IMAGE CLASSIFICATION AND SEGMENTATION FOR EFFICIENT SURVEILLANCE APPLICATIONS

MARYAM ASADZADEH KALJAHI

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

IMAGE CLASSIFICATION AND SEGMENTATION FOR EFFICENT SURVELLIENCE APPLICATIONS

MARYAM ASADZADEH KALJAHI

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Maryam Asadzadeh kaljahi

Matric No: WHA150055

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("Image classification and segmentation for efficient surveillance applications"):

Field of Study: Image Processing, Visual Sensor Network

 I do solemnly and sincerely declare that:

(1)   I am the sole author/writer of this Work;
(2)   This Work is original;
(3)   Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4)   I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5)   I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6)   I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature   Date:

Subscribed and solemnly declared before,

Witness's Signature   Date:

Name:

Designation:

IMAGE CLASSIFICATION AND SEGMENTATION FOR EFFICIENT SURVEILLANCE APPLICATIONS

ABSTRACT

Image or video-based surveillance systems are playing a vital role in developing smart city and disaster management, such as flood and air pollution, etc. The need for the above surveillance systems is increasing exponentially. As a result, there is a demand for developing an accurate, efficient and safe system. There are existing systems for solving the above issues but the performance of the systems degrades or inconsistent for the different applications and situations. Besides, most existing systems do not aim to combine image processing and networking as one system for addressing the challenges. Therefore, there is immense scope for developing a new image-based surveillance system, which can cope with the causes of different applications and situations with minimum changes. To address the above research challenge, the proposed work is divided into three sub-challenges, namely, classification of images for developing a generalized system, segmentation for image size reduction and detecting region for safe landing for the purpose of safety. To solve the above challenge-1, in the past, the methods are developed, which include content-based image retrieval, scene categorization, and deep learning-based. The main issue of these methods is that the methods are limited to particular shapes of the objects in the images. In the same way, deep learning-based methods expect a large number of labeled samples and high computations. Therefore, the methods are limited to specific classification but not the classification considered in this work, which requires the generalized method. Thus, the proposed work aims at developing a new method for extracting edge strength and sharpness for classification of different image classes namely, soil, flood, air pollution, plant growth and garbage scene images. The reason to choose the above features is that these features can be used to extract unique observation in the images irrespective of objects shape. To address the

challenge-2, edge, texture, color and deep learning-based methods are proposed in the past. However, the methods are sensitive to background complexity and may not work well for the proposed image classes because each image can contain multiple colors, texture, etc. Therefore, the proposed work introduces a general saliency-based method for segmenting common region of the images. To find a solution to the above challenge-3, the existing methods extract texture, edges, and color for detecting flat region (safe landing zone) in the images. However, these methods are not adequate for the proposed images of complex background. Hence, the proposed work explores Gabor orientation responses for studying flat and rough region instead of magnitude values. The developed methods would be evaluated on different datasets to validate their performance.

Keywords: Classification; Segmentation; Surveillance applications

BAHAGIAN KLASIFIKASI DAN SEGMENTASI UNTUK APLIKASI

PENYELESAIAN KECEKAPAN

ABSTRAK

Sistem pengawasan berasaskan imej atau video memainkan peranan penting dalam membangunkan bandar pintar dan pengurusan bencana, seperti banjir dan pencemaran udara, dan lain-lain. Keperluan untuk sistem pengawasan di atas semakin meningkat secara eksponen. Akibatnya, terdapat permintaan untuk membangunkan sistem yang tepat, cekap dan selamat. Terdapat sistem sedia ada untuk menyelesaikan isu-isu di atas tetapi prestasi sistem tersebut tidak memuaskan atau tidak konsisten apabila digunakan untuk aplikasi dan situasi yang berlainan. Selain itu, kebanyakan sistem sedia ada tidak bertujuan untuk menggabungkan pemprosesan imej dan rangkaian sebagai satu sistem untuk menangani cabaran tersebut. Oleh itu, terdapat skop yang besar untuk membangunkan sistem pengawasan berasaskan imej yang baru, yang dapat menangani penggunaan aplikasi dan situasi yang berbeza dengan perubahan minimum. Untuk menangani cabaran penyelidikan di atas, kerja-kerja yang dicadangkan dibahagikan kepada tiga sub-cabaran, iaitu, klasifikasi imej untuk membangunkan sistem umum, segmentasi untuk pengurangan saiz imej dan mengesan rantau untuk pendaratan selamat untuk tujuan keselamatan. Untuk menyelesaikan cabaran-1 di atas, kaedah-kaedah yang telah dibangunkan pada masa lalu adalah termasuk imej berasaskan temu-balik kandungan (*information retreival*), pengkategorian latar, dan berdasarkan pembelajaran-mendalam (*deep learning*). Isu utama kaedah ini adalah, kaedah ini adalah terhad kepada bentuk tertentu objek dalam imej. Dengan cara yang sama, kaedah berasaskan pembelajaran-mendalam memerlukan bilangan sampel berlabel yang besar dan perhitungan yang tinggi. Oleh itu, kaedahnya adalah terhad kepada pengelasan tertentu tetapi tidak klasifikasi yang dipertimbangkan dalam kerja ini, yang memerlukan kaedah

umum. Oleh itu, kerja yang dicadangkan ini bertujuan untuk membangunkan kaedah baru untuk mengekstrak kekuatan dan ketajaman tepi untuk klasifikasi kelas imej yang berbeza iaitu tanah, banjir, pencemaran udara, pertumbuhan tumbuhan dan latar imej sampah. Ciri-ciri di atas dipilih kerana ciri-ciri ini boleh digunakan untuk mengekstrak pemerhatian unik dalam imej tanpa mengira bentuk objek. Untuk menangani cabaran-2, tepi, tekstur, warna dan kaedah berasaskan pembelajaran-mendalam dicadangkan pada masa lalu. Walau bagaimanapun, kaedah ini sensitif terhadap latar belakang yang rumit dan mungkin tidak berfungsi dengan baik untuk kelas imej yang dicadangkan kerana setiap imej boleh mengandungi pelbagai warna, tekstur, dan lain-lain. Oleh itu, kerja yang dicadangkan memperkenalkan kaedah umum berdasarkan-ketaraan (*general saliency-based*) untuk mengasingkan kawasan umum imej. Untuk mencari penyelesaian kepada cabaran-3 di atas, kaedah sedia ada mengekstrak tekstur, tepi, dan warna untuk mengesan kawasan rata (zon mendarat selamat) dalam imej. Walau bagaimanapun, kaedah ini tidak mencukupi untuk imej latar belakang kompleks yang dicadangkan. Justeru, kerja yang dicadangkan ini meneroka tindak balas orientasi Gabor untuk mengkaji kawasan rata dan kasar dan bukannya nilai magnitud. Kaedah yang dibangunkan akan dinilai pada dataset yang berbeza untuk mengesahkan prestasi mereka.

Kata kunci: Klasifikasi; Segmentasi; Aplikasi pengawasan

ACKNOWLEDGEMENTS

I would like to express my immeasurable appreciation and deepest gratitude to my supervisors Dr. Palaiahnakote Shivakumara, Associate Prof. Dr. Mohd Yamani Idna Idris and Associate Prof. Dr. Mohammad Hossein Anisi, for their kind support, guidance, and encouragement to enable me to complete this research on time and in the best possible way.

I would like to express my sincere thanks to Associate Prof. Dr. Mohammad Hossein Anisi and Prof. Dr. Mohsen Jahanshahi for providing me the opportunity of studying the Ph.D in the University of Malaya.

Last, and most important, my profound gratitude goes to my lovely family. Words cannot express that how grateful I am to my heroine, my mother Akram Shahmirzadi, my father Hassan, my sister Marzieh and my brother Mohammad Ali for their unconditional love, endless support and all that they did for saving my life three years ago. To them, I heartily dedicate this thesis.

