tilde{\Psi}_n - \mu_m) \right\} \\ & \cdot \left\{ \sum_{i=1}^n \beta_i^l k(\Psi_i, \Psi_n) \frac{(\Psi_n - \Psi_i)}{\sigma^2} \right\} \end{aligned} \quad (4.47)$$

Rewrite the geodesic active contour function Eqn. (2.21) in an iterative equation also

$$\Psi_{n+1} = \Psi_n + \Delta t_2 \cdot (g(c + \kappa) |\nabla \Psi_n| + \nabla \Psi_n \cdot \nabla g) \quad (4.48)$$

Combining Eqns. (4.46) and (4.48) together, we can obtain the final expression of our curve evolution function in an iterative equation

$$\Psi_{n+1} = \Psi_n + \Delta t_2 \cdot (g(c + \kappa) |\nabla \Psi_n| + \nabla \Psi_n \cdot \nabla g) + \Delta t_1 \cdot A \quad (4.49)$$

The image gradient information in Eqn. (4.48) provides an edge-driven component to Eqn. (4.49), while the shape energy term in Eqn. (4.46) provides a knowledge-driven component to Eqn. (4.49). Thus, the curve can not only capture the local deformation based on edge information, but also preserve a global shape consistency under shape prior model.

4.4.3 Experimental Results

In this section, the proposed shape prior model is evaluated. Synthetic images and real human shapes collected from video sequence are used for testing. Since it will be difficult for us to adjust the pose of the shape prior model to the evolving curve if they are too dissimilar (e.g. the initial contour of the object is far away from the object), it is proposed that the curve is moved according to the image gradient first, after becoming stationary, then include the shape prior model to constrain further curve evolution.

4.4.3.1 Synthetic images

The synthetic images as shown in Figure 4.1 are tested first. These shapes are triangle, cross and circle. Their level set functions are computed and used as training data. They are projected onto a low dimensional feature subspace using KPCA (in order to retain sufficient total variance, data are projected onto the first 11 largest eigenvectors), and their distribution is estimated using GMM. Figure 4.34 (a) shows the data distribution in the feature subspace after KPCA (data are viewed in a 3D space), and Figure 4.34 (b) show the corresponding Gaussian components in the subspace (data are viewed in a 2D space).

Now we partially occlude the objects with another object as shown in Figure 4.35 column (a). It is obvious that the original shapes of objects are destroyed, and thus the traditional geodesic active contour method which evolves the curve based on edge information could not find the real contour of objects and converged to wrong location as shown in Figure 4.35 (b). However, by incorporating the proposed shape prior model with geodesic active contours, the curve evolves under the influence of both the edge information and the shape prior knowledge. Hence, it can correctly recover the original contours of objects as shown in Figure 4.35. Figure 4.42 show the energy minimization process of the proposed nonlinear shape prior model. Since Eqn. (4.45) is reformulated by a “classification” term, the curve is forced to converge to the mean of each corresponding cluster instead of converging to the local minimum of the function.

Figure 4.34: Results of performing dimensionality reduction using KPCA and density estimation using Gaussian mixture model on the synthetic data.

Table 4.1: Computational time of the proposed method on the partially occluded synthetic images.

Partially occluded circle	Partially occluded cross	Partially occluded triangle
6.528 sec	8.352 sec	6.779 sec

From Figure 4.42 it can be seen that each shape (upper line corresponds to the circle and lower line corresponds to the cross) moved exactly towards its shape cluster and finally converged to the center. The detection results of our proposed method match with the energy minimization results and the shape of each occluded object is recovered by the information from prior knowledge.

Finally, the computation time of the proposed method on the synthetic images is evaluated. It is noted that the final curve evolution equation (4.46) involves an amount of calculations for the shape prior term Eqn. (4.47). Hence, the proposed method generally needs more time to detect the object boundary. We perform the proposed method on the partially occluded synthetic images as shown in Fig. 4.35 on a AMD Athlon 64 4000+ dual core CPU (2.1GHz). Table 4.1 shows the computational

Figure 4.35: Detection results of the proposed nonlinear shape prior model and the traditional geodesic active contour method on partially occluded objects. Column (a) is the examples of partially occluded objects. Column (b) is the detection results of the traditional geodesic active contours. Column (c) is the detection results of the proposed method.

time of the detection process. It can be seen that for synthetic images (size 64×64), it takes around 6 to 8 seconds for the proposed method to finally find the real object boundary.

4.4.3.2 Real human shape from surveillance video sequence

The proposed shape prior model is then tested on the images from real video sequence which are occluded by artificially created obscuration. The video sequences are obtained from CAVIAR database [139]. Five human shape states, such as walking, standing and raising arms during a walking process are selected as training shapes to build the shape prior model. Figure 4.37 shows the examples of these shapes after alignment and Figure 4.38 shows the corresponding level set functions. These data are projected by KPCA onto a low dimensional feature subspace as shown in Figure 4.39 (in order to retain over 90% of the total variance, the number of principal components k in KPCA is chosen to be 5), and their distribution is approximated by GMM. Figure 4.40 illustrates the result of fitting GMM with the training data in the feature subspace.

Figure 4.36: Energy minimization process of the proposed method. Upper line corresponds to the minimization of partially occluded circle. Lower line corresponds to the minimization of partially occluded cross. Finally, they both converged to the center of the most similar shape cluster respectively.

Figure 4.37: Examples of alignment training shapes.

The human shapes are partially occluded with a horizontal bar as shown in Figure 4.41 column (a). It can be seen that the original human shapes are destroyed. Hence the traditional geodesic active contour method which evolves the curve only based on edge information could not find the correct human contour and converged to wrong location as shown in Figure 4.41 column (b). By incorporating the shape prior model with GAC algorithm and evolving the curve both under the influence of edge information and the shape prior knowledge, it can be seen that the original human contours can

Figure 4.38: Examples of training data. Each image is the level set representation of a corresponding shape.

Figure 4.39: Training data in feature subspace after KPCA. Different color corresponds to different shape state (data are viewed in a 3D space).

be successfully recovered.

