Syed Abdul Mateen

Nº 46862

Sequential Extraction Thresholding Clustering for Segmentation of Coastal Upwelling on Sea Surface Temperature Images

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática

Advisor: Susana Nascimento, Profª. Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

December, 2017

Sequential Extraction Thresholding Clustering for Segmentation of Coastal Upwelling on Sea Surface Temperature Images

Abstract

Coastal upwelling is a process when cold and nutrient-rich water dynamically appears over the surface of the ocean by replacing the warm water. The oceanographers are interested to detect the upwelling regions and corresponding boundaries but to examine the whole process of upwelling they have to work manually on each image, therefore; it increases the workload. The main purpose of this application is to automatically detect the upwelling regions, monitoring environmental changes and the study of fishery resources.

The Seed Expanding Clustering algorithm (SEC) (Nascimento et al., 2015) is a thresholding clustering method for automatic detection of upwelling and delineation of its fronts. The self-tuning thresholding is derived from the clustering criterion and serves as a boundary regularizer of the growing clusters. The SEC algorithm is shown more than 80% of accuracy rate on the unsupervised automatic recognition of the phenomenon.

The main contribution of this dissertation is threefold. First, the development of a sequential extraction version of the SEC algorithm with a stop condition that takes advantage of the knowledge domain to select seeds and model extracted features. Second, the development of an explosion control procedure to detect the so-called leakage problem. Third, the development of a fusion scheme of unsupervised clustering validation measures.

The experimental comparison of the new iterative version of the SEC algorithm with a new developed iterative version of Adams & Bischof SRG on the unsupervised segmentation of upwelling regions on SST images from different regions of the globe show their competitiveness comparing to other conventional SRG methods.

Keywords: Image segmentation; automatic thresholding; seeded region growing; control leakage problem; unsupervised validation; Sea Surface Temperature (SST) images.

Acknowledgement

Firstly, I am grateful to the Almighty God for establishing me to complete this thesis promptly. I would like to express my sincere thanks to Prof. Susana Nascimento her valuable guidance and encouragement extended to me. She provided me the motivation to do this research and provided monitoring throughout this thesis work as well as the accuracy that is required on all steps. I would like to thank the Professors of Centro de Oceanografia and Department de Engenharia Geográfica, Geof\'isica e Energia (DEGGE), Faculdade de Ciências, Universidade de Lisboa for providing the SST images of Portugal examined in this study, as well as Professors Paulo Relvas and Joaquim Luís from Faculty of Science and Tecnology, University of Algarve for providing the collection of SST images of Canary Island.

I take this opportunity to record my sincere thanks to all those who directly or indirectly, have lent their helping hands in this research. I would like to extend my thanks to the department FCT-UNL by giving me the opportunity to develop my skills throughout the course. Lastly, I am thankful to my parents for their unceasing encouragement and support.

Table of Contents

List of Figures vi

List of Tables x

1 Introduction 1

1.1 Motivation 1

1.2 Problem Description 3

1.3 Main Contributions 4

1.4 Organization of the Document 5

2 Literature Review 6

2.1 Introduction to Image Segmentation 6

2.2 Clustering Automatic Thresholding (CAT) Methods for Image Segmentation 6

2.2.1 Ridler and Calvard's method 8

2.2.2 Otsu's method 8

2.2.3 Kittler and Illingworth's method 9

2.3 Seeded Region Growing (SRG) Methods for Image Segmentation 9

2.3.1 Adam's Seeded Region Growing Method 9

2.3.2 Adaptive Seeded Region Growing using Automatic Thresholding 10

2.4 Domains of Application 13

2.5 Strategies for Controlling Explosion in SRG 14

2.6 Clustering Validation Approaches 17

2.6.1 Supervised vs Unsupervised Validation 17

2.6.2 Unsupervised Validation Measures 18

3 Extending the Seed Expanding Clustering and Related Methods 21

3.1 The Seed Expanding Cluster (SEC) Method and its Algorithms 21

3.2 The Iterative Seed Expanding Cluster (ISEC) Algorithm 23

3.3 Improved Iterative Seed Expanding Clustering Algorithm 24

3.3.1 Inner Stop Condition: the Revised SEC 24

3.3.2 Outer Stop Condition: the Revised ISEC 27

3.4 Sequential Iterative version of Adams SRG Algorithm 30

3.5 Proposed Strategy for Explosion Control 32

3.6 Fusion Strategy for Unsupervised Clustering Validation 34

4 Experimental Study 35

4.1 Goals of the Study 35

4.2 Imagery Data and Parameterization 36

4.3 Analysis of Iterative Seed Expanding Clustering 39
4.3.1 Comparing ISEC-V2 vs ISEC 39
4.3.2 Comparing ISEC-V2 vs I-Adams SRG 41
4.3.3 Comparing ISEC-V2 vs Conventional SRG Methods 43
4.4 Fusion Strategy for Unsupervised Clustering Evaluation 46
4.4.1 Unsupervised Fusion Analysis of ISEC-V2 vs I-Adams SRG 46
4.4.2 Unsupervised Fusion Analysis of ISEC-V2 vs Conventional SRG Methods 48
4.5 Analysis of the Explosion Control in ISEC-V2 51
4.6 Summary of the Results 55
5 Conclusion and Future Work 57
Bibliography 58
A The Results 62
A.1 Matlab and R Correlation Analysis 62
A.2 ISEC-V2 Comparative Results with ISEC 63
A.3 SEC vs RSEC Results of F-measure Index 64
A.4 Segmentation Results with SST images (Explosion) 65
A.5 Segmentation Results with SST images (No Explosion) 70
A.6 Segmentation Results with SST images of Canary 74
A.7 Segmentation Results with Contour Strength (CS) and Derivative 77

List of Figures

1.1 Applying SEC on SST image, (1.1a) original SST image, (1.1b) corresponding ground truth map, (1.1c) segmentation result by applying the ST-SEC algorithm 2

1.2 Applying SEC on SST image, (1.2a) original SST image, (1.2b) corresponding ground truth map, (1.2c) segmentation result by applying the ST-SEC algorithm 2

3.1 Resulting images, left one is the segment result from RSEC algorithm and the right one is the result from SEC algorithm 25

3.2 F-measure results using SEC and RSEC algorithms for the images of Canary 26

3.3 Feature firstmean-min value for sequential extracted clusters on ISEC-V2 28

3.4 ISEC-V2 combine results in terms of F-measures and the number of iterations for images of 1998 and 1999 (Portugal). The left image is the line graph of F-meaure analysis and the right image is the bar chart of number of outer iterations 29

3.5 F-measure as a result of Iterative Adams for the images 1998 and 1999 of Portugal. 30

3.6 left: the image of Ground-truth, Right: the resulting image from I-Adams 31

3.7 Bar chart showing outer iterations comparing ISEC and I-Adams for the images 1998 and 1999 of Portugal. 31

3.8 Contour Strength trend for the images of No Explosion 1998-1999 (Portugal) 33

3.9 Contour Strength trend for the images of No Explosion 1998-1999 (Portugal) 33

4.1 Two SST images, the left one is from Portugal (1998-07-11) and the right one is from Canary (img_262) 36

4.2 Three SST images, the one in the left is the image with strong gradient, the middle one is with weak gradient and the right one is the noisy SST image 37

4.3 The ground-Truth of the above image with strong gradient 37

4.4 Coordinates of the seeds and the coastline 39

4.5 Silhouette Index analysis comparing ISEC-V2 with I-Adams using the whole data set of SST images (Portugal) 41

4.6 Calinski, SDbw and Davies Bouldin indices analysis comparing ISEC-V2 with I-Adams using the whole dataset of SST images (Portugal) 42

4.7 SI/CH/DB/SDbw indices comparing ISEC-V2 with I-Adams using the images of Canary 43

4.8 SI/CH indices analysis comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal) 43

4.9 DB/SDbw indices analysis comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal) 44

4.10 SI/CH indices DB/SDbw indices analysis comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal) 44

4.11 DB/SDbw indices comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal) 45

4.12 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with I-Adams using the whole dataset of SST images (Portugal) 46

4.13 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with I-Adams for the images of Canary. 47

4.14 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal). 48

4.15 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal). 49

4.16 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Shih SRG using the images of Canary 49

4.17 SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Verma SRG using the images of Canary 50

4.18 SST images, left image is the original image, middle one is the ground truth and the right image is segmented image with explosion 51

4.19 Contour Strength corresponding to the number of iterations, the left image is the one with explosion and the right is without explosion.. 52

4.20 Image 19990914 Portugal, Contour Strength with no-explosion 52

4.21 First derivative corresponding to the number of iterations of the image 19980612 with explosion 53

4.22 Explosion Analysis for the images of 1998-1999, Iterations with actual algorithm and after the CS apply 54

4.23 Two SST images, the left one is the ground truth and the right one is the segmented image with explosion 55

4.24 First Derivative corresponding to the number of iteration for the image 19980711 (Portugal) 55

4.25 Resulting image when RSEC algorithm stops at iteration 51 and ground truth in the background 56

A.1 Correlation of Matlab and R using Silhouette index with F-measure 62

A.2 Correlation of Matlab and R using Davies Bouldin index with F-measure 62

A.3 Correlation of Matlab and R using Calinski Harabsz index with F-measure 63

A.4 SST image (1998-06-14) in the left and Ground Truth in the right 63

A.5 Segmentation results, left image is the result of ISEC and the right one is the result of ISEC-V2 64

A.6 Iterative result images by ISEC algorithm 64

A.7 Iterative result images by ISEC-V2 algorithm 64

A.8 SST Image 1998-07-15 65

A.9 GT 1998-07-15 65

A.10 ISEC-V2 1998-07-15 65

A.11 I-Adams 1998-07-15 65

A.12 Shih-SRG 1998-07-15 66
A.13 Verma-Otsu 1998-07-15 66
A.14 SST image 1998-07-11. 66
A.15 GT 1998-07-11. 66
A.16 ISEC-V2 1998-07-11. 66
A.17 I-Adams 1998-07-11.. 66
A.18 Shih SRG 1998-07-11 67
A.19 Verma-Otsu 1998-07-11. 67
A.20 SST image 2002-07-31 67
A.21 ISEC-V2 2002-07-31. 67
A.22 I-Adams 2002-07-31. 67
A.23 Shih SRG 2002-07-31 68
A.24 Verma-Otsu 2002-07-31. 68
A.25 SST image 1998-06-12 68
A.26 GT 1998-06-12. 68
A.27 ISEC-V2 1998-06-12. 68
A.28 I-Adams 1998-06-12. 68
A.29 Shih SRG 1998-06-12 69
A.30 Verma-Otsu 1998-06-12. 69
A.31 SST image 1998-06-14. 69
A.32 GT 1998-06-14. 69
A.33 ISEC-V2 1998-06-14. 69
A.34 I-Adams 1998-06-14 69
A.35 Shih SRG 1998-06-14 70
A.36 Verma-Otsu 1998-06-14 70
A.37 SST image 1998-08-01 70
A.38 GT 1998-08-01 70
A.39 ISEC-V2 1998-08-01 70
A.40 I-Adams 1998-08-01 70
A.41 Shih SRG 1998-08-01 71
A.42 Verma-Otsu 1998-08-01 71
A.43 SST image 1998-08-02 71
A.44 GT 1998-08-02 71
A.45 ISEC-V2 1998-08-02 71

A.46 I-Adams 1998-08-02 71
A.47 Shih SRG 1998-08-02 72
A.48 Verma-Otsu 1998-08-02 72
A.49 SST image 1999-09-01 72
A.50 GT 1999-09-01 72
A.51 ISEC-V2 1999-09-01 72
A.52 I-Adams 1999-09-01 72
A.53 Shih SRG 1999-09-01 73
A.54 Verma-Otsu 1999-09-01 73
A.55 SST images 2000-08-08 73
A.56 ISEC-V2 2000-08-08 73
A.57 I-Adams 2000-08-08 73
A.58 Shih SRG 2000-08-08 74
A.59 Verma-Otsu 2000-08-08 74
A.60 SST image 177 74
A.61 GT 177 74
A.62 ISEC-V2 image 177 74
A.63 I-Adams image 177 74
A.64 Shih SRG image 177 75
A.65 Verma-Otsu image 177 75
A.66 SST image 117 75
A.67 GT image 117 75
A.68 ISEC-V2 image 117 75
A.69 I-Adams image 117 75
A.70 Shih SRG image 117 76
A.71 Verma-Otsu image 117 76
A.72 SST image 237 76
A.73 GT image 237 76
A.74 ISEC-V2 image 237 76
A.75 I-Adams image 237 76
A.76 Shih SRG image 237 77
A.77 Verma-Otsu image 237 77
A.78 SST image 1998-06-12 77
A.79 GT 1998-06-12 77

A.80 ISEC-V2 1998-06-12 77
A.81 Contour Strength (CS)'s first derivative, 1998-06-12 78
A.82 GT and Segmentation, 1998-06-12 78
A.83 Segmentation at it=37, 1998-06-12 78
A.84 SST image 1998-06-18 78
A.85 GT 1998-06-18 78
A.86 ISEC-V2 1998-06-18 79
A.87 Contour Strength (CS)'s first derivative, 1998-06-18 79
A.88 GT and Segmentation, 1998-06-18 79
A.89 Segmentation at it=49, 1998-06-18 79

List of Tables

3.1 Percentage of improved F-measure using the images of 1998 and 1999. ……..…… 25

3.2 Images of CANARY island with result of f-measures, using SEC and the new revised version RSEC.……...……...…… 26

4.1 The whole dataset of Portugal and the Canary Island ……....…… 37

4.2 Unsupervised Indices …….....…… 41

4.3 Visual analysis (Explosion and Under-segmentation) for the whole dataset of Portugal and the Canary. …….…….....…… 45

4.4 Fusion Analysis for the whole dataset of Portugal as well as Canary. ….....…… 50

A.1 Images of 1998 and 1999 with result of f-measures, using SEC and the new revised version RSEC ….....……..… 64

CHAPTER 1

INTRODUCTION

1.1 Motivation

Upwelling occurred when cold and nutrient–rich water appears over the surface of the ocean by replacing the warm water or nutrient-depleted water. The nutrient-rich water is fertilized that produces a good fishing ground; therefore, upwelling detection is directly related to the maritime economy or blue economy. Due to the presence of cold water in these regions, upwelling areas can be identified by cold Sea Surface Temperatures (SST) and high concentrations of chlorophyll-a. The higher availability of upwelling regions results in high levels of primary fishery production. Roughly, 25% of the total global marine fishing is coming from five different upwelling areas that are 5% of the total oceanic area.

The oceanographers have been using SST images, as described by Nascimento and Franco (2009), and Nascimento et al. (2012), for the identification of conversion zone between the colder and warmer oceanic waters. They applied high scale resolution on SST images, which used to take a lot of time and effort to process one by one. In order to get the best results, a good visualization of the whole phenomena is necessary. The detection and continuous monitoring of upwelling might be an extensive process, therefore, automatic tools are required, because not only a large quantity of data collected daily but also to predict the trend of upwelling areas in different regions and seasons, thus an objective approach to extract that region is necessary.

In the past, different approaches have been adopted in order to perform automatic upwelling detection from SST images. The artificial neural networks were applied to wind and SST data for the prediction of coastal upwelling (Kriebel et al., 1998). Neural network algorithm was used for detection and segmentation of upwelling regions (Chaudhari et al., 2008). The author used k-means clustering results to determine the presence of upwelling. Marcello et al. (2005), based on coarse-segmentation method proposed automatic detection of upwelling. The semi-automated method used for detection of upwelling areas (Plattner et al., 2006). Automatic detection of frontal activity was applied using edge detection algorithm (Nieto et al., 2012). The upwelling extracted by means of Otsu's automatic thresholding method and Fuzzy C-means (Tamim et al., 2013).

According to Nascimento et al. (2005), the Fuzzy C-means algorithm was used for the SST image segmentation, however the process was not automatic therefore the problem was resolved by adding one phase in the previous algorithm that is frontier detection after the segmentation as described by Nascimento and Franco (2009). Later on, a fully automated fuzzy clustering method was developed to resolve the problem of automatic detection of upwelling areas (Nascimento et al., 2012). The system named FuzzyUPWELL provided an automatic framework for detection of upwelling areas. Although this algorithm was able to solve the problem of automatic detection but it only operates over the temperature data during the

segmentation process, not taking into account the geographical information about the extracted clusters. Moreover, detection of the frontier was entirely separate from the segmentation phase.

Therefore, Nascimento et al. (2015) adapted the Seeded Region Growing (SRG) (Adams and Bischof, 1994) algorithm in order to propose a new algorithm named Seed Expanding Cluster algorithm (SEC). This algorithm not only considered the temperature value of the pixel but also its spatial context in order to combine pixels for segmentation. It grows regions according to the similarity criterion that is, the temperature of the region to the temperature of a seed pixel. The seed pixel is selected in the beginning, which is the pixel with the lowest temperature.

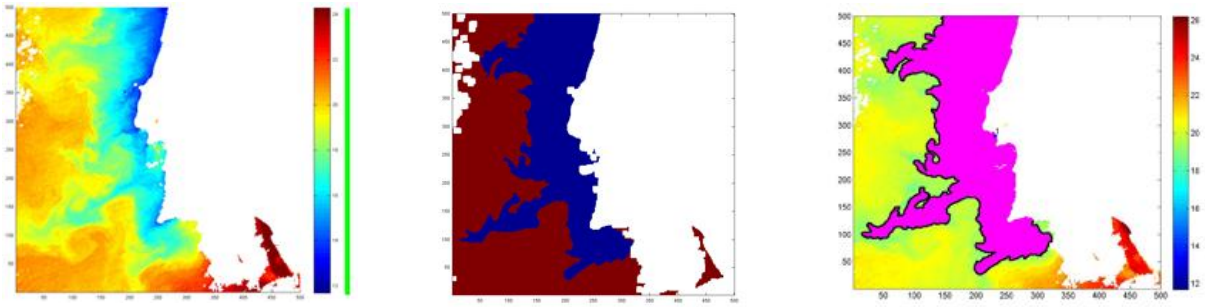


Figure 1.1.a **Figure 1.1.b** **Figure 1.1.c**

Figure 1.1: Applying SEC on SST image, (1.1a) original SST image, (1.1b) corresponding ground truth map, (1.1c) segmentation result by applying the ST-SEC algorithm.

The images in this section have been taken from Nascimento et al. (2015), the Figure (1.1.a) is the original image, Figure (1.1.b) is the ground-truth map and Figure (1.1.c) is the resulting image. The SEC algorithm on these images has shown promising results in order to recognize the upwelling area automatically and the frontline. The result with Self-T SEC algorithm is very satisfactory but only in the case of a strong gradient. In the case of smooth gradients explosion has been observed, which can be seen in Figure (1.2.c).

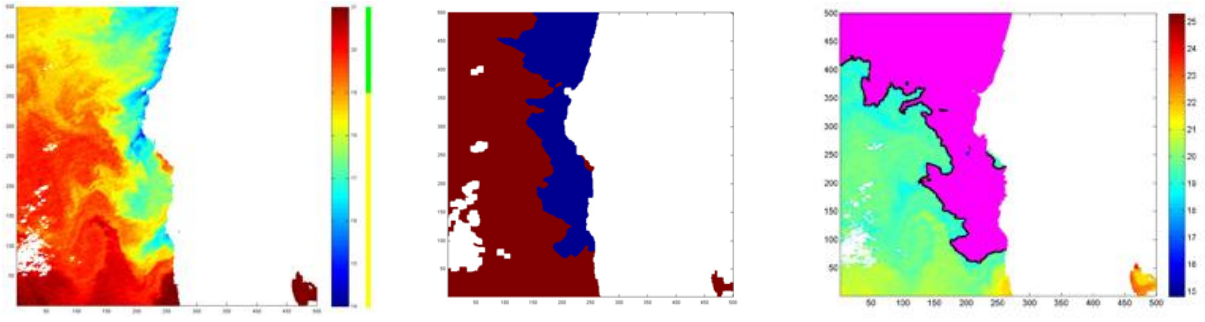


Figure 1.2.a **Figure1.2.b** **Figure 1.2.c**

Figure 1.2: Applying SEC on SST image, (1.2a) original SST image, (1.2b) corresponding ground truth map, (1.2c) segmentation result by applying the ST-SEC algorithm.

It is difficult to handle upwelling regions with transition zones characterized by smooth gradient boundaries because it is hard to make the distinction between the objects of interest and the background. Segmentation leak or explosion is one of the biggest issues in the detection of upwelling areas.

1.2 Problem Description

The algorithms, which solve the problem of automatic detection of upwelling region, are Seed Expanding Cluster (SEC), Self-tuning Seed Expanding Cluster (ST-SEC) (Nascimento et al., 2015) and the Iterative Seed Expanding Cluster (I-SEC) (Lopes, 2015). All the above-mentioned authors adapted the classical Seeded Region Growing (SRG) algorithm but with different criterion of homogeneity. In SRG, region grows when the homogeneity criterion matches i.e. the difference between testing pixel and the seed pixel. It starts with one pixel or the set of multiple pixels if the objective is to segment multiple areas by adding the similar pixels into the upwelling region according to the homogeneity criteria.

The SEC-algorithm starts with initial seeds but it takes only one pixel to start region growing. The main difference between SEC and SRG is the calculation of homogeneity criterion (threshold) that is, the product of the considering pixels instead of the conventional difference. Therefore, the SEC and its family algorithms (ST-SEC and I-SEC) are different from novel Seeded Region Growing (SRG) in terms of a threshold. In SRG, a threshold is defined manually in order to stop the expansion of growing region, which is not appropriate for automatic detection process, therefore, ST-SEC proposed an automatic calculation method for a threshold. In other versions of SEC, threshold is also calculated automatically from the known clustering automatic threshold methods named Ridler and Calvard (1978); Otsu (1979); Kittler and Illingworth (1986).

Lopes (2015), in his master thesis, developed and experimentally test a preliminary iterative version of the SEC algorithm. The SEC algorithm only grows one region, however; I-SEC will extract several regions by an iterative process. The new proposed algorithm improves the iterative version by setting up a good stop condition in order to improve the convergence rate. An experimental study will be executed to test the effectiveness of proposed algorithm by validating the performance of the results with different supervised and unsupervised measures. The results will also be compared with classical SRG method.

1.3 Main Contributions

The main contributions of this dissertation are:

(i) A sequential extraction version of the Seed Expanding Cluster (SEC). This algorithm is composed by two-nested iterative cycles: the inner one responsible for the construction of the 'core' cluster; and the outer one that extracts clusters one-by-one from the residual data. The external stop condition takes advantage of the knowledge domain by defining the seeds selection region and a modeled extracted feature. This new iterative version of SEC Algorithm (I-SECV2) had been experimented on SST images from different regions on the globe having very diverse upwelling patterns;

(ii) The development of an explosion control procedure and its incorporation in the previous algorithm, and the study of the effectiveness of the new version of the algorithm in avoiding the so-called leakage problem.

(iii) The development of a sequential extraction version of the benchmark Seeded Region Growing algorithm (Adams and Bischof, 1994), following the architecture of I-SECV2.

(iv) The development of a 'consensus' scheme of unsupervised clustering validation measures, since the last explored strategy is far from being satisfactory. Not surprisingly, the obtained results using several validation indices are not concordant between each other. Therefore, it was implemented a scheme of fusion voting for unsupervised validation.

(v) To perform an experimental study comparing the I-SECV2 clustering algorithm with I-Adams SRG and other thresholding SRG methods on automatic recognition of upwelling regions from different regions of the globe. A collection of validation indices had been applied to evaluate the segmentation results as well as the developed fusion method for unsupervised validation.

1.4 Organization of the Document

The rest of the document has been organized as Chapter 2 covers the state of the art topics that are related to this work. Then it follows the automatic clustering techniques for image segmentation. After, this chapter describes the basic idea of classical Seeded Region Growing (SRG) by Adams and Bischof (1994) algorithm and adaptive SRG methods. The SRG domain applications were also covered in chapter 2. In chapter 3, the SEC family algorithm and its revised version RSEC followed by I-SEC and the revised version of ISEC called ISEC-V2 were described. The proposed explosion strategy had also described in this Chapter. All the comparative experimental work had done in Chapter 4.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Image Segmentation (IS)

Image segmentation is a process of converting an image into partitions called segments. Each segment contains information about color, motion, texture etc. The segments are homogeneous according to some criterion. Image segmentation always plays a leading role in image processing research and it is the first step for image analysis.

The results depend on the way of applying image segmentation methods that means good analysis is directly related to the IS. Mainly there are two objectives of image segmentation, first is to decompose the image into segments and the second is more important i.e., to arrange the pixels into an efficient and meaningful way for further analysis.

It is not a practical approach to process the whole image directly. Therefore, several image segmentation algorithms had been proposed, in the field of image processing before recognition. Image segments classify an image into clusters or regions according to the same features. Now a day, lots of image segmentations algorithm exists and are applied in different fields of science and our daily life. We can categorize these algorithms according to the methods used, like edge-based, region-based and data-clustering based segmentation.

Automate the image segmentation process makes all stages of image processing more efficient and easy. The proposed study is more focused on the clustering automatic threshold for image segmentation that would be discussed in the next section.

Thresholding is one of the simplest methods in IS where pixels are assigned to a category in which the value lies. Each pixel allocated to some category based on a threshold. Furthermore, we have Region-based segmentation in which region grows from one seed or multiple seeds. All of them expand each region pixel by pixel based on the homogeneity criteria. In data clustering, the concept of growing region is based on the distance between each pixel.

2.2 Clustering Automatic Thresholding (CAT) Methods for Image Segmentation

Thresholding is one of the most popular and widely used methods for image segmentation. The basic idea is to separate foreground from the background by selecting the value for a threshold. This depends on the image features of interest.

The threshold is set before the segmentation starts and it is called manually tuned threshold. Therefore, there always be a need for automatic threshold that is deriving the value without human intervention.

Sezgin and Sankur (2004) grouped threshold into six main categories:

1) Histogram-Based, where peaks and valleys of a smooth histogram analyzed. It is also called shaped based methods of thresholding.
2) Clustering based, where grey level samples grouped into two parts, foreground, and the background.
3) Object attribute-based, it includes the methods that measure the similarity between the gray level and the binary image, using fuzzy shape similarity, edge coincidence etc.
4) Entropy-based, analysis features similarity according to the foreground and background entropy.
5) The spatial-based, methods used higher-order probability distribution and/or correlation between pixels.
6) Local methods, calculate a threshold for each pixel from the local image characteristics.

Now we will have a look at three main methods for automatic thresholding of an image that belongs to above discussed clustering-based methods.

- Ridler and Calvard's method
- Otsu's method
- Kittler and Illingworth's method

Formulation:

Let the pixels of an image represented in terms of L gray levels $[1,2,\dots L]$. The number of pixels at level i is denoted by n_i and the total number of pixels by $N = n_1 + n_2 + \cdots n_L$. The gray-level histogram normalized and the probability distribution is:

$$p_i = \frac{n_i}{N} \ , \ p_i \geq 0 \ , \ \sum_{i=1}^{L} p_i = 1 \ . \tag{2.1}$$

Now assume that the pixels divided into two classes background and foreground, and threshold at level k ; background denotes pixels with levels $[1,\dots k]$ and foreground denotes pixels with levels $[k+1,\dots L]$.

$$P_1 = \sum_{i=1}^{k} (p_i) \ , \tag{2.2}$$

$$P_2 = \sum_{i=k+1}^{L} (1 - P_1). \tag{2.3}$$

$$\mu_1 = \sum_{i=1}^{k} \big(i.P_1(i)\big), \tag{2.4}$$

$$\mu_2 = \sum_{i=k+1}^{L} \big(i.P_2(i)\big). \tag{2.5}$$

$$\sigma_1^2 = \sum_{i=1}^{k}(i - \mu_1(k))^2 \,.\, P_1(i), \quad (2.6)$$

$$\sigma_2^2 = \sum_{i=k+1}^{L}(i - \mu_2(k))^2 \,.\, P_2(i). \quad (2.7)$$

2.2.1 Ridler and Calvard's method

Ridler and Calvard (1978) proposed an iterative process for an image thresholding method. In this method, an optimal value of threshold has chosen automatically because of an iterative process. Iterations provide a cleaner extraction of the object region. They introduce a signal controlling function in which if the function received zero value, the image will signal the background and vice versa.

The switching function (threshold) is defined as the average of the foreground and background class means.

$$T_{n+1} = [\mu_1(T_n) + \mu_2(T_n)]/2 \quad (2.8)$$

The process starts by reading the image pixel by pixel. It stops when the switching function remains constant for further iterations.

Selection of an optimal threshold is very difficult to achieve therefore this method proposed maximum four iterations.

2.2.2 Otsu's method

According to Otsu (1979), automatic selection of threshold is based on unsupervised and nonparametric method. It assumes that the image contains two classes of pixels following bi-modal histogram (foreground and background), then it calculates the optimum threshold mathematically by separating classes in a way that their combined intra-class variance is minimal, or equal to their inter-class variance is maximal.

The total variance is the sum of the within-class variances and the between-class variances.

$$\sigma^2_{WithinClass} = P_1(t)\sigma^2{}_1(t) + P_2(t)\sigma^2{}_2(t) \quad (2.9)$$

where " P_i " is the probablity of backgroud class and foreground class respectively.

Since the total variance is constant, the effect is merely the same. It means when intra class variance is minimal or inter class variance is maximal it has the same effect.

2.2.3 Kittler and Illingworth's method

The method proposed by Kittler and Illingworth (1986), starts by calculating the bi-model histogram of the grey level image h(g) then it estimates the prior probability P of the calculated histogram. In the end, the mean is equal to the total probability.

It is the initial Threshold of the given image and separates the image into classes foreground and background.

