JADAVPUR UNIVERSITY

An Automated System for Segmentation of Natural Scene

by

Soumik Sarkar

Department of Information Technology
Registration No. : 129046 of 2014-15
Examination Roll No. : M4SWE19027

A thesis submitted in partial fulfillment for the
degree of Master of Engineering in Software Engineering

under the guidance of

Prof. Sanjoy Kumar Saha
Department of Computer Science and Engineering

2019

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

Certificate of Recommendation

This is to certify that the thesis entitled "An Automated System for Segmentation of Natural Scene" has been carried out by Soumik Sarkar(University Registration No.: 129046 of 2014-15, Examination Roll No.: M4SWE19027) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Engineering in Software Engineering. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree in any other University or Institute.

....
Prof. Sanjoy Kumar Saha (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-32

Countersigned

....
Dr. Bhaskar Sardar
Head, Department of Information Technology
Jadavpur University, Kolkata-98

....
Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology
Jadavpur University, Kolkata-32

FACULTY OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

Certificate of Approval*

This is to certify that the thesis entitled "An Automated System for Segmentation of Natural Scene" is a bona fide record of work carried out by Soumik Sarkar in partial fulfillment of the requirements for the award of the degree of Master of Engineering in Software Engineering in the Department of Information Technology, Jadavpur University during the period of June 2018 to May 2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

....

Signature of Examiner 1

Date:

....

Signature of Examiner 2

Date:

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis entitled "An Automated System for Segmentation of Natural Scene" contains literature survey and original research work by the undersigned candidate, as part of his Degree of Master of Engineering in Software Engineering.All information have been obtained and presented in accordance with academic rules and ethical conduct.I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name : Soumik Sarkar
Reg.No. : 129046 of 2014-15
Exam Roll No. : M4SWE19027
Thesis Title : An Automated System for Segmentation of Natural Scene

. . . .
Soumik Sarkar
May 2019

JADAVPUR UNIVERSITY

Abstract

Prof. Sanjoy Kumar Saha
Department of Computer Science and Engineering

Master of Engineering in Software Engineering

by Soumik Sarkar

Department of Information Technology
Registration No. : 129046 of 2014-15
Examination Roll No. : M4SWE19027

Segmentation is one of the most important aspects of computer vision that is tasked with dividing an image into regions that are semantically different from one another with respect to some chosen context. The context can be purely mathematical in nature or based on human understanding. Semantic segmentation describes the process of associating each pixel of an image with a class label.The following thesis investigates about existing technologies relating to segmentation in a natural scene images briefly before proceeding towards the development of a segmentation method. Here we try to develop a method for natural image segmentation with a supervised statistical approach. It first uses the concept of superpixel segmentation to divide the image into regions which have little intrapixel variance. Then a neighbourhood graph is created for further processing which allows testing of neighbouring regions for closeness based on their texture, hue and color. The amount of closeness is supervised as a confidence level. In a further accurate but computationally heavy version, closeness of regions are tested by various distance functions, and then regions are merged if allowed by consensus. The novelty of the approach is its middle ground as it allows more control and diversity like statistical approaches but with partial supervision can go near accuracy levels of deep learning approaches without the model bias. The method is also compared with existing popular methods of object segmentation and stress was laid on the accuracy of the method rather than the speed....

Acknowledgements

I would like to take this opportunity to thank the people who taught me, guided me and supported me throughout the duration of the entire thesis work.

First and foremost, I would like to express my sincere gratitude towards my thesis supervisor, Prof. Sanjoy Kumar Saha of Computer Science and Engineering department at Jadavpur University for his guidance, encouragement and his confidence towards my work for this thesis.

Then I would like to thank all the teachers of our Information Technology Department, specially Prof. Bibhas Chandra Dhara, Dr. Kartick Chandra Mondal for their continuous support. I am also grateful to Dr. Bhaskar Sardar, Head of the department, Information Technology, Jadavpur University.

I feel very lucky to have a mentor like Mr. Aritra Mukherjee who not only guided me in my research work but also assisted me whenever I needed any help. I have been learning so many things from him since I met him.

I would also like to thank all the members of Neptune Lab under the Mechanical Department for encouraging me throughout the duration. I would also like to thank Ms. Aindrila Sur, Mr. Mausam Kundu, and Mr. Mohit Mahata for their support, encouragement, and the amount of believe have in me.

Lastly, I would like to thank my parents and elder brother for their constant support and love. This accomplishment would not have been possible without them. Thank you.

. . . .

SOUMIK SARKAR
Registration No.: 129046 of 2014-15
Examination Roll No.: M4SWE19027
Department of Information Technology
Jadavpur University

Contents

List of Figures

Dedicated to my parents...

Chapter 1

Introduction

Computer vision is an interdisciplinary field that deals with how computers can be made to gain a high-level understanding from digital images or videos. It has a dual goal. From the biological science point of view, computer vision aims to come up with computational models of the human visual system. From the engineering point of view, computer vision aims to build autonomous systems which could perform some of the tasks which the human visual system can perform."Computer vision is concerned with the automatic extraction, analysis, and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding."[8][9][10]. Lawrence Roberts also known as the father of Computer Vision, in 1963 he published Machine Perception Of Three-Dimensional Solids[1] in his Ph.D. thesis which started it all for us. There he discussed extracting 3D information about solid objects from 2D photographs of line drawings. He mentioned things such as camera transformations(Figure 1.1), perspective effects, and the rules and assumptions of depth perception things that we discuss to this very day. At that time, AI became an academic discipline and millions were pumped into research with the intention of developing a machine as intelligent as a human being. One of the co-founders of the Artificial Intelligence Lab at MIT, Marvin Minsky in 1966 gave an assignment to his student, link a camera to a computer and let the computer describe what it sees. After this incident, computer vision gradually founds its place in this ambitious time of AI as well. People then thought that constructing a machine to mimic the human visual system was going to be an easy task on the road to finally building a robot with human-like intelligent behavior. In the same year, Seymour Papert organized The Summer Vision Project[11] at the MIT, the main idea of this project was background/foreground segmentation of real-world images with a final goal of extracting non-overlapping objects from them. After this incident researchers approach towards computer vision had changed so much.From then new innovations at this field takes it

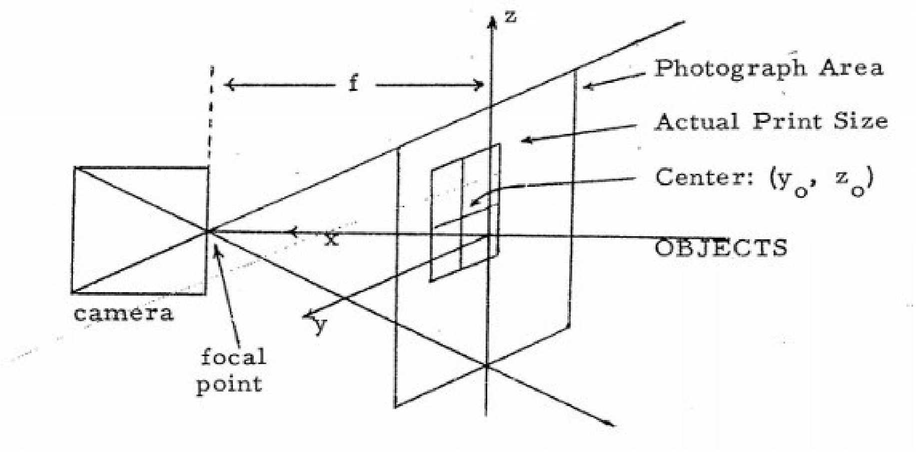

FIGURE 1.1: Camera Transformation, [1]

to an another level. Most of the computer vision algorithms those exist today, such as extraction of edges from images, labeling of lines, non-polyhedral and polyhedral modeling, representation of objects as interconnections of smaller structures, optical flow, and motion estimation had created the platform for computer vision.[2]

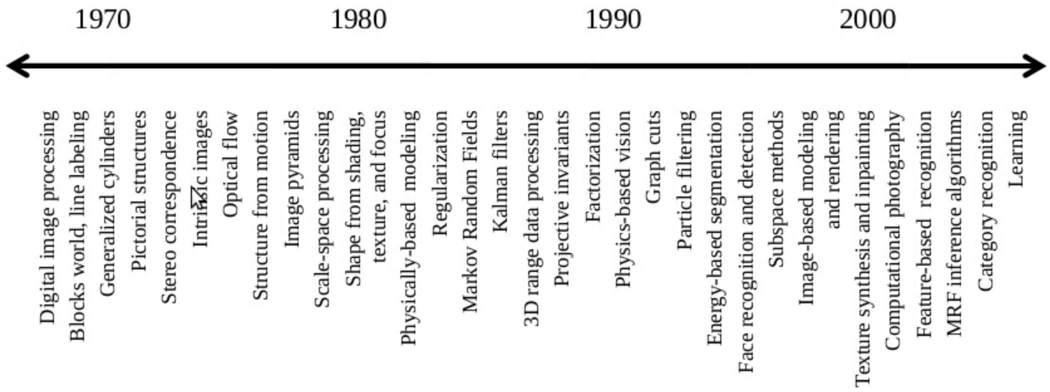

FIGURE 1.2: A rough timeline of some of the most active topics of research in computer vision., [2]

In the next decade computer vision technology got its first commercial applications through Optical Character Recognition (OCR) technology, the purpose of that application was to help blind people to read via OCR computer programs. From the next decade, researchers gave more attention towards sophisticated mathematical techniques for performing quantitative image and scene analysis. These include some of the famous research topics such as the concept of scale-space, the inference of shape from various cues such as shading, texture, and focus, and contour models known as snakes. In the first phase of 1990 research in projective 3-D reconstructions led to a better understanding of camera calibration. Progress was made on the dense stereo correspondence problem and further multi-view stereo techniques. At the same time, different types of graph cut were used to solve image segmentation. This decade also marked the first time statistical learning techniques were used in practice to recognize faces in images using the concept of Eigenface [12]. At the end of the 1990s, a significant change came about with the increased interaction between the fields of computer graphics and computer vision. This included image-based rendering, image morphing, view interpolation, panoramic image stitching, and early light-field rendering. In the next section, a brief overview of image segmentation was discussed, which is our area of interest.