Figure 4.42 show the examples of energy minimization process of two shapes. The two traces in the figure represent the convergence process of two human shape states. It can be seen that each shape moves exactly towards its corresponding shape cluster and finally converges to its center.

Finally, the computational time of the proposed method on the real human shapes are evaluated. Five partially occluded human shapes shown in Fig. 4.41 are used for

Figure 4.40: Fitting GMM to our training data in subspace. Different colors correspond to the different shape clusters (Data are viewed in a 2D space).

Table 4.2: Computational time of the proposed method on the partially occluded real human shapes.

Shape in row 1	Shape in row 2	Shape in row 3	Shape in row 4	Shape in row 5
12.074 sec	11.546 sec	10.952 sec	12.676 sec	10.421 sec

testing. Table 4.2 shows the computational time of the detection process. It can be seen from the data that the computational time for real human shapes is around 10 to 12 seconds.

The detection process of the real human shapes are a little slower than the detection of the synthetic images. It is because firstly, the size of the training shape template of the real human shapes is larger than the synthetic images. For real human shapes, the size of the training shape template is 80×80 , while for synthetic images, it is only 64×64 . The parameter n in Eqn. (4.47) of the real human shapes is larger than the one in the synthetic images. Secondly, there are more shape clusters in the real human shapes (5 clusters) than in the synthetic images (3 clusters). The parameter k in Eqn. (4.47) of the real human shapes is larger than the one in the synthetic images as well.

Figure 4.41: Detection results of proposed method on different human shape patterns. Column (a) is the partially occluded human shapes, column (b) is the detection results of traditional GAC, column (c) is the detection results of our proposed method.

Hence, there are more calculations involved in the curve evolution process of the real human shapes than the synthetic images, and the detection process is slowed down by

Figure 4.42: Energy minimization process of our proposed method. The two traces in the figure represent the convergence process of two human shape states. They start from an obscured shape and finally converged to the center of each shape cluster.

the extra calculations correspondingly.

4.5 Comparison of the three proposed shape prior models

In this chapter, three shape prior models are proposed to solve the missing boundary problem encountered when detecting the object boundary. In the following, these methods are compared with each other in several aspects to evaluate their detection performance.

4.5.1 Construction process

Table 4.3 shows the comparison of the construction process of all three shape prior models. Generally speaking, all the shape prior models need a dimensionality reduction and a density estimation step in the construction process. And the shape prior models are all built from a probability density function.

Table 4.3: Comparison of the construction process of the three proposed shape prior models.

Shape prior model built on PCA + a single Gaussian	Shape prior model built on PCA + GMM	Shape prior model built on KPCA + GMM
The training shapes are projected onto a low dimensional feature subspace using principal component analysis, and then the data distribution is approximated by a single Gaussian function. The obtained probability density function is used as shape prior model.	The training shapes are projected using PCA, and then the data distribution is approximated by a Gaussian mixture model. A Bayesian classifier is created based on GMM. The Bayesian classifier first designates the evolving curve to a most similar shape cluster in GMM, then utilizes the shape prior information from that cluster to constrain the curve evolution.	The training shapes are projected using kernel principal component analysis. In the feature subspace, the data distribution is approximated by GMM. The obtained probability density function is revised and used as a shape prior model.
Simple	Medium	Complex

However, the differences among them are that, the first and second shape prior models are created by linearly projecting the training data onto a low dimensional subspace, while the third one involves a nonlinear mapping of the training data first, then performs the projection in the kernel space. Hence, the third shape prior model is able to handle shape patterns under nonlinear shape deformations. Secondly, the first shape prior model is constructed using a single Gaussian estimation, while the second and third ones using a Gaussian mixture model. Hence, the second and third shape prior models are able to generalize to multiple shape patterns and are able to handle larger shape variations than the first one.

Table 4.4: Comparison of the application area of the three proposed shape prior models.

Shape prior model built on PCA + a single Gaussian	Shape prior model built on PCA + GMM	Shape prior model built on KPCA + GMM
As the training shapes are collected from a single shape pattern, it can only be used to detect a single shape pattern.	As the training shapes are collected from multiple different shape patterns, it can be used to detect multiple shape patterns either from multiple objects or from multiple shape states of a single object.	As the training shapes are collected from object shapes under nonlinear deformations, it can be used to detect the shapes under nonlinear deformations.

4.5.2 Application area

It is noted from Table 4.4 that for the first proposed shape prior model, the training shapes are collected from a single shape pattern. Hence, this shape prior model can be applied to detect one single shape pattern only. While for the second shape prior model, there are multiple different shape patterns collected from either multiple objects or multiple shape state of a single object. Hence, it is able to generalize to multiple shape patterns, and can be applied to detect multiple different shapes. Finally, for the third proposed shape prior model, the training shapes are collected from shape patterns under nonlinear deformation (such as a human shape moving in a video sequence). Hence, this shape prior model is able to detect the shapes under nonlinear deformations.

4.5.3 Computational time

Table 4.5 shows the comparison of the computational time of all three shape prior model. Since the first proposed shape prior model is the simplest one, there are less calculations involved in the construction process. Hence, its detection process is relatively fast, usually around 3 to 4 seconds. However, for the third shape prior model, there

Table 4.5: Comparison of the computational time of the three proposed shape prior models.

Shape prior model built on PCA + a single Gaussian	Shape prior model built on PCA + GMM	Shape prior model built on KPCA + GMM
The average computational time is around 3 to 4 seconds.	The average computational time is around 4 to 6 seconds.	The average computational time is around 10 to 12 seconds.
Fast	Medium	Slow

Table 4.6: Comparison of the cognition level of the three proposed shape prior models.

Shape prior model built on PCA + a single Gaussian	Shape prior model built on PCA + GMM	Shape prior model built on KPCA + GMM
Can detect object shapes with small shape variations from images. The training shapes must be approximatable after a linear projection.	Can detect object shapes with small shape variations from images. The training shapes must be approximatable after a linear projection. Can detect multiple different shape patterns.	Can detect object shapes under nonlinear shape deformations from video sequences.
Low	Medium	High

are many calculations involved in its curve evolution equation. Its detection process is much slower than the other two, usually around 10 to 12 seconds. For the second shape prior model, as it involves an extra classification step comparing with the first shape prior model, it is a little slower than the first one, but still faster than the third shape prior model.