Kittler proposed the criterion function:

$$T = 1 + 2[P_1(T)\log(\sigma_1(T)) + P_2(T)log(\sigma_2(T))] - 2[P_1(T)log\ P_1(T) + P_2(T)log\ P_2(T)] \quad (2.10)$$

and the desired threshold is based on the minimization of criterion function that is:

$$Threshold = arg_{\,1<t>n}\ \min[T] \quad (2.11)$$

2.3 Seeded Region Growing (SRG) Methods for Image Segmentation

2.3.1 Adam's Seeded Region Growing Method

Adam and Bishof (1994) proposed a region-growing algorithm called SRG, which is widely used in a different kind of application nowadays. The SRG is based on the conventional region growing assumption, that is, region grows based on similarity of pixels. It is a simple and robust method for growing region systematically. The good results can be achieved in the first step but it majorly depends upon the selection of seeds, therefore, high-level knowledge of image is the root of selection seeds.

It starts with placing the initial seeds in the image, where each seed could be a single pixel or set of pixels. Regions grow with these pixels by adding neighboring pixels to them who qualify the criterion. The SRG stops when all the pixels have allocated some region (only one).

Mainly SRG based on two factors:

1) The seed selection, and
2) The similarity criterion

The homogeneity or similarity criterion is defined as the difference between the testing pixel and the pixel of interesting region R.

$$\delta_o(x, R) = |g(x) - \frac{1}{n\ \sum_{i=1}^{n}\ g(r_i)}| \quad (2.12)$$

where $g(x)$ is the gray value of the pixel x.

Main issues with SRG are:

1) To start the algorithm, how to select a good initial seed?
2) What is the threshold for the region to grow?
3) How to manage the labeling of the pixels?

2.3.2 Adaptive Seeded Region Growing using Automatic Thresholding

Seeded Region Growing (SRG) is a fast and robust method for image segmentation and has the ability to adapt other techniques to make the process more efficient. In this section, we will discuss the adaptive SRG using homogeneity criterion (threshold) more automatic.

2.3.2.1 Linear SRG and Quadratic SRG

Linear and Quadratic SRG algorithm by Fan and Lee (2015), relaxed the grey level assumption of the original SRG. Since the SRG (original) does not impose any restriction on the growing regions, therefore it would produce very rough segmentation boundaries. They also introduced a stabilized SRG that encourages smoother boundaries and prevents the so-called leakage or explosion problem.

In SRG, similarity criterion $\delta_o(x, R)$ is assumes that the grey value of any region does not change and can be a single constant value. Linear SRG relaxes this assumption by modeling the grey values with linear plane.

Let the numbers of rows and columns of an image be n_r and n_c, and the coordinates of a pixel located at that point are $(\frac{i}{n_r}, \frac{j}{n_c})$ respectively. The new Linear SRG method is modeled as $a_1\left(\frac{i}{n_r}\right) + a_2\left(\frac{j}{n_c}\right) + a_3 + \varepsilon$ where $a_1, a_2\ and\ a_3$ are coefficients of the corresponding plane, and ε is error term, usually assumed to be identically and independently distributed. The new homogeneity criterion based on linearity is:

$$\delta_L(x, R) = |g(x) - \{a_1\hat{}\left(\frac{i_x}{n_r}\right) + a_2\hat{}\left(\frac{j_x}{n_c}\right) + a_3\hat{}\} \tag{2.13}$$

Quadratic SRG is similar to the Linear SRG but the only difference is that it is modeled in quadratic planes. The homogeneity criterion in this contrast is:

$$\delta_Q(x, R) = |g(x) - \{b_1\hat{}\left(\frac{i_x}{n_r}\right)^2 + b_2\hat{}\left(\frac{i_x j_x}{n_r n_c}\right)^1 + b_3\hat{}\left(\frac{j_x}{n_c}\right)^2 + b_4\hat{}\left(\frac{i_x}{n_r}\right)^1 + b_5\hat{}\left(\frac{j_x}{n_c}\right)^1 + b_6\hat{}\} \tag{2.14}$$

Both Linear and Quadratic SRG adapted the original SRG.

2.3.2.2 Seed Expanding Cluster (SEC)

The single-seeded region-growing algorithm by Nascimento et al. (2015) was inspired by SRG (classical model) but with different homogeneity criterion in the format of a product. The growth is controlled by a similarity threshold and it stops when no more pixels remain in the frontier boundary. We will discuss SEC algorithm and its versions in detail in section 3.

2.3.2.3 The Verma Seeded Region Growing

Verma et al. (2011) proposed a single-seeded region growing method similar to the SEC algorithm for color image segmentation. The algorithm starts from the central pixel of the image as seed. It grows only one region at a time and in order to tune the automatic threshold value, it uses Otsu's adaptive thresholding technique.

The similarity criterion is based on the intensity of the pixels, that is:

$$\sqrt{(g(i,j) - g(i_s, j_s))^2} < \pi \tag{2.15}$$

where $g(i,j)$ is the intensity value of the testing pixel $g(i,j) \in \mathrm{N}(i,j)$, and $g(i_s, j_s)$ is the intensity value of the seed. The π value is the threshold that derived from Otsu's method of automatic thresholding to grow cluster.

2.3.2.4 The Shih and Cheng Seeded Region Growing

Shih and Cheng (2005) proposed an automatic seed region-growing algorithm derived from the classical SRG (Adams and Bischof, 1994). It starts with more than two seeds to grow the regions. It resolves the problem of seed selection in classical SRG by introducing an automatic selection of initial seeds. For automatic seed selection, three criteria must be satisfied.

1) Seed pixel must have high similarity to its neighbors.
2) At least on seed is selected from expected region.
3) Seeds from different regions must be disconnected.

The homogeneity criterion in order to combines regions is defined by:

$$\delta\left((i,j), C_k\right) = \sqrt{(g(i,j) - mean\{\, g(i_k, j_k)\})^2} \tag{2.16}$$

where $g(i,j)$ is the intensity value of the testing pixel and $g(i_k, j_k)$ is the intensity value of the growing Cluster C_k.

2.3.2.5 The Zanaty and Asaad Seeded Region Growing

This method called a probabilistic Region Growing because it was based on the probability of the pixels. The algorithm presented by Zanaty and Asaad (2013) depends on the different homogeneity criterion. It starts the growth with the seed pixel and stops when pixel does not match the homogeneity criterion. The pixels that pass the criteria move from the frontier F to the cluster C.

$$\delta\,(F, C) = |\, mean\,\{g(i_c\,, j_c)\} - g(i, j)\,| \tag{2.17}$$

where $(i, j) \in$ F and $g(i_c\,, j_c)$ is the mean intensity of the pixels in the cluster C. The pixel assigned to the cluster C if it passes the following similarity criterion:

$$|\, I_z \,-\, mean\,\{g(i_c\,, j_c)\}\,| \leq T\,(I_z\,, \Pr(I_z\,)) \tag{2.18}$$

where I_z is the intensity of the testing pixel and $\Pr(I_z\,)$ is the probability of that intensity value. It calculated the threshold dynamically by:

$$T\,(I_z\,, \Pr(I_z\,)) = \, T_1(I_z\,).T_2\,(\Pr(I_z\,)) \tag{2.19}$$

where T1 and T2 are two threshold values that are mathematically derived.

2.4 Domains of Application

In past, great scientific work had done where Seeded Region Growing used to solve the complex problems. Due to simplicity and robustness of SRG algorithm, now days, it has been used in different domains with conjunction of other algorithms.

2.4.1 SRG in Industrial Application

SRG used in many industrial applications at different domains. Lachance et al. (2004) presented a region growing technique to measure the wear flat area through grinding machine. The process controls automatically the position of the wheel and captures digital images of the wheel between grinding cycles. Pottmann et al. (2005) used SRG for the problems of geometric optimization in Geometry applications. Zhengtao (2011) proposed Capsule Image Segmentation Based on Linear Region Growing by analyzing the traditional image segmentation method SRG. Hadwiger et al. (2008) presented a novel method for interactive exploration of industrial CT volumes such as cast metal parts, with the goal of interactively detecting, classifying, and quantifying features using a visualization-driven approach.

2.4.2 SRG in Medical Image Processing

In the field of Medical Image Processing, Seeded Region Growing algorithm is used for the detection of tumor and also used in brain MRI. Stokking et al. (2000) applied SRG to brain MRI images to visualize and quantify the segments. The method is called morphology-based brain segmentation. As the brain, tissues are very connected to each other so other algorithms are also used with SRG in order to find good results. Pohle and Toennies (2001) presented a new self-learning, fully automatic region-growing segmentation of medical images.

Mat-Isa et al. (2005) applied SRG on digital images and called the method a Seeded Region Growing Feature Extraction. This method used to extract the size of the nucleus, size of cytoplasm, grey level of nucleus and grey level of cytoplasm. Wong and Zrimec (2006) presented a novel technique, which uses a seeded region-growing algorithm to guide the classifier to regions with potential honeycombing. The classification used for analyzing the patterns of lung diseases. Chen et al. (2006) proposed a sketch-based interface for seeded region growing volume segmentation. A user freely sketches regions of interest (ROI) directly over the 3D volume. Parts of the volume outside the ROIs are then automatically cut out in real-time.

Wang and Chen (2012) established Automatic Vector Seeded Region Growing for Parenchyma Classification in Brain MRI. Nuclear magnetic resonance (NMR) can be used to measure the nuclear spin density, the interactions of the nuclei with their surrounding molecular environment and those between close nuclei, respectively. Al-Faris et al. (2013) used a system with automated features for MRI breast tumor segmentation.

2.4.3 SRG in Remote Sensing

There are multiple Remote sensing applications where SRG was used as Bins et al. (1996), presented a segmentation method based on a region growing approach. The technique is

applied to segment images, which are being used to assess land use changes in the Amazon region. Bagli et al. (2004) presented Automatic delineation of shoreline and lake boundaries from Land sat satellite images. Gao et al. (2011) established different segmentation methods in multispectral Landsat images to achieve object based image classification, and used SRG as one of the method. Wang and Chen (2012) proposed an hybrid algorithm that contains different steps including clustering k-means, segment initialization, seed generation, region growing, and region merging. The algorithm used widely in remote sensing data, and also used in urban and regional planning. Stroppiana et al. (2012) introduced a method for extracting burned areas from landsat images using some techniques, which include a region-growing algorithm. Zhang et al. (2013) extracted coastline in aquaculture zones by using region growing segmentation with multiple steps. Mishra and Susaki (2013) proposed some methodologies based on the analysis of multi-temporal Synthetic Aperture Radar images.

2.5 Strategies for Controlling Explosion in SRG

In the process of region growing, the original classical method (SRG) does not impose shape restriction on the contour (boundary) of the region. When there is a weak gradient between the target and the neighbor region, the results could have a very large size of the region and also could have very rough boundaries. In addition, the so-called Leakage or the explosion problem could occur. This leakage problem refers to the situation when the grey values of targeted and the neighbor objects are very similar, and the growing region of one object breaks the true boundary and enters to the other object's region.

We are going to explore different strategies in this section, which controls the explosion in adaptive SRG methods.

2.5.1 Stabilized Seeded Region Growing

Fan et al. (2014) proposed a variant of SRG as discussed in section (2.3.2.1), that encourages smoother boundaries and the aim is to prevent the explosion problem. During the growth process, Stabilized-SRG not only considers the grey value of x, but it also takes into account the grey values of neighboring pixels. This set of neighboring pixels are denoted by the square of size (2L+1) * (2L+1) centered as x.

The neighboring pixels define as:

$$Y_L = \{y_{jk} | -L \leq j,k \leq L\} \quad \text{where } y_{00} = x \qquad (2.20)$$

Note that when $L = 0$, Stabilized-SRG $\delta_{Stablisied}(x, R, Y_L)$ reduce to the Original SRG $\delta_{Original}(x, R)$. Moreover, Stabilized-SRG can be paired with Linear (eq.2.13) or Quadratic SRG (eq.2.14) by modifying in a similar fashion.

The parameter L determines the smoothness of the boundary. Larger the value of L smoother will be the boundary. In practice, it can be chosen by a user (operator) in an attractive manner, moreover prior knowledge about the image is necessary.

2.5.2 New Region Growing Based on Selection of Optimal Threshold and Seeds

Afifi and Ghoniemy (2015) proposed an algorithm that works with local search process to achieve the optimal threshold and better seed selection. The output seeds are the input of local search algorithm to extract the best seeds around initial seeds. Seed selection and get optimum threshold overcomes the limitations of classical SRG as described in section (2.3.1). The algorithm works automatically that means it works without any predefined parameters.

In both histogram-based and region-based segmentation techniques, if the threshold is not correct or not optimum, the contour of the object will destroy and causes an explosion. The algorithm hybridized the seed selection, local search, and thresholding algorithms with the region growing technique in order to get good segmentation results.

It iteratively merges similar pixels into regions in 3 main steps:

1- Choice of the seed pixels;
2- Local search according to a similarity rule;
3- Thresholding algorithm for growing the regions by including adjacent pixels that satisfy the similarity rule.

The Proposed method resolved the issue of optimum threshold of SRG by using the homogeneity test $I(P_x) - RA \leq T$ where RA is the seed pixel. The seed pixel has maximum amplitude from grey level histogram.

The author defines the thresholding algorithm as follows:

The image is divided into two parts using initial threshold T old. The average grey level values for each part (mean1, mean2) is computed then updates threshold value by:

$T_{new} = (mean1 + mean2)/2$ and stop when the condition $|T_{new} - T_{old}|$ <delta satisfied.

Where delta = $0.1 * (h_{max} + h_{min})/2$.

2.5.3 Leak Detection using Distance Transformation in SRG

A region growing method based on the concept that leaks are caused by a narrow bottleneck connection to the seed area (Heimann et al., 2004). This method needs a user intervention; therefore, it allows the user to specify a single point somewhere within the erroneous area and traces a path back to the seed point along the contour of the segmentation region in order to detect the bottleneck.

This method used region grower tool that is, a user defines seed then the segmentation starts and regions grow while neighboring pixels lie within a specific grey value range. This works very well when the segmented image contrast is good but when the contrast is not sufficient at the contour the algorithm produces leaks. This is because of similar grey values of the targeted region and the neighborhood. Often the origin of the leak is only a narrow connection between the boundaries of targeted and neighbor pixels.

This method used Distance Transformation technique in order to identify the leak regions and then remove that additional area from the segmentation. The basic idea was to calculate the path from a random point within the erroneous area to the seed point, which supposed to maintain the maximum possible distance. The bottleneck is the point whose local maximum has the shortest distance hence the origin of the leak. For every pixel, repeat the same process then search the two nearest, opposing points on the contour and separate the segmented area along a line between these two points. The part of the area where the user clicked is then removed from the segmentation.

2.5.4 Automatic Detection by gradient magnitude likelihood classification and Correction of Segmentation Leaks

Kronman et al. (2011) proposed a method that identifies the segmentation leak basis boundary by gradient magnitude likelihood classification. The leak basis boundary then fits the surface and leaks has been removed from the targeted structure by finding the common boundary.

Segmentation leaks are one of the most invasive segmentation errors that can be found in any algorithm. According to the author, leaks produce in the segmentation when the pixels gradient intensity magnitudes of the target and neighboring region boundaries are too small or the characteristics are very similar. After the leaks are found the most important task is to remove them from the results, therefore extensive manual user interaction is required. Many prior shape knowledge-based models have been proposed in order to reduce segmentation leaks. These models have drawbacks as they are structure specific, relies on the experts (manual intervention) and are time-consuming because of the prior generation of the shape.

Heimann et al. (2004) introduced a method, as described in section 2.5.3 that explicitly detects a segmentation leak by computing a path by shortest distance procedure between two user-defined points but this method automatically detects and corrects leaks. The method first finds out the leak basis boundary then fits a surface to pixels of this boundary. The leak is then separated from the target structure by re-labeling the leak basis pixels as background. Since the leak is the actual boundary between targeted and the neighbor structure, so the goal is to find the segmentation leak basis.

The method first computes leak front (LF), derive from it the leak basis boundary (LBB) and then obtains the leaks basis from it.

The leak L between the target structure T and a neighbor structure N is a set of pixels in I that are classified by ST.

$$L = ST \cap N \tag{2.21}$$

where S is the set of structures of interest.

If $L = \emptyset$ there is no leak.

The leak detection is consists of 4 steps.

1- $gradient\ magnitude\ liklehood\ histogram = GL(|\nabla I|)$
2- $LF = N \cap boundary(ST)$
3- $LB = L \cap boundary(T)$
4- $LBB = ST \cap LB$

There are some major advantages of this method.

- It is independent of the segmentation method used;
- It does not require any prior shape, location and/or intensity information;
- It is fully automatic;

After the identification of LBB the correction could be achieved by eliminating the leaks. For each segmentation leak L, it re-labels the leak basis pixels as background and finds the target-connected component.

2.6 Clustering Validation Approaches

2.6.1 Supervised vs. Unsupervised Evaluation

Image segmentation is the first important step in many multimedia applications. In this area, many different approaches and algorithms were proposed, but no one guaranteed to get the best results. To address this problem evaluation criterion was used to quantify the quality of the results since last few years. Supervised evaluation is the one in which user assistants is involved that means it needs some prior knowledge (ground-truth) required by experts to compare the results. Whereas in unsupervised evaluation no user assistant is required whereas some statistics are computed from the segmentation result.

Evaluation methods that require user assistance, are infeasible in many computer vision applications, so unsupervised methods are necessary.

Segmentation results are evaluated with ground-truth made by the experts in the area of supervised evaluation. The Zhang et al. (2007) categorize segmentation evaluation into analytical, empirical goodness and the empirical discrepancy. The supervised evaluation also knows empirical discrepancy that takes into account the difference between the segmented image and the reference ground-truth image.

The very well known measure that was used in supervised evaluation is F-measure (Rijsbergen, 1979). It combines precision and recalls then calculate the values from confusion matrix and cross-validate the results with the ground-truth. Precision is the proportion of predicted positive cases that are correctly real positives, while recall is the proportion of real positive cases that are correctly predicted positive. F-measure gives the value range from 0 to 1, the highest value shows better results we get. The second most important measure is Adjusted Rand Index (ARI), presented by Hubert and Arabie (1985) that was based on the similarity between two data clusters. ARI can score from negative values to 1 and the highest value means good results. In this study, we will use the F-measure as supervised evaluation for the experimental results.

2.6.2 Unsupervised Validation Measures

In external validation, one can use external information not present in the data but when we do not have this information then internal validation is used. Internal validation relies on the information inside data. Data characteristics like noise, monotonicity, density etc. are the basic information that is used for the internal validation indices.

Internal validation takes into account the compactness and the separation of the clusters. How closely the objects are in the cluster is compactness and it is based on variance. Lower variance means high compactness of the clusters. The separation is about how clusters are well separated to each other.

Internal validation process starts by applying the clustering algorithm to the data set. Each clustering algorithm then uses different combinations of parameters to get different clustering results. Now compute internal validation index of each partition or cluster. The best partition has the optimum cluster number.

One of the most important challenges in data clustering now days is how to evaluate the results without auxiliary information. Esendira et al. (2011) address this problem by comparing different internal validation indices. As the internal indices depend upon the intrinsic information present in the data so, the results can be different for different type of data.

According to Liu (2010), S-Dbw is one of the best indices between other unsupervised indices but the data he used was very simple in nature. Chouikhi (2015) compared 30 different internal indices and found that CH and DB perform the best.

Below we have some of the most popular validation indices those are included in the proposed study.

2.6.2.1 Calinski-Harabasz index

The Calinski-Harabasz index, proposed by Calinski and Harabasz (1974), is also called the variance ratio validity index. It takes into account the between cluster variance and within cluster variance.

$$S_{CH} = \frac{trace(S_B)}{trace(S_W)} . n_p - 1 / n_p - k \tag{2.22}$$

where S_B is the between-cluster scatter matrix, S_W is the within-cluster scatter matrix, n_p is the number of cluster points and k is the number of clusters. Maximal value of S_{CH} indicates that the results are good.

The S_{CH} index evaluates the results based on the average between- and within-cluster sum of squares.

2.6.2.2 Dunn index

The Dunn validity index (Dunn, 1974), sometime called distance ratio index because it takes into Account the min and max distances between two points.

$$S_{Dunn} = \frac{d_{min}}{d_{max}} \tag{2.23}$$

where d_{min} is the minimum distance between two points belonging to different clusters, and d_{max} is the maximum distance between any two points selected from the same cluster. The maximal value of S_{Dunn} will indicates better candidate.

The S_{Dunn} index uses the minimum pair wise distance between objects in different clusters as the inter-cluster separation and the maximum diameter among all clusters as the intra-cluster compactness.

$$Index = (a. Separation)/(b. Compactness)$$

where a and b are the weights.

2.6.2.3 Davies-Bouldin index

This internal validation index, proposed by Davies and Bouldin (1979), used to evaluate the clustering results. It is based on the similarities of the obtained clusters.

$$S_{DB} = \frac{1}{k} \sum_{i=1}^{k} max_{j=1,\ldots\ldots k; i \neq k} \ (d_i + d_j)/d(c_i, c_j) \tag{2.24}$$

where k denotes the number of clusters. If i, j are cluster labels, then d_i and d_j are average distance of all patterns in clusters i, j to their respective cluster centroids, and $d(c_i, c_j)$ is the distance between these centroids.

The S_{DB} index is calculated for each cluster C as, compute the similarities between C and all other clusters, and the highest value is assigned to C as its cluster similarity. Then S_{DB} index can be obtained by averaging all the cluster similarities. The smaller value of the index shows good results.

2.6.2.4 Silhouette index

Rousseeuw (1987), proposed the Silhouette index in order to validates the clustering results based on the pair wise difference of between and within-cluster distances as:

$$S = \frac{1}{k}\sum i \ \{ \ \frac{1}{n_i}\sum x \in C_i \ . b(x) - a(x) / \max [b(x), a(x)] \tag{2.25}$$

where k is the number of clusters. n_i is the number of objects in ith cluster. Moreover, the optimal value of the index shows best results.

2.6.2.5 S_Dbw Validity Index

Halkidi et al. (2001) proposed a new clustering validity index based on density.

$$S_{cat}\ (NC) + D_{ens\_bw}\ (NC) \tag{2.26}$$

where S_{cat} is the inter-cluster separation and D_{ens_bw} is the intra-cluster density.

This validity index is the combination of inter-cluster separation and intra-cluster compactness, which obtained from the density between and the variance of the cluster objects respectively. The minimum value of this index is indicating the optimum number of cluster.

CHAPTER 3

EXTENDING THE SEED EXPANDING CLUSTERING AND RELATED METHODS

3.1 The Seed Expanding Cluster (SEC) Method and its Algorithms

The Seed Expanding Clustering algorithm (SEC) (Nascimento et al., 2015) is a new algorithm extending the Seeded Region Growing (SRG) by defining a homogeneity criterion inspired on the concept of approximate clustering (Mirkin, 1996). The approach differs in that the algorithm thresholding values are not expert-driven but rather derived from the approximate clustering model.

The algorithm starts from a pixel with the lowest temperature value in the SST map and uses it as the initial seed. Then, it grows a region by labeling the boundary pixels and expanding simultaneously. This method resolves the problem of boundary pixels labeling and the dependency of pixel sorting order, therefore the SEC algorithm performs these in parallel to speed up the procedure.

The algorithm can be summarized as follows: it receives temperature map T(R, L) as an input where R is the set of rows and L the set of columns and elements of R × L are pixels.

Step 1: In the first step *pre-processing stage,* the data is normalized by taking each pixel in the image and subtract it from the average temperature.

Step 2: The second step of the algorithm is the *Cluster Initialization,* each pixel with the exploring window W centered at the seed pixel $o(i_o, j_o)$ go into the cluster if homogeneity criterion satisfies.

$$c * t(i,j) \geq \pi \tag{3.1}$$

where c is the temperature of the seed pixel, $t(i,j)$ is the temperature of tested pixel and π is the temperature similarity threshold.

Step 3: Set *Cluster Boundary* is the third step in which a set F is define as:

$$F = \{(i',j') \ni C \mid N(i',j') \cap C \neq \emptyset\} \tag{3.2}$$

where $N(i',j')$ is the set of 8-neighborhood pixels.

Step 4: In the fourth step that is, *Cluster Expansion,* cluster C starts to grow and join pixels in iterative way until the stop condition satisfy i.e., the boundary set F as describe in equation (3.2), becomes empty.

In SEC, the homogeneity criterion comprises by two separate conditions, the temperature similarity and the density condition. The temperature similarity condition, equation (3.3), makes SEC different from the conventional SRG and ensures that the expansion of the cluster is smooth with the temperature variation, and the density condition equation (3.4), covers a continuous fragment of the ocean.

$$c^* \times t(i', j') \geq \pi \tag{3.3}$$

where the temperature c^* define as $c^* = mean(T(W(i', j') \cap C))$ and $t(i', j')$ representing the current boundary pixel temperature.

The density condition is defined as the total number of pixels those are in the cluster and intersect the exploring window $W(i', j')$, divided by the number of total pixels in the window$W(i', j')$:

$$\frac{|W(i', j') \cap C|}{W(i', j')} \geq \alpha \tag{3.4}$$

where α is the density threshold and if both of the conditions satisfy then that pixel is enter into the cluster.

The major innovation of the algorithm is its similarity criterion (3.3.) that takes the form of a product rather than a difference as in SRG algorithms. The Self-tuning version of the algorithm (ST-SEC) dynamically calculates the threshold values, directly derived from the clustering criterion (Nascimento et al., 2015), (Nascimento and Mirkin, 2017). Its value changes depending on the state of the cluster C and its interception with the window $W(i', j')$.

The ST-SEC is similar in structure to the other version of the algorithm, except the calculation of threshold value π that define as:

$$mean(T(W(i', j') \cap C)))^2/2 \tag{3.5}$$

There were developed distinct versions of the SEC algorithm according to the adopted method to calculate the threshold values. Specifically: SEC-Otsu (Nascimento et al., 2015), the method used in this version is the one derived from Otsu (1979). In SEC- Kittler, and SEC-Ridler (Lopes, 2015), the threshold is calculated from Kittler and Illingworth (1986), and Ridler and Calvard (1978) respectively.

A preliminary proposal of a sequential iterative version of the algorithm (ISEC) was developed by Lopes (2015), where clusters are extracted one by one, to retrieve discontinuous upwelling regions. However, the stop condition of that version faced some problems that demand a revised version.

3.2 The Iterative Seed Expanding Cluster (ISEC) Algorithm

The upwelling area sometime appears in different coastal regions. The SEC algorithm only grows one region because it takes only one seed; however, for tackling the problem of extracting the multiple upwelling areas, it is necessary to run the region growing procedure more than once. An iterative version of the SEC algorithm (I-SEC) was developed and experimentally studied by Lopes (2015) in his master thesis, had to treat the problem of discontinuous upwelling areas.

The proposed iterative SEC (ISEC-V2) modifies the stop condition of the previous ISEC in order to enhance the efficacy and efficiency of the algorithm. The ISEC-V2 stop condition comprises of two sub-conditions.

C1: Coastline distance calculation is the first stop condition, in which the selected seed distance is calculated from the coastline if the seed is far from the coastline the algorithm stops. How the coastline and the distance are calculated will be described in section (4.3.1). This removes completely the stop condition in the previous algorithm where the iterations were fixed to five.

C2: Feature extraction (first mean minus minimum), the mean of first cluster is recorded and check the next cluster if the difference of first mean and its minimum temperature value is greater than epsilon value that was fixed to 1.068 in ISEC, the algorithm stops. The ISEC-V2 modifies this stop condition by introducing the epsilon value calculation from the well known Otsu (1979) method.

3.3 Improved Iterative Seed Expanding Clustering Algorithm

The ISEC algorithm comprises inner and outer stop conditions. The inner stop condition is related to the SEC algorithm as described in section (3.1), the algorithm for the core cluster formation. The outer stop condition ISEC stop condition itself and responsible for the formation of more than one upwelling areas.

3.3.1 Inner Stop Condition: the Revised SEC

The revised SEC algorithm is the same as described in section (3.1) in structure but we made few changes in order to increase the effectiveness of the segmentation results. In addition, we had compared the results with SEC method and found the results with improvements. The original SEC algorithm has been revised in the following aspects:

i) The neighborhood to explore the cluster boundary, F (equation (3.2)) was set to a window of 4-neighborhood instead of 8-neighborhood;

ii) It was adopted a pixel-to-pixel update of the cluster during the dilatation of the boundary F;

iii) The inner stop condition is defined by the stability of the cluster, substituting the condition of the empty boundary.

This revised version, RSEC, allows a better convergence rate of the algorithm as well as slightly improves the quality of the segmentation results those were examined by experiments as shown in tables (3.1) (Portugal) and (3.2) (Canary). We also observed that with the 4-neighbors the number of iterations increased which results in a good segmentation. In figure (3.1), the left image is the result of new RSEC and the right one is the resulting image of SEC. It observed that with 8-neighbors the segmentation region took more pixels from the neighbor those were not in interest, the pixels with the circle marked were those, which removed in the RSEC.

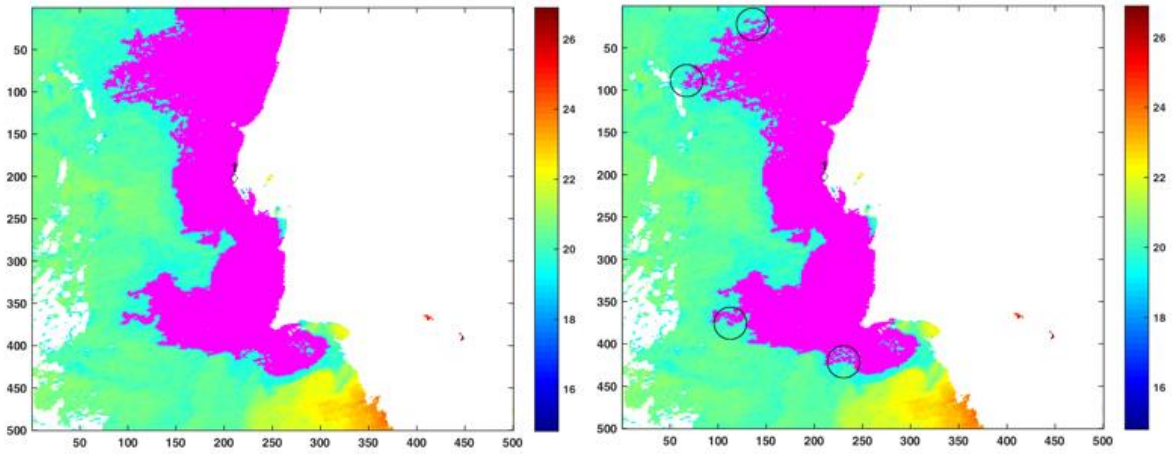


Figure 3.1: Resulting images, left one is the segment result from RSEC algorithm and the right one is the result from SEC algorithm.