1.1 Image Segmentation

In Computer Vision, Image Segmentation is one of the oldest and most studied problems.The simplest definition of image segmentation could be the process of dividing an image into segments, but some questions arise from this definition. What properties should these segments have? Should they be homogeneous in color? That is why image segmentation is often referred to as an ill-posed problem. However, this has not prevented researchers to keep their research on this topic and providing more and more advanced techniques of image segmentation. In simple word, it is a process of partitioning an image into regions or segments, which are semantically different from one another. Each segment corresponds to one object or some part of that object. Every pixel in an image belongs to one segments. The main aim of segmentation is to understand the scene of the image and reduce the complexity of the scene so that it is easily understandable to the observer and it helps in further processing. When we discuss image segmentation one term come in our mind, semantic segmentation. So what is this semantic segmentation? It attempts to partition the image into semantically meaningful parts and to classify each part into one of the predetermined classes. We can also classify each pixel. In that case, we have to do a pixel-wise classification, which leads to the same end result.

1.2 Types of Image Segmentation

There are several techniques available for image segmentation. In this part we will discuss some the techniques which are mostly used for segmentation.

- Thresold method
- Edge based method
- Region based method

- Clustering based method
- Watershed based method
- PDE based method
- ANN based method

1.2.1 Thresold based method

Thresholding methods are the simplest methods for image segmentation. These methods divide the image pixels with respect to their intensity level. These methods are used over images having lighter objects than background. The selection of these methods can be manual or automatic. There are basically three types of thresholding [13] [14].

- **Global thresolding method**

  Here we first assume a global threshold value T which will be constant for whole image. On the basis of T the output image $q(x, y)$ can be obtained from original image $p(x, y)$ as:

$$q(x,y) = \begin{cases} 1, & \text{if } p(x,y) > T \\ 0, & \text{if } p(x,y) \leq T \end{cases} \tag{1.1}$$

- **Variable thresolding method**

  In this type of thresholding, the value of T can vary over the image. It can further be divided into two catagory. First one is **Local thresold** where value of T depends upon the neighborhood of x and y. Another one is **Adaptive Threshold** where value of T is a function of x and y.

- **Multiple thresolding method**

  In this type of thresholding, there is more than one threshold value of an image. Suppose T0 and T1, so we have 3 labels say m, n, o. On the basis of following equation the output image can be computed:

$$q(x,y) = \begin{cases} m, & \text{if } p(x,y) \leq T0 \\ n, & \text{if } p(x,y) \leq T1 \\ o, & \text{if } p(x,y) > T1 \end{cases} \tag{1.2}$$

1.2.2 Edge based method

The edge detection is a well known technique of image segmentation. These methods try to find pixels location where the image brightness changes sharply or, more formally, has discontinuities because a single intensity value does not provide good information about edges. It first locate the edges where either the first derivative of intensity is greater than a particular threshold or the second derivative has zero crossings. In edge

FIGURE 1.3: An example of Edge detection(image courtesy:wikipedia)

based segmentation methods, first of all the edges are detected and then are connected together to form the object boundaries to segment the required regions. The basic two edge based segmentation methods are: Gray histograms and Gradient based methods. These are the structural techniques based on discontinuity detection [15][16].

1.2.3 Region based method

This type of segmentation methods divide the whole image into multiple segments based on some characteristic. There are two basic techniques based on this method [17] [18].

- **Region growing methods**: These methods segment the image into various regions based on the growing of seeds (initial pixels). These seeds can be selected manually (based on prior knowledge) or automatically (based on particular application). The growing of seeds is controlled by connectivity between pixels and with the help of the prior knowledge of problem this can be stopped.
- **Region splitting and merging methods**: The region splitting and merging based segmentation methods follow splitting and merging techniques for segmentation of an image. Splitting technique iteratively divides an image into regions having similar characteristics and then merging technique to combines the adjacent similar regions[19].

1.2.4 Clustering based method

The clustering based techniques segment the image into clusters having pixels with similar characteristics. Data clustering is the method that divides the pixels into clusters such that pixels in same cluster are more similar to each other than others. There are two basic categories of clustering methods: Hierarchical method and Partition based method. The hierarchical method works based on the concept of trees where root of the tree represents the whole database and the internal nodes represent the clusters. On the other side the partition based methods use optimization methods iteratively to minimize an objective function. Based on these two methods there are various algorithms to find clusters. [20][21]

1.2.5 Watershed based method

Any grayscale image can be viewed as a topographic surface where high intensity denotes peaks while low intensity denotes valleys. Watershed method starts filling every isolated valleys (local minima) with different colored water(labels). As the water level rises, depending on the peaks (gradients) nearby, water from different valleys with different colors will start to merge and then it creates barriers where water merges. This process filling water and building barriers will going on until all the peaks are under water. After that the barriers gives the final segmentation result[22].

1.2.6 PDE based method

PDE (Partial Differential Equations) equations or PDE models are used widely in image segmentation. They uses active contour model for segmentation purpose. Active Contour model or Snakes transform the segmentation problem into PDE. Some famous methods of PDE used for image segmentation are Snakes, Level-Set, and Mumford shah method [23].

1.2.7 ANN based method

Artificial neural network based segmentation methods simulate the learning strategies of human brain for the purpose of decision making. In ANN, every neuron is corresponding to the pixel of an image. First image is mapped to the neural network then it is trained using training samples then only it finds the connection between neurons, i.e., pixels. Then the new images are segmented from the trained image[24]. Some of the mostly used neural networks for image segmentation are Hopfield, BPNN, FFNN, MLFF, MLP,SOM, and PCNN etc. Segmentation of image using neural network mainly has two steps, one is pixel classification and another one is edge detection[25].

1.3 Importance of Image Segmentation and its Applications

Segmentation is the most important step for analyzing an image. Segmentation bridges the gap between low-level image processing and high-level image processing. There are many kinds of segmentation techniques available which will help to extract objects and detailed information about the objects in the images. The success or failure of any image application depends on the type of approach used for that application. However, there is no segmentation technique available which gives good result for a different type of images. Some of the applications of image segmentation are following.

1. **In medical imaging**
   (a) Locate tumours and other pathologies
   (b) Diagnosis, study of anatomical structure
   (c) Measure tissue volumes
2. **In Object detection**
   (a) Pedestrian detection
   (b) Face detection
   (c) Locate objects in satellite images (roads, forests, crops, etc.)
3. **Machine Vision**
4. **Robot Vision**
5. **In Traffic control systems**
6. **Video surveillance**
7. **Recognition Tasks**
   (a) Face recognition
   (b) Fingerprint recognition
   (c) Iris recognition

These are so many applications available in market where segmentation is required. In spite of these, a lot of application is present in the market which uses segmentation techniques to get their results.

1.4 Scope of the thesis

The following thesis is about investigation of different existent techniques of image segmentation of natural color images and its related works and to develop a comparatively simple technique to do the same. In previous study section a brief survey of related works are discussed. Then in the implementation section a partially novel technique is developed and discussed which is further explained by software implementation and analysis.

The goal of the thesis is to develop an automated system which will take an image as input and it gives a colored segmented image of the input image as a output. The algorithm which is developed here is giving quite satisfactory results in a semi controlled environment, but the processing speed is not fast enough for real-time applications.

Chapter 2

Related Works

Already a lot of work has done on image segmentation. There are many approaches to image segmentation. In this portion, we will briefly discuss previous works that are done by eminent researchers.