4.5.4 Cognition level

At last, the cognition level of all three shape prior models are compared in Table 4.6. It is noted from the previous analysis that, the first shape prior model is the simplest and fast one among all. It can only detect a single shape pattern with small shape variations from images. Hence, its cognition level is low. For the second shape prior model, it can be applied to more generalized cases and can detect multiple different shape patterns. Hence, its cognition level is higher than the first one. For the third shape prior model, the training shapes are considered to form clusters in a submanifolds after a nonlinear dimensionality reduction using KPCA. It can handle the case of nonlinear shape deformations which cannot be solved by the previous two methods. It is more powerful and its cognition level is high.

4.6 Discussions and Conclusions

In this chapter, three new methods that incorporate the shape prior model to solve the “missing boundary” problem are proposed. The shape prior information is incorporated into the geodesic active contour framework to regularize the curve evolution process.

In the first shape prior model, the training shapes are collected from a single shape pattern. The data distribution is modeled by a single Gaussian function after the principal component analysis (PCA). The constructed shape prior model can only encode a single shape pattern. In the second proposed method, we extend the first method by modeling the distribution of training data with a Gaussian mixture model after the PCA projection. Hence, the second shape prior model is capable to generalize to multiple shape patterns from either multiple objects or multiple shape states of a single object. In the third proposed method, the training shapes are collected from a series of object shapes seen under different viewing angles or silhouette deformations. They are considered to form clusters in a submanifolds after a nonlinear dimensionality

reduction using kernel principal component analysis (KPCA). A Gaussian Mixture Model (GMM) is used to estimate the distribution of multiple clusters in the subspace.

To accelerate the convergence under the influence of shape prior knowledge of these proposed method, an explicit alignment method is used to adjust the shape model to the same pose as the shape of curve under current evolution. Moreover, a weighting scheme is developed to balance the influence between the image gradient and shape prior information by quantifying their differences.

The first and second proposed shape prior models have been tested on the synthetic dataset, real human shape dataset or the sign language dataset in the experimental part to demonstrate their detection ability. The results show that the proposed methods can successfully recover the original object shapes in the presence of partial occlusions and the cluttered background. The whole algorithm can be implemented fast, with no heavy computation induced. The third proposed method has been tested on synthetic images and real human shapes from surveillance video sequence. The results demonstrate that the proposed shape prior model can successfully handle the object occlusion problems in video sequences.

Chapter 5

Conclusions and Future Work

*I could be bounded in a nutshell
and count myself a king of infinite space...*

— Shakespeare, Hamlet, Act 2, Scene 2

5.1 Conclusions

In this thesis, robust object boundary detection by the variational methods is studied. The variational methods fall into two categories: boundary-based and region-based. This thesis focuses on the boundary-based variational methods. However, as the boundary-based variational methods depend only on the boundary information to locate the object, they are very sensitive to the disruptions to the object boundary. If the object is partially occluded by other objects or is interfered by cluttered background, the evolving curve of the variational methods may be pulled away from the real boundary and converge to the wrong place. This is regarded as the “**missing boundary**” problem in this thesis. In order to achieve a robust object detection, additional boundary constraints need to be incorporated into the variational methods to regularize the curve evolution. In this thesis, two new boundary constraints are proposed.

5.1.1 Regularizing Variational Methods by Adding Temporal Information

In the application of boundary-based variational method to the detecting of inner heart wall boundary of left ventricle from the echocardiographic image sequence, we proposed utilizing the temporal information to regularize the variational methods.

The echocardiographic data suffer from signal dropouts which weaken or totally destroy the ventricular boundary. Hence, part of the heart wall boundary may be missing due to the dropouts. The traditional boundary-based variational methods cannot find the correct heart wall boundary. To solve this “missing boundary” problem, a new approach is proposed which incorporates the temporal information into the variational methods to regularize the curve evolution process. The ventricular boundary is divided into strong and weak segments. The weak segments are considered to be destroyed by signal dropouts and they are hard to be detected. Temporal information from neighboring frames is incorporated into the variational methods to re-construct the missing boundary and strengthen the gradient force in the weak segments. The new aspects in this approach are the following: (i) a Gaussian Mixture Model (GMM) is employed to calculate the thresholds that quantitatively partition the strong and weak segments; (ii) the temporal information is incorporated as a regularizer into the external energy of active contour algorithm and the speed function of level set algorithm as shown in Eqn. (3.10) and Eqn. (3.14); (iii) a weight is developed to balance the strength between image gradient force and temporal information, and a strengthening factor is introduced to boost the total energy (image gradient force and temporal information) up to the average level.

To evaluate the detection performance of the proposed methods, the active contour method using temporal information was compared with the ground truth, the traditional active contour method, and two other methods that also incorporate temporal information into active contour equation. The level set method using temporal information was compared with the ground truth, the traditional level set method and the

active contour without edges method which employs region information as a regularizer. The detection results are assessed according to their detection performance with respect to the mean absolute distance (MAD), Jaccard similarity coefficient $\mathcal{J}$, and the computational burden. The experiments demonstrate that the proposed methods are superior to the compared methods with little additional computational burden.

5.1.2 Regularizing Variational Methods by Adding Shape Prior Model

Since there are cases of the “missing boundary” problem that cannot be handled by only using the temporal information, in the second approach, the shape prior information is used to regularize the curve evolution process of the variational methods. Three shape prior models are proposed.

In the first proposed shape prior model, the training data are projected to a low dimensional subspace using principal component analysis (PCA), and the data distribution is approximated by using a single Gaussian function. In the second proposed shape prior model, we extend the first one by modeling the distribution of training data in the subspace using a Gaussian mixture model (GMM). Hence, the second proposed method is able to generalize to multiple shape patterns from either multiple objects or multiple shape states of a single object.