We took 30 SST images of 1998 and 31 of 1999 from Portugal and also 10 images from Canary island. We applied SEC and RSEC for this data set and observed that with our new revised version (RSEC) the F-measure improved to some extent.

Table 3.1: Percentage of improved F-measure using the images of 1998 and 1999.

	R-SEC	SEC
R-SEC		85%
SEC	15%	

Table (3.1) shows 51 out of 61 images from Portugal (85%) and table (3.2) shows 8 out of 10 images from Canary (80%), the f-measure increases. The complete results with f-measure can be seen in appendix A.3. The revision in the SEC algorithm showed slightly improved results can be seen in figure (3.1) where the circle marked in black are those pixels those should not be a part of a cluster hence RSEC removed those pixels from the final segmentation.

Table 3.2: Images of CANARY island with result of f-measures, using SEC and the new revised version RSEC.

		ISEC	RSEC
S#	***image***	***f-measure***	***f-measure***
1	img_58.mat	**0.7776**	**0.7781**
2	img_117.mat	0.8063	0.8034
3	img_152.mat	**0.7754**	**0.7828**
4	img_177.mat	**0.8013**	**0.8084**
5	img_214.mat	**0.8451**	**0.8485**
6	img_237.mat	**0.8268**	**0.8307**
7	img_262.mat	0.8136	0.8054
8	img_310.mat	**0.8061**	**0.8077**
9	img_334.mat	**0.7237**	**0.7265**
10	img_336.mat	**0.7231**	**0.7267**

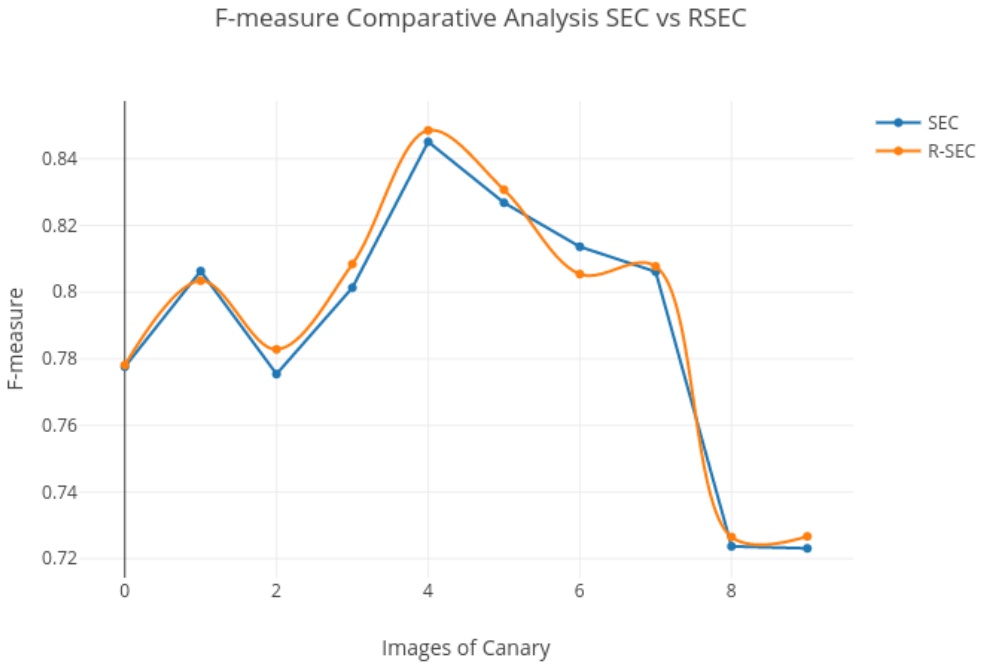


Figure 3.2: F-measure results using SEC and RSEC algorithms for the images of Canary.

The images of Canary were very different from the Portugal images but we had the same results when applied the RSEC on these images as shown in figure (3.2).

3.3.2 Outer Stop Condition: the Revised ISEC

The ISEC algorithm resolved the problem of discontinuity but had the problem of complexity of the algorithm due to the several stop conditions as described in section (3.2). The algorithm contains inner and outer stop conditions. The inner stop condition is related to the revised SEC algorithm as described in section (3.3.1), the algorithm for the core cluster formation. The outer stop condition is related to ISEC and is responsible for the formation of more than one upwelling areas. In this section, we will discuss the revision of outer stop condition in ISEC to form a new iterative version named Iterative Seed Expanding Cluster (ISEC-V2) and will compare the results with the previous ISEC version.

The ISEC-V2 took advantage of the domain knowledge that seeds exist near the coast because the water is coolest in that area. The revised version ISEC-V2 calculates the coastline and records the coordinates and the spatial values of each coastline pixel.

The coastline once calculated at the start of the algorithm and took these coordinates in the iterative process for the distance calculation between the seed and the coastline pixel at that position. The coastline formation will be discussed in detail in section (4.3.1); the formation of coastline is the base of ISEC-V2 because this revised version main stop condition is related to the coastline. If the seed exists near the coastline, that seed would be the mature seed and take into consideration but if not then the algorithm will stop. The distance is calculated with respect to the coastline coordinates.

The revised ISEC-V2 is eliminated unnecessary stop conditions and make it more simple and efficient. If the seed occurs near the coastline than it passes to the second criterion i.e., *first mean minus min*, this feature is, the difference between the mean temperature of the first cluster retrieved and the minimum temperature of the current cluster. This stop criterion is similar to the previous ISEC but the basic difference is the calculation of the threshold (epsilon). In ISEC, this epsilon value was fixed to 1.068 but in this revised version ISEC-V2, the threshold calculated from the automatic clustering threshold method by Otsu (1979). It was experimentally found that the value of the feature *(firstmean-min)* decreased with the increase in the number of clusters.

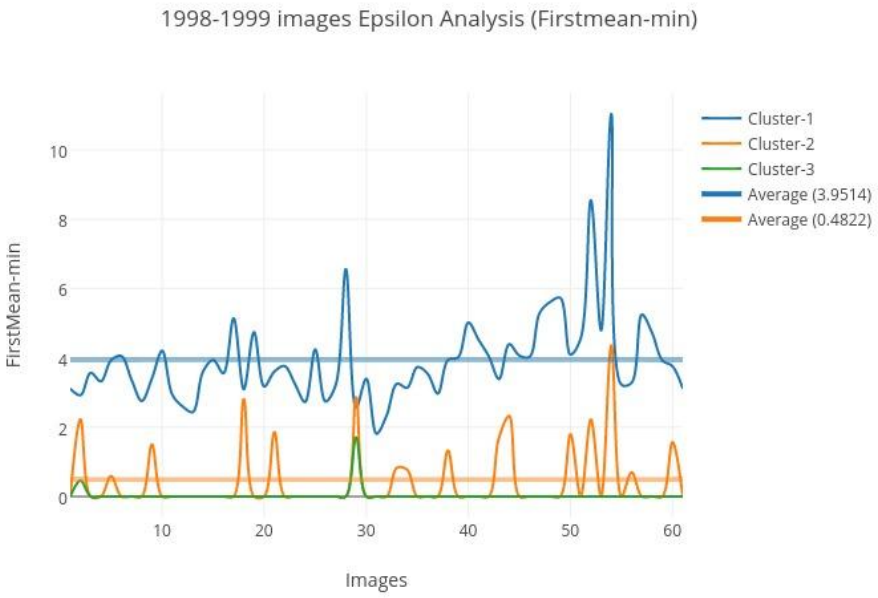


Figure 3.3: Feature firstmean – min value for sequential extracted clusters on ISEC-V2 algorithm.

We tested the approach with two subsets of SST images for the upwelling seasons of 1998 and 1999 for this calculation as shown in figure (3.3), and found that for the first cluster, the average value was 3.9 and it decreased to 0.48 an average in the second cluster. Hence, we used Otsu, to calculate the good threshold value, in our experimental study instead of the fixed threshold.

The ISEC-V2 showed promising results in terms of segmentation and the efficiency. We did experiments with ISEC-V2 for the images of 1998 and 1999 and found that the f-measure remains stable and the execution time to process each image decreased because of a number of iterations of the outer loop.

In figure (3.4), we can see that the number of iterations is not more than 2 but in the previous version, we have 5 iterations fixed. This revision not only got good segmentation results but also to enhance the efficiency of the algorithm in terms of executed time. Most of the images had the stable results but few images improved the f-measure. For example image 24 in figure (3.4), the left one, f-measure improved from 0.69 to 0.71, image 33 improved from 0.38 to 0.40, image 37 improved from 0.42 to 0.44, image 52 improved from 0.78 to 0.80 and image 59 f-measure improved from 0.61 to 0.78.

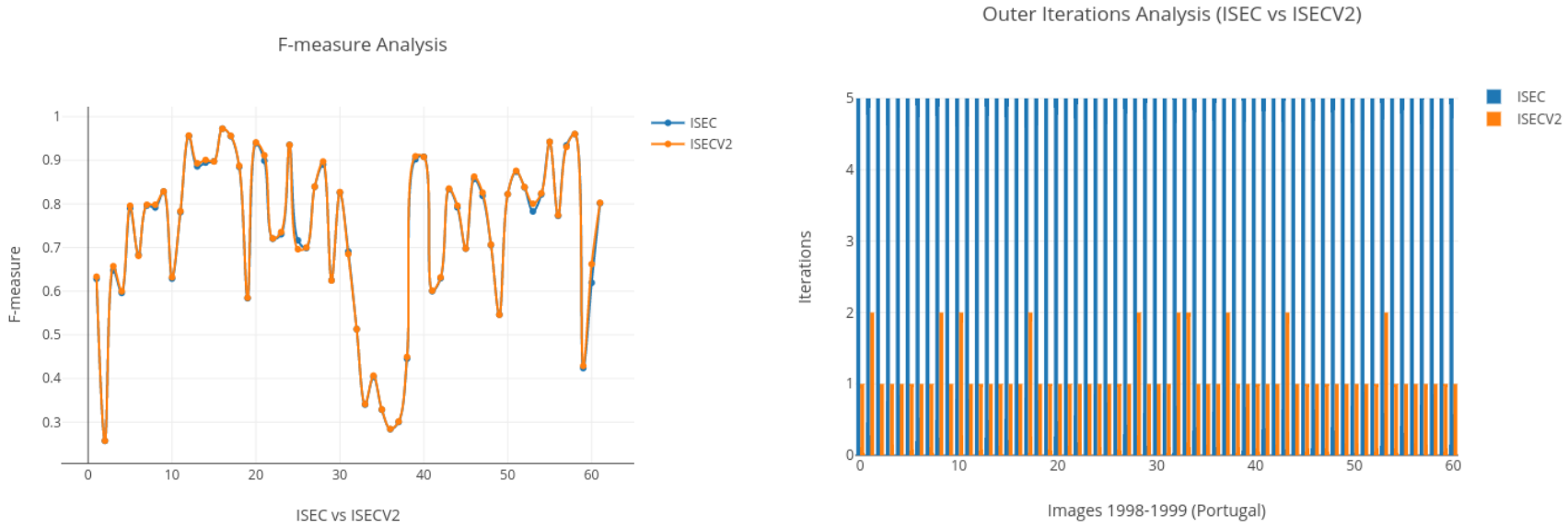


Figure 3.4: ISEC-V2 combine results in term of F-measures and the number of iterations for images of 1998 and 1999 (Portugal). The left image is the line graph of F-measure analysis and the right image is the bar chart of number of outer iterations

Hence, from this comparative analysis, we observed that with ISEC-V2 we achieved efficiency and robustness that was not in the previous version. Moreover, the results were comparatively good in terms of f-measure. We applied the new revised version ISEC-V2 on all the data set of Portugal as well as the images of Canary Island and the results were promising.

3.4 Sequential Iterative Version of Adams SRG (I-Adams)

Adams and Bischof (1994) proposed an SRG method that based on the similarity of the pixels in order to grow the region as described in section (2.3.1). The Adams SRG is simple and continues being an extensively used SRG algorithm. The similarity criterion of Adams is; the difference of the testing pixel and the pixel of an interested region as mentioned in equation (2.12).

The first main drawback or the problem of Adams was the selection of the initial seed or seeds because it based on the single seed as well as multiple seeds. The second problem of Adams was to identify the number of regions that means it does not found that how many clusters are in the data. Hence, to overcome these problems we developed an iterative version of Adam's SRG, homologous to ISEC-V2: sequentially extracting clusters one by one from the residual SST map until the stop condition holds. This version is called, I-Adam SRG.

The seed selection in ISEC-V2 is the pixel near the coastline as described in section (3.3.2) that removed the problem of seed selection in Adams. Moreover, the iterative version resolved the second problem i.e., how many regions or clusters are in the data.

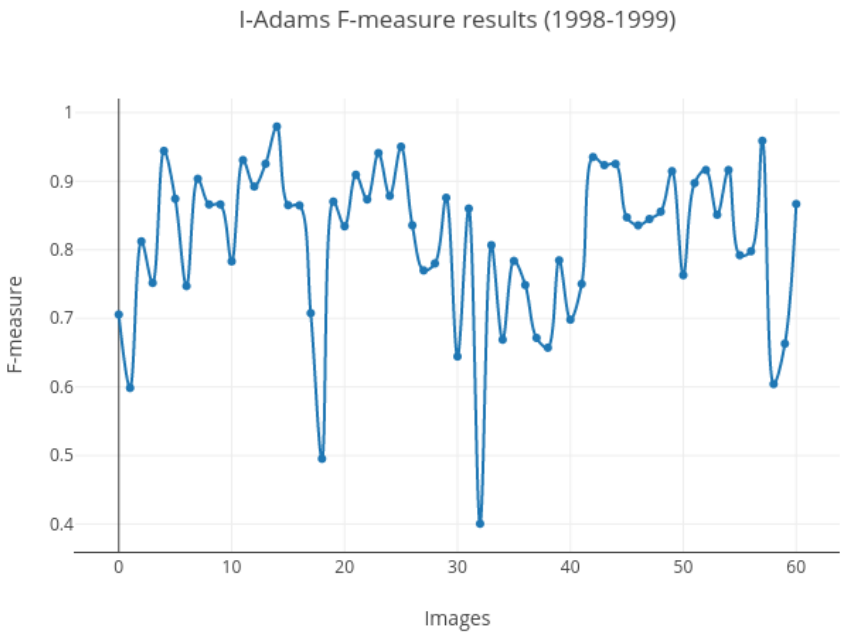


Figure 3.5: F-measure as a result of Iterative I-Adams for the images 1998 and 1999 of Portugal.

In figure (3.5), we can see that the results of f-measure are very good as higher the f-measure value show goodness of the results. The results of iterative Adams now become competitive therefore we used it for most of the comparative study.

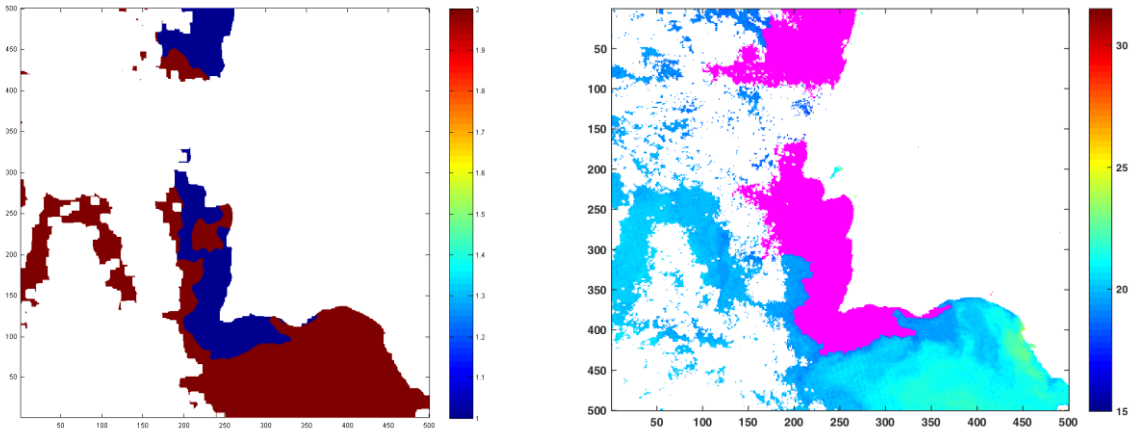


Figure 3.6: left: the image of Ground-truth, Right: the resulting image from I-Adams

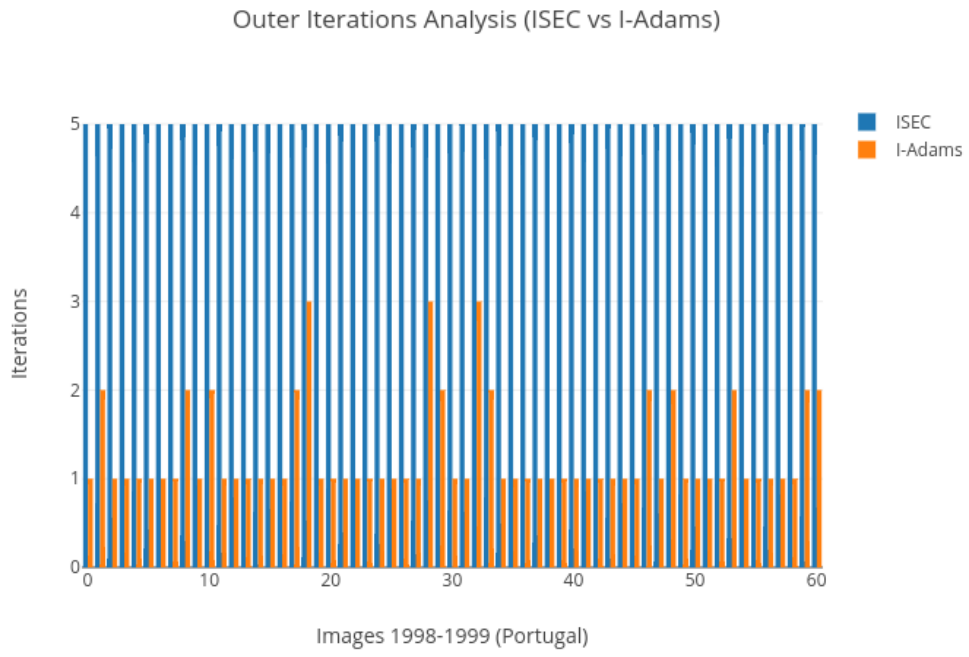


Figure 3.7: Bar chart showing outer iterations comparing ISEC and I-Adams for the images 1998 and 1999 of Portugal.

The figure (3.7) shows the efficiency of new I-Adams in terms of the total number of outer iterations. The previous iterative model was fixed the iteration to five whereas in new Adams it is not fixed that increases the efficiency of the algorithm.

3.5 Proposed Strategy for Explosion Control

It is well known that region-based segmentation algorithm suffers from the problem of an explosion, also denoted as leaking problem. In intensity-based thresholding and region-growing methods, they appear when the pixels intensity of the leak boundaries and inside the target structure is very close. In adaptive region growing methods, like SEC they appear when the intensity distribution of the leak and the target structure are similar.

In this section, we are going to introduce a strategy for explosion control and will implement in conjunction with SEC algorithm to cope up the problem of leakage (explosion).

3.5.1 The Contour Strength (CS) Criterion

We choose the contour strength (CS) measure, adapted from the literature (Siebert, 1997), as the measure to detect explosion in the expanding process of the clusters. The reason for the selection of this strategy in my proposed work is that it introduces a very good measure for the quality of growing region by maximizes the contour strength

The contour strength of a region R is the sum of the absolute differences between each pixel p_i on the contour and the neighbor of contour points q_i that do not belong to the region of interest, i.e.

$$cs(R) = \frac{1}{n} \sum_{pi \in C_R} | p_i - q_i | \tag{3.6}$$

where C_R is the set of pixels on the contour of R.

The general idea is that regions are bounded by strong contour, therefore regions grow such as to maximize $cs(R)$. We assumed the strategy that the cluster expands by maximizing its contour strength.

We calculate the CS as mentioned in equation (3.6), by taking the difference of the pixel on contour (boundary (F)) and the neighbors those do not belong to the cluster. We tested CS on all those images of 1998 and 1999, which had an explosion and all those, where explosion did not appear. The detail experimental results will be discussed in section (4.5), where we will also examine the CS in conjunction with RSEC. The study has been done systematically because getting good results were not possible in one-step. We first calculate the first derivative of the CS of cluster R and analyze the trend of its values for the SEC segmentations facing explosion as opposed to the ones without explosion. We observed that few images not shown good results as expected i.e. the images with no explosion shows downward slope (weak CS). To cope up the problem we used the concept of Moving Average will be discussed in detail in section (4.5) so that we can get the smooth boundary.

The main goal was to find a trend between images with explosion and with no explosion. That trend we can see in figures (3.8) and (3.9).

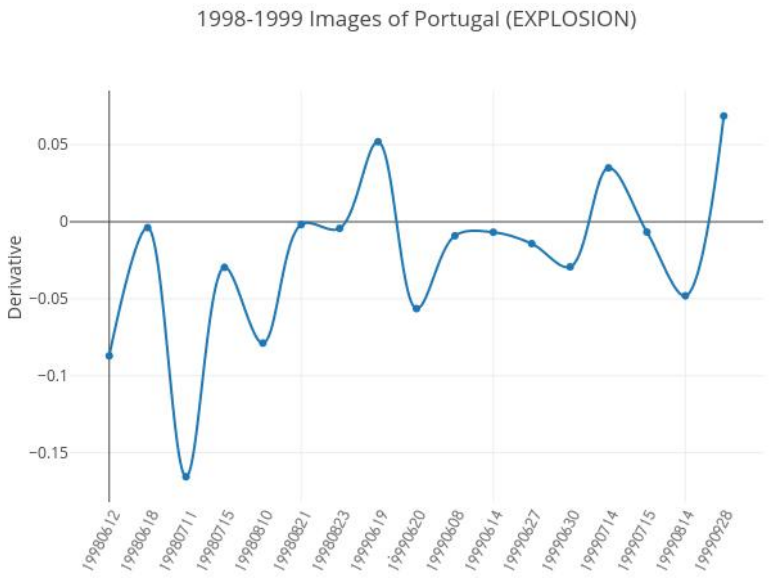


Figure 3.8: Contour Strength trend for the images with Explosion 1998-1999 (Portugal).

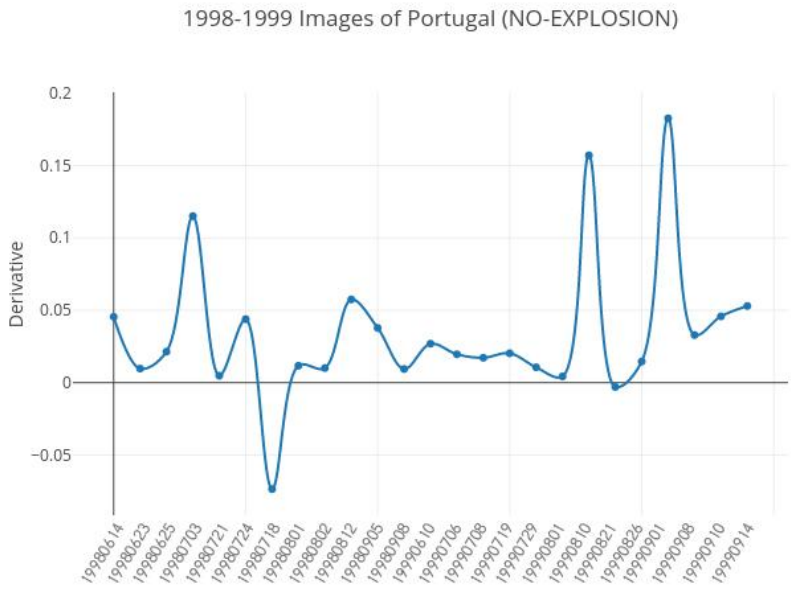


Figure 3.9: Contour Strength trend for the images with No Explosion 1998-1999 (Portugal).

We observed from figures (3.8) and (3.9) that the images of explosion we had the negative first derivative of CS and for the images of no explosion we had positive derivate accordingly. The image like 19990619 in figure (3.8) is the one that has an explosion but positive derivate is because of outliers. The negative value of the derivate of contour strength indicates the starting of the explosion. We took the images of 1998 and 1999 with explosion and found the negative CS trend similarly for the images with no explosion we had the positive trend, figures (3.8) and (3.9) respectively.

3.6 Fusion Strategy for Unsupervised Clustering Validation

As discussed in Section (2.6.3), it is well known that internal validation indices are not concordant among them, therefore; we decide to introduce a fusion procedure in which all the unsupervised indices as mentioned in section (2.6.3) are used separately. In this section, we will discuss those fusion indices and the results we got from them.

3.6.1 Fusion method for Clustering Validity Indices (CVI)

We explored a fusion approach for clustering validation adapted from Kryszczuk and Hurley (2010). The method called fusion because it calculates the index based on the multiple unsupervised indices. In the proposed study, we will use this method by using the normalized data of Silhouette index, Davies Bouldin index, Calinski-Harabasz index and S_Dbw index as described in section (2.6.2).

Silhouette index normally ranges from 0 to 1. The Davies Bouldin is considering being good when it has a lower index and it normally ranges from 0 to 2. The S_Dbw is similar to the Davies Bouldin that means lower index value considering being good. The Calinski Harabasz is different from all three indices because its value ranges from 0 to any whole number. Its goodness depends upon the higher value of the index. As the values range is diversified for the different indices, so for the fusion we normalized the data ((value – min)/max) to keep its range from 0 to 1. The goodness of fusion indices depends on the higher score value. This method comprises four different fusion measures named as SF-A, SF-G, SF-H, and SF-Med.

$$Sf - A = \frac{1}{i} \left(\sum_{m=1}^{i} Sm\right) \tag{3.7}$$

where i is the number of indices used and S represents the index. As mentioned above we used four unsupervised indices with their normalized values therefore, this index takes the sum of these values (SI + DB + CH + SDbw) and divided by 4.

$$Sf - G = \left(\prod_{m=1}^{i} Sm\right)^{1/i} \tag{3.8}$$

where i is the number of indices used and S represents the index. This fusion index uses product (SI x DB x CH x SDbw) instead of the sum as in above index and take the root with value 4, as 4 is the total number of indices we used. The higher value of the index shows goodness of the results.

$$Sf - H = i / \sum_{m=1}^{i} \frac{1}{Sm} \tag{3.9}$$

where i is the number of indices and S represents the index. It takes reciprocal of each unsupervised index value (1/SI + 1/DB + 1/CH + 1/SDbw) and sum up them. In the last, this term is divided with total number of unsupervised indices i.e. 4.

$$Sf - Med = median\ (Sm) \tag{3.10}$$

It is the median of all the unsupervised indices those are used in fusion method. Where i is the number of indices and S represents the index.

CHAPTER 4

EXPERIMENTAL STUDY

4.1 Goals of the Study

The main goals of the experimental study are:

(i) To analyze the effectiveness of the new Iterative SEC algorithm, ISEC-V2, comparing it against the previous one, as well as with the proposed Iterative Adam's SRG algorithm I-Adams SRG;

(ii) To experimentally fine-tune the parameters of the strategy to control the explosion problem on SRG, incorporate it in the ISEC-V2, and experimentally analyze the effectiveness of the algorithm in preventing the leakage problem;

(iii) To develop a fusion strategy of internal clustering validation indices to perform the unsupervised evaluation of the segmentation results;

(iv) To perform an extensive experimental study comparing ISEC-V2 with the new developed I-Adams SRG as well as other conventional SRG algorithms.

4.2 Imagery data and Parameterization

The experimental study is performed on high-resolution SST images. For Portugal, each image contains 500 by 500 pixels map, each pixel contains 1km by 1km, and for the Canary Island images contain 350 by 570 pixels map as presented in figure (4.1). The reason why Canary images have fewer pixels is that in that region, the upwelling phenomena is a very thin line as compared to the Portugal images so most of the area in Canary images is not under consideration.

In the map, each pixel has a value in temperature Celsius degree and the value NaN if there is a cloud in the sky (transmission error) or if it is the land. NaN shows that we have no data at that point (pixel).

SST images are characterized by strong and weak gradients, as presented in figure (4.2), due to the occurrence of upwelling in different regions diversely. If the boundary between the cold and warm water is very clear then the images are with strong gradient and if the boundary is smooth or very hard to differentiate between the boundary and the targeted area then the images are called weak gradient images. The explosion problem as described in section (3.4) is more related to these weak gradient images. Some images are noisy because of the transmission error or due to the cloud in the sky.