2.1 Segmentation based on Histogram Thresholding

In the early days of Computer Vision histogram thresholding method was a good choice for image segmentation. Many researchers used this approach in the field of segmentation. In 1990 Y.W Lim et al.[26] worked on segmentation and proposed a method for color images. The method works based on thresholding and fuzzy c-means(FCM) [27] technique. The proposed methodology uses a coarse-fine concept to reduce the computational complexity of FCM. The coarse segmentation attempts to segment coarsely using the thresholding technique, while the fine segmentation assigns the unclassified pixels a class having membership value maximum. M.Cheriet et al.[28] proposed a method which uses the concept of Otsus method[29]. They repetitively applied the Otsus method on different regions of the image until no new peaks can found. In 2006, Arifin et al. proposed a segmentation method[30] based on the concept of histogram thresholding where they calculated the threshold value using hierarchical cluster analysis. Initially, they considered ever y non-empty gray level as a separate cluster then calculate the distance between every pair of adjacent clusters. The distance indicates the dissimilarity of the adjacent clusters. The pair of clusters having minimum distance are merged. This process will continue (K-t) times where K is a number of cluster and t is the level of thresholding. S Sarkar et al.[31] proposed a multi-level image segmentation method by using the 2D Tsallis entropy to incorporate advantages of both 2D histogram and multi-level thresholding. But a limitation of histogram thresholding based approach is the image should have taken in a good lighting condition with minimal glare, i.e for segmentation of complex images this approach is not a good choice.

2.2 Segmentation based on Graph

In 2000 J Shi et al.[32] proposed a novel approach to solve the perceptual grouping problem of vision. Instead of focusing on local features, they extracted the global impression of an image. They treat image segmentation as a graph partitioning problem and propose a global criterion, the normalized cut, for segmenting the graph. Their approach was very similar like grouping in graph theory where a set of points in arbitrary feature space are represented as a weighted undirected graph $G(V, E)$, where the nodes of the graph are the points in feature space. The weight of each edge is a function of the similarity between nodes, a normalized cut criterion which measures both total dissimilarities between the different groups as well as total similarities within the groups. They applied this approach to segmenting static images. Felzenszwalb et al.[4] address the problem of segmenting an image into regions. They defined a predicate for measuring the evidence for a boundary between two regions using a graph-based representation of the image then developed an efficient segmentation algorithm based on this predicate. An important characteristic of the method is its ability to preserve details in low-variability image regions while ignoring details in high-variability regions. Another method based on graph cut was proposed by Eriksson et al.[33] for segmentation and classification of images. The concept of optimality of such cuts was introduced by associating energy to each cut. Graph cut is a very popular approach for image segmentation. It minimizes an energy function consisting of a data term (which is computed using color likelihoods of foreground and background) and a spatial coherency term. But it has problems with segmenting thin elongated objects due to the shrinking bias. Sara Vicente et al.[34] proposed a different way to solve the task of segmenting challenging objects with very thin, elongated parts. They gave a heuristic algorithm named DijkstraGC which actually merges the Dijkstra algorithm and graph cut concept. We know that the Dijkstra method has already been used for extracting thin objects. Also the addition of graph cut which allows to explicitly use the MAP-MRF formulation which proved to be very successful. They also provide another slow method based on problem decomposition which provides a lower bound on the problem. It verifies that for some practical examples DijkstraGC is able to find the global minimum. By incorporating the advantages of the mean shift (MS)[35] segmentation and the normalized cut (Ncut)[32] partitioning methods, W Tao et al.[36] proposed a method which requires low computational complexity and therefore this method was very feasible for real-time image segmentation processing. It first preprocesses an image by using the MS algorithm to form segmented regions that preserve the desirable discontinuity characteristics of the image. The segmented regions then represented by using the graph structures, and the Ncut method is applied to perform globally optimized clustering. Zhenguo Li et al.[37] proposed a segmentation framework which aggregates multi-layer superpixels in an effective manner. It can incorporate additional group constraints such as those specifying a set of pixels belong to the same group. They have shown that spectral clustering can be highly efficient on unbalanced bipartite graphs compare to general graphs. Another work based on graph cut approach proposed by X Wang et al.[38] , where they segmented images by constructing an affinity graph using l_0 sparse representation. A well known classical approach to obtain a single meaningful image partition from a given hierarchy is to cut it in an optimal way, based on the approach of the scale-set theory. The resulting segmentation, being a non-horizontal cut, is limited by the structure of the hierarchy. Yongchao Xu et al.[39] proposed a novel approach that transforming an input hierarchy into a new saliency map. A graph representation of a set of regions extracted from the image. Each region is associated with an attribute. They weighed the boundaries of a subset of meaningful regions (local minima) in the shape space by extinction values based on the attribute. This extinction-based saliency map represents a new hierarchy of segmentation. Each threshold of this map represents a segmentation which is generally different from any cut of the original hierarchy. This new approach thus enlarges the set of possible partition results that can be extracted from a given hierarchy.

2.3 Segmentation based on Spliting and Merging

In this approach, image is successively split into quadrants based on some homogeneity criterion and then based on the similarity between regions, some of them are merged to create the segmented result. This technique works based on a quadtree data structure where the total region is a parent, and each of the four splits is a child.

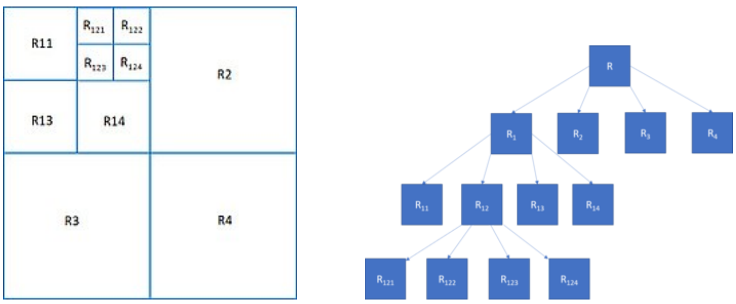

FIGURE 2.1: a] Spliting process of any region. b] Quad tree structure (image courtesy:wikipedia)

In 2002 H.D Cheng et al.[40] proposed a method based on homogram thresholding followed by region merging. It's a two steps process wherein the first step they applied fuzzy homogeneity approach to three color component to find the threshold for each of them then the segmentation result of each of them are combined to partition the color space into several clusters. Some of the clusters are merged in the next step. F.Y Shih et al.[41] proposed a simple approach for segmentation based on region growing approach. Instead of RGB, they used YCbCr color space. They applied automatic seed selection to obtain initial seeds then applied a seeded region growing algorithm to segment the image into regions, in the next step region merging was used to merge similar regions. J Tang et al.[42] worked on region growing approach to segment color images. They proposed a method based on an automatic seed region growing method with the combination of the watershed algorithm and seed region growing algorithm. S Liu et al.[43] proposed a new region-based active contour model, namely local region-based ChanVese (LRCV) model. By considering local characteristics of images, the proposed model can effectively and efficiently segment images with intensity inhomogeneity. To reduce manual interaction in many active contour models and for automatic segmentation, a degraded CV model was proposed, whose segmentation result was the initial input of the LRCV model. B Peng et al.[44] addressed the automatic image segmentation problem in a region merging style. They started with an oversegmented image having many regions with homogeneous color, and then iteratively merging the regions according to a statistical test. The merging order follows the principle of dynamic programming which formulates the image segmentation as an inference problem, where the final segmentation is established based on the observed image. Two essential issues in a region-merging algorithm: order of merging and the stopping criterion are solved with the help of sequential probability ratio test and minimal cost criterion. Sometimes spectrally similar objects often appear in spatially separate locations. J C. Tilton et al.[45] proposed an approach for tightly integrating best merge region growing with nonadjacent region object aggregation, which they named as hierarchical segmentation or HSeg. However, the original implementation of HSeg required excessive computing time even for moderately sized images. They introduced a refined implementation of nonadjacent region object aggregation in HSeg that reduces the computational requirements of HSeg and also this version of HSeg can process moderately sized images in less amount of time.