This work is also dedicated to all dear scholar women struggling but pursuing their dreams in spite of all social implicit bias and inequalities against them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

2D/ 3D   :   2-Dimensional/ 3-Dimensional

AP   :   Air Pollution

B   :   Bit Plane

BS   :   Based/Transmitter Station

C   :   Node/Camera Node

CA   :   Canny Image/Map

CCTV   :   Closed-Circuit Television

CF   :   Cost Function

CH   :   Cluster Head

CNN   :   Convolutional Neural Network

CP   :   Candidate Pixel

CLR   :   Classification Rate

CR   :   Candidate Region

CSR   :   Compressed Significant Region

CT   :   Coast

CV   :   Covariance

d   :   Distance

DOR   :   Dominant Overlapping Region

DR   :   Detection Rate

DSC   :   Distributed Source Coding

e   :   minutes

ER   :   Error Rate

ES   :   Edge Strength

ET   :   Entropy

F	:	F-measure
FD	:	Flood
FOV	:	Field of View
FT	:	Forest
FV	:	Feature Vectors
fn	:	False Negative
fp	:	False Positive
g	:	Gradient Magnitude
G	:	Gabor
GE	:	Garbage
GPS	:	Global Positioning System
h	:	hours
H	:	Histogram
HF	:	Histogram Highest Frequency
HY	:	Highway
I		Input Image
IC	:	Inside City
IOT	:	Internet of Things
KNN	:	K Nearest Neighbor
k	:	Coefficient
l	:	Frequency of The Wave Propagating
L	:	Summation of the Pixel Value and Its Neighboring Pixels
m	:	Image Width
M	:	Mean
MN	:	Mountain
MCC	:	Markov Chain Code

n	:	Image Height
N	:	Number
OC	:	Open Country
P	:	Pixel
Pr	:	Probability
Pg	:	Percentage
PT	:	Plant
PSNR	:	Peak Signal to Noise Ratio
QoS	:	Quality of Service
ROI	:	Region of Interest
RGB	:	Red Green Blue
s	:	Second
SA	:	Saliency
S	:	Size
SD	:	Standard Deviation
SIFT	:	Scale Invariant Feature Transform
SH	:	Sharpness
SPIHT	:	Set Partitioning in Hierarchical Tree
SL	:	Soil
ST		Street
SSIM	:	Structural Similarity Index
SVM	:	Support Vector Machine
SW		Slepian-Wolf
tp	:	True Positive
tn	:	True Negative
T	:	Time

TB	:	Tall Building
TCP	:	Transport Control Protocol
TH	:	Threat
u	:	Current
UAV	:	Unmanned Aerial Vehicle
UDP	:	User Datagram Protocol
WSN	:	Wireless Sensor Network
VSN	:	Visual Sensor Network
WZ	:	Wyner Ziv
v	:	Voltage
V	:	Variance
ρ	:	Threshold
δ	:	Gaussian Envelope
∂	:	Directional Derivative
α	:	Angle
τ	:	Constant Value

CHAPTER 1: INTRODUCTION

1.1 Background

One of the smart city goals is to improve the quality of living through advanced technologies. To achieve this, surveillance technologies play a vital role and are integral parts of green city, eco city, safe city and digital city development (Jang and Cha 2014). For instance, visual based surveillance system provides monitoring services, such as crime identification, suspicious things identification and helps in disaster management. However, to use the visual surveillance systems efficiently for different situations and purposes, it is necessary to choose relevant information, reduce the data size and take system safety into account such that visual surveillance systems can save resources, such as energy, network lifetime and cost. Therefore, in order to diminish image data size, choose relevant information and keep system safety, image processing based methods play a crucial role. Hence, it can be concluded that visual surveillance requires image processing, at the same time, image processing requires visual surveillance application to expand the image processing strengths. In this thesis, classification of images for multipurpose surveillance applications, segmentation of the Region of Interest (ROI) to reduce size of the data, detection of Safe Landing Zone (SLZ) for flying drones are considered.

1.2 Importance of Images for Surveillance Applications

In this section, general surveillance applications are presented including the importance of visual monitoring based on video and image.

1.2.1 General Surveillance Applications

An intelligent surveillance system is a systematic process of real-time monitoring, analysis, and transmission of the data acquired by sensors for the purpose of managing, investigation and protection (Valera and Velastin 2005). Figure 1.1 shows a general

overview of the surveillance system combining multiple disciplines. The system includes data collection, processing, and transmission to conduct numerous monitoring tasks. The environmental scalar data such as humidity, temperature, pressure, pollution level, sound, and also visual data are sensed via different sensor node types. A node can be deployed carefully or randomly and can move and/or fly over an area of interest to collect data according to its defined acquisition plan. In most of the scenarios, there are large certain geographic sections such that an individual sensor node is neither sufficient nor reliable to cover the whole area of interest (Yan, He et al. 2003). In this case, multiple sensors cooperate to monitor the location with a higher degree of coverage and confidence. Wireless Sensor Network (WSN) have been developed including a large number of low-cost battery-powered collaborative sensing nodes as an advanced information sensing paradigm and the basis for intelligent surveillance systems (Mostafaei, Chowdhury et al. 2018). Afterward, a variety of processing methods can be applied to the sensed data to extract useful information according to the application requirements. The output of the processing stage provides the system with extracted features or alarm message to take immediate actions in the station. Even the node can decide independently in urgent circumstances without waiting for the station's command in a timely manner (Hampapur, Brown et al. 2003, Patterson, McClean et al. 2014). The information is transmitted through different communication technologies and channels toward the station (Memos, Psannis et al. 2018). Many surveillance applications such as security, protection, health, agriculture, environment, and urban monitoring would be provided based on the smart surveillance systems.
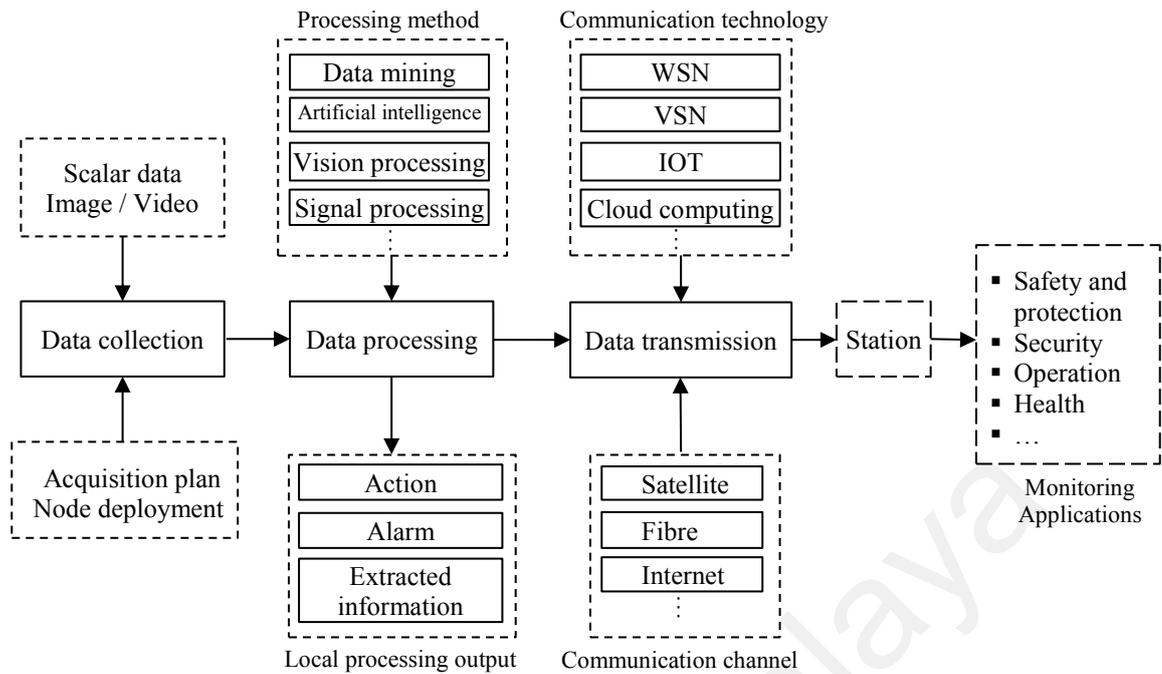
Figure 1.1: Surveillance system overview

1.2.2 Video based Surveillance Applications

In many surveillance applications, only scalar data collected by the traditional sensors is not sufficient. Moreover, in many scenarios, the area of interest is remote or dangerous to approach and deploy nodes. Another scenario is monitoring disasters like flood or earthquake requiring full 3D coverage immediately. To meet all these requirements, sensors are equipped with video sensors to provide that rich source of 3D information even from a very far distance and any desired viewpoint enhances decision-making process in the smart surveillance systems (Verma, Gautam et al. 2018). The video based surveillance systems have been evaluated technologically with analogue CCTV systems (Velastin and Remagnino 2006). In conventional video based surveillance systems, video streams continuously are sent towards a processing station to be verified by a human operator (Peixoto and Costa 2017). Human monitoring of surveillance video is prone to error and takes a very long time especially in surveillance scenarios covering large public areas for a long time (Hampapur, Brown et al. 2003). On the other hand, the lifetime of battery-operated nodes is limited. Given the huge amount of frame data sensed by the