To accelerate the convergence under the influence of shape prior knowledge, an explicit alignment method is used to adjust the shape model to the same pose as the shape of curve under current evolution. Thus, less iterations are needed than the Leventon’s method. Also, a weighting scheme has been developed to balance the influence between the image gradient and shape prior information by quantifying their differences.

The proposed shape prior models have been tested on synthetic dataset, real human shape dataset and sign language dataset in the experimental part to display their

detection abilities. The results show that the proposed method can successfully recover the original object shapes in the presence of partial occlusions and cluttered background. The entire framework can be implemented fast, with little additional computation incurred.

Although the linear shape prior models are efficient to solve the object occlusion and cluttered background, however, the shape prior models are constructed by the principal component analysis and Gaussian density estimation of the training shapes in the subspace. This means that the variations of training shapes are assumed to correspond to the linear combination of a set of eigenvectors. However, very often the object shapes may undergo nonlinear deformations, such as in the cases that person walks under a surveillance camera. The shapes may follow a nonlinear change due to the change of viewing angles and silhouette deformation. In this case, a linear projection such as PCA cannot explore the intrinsic structure of data distribution.

Hence, in the third method, a nonlinear shape prior model is proposed to regularize the curve evolution of the variational methods. The shapes that are collected from either multiple objects or multiple shape states of a single object are used as training data. Their level set functions are computed and assumed to lie in a low dimensional subspace after a nonlinear mapping (KPCA). As the data may form clusters in the subspace, the Gaussian mixture model (GMM) is employed to approximate the data distribution. The obtained probability density function is treated as a shape energy term and is incorporated into the geodesic active contour function to constrain the curve evolution. Experimental results show that the third shape prior model can successfully handle the object occlusion problems in video sequences and show good detection performance.

5.2 Future Work

The variational methods play a major role in the computer vision for object detection and tracking. Improving the detection performance by incorporating additional shape constraints into the variational methods to regularize the curve evolution process is still considered an effective way. The promising research areas of adding regularizer to the variational methods in the future include:

- **A fast, and real-time detection process:** Recently, more computer vision techniques have been applied to the daily life. For example, face detection and recognition have been adopted in digital cameras for automatically locating and focusing the human faces. It is believed that the variational methods will be applied to the daily life soon. The variational methods can be commonly used in the surveillance system in banks, stores, hospitals, etc. for real-time object detection and tracking. In order to meet the requirement of real-time detection, a fast algorithm is preferred.

However, when additional information is incorporated into the variational methods, more computation time is usually needed as the algorithms have to handle more parameters or calculations. The algorithm will be slowed down. It is, therefore, necessary to develop new approaches which can greatly reduce the computational burden and meet the requirement of real-time processing.

- **The ability to detect multiple objects simultaneously:** The state-of-the-art variational methods (such as level sets, geodesic active contours) are all capable of detecting multiple objects simultaneously. However, when the shape prior constraints are imposed on these methods, the variational methods may lose their capability of topological changes. They are only capable of detecting one single object which has the same shape prototype as the prior model simply. Although Cremers et al. [124] has proposed a dynamic labelling method to identify and match each object with the shape prior model when solving for multiple

object detection, there are cases (such as mutual occlusion of two target objects) that cannot be solved by their method. Hence, it is imperative to explore new methods which can be more effective.

- **Combine temporal information and shape prior information together to regularize the variational methods:** As shown in this thesis, the temporal information and shape prior information are both efficient regularizers for variational methods. Since the object detection and tracking are commonly used in surveillance videos, it is possible to develop new methods that can combine both temporal information and shape prior information as a mixed regularizer to constrain the curve evolution.

To sum up, temporal information and shape prior information have been explored to regularize the curve evolution process of variational methods for the respective applications in this thesis. They are incorporated into the variational methods to regularize the curve evolution process in the presence of the “missing boundary” problem. Since it is preferred that the variational methods can be used in real-time video sequences for object detection and tracking, it is recommended that new methods which can meet the requirement of real-time processing should be developed. Moreover, as the variational methods may lose the capability of topological changes when the regularizers are incorporated, it is necessary to explore new methods which can detect multiple objects simultaneously. In conclusion, it is vital to combine both temporal information and shape prior information as a mixed regularizer to limit the curve evolution process when the objects are detected from a video sequence.

Author's Publications

1. Wen Fang, Kap Luk Chan, Sheng Fu, Shankar Muthu Krishnan, "Incorporating Temporal Information into Active Contour Method for Detecting Inner Heart Wall Boundary from Echocardiographic Image Sequence", accepted by Journal of Computerized Medical Imaging and Graphics, June 2008.
2. Wen Fang, Kap Luk Chan, Sheng Fu, Shankar Muthu Krishnan, "Incorporating Temporal Information into Level Set Functional for Robust Ventricular Contour Detection from Echocardiographic Image Sequence", accepted by IEEE Transactions on Biomedical Engineering, in press, 2007.
3. Wen Fang and Kap Luk Chan, "Incorporating shape prior into geodesic active contours for detecting partially occluded object", Pattern Recognition, vol. 40, Issue 8, pp. 2163-2172, Aug. 2007.
4. Jia-Yin Zhou, Wen Fang, Kap-Luk Chan, V.F.H. Chong, James B.K. Khoo, "Extraction of Metastatic Lymph Nodes from MR Images Using Two Deformable Model-based Approaches", Journal of Digital Imaging. Springer, vol.20, No. 4, pp. 336-346, January 2007.
5. Wen Fang, Kap Luk Chan, Sheng Fu, S M Krishnan, "Incorporating Temporal Information for Ventricular Contour Detection in Echocardiographic Image Sequences", Proceedings of the IEEE 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS'05), pp. 1099-1102, 1-4 Sept. 2005 Shanghai, PR China. (Student Paper Competition Open Finalist, Certificated by IEEE Engineering in Medicine and Biology Society).
6. Wen Fang, Kap Luk Chan, "Using Statistical Shape Priors in Geodesic Active Contours for Robust Object Detection", Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Vol. 2, pp. 304-307, Aug. Hong

Kong 2006. (Awarded Certificate by the Pattern Recognition and Machine Intelligence Association (PREMIA)).