2.4 Segmentation based on Clustering

Chi Zhang et al.[46] proposed a method where they applied the K-means algorithm[47] in HSI space instead of RGB space. In the process of hue clustering, the special cyclic property of the hue component was taken into consideration. T W Chen et al.[3] proposed a new quantization technique where they chose HSV color space to generate a color histogram and a gray histogram for K-Means clustering[47]. The initialization of centroids and the number of clusters automatically estimated in the proposed method. Also, a post-processing filter was introduced to eliminate small spatial regions. The proposed segmentation algorithm(In Fig 2.2) achieves high computational speed. Li-Hong Juang

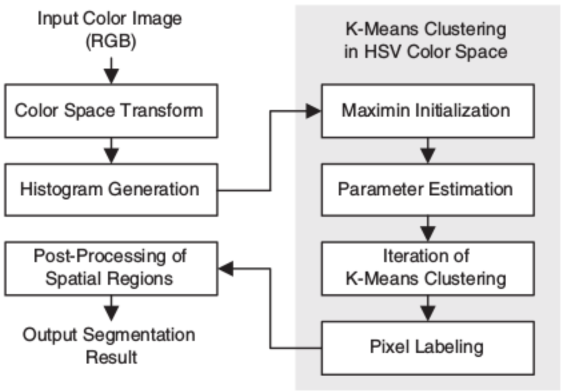

FIGURE 2.2: Proposed algorithm[3]

et al.[48] proposed a method for tumor objects tracking in magnetic resonance imaging (MRI) brain images that incorporate a color-converted segmentation algorithm with K-means clustering technique. Addition of color-based segmentation operation helps to find difficult lesion objects problem in MRI image. The key idea was to convert the input gray-scale image into an RGB image and operating the image labeled by a K-means clustering[47]. Segmentation of real-world images is a challenging task because of the presence of complex noise. To address these problems, Lingfeng Wang et al.[48] propose a novel segmentation algorithm via a local correntropy-based K-means (LCK) clustering. Due to the correntropy criterion, the clustering algorithm can decrease the weights of the samples that are away from their clusters and also increase the weights of samples close to their clusters As a result, this algorithm was robust to the outliers. Another clustering based segmentation approach in medical imaging proposed by E.A.Maksoud et al.[49] They integrated the concept of K-means clustering[47] with Fuzzy C-means[27] algorithm to provide an accurate brain tumor detection system. The proposed technique took the advantages of the K-means clustering in the aspects of minimal computation time and also get advantages of the Fuzzy C-means in the aspects of accuracy. Choy S K et al.[50] proposed an unsupervised learning algorithm based on a fuzzy Generalized Gaussian Density (GGD) segmentation model and the GGD-based agglomerative fuzzy algorithm for grouping image pixels. The advantage of this algorithm is initial parameters and the number of groups can be estimated via the validation technique. To minimize the objective function they defined a dissimilarity measure based on the KullbackLeibler divergence of the GGDs.

2.5 Segmentation based on Texture

Malik Jitendra et al.[51] proposed an image segmentation algorithm based on the framework of Normalized Cuts. They considered image as a weighted graph where the nodes i and j are pixels and edge weights, W_{ij}, denote a similarity between the two pixels then performed grouping by finding eigenvectors of the normalized Laplacian of this graph. The proposed method analyzes the image using contour and texture. The local similarity measure between pixels i and j due to the contour cue was computed in the intervening contour framework using peaks in contour orientation energy. They analyzed the texture using textons.A histogram of texton densities was used as the texture descriptor. Here similarity, is measured using the χ^2 test on histograms. The edge weights estimated by combining both contour and texture information. Y Deng et al.[52] proposed a method for unsupervised segmentation of color-texture regions named JSEG. The proposed method was a two-step procedure, color quantization followed by spatial segmentation. In the first step, colors in the image were quantized to several representative classes then each pixel was replaced by their corresponding color class labels and formed a class-map of the image. In the second step, the focus was on spatial segmentation where a criterion for good segmentation using the class-map is proposed. They applied the criterion to local windows in the class-map results in the J-image, in which high value correspond to possible boundaries and low values to interiors of colour texture regions. Then they used a region growing method to get the final segmented result. J Cui et al.[53] proposed an iris localization algorithm based on texture segmentation in the field of medical imaging. An iris recognition system consists of two steps preprocessing and iris localization. Iris localization includes finding the iris boundaries and eyelids. The proposed iris localization algorithm first uses the information of low frequency of wavelet transform of the iris image for pupil segmentation and then localize the iris with a differential integral operator. Then the upper eyelid edge was detected after eyelash is segmented. Finally, the lower eyelid was localized using a parabolic curve fitting based on gray value segmentation. Chang Y.C et al.[54] proposed another texture-based segmentation algorithm for color images. Many powerful color image segmentation algorithms such as J-Segmentation (JSEG)[52] suffer from over-segmentation. So they developed an improved JSEG method called improved contrast JSEG, or IC-JSEG. The proposed method first builds a contrast map to obtain the basic contours of the homogeneous regions in the image. The noise was removed by applying different filters and then a seed growing-merging method applied to both improved contrast map and the original J-map constructed by JSEG to segment the image. Houhou N et al.[55] presented an approach for unsupervised segmentation of natural and textural images based on the extraction of image features and a fast active contour segmentation model. The main problem of textural images is neither the gray-level information nor the boundary information is adequate for object extraction because these images cannot be processed directly by the gray-level information so they proposed a new texture descriptor which intrinsically defined the geometry of textures using semi-local image information and tools from differential geometry. Then, to design an active contour model they used the popular Kullback-Leibler distance which distinguishes the background and textures of interest. Finally, a texture segmentation algorithm based on the Split-Bregman method was introduced to extract meaningful objects in a fast way. Brox T. et al[55] presented an approach for the extraction and combination of different cues in a level set based image segmentation framework. The proposed method had two steps. In the first step, the input features were extracted and enhanced by applying coupled nonlinear diffusion and in the next step, the resulting features were then employed for a vector-valued front propagation based on level sets and statistical region models that approximate the distributions of each feature. J Yuan et al.[56] worked on a factorization based approach and proposed an efficient method for segmenting textured images. Wu Q et al.[57] presented another texture image segmentation where they proposed a convex texture image segmentation model. They first extracted the texture features of Gabor and GLCM (gray level co-occurrence matrix) from the original image and then fused together to effectively construct a discriminative feature space by concatenating with each other. Then in the segmentation step, a convex energy function was defined by taking the non-convex vector-valued model of Active Contour without Edges (ACWE) into a global minimization framework (GMAC). The proposed global minimization energy function with fused textures (GMFT) can avoid the existence of local minima in the minimization of the vector-valued ACWE model.

2.6 Segmentation of natural scenes

Lazebnik, S et al.[58] proposed a method that recognized the category of natural scenes based on global geometric information. This technique first partitioned the image into increasingly sub-regions and computed histograms of local features for each sub-region. So the resulting spatial pyramid was an extension of orderless bag-of-features image representation. The proposed algorithm also achieved high accuracy on a large database of fifteen natural scene categories of Caltech-101 database. N Kulkarni et al.[59] proposed a multilevel color thresholding method for natural scene segmentation. The proposed technique was being carried out based on a slight modification of the grey level thresholding algorithm. Multilevel thresholding had been conducted to the RGB color information of the object and extract it from the background and other objects. Another work presented by H Mobahi et al.[60] for the segmentation of natural scenes. This method was based on observations that how a homogeneously textured region of a natural scene can be well modeled by a Gaussian distribution and then region boundary was coded by an adaptive chain code. Yu Han et al.[61] proposed a method based on the fuzzy framework, which solved the problem of segmenting multi-region color-scale images of natural scenes. The advantage of this method was the introduction of PCA descriptors, that can partition color-texture images better than classical variational-based segmentation models. J. T. Barron et al.[62] extended their previous work, which was shape, illumination, and reflectance from shading (SIRFS) model[63, 64], which recovered intrinsic scene properties from a single image. Though SIRFS performs well on images of segmented objects, it performs poorly on images containing occlusion and spatially-varying illumination. So they proposed Scene-SIRFS, a generalization of SIRFS which performs well with images having a mixture of shapes and a mixture of illuminations were embedded. Md. Abul Hasnat et al.[65] proposed an unsupervised method for indoor RGB-D image segmentation which consists of the following things, a joint color-spatial-directional clustering method, and a statistical planar region merging method. They considered a statistical image generation model based on the color and geometry of the scene. J Redmon et al.[66] proposed an approach for object detection. They treated object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. They proposed a single neural network that predicts the bounding boxes and class probabilities directly from full images in one evaluation. V Badrinarayanan et al.[67] presented a practical deep fully convolutional neural network architecture named SegNet for semantic pixel-wise segmentation which consist of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. Mukherjee, Aritra et al.[68] also worked on natural scene segmentation. First, they applied the K-means clustering algorithm on the original image. The number of clusters was automatically computed from the color histogram of the input image by finding the local minima in it. Then they merged the pixel groups where one group is contained within another based on spatial inclusiveness criteria. At last, a graph-based approach was followed by them for merging the adjacent regions based on color gradient.

After going through some research works done by eminent researchers in the field of natural scene segmentation we are moving forward to the next chapter of this thesis where we are going to describe our proposed methodology.

Chapter 3

Proposed Methodology

In this chapter, the proposed methodology is discussed briefly. Here we are trying to build an automated system with less human interaction. Our input to the proposed system is a natural color image. The ideal output of this system is a segmented image where each segment which is semantically different from the other segments, are assigned with a unique label. And our desired output of this system is a well-segmented image where a set of pixels are assigned with a unique label based on their RGB color distribution, hue values, and some spatial distances. Broadly our algorithm is divided into three steps, (a) preprocessing of the input image, (b) superpixel segmentation and (c) graph creation and merging the neighboring segments based on some criteria. A block diagram of the proposed system is shown in Figure 3.1.