3

cameras, transmitting all video data is very much costly in terms of network resources such as energy and bandwidth (Soro and Heinzelman 2009). Therefore, a network of smart cameras called Visual Sensor Network (VSN) has emerged in which camera nodes capture, locally process and transmit visual data. In VSN, smart nodes integrate the tiny cameras, embedded processor and wireless transceiver. After data acquisition, video processing techniques provide the system with descriptions of captured events (Soro and Heinzelman 2009). To this aim, firstly, temporal data of the frame sequences are processed and then in a higher level, key frames are analyzed to extract the information (Kim and Hwang 2002, Chandra, Couprie et al. 2018). The output will be sent to the station to monitor and track human, objects and events during the time. The local video processing significantly reduces data transmission burden throughout the network to enhance its lifetime as well as system intellectuality (Goswami, Paswan et al. 2016, Kumar and Priya 2018). However, processing huge video data locally on nodes for long time is neither efficient nor even possible. Figure 1.2 shows an overall framework of video based surveillance system.
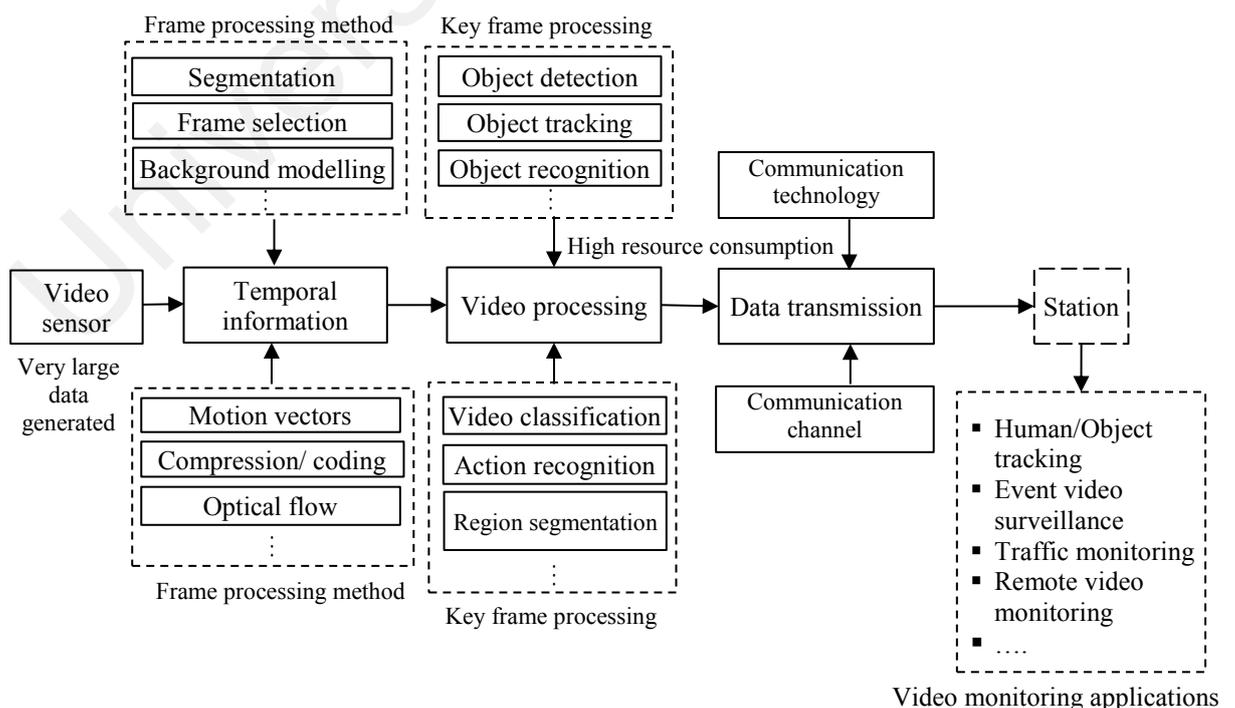
Figure 1.2: Video based surveillance system overview

1.2.3   Image based Surveillance Applications

As mentioned in the previous section, given the limited processing capabilities, bandwidth, battery power and small memory in the nodes, processing large amount of video stream data degrades performance of the system. This problem is even worse in surveillance scenarios since the nature of these applications requires a long lifetime monitoring. One approach to overcome to this problem is image based surveillance system in which images are taken periodically or other sensors can provoke the camera only once they sense any event instead of capturing all huge video data (Ahmad, Mehmood et al. 2017). After image acquisition, they are exposed to different processing methods such as compression, segmentation, detection, recognition and classification according to the application target (Ahmad, Mehmood et al. 2017, Lopez and Stilla 2018, Memos, Psannis et al. 2018).The output will then be sent as extracted descriptions or alarm to the station or it may assist the node to take an immediate decision accordingly (Patterson, McClean et al. 2014, Memos, Psannis et al. 2018). Therefore, image based surveillance system is a very important research topic since it does cope more with surveillance system constraints and can extend its lifetime and efficiency. However, designing image based surveillance systems still imposes a number of engineering problems which three of them in term of resources, energy (Sukumaran, Sankararajan et al. 2017) and safety (Patterson, McClean et al. 2014) are discussed in this study. Different image processing schemes are proposed to address these limitations. Figure 1.3 shows an overview of the image based system.
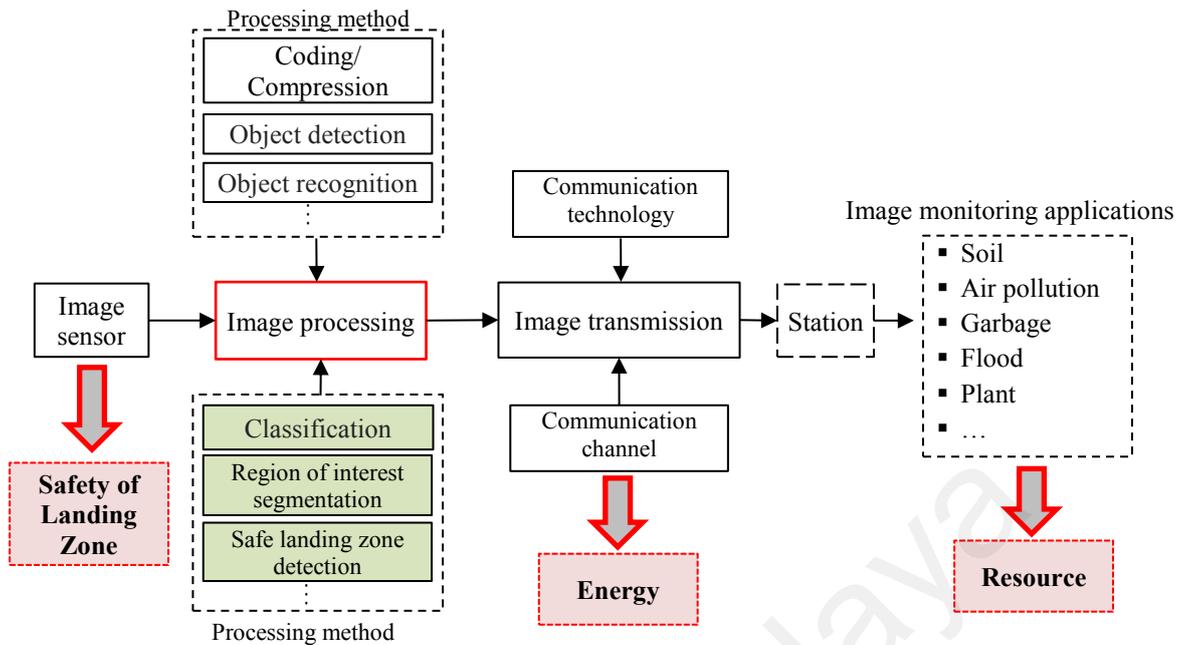
Figure 1.3: Image based surveillance system overview

1.3   Motivation

It is noted from the previous sections, there are several real time applications that require the design of an efficient and accurate surveillance system for transmitting image data. In addition, transmitting the whole image through a single purpose network is not advisable in terms of surveillance systems performance. Therefore, there is a need for classification of the different type of images and segmenting ROI to save system resources. In case of monitoring during a disaster, it is necessary to find a safe zone for landing drones or small helicopter. Therefore, the proposed work helps surveillance system to detect flat zone for a safe landing. This section presents the importance of the above issues to improve the surveillance system performance.