7. Peng Li, Kap Luk Chan, Wen Fang, “Hybrid Kernel Machine Ensemble for Imbalanced Data Sets”, Proceedings of the 18th International Conference on Pattern Recognition (ICPR’06), Hong Kong, August 2006.
8. Wen Fang, Kap Luk Chan, “Learning a New Statistical Shape Prior Model for Object Detection by Geodesic Active Contours”, Proceedings of IEEE International Conference on Video and Signal Based Surveillance (AVSS’06), Nov. 22-24 2006. Sydney, Australia.
9. Wen Fang, Kap Luk Chan, “Object Detection with Geodesic Active Contours Incorporated with Shape Prior”, Proceedings of Asia-Pacific Workshop on Visual Information Processing (VIP2005), 11-13 Dec 2005, Hong Kong.
10. Wen Fang, Kap Luk Chan, V. Anantharaman, “An Improved Active Contour Method for Heart Wall Boundary Detection in Echocardiographic Image Sequence”, Proceedings of 1st International Bioengineering Conference 2004, Sep. 8-10, 2004 Singapore.

Bibliography

- [1] J. Sethian, *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [2] C. Xu and J. L. Prince, “Gradient vector flow: a new external force for snakes,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 66–71, June 1997.
- [3] B. Scholkopf, A. Smola, and K. R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, pp. 1299–1319, 1998.
- [4] W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of objection motion and behaviors,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 3, pp. 334–351, Aug. 2004.
- [5] B. Ma, Z. Chi, and T. Zhang, “Motion detection and tracking based on level set algorithm,” *International Conference on Control, Automation, Robotics and Vision*, vol. 1, pp. 659–664, Dec. 2004.
- [6] L. A. Vese, “Multiphase object detection and image segmentation,” *Geometric Level Set Methods in Imaging, Vision and Graphics*, pp. 175–194, 2003.
- [7] M. Holtzman-Gazit, R. Kimmel, N. Peled, and D. Goldsher, “Segmentation of thin structures in volumetric medical images,” *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 354–363, Feb. 2006.
- [8] D. P. Mukherjee, N. Ray, and S. T. Acton, “Level set analysis for leukocyte detection and tracking,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 562–572, April 2004.
- [9] A. S. Pednekar and I. A. Kakadiaris, “Image segmentation based on fuzzy connectedness using dynamic weights,” *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1555–1562, June 2006.
- [10] N. Paragios and R. Deriche, “Geodesic active contours and level sets for the detection and tracking of moving objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 266–280, March 2000.
- [11] M. R. Morelande, C. M. Kreucher, and K. Kastella, “A Bayesian approach to multiple target detection and tracking,” *IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] Signal Processing*, vol. 55, no. 5, pp. 1589–1604, May 2007.
- [12] L. Havasi, Z. Szlvik, and T. Szirnyi, “Detection of gait characteristics for scene registration in video surveillance system,” *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 503–510, Feb. 2007.

- [13] J. Wang, A. Vassilis, S. Stan, and B. Margrit, "Detecting objects of variable shape structure with hidden state shape models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* : Accepted for future publication, vol. DOI:10.1109/TPAMI.2007.1178, 2007.
- [14] J. M. B. Dias and J. M. N. Leitaó, "Wall position and thickness estimation from sequences of echocardiographic images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 25–38, Feb. 1996.
- [15] J. Klingler, C. Vaughan, T. Fraker, and L. Andrews, "Segmentation of echocardiographic images using mathematical morphology," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 11, pp. 925–934, Nov. 1988.
- [16] N. Friedland and D. Adam, "Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 344–353, Dec. 1989.
- [17] V. Chalana, D. T. Linker, D. R. Haynor, and Y. Kim, "A multiple active contour model for cardiac boundary detection on echocardiographic sequences," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 290–298, June 1996.
- [18] G. Jacob, J. A. Noble, C. Behrenbruch, A. D. Kelion, and A. P. Banning, "A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 226–238, March 2002.
- [19] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. C. Reiber, "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 1374–1383, Nov. 2002.
- [20] J. S. Suri, K. Liu, S. Singh, S. N. Laxminarayan, X. Zeng, and L. Reden, "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review," *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 8–28, March 2002.
- [21] N. Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *IEEE Transactions on Medical Imaging*, vol. 1, no. 8, pp. 773–776, June 2003.
- [22] D. Comaniciu, X. S. Zhou, and S. Krishnan, "Robust real-time myocardial border tracking for echocardiography: an information fusion approach," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 849–860, July 2004.
- [23] B. Georgescu, X. S. Zhou, D. Comaniciu, and B. Rao, "Real-time multi-model tracking of myocardium in echocardiography using robust information fusion," *Medical Image Computing and Computer Assisted Intervention (MICCAI'04)*, vol. 3217, pp. 777–785, Saint-Malo, France, Sept. 2004.
- [24] M. Rousson and R. Deriche, "A variational framework for active and adaptative segmentation of vector valued images," *Proceedings of IEEE Workshop on Motion and Video Computing*, pp. 56–62, Dec. 2002.
- [25] —, "A variational framework for active and adaptative segmentation of vector valued images," *RR 4515 INRIA*, July 2002.