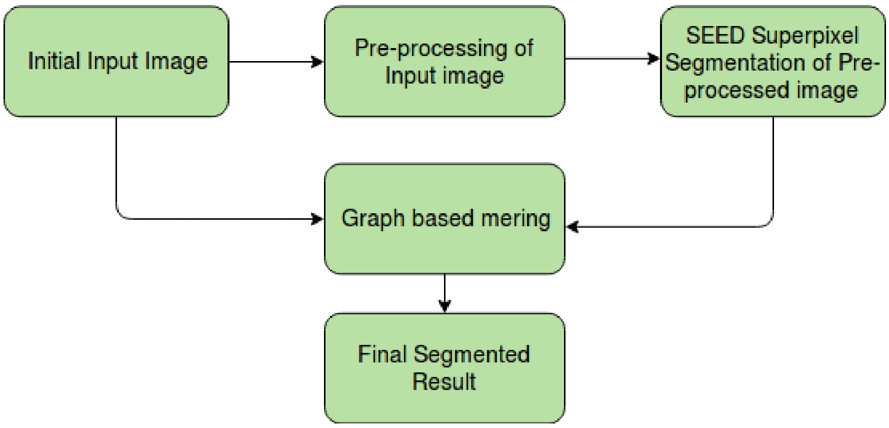

FIGURE 3.1: Block diagram of the proposed system

FIGURE 3.2: Initial input image to the system

3.1 Pre-processing of the input image

In this phase, the main motive is to reduce the noise from the input image and retaining the details of the edges. This noise removal process is very important for future operations. The non-local means denoising technique[69] is used for noise reduction. The basic idea of this technique is replacing the color of each pixel with an average of the colors of similar pixels but not necessarily the location of similar pixels are close to that pixel. So a huge number of a scan of the image is required to find these similar pixels. Then denoising is performed by computing the average color of these similar pixels. This filter can be written as

$$NLu(p) = \frac{1}{C(p)} \int f(d(B(p), B(q))u(q)dq \tag{3.1}$$

where $d(B(p), B(q))$ is an Euclidean distance between image patches centered respectively at p and q. f is a decreasing function and $C(p)$ is the normalizing factor.

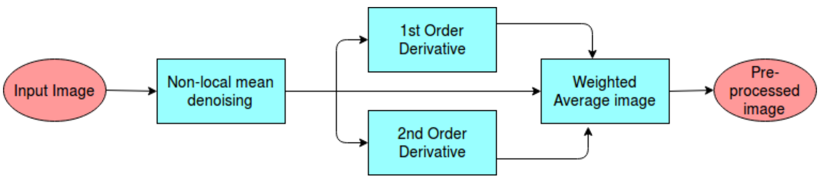

FIGURE 3.3: Block diagram of the pre-processing phase

After applying non-local means filter to the input image we have got a denoised image. The superpixel segmentation depends on the edges of image. Due to the denoising operation some edges may have been weaken so enhancement of those edge is required.

So first order and second order derivatives of the denoised image is computed. The final pre-processed image which is going to be the input of superpixel segmentation is a weighted average of the denoised image and two derivatives that we got from the denoised image. The pre-processed image is shown in Fig: 3.4.

FIGURE 3.4: The pre-processed input image

3.2 Superpixel Segmentation

Here a superpixel segmentation is applied on the pre-processed image which we got from the previous phase. SEED superpixel algorithm proposed by Bergh et al.[70] is used in this phase. This algorithm simply over-segment the input image by grouping pixels belong to the same object. They presented the SEED superpixel algorithm as an energy maximization problem where each superpixel was defined as a region with a color distribution and a shape of the boundary. They defined the energy function as the sum of two terms $H(s)$ and $G(s)$. The energy function is

$$E(s) = H(s) + \gamma G(s) \tag{3.2}$$

Where $H(s)$ focuses on the color distribution in the superpixel. $G(s)$ is called boundary term that takes care of the shapes of the superpixel boundaries. γ weighs the influence of each term.

They approximated the color density distribution term by building a binned histogram. And for boundary term computation, they placed a patch of fixed size around each pixel in the image and for each patch it counts the number of different superpixels. A patch corresponding to a pixel near the boundaries contains several superpixels, on the other side a patch of a non-boundary pixel contains unique superpixel. Initially, the process starts with an image divided into non-overlapped blocks where each block is called as superpixel. To maximize the objective function they have used Hill-climbing optimization algorithm which is a very fast and efficient method. This algorithm iteratively updates the solution by proposing small local changes at each iteration. If the energy function of a proposed solution increases, the solution will be updated. Suppose at any instant time t, $s_t \in S$ (S is the set of all sets of partitions) is the current partition and $s \in S$ as is the new proposed partition which can get by updating some local changes at s_t. Now if the objective function of s increases then updates s_t with s. An overview of the hill-climbing algorithm is shown in Algorithm 1.

s_t =initialize();
while $t < t_{stop}$ **do**
    $s = Propose(s_t)$;
    if $E(s) > E(s_t)$ **then**
        $s_t = s$;
    end
end
S=s_t;

Algorithm 1: Overview of Hill Climbing Algorithm

FIGURE 3.5: Result after applying of the SEED superpixel algorithm

This SEED superpixel method is already been implemented in OpenCV library. We have used that inbuilt function. It takes a few parameters as input to the function like initial number of superpixels and the number of iterations etc. Higher the number of superpixels or iteration more is the time requirement. Here we have considered the value of initial number of superpixels and number of iterations are 200 and 10 which will be fixed for full scale images. This function returns a labeled image where the value ranges from 0 to N_{sp} 1 where N_{sp} is the number of superpixels detected by the algorithm.

Also, it returns the mask of the labeled image. The result of the SEED superpixel segmentation is shown in Fig: 3.5.

3.3 Graph Building and Merging

At first, a graph is built with the help of labeled image L and its mask M which were the outputs of the SEED superpixel algorithm. Where Each node represents a superpixel and an edge between two nodes indicate that the nodes are spatially close to each other. We have represented the graph as an adjacency matrix G with size N_{sp} x N_{sp} where N_{sp} is the total number of superpixel. $G_{ij} = 1$ indicates that $i - th$ and $j - th$ superpixels are spatially very close and $G_{ij} = 0$ otherwise. For the initial graph building process, we followed this approach. Scanning M for each pixel whenever an edge pixel is encountered simply compare the value of the corresponding position in L with its 8 neighborhood(north, south, east, west, north-east, south-east, south-west, north-west) if a mismatch found we immediately set the value of corresponding position of graph as 1 i.e $G_{current,neighbor}$= 1 and otherwise 0.

After completion of the graph building process, the next step is merging the neighborhood superpixel. This step is important because it may happen that two superpixels are spatially close but texturally different and color-distributions are also different then we cannot merge those superpixel, so we have to take care of these situation. For merging we have focused mainly on textural and color similarity. So to compute the texture features we have used local binary pattern(LBP)[71] which captures the textural features around each pixel of an image. In our method, we have applied the LBP on L and got L_{LBP}. From L_{LBP} we computed the histogram of each superpixel. Let i and j are two connected superpixels and X_i, Y_j are the histograms of them. For measuring the texture similarity between two superpixels we have used Chi-square distance[72, 73]. It is computed as:

$$CSD_{ij} = \sum_{k=1}^{b} \frac{(X(k)_i - Y(k)_j)^2}{(X(k)_i + Y(k)_j)} \tag{3.3}$$

The LBP histograms are divided into b bins. In our experiment we have taken b=5. Low value of CSD_{ij} indicates high similarity in terms of texture. For color similarity, we measured three types of distance, (a) Euclidean distance between the average RGB value of two superpixels (b) Bhattacharya Distance and (c) Hue based distance. For a Euclidean distance calculation between two superpixel i and j we used this simple formula.

$$CBD_{ij} = \sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2} \tag{3.4}$$

Where R_i, G_i, B_i are the average color value of $i-th$ superpixel and R_j, G_j, B_j are the average color value of $j-th$ superpixel.

Another measure called Bhattacharyya distance[74] which can measure the similarity between two distribution. Here we used this for measure similarity between two histogram H_i and H_j of $i-th$ and $j-th$ superpixels. The formula for Bhattacharyya distance is following

$$BD_{ij} = \sqrt{1 - \frac{\sum c \in r, g, b(\sum_{k=1}^{b} \sqrt{H_i{}^c(k) H_j{}^c(k)})}{3}} \tag{3.5}$$

Where $H_i{}^c$ and $H_j{}^c$ represent the histograms of color channel c ($\in R, G, B$) for superpixels i and j respectively. Value of bin b is considered as 5. Sometimes it may happen that color values of superpixel gets effected by noise. For this reason we have taken the last distance measure which is the absolute difference between the average hue value of two superpixel i and j.

$$HD_{ij} = |h_i - h_j| \tag{3.6}$$

Where h_i and h_j are the average hue value of i and j superpixel.