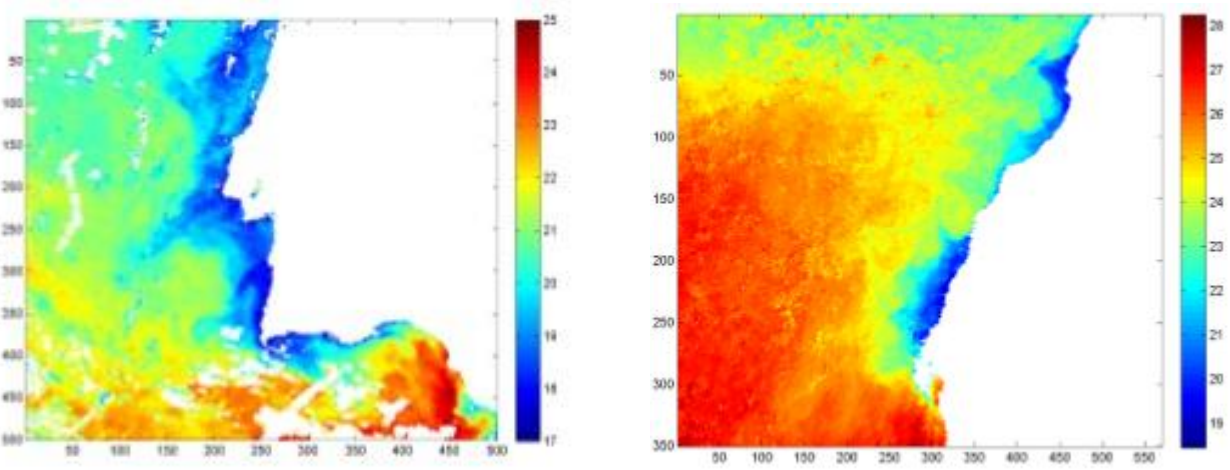


Figure 4.1: Two SST images, the left one is from Portugal (1998-09-08) and the right one is from Canary (img_262).

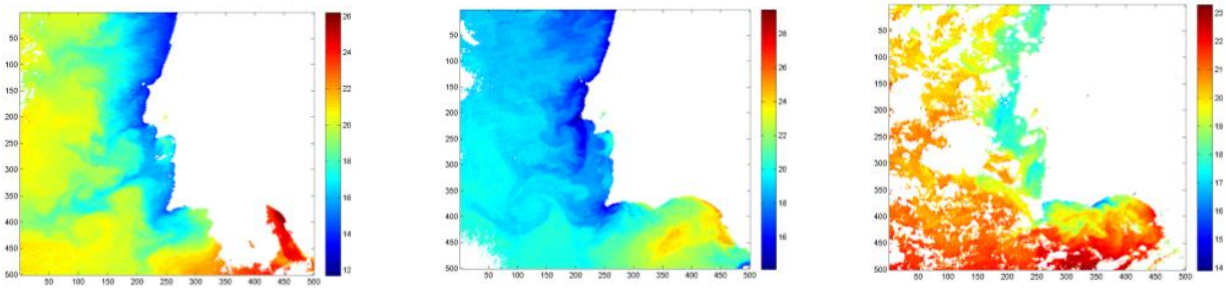


Figure 4.2: Three SST images, the one in the left is the image with strong gradient, the middle one is with weak gradient and the right one is the noisy SST image.

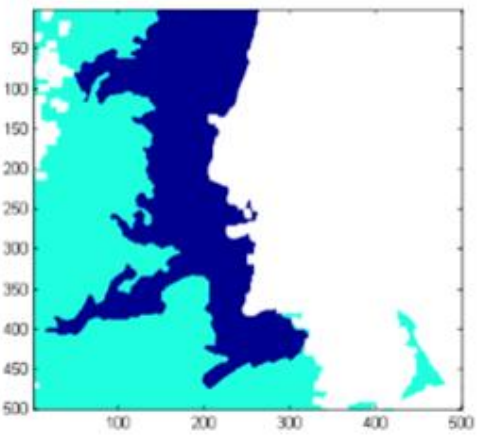


Figure 4.3: The ground-Truth of the above image with strong gradient.

As shown in table (4.1), we deal with the images of Portugal for the years 1998 (30 + 52 = 82 images), 1999 (31 images), 2000 (32 images), 2001 (30 images) and 2002 (22 images). The sample of 10 images from Canary was also the part of our experiments because those images were very different from the images of Portugal.

The images of the year 1998 and 1999 are those, which have ground truth, the image made by an expert in order to match the result. Figure (4.3) is the ground truth image that will use to compare the result.

	Year	Total Images	Ground Truth
PORTUGAL	1998	30	Yes
	1998	52	No
	1999	31	Yes
	2000	32	No
	2001	30	No
	2002	22	No
Canary	NA	10	Yes

Table 4.1: The whole dataset of Portugal and the Canary Island.

4.2.1 Settings of the Experiments

This study is more focused on the comparative analysis of the new version of ISEC (ISEC-V2) with its previous version and other classical SRG methods. The baseline SEC algorithm has different versions in terms of using popular threshold methods (Ridler, Otsu, Kittler) namely (SEC-Ridler, SEC-Otsu, SEC-Kittler). In our experiments, we used another version of SEC called SEC-Self tuning, as described in section (3.1), equation (3.5); the algorithm calculates the threshold dynamically in each iteration.

The data set of 207 images were used for experiments in which most of the images were taken from different areas of Portugal in different years, as well as few images from Canary Island, as shown in the table (4.1), in order to test the adaptability of the newer version. The images from Canary were very different from Portugal geographically so a separate analysis required for those images. The images from 1998 and 1999 have ground truth; this is why we apply supervised evaluation (F-measure) for this dataset.

The proposed study gives more emphasis to unsupervised analysis because nowadays we have a lot of data but with no ground truth. We used well-known unsupervised indices for our experiments, as described in section (2.6); those indices include Silhouette index, Davies Bouldin index, Calinski harabsz index and SDbw index. For the comparative experiments, we used the newer version ISEC-V2 and other SRG methods like Adams, Shih SRG and Verma-Otsu SRG by using the whole dataset.

We used Matlab for our experimental study but we also used R due to its robustness using unsupervised validation indices. We can see the correlation comparison of Matlab and R in Appendix A.1, figures (A.1, A.2, and A.3). Moreover, the SDbw index execution was not available on Matlab, hence no need to compare it.

The images, which have a strong gradient as in figure (4.2), the middle one, have the problem of the explosion that is why we used those images for our experiments of explosion detection. We used a very good strategy known as Contour Strength (CS), described in section (3.5).

4.3 Analysis of Iterative Seed Expanding Clustering

4.3.1 Comparing ISEC-V2 vs. ISEC

In previous section (3.2), we discussed the stop conditions of ISEC and its complexity. Moreover, we also explained in section (3.3) the new version, ISEC-V2, and its stop conditions. In this section, we will go one-step ahead by explaining the formation of the coastline and how will the new algorithm works.

One of the main contributions of this study was to determine a good stop condition for an iterative version of SEC (ISEC). As in section (3.2), we reviewed that there are many stop conditions of ISEC, which did not only increases the granularity of the solution but also create complexity.

The proposed new version of ISEC (ISEC-V2) is more simple and efficient in terms of execution. According to the expert domain knowledge, the seeds exist near the coast so there was a need to determine the coastline for better analysis of the stop condition.

ISEC-V2 starts similarly with the initial seed, the seed with minimum temperature value. Before the seed selection, it calculates the coastline pixels, their spatial values, and the coordinates. By taking the advantage of the domain knowledge, it calculates the distance of the seed pixel from the coastline and if the seed is far from the coastline like 10km, the algorithm will stop. We experimentally tested that those seeds that were far from the coastline they were not mature enough to grow the cluster.

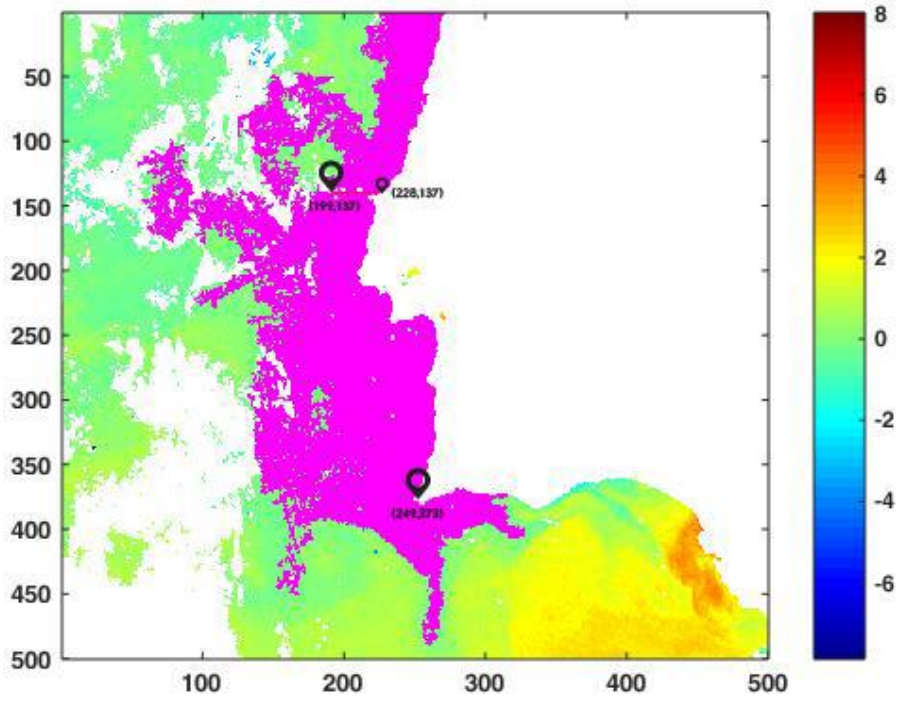


Figure 4.4: Coordinates of the seeds and the coastline.

The coastline distance is the main stop condition of ISEC-V2. In figure (4.4), the pixel with coordinates (249,373) is the initial seed and it grows the cluster. The second seed has coordinates (199, 137) and the algorithm calculates the distance between these pixels from the coastline, which has been recorded already. At this point the coastline has coordinates (228,137), hence the x-axis distance is calculated 29 (228-199), and the distance is more than 10km this is why the algorithm will stop the execution.

Now if seed passes the above criterion of distance it goes to the feature extraction criterion that is; *first mean minus minimum.* We did not fix the 'epsilon' value but calculate it from automatic threshold method Otsu (1979).

We completely removed the extra stop conditions from the previous version. The experiments were performed for the whole set of data and the results were stable. We were not only decreased the granularity of the solution but also increased the efficiency of the algorithm in terms of iterations. In the previous version, the iterations were fixed to run the algorithm five times by no means, instead of, the newer version had maximum two or three iterations, depending upon the total number of clusters as shown in the appendix, figure (A.6) and (A.7).

4.3.2 Comparing ISEC-V2 vs. I-Adams SRG

It is always very hard to find out a good index for the unsupervised analysis because we do not have any ground truth and it is a very challenging problem nowadays. In this study, we took most popular indices as described in 'state of the art' section. How these indices better depicting the results, we had two tests. One was the correlation test with f-measure and the second was the visual inspection. By doing these two tests, we found that Silhouette index (SI) had edge over all three indices we used, like Davies Bouldin (DB), Calinski Harabsz (CH) and SDbw index. A comparative study by Chouikhi et al. (2015) also concluded that Silhouette Index typically perform better than the others.

S.No	Index	Abbreviation	Best	Protocol
1	Silhouette index	SI	MAX	Matlab
2	Davies Bouldin	DB	MIN	R
3	Calinski Harabsz	CH	MAX	R
4	SDbw Index	SDbw	MIN	R

Table 4.2: Unsupervised Indices

In our comparative study, we first apply ISEC-V2 and I-Adams's SRG over all the images of Portugal as well as the images of Canary. We took the advantage of two years 1998 and 1999 with ground truth and visually inspect the segmentation results. We observed that ISEC-V2 perform better than Adams did, as we can see in the figure (4.5). The Silhouette index got better results for the years 1998 NGT (no ground truth), 2000 and 2001. The Silhouette index for the years 1998, 1999 (with ground truth) and 2002 has the same results.

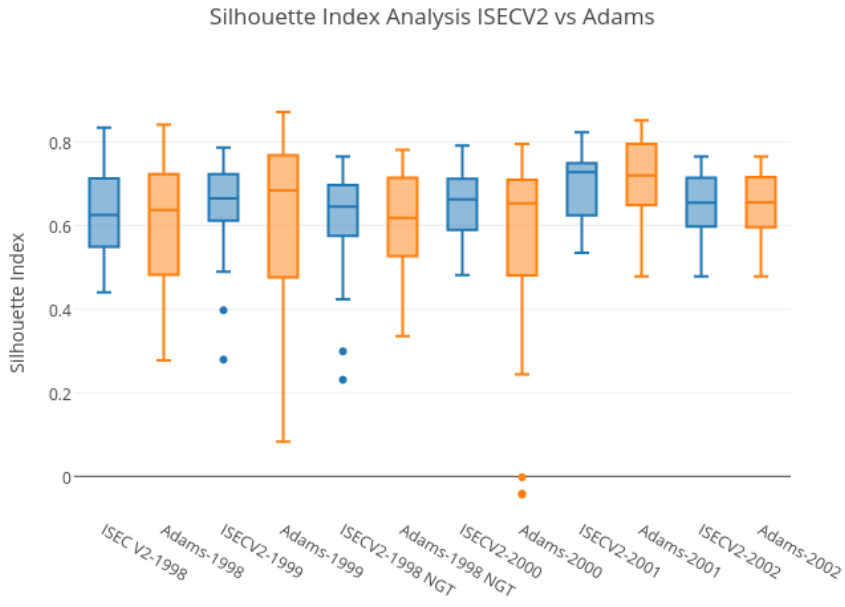


Figure 4.5: Silhouette Index analysis comparing ISEC-V2 with I-Adams using the whole data set of SST images (Portugal).

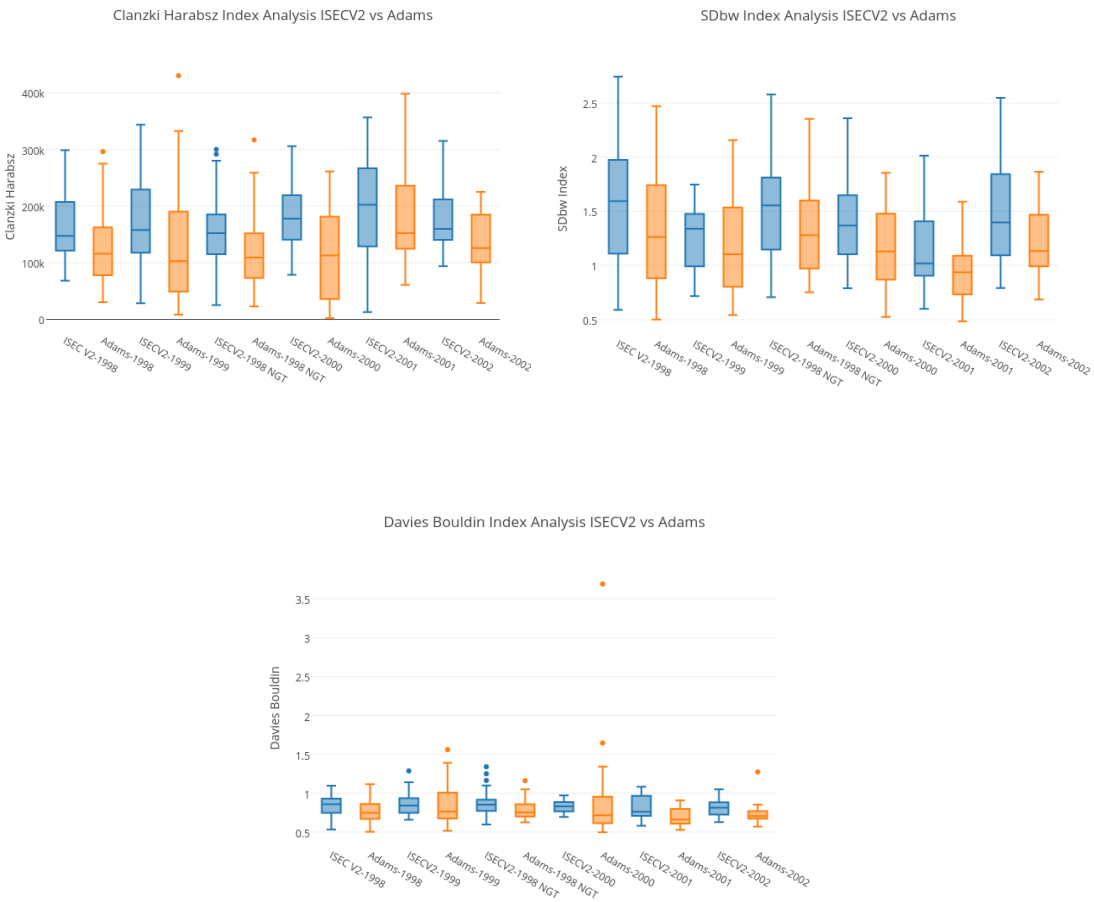


Figure 4.6: Calinski, SDbw and Davies Bouldin indices analysis comparing ISEC-V2 with I-Adams using the whole data set of SST images (Portugal).

The results of unsupervised indices showed that Silhouette index and Calinski Harabsz have good values whereas the Davies Bouldin and SDbw have not scored well as compared to the other two. Davies Bouldin and SDbw are considered good when they have minimum score values whereas Silhouette and Calinski Harabsz are good when they got maximum values. According to the overall results, ISEC-V2 won from I-Adams using SI and CH whereas I-Adams got better results from ISEC-V2, using DB and SDbw. As shown in figure (4.6), top left, all the years ISEC-V2 has higher values than I-Adams did whereas in the same figure the top right and the bottom image I-Adams has performed well.

The images of Canary had different geographical aspects as compared to the Portugal images so separate analysis was required. In the figure (4.7), you can see Silhouette index and SDbw have good results with compare to the I-Adams whereas Calinski and DB have lower results.

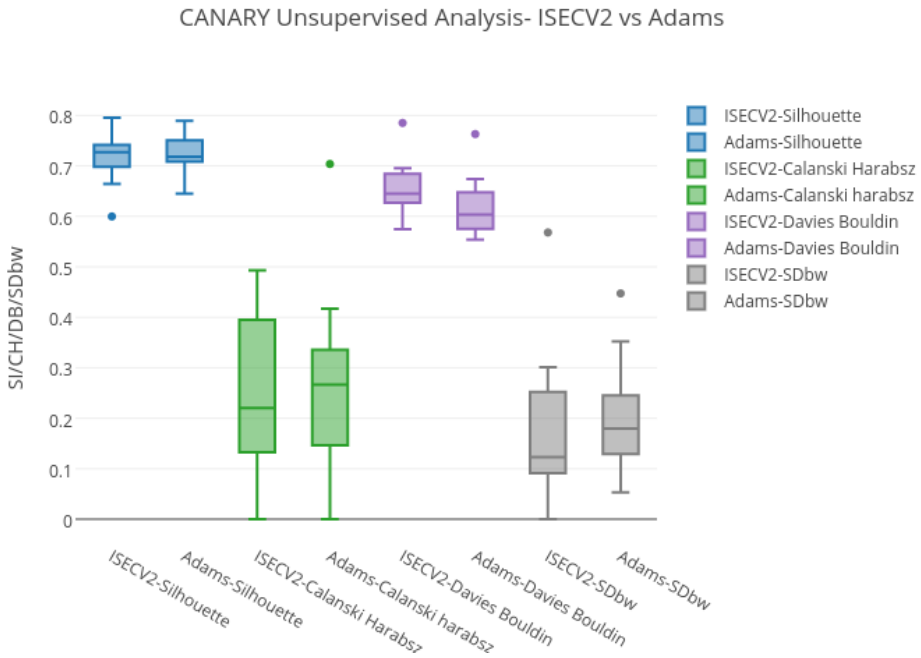


Figure 4.7: SI/CH/DB/SDbw indices comparing ISEC-V2 with I-Adams using the images of Canary.

4.3.3 Comparing ISEC-V2 vs. Conventional SRG Methods

In this section, we compared the results of ISEC-V2 with Shih and Cheng (2005), and Verma (2011) using the whole data set. The purpose of this comparative analysis was that how ISEC-V2 performed well from other classical SRG methods.

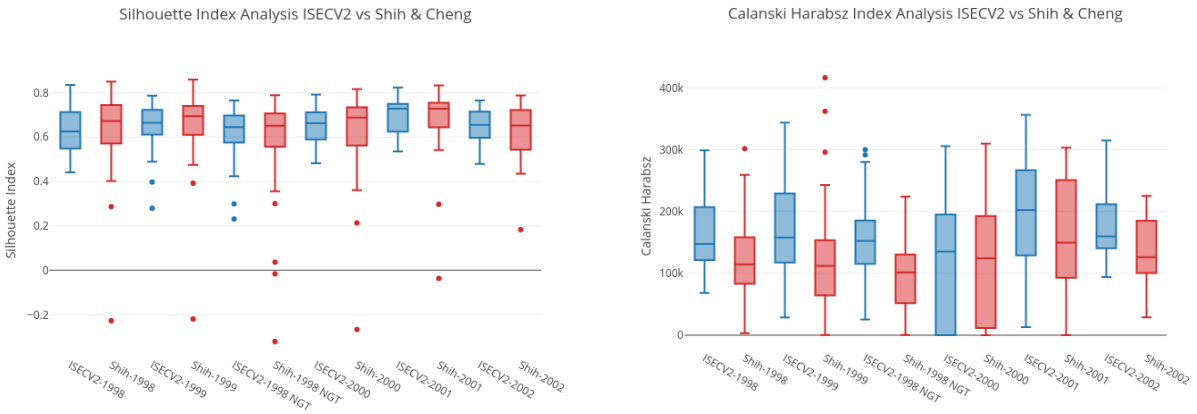


Figure 4.8: SI/CH indices analysis comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal).

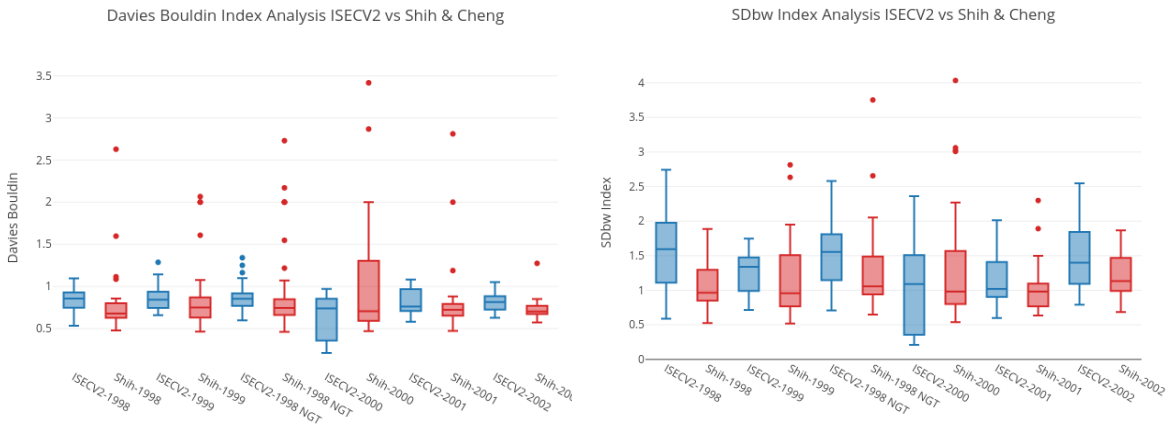


Figure 4.9: DB/SDbw indices analysis comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal).

As described in section (3.6), the unsupervised indices are not concordant for all data as in this case Calinski Harabsz performed well instead of Silhouette index. The indices DB and SDbw for ISEC-V2 are still not good. By inspecting visually, we found that 40% images of Portugal (1998) had explosion when using Shih SRG as shown in the table (4.3), whereas ISEC-V2 had only 20% images with an explosion. Similarly, 38% images (1999) had an explosion using Shih SRG, whereas ISEC-V2 had only 22% images with the explosion.

We apply the same experiments over Verma-Otsu and found that none of the unsupervised indices performs well for ISEC-V2, can be seen in figures (4.10 and 4.11).

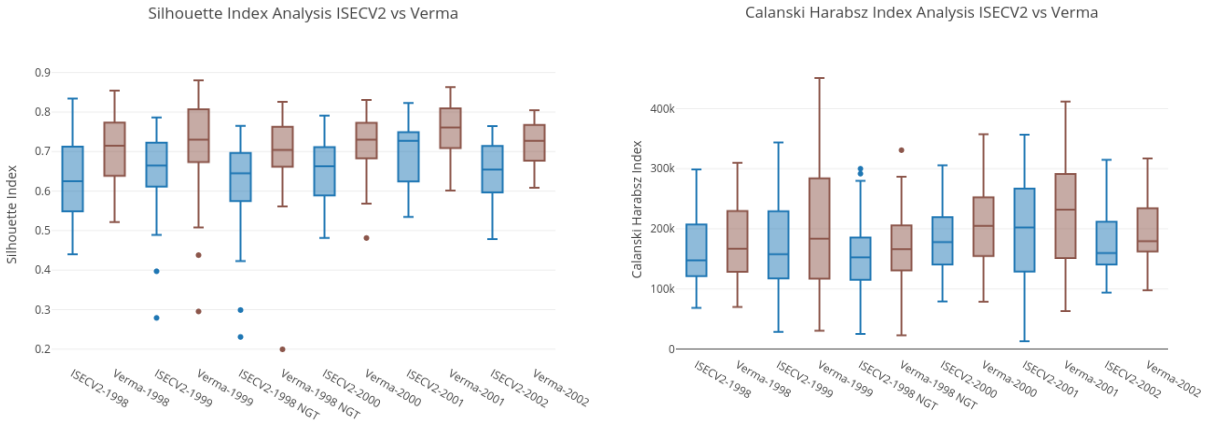


Figure 4.10: SI/CH indices analysis comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal).

Images		Years	ISEC-V2	Shih SRG	Verma SRG	I-Adams
30	Portugal	1998	20%	40%	46%	33%
31		1999	22%	38%	25%	32%
52		1998 NGT	13%	32%	36%	23%
32		2000	18%	37%	31%	18%
30		2001	6%	23%	16%	6%
22		2002	22%	4%	40%	9%
10	Canary	N/A	0%	20%	50%	20%

Table 4.3: Visual analysis (Explosion and Under-segmentation) for the whole dataset of Portugal and the Canary.

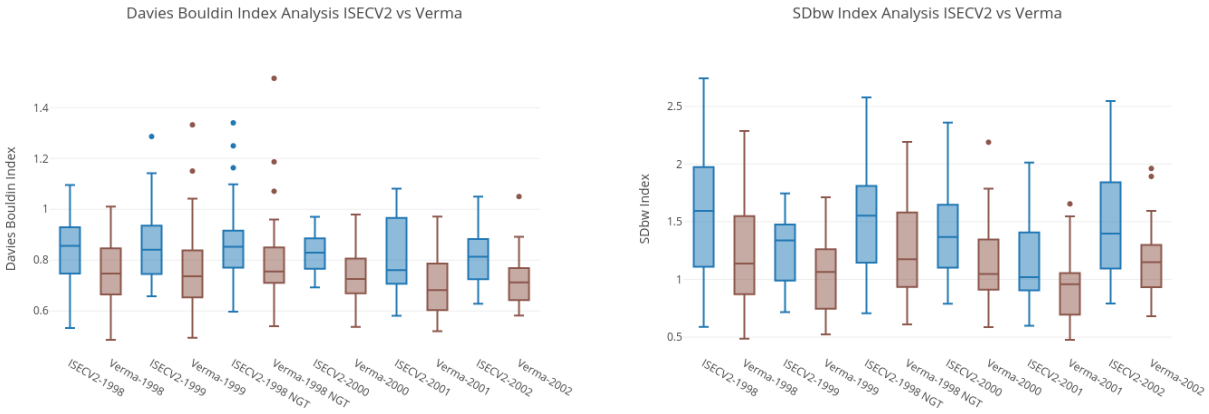


Figure 4.11: DB/SDbw indices comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal).

By inspecting the segmentation results visually, we found that 46% images (1998) had explosion using Verma SRG as shown in the table (4.3), whereas ISEC-V2 had only 20% images with the explosion. Similarly, 36% images of the year (1998 NGT) had explosion when used Verma whereas only 13% images had explosion in case of ISEC-V2. Hence, we can say that the visual results and the results from these unsupervised indices are not concordant, therefore; a good strategy needed to evaluate the results.

4.4 Fusion Strategy for Unsupervised Clustering Evaluation

In above section, we conclude that unsupervised indices like SI, DB, CH and SDbw performed differently for Adams, Shih and Verma SRG. As these indices are based on the internal information, so for different data the results are different (Arbelaitz, 2012). Hence, there was a need to find a good unsupervised solution that is why we introduced the fusion strategy to analyze the results with more accuracy. As discussed in section (3.6), for fusion measures we need to normalized data before applying the strategy. In this section, we will analyze this fusion strategy for the whole data set of Portugal as well as Canary. Moreover, fusion indices consider being good when they have the maximum score.

4.4.1 Unsupervised Fusion Analysis of ISEC-V2 vs. I-Adams SRG

We applied fusion indices SF-A, SF-G, SF-H and SF-Med on the results we got with ISEC-V2 and I-Adams and found very good results. The higher fusion score indicated the good results.