1.3.1   Classification of Images

Nowadays, there is a great interest in surveillance systems especially in applications related to natural disasters, such as flood and air pollution monitoring (Messer and Sendik 2015). Additionally, plant growth and soil monitoring for the purpose of studying the health of plants and identifying diseases as well as garbage identification to control

dangerous diseases such as Dengue, and Malaria are important applications. However, once these surveillance applications change, the systems have to be re-designed again with the help of human intervention. This is not feasible for smart city projects, where there is infinite number of applications (Jang and Cha 2014). For example, the set-up designed for soil monitoring cannot be adapted for monitoring flood and garbage without human intervention. In this situation, it is necessary to develop an automatic system to facilitate surveillance system to monitor different classes, which results in the saving of resources, manpower, cost, time, energy, etc. Therefore, there is immense scope for processing images to classify them such that we can use the same developed network design as illustrated in Figure 1.4., where we can see image processing playing a key role in identifying image class before sending it to a networked system. It is noted from Figure 1.4 that if an image processing system does not identify the image type or class, the designed network transmits the image as it is without classification because the surveillance system is designed for a specific task or image class but not multiple classes.

It is evident from the study by (Messer and Sendik 2015) that a generalized design for a surveillance system for different situations is challenging. Davis, Liang et al. (2012) designed a network for soil analysis. Shaban, Kadri et al. (2016) proposed a networked system for monitoring air pollution. Afsharinejad, Davy et al. (2016) proposed a system for plant monitoring. In (Longhi, Marzioni et al. 2012), the author proposed a WSN for garbage/waste management. El Bastawesy, Ali et al. (2012) proposed an image processing technique to analyze soil images. Also, a Bayesian-based technique for flood detection was developed in (D'Addabbo, Refice et al. 2016). ZainEldin, Elhosseini et al. (2015) designed an edge preserving-based technique for haze image detection. In (Guerrero, Guijarro et al. 2013) an image processing-based technique is presented for crop row detection in maize fields. Furthermore, the study by (Rao and Kumar 2012) dealt with a computer-aided diagnosis for Dengue fever detection. Therefore, the

currently available systems are developed for particular dataset and class. This was the motivation to focus on image processing to identify the different classes such that the same surveillance system can be used for several applications to save time, resources, energy etc.
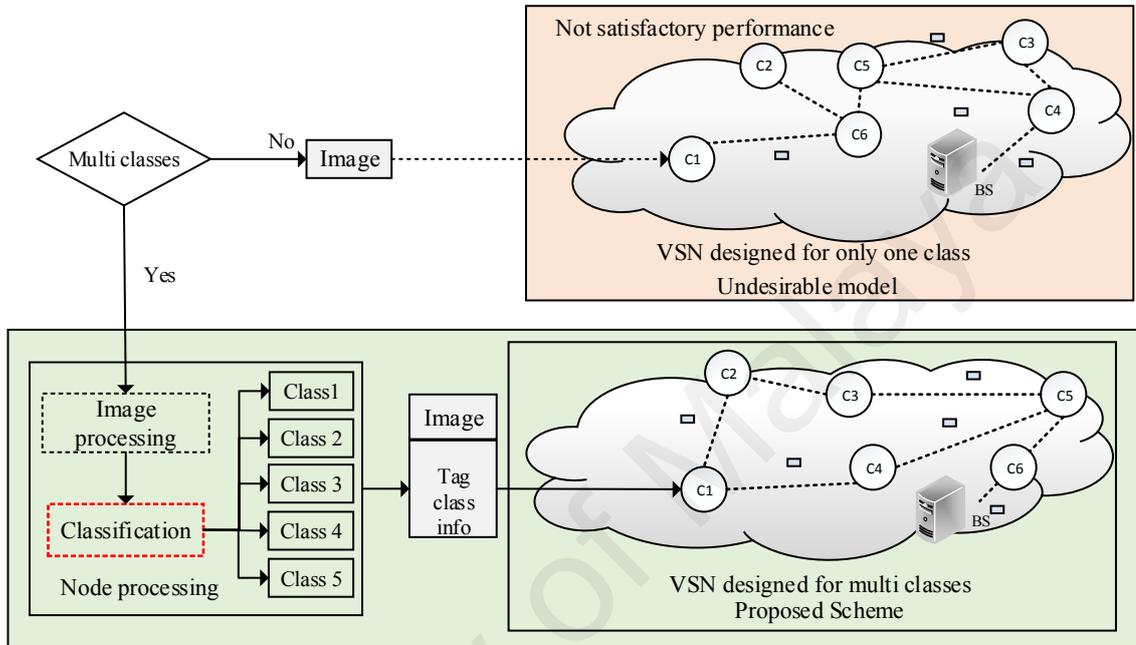
Figure 1.4: Need for image type classification to design a smart surveillance system

1.3.2   Segmentation of Region for Image Size Reduction

Expanding VSN to surveillance monitoring is an interesting but challenging task as these applications require generating continues high quality information while VSN has inherent limitations of energy, bandwidth, and life of the nodes. These applications involve several transmissions of images and as network coverage increases, the number of captured images and the complexity of resource management increases, significantly. Moreover, the transmission of multimedia data without losing quality requires high energy and network resources (Lin, Rodrigues et al. 2011, Mammeri, Hadjou et al. 2012, ZainEldin, Elhosseini et al. 2015). Therefore, there is immense scope for image size reduction such that a network can use resources efficiently for longer life and less information loss in critical surveillance applications. There are numerous methods in the

literature to find solutions dealing with VSN constraints (Lin, Rodrigues et al. 2011, Bhandary, Malik et al. 2016, Liu, Sridharan et al. 2016, Shen and Bai 2016, Al-Ariki and Swamy 2017, Yap and Yen 2017). Most of the methods focus only on network parameters and requirements but not images and they treat images just as black boxes or packets for transmission. However, VSNs consume more power for data transmission compared to image acquisition and pre-processing (Lu and Manduchi 2011, Güngör and Hancke 2013, Nirmala, Vignesh et al. 2013, Ozger, Fadel et al. 2016). Therefore, these methods that do not focus much on removing unwanted information to reduce the large size of visual data, will lead to decreasing network lifetime. This fact is the motivation to propose new image size reduction approaches in this study. An example is provided in Figure 1.5, where we can see when the color of images are passed through the general network, there are high-energy consumption and high chances of information loss due to huge size of image, which split into several packets during the transmission. The problem can be alleviated once the segmented ROI instead of the whole image is transmitted.
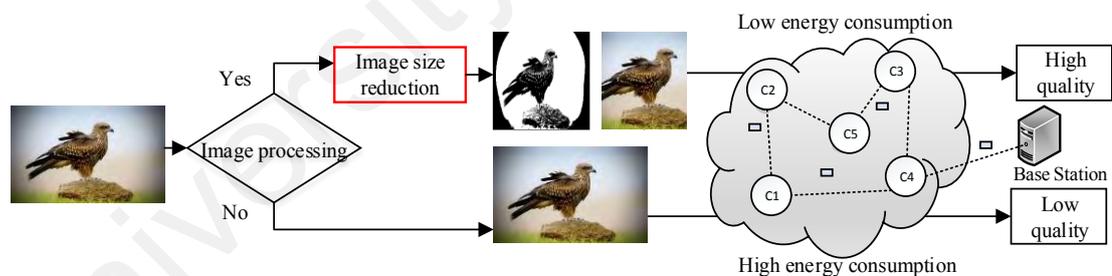
Figure 1.5: Need for region of interest segmentation to design a smart surveillance system