- [26] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia, "A variational model for image classification and restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 5, pp. 460–472, May 2000.
- [27] H. K. Zhao, T. Chan, B. Merriman, and S. Osher, "A variational level set approach to multiphase motion," *Journal of Computational Physics*, vol. 127, pp. 179–195, 1996.
- [28] N. Paragios, "Geodesic active regions and level set methods: Contributions and applications in artificial vision," *PhD Thesis, Ecole Supérieure en Science Informatique, Université de Nice-Sophia Antipolis/I.N.R.I.A.*, 2000.
- [29] M. Kass, A. Witkin, and D. Terzopolous, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
- [30] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [31] R. Malladi, J. Sethian, and B. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, Feb. 1995.
- [32] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, Feb. 1997.
- [33] N. Paragios and R. Deriche, "Geodesic active regions for supervised texture segmentation," *IEEE International Conference on Computer Vision*, vol. 2, pp. 926–932, Sept. 1999.
- [34] —, "Unifying boundary and region-based information for geodesic active tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 300–305, 1999.
- [35] D. Cremers, C. Schnorr, and J. Weickert, "Diffusion-snakes: combining statistical shape knowledge and image information in a variational framework," *IEEE International Workshop on Variational, Geometric and Level Set Methods*, pp. 137–144, July 2001.
- [36] S. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.
- [37] Y. Shi and W. C. Karl, "Real-time tracking using level sets," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 34–41, June 2005.
- [38] C. Zimmer and J. C. Olivo-Marin, "Coupled parametric active contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1838–1842, Nov. 2005.
- [39] M. Niethammer and A. Tannenbaum, "Dynamic geodesic snakes for visual tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 660–667, June 2004.
- [40] D. Cremers, N. Sochen, and C. Schnorr, "A multiphase dynamic labeling model for variational recognition-driven image segmentation," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 67–81, 2006.
- [41] W. J. Niessen, B. Romeny, and M. A. Viergever, "Geodesic deformable models for medical image analysis," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 634–641, Aug. 1998.

- [42] E. Brandt, L. Wigstrom, and B. Wranne, "Segmentation of echocardiographic image sequences using spatio-temporal information," *Medical Image Computing and Computer Assisted Intervention (MICCAI'99), Lecture Notes on Computer Science, Springer, Cambridge, UK*, vol. 1679, pp. 410–419, 1999.
- [43] D. Kucera and R. W. Martin, "Segmentation of sequences of echocardiographic images using a simplified 3D active contour model with region-based external forces," *Computerized Medical Imaging and Graphics*, vol. 21, no. 1, pp. 1–21, 1997.
- [44] E. D. Angelini, A. F. Laine, S. Takuma, J. W. Holmes, and S. Homma, "LV volume quantification via spatiotemporal analysis of real-time 3-D echocardiography," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 457–469, June 2001.
- [45] M. Choy and J. Jin, "Extracting endocardial borders from sequential echocardiographic images," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, pp. 116–121, Jan.-Feb. 1998.
- [46] M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 316–323, June 2000.
- [47] M. Rousson and N. Paragios, "Shape priors for level set representation," *Proceedings of the European Conference on Computer Vision*, pp. 78–92, May 2002.
- [48] D. Cremers, "A variational framework for image segmentation combining motion estimation and shape regularization," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 53–58, June 2003.
- [49] D. Cremers and C. Schnorr, "Motion competition: variational integration of motion segmentation and shape regularization," *Pattern Recognition*, vol. 2449, pp. 472–480, Sept. 2002.
- [50] X. Bresson, P. Vanderghenst, and J. Thiran, "A priori information in image segmentation: energy functional based on shape statistical model and image information," *IEEE International Conference on Image Processing*, vol. 3, pp. 425–428, Sept. 2003.
- [51] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations of shape gradients?" *SIAM Journal of Applied Mathematics*, vol. 63, no. 6, pp. 2128–2154, 2003.
- [52] T. Brox and J. Weickert, "Level set based image segmentation with multiple regions," In *C. E. Rasmussen, H. Bulthoff, M. Giese, and B. Scholkoph (Eds.): Pattern Recognition. Lecture Notes in Computer Science*, pp. 415–423, Aug. 2004.
- [53] J. Kim, J. W. Fisher, A. Yezzi, M. Cetin, and A. Willsky, "Nonparametric methods for image segmentation using information theory and curve evolution," *International Conference on Image Processing*, vol. 3, pp. 797–800, 2002.
- [54] T. Chan and W. Zhu, "Level set based shape prior segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1164–1170, June 2005.
- [55] G. Charpiat, O. Faugeras, and R. Keriven, "Approximations of shape metrics and application to shape warping and empirical shape statistics," *Journal of Foundations of Computational Mathematics*, vol. 5, no. 1, pp. 1–58, 2005.

- [56] A. Chakraborty, H. Staib, and J. Duncan, "Deformable boundary finding in medical images by integrating gradient and region information," *IEEE Transactions on Medical Imaging*, vol. 15, no. 6, pp. 859–870, 1996.
- [57] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1531–1536, Nov. 2004.
- [58] D. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding*, vol. 53, pp. 211–218, 1991.
- [59] N. Paragios and R. Deriche, "Geodesic active regions for motion estimation and tracking," *IEEE International Conference on Computer Vision*, vol. 1, pp. 688–694, Sept. 1999.
- [60] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *International Journal of Computer Vision*, vol. 72, pp. 195–215, April 2007.
- [61] H. Delingette and J. Montagnat, "New algorithms for controlling active contours shape and topology," *Proceedings of the European Conference on Computer Vision*, vol. 1843, pp. 381–395, 2000.
- [62] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi, "Gradient flows and geometric active contour models," *IEEE International Conference on Computer Vision*, pp. 810–815, 1995.
- [63] J. O. Lachaud and A. Montanvert, "Deformable meshes with automated topology changes for coarse-to-find three-dimensional surface extraction," *Medical Image Analysis*, vol. 3, no. 2, pp. 187–207, 1999.
- [64] M. Rousson, N. Paragios, and R. Deriche, "Implicit active shape models for 3D segmentation in MRI imaging," *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 209–216, 2004.
- [65] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Evolutionary fronts for topology-independent shape modeling and recovery," *Proceedings of European Conference on Computer Vision*, vol. 1, pp. 3–13, 1994.
- [66] —, "A topology independent shape modeling scheme," *SPIE Conference on Geometric Methods in Computer Vision II*, vol. 2031, pp. 246–258, 1994.
- [67] X. Han, C. Xu, and J. L. Prince, "A topology preserving level set method for geometric deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 755–768, June 2003.
- [68] V. Caselles, F. Catte, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Mathematik*, vol. 66, pp. 1–31, Oct. 1993.
- [69] S. Jehan-Besson, M. Barlaud, and G. Aubert, "DREAM2S: deformable regions driven by an Eulerian accurate minimization method for image and video segmentation," *International Journal of Computer Vision*, vol. 53, no. 1, pp. 45–70, 2003.