So based on this above mentioned measures neighboring superpixels have merged. Each of this measures is normalized between 0 and 1, which indicate a confidence value. A high confidence value indicates low distance and high similarity. For each edge, we computed the confidence value $C_{i,j}$ as

$$\begin{aligned} C_{i,j} = (C_{CSD_{i,j}} - thresh_{CSD}) + (C_{CSD_{i,j}} - thresh_{CSD}) \\ +(C_{BD_{i,j}} - thresh_{BD}) + (C_{HD_{i,j}} - thresh_{HD}) \end{aligned} \tag{3.7}$$

Where $C_{CSD_{i,j}}$, $C_{CBD_{i,j}}$, $C_{BD_{i,j}}$, $C_{HD_{i,j}}$ are the confidence corresponding to $CSD_{i,j}$, $CBD_{i,j}$, $BD_{i,j}$, $HD_{i,j}$. And $thresh_{CSD}$, $thresh_{CBD}$, $thresh_{BD}$, $thresh_{HD}$ are the threshold values of confidence measure of texture and color. The neighborhood graph of image after merging is shown in Fig: 3.6.

3.4 Requirement of Ultrametric Contour Map (UCM)

Our proposed method's outline is performed superpixel segmentation followed by a graph-based merging and then extraction of regions. But some segmentation methods also follow edge-based strategy. For performance comparison with edge-based methods, we have built our method on the multi-scale version and generated Ultrametric Contour Map (UCM) which represents an edge map with confidence. The discussion so far

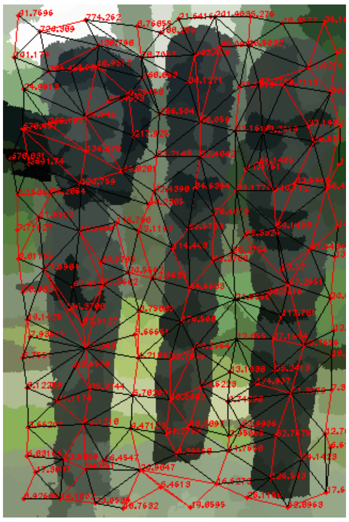

Figure 3.6: The neighboring graph showing the edges between nodes. Where Red colored edge denotes that two superpixels should be merged and Black edge denotes two superpixel cannot be merged

corresponds to scale 1. We have applied the superpixel segmentation followed by graph-based merge process on the 80%, 60%, 40% and 20% of the pre-processed image and generated edge maps. Initial number of superpixels (input to SEED) is also adjusted proportionately according to the scale. To compute UCM all edge maps are scaled to the same size and then skeletonized and averaged. In UCM the weak edges have less confidence value and the strong edges having a high confidence value in all scale. The final result and its Ultrametric Contour Map are shown in Fig: 3.7.

Figure 3.7: Left side image is the Final segmented image of the proposed method. Right side image is Ultrametric Contour Map of final result.

Chapter 4

Implementation, Results and Analysis

In this chapter, we are going to discuss the computational implementation details, the results of our proposed methodology and analysis of the results. The hardware and software specifications of the computer systems for the implementation of the proposed methodology are shown in Table 4.1. We did not use GPU for this implementation.

Parameter	**Specification**
Processor	Intel(R) Core(TM) i3-2350M (2nd Gen) CPU @ 2.30GHz L3 cache: 3 MB
Main Memory	4 GB 1333 MHz DDR3 SDRAM*2
Hard Disk Drive	320 GB*4 (Serial ATA, 5400 rpm)
Graphics Memory	NVIDIA GeForce 410M GPU, 512 MB DDR3
Operating System	Linux, Ubuntu 16.04 LTS 64 bit

Table 4.1: Configuration of Experimental Machine

We have used C++ language and OpenCV 4.0.0 library[75] for the implementation of the proposed methodology. OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library developed by Intel, mainly aimed to build real-time computer vision application. BSDS300 dataset is used for our experiment. It consists of 300 images with a resolution of 481×321 for landscape images and 321×481 for portrait images. But for our experiment purpose, we have increased the resolution by 1.5 times to adjust with the standard webcam resolution. The dataset contains a huge variety of images like natural scenes, wildlife scenes, scenes focused on human subjects, etc. The dataset has provided unlabeled human annotated ground truth image. For each image, it gives single human annotation(single ground truth for an image) and multi-human annotation(multiple ground truth by the individuals). An edge contour map is computed for each image by averaging the ground truths of multi-human annotation and treated as ground truth UCM.

For performance analysis of our results, we have compared it with the methods proposed by Felzenszwalb et al.[4], Bosch et al.[5], Arbelaez et al.[6], and Pont et al.[7]. The outputs of [6, 7] methods are in UCM form(Ultrametric Contour Map) but the outputs of [4, 5] are of edge image form so we have computed our outputs in UCM form(averaging the edge maps at various scale) as well as edge form.

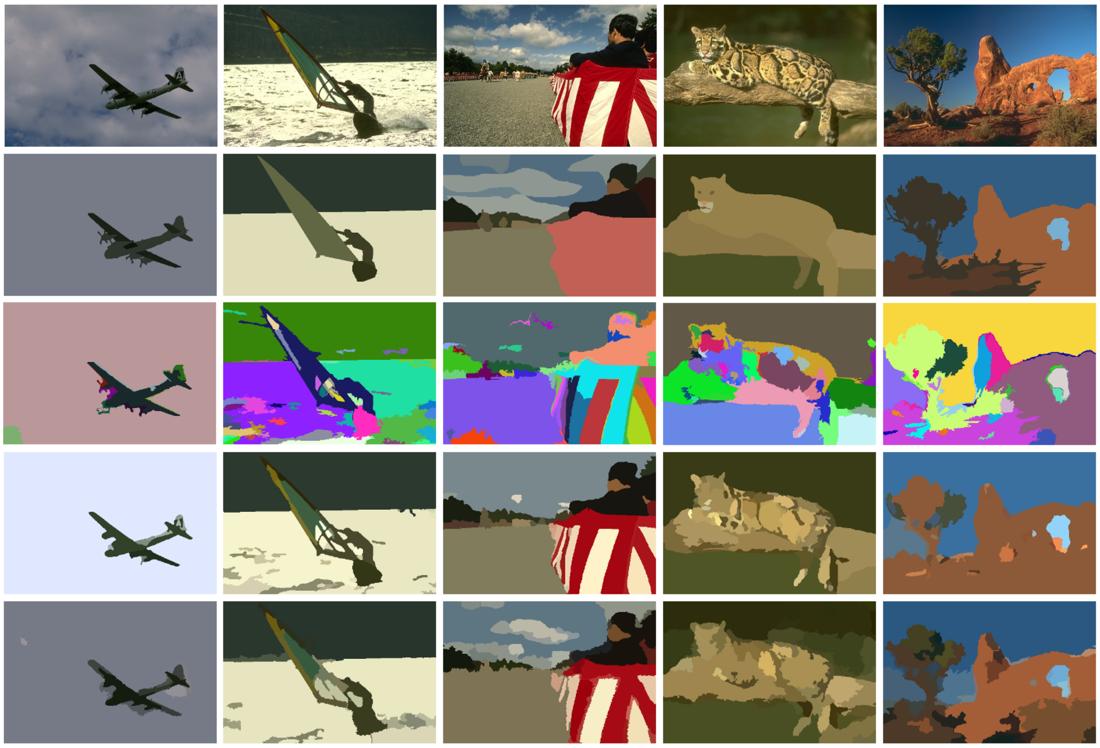

FIGURE 4.1: Output of different methods in single scale mode, from top to bottom rows are: Input image, single human annotated ground truth, result of Felzenszwalb et al.[4], result of Bosch et al.[5] and result of proposed method

Figure 4.1 shows a visual comparison of the output between Felzenszwalb et al.[4] works, Bosch et al.[5] works, and our proposed methodology. In the output images of ground truth and Bosch et al.[5], the segments are filled with the average color of corresponding pixels in the original image. And the outputs of Felzenszwalb et al.[4] are filled with a random color by their source code. Visually we can say that the output of the proposed methodology is better for some images. Figure 4.2 shows the comparison with the