1.3.3 Detecting Zone from the Segmented Region of Interest for UAV Safe Landing

Unmanned Aerial Vehicle (UAV) is commonly known as drone that indicate the aircraft platforms without a human pilot onboard (Huang, Chen et al. 2016). UAVs are used for surveillance of many cases such as military and battlefield reconnaissance, disaster and damage assessment, border and environment exploration due to low cost,

high mobility, flexibility, safety and customizability (Patterson, McClean et al. 2010, Lee, Morton et al. 2017, Al-Kaff, Martín et al. 2018). The observation and monitoring of the above-mentioned situations have become much easier and faster because UAVs can reach an area of interest in a short time (Fan, Lu et al. 2018). It is true that usually, UAVs perform tasks according to the instructions given by a planned path using a satellite-based system, such as the global positioning system (GPS). However, there are situations where GPS may not function well due to adverse and poor environmental factors which affect the signal strength and reliability (Patterson, McClean et al. 2010, Garcia-Pulido, Pajares et al. 2017). As a result, one cannot expect GPS based systems to work well at all the times. More details can be found in (Lee, Morton et al. 2017). On top of this, sometimes, static and dynamic obstacles, engine failures and security attacks may create emergencies in which UAV is not able to recover or even go back to the station. In order to find a solution controlling these critical situations, one needs an alternative way of finding landing safe sites with the help of image processing techniques. This urgent need is the motivation of this study in order to propose a new segmentation model to address safety requirement of surveillance systems as well. Figure 1.6 shows an example in which a UAV faces emergencies and looks for a safe zone to land.
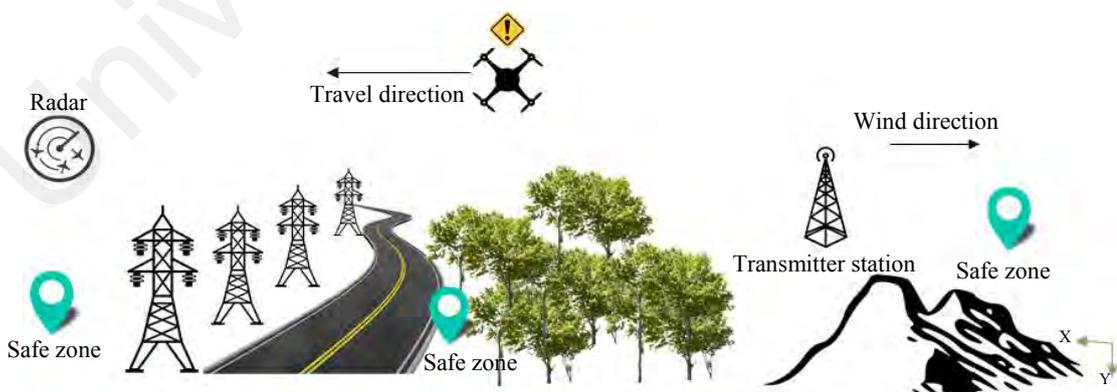
Figure 1.6: Example of emergencies whilst a UAV is flying

1.4   Problem Statement

In the previous section, the importance of three image-processing issues for enhancing the surveillance systems is discussed according to different real time applications. Based on the discussion, one can list many challenges for classification, segmentation and safe landing for UAV detection with respect to surveillance applications, which are elaborated in the following section.

1.4.1   Challenges of Classification Methods

The deep learning based scene image classification methods that have been developed recently (Zuo, Wang et al. 2015, Bai 2017, GOOGLE API) or most of other existing methods, are defined according to the shapes of the objects in the images (Bosch, Zisserman et al. 2008, Dunlop 2010, Du and Ling 2016, Qin, Shivakumara et al. 2016). For instance, the method proposed in (Nogueira, Penatti et al. 2017) is used to extract features that represent outdoor elements such as sky, ocean, mountain, tree for classifying outdoor images. Similarly, the methods in (Hayat, Khan et al. 2016) extracted features that represent the indoor environment, such as objects and humans for classifying. In spite of the ability to solve complex issues, deep learning-based methods, cannot be used directly for scene images of this study namely soil, plants, flood, air pollution and garbage. This is because, it is hard to define the shape of objects in these images. For example, air pollution images may not contain any objects or it may contain tall buildings, street, sky, ocean, etc. Similarly, garbage images may contain stagnant water with waste and different trash items, which do not have any particular shapes. As a result, learning becomes complex for deep learning-based methods. Therefore, classifying the above-mentioned scene images with respect to different situations for monitoring systems is both challenging and interesting. Figure 1.7 shows samples of misclassified images of some existing methods.

Flood as Plant	Air pollution as Flood	Garbage as Soil	Soil as Flood
(a) Bosch, Z et al. (2008)		(b) Dunlop, H. (2010)	

Soil as Plant	Garbage as Soil	Garbage as vehicle	Air pollution as Hand
(c) Qin, L et al. (2016)		(d) Google API	

Figure 1.7: Sample of misclassified results by the existing methods

1.4.2 Challenges of Image Size Reduction Methods

There are some methods for image size reduction in the literature, which focus on coding, decoding criteria to reduce image size such that they can be retrieved accurately and efficiently at the station (Yan, Zhang et al. 2014, Sastra and Hendrantoro 2015, Rein and Reisslein 2016, Sofi and Naaz 2016, Paek and Ko 2017, Yan, Xie et al. 2018). However, these methods consider the whole image for processing and hence there is not much reduction in image size. Besides, as size of the image decreases, there is potential for losing quality. In order to reduce size without losing quality, there are techniques to binarize the images, which results in great image size reduction as shown in Figure 1.8. One such approach divides color image into bit planes, which provides binary information (Felemban, Sheikh et al. 2014). The size of the binary planes is much lower than that of the color images. However, sometimes, binarized planes may lose quality and significant information due to large variations in the images. In addition, there is no guarantee that always the plane of the most significant bit always provides significant information because the plane information depends on the complexity of the images. To alleviate this problem, these methods are developed for significant plane detection (Dutta, Mandal et al. 2007, Chen, Ma et al. 2017, Raghunandan, Shivakumara et al. 2018) while their

methods are not adequate to deal with the complexity of scene and general images as shown in Figure 1.8 (b)-(d). In other words, the existing plane detection schemes work well for specific images with *a prior* knowledge of the images and dataset.

Sample classified images

(a) The most informative bit plane

(b) Detected bit plane by Felemban, Sheikh et al. (2014)

(c) Detected bit plane by Raghundan et al. (2018)

(d) Detected bit plane by Dutta et al. (2007)

Figure 1.8: Sample of binarized image by the existing methods

For complex network systems where many cameras are mounted, one can still expect redundant and unwanted information between images because of the common regions given by multiple cameras. There are methods which reduce the image size considering these overlapping regions as redundant information. For example, the methods presented in several studies (Wang, Li et al. 2007, Imran, Khursheed et al. 2010, Wang, Peng et al.

2010, Jayshree, Biradar et al. 2012, Khursheed, Ahmad et al. 2012, Imran, Khursheed et al. 2013, Coşar and Çetin 2015) identify the common region with the help of cameras' fields of view as redundant information. These methods save node's energy compared to the previous methods working on single images and intra redundancies. However, for complex network systems including many cooperative cameras, the trade-off between network limitations and the image quality is not taken into account well. Figure 1.9 shows sample results of methods presented in (Wang, Li et al. 2007, Wang, Peng et al. 2010). Therefore, developing a new method, which can achieve both image size reduction and high quality is very challenging.

(a) Sample image   (b) Wang. D et al (2010)   (c) Wang. D et al (2007)

Figure 1.9: Sample image size reduction result of existing methods

1.4.3 Challenges of UAV Safe Landing Zone Detection Methods for Disaster Management

Finally, in the case of surveillance applications using UAVs, there are existing systems to handle the emergency situations which follow the steps as shown in Figure 1.10. During emergencies as mentioned in Figure 1.10, an automatic system called a fault detection unit sends an alert message to another unit for searching out safe zones.
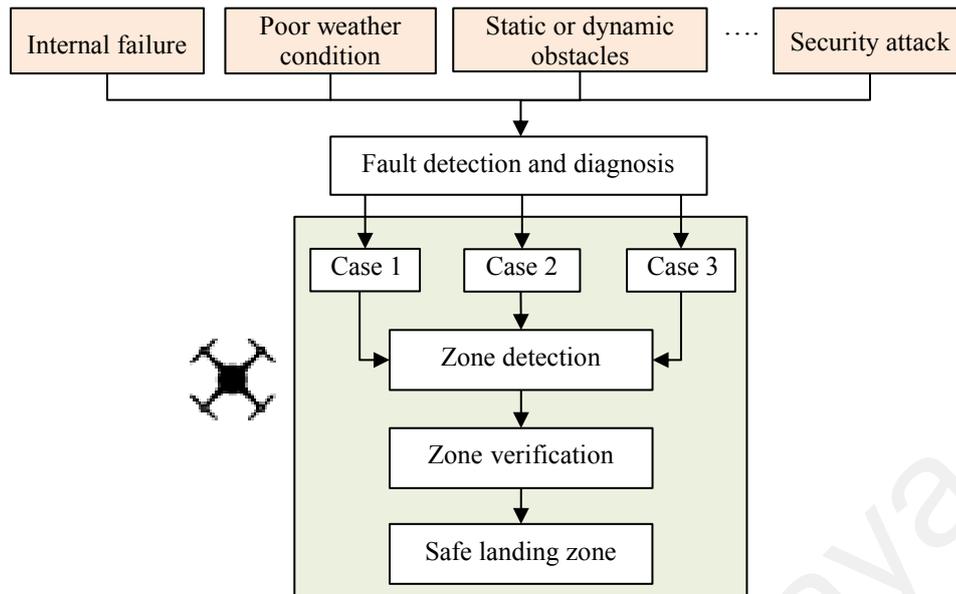
Figure 1.10: General framework for safe landing of UAVs