- [70] M. Rousson, C. Lenglet, and R. Deriche, "Level set and region based surface propagation for diffusion tensor MRI segmentation," *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, May 2004.
- [71] J. Ivins and J. Porrill, "Active region models for segmenting medical images," *IEEE International Conference on Image Processing (ICIP'94)*, vol. 2, pp. 227–231, Nov. 1994.
- [72] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, pp. 577–685, 1989.
- [73] D. Cremers, "Statistical shape knowledge in variational image segmentation," *PhD thesis, Department of Mathematics and Computer Science, University of Mannheim, Germany*, 2002.
- [74] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [75] A. Blake and A. Zisserman, "Visual reconstruction," *MIT Press*, 1987.
- [76] A. Blake and M. Isard, "Active Contours," *Springer*, 1998.
- [77] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.
- [78] L. A. Vese and t. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of computer vision*, vol. 50, no. 3, pp. 271–293, 2002.
- [79] N. Paragios and R. Deriche, "Geodesic active regions: a new paradigm to deal with frame partition problem in computer vision," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1/2, pp. 249–268, 2002.
- [80] J. Lie, M. Lysaker, and X. Tai, "A binary level set model and some applications to Mumford-Shah image segmentation," *IEEE Transactions on Image processing*, vol. 15, no. 5, pp. 1171–1181, May 2006.
- [81] J. Kauffhold and A. Hoogs, "Learning to segment images using region-based perceptual features," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 954–961, 2004.
- [82] S. K. Setarehdan and J. J. Soraghan, "Automatic cardiac LV boundary detection and tracking using hybrid fuzzy temporal and fuzzy multiscale edge detection," *IEEE Transactions on Biomedical Engineering*, vol. 46, pp. 1364–1378, Nov. 1999.
- [83] M. Suhling, M. Arigovindan, C. Jansen, P. Hunziker, and M. Unser, "Myocardial motion analysis from B-mode echocardiograms," *IEEE Transactions on Image Processing*, vol. 14, pp. 525–536, April 2005.
- [84] J. Nascimento, J. S. Marques, and J. Sanches, "Estimation of cardiac phases in echocardiographic images using multiple models," *Proceedings of International Conference on Image Processing*, vol. 2, pp. 149–152, Sept. 2003.

- [85] S. Malassiotis and M. G. Strintzis, "Tracking the left ventricle in echocardiographic images by learning heart dynamics," *IEEE Transactions on Medical Imaging*, vol. 18, no. 3, pp. 282–290, March 1999.
- [86] I. Mikic, S. Krucinski, and J. D. Thomas, "Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 274–284, April 1998.
- [87] C. H. Chu, E. J. Delp, and A. J. Buda, "Detecting left ventricular endocardial and epicardial boundaries by digital two-dimensional echocardiography," *IEEE Transactions on Medical Imaging*, vol. 7, pp. 81–90, June 1988.
- [88] R. N. Czerwinski, D. L. Jones, and W. D. O'Brien, "Detection of lines and boundaries in speckle images - application to medical ultrasound," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 126–136, Feb. 1999.
- [89] G. Storvik, "A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 976–986, Oct. 1994.
- [90] M. Mignotte, J. Meunier, and J. C. Tardif, "Endocardial boundary estimation and tracking in echocardiographic images using deformable templates and Markov random fields," *Pattern Analysis and applications*, vol. 4, pp. 256–271, Nov. 2001.
- [91] I. A. Hunter, J. J. Soraghan, and T. McDonagh, "Fully automatic left ventricular boundary extraction in echocardiographic images," *Proceedings of Computers in Cardiology*, pp. 741–744, Sept. 1995.
- [92] Y. Xin, P. Boekaerts, M. Bister, Y. Taeymans, and J. Cornelis, "Left ventricular boundary detection in echocardiographical images using artificial neural networks," *Proceedings of 6th European Signal Processing Conference*, pp. 1199–1202, 1992.
- [93] Y. N. Sun and C. T. Tasi, "Segmentation of echocardiograms using a neural network," *Microprocessing Microprogram*, vol. 35, pp. 791–798, 1992.
- [94] G. E. Mailloux, F. Langlois, P. Y. Simard, and M. Bertrand, "Restoration of the velocity field of the heart from two-dimensional echocardiograms," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 143–153, June 1989.
- [95] Q. Zhang and R. Pless, "Segmenting cardiopulmonary images using manifold learning with level sets," *IEEE Workshop on Computer Vision for Biomedical Image Applications*, vol. 3765, pp. 479–488, 2005.
- [96] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis," *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, vol. 1, pp. 171–180, June 1996.
- [97] D. Rueckert, P. Burger, S. M. Forbat, R. D. Mahiaddin, and G. Z. Yang, "Automatic tracking of the aorta in cardiovascular MR images using deformable models," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 518–590, Oct. 1997.