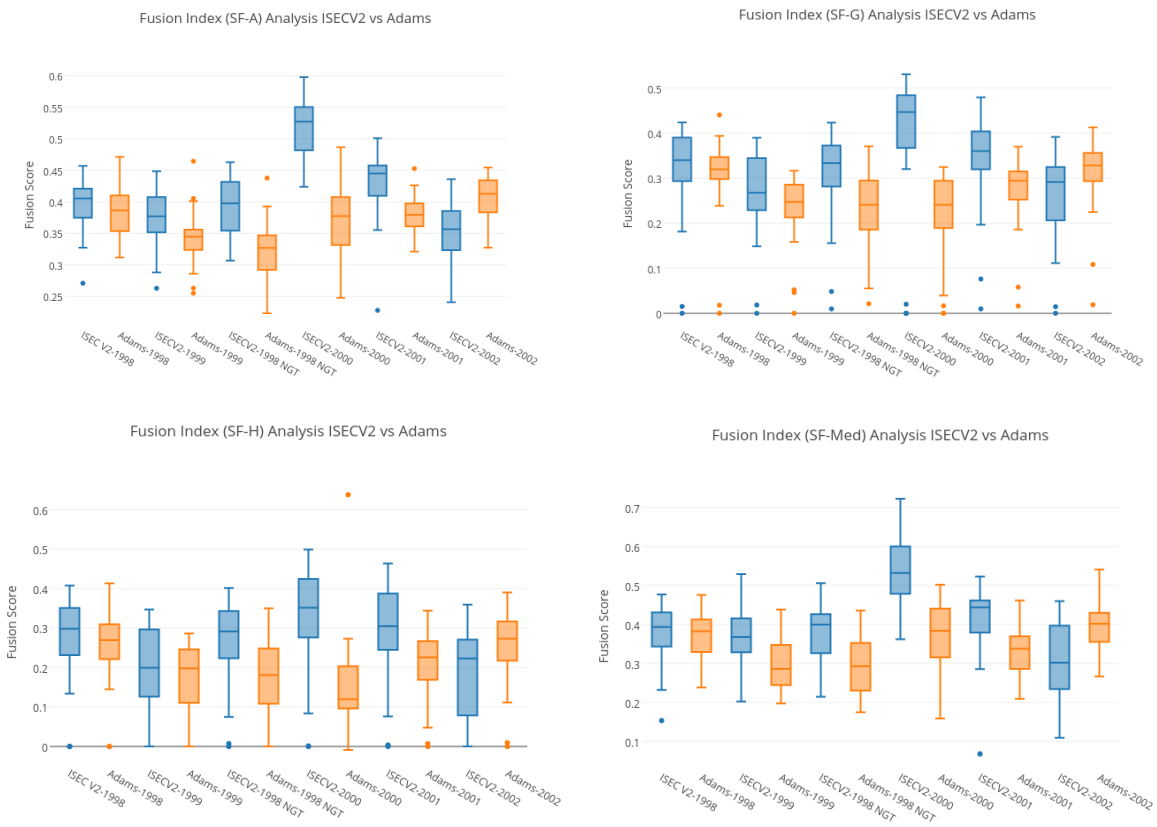


Figure 4.12: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with I-Adams using the whole dataset of SST images (Portugal).

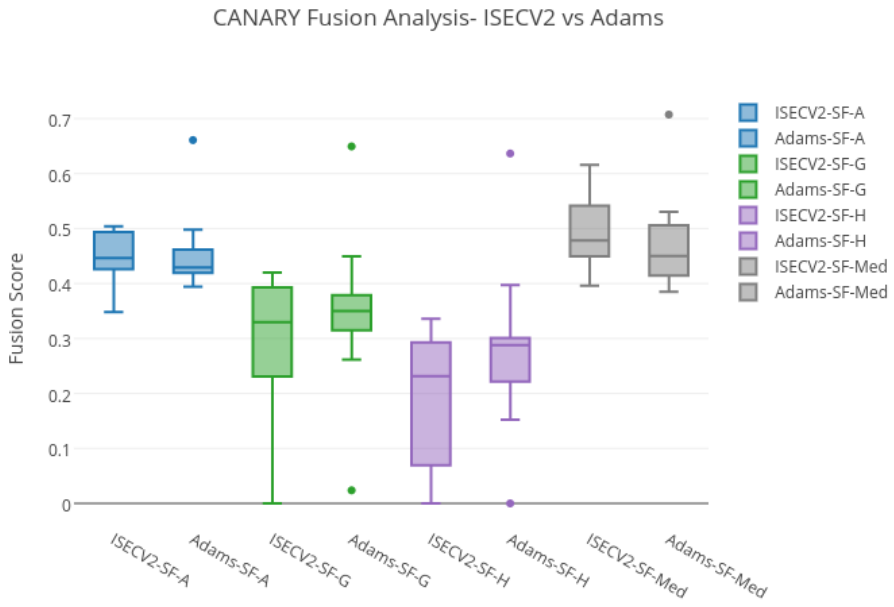


Figure 4.13: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with I-Adams for the images of Canary.

As described in section (3.6), SF-A is the fusion measure that takes the average of all the unsupervised indices. In figure (4.12), top left, we found that for all the images of Portugal except 2002 ISEC-V2 fusion scores were higher than the I-Adams was. The results we got from SF-G, which is the root of indices product, were pretty same as SF-A because only the year 2002 we got less fusion score whereas all the years the score was higher than the I-Adams as shown in figure (4.12), top right. The indices SF-H and SF-Med had the same results as the above two indices and also shown in figure (4.12), bottom left and right.

The same fusion measures applied to Canary images and the results were little different as shown in figure (4.13). The indices SF-A and SF-Med got the higher fusion score as compare to I-Adams whereas the other indices say SF-G and SF-H got lower fusion score.

4.4.2 Unsupervised Fusion Analysis of ISEC-V2 vs. Conventional SRG

All fusion indices applied over classical SRG methods (Shih and Cheng, 2005), (Verma et al., 2011) and the results were promising. When apply fusion on Shih SRG with all the images of Portugal we found that SF-A fusion score was higher for all the images of Portugal except the year 2002 as shown in figure (4.14), top left.

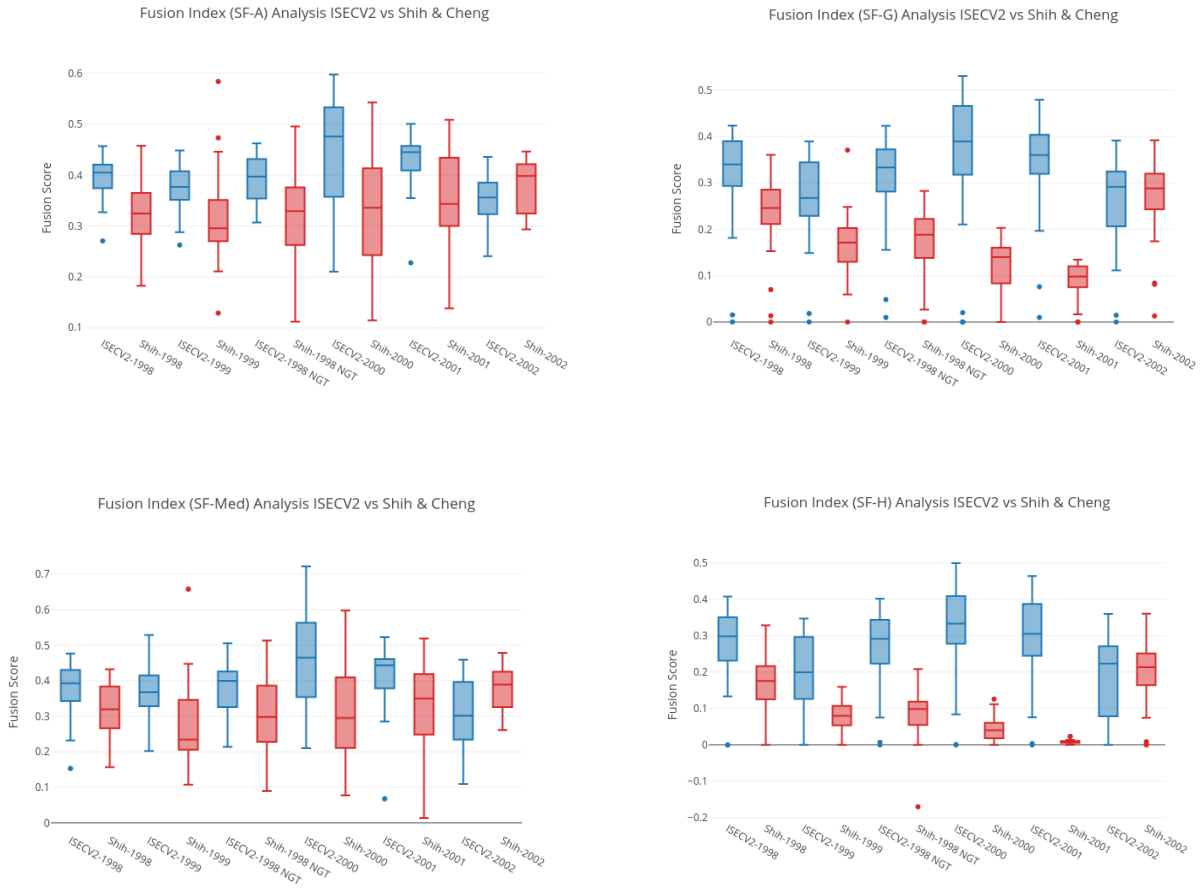


Figure 4.14: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Shih SRG using the whole dataset of SST images (Portugal).

For all other indices SF-G, SF-H and SF-Med had the same trend as SF-A shown in figure (4.14). The year 2002, Shih performed well as compare to ISEC-V2 and when we did the visual inspection the same results we found, can be seen in the table (4.3), that only 4% explosion found in this year using Shih SRG. The images of Canary, ISEC-V2 got higher fusion scores by using all the four fusion measures as shown in figure (4.16).

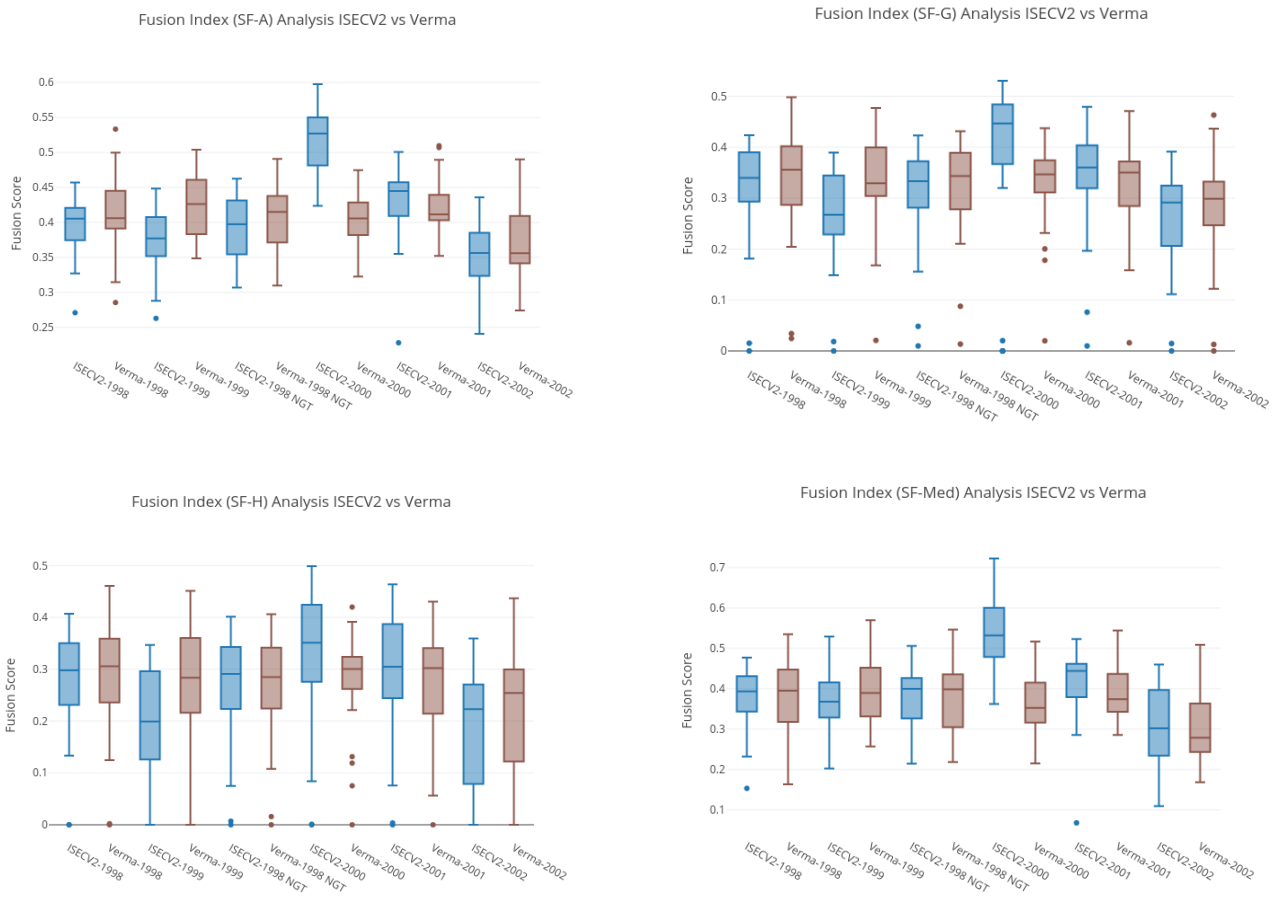


Figure 4.15: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Verma SRG using the whole dataset of SST images (Portugal).

Similarly, when fusion applied to Verma-Otsu, we found that the results were different from the Shih SRG. SF-A score for the year 1999 was lower when used ISEC-V2 whereas all the other years ISEC-V2 leads, as shown in figure (4.15). When we did a visual inspection, we found that Verma had only 25% explosion in the year 1999 as shown in the table (4.3), and for the year 2000 ISEC-V2 score prominently high as compare to Verma. The other fusion indices had the same result as SF-A did.

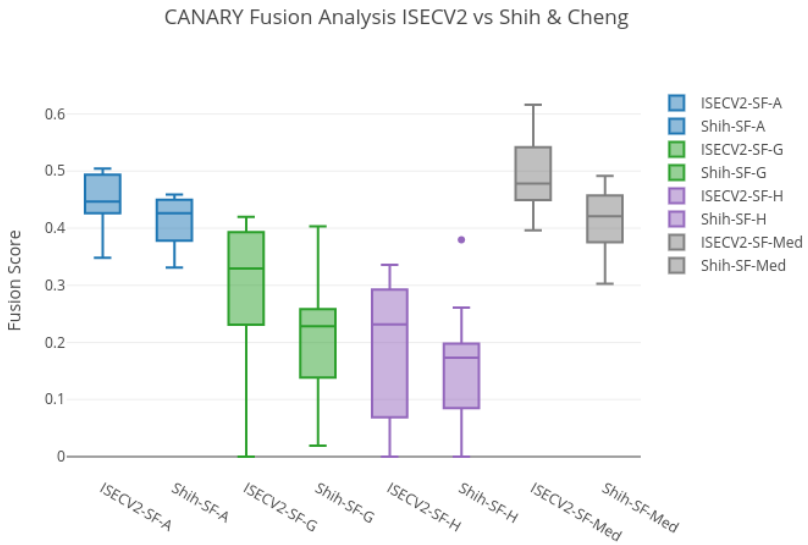


Figure 4.16: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Shih SRG using the images of Canary.

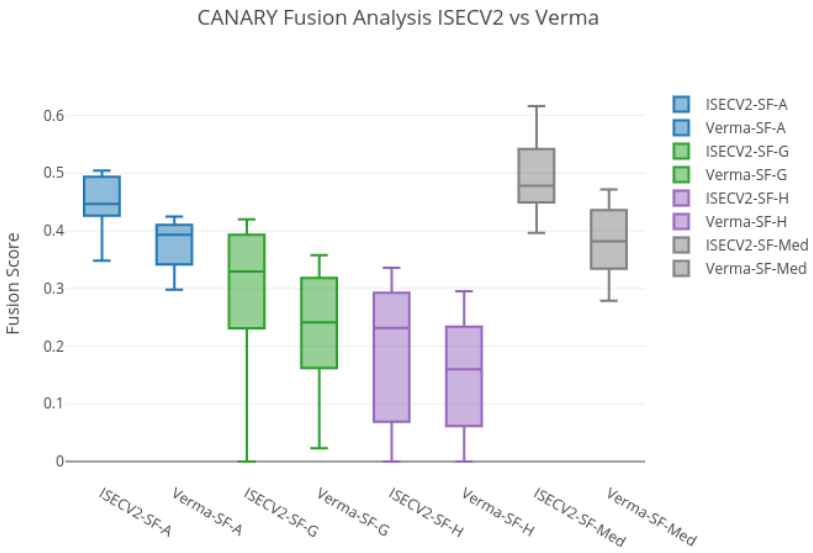


Figure 4.17: SF-A/SF-G/SF-H/SF-Med fusion measures comparing ISEC-V2 with Verma SRG using the images of Canary.

Canary images were different from Portugal images so we need separate experiments. We applied fusion indices on Canary images and found good results with ISEC-V2. In figure (4.17), we had the results with Verma SRG and ISEC-V2 and you can see it clearly that ISEC-V2 scored better than Verma.

Algorithm	SF-A	SF-G	SF-H	SF-Med
ISEC-V2	**83%**	**47%**	**23%**	**72%**
I-Adams	61%	12%	3%	47%
Shih SRG	42%	4%	1%	40%
Verma SRG	82%	37%	15%	55%

Table 4.4: Fusion Analysis for the whole dataset of Portugal as well as Canary.

4.5 Analysis of the Explosion Control in ISEC-V2

As described in section (4.2), the images with weak gradient had the problem of explosion or so-called leakage. The reason for this problem is to make difference between the boundary and the region of interest. In our data, 15 to 20 percent of images had this problem of the explosion, as shown in the table (4.3). In figure (4.18), the right side image is the segmentation result showing the explosion problem.

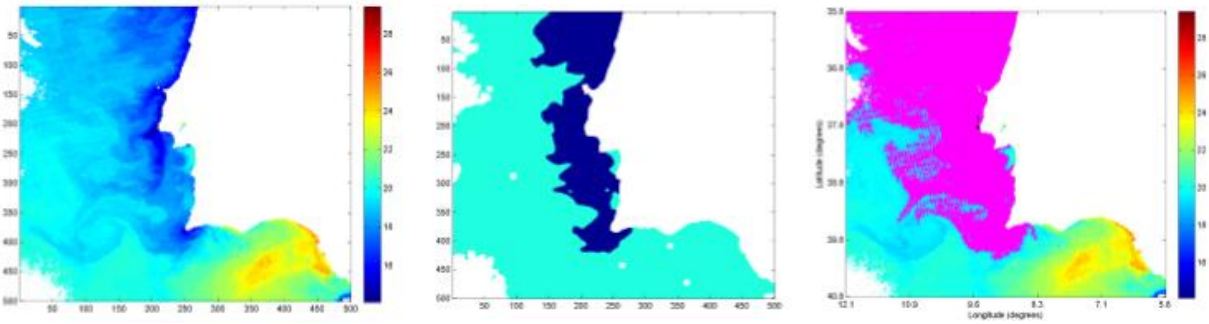


Figure 4.18: SST images, left image is the original image, middle one is the ground truth and the right image is segmented image with explosion.

The problem of the explosion is very challenging in SRG methods and there is a need to identify it. We did experiments to find out a good strategy for leakage identification. First, it was very important to find the start of the leakage then stop the algorithm to grow the region further. The Contour Strength (CS) strategy as described in section (3.5) showed good results, for the problem of the explosion.

DRG introduced a Contour Strength (CS) strategy, according to that the region grows when they have a strong CS. The selection of CS for our data was empirical because we took images with explosion and without explosion and visualized the results. In figure (4.19), we observed that in the left image CS starts with increasing rate but at certain point, it goes down, whereas in the image with no explosion CS continuously moves upwards. According to DRG, the start of the explosion is when CS goes down or becomes weaker because when region grows the CS maximizes. We applied CS to all the images of 1998 and 1999 and found the same trend.

After performing the experiments on all the images of 1998 and 1999, we also found that at some points, images with no explosion had weak CS for a very short time, as shown in figure (4.20). In order to solve this problem, we used the concept of Moving Average so that the curve could be smooth.

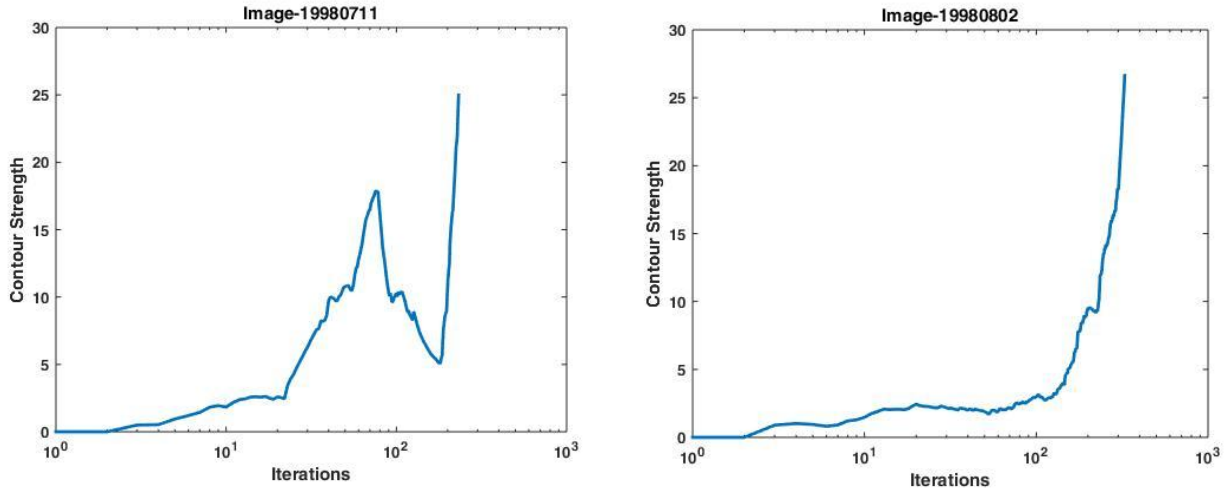


Figure 4.19: Contour Strength corresponding to the number of iterations, the left image is the one with explosion and the right is without explosion.

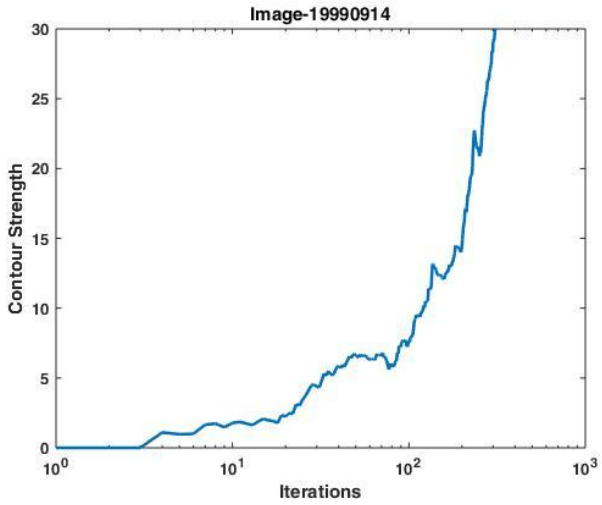


Figure 4.20: Image 19990914 Portugal, Contour Strength with no-explosion.

At this point, we need to find out a benchmark value that could depict the starting point of the explosion.

4.5.1 Contour Strength's first Derivative

We calculated the first derivative of CS values, which already been recorded and we found very interesting results. We performed experiments and found that when CS goes down the first derivative becomes negative, hence we set the benchmark that; *explosion starts when the first derivative becomes negative.*

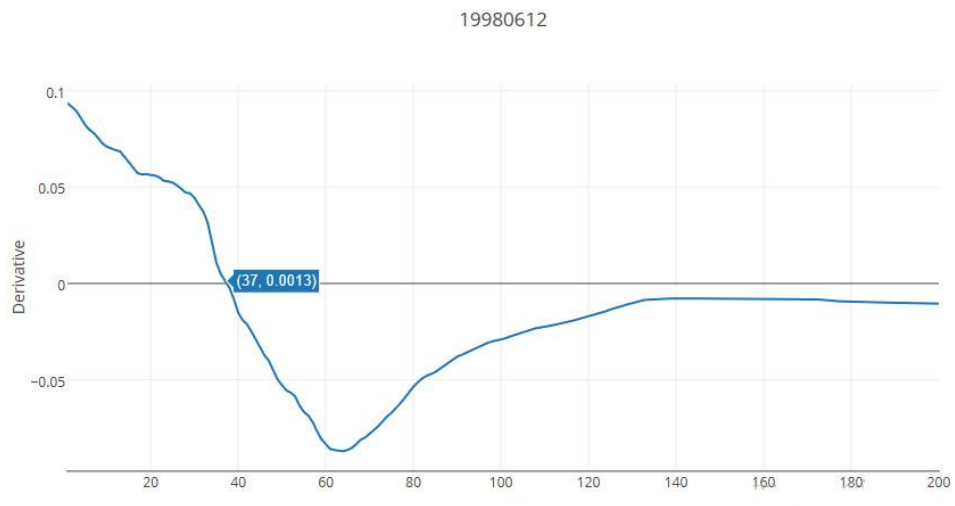


Figure 4.21: First derivative corresponding to the number of iterations of the image 19980612 with explosion.

We also apply the moving average on the first derivative values that already been calculated and recorded but the problem was to find out a good size of the window for moving average. We did that experimentally by applying different sizes of the window and check that which size of the window had the best match for our data. In figure (4.21), the image with explosion shows that at iteration 37, the derivative starts to go negative, hence shows the start of the explosion.

4.5.2 Modification in RSEC: Explosion Control

As described in section (3.3), ISEC is the combination of outer and inner stop conditions and inner stop condition is related to the cluster formation. In order to implement the CS strategy, we did some modifications in the inner stop condition, such as:

1) Calculate Contour Strength (CS).
2) Calculate first derivative of the CS.
3) Apply moving average on the derivative, of window size 80.
4) Record that number of iteration when derivative goes negative.

After these calculations, we are now able to modify the inner stop condition as described in section (3.1), i.e., *when clusters become stable*. Now the new stop condition becomes:

when clusters are stable or when CS becomes negative.

We implemented and applied this new stop condition on all the images of explosion and without explosion, we found that the images with the explosion, the algorithm stops when CS < 0. Moreover, the images with no explosion, the algorithm had no effect and the segmentation results were same like before, because CS was positive.

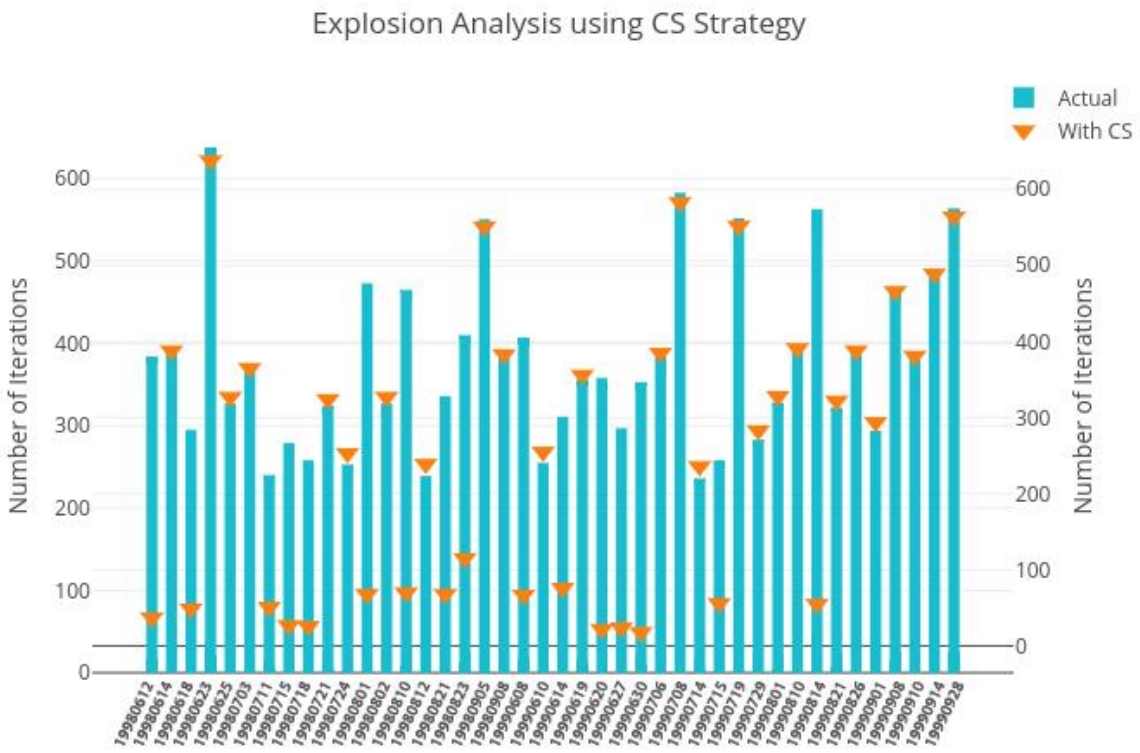


Figure 4.22: Explosion Analysis for the images of 1998-1999, Iterations with actual algorithm and with the CS.

As shown in figure (4.22), the images with explosion did not complete the total number of iterations (blue bars), whereas they stop growing when explosion starts or when CS < 0. The blue bars are showing the total number of iterations when clusters become stable, whereas the orange triangular mark represents that the algorithm stops when CS < 0. The images with no explosion completed the total number of iterations and vice versa. Hence, the new version of RSEC detects the explosion with CS strategy efficiently by stopping the algorithm to grow further.

4.6 Summary of the Results

We took the images with explosion; calculated the CS first, after getting the first derivative of CS, applied moving avg. When the derivative became negative, we stop the algorithm in order to control explosion.

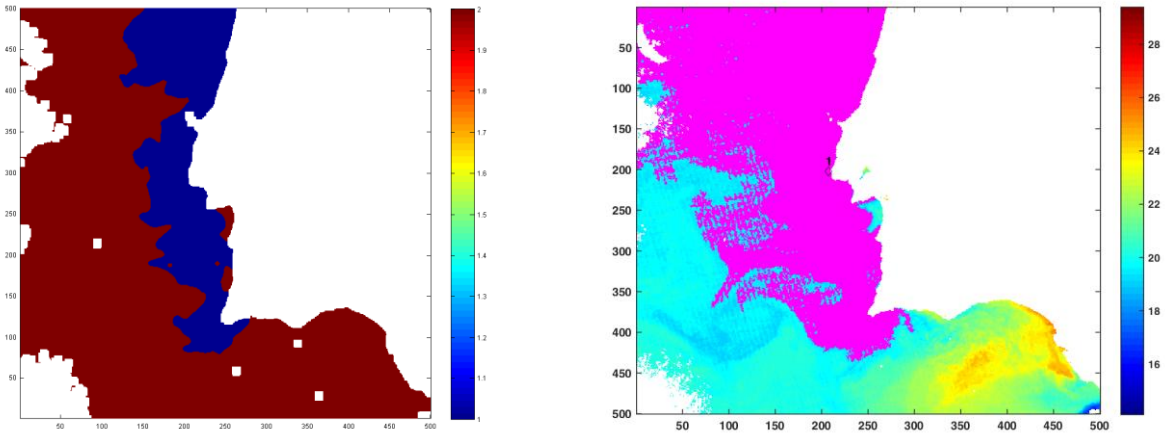


Figure 4.23: Two SST images, the left one is the ground truth and the right one is the segmented image with explosion.