To identify a safe zone for landing, existing image processing-based systems focus on segmentation of regions of interest (zone detection), such as greenery, forest, river and mountains. Then, the systems classify segmented regions to find safe and flat zones; this can be considered as zone verification for the detected zones. In Figure 1.10, case 1, case 2 and case 3 refer to emergencies created due to communication failures from ground stations, GPS failures and software-hardware/energy failures, respectively (Patterson, McClean et al. 2014). It is noted that defining regions of interest and the classification of particular regions (zones) is good when we have a limited number of regions. However, it is not necessarily true for all situations, where we can anticipate unexpected regions such as buildings, towers, small plants, and flat regions with cars or other objects. In addition, one can say that shapes and the nature of regions could be infinite and therefore methods depending on some specific features cannot segment all safe sites. It is evident from the sample results of an existing system (Patterson, McClean et al. 2014) in Figure 1.11. In this Figure, different input images in (a) and white pixels representing safe zones in (b) can be seen, the existing system fails in the case of the fourth image and does not detect accurately the first three images shown in Patterson et al, (2014)

Figure 1.11 (c). The main reason is that the existing systems have inherent limitations of segmentation and classification and the features are specified. Therefore, segmentation of safe region irrespective of the image content dealing with unexpected landing terrains is challenging.

(a) Input images captured by UAV of different regions

(b) Safe land segmentation

(c) Safe land segmentation in Patterson et al, (2014)
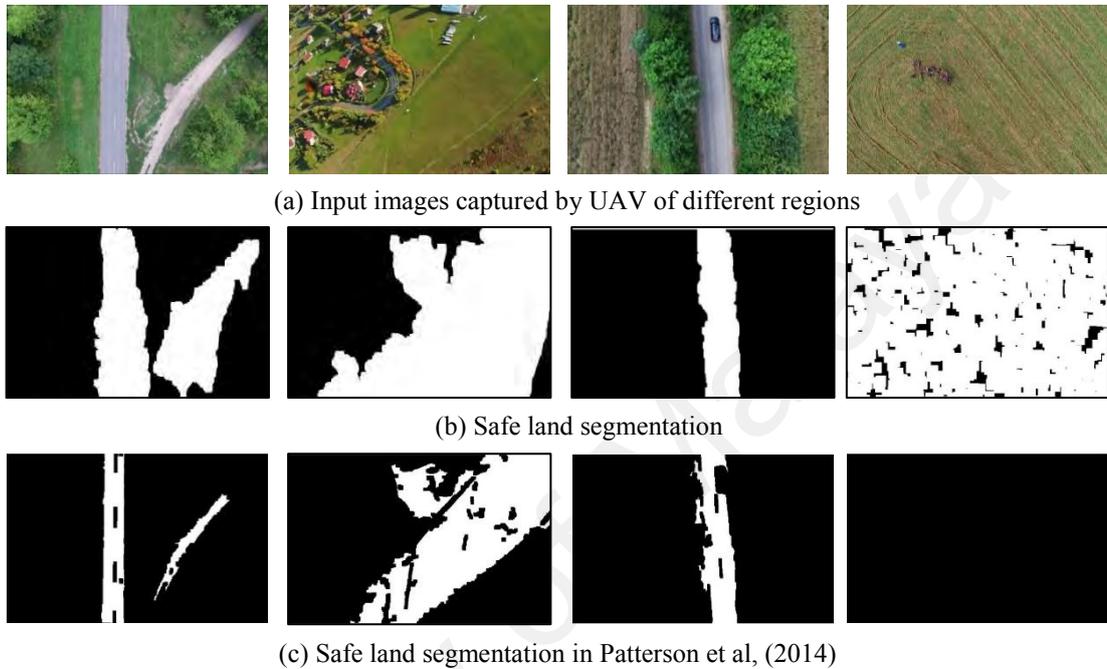
Figure 1.11: Challenges of safe land segmentation.

1.5   Research Questions

In order to find solutions to the problems mentioned in the previous section, the following are the research questions to be framed.

- How to extract the features for multi-application classification with respect to different surveillance classes?

- How to segment region of interest to reduce image size such that visual surveillance network can save resources?

- How to detect UAV safe landing zone for rescuing UAV from the segmented region of interests?

1.6   Research Objectives

To address the above problems and questions as discussed, the following objectives are set to achieve the goal as are depicted in Figure 1.12.

1. To propose an image classification technique based on detecting the information that represents focus edges through sharpness and edge strengths.

2. To propose a segmentation of ROI method by finding the information that represent dominant information in the images.

3. To propose for safe zone landing detection method from segmented ROI by studying direction based responses of flat and rough regions.
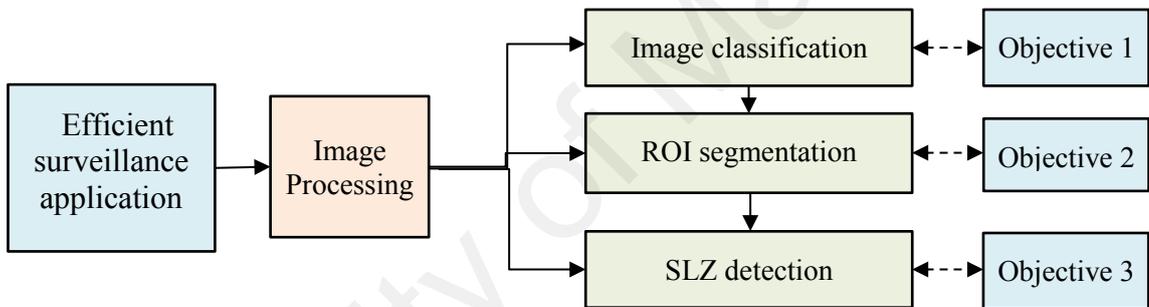
Figure 1.12: The thesis objectives overview

1.7   Significance of Research

In the first objective, five image classes are chosen, namely: (1) soil, (2) plant, (3) air pollution, (4) flood and (5) garbage. The research on these five classes is essential since, in Malaysia, most ubiquitous surveillance systems are installed for monitoring of the above situations. In this country, due to unpredictable weather and rain, it is hard to grow crops, which yields a profit for farmers. According to the agricultural department, there is a need to monitor soil and plants of the crops to protect the crop Afsharinejad, Davy et al. (2016). Similarly, for the same reason, unexpected floods destroy the crops in the fields and people's lives. In the same way, air pollution due to haze, smoke and fire is hampering the public in Kuala Lumpur. Recently, dangerous diseases such as Dengue and Malaria

are quite common in Malaysia as well as in the world. To control and prevent such disease outbreaks, there is a need for a system to identify garbage and waste. Therefore, developing a multi-purpose surveillance system for these classes is essential for Malaysia as well as other developing countries as their future target is to develop smart cities.

In the second objective, these classified images are not desirable to be sent simply through the network since it causes high energy consumption and decreasing network lifetime. This is one of the main challenges of VSNs in surveillance applications. Therefore, designing a comprehensive scheme to segment informative regions considering the quality of image can improve and enhance the efficiency of the system significantly leading to monitoring the area of interest for a longer time.

Finally, in the third objective, in order to have an efficient surveillance system, not only resources and nodes' lifetime are required to be taken into account, but also the safety of nodes, properties and people in the system is very important. UAVs flying and monitoring the area of interest may face emergency situations in which they need to decide without a human assistance as soon as possible. Otherwise, it may be too late or impossible to wait for a command from the base station. Therefore, developing a detection model to find the safest zone regarding the emergency types would increase the safety of system so that the node can monitor as long and safe as possible.

1.8   Layout of the Thesis

The organization of the chapters in this thesis is as follows:

- In Chapter 1, a preview of the whole thesis has been provided including the motivation of undertaking this research, research challenges, problem statement and research questions, objectives, thesis contribution and significance of the research.

- In Chapter 2, the existing methods of scene image classification, image size reduction, bit plane detection and safe landing zone detection are reviewed in an elaborate manner in order to understand the necessity of proposing new classification and segmentation schemes.

- In Chapter 3, the proposed classification algorithm is technically explained including edge strength and sharpness extraction and integration. Templates and rule based classifications are explained. In the experimental section, key steps of the proposed method, then SVM, KNN and random forest classifiers and comparative study are elaborated.