- [98] D. Cremers and S. Soatto, "Variational space-time motion segmentation," *IEEE International Conference on Computer Vision*, vol. 2, pp. 886–892, Oct. 2003.
- [99] —, "Motion competition: a variational framework for piecewise parametric motion segmentation," *Internal Journal of Computer Vision*, vol. 62, no. 3, pp. 249–265, May 2005.
- [100] Y. Lin and W. Chen, "An adaptive speed term based on generalized fuzzy operator for level set segmentation," *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, vol. 1, pp. 141–144, April 2004.
- [101] S. Ho, E. Bullitt, and G. Gerig, "Level set evolution with region competition: automatic 3-D segmentation of brain tumors," *International Conference on Pattern Recognition*, vol. 1, pp. 532–535, Aug. 2002.
- [102] Y. Jin, A. Laine, and C. Imielinska, "An adaptive speed term based on homogeneity for level-set segmentation," *Proceedings of SPIE, Medical Imaging 2002: Image Processing*, vol. 4684, pp. 383–390, May 2002.
- [103] X. Xie and M. Mirmehdi, "RAGS: region-aided geometric snake," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 640–652, May 2004.
- [104] J. S. Suri, "Leaking prevention in fast level sets using fuzzy models: An application in MR brain," *IEEE EMBS International Conference on Information Technology applications in Biomedicine*, pp. 220–226, Nov. 2000.
- [105] K. Siddiqi, Y. B. Lauziere, A. Tannenbaum, and S. W. Zucker, "Area and length minimizing flows for shape segmentation," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 433–443, March 1998.
- [106] Y. Pan, J. D. Birdwell, and S. M. Djouadi, "A new gradient and region based geometric snake," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 745–748, March 2005.
- [107] D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, M. Lovelock, and G. T. Y. Chen, "Model-based segmentation of medical imagery by matching distributions," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 281–292, March 2005.
- [108] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137–154, Feb. 2003.
- [109] C. Pluempitiwiriyaew, J. M. F. Moura, Y. L. Wu, and C. Ho, "STACS: new active contour scheme for cardiac MR image segmentation," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 593–603, May 2005.
- [110] W. Fang, K. L. Chan, S. Fu, and S. M. Krishnan, "Incorporating temporal information for ventricular contour detection in echocardiographic image sequences," *Proceedings of IEEE Engineering Medicine Biology Society Conference (EMBC'05)*, pp. 1099–1102, Sept. 2005.
- [111] T. Koga, K. Linuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated inter-frame coding for video conferencing," *Proceedings of National Telecommunications Conference (NTC'81)*, vol. 4, pp. G5.3.1–G5.3.5, Dec. 1981.

- [112] M. A. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.
- [113] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society*, vol. 39, pp. 1–38, 1977.
- [114] http://www.med.yale.edu/intmed/cardio/echo_atlas/index.html.
- [115] <http://www.e-echocardiography.com>.
- [116] M. Pardas and E. Sayrol, "A new approach to tracking with active contours," *Proceedings of International Conference on Image Processing*, vol. 2, pp. 259–262, Sept. 2000.
- [117] Y. T. Lin and Y. L. Chang, "Tracking deformable objects with the active contour model," *IEEE International Conference on Multimedia Computing and Systems*, pp. 608–609, June 1997.
- [118] F. Segonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn, and B. Fischl, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage*, vol. 22, no. 3, pp. 1060–1075, 2004.
- [119] D. Cremers and G. Funka-Lea, "Dynamical statistical shape priors for level set based tracking," *IEEE International Workshop on Variational and Level Set Methods, Springer, LNCS*, vol. 3752, pp. 210–221, 2005.
- [120] M. Gastaud, M. Barlaud, and G. Aubert, "Combining shape prior and statistical features for active contour segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 726–734, May 2004.
- [121] Y. Chen, S. Thiruvenkadam, H. D. Tagare, F. Huang, D. Wilson, and E. A. Geiser, "On the incorporation of shape priors into geometric active contours," *IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pp. 145–152, July 2001.
- [122] Y. Chen, H. D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. A. Geiser, "Using prior shapes in geometric active contours in a variational framework," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 315–328, 2002.
- [123] T. Zhang and D. Freedman, "Tracking objects using density matching and shape priors," *IEEE International Conference on Computer Vision*, vol. 2, pp. 1056–1062, Oct. 2003.
- [124] D. Cremers, N. Sochen, and S. C, "Towards recognition-based variational segmentation using shape priors and dynamic labelling," *International Conference on Scale Space Theories in Computer Vision*, vol. 2695, pp. 388–400, 2003.
- [125] D. Cremers, T. Kohlberger, and C. Schnorr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition*, vol. 36, no. 9, pp. 1929–1943, Sept. 2003.
- [126] Y. Wang and L. Staib, "Boundary finding with correspondence using statistical shape models," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 338–345, June 1998.
- [127] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "Model-based curve evolution technique for image segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 463–468, 2001.

- [128] X. Bresson, P. Vandergheynst, and J. Thiran, "A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional," *International Journal of Computer Vision*, vol. 28, no. 2, pp. 145–162, June 2006.
- [129] M. B. Stegmann and D. D. Gomez, "A brief introduction to statistical shape analysis," p. 15, Mar 2002. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?403>
- [130] J. Xie, H. Tsui, and W. M. Lam, "Shape prior based segmentation for organ deformation correction," *IEEE International Conference on Image Processing*, vol. 4, pp. 2575–2578, Oct. 2004.
- [131] Z. Zhang, K. L. Chan, Y. Wu, and C. Chen, "Learning a multivariate Gaussian mixture model with the reversible jump MCMC algorithm," *Statistics and Computing*, vol. 14, no. 4, pp. 343–355, Oct. 2004.
- [132] N. Acito, G. Corsini, and M. Diani, "An unsupervised algorithm for hyperspectral image segmentation based on the gaussian mixture model," *IEEE International Conference on Geoscience and Remote Sensing Symposium*, vol. 6, pp. 3745–3747, July 2003.
- [133] P. C. de-la Higuera, J. I. Arribas, E. Munoz-Moreno, and C. Alberola-Lopez, "A comparative study on microcalcification detection methods with posterior probability estimation based on gaussian mixture models," *Proceedings of IEEE Engineering in Medicine and Biology Conference (EMBS'05)*, pp. 49–54, Sept. 2005.
- [134] M. Rousson and D. Cremers, "Efficient kernel density estimation of shape and intensity priors for level set segmentation," *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 1, pp. 757–764, 2005.
- [135] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780–788, June 2002.
- [136] T. Cover and J. Thomas, "Elements of Information Theory," *New York: John Wiley and Sons*, 1991.
- [137] <http://www.vision.auc.dk/tbm/Gestures/>.
- [138] http://en.wikipedia.org/wiki/Kernel_principal_component_analysis/.
- [139] <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.
- [140] http://en.wikipedia.org/wiki/Papillary_muscle/.
- [141] <http://www.medterms.com/script/main/art.asp?articlekey=4755/>.