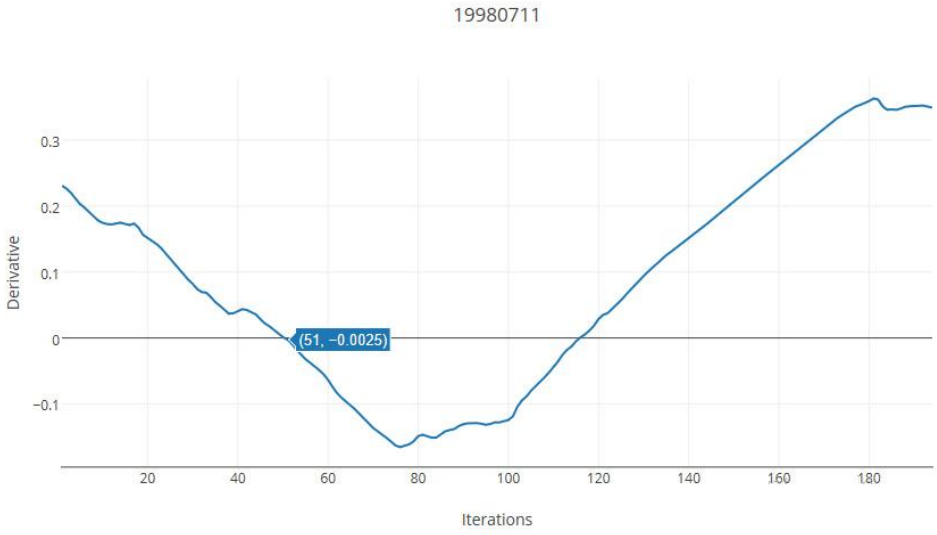


Figure 4.24: First Derivative corresponding to the number of iteration for the image 19980711 (Portugal).

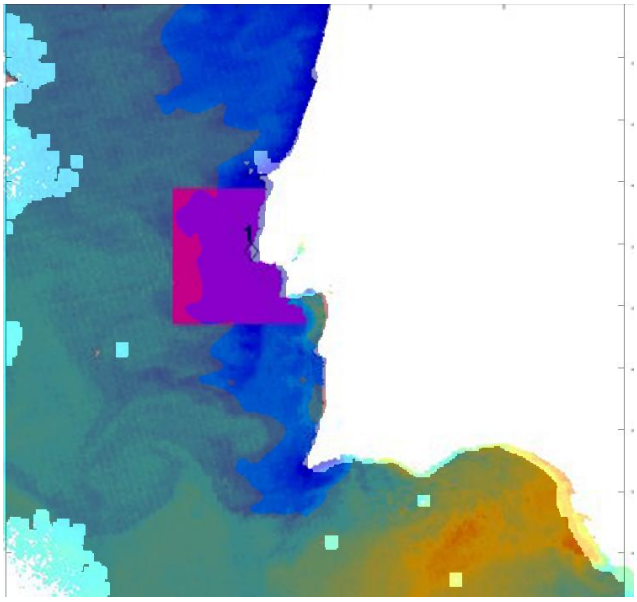


Figure 4.25: Resulting image when RSEC algorithm stops at iteration 51 and ground truth in the background.

We can easily find the fact from the figure (4.25), that the condition CS < 0 (negative) stops the algorithm at the point of explosion. The purplish area is the region of interest and the pinkish area is the segmented region with 51 iterations, as shown in figure (4.25). Hence, if the algorithm not stops, the segmentation continues to grow, that results in the explosion as shown in figure (4.23), the right side image.

CHAPTER 5

CONCLUSION AND FUTURE WORK

Automatic detection of upwelling phenomena is always a need of oceanographers. The SEC algorithm resolved this problem efficiently but it finds only one region at a time. An iterative version of SEC (ISEC) successfully applied for all the images of Portugal as well as the images of Canary Island in order to find the discontinuous regions. A special version of SEC (SEC-ST) that used in ISEC and further in ISEC-V2 calculates the threshold dynamically.

The iterative version ISEC although resolved the problem of a discontinuous region but it had the problem of complexity of the algorithm that was one of the important contributions of this work. The proposed version ISEC-V2 is simple and efficient with good segmentation results. The stop condition, which makes the algorithm simple, is the coastline distance. It used the domain knowledge that seeds near the coast have the potential to grow regions. ISEC-V2 has compared to the previous version ISEC and other SRG methods and the results were improved in terms of efficiency and efficacy. Due to the modification in outer stop condition, the results efficiency improved as shown in figure (3.4), where we can see that the iterations are not more than two that reduces the total time of execution. On the other hand, the efficacy also improved can be seen in the table (4.4), where ISEC-V2 prominently wins.

The segmentation results evaluated by unsupervised indices like Silhouette Index, Davies Bouldin, Calinski Harabsz and SDbw. By the comparative experimental study of ISEC-V2 with Adams and other classical SRG's like Shih and Verma, it concluded that none of the above indices performed consistently hence there was a need to find a good strategy to evaluate the results with more accuracy. In this work, we introduced a very new fusion strategy based on the above unsupervised indices and compared the results with not only the index values but also visual inspection. The results were promising for the whole set of data.

In all SRG methods, the leakage or the so-called explosion is a very challenging problem. In this study different novel explosion, control strategies were introduced and the purpose was to remove it from the segmentation. The removal of explosion was the target of this study therefore a very strong strategy introduced called Contour Strength (CS). It based on the fact that regions grow with strong contour i.e. if the region has a weak contour that shows the start of explosion. The problem of explosion divided into two parts first was the detection of the point of explosion from where the leakage would start and the second part was to remove it from the final segmentation result.

In this dissertation, the first part of the explosion control has achieved by finding the start of the explosion using CS and stop the algorithm at that point. The future work can be linked with this by start the algorithm again and continue to find the segmentation.

Bibliography

Adams, R., and Bischof, L. (1994). "Seeded region growing". In: Pattern Analysis and Machine Intelligence, IEEE Transactions on 16(6), pp. 641–647.

Al-Faris, A. Q., Ngah, U. K., Isa, N. A., and Shuaib, I. L. (2013). "Computer-Aided Segmentation System for Breast MRI Tumour using Modified Automatic Seeded Region Growing (BMRI-MASRG)". In: Journal of digital imaging 27(1), pp. 133–144.

Afifi, A., and Ghoniemy, S. (2015). "New Region Growing Based on Thresholding Technique Applied to MRI Data". In: I.J. Computer Network and Information Security, 2015, pp.7, 61-67.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Perez, J. M., and Perona, I. (2012). "An extensive comparative of cluster validity indices" by Department of Computer Architecture and Technology, University of the Basque Country.

Bins, L. S., Fonseca, L. M., Erthal, G. J., and Ii, F. M. (1996). "Satellite imagery segmentation: a region growing approach". In: Simpósio Brasileiro de Sensoriamento Remoto 8, pp. 677–680.

Bagli, S., Soille, P., and Fermi, E. (2004). "Automatic delineation of shoreline and lake boundaries from Landsat satellite images". In: Proceedings of initial ECO-IMAGINE GI and GIS for Integrated Coastal Management, Seville, pp. 13–15.

Chen, H. L., Samavati, F. F., Sousa, M. C., and Mitchell, J. R. (2006). "Sketch-based volumetric seeded region growing". In: Proceedings of the Third Eurographics conference on Sketch-Based Interfaces and Modeling. Eurographics Association, pp. 123–130.

Calinski, T., and Harabasz, J. (1974). "A dendrite method for cluster analysis". In: Comm. In Statistics, vol. 3, no. 1, pp. 1-27.

Chouikhi, H., Charrad, M., and Ghazali, N. (2015). "A Comparision Study of Clustering Validity Indices". In: IEEE 978-1-4673-6587-1/15.

Chaudhari, S., Balasubramanian, R., and Gangopadhyay, A. (2008). "Upwelling detection in AVHRR sea surface temperature (SST) images using neural-network framework". In: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International. Vol. 4. IEEE, pp. IV–926.

Davies, D. and Bouldin, D. (1979). "A cluster separation measure". In: Pattern Analysis and Machine Intelligence, IEEE Transactions on(2), pp. 224–227.

DErrico, J (2006). "Surface fitting using gridfit". In: Matlab Central File Exchange (2006).

Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". In: Journal of Cybernetics. 3 (3): pp. 32–57.

Fan, M., and Lee, T. C. M. (2015). "Variants of Seeded Region Growing". In: IET image processing.

Gao, Y., Mas, J. F., Kerle, N., and Navarrete, P. J. A. (2011). "Optimal region growing segmentation and its effect on classification accuracy". In: International Journal of Remote Sensing 32(13), pp. 3747–3763.

Hadwiger, M., Laura, F., Rezk-Salama, C., Hollt, T., Geier, G., and Pabel, T. (2008). "Interactive volume exploration for feature detection and quantification in industrial CT data". In: Visualization and Computer Graphics, IEEE Transactions on 14(6), pp. 1507– 1514.

Halkidi, M., and Vazirgiannis, M. (2001). "Clustering validity assessment: finding the optimal partitioning of a data set ". In: ICDM, Washington, DC, USA, pp. 187-194.

Hubert, L. and Arabie, P. (1985). "Comparing partitions". In: Journal of Classification 2(1), pp. 193–218. issn: 0176-4268.

Heimann, T., Thorn, M., Kunert, T., Meinzer, H. P. (2004). "New Method for Leak Detection and Contour Correction In Seeded Region Growing Segmentation."

Kryszczuk, K., and Hurley, P. "Estimation of the number of clusters uisng multiple clustering validity indices". IBM Zurich Research Laboratory, Switzerland.

Kittler, J., and Illingworth, J. (1986). "Minimum error thresholding". In: Pattern recognition 19(1), pp. 41–47.

Kriebel, S. K., Brauer, W., and Eifler, W. (1998). "Coastal upwelling prediction with a mixture of neural networks". In: Geoscience and Remote Sensing, IEEE Transactions on 36(5), pp. 1508–1518.

Kronman, A., Joskowicz, L., Sosna, J. (2011). "Automatic Detection and Correction of Segmentation Leaks in Medical Images". In: Conference Paper.

Lachance, S., Bauer, R., and Warkentin, A. (2004). "Application of region growing method to evaluate the surface condition of grinding wheels". In: International Journal of Machine Tools and Manufacture 44(7), pp. 823–829.

Liu, Y., Li, Z., Xiong, H., Gao, X., and Junjie, W. (2010). "Understanding of Internal Clustering Validation Measures". In: IEEE Internation Conference on Data Mining.

Lopes, V. (2015). "Seeded Region Growing Methods for Automatic Upwelling Detection from Sea Surface Temperature Images". In: Master Thesis, Supervised by (Nascimento, S.).

Mishra, B., and Susaki, J. (2013). "Coupling of Thresholding and Region Growing Algorithm for Change Detection in SAR Images". In: Progress In Electromagnetic Research 143, pp. 519–544.

Marcello, J., Marques, F., and Eugenio, F. (2005). "Automatic tool for the precise detection of upwelling and filaments in remote sensing imagery". In: Geosciences and Remote Sensing, IEEE Transactions on 43(7), pp. 1605–1616.

Mat-Isa, N. A., Mashor, M. Y., and Othman, N. H. (2005). "Seeded region growing features extraction algorithm; its potential use in improving screening for cervical cancer". In: International Journal of The Computer, the Internet and Management 13(1), pp. 61–70.

Nieto, K., Demarcq, H., and McClatchie, S. (2012). "Mesoscale frontal structures in the Canary Upwelling System: New front and filament detection algorithms applied to spatial and temporal patterns". In: Remote Sensing of Environment 123, pp. 339–346.

Nascimento, S., and Franco, P. (2009). "Segmentation of upwelling regions in sea surface temperature images via unsupervised fuzzy clustering". In: Intelligent Data Engineering and Automated Learning-IDEAL 2009. Springer, pp. 543–553.

Nascimento, S., Franco, P., Sousa, F., Dias, J., and Neves, F. (2012). "Automated computational delimitation of SST upwelling areas using fuzzy clustering". In: Computers & Geosciences 43, pp. 207–216.

Nascimento, S., Sousa, F., Casimiro, H., and Boutov, D. (2005). "Applicability of fuzzy clustering for the identification of upwelling areas on sea surface temperature images".

Nascimento, S., Casca, S., and Mirkin, B. (2015). "A Seed Expanding Cluster Algorithm for Deriving Upwelling Areas on Sea Surface Temperature Images". Special issue on "Statistical learning in geosciences modeling: novel algorithms and challenging case studies". In: Computers & Geosciences, Vol. 85, pp. 74-85.

Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms". In: Systems, Man and Cybernetics, IEEE Transactions on 9(1), pp. 62–66. issn: 0018-9472.

Pohle, R., and Toennies, K. D. (2001). "A new approach for model-based adaptive region growing in medical image analysis". In: Computer Analysis of Images and Patterns. Springer, pp. 238–246.

Pottmann, H., Leopoldseder, S., Hofer, M., Steiner, T., and Wang, W. (2005). "Industrial geometry: recent advances and applications in CAD". In: Computer-Aided Design 37(7), pp. 751–766.

Plattner, S., Mason, D. M., Leshkevich, G. A., Schwab, D. J., and Rutherford, E. S. (2006). "Classifying and forecasting coastal upwelling in Lake Michigan using satellite derived temperature images and buoy data". In: Journal of Great Lakes Research 32(1), pp. 63–76.

Ridler, T. W., and Calvard, S. (1978). "Picture thresholding using an iterative selection method". In: IEEE transactions on Systems, Man and Cybernetics 8(8), pp. 630–632.

Rijsbergen, V. C. J. (1979). Information Retrieval. 1979.

Rousseeuw, P. J., (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". In: Computational and Applied Mathematics. 20: pp. 53–65.

Siebert, A. "Dynamic Region Growing" by Department of Computer Science, University of British Columbia.

Shih, F. Y., and Cheng, S. (2005). "Automatic seeded region growing for color image segmentation". In: Image and Vision Computing 23(10), pp. 877–886.

Sezgin, M., and Sankur, B. (2004). "Survey over image thresholding techniques and quantitative performance evaluation". In: Journal of Electronic imaging 13(1), pp. 146–168.

Stokking, R., Vincken, K. L., and Viergever, M. A. (2000). "Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data". In: Neuro Image 12(6), pp. 726–738.

Stroppiana, D., Bordogna, G., Carrara, P., Boschetti, M., Boschetti, L., and Brivio, P. A. (2012). "A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple Spectral Indices and a region growing algorithm". In: ISPRS Journal of Photogrammetry and Remote Sensing 69, pp. 88–102.

Tamim, A., Minaoui, K., Daoudi, K., Yahia, H., Atillah, A., Smiej, M. F., and Aboutajdine, D. (2013). "A simple and efficient approach for coarse segmentation of Moroccan coastal upwelling". In: Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European. IEEE, pp. 1–5.

Verma, O. P., Hanmandlu, M., Susan, S., Kulkarni, M., and Jain, P. K. (2011). "A simple single seeded region growing algorithm for color image segmentation using adaptive thresholding". In: Communication Systems and Network Technologies (CSNT), 2011 International Conference on. IEEE, pp. 500–503.

Wang, C. M., and Chen, R. M. (2012). "Automatic Vector Seeded Region Growing for Parenchyma Classification in Brain MRI".

Wong, S. J., and Zrimec, T. (2006). "Classification of lung disease pattern using seeded region growing". In: AI 2006: Advances in Artificial Intelligence.

Zanaty, E. A., and Asaad, A. (2013). "Probabilistic region growing method for improving magnetic resonance image segmentation". In: Connection Science 25(4), pp. 179–196.

Zhengtao, Z., Xiongyi, Y., Liuqian, H., and De, W. (2011). "Fast capsule image segmentation based on linear region growing". In: Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on. Vol. 2. IEEE, pp. 99–103.

Zhang, H., Fritts, J. E., and Goldman, S. A. (2007). "Image segmentation evaluation: A survey of unsupervised methods". In: Computer Vision and Image Understanding 110 (2008) 260–280.

Zhang, T., Yang, X., Hu, S., and Su, F. (2013). "Extraction of Coast line in Aqua culture Coast from Multi spectral Remote Sensing Images: Object-Based Region Growing Integrating Edge Detection". In: Remote Sensing 5(9), pp. 4470–4487.

APPENDIX A

THE RESULTS

A.1 Matlab and R Correlation Analysis

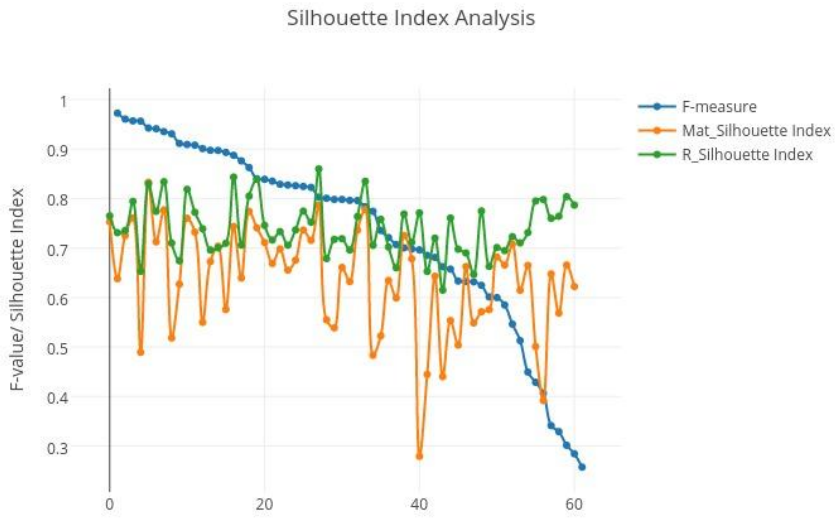


Figure A.1: Correlation of Matlab and R using Silhouette index with F-measure.

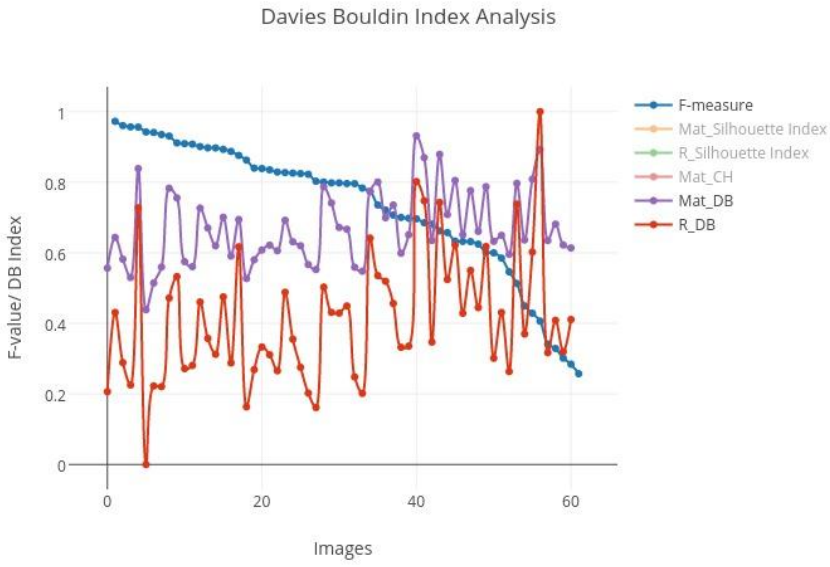


Figure A.2: Correlation of Matlab and R using Davies Bouldin index with F-measure.

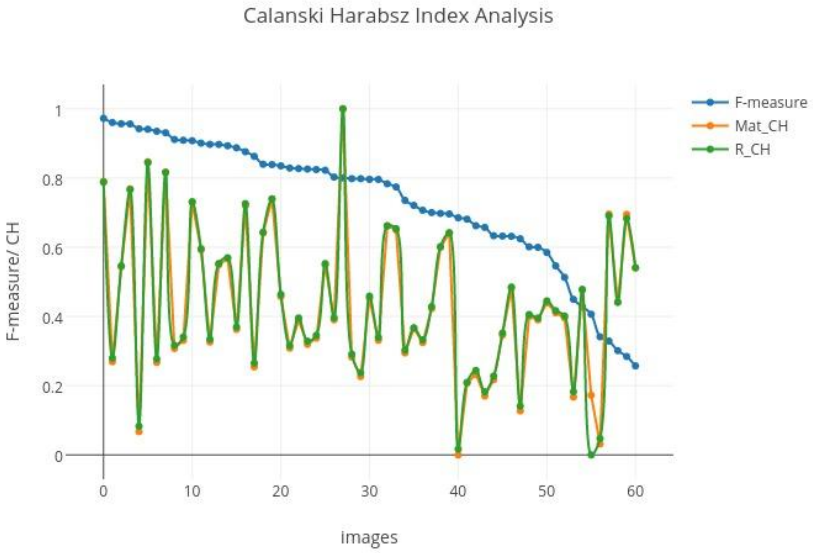


Figure A.3: Correlation of Matlab and R using Calinski Harabsz index with F-measure.

A.2 ISEC-V2 Comparative Results with ISEC

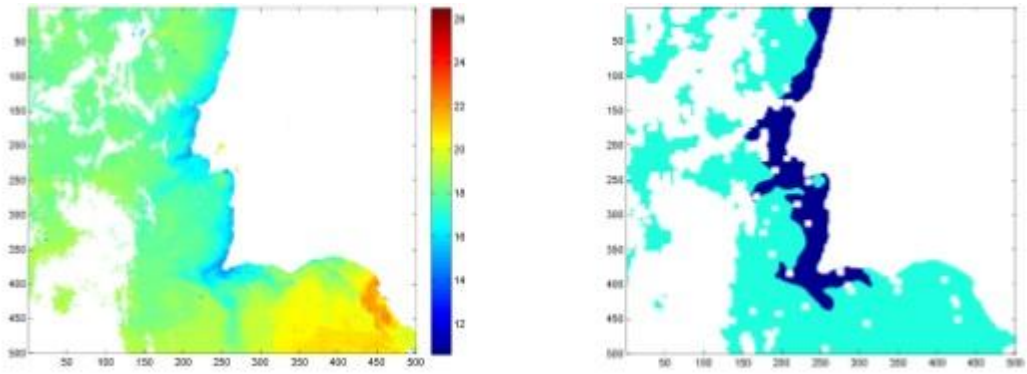


Figure A.4: SST image (1998-06-14) in the left and Ground Truth in the right.

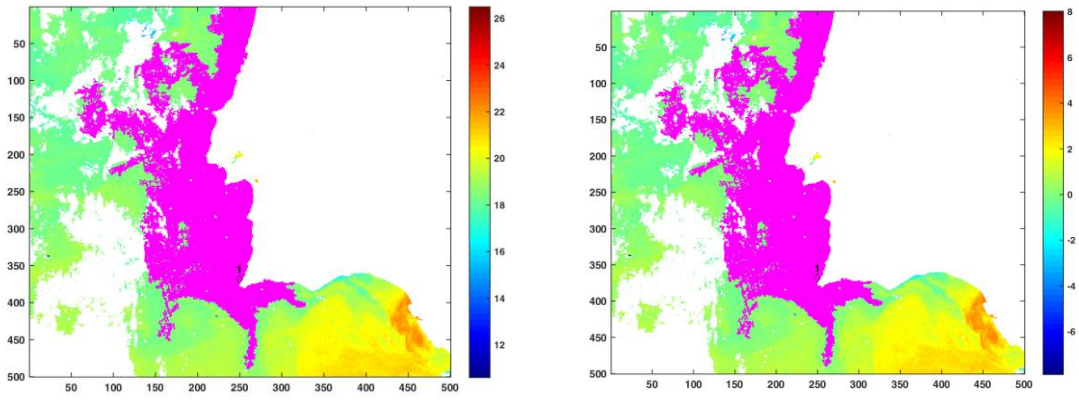


Figure A.5: Segmentation results, left image is the result of ISEC and the right one is the result of ISEC-V2

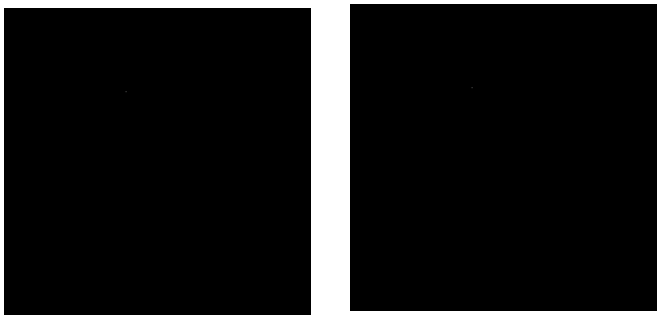


Figure A.6: Iterative result images by ISEC algorithm.

Figure A.7: Iterative result images by ISEC-V2 algorithm.

A.3 SEC vs. RSEC Results of F-measure Index

Table A.1: Images of 1998 and 1999 with result of f-measures, using SEC and the new revised version RSEC.

		SEC	RSEC		SEC	RSEC
S#	*image*	*f-value*	*f-value*	*image*	*f-value*	*f-value*
1	19980609.1813.n12.mat	**0.6285**	**0.6332**	19990602.1906.n15.mat	0.6911	0.6851
2	19980612.1528.n14.mat	**0.2569**	**0.2573**	19990608.2013.n15.mat	0.5139	0.5139
3	19980614.1803.n12.mat	**0.6478**	**0.6573**	19990610.1929.n15.mat	**0.3401**	**0.3411**
4	19980618.0656.n12.mat	**0.5962**	**0.6008**	19990614.1559.n14.mat	**0.4046**	**0.4064**
5	19980623.0341.n14.mat	**0.7901**	**0.7968**	19990619.1930.n15.mat	**0.3281**	**0.3289**
6	19980625.1444.n14.mat	0.6835	0.6811	19990620.1908.n15.mat	0.2847	0.2843

7	19980628.1755.n12.mat	**0.7961**	**0.7981**	19990627.1953.n15.mat	**0.3003**	**0.3014**
8	19980703.1456.n14.mat	**0.7928**	**0.7982**	19990630.1620.n14.mat	**0.4453**	**0.4494**
9	19980707.1757.n12.mat	**0.8281**	**0.8297**	19990706.1954.n15.mat	**0.9022**	**0.9091**
10	19980711.1808.n12.mat	**0.6298**	**0.6322**	19990708.0505.n14.mat	**0.9081**	**0.9087**
11	19980715.1424.n14.mat	**0.7811**	**0.7833**	19990714.1522.n14.mat	**0.6001**	**0.6011**
12	19980718.1855.n12.mat	**0.9552**	**0.9567**	19990715.1955.n15.mat	**0.6296**	**0.6316**
13	19980721.1748.n12.mat	**0.8861**	**0.8931**	19990719.2006.n15.mat	**0.8333**	**0.8351**
14	19980724.1425.n14.mat	**0.8947**	**0.9018**	19990721.1922.n15.mat	**0.7921**	**0.7961**
15	19980728.1521.n14.mat	0.8978	0.8973	19990729.1945.n15.mat	**0.6976**	**0.6988**
16	19980801.1846.n12.mat	**0.9724**	**0.9728**	19990731.1900.n15.mat	**0.8572**	**0.8625**
17	19980802.1426.n14.mat	**0.9549**	**0.9565**	19990801.1520.n14.mat	**0.8187**	**0.8258**
18	19980805.1533.n14.mat	**0.8849**	**0.8872**	19990810.1519.n14.mat	**0.7056**	**0.7078**
19	19980810.0313.n14.mat	**0.5839**	**0.5856**	19990814.1615.n14.mat	0.5464	0.5461
20	19980812.1806.n12.mat	**0.9391**	**0.9412**	19990817.1541.n14.mat	0.8224	0.8224
21	19980819.1438.n14.mat	**0.8988**	**0.9119**	19990821.1935.n15.mat	**0.8737**	**0.8768**
22	19980821.1416.n14.mat	**0.7202**	**0.7219**	19990823.1851.n15.mat	**0.8377**	**0.8397**
23	19980823.1535.n14.mat	**0.7302**	**0.7363**	19990826.1924.n15.mat	**0.7826**	**0.8004**
24	19980830.1807.n12.mat	**0.9353**	**0.9359**	19990830.1935.n15.mat	**0.8216**	**0.8243**
25	19980905.1451.n14.mat	0.7161	0.6964	19990901.1851.n15.mat	**0.9421**	**0.9423**
26	19980908.0434.n14.mat	**0.6982**	**0.7001**	19990908.1635.n14.mat	**0.7726**	**0.7741**
27	19980911.0400.n14.mat	**0.8319**	**0.8402**	19990910.0447.n14.mat	0.9341	0.9306
28	19980915.1441.n14.mat	**0.8905**	**0.8972**	19990914.1903.n15.mat	**0.9599**	**0.9604**
29	19980924.0317.n14.mat	**0.6244**	**0.6258**	19990928.1609.n14.mat	**0.4238**	**0.4297**
30	19980930.0703.n12.mat	**0.8252**	**0.8271**	19990930.1546.n14.mat	**0.6194**	**0.6621**
31				19991003.1720.n12.mat	**0.8015**	**0.8027**

A.4 Segmentation Results with SST images (Explosion)