- In Chapter 4, the method for image size reduction is presented, which include the algorithm for Dominant Overlapping Regions (DORs) detection and partitioning for optimal transmission. This method works well when the image contains overlapping regions. To alleviate this limitation, the proposed work introduces bit plane detection for image size reduction. The bit plane detection method works based on saliency detection and canny edge detector. The results are provided for evaluating the proposed and existing methods.

- In Chapter 5, the proposed method for safe landing zone is illustrated including Gabor transform, Markov Chain Code (MCC) process, similarity estimation and Chi square distance. The experimental part reports the validation of parameters and detection rate for safe landing zone in comparison with existing methods in different emergency circumstances.

- In Chapter 6, a summary of the main contributions is given, afterward, limitations and future works are explained.

CHAPTER 2: LITERATURE REVIEW

2.1   Background

In the previous chapter, the importance of the problem, applications and need for solutions to the issues are presented. This chapter presents the review of the existing methods for classification, segmentation and safe landing zone detection to understand the state-of-the-art methods, which helps to propose new methods for finding solutions to the issues.

2.2   Image Classification based Methods for Surveillance Applications

Classification is a fundamental and common problem in computer vision for video and image understanding. It provides the basis for many high-level vision tasks. There are mainly three categorizations of classification approaches; video, scene image and visual scene image based classification methods. Since the scope of the first objective is limited to classification of images of different types, this section presents a review of the mentioned categories.

2.2.1   Video based Classification Methods

Video as a collection of images called frames can be combined to get the original video. In addition to the spatial features presented in images, video has the property of temporal features. So, a problem related to video data is not that different from an image classification problem else one additional step of frames extraction from the video. Video classification has been progressed using static single-frame based visual features as well as dynamic information such as motion and changes of scenes to take the time dimension into account (Zhang, Mei et al. 2018). For example, Tian, Sun et al. (2016) proposed keypoint trajectory coding on a compact descriptor for video analysis. In this work, keypoints given by the Scale-invariant Feature Transform (SIFT) descriptor are explored for video classification. The method works well when an image contains high contrast.

For the images with low contrast, the method does not perform well. A scene classification of images and video via semantic segmentation is developed in (Liu, Chen et al. 2016). The method segments shots for classification of indoor and outdoor scenes. The performance of the method relies on the segmentation results. In other words, if the segmentation is successful, the method gives better results. However, segmenting region of interest is difficult because images may not contain the objects with specific shapes. Liu, Chen et al. (2016) proposed video classification via weakly supervised sequence modeling. This method combines multiple instance learning and conditional random fields for classification. The method requires a large number of samples for training. Therefore, it works well for limited applications. A new shape feature for vehicle classification in thermal video sequences is proposed in (Yang and Park 2016). The method uses target trait context features for classification. The method focuses on extracting object shapes in the images. The performance of the method depends on the shape of the objects. It is not necessarily true for the proposed problem in this research. Moreover, there are deep learning based video classification methods. For instance, Karpathy, Toderici et al. (2014) proposed large-scale video classification with convolutional neural networks. This method extracts local spatio-temporal information for achieving classification. A temporal action localization in untrimmed videos via multi-stage CNNS is developed in (Shou, Wang et al. 2016). This work focusses on actions in the image. Wu, Fu et al. (2016) proposed harnessing objects and scene semantics for large-scale video understanding. It combines three streams of information using three-layer neural networks which require large training dataset. A deep spatio-temporal importance prediction in driving videos is presented in (Ohn-Bar and Trivedi 2017). This work aims to understand the context of the surrounding agents. Qin, Shivakumara et al. (2016) proposed video scene text frame categorization for text detection and recognition. This work is similar to the proposed classification method

because the classification is performed in order to improve text detection and recognition in the same way; here a classification method for improving surveillance system performance is proposed. However, its performance depends on fine edge in the images. In the case of video-based classification methods, since the methods works based on the fact that video provides temporal information for estimating motions and time sequence. Besides, the above process involves several duplicate frames, the methods are computationally expensive. Therefore, the methods use keyframe or still frames or images for processing to understand scene images, which will be discussed in the subsequent section.

2.2.2 Scene Image based Classification Methods

Scene image classification is a vital process in the computer vision system to understand the surrounding areas such as mountains, forests or office rapidly and effectively. However, classifying scene images is not a simple task due to their variability, ambiguity, and the wide range of illumination and scale conditions that may apply (Lang, Xi et al. 2014). There are many existing methods to address this problem. For example, a scene classification using a hybrid generative/discriminative method in (Bosch, Zisserman et al. 2008) is developed. This method explores probabilistic latent semantic analysis and SIFT features in a different color space. This method is good for the images, which contain objects with clear shape. Du and Ling (2016) designed a dynamic scene classification using redundant spatial scenelets. The method defines scenelets, which represent the unique properties of the scenes in the images. The spatial relationship between the scenelets is used for classification of the images. When the images have objects with clear shapes, it is easy to define spatial relationship otherwise, it is hard to define. Hayat, Khan et al. (2016) proposed a spatial layout and scale invariant feature representation for indoor scene classification. This method is limited to indoor scene images but not the scene complex images considered in the proposed work. This is valid

because when the images have a complex background, it is hard to find the shape of the objects in the general images. Zhu, Wu et al. (2016) presented a reconfigurable tangram model for scene representation and categorization. The method defines primitives based on the shapes of the objects for classification. This method also expected objects with a clear shape in the images. Sun, Liu et al. (2018) proposed perceptual multi-channel visual features for scene categorization, which explores multiple visual features at both low level and high level for classification. A kernel based human gaze estimation technique is designed to find the regions that human attends within an image to classify it. The extracted features feed to a SVM classifier. However, the proposed techniques depend on the clear objects in the image attracting attention. A bag of feature-based method for image classification is designed in Xie, Tian et al. (2016). This method explores local and spatial-based features for classification. The method works well for high contrast images. Since the proposed work can have images with contrast variations, it may not perform well for the dataset considered in this research. Scene classification is proposed by using feature and kernel combination with adaptive weights in Yuan, Chen et al. (2015). The method uses multi-category classifiers for classification. Since the method involves multiple learning stage, it is limited to specific applications. A multi-scale context for scene labeling via a flexible segmentation graph is developed in (Zhou, Zheng et al. 2016). It segments regions of interest and then finds spatial relationships between the segmented regions for classification. Since the images considered in this work for classification do not provide clear shaped objects, the method may not perform well. In (Shrivastava, Bhoyar et al. 2017) a scene classification system is built inspired by the perceptual ability of human vision. The structure and content of a natural scene are extracted via dominant color, direction, openness and roughness features. Afterward, they are inputs to the process of distance evaluation and the results as the most discriminating features are feed to a SVM classifier. This work classifies images with predefined and

specified features. In addition, there are methods that focus on different types of deep learning to address scene image classification problem. For example, Growing random forest on deep convolutional neural networks for scene categorization is proposed in Bai (2017). This method explores deep learning for the extraction of features and then extracted features are fed to random forest for scene classification. The method requires a large number of predefined samples for labeling. With different images of different applications, it is hard to adapt this method. An exemplar-based deep learning, as well as discriminative and shareable feature learning for scene image classification is presented in (Zuo, Wang et al. 2015). Moving towards better exploitation of convolutional neural networks for remote sensing scene classification is proposed in Nogueira, Penatti et al. (2017). This work focuses on a fine tuning method and parameters rather than increasing the number of samples to achieve better results. Liu, Wang et al. (2017) proposed a survey of deep neural network architectures and their applications. According to the discussions in this study, despite deep learning approaches being able to help in solving complex issues, they have inherent limitations (Sharma 2015), such as generalized framework, a large number of training samples, and optimizing parameters to name a few. The above methods works based on an institution that the shape of the objects in the images provides cues for classification. However, one cannot expect the correct shapes of the objects in the images considered in this work.

2.2.3   Visual Scene Image based Classification Methods

In addition to general scene classification methods, many visual scene image classification models developed for enhancing monitoring services efficiency (Cohen, Afshar et al. 2018). For example, classifying and assigning different priorities to images can assist VSN to manage the visual data flow (Obraczka, Manduchi et al. 2002, Khedo,

Perseedoss et al. 2010). This section focus on reviewing the methods, which proposed specifically for classification and enhancement with respect to monitoring.