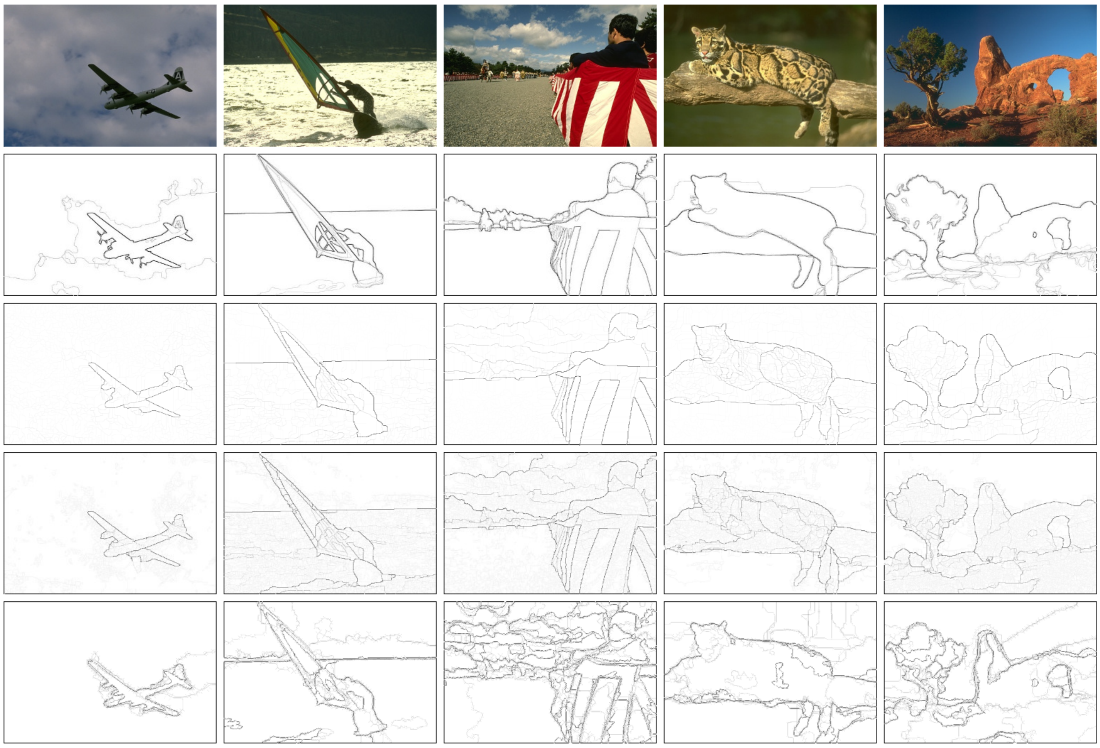

Figure 4.2: Output of different methods in UCM mode, from top to bottom rows are: input image, multi user ground truth UCM, result of Arbelaez et al[6], result of Pont et al[7] and our result

methodologies of Arbelaez et al[6], Pont et al[7].In this case the outputs are considered in UCM form. Visually, it is clear that the proposed methodology outperforms others.

Methodology	**Average F-measure**	**Average precision**	**Average recall**
Felzenszwalb *et al*[4]	0.39	0.51	0.35
Bosch *et al*[5]	0.51	0.55	0.50
Proposed Methodology	0.53	0.44	0.74

Table 4.2: Comparison of performance based on edge map

In Table 4.2 a quantitative comparison is given between the proposed method and other methods based on precision, recall, and F-measure. After considering all images of the dataset it gives an overview of average precision, recall, and F-measure for the works of Felzenszwalb et al[4] and Bosch et al.[5] and proposed methodology. From the table, we can say that with respect to F-score and recall the proposed methodology is better than others but the average precision value of the proposed methodology is not good with respect to others. Over-segmentation is the reason behind a less precision value.

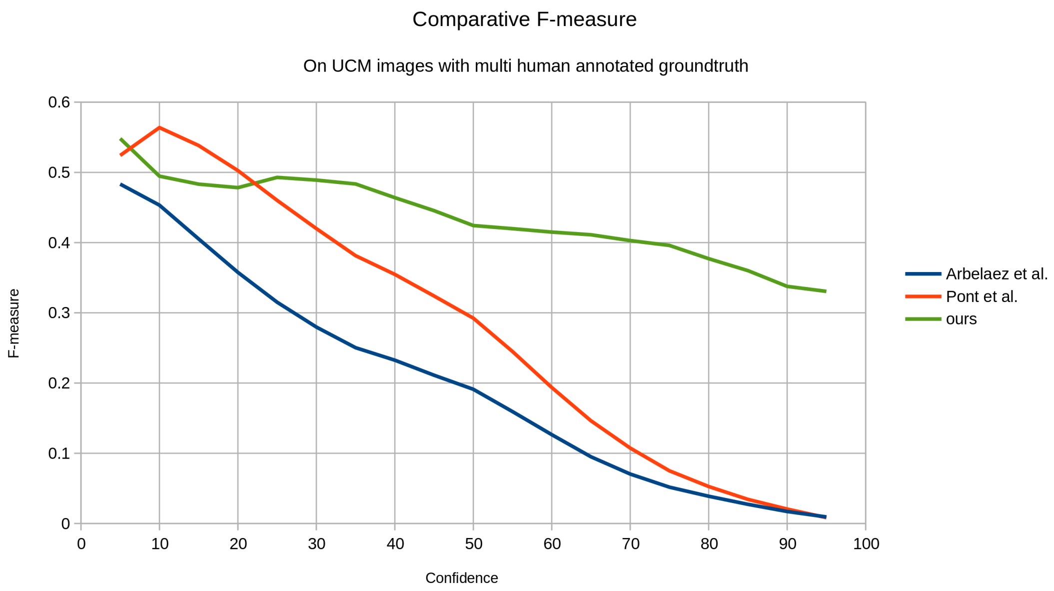

FIGURE 4.3: Comparison of F-score based on UCM

Figure 4.3 shows the comparison between the methods of Arbelaez et al.[6], Pont et al.[7] and the proposed method based on the F-score measures of their corresponding UCMs with respect to ground truth UCM. The plotting of Figure 4.3 has shown that the F-score measure of the proposed methodology is higher than the other methods for most of the images. Another time-based analysis has been shown in Table 4.3. The time required

Scale of image	**minimum time (in ms)**	**maximum time (in ms)**	**average time (in ms)**
1.0 (full scale)	287	359	317
0.8	182	228	200
0.6	104	127	111
0.4	45	56	48
0.2	15	18	15
All scales	633	788	691

TABLE 4.3: Time requirement for proposed methodology

by our proposed methodology for all images at different scales is given. Although the processing time of the proposed method for segmenting a full-scale image is not good enough for real-time application. But at a slightly reduced scale, its performance is good enough for real-time application. So we can say that based on a few comparisons and analysis proposed methodology performs well.

Chapter 5

Conclusion

In this thesis, we have proposed a novel strategy for natural scene segmentation. Where we combine the approaches of some elegant methods for segmenting images. In the pre-processing phase, the initial input to the system is denoised by applying non-local means denoising algorithm. Then computed the first order, second order derivatives of that denoised image and finally, a weighted average of all three images is taken. SEED superpixel algorithm is applied to the weighted average image which gives a labeled image and its mask as output. After that, a neighboring graph is formed from the outputs of previous steps. Finally, merging some of the neighboring regions based on their similarity of texturedness and similarity of color distribution. In future, this methodology can be extended further for semantic labeling of segments created by statistical methods. The hybrid approach can lead to higher accuracy along with semantic labeling without the risk of model bias of deep learning methods. Segments histogram calculation is an expensive affair and is not coupled among segments, thus partial parallel implementation of the methodology will lead to a higher speed, which again is a future prospect.

Bibliography

[1] Lawrence G Roberts. *Machine perception of three-dimensional solids*. PhD thesis, Massachusetts Institute of Technology, 1963.

[2] Richard Szeliski. *Computer Vision - Algorithms and Applications*. Texts in Computer Science. Springer, 2011. ISBN 978-1-84882-934-3. doi: 10.1007/978-1-84882-935-0. URL `https://doi.org/10.1007/978-1-84882-935-0`.

[3] Tse-Wei Chen, Yi-Ling Chen, and Shao-Yi Chien. Fast image segmentation based on k-means clustering with histograms in hsv color space. In *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pages 322–325. IEEE, 2008.

[4] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

[5] Marc Bosch, Christopher M Gifford, Austin G Dress, Clare W Lau, Jeffrey G Skibo, and Gordon A Christie. Improved image segmentation via cost minimization of multiple hypotheses. *arXiv preprint arXiv:1802.00088*, 2018.

[6] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2010.

[7] Jordi Pont-Tuset, Pablo Arbelaez, Jonathan T Barron, Ferran Marques, and Jitendra Malik. Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE transactions on pattern analysis and machine intelligence*, 39(1):128–140, 2016.

[8] Dana H Ballard and Christopher M Brown. Computer vision. englewood cliffs. *J: Prentice Hall*, 1982.

[9] T Huang. Computer vision: Evolution and promise. 1996.

[10] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[11] Seymour A Papert. The summer vision project. 1966.

[12] Matthew A Turk and Alex P Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591. IEEE, 1991.

[13] Yu-Jin Zhang. An overview of image and video segmentation in the last 40 years. In *Advances in Image and Video Segmentation*, pages 1–16. IGI Global, 2006.

[14] Tony Lindeberg and Meng-Xiang Li. Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Computer Vision and Image Understanding*, 67(1):88–98, 1997.

[15] Salem Saleh Al-Amri, NV Kalyankar, and SD Khamitkar. Image segmentation by using edge detection. *International Journal on computer science and engineering*, 2(3):804–807, 2010.

[16] Dilpreet Kaur and Yadwinder Kaur. Various image segmentation techniques: a review. *International Journal of Computer Science and Mobile Computing*, 3(5): 809–814, 2014.

[17] Muhammad Rizwan Khokher, Abdul Ghafoor, and Adil Masood Siddiqui. Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system. *IET image processing*, 7(3):201–211, 2013.