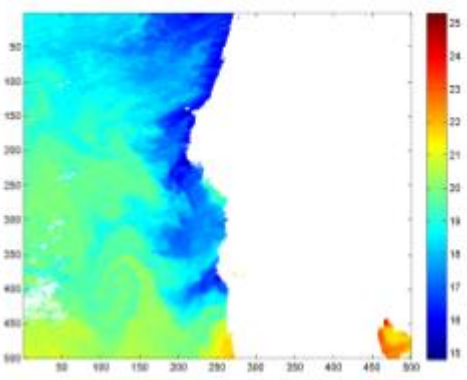


Figure A.8: SST Image 1998-07-15

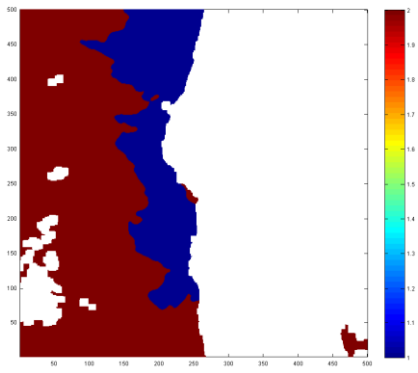
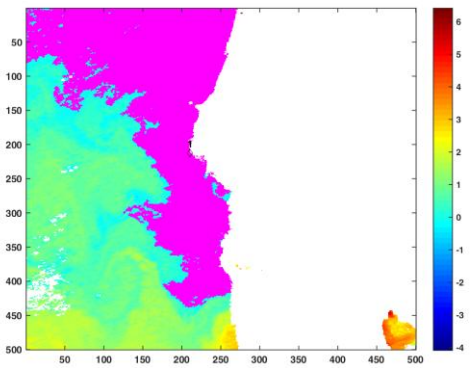


Figure A.9: GT 1998-07-15



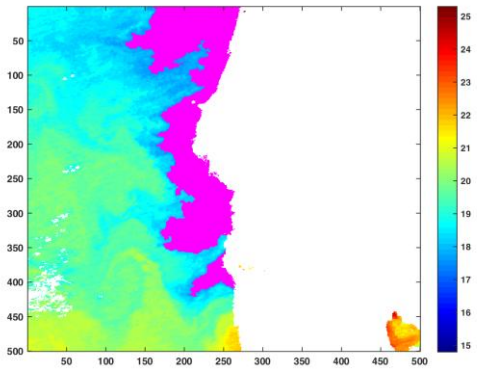


Figure A.10: ISEC-V2 1998-07-15

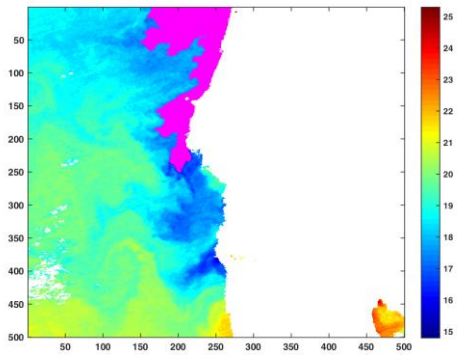


Figure A.11: I-Adams 1998-07-15

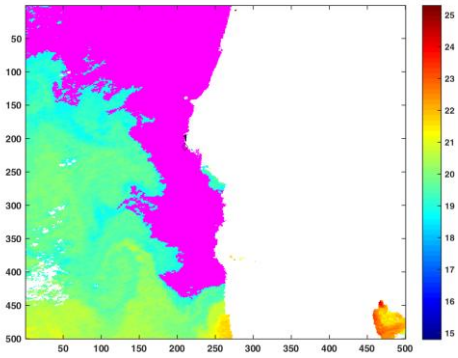


Figure A.12: Shih-SRG 1998-07-15

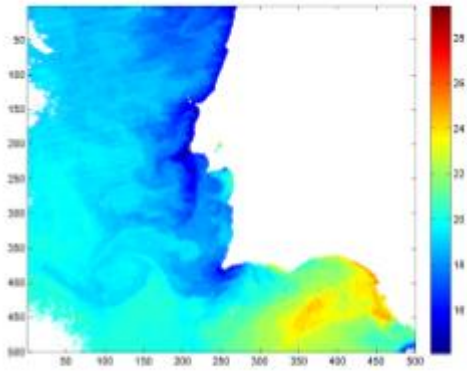


Figure A.13: Verma-Otsu 1998-07-15

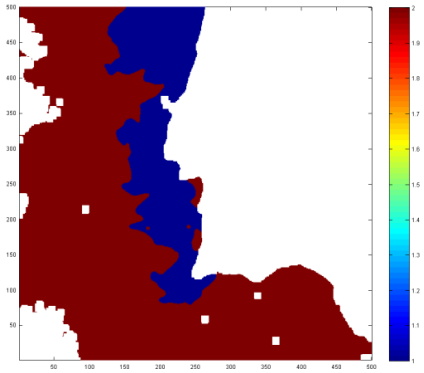


Figure A.14: SST image 1998-07-11

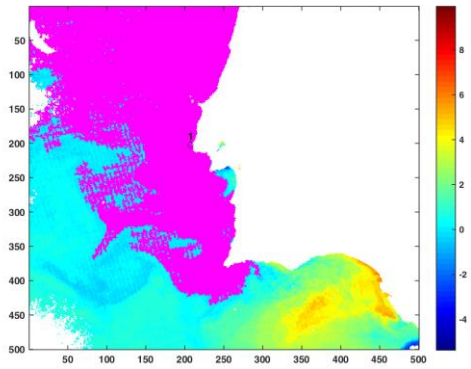


Figure A.15: GT 1998-07-11

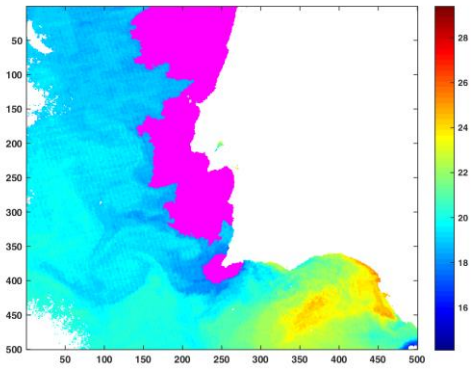


Figure A.16: ISEC-V2 1998-07-11

Figure A.17: I-Adams 1998-07-11

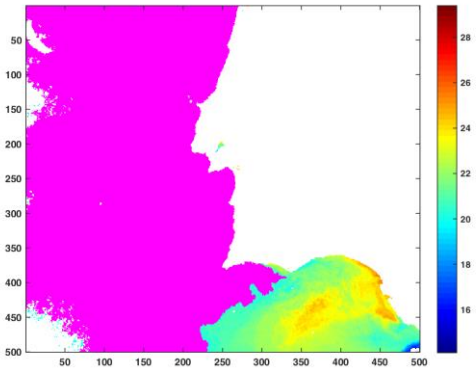


Figure A.18: Shih SRG 1998-07-11

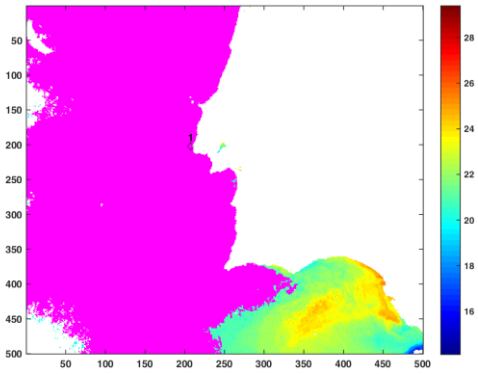


Figure A.19: Verma-Otsu 1998-07-11

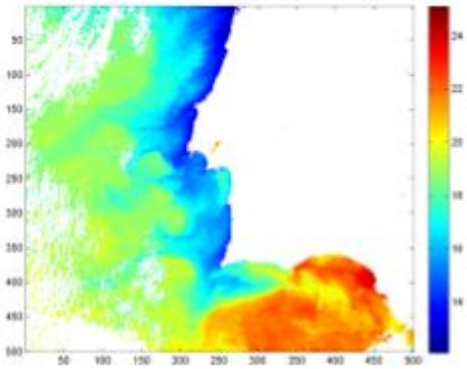


Figure A.20: SST image 2002-07-31

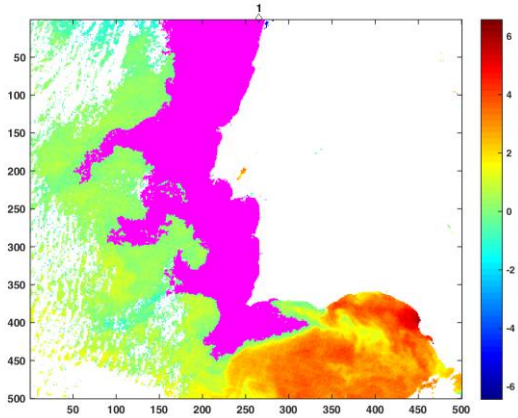


Figure A.21: ISEC-V2 2002-07-31

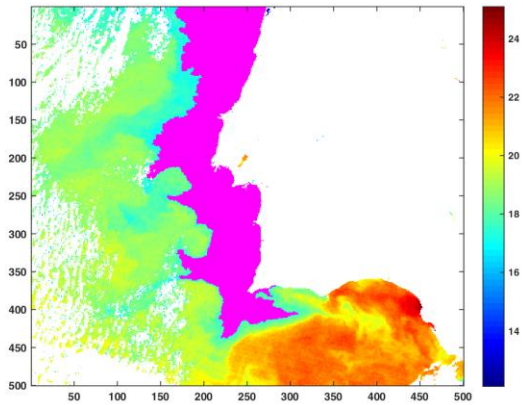


Figure A.22: I-Adams 2002-07-31

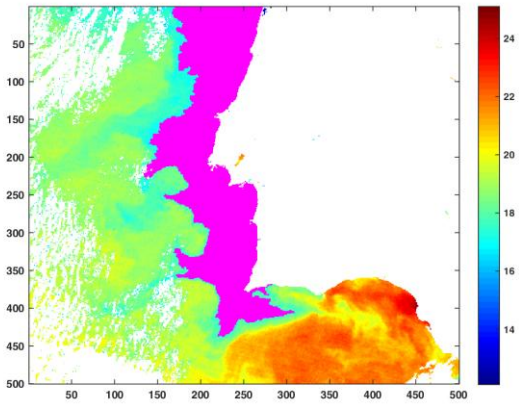


Figure A.23: Shih SRG 2002-07-31

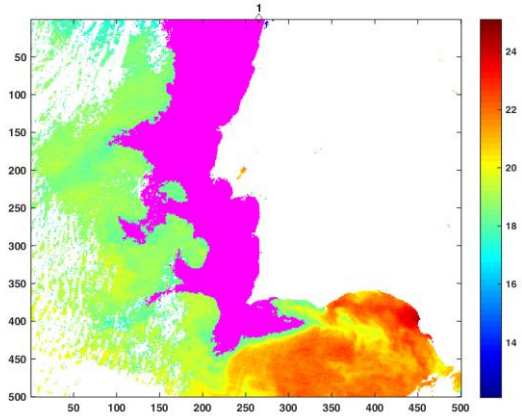


Figure A.24: Verma-Otsu 2002-07-31

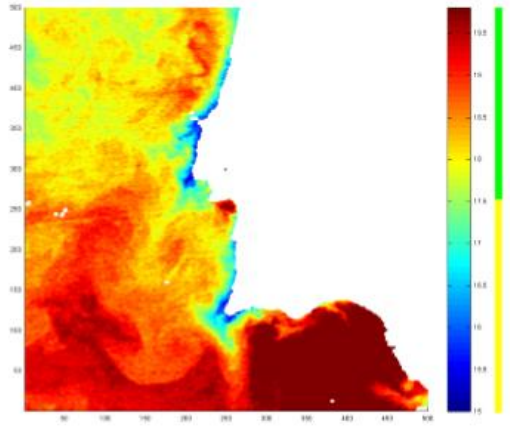


Figure A.25: SST image 1998-06-12

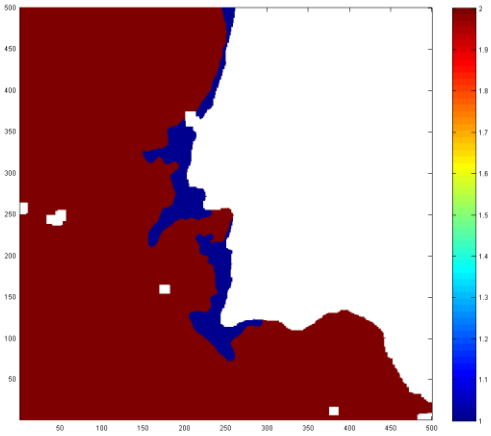


Figure A.26: GT 1998-06-12

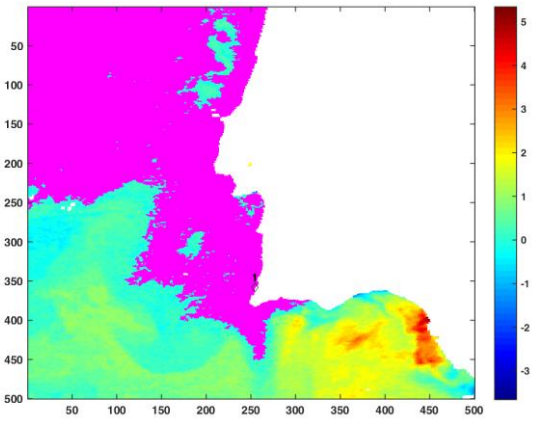


Figure A.27: ISEC-V2 1998-06-12

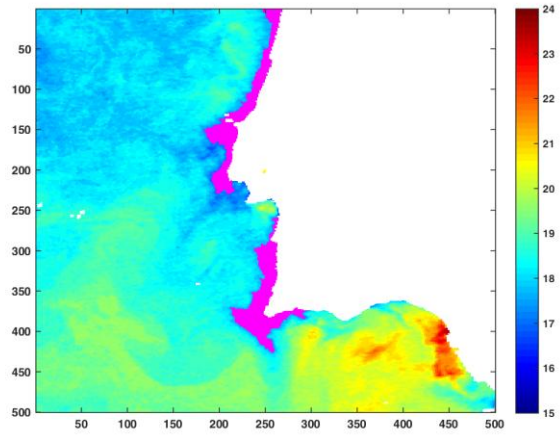


Figure A.28: I-Adams 1998-06-12

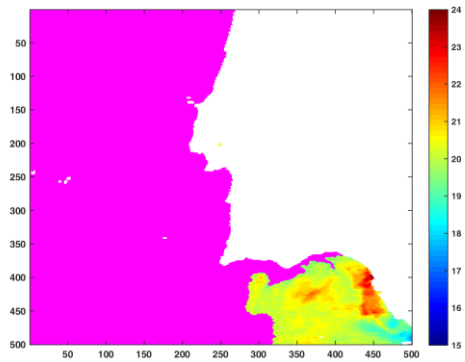


Figure A.29: Shih SRG 1998-06-12

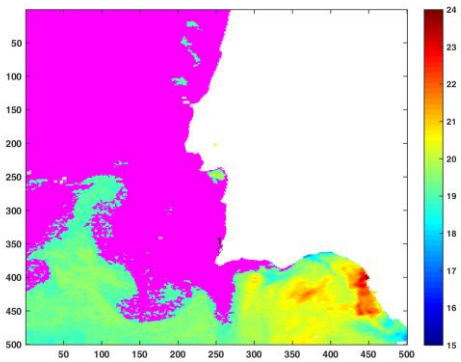


Figure A.30: Verma-Otsu 1998-06-12

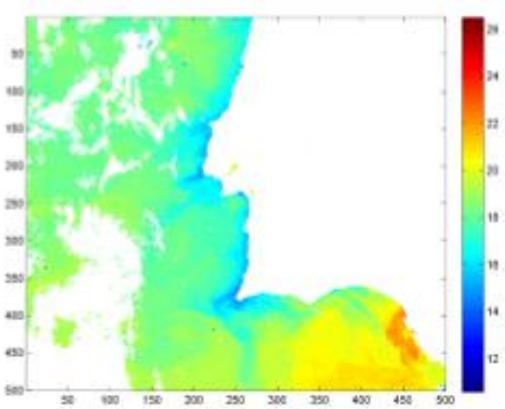


Figure A.31: SST image 1998-06-14

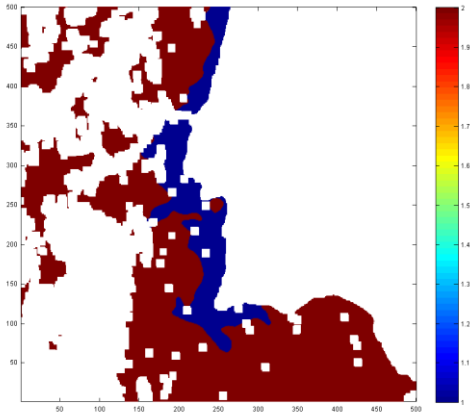


Figure A.32: GT 1998-06-14

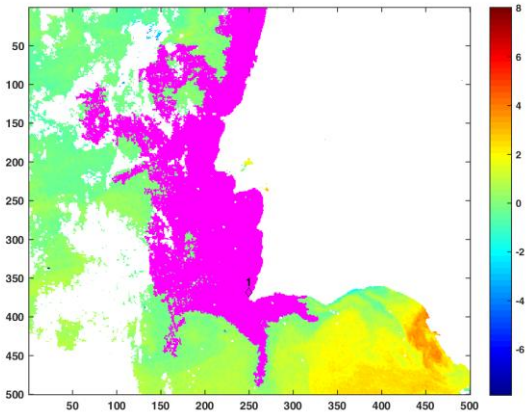


Figure A.33: ISEC-V2 1998-06-14

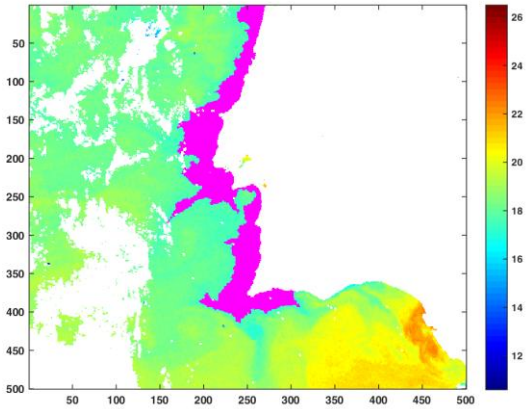


Figure A.34: I-Adams 1998-06-14

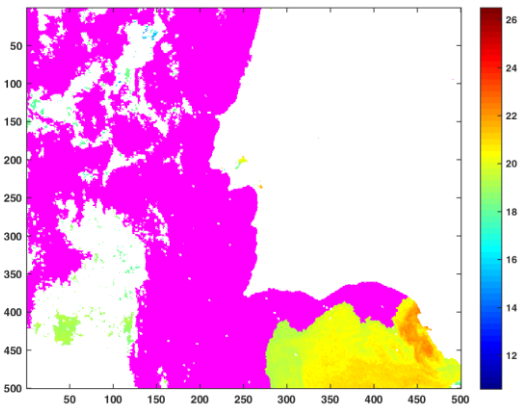


Figure A.35: Shih SRG 1998-06-14

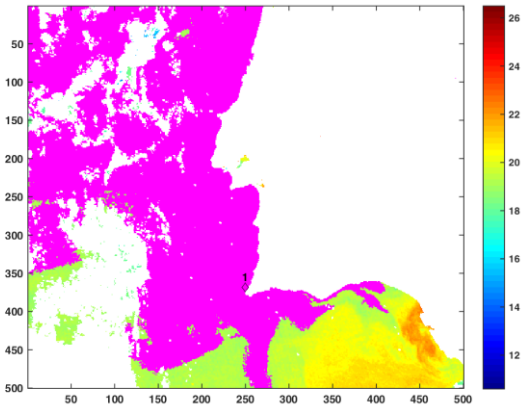


Figure A.36: Verma-Otsu 1998-06-14

A.5 Segmentation Results with SST images (No Explosion)