[18] S Angelina, L Padma Suresh, and SH Krishna Veni. Image segmentation based on genetic algorithm for region growth and region merging. In *2012 international conference on computing, electronics and electrical technologies (ICCEET)*, pages 970–974. IEEE, 2012.

[19] Hassana Grema Kaganami and Zou Beiji. Region-based segmentation versus edge detection. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1217–1221. IEEE, 2009.

[20] Mahesh Yambal and Hitesh Gupta. Image segmentation using fuzzy c means clustering: a survey. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(7), 2013.

[21] Vinod Kumar Dehariya, Shailendra Kumar Shrivastava, and RC Jain. Clustering of image data set using k-means and fuzzy k-means algorithms. In *2010 International Conference on Computational Intelligence and Communication Networks*, pages 386–391. IEEE, 2010.

[22] Wen-Xiong Kang, Qing-Qiang Yang, and Run-Peng Liang. The comparative research on image segmentation algorithms. In *2009 First International Workshop on Education Technology and Computer Science*, volume 2, pages 703–707. IEEE, 2009.

[23] Waseem Khan. Image segmentation techniques: A survey. *Journal of Image and Graphics*, 1(4):166–170, 2013.

[24] Berend Jan Van Der Zwaag, C Slump, and Lambert Spaanenburg. Analysis of neural networks for edge detection. In *Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing*, pages 580–586, 2002.

[25] D Suganthi and S Purushothaman. Fmri segmentation using echo state neural network. *International Journal of Image Processing*, 2(1):1–9, 2008.

[26] Young Won Lim and Sang Uk Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern recognition*, 23(9):935–952, 1990.

[27] Pi Yeong Chi and J Van Ryzin. A simple histogram method for nonparametric classification. In *Classification and Clustering*, pages 395–421. Elsevier, 1977.

[28] Mohamed Cheriet, Joseph N Said, and Ching Y Suen. A recursive thresholding technique for image segmentation. *IEEE transactions on image processing*, 7(6): 918–921, 1998.

[29] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[30] Agus Zainal Arifin and Akira Asano. Image segmentation by histogram thresholding using hierarchical cluster analysis. *Pattern Recognition Letters*, 27(13):1515–1521, 2006.

[31] Soham Sarkar and Swagatam Das. Multilevel image thresholding based on 2d histogram and maximum tsallis entropya differential evolution approach. *IEEE Transactions on Image Processing*, 22(12):4788–4797, 2013.

[32] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Departmental Papers (CIS)*, page 107, 2000.

[33] Anders P Eriksson, Olof Barr, and Kalle Astrom. Image segmentation using minimal graph cuts. 2006.

[34] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Graph cut based image segmentation with connectivity priors. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[35] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5):603–619, 2002.

[36] Wenbing Tao, Hai Jin, and Yimin Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1382–1389, 2007.

[37] Zhenguo Li, Xiao-Ming Wu, and Shih-Fu Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 789–796. IEEE, 2012.

[38] Xiaofang Wang, Huibin Li, Charles-Edmond Bichot, Simon Masnou, and Liming Chen. A graph-cut approach to image segmentation using an affinity graph based on 0-sparse representation of features. In *2013 IEEE International Conference on Image Processing*, pages 4019–4023. IEEE, 2013.

[39] Yongchao Xu, Edwin Carlinet, Thierry Géraud, and Laurent Najman. Hierarchical segmentation using tree-based shape spaces. *IEEE transactions on pattern analysis and machine intelligence*, 39(3):457–469, 2017.

[40] Heng-Da Cheng, XH Jiang, and Jingli Wang. Color image segmentation based on homogram thresholding and region merging. *Pattern recognition*, 35(2):373–393, 2002.

[41] Frank Y Shih and Shouxian Cheng. Automatic seeded region growing for color image segmentation. *Image and vision computing*, 23(10):877–886, 2005.

[42] Jun Tang. A color image segmentation algorithm based on region growing. In *2010 2nd International Conference on Computer Engineering and Technology*, volume 6, pages V6–634. IEEE, 2010.

[43] Shigang Liu and Yali Peng. A local region-based chan–vese model for image segmentation. *Pattern Recognition*, 45(7):2769–2779, 2012.

[44] Bo Peng, Lei Zhang, and David Zhang. Automatic image segmentation by dynamic region merging. *IEEE Transactions on image processing*, 20(12):3592–3605, 2011.

[45] James C Tilton, Yuliya Tarabalka, Paul M Montesano, and Emanuel Gofman. Best merge region-growing segmentation with integrated nonadjacent region object aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4454–4467, 2012.

[46] Chi Zhang and Patrick Wang. A new method of color image segmentation based on intensity and hue clustering. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 613–616. IEEE, 2000.

[47] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[48] Li-Hong Juang and Ming-Ni Wu. Mri brain lesion image detection based on color-converted k-means clustering segmentation. *Measurement*, 43(7):941–949, 2010.

[49] Eman Abdel-Maksoud, Mohammed Elmogy, and Rashid Al-Awadi. Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, 16(1):71–81, 2015.

[50] Siu Kai Choy, Shu Yan Lam, Kwok Wai Yu, Wing Yan Lee, and King Tai Leung. Fuzzy model-based clustering and its application in image segmentation. *Pattern Recognition*, 68:141–157, 2017.

[51] Jitendra Malik, Serge Belongie, Thomas Leung, and Jianbo Shi. Contour and texture analysis for image segmentation. *International journal of computer vision*, 43(1):7–27, 2001.

[52] Yining Deng and BS Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE transactions on pattern analysis and machine intelligence*, 23(8):800–810, 2001.

[53] Jiali Cui, Yunhong Wang, Tieniu Tan, Li Ma, and Zhenan Sun. A fast and robust iris localization method based on texture segmentation. In *Biometric Technology for Human Identification*, volume 5404, pages 401–409. International Society for Optics and Photonics, 2004.

[54] YC Chang, James K Archibald, Yong-gang Wang, and Dah-jye Lee. Texture-based color image segmentation using local contrast information. *International Journal of Information Technology and Intelligent Computing*, 2(4):12, 2007.

[55] Nawal Houhou, Jean-Philippe Thiran, and Xavier Bresson. Fast texture segmentation based on semi-local region descriptor and active contour. *Numerical Mathematics: Theory, Methods and Applications.*, 2(ARTICLE):445–468, 2009.

[56] Jiangye Yuan, Deliang Wang, and Anil M Cheriyadat. Factorization-based texture segmentation. *IEEE Transactions on Image Processing*, 24(11):3488–3497, 2015.

[57] Qinggang Wu, Yong Gan, Bin Lin, Qiuwen Zhang, and Huawen Chang. An active contour model based on fused texture features for image segmentation. *Neurocomputing*, 151:1133–1141, 2015.

[58] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2169–2178. IEEE, 2006.

[59] Nilima Kulkarni. Color thresholding method for image segmentation of natural images. *International Journal of Image, Graphics and Signal Processing*, 4(1):28, 2012.

[60] Hossein Mobahi, Shankar R Rao, Allen Y Yang, Shankar S Sastry, and Yi Ma. Segmentation of natural images by texture and boundary compression. *International journal of computer vision*, 95(1):86–98, 2011.

[61] Yu Han, Xiang-Chu Feng, and George Baciu. Variational and pca based natural image segmentation. *Pattern Recognition*, 46(7):1971–1984, 2013.

[62] Jonathan T Barron and Jitendra Malik. Intrinsic scene properties from a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24, 2013.

[63] Jonathan T Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *European Conference on Computer Vision*, pages 57–70. Springer, 2012.

[64] Jonathan T Barron and Jitendra Malik. Shape, albedo, and illumination from a single image of an unknown object. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 334–341. IEEE, 2012.

[65] Md Abul Hasnat, Olivier Alata, and Alain Tremeau. Joint color-spatial-directional clustering and region merging (jcsd-rm) for unsupervised rgb-d image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2255–2268, 2015.

[66] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[67] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[68] Aritra Mukherjee, Soumik Sarkar, and Sanjoy Kumar Saha. Segmentation of natural image based on colour cohesion and spatial criteria. In *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, pages 1–7. IEEE, 2017.

[69] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Non-local means denoising. *Image Processing On Line*, 1:208–212, 2011.

[70] Michael Van den Bergh, Xavier Boix, Gemma Roig, Benjamin de Capitani, and Luc Van Gool. Seeds: Superpixels extracted via energy-driven sampling. In *European conference on computer vision*, pages 13–26. Springer, 2012.

[71] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[72] Yossi Rubner and Carlo Tomasi. *Perceptual metrics for image database navigation*, volume 594. Springer Science & Business Media, 2013.

[73] Dengsheng Zhang and Guojun Lu. Evaluation of similarity measurement for image retrieval. In *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003*, volume 2, pages 928–931. IEEE, 2003.

[74] Frank J Aherne, Neil A Thacker, and Peter I Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4): 363–368, 1998.

[75] Gary Bradski. The opencv library. *Dr Dobb's J. Software Tools*, 25:120–125, 2000.