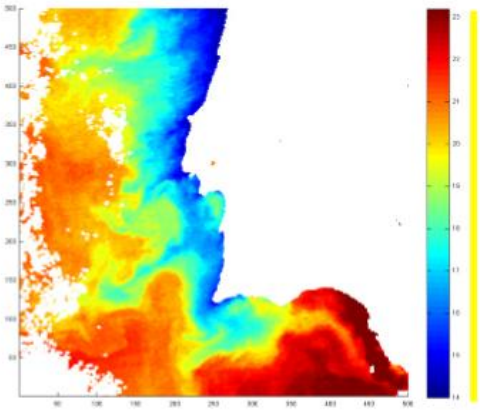


Figure A.37: SST image 1998-08-01

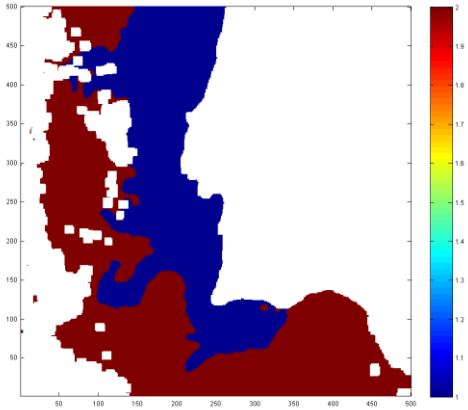


Figure A.38: GT 1998-08-01

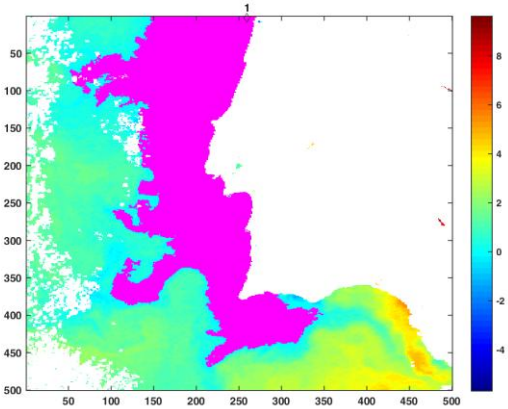


Figure A.39: ISEC-V2 1998-08-01

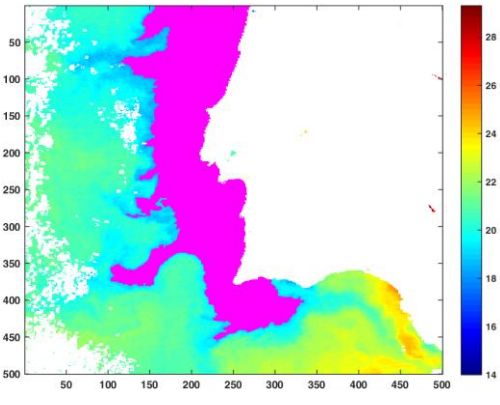


Figure A.40: I-Adams 1998-08-01

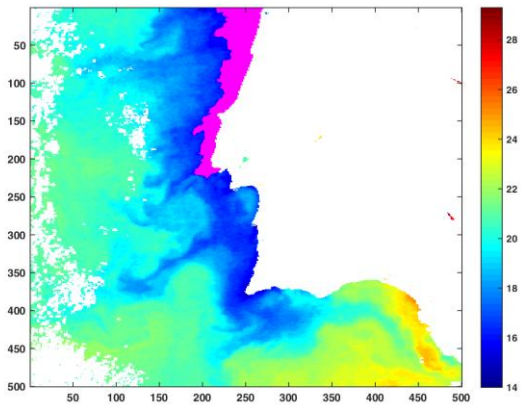


Figure A.41: Shih SRG 1998-08-01

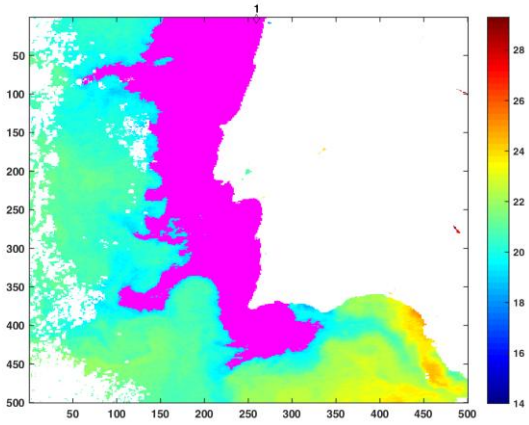


Figure A.42: Verma-Otsu 1998-08-01

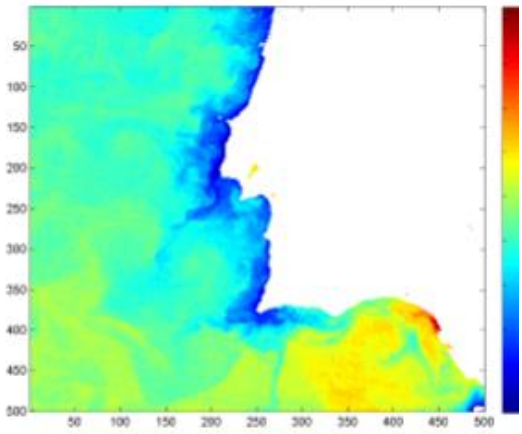


Figure A.43: SST image 1998-08-02

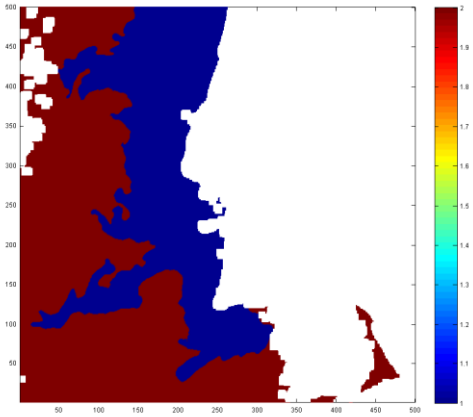


Figure A.44: GT 1998-08-02

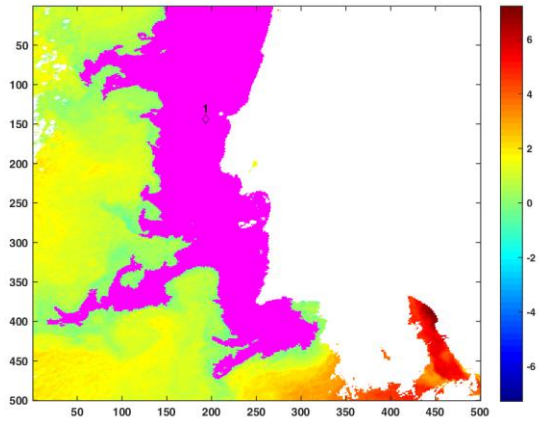


Figure A.45: ISEC-V2 1998-08-02

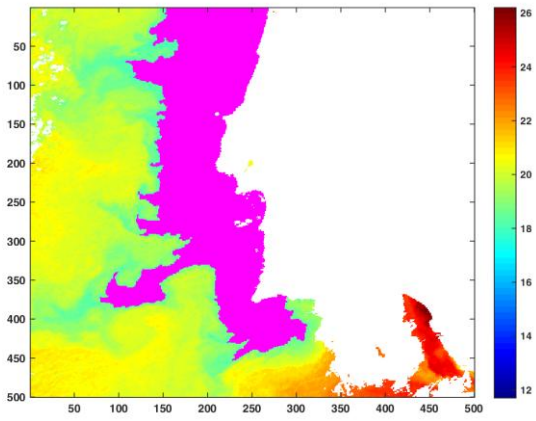


Figure A.46: I-Adams 1998-08-02

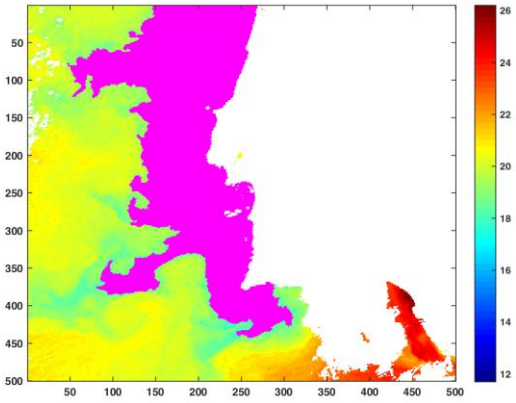


Figure A.47: Shih SRG 1998-08-02

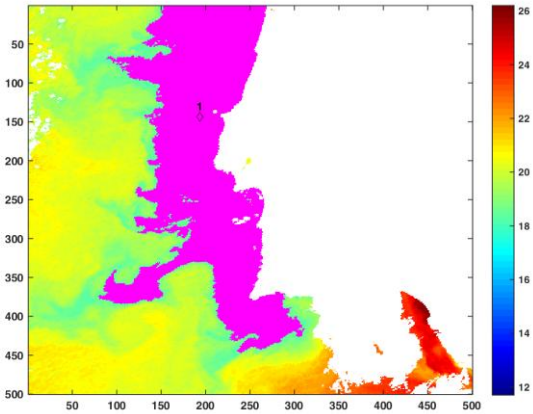


Figure A.48: Verma-Otsu 1998-08-02

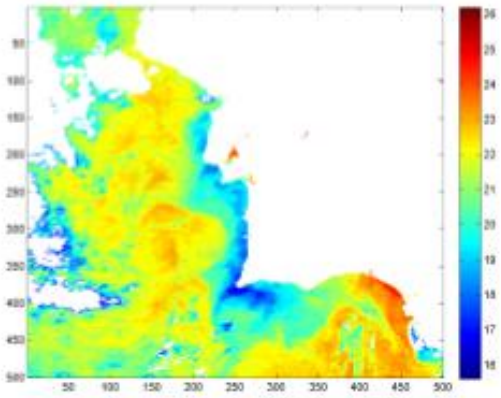


Figure A.49: SST image 1999-09-01

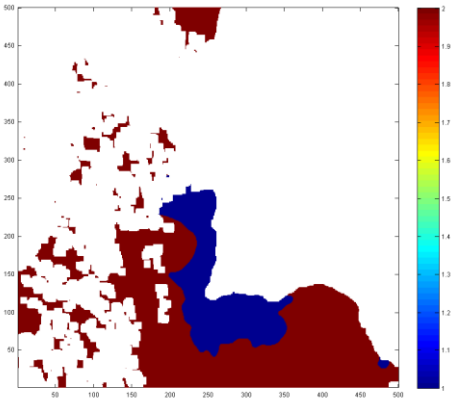


Figure A.50: GT 1999-09-01

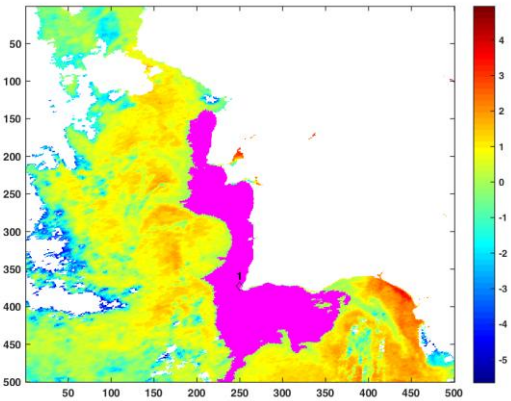


Figure A.51: ISEC-V2 1999-09-01

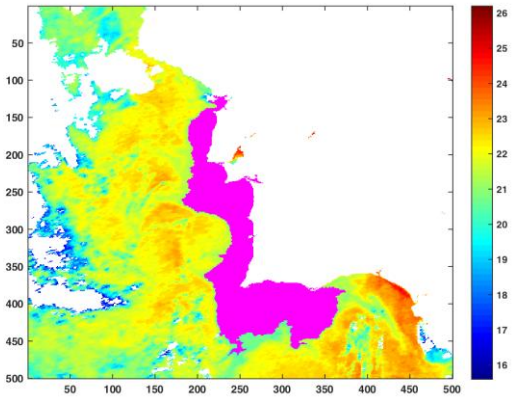


Figure A.52: I-Adams 1999-09-01

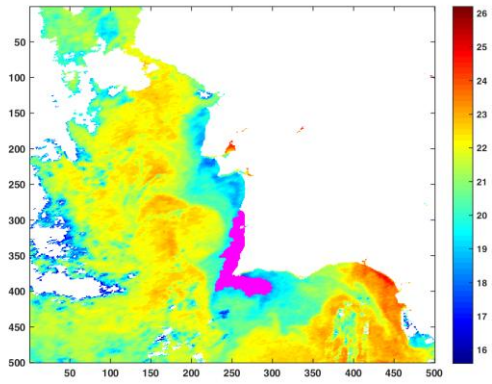


Figure A.53: Shih SRG 1999-09-01

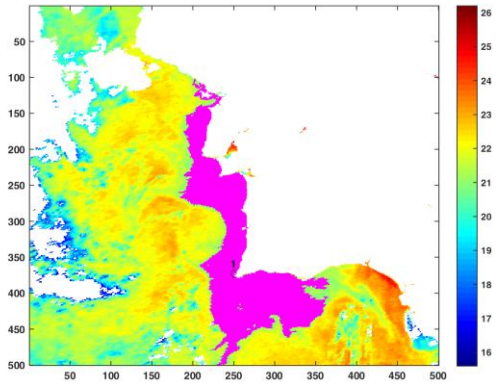


Figure A.54: Verma-Otsu 1999-09-01

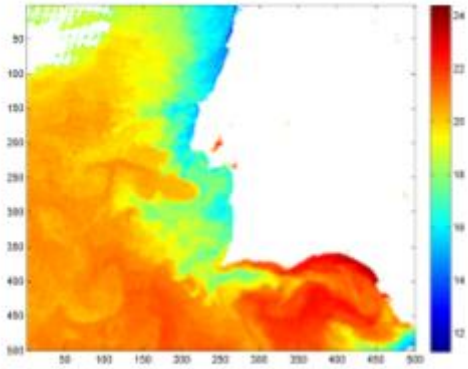


Figure A.55: SST images 2000-08-08

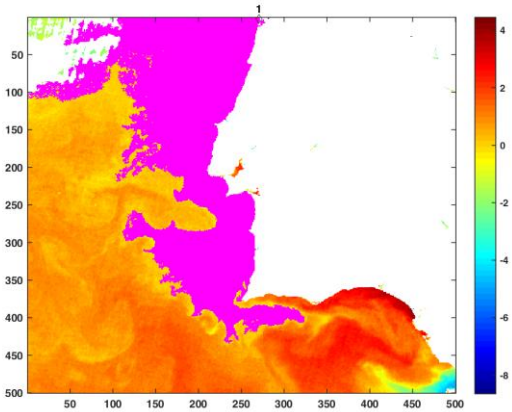


Figure A.56: ISEC-V2 2000-08-08

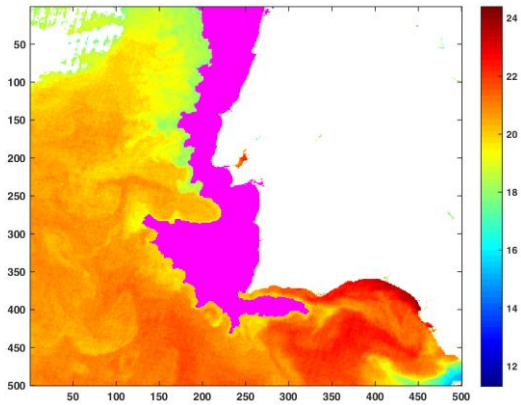


Figure A.57: I-Adams 2000-08-08

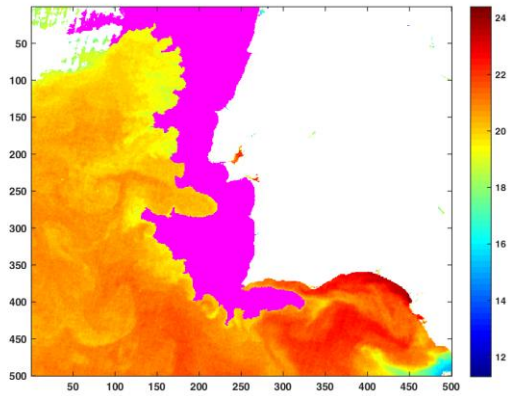


Figure A.58: Shih SRG 2000-08-08

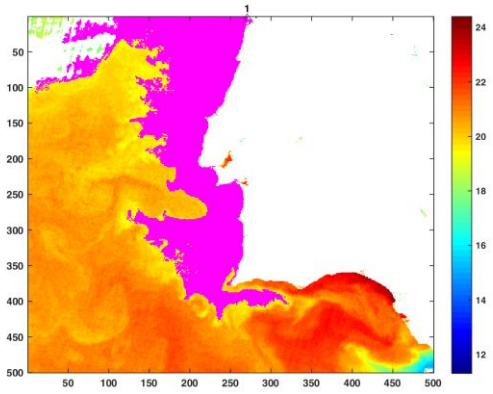


Figure A.59: Verma-Otsu 2000-08-08

A.6 Segmentation Results with SST images of Canary

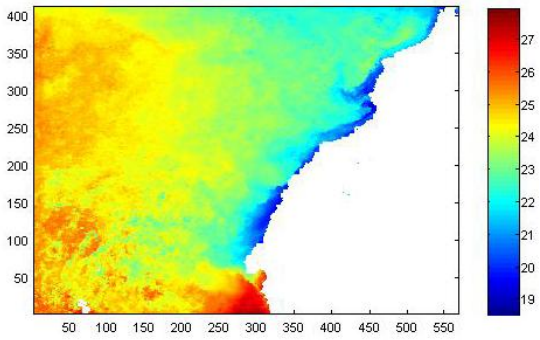


Figure A.60: SST image 177

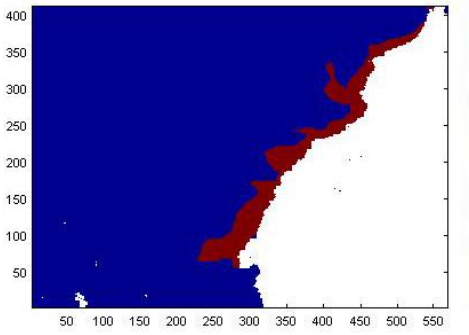


Figure A.61: GT 177

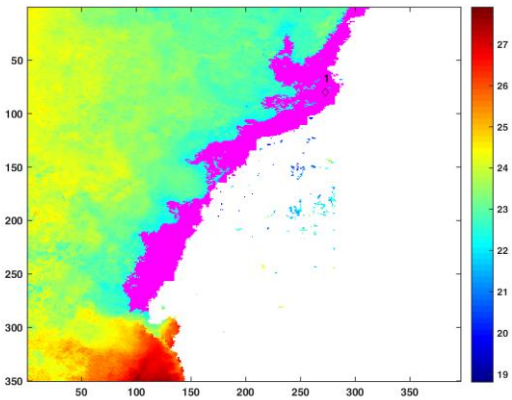


Figure A.62: ISEC-V2 image 177

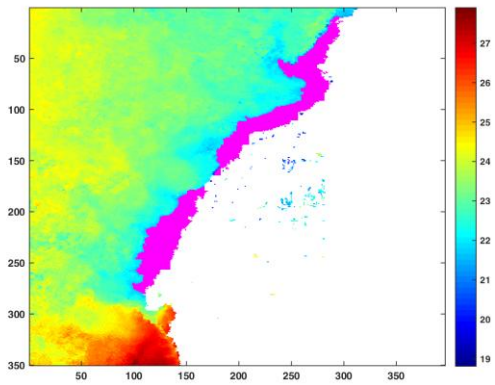


Figure A.63: I-Adams image 177

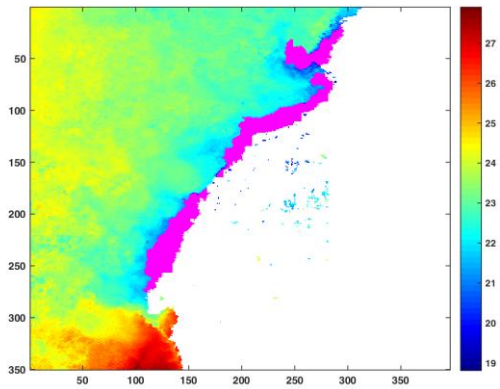


Figure A.64: Shih SRG image 177

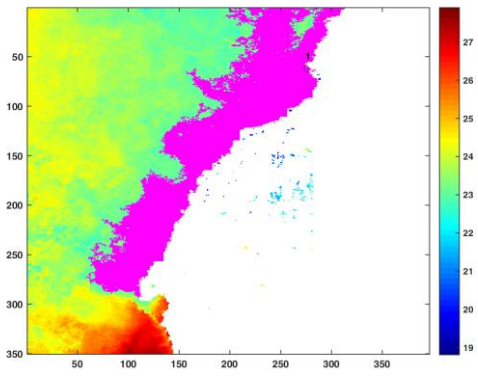


Figure A.65: Verma-Otsu image 177

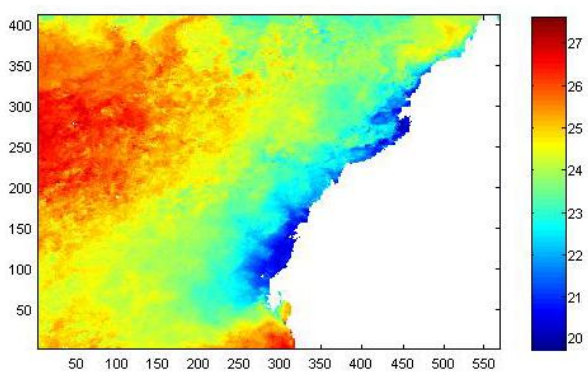


Figure A.66: SST image 117

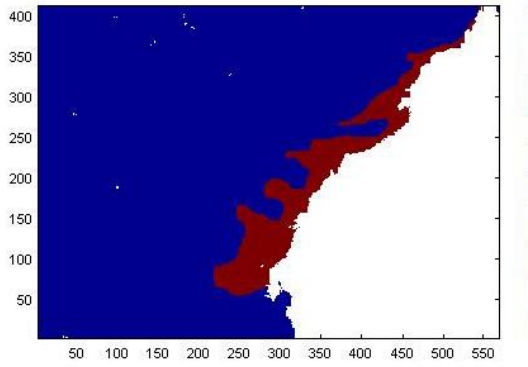


Figure A.67: GT image 117

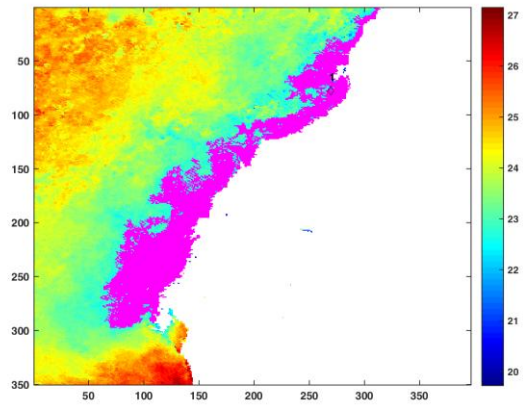


Figure A.68: ISEC-V2 image 117

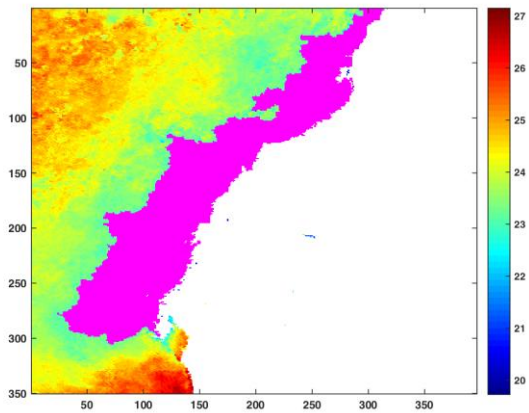


Figure A.69: I-Adams image 117

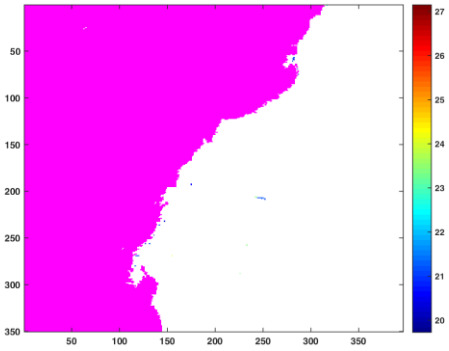


Figure A.70: Shih SRG image 117

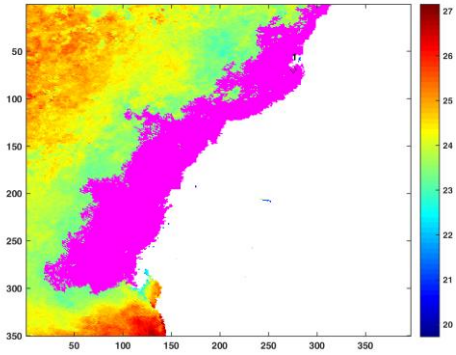


Figure A.71: Verma-Otsu image 117

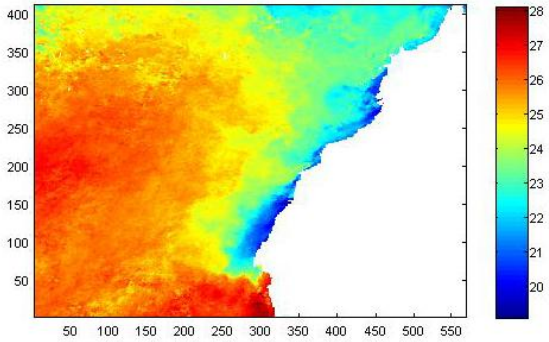


Figure A.72: SST image 237

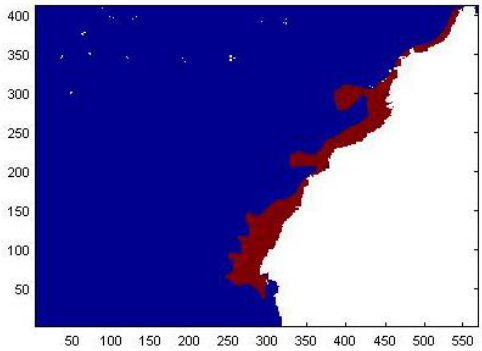


Figure A.73: GT image 237

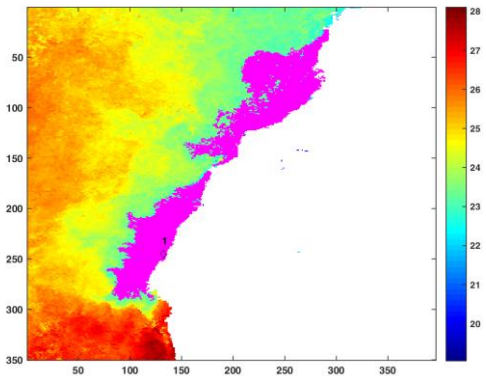


Figure A.74: ISEC-V2 image 237

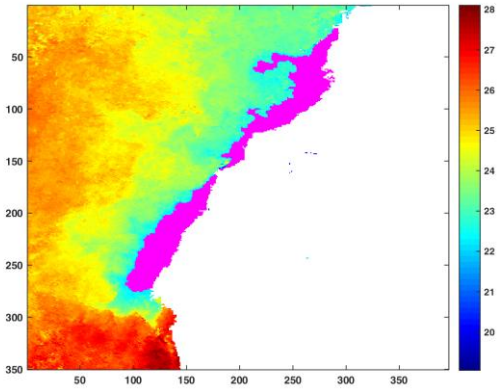


Figure A.75: I-Adams image 237

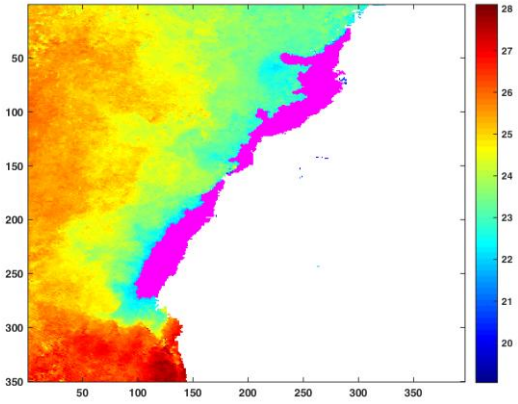


Figure A.76: Shih SRG image 237

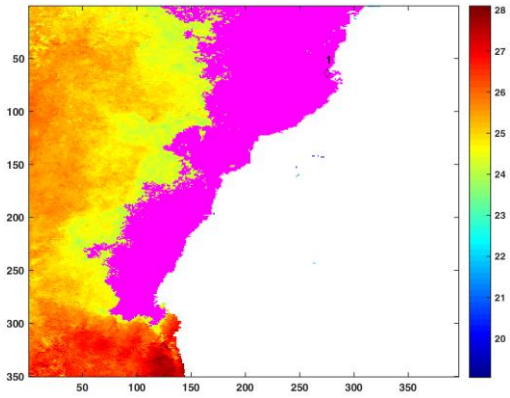


Figure A.77: Verma-Otsu image 237

A.7 Segmentation Results with Contour Strength (CS) and Derivative

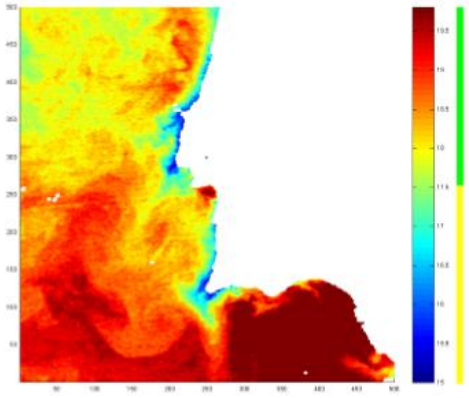


Figure A.78: SST image 1998-06-12

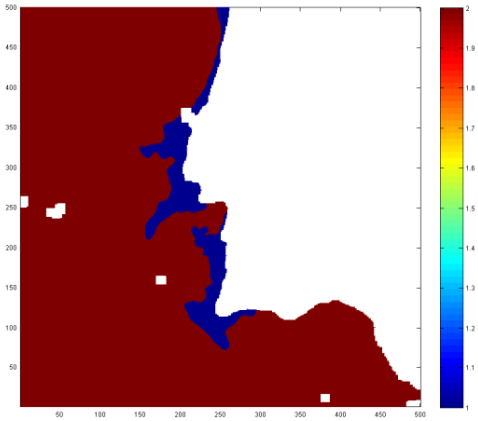


Figure A.79: GT 1998-06-12

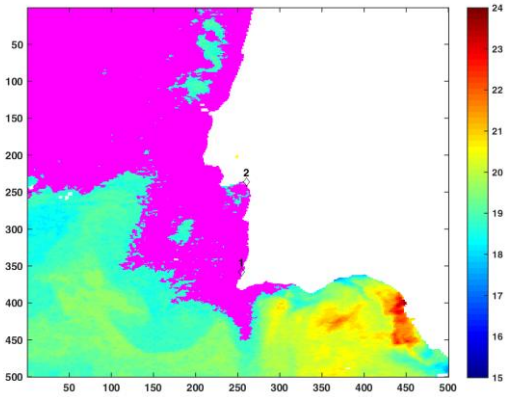


Figure A.80: ISEC-V2 1998-06-12

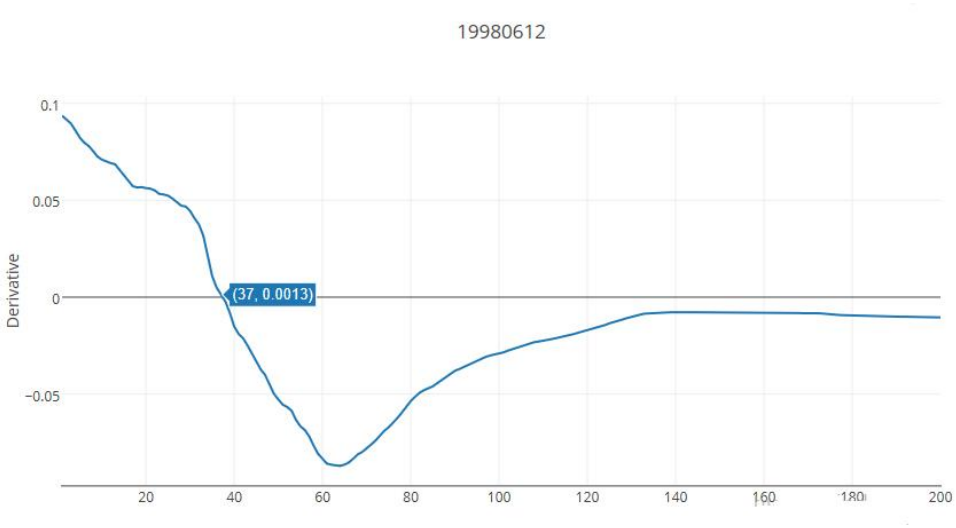


Figure A.81: Contour Strength (CS)'s first derivative, 1998-06-12

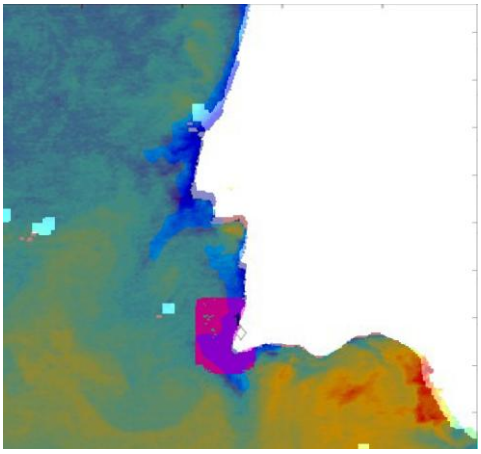


Figure A.82: GT and Segmentation, 1998-06-12

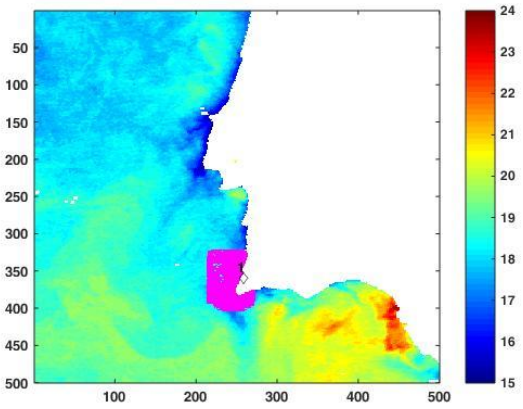


Figure A.83: Segmentation at it=37, 1998-06-12

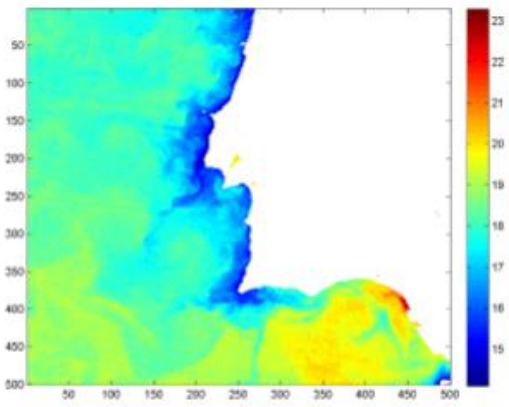


Figure A.84: SST image 1998-06-18

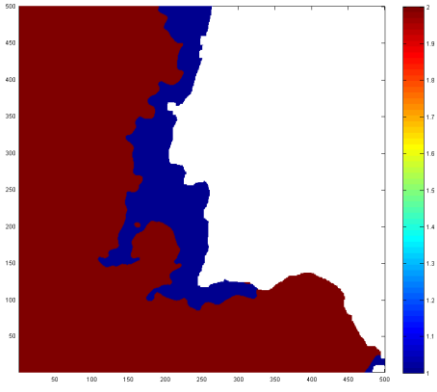


Figure A.85: GT 1998-06-18

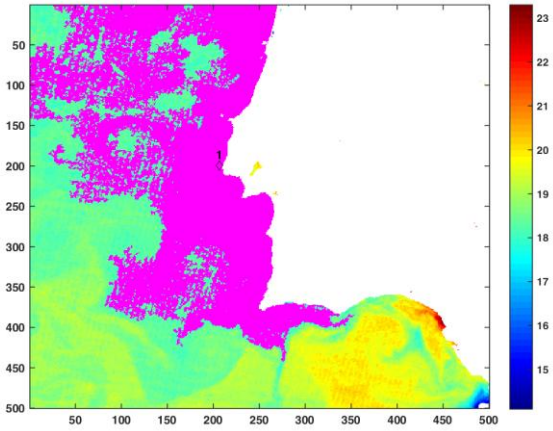


Figure A.86: ISEC-V2 1998-06-18

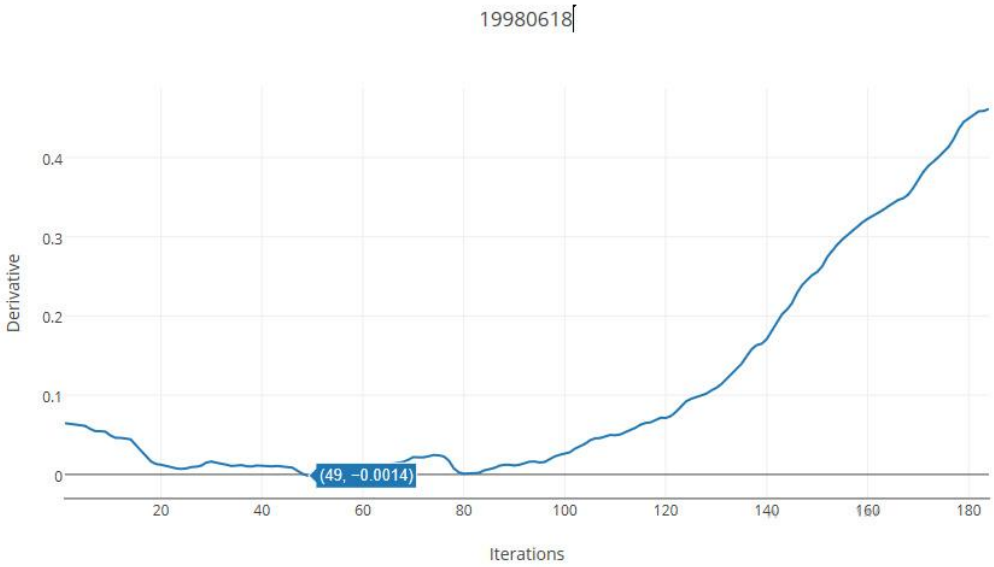


Figure A.87: Contour Strength (CS)'s first derivative, 1998-06-18

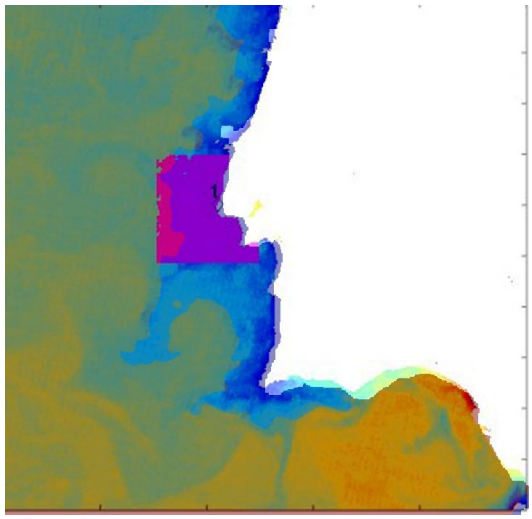


Figure A.88: GT and Segmentation, 1998-06-18

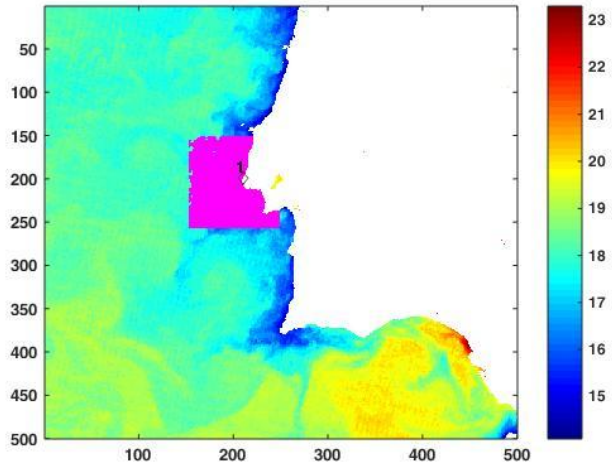


Figure A.89: Segmentation at it=49, 1998-06-18