Balla-Arabé, Gao et al. (2014) proposed classification of remote sensing imagery to segment and label area of interest such as the active volcano or the flooded area in the image. This work applied multi-kernel implicit curve evolution for selected texture region segmentation in VHR satellite images. However, the method is good for high resolution and high contrast images. Since the proposed classification scheme can have images with contrast variations, the method may not perform well for general images. Romero-González, Martínez-Gómez et al. (2017) developed an indoor scene classification model for improving robots understanding so that they behave and interact with its environment smarter. An image descriptors generation approach is designed based on the use of dense SIFT features and spatial pyramid representation to obtain a set of initial descriptors as the input of SVM. They focus on size and descriptor quality to deal with real time robotic platforms scenarios. Verma, Sharma et al. (2018) also proposed another indoor scene classification method for improving artificial intelligent applications like remote navigation. Features descriptors like texture, pattern and color and Histogram of Oriented Gradients (HoG) are formed the input of a classifier. Another real-time indoor scene classification method of mobile devices is proposed in (Wald, Tateno et al. 2018). Input depth images in a global 3D model provide geometric segments and semantic annotations and map. Different incremental descriptors such as geometry, color or texture encode these properties. Random forest classifier is used to predict the semantic label and class of each frame. A visual classification method for nuclear waste material is designed in (Shaukat, Gao et al. 2016). Objects are characterized by their 2D shape representations and using a random forests learning classifier. All these proposed feature descriptors do not work for the target complex images of outdoor surveillance applications due to unlimited nature of their content. A monitoring system of outdoor environments for

surveillance applications is proposed in (Teschioni, Oberti et al. 1999) to classify vehicles, village and civilian pedestrians in a zone in order to provide alarms accordingly. Features are modeled but limited based on objects, shape and color such as colorful and orange dress for the village people class. Mora, Lopera et al. (2018) presented an automatic surveillance system for smart parking; where a binary classification algorithm is designed to check whether a parking lot space is occupied or not. First, a method based on bag of features such as color, corners, edges and SIFT algorithm extracts keypoints for feeding SVM classifier which is limited to their predefined features. Then, the results are feed to a CNN classifier. Moreover, Liu, Lau et al. (2016) designed a crowd scene classification algorithm to analyze pedestrians' behaviors in real-world crowd scenes. An exemplar-based motion model is designed to extract motion features at the individual and the holistic levels. Crowd features are extracted based on an iterative optimization algorithm, an Extended Kalman smoother and KL-divergence. However, the features are not applicable to static images with no motion.

In summary, the existing classification methods are developed for a particular image type and dataset containing objects with a clear feature. The methods assume that image contains objects with clear shapes for classification. This is not true for the images considered in this work where it can be expected unpredictable shapes in the images. Therefore, classification of visual scene images of this study remains a challenge.

2.3   Segmentation for Image Size reduction

In this section, the existing image size reduction methods for VSN are reviewed. The literature on image size reduction can be classified widely as intra-redundancy based and inter-redundancy based methods. Since the proposed work focuses on visual scene applications in large geographical areas such as watching animal behavior in forests and monitoring floods requiring both of these classes, this section reviews these redundancy-

based methods.

2.3.1 Intra Redundancy Based Methods

The intra-redundancy based methods consider a single image and propose methods to remove unwanted information within the same image. These methods are divided into two parts, compression based and region of interest based methods as are explained in the following sub-sections. They are effective in simple situations where only one camera is mounted for specific applications.

2.3.1.1 Compression based methods

Image compression refers to the task of encoding or converting an image in a way that it consumes less memory than the original file without affecting its quality to a greater extent. There are many existing methods attempting to reduce a single image size via image compression. Paek and Ko (2017) proposed k-means clustering-based learning by selecting common colors such that the image size is reduced. However, the method ignores the quality of the images as the size of the image is reduced. Rein and Reisslein (2016) proposed a scalable tree-based wavelet image compression algorithm to reduce the size of the image for transmission. Since the method uses a wavelet transform, there is a risk of losing significant information due to a blocking effect. In (Sofi and Naaz 2016) several combinations of wavelets is proposed for compression to save energy consumption and lifetime of the network. The method does not focus on image quality. Similarly, the work in (Sastra and Hendrantoro 2015) proposed an improved JPEG 2000 compression algorithm that selects optimal parameters such as bit rate of the network and DTW level to reduce energy consumption. However, this method does not change the parameters according to the quality of the image. Aurangzeb, Alhussein et al. (2018) presented a combination of bi-level image compression standards with a variety of change coding for sizes and locations of white objects in adjacent frames. Their analysis shows

these joint compression methods reduce more image size and it is better than image coding alone in VSN. Although the impact of illumination and noise are tested, the quality analysis is ignored. Zhang, Gu et al. (2018) developed a compression scheme for the transmission of remote sensing images given resource limitations. A deep learning based model is designed fusing the high frequency and low frequency information in a nonlinear scheme to handle complicated image compression artifacts while gaining visual quality. A lossy image compression scheme based on deep learning approach is proposed in (Fan, Lin et al. 2019) to predict satisfied user ratio curves. A Siamese convolutional neural network, feature pooling, a fully connected regression-head, and transfer learning are applied to achieve this aim. Another convolutional network (CNN)-based lossy image compression system is developed in (Li, Zuo et al. 2018). The binarizer is adopted to quantize the output of encoder and a proxy function is introduced for approximating binary operation in backward propagation to make it differentiable. However, in these methods image size is not taken into account carefully and the quality of all pixels have the same priority which is not an applicable solution in VSN systems. More compression methods designed for VSN are survived in (Mammeri, Hadjou et al. 2012, Suseela and Phamila 2018).

It is noted from the existing literature that the compression based methods developed for the purpose of networking systems consider the whole image for processing. As a result, the methods are expensive and unfeasible in terms of cost and energy consumption for real-time applications, such as animal surveillance, monitoring floods, garbage. The methods works based on the finding redundant information in the images. When the image does not contain redundant information, the performance of the methods degrade.

2.3.1.2　Region of Interest based Methods

To overcome the issues of the compression methods, there are other methods which

identify the region of interest with the help of feature extraction. The methods transmit only regions of interest (ROI) rather than transmitting the whole image or compressed image as in the previous methods. For example, in (Redondi, Baroffio et al. 2013) the method uses BRISK features for identifying ROI and sends it to a compression algorithm before passing it to the network. Similarly, the method proposed in (Van Opdenbosch and Steinbach 2018) extracts ORB, BRISK and FREAK features to avoid transmission of all the visual information. A rate allocation scheme along with a utility function categorizes features into classes according to their usefulness in order to maximize the overall performance while staying below a target bitrate. These methods work well when we have a priori knowledge of the data. A tracking algorithm is proposed in (Coşar and Çetin 2014) and a likelihood function, as significant coefficients are sent to the network, instead of transmission of the whole image, in order to increase the network performance. In (Coşar and Çetin 2015), the method is designed to use color and motion features for identifying ROI. These methods are effective when the image contains clear objects. In (Imran, Khursheed et al. 2012, Khursheed, Ahmad et al. 2012) the method identifies the region of interest by removing noise and background information. As mentioned earlier, these methods expect clear objects and shapes for achieving accurate results. Similarly, in (Imran, Khursheed et al. 2010), the method considers background subtraction, segmentation and finally Tiff Group4 compression as local processing steps. In (Felemban, Sheikh et al. 2014), the method uses three bit planes with the most significant bit information for increasing response time. Liu, Qi et al. (2019) proposed ROI detection scheme based on the statistical distinctiveness incorporating both the lower and the higher orders. The model identifies regions of interest that show high distinctive features in comparison with the rest of the scene. These methods do not perform well for complex general images.

There are also deep learning based proposed models to segment region of interest in different applications. For instance, Jiang, He et al. (2019) detect pesticide residues in apples via segmentation of the foreground and background regions of the apple image. By calculating the roundness value and extracting the region with the highest roundness value in the connected region, a region of interest (ROI) mask is created for the apple. All the extracted hyperspectral images of apples are the input of the CNN network for pesticide residue detection. However, this method works only for their images and roundness is not a general feature for images in this thesis.

It is observed from the above discussion that these methods may help in reducing the size of the image, the methods works based on the pixel uniformity for finding the region of interest. However, in the case of the proposed work, images may contain uniform values.

2.3.2   Inter Redundancy Based Methods

In recent years most of the monitoring applications are performed via a network of cameras to enlarge the surveillance coverage. Although, multiple coexisting camera nodes may capture images of the same area from different views to provide all vital information, overlapped sensors introduce inter redundancy as well (Jayashree, Biradar et al. 2012). Redundant processing and transmission of these common regions unnecessarily consume system resources. The aforementioned intra-redundancy based methods are not enough solutions for these complex situations. A number of solutions exist in the research literature that extracting a mutual sensed region of images using geometric-based approaches, correlation models (Awad, Mowafi et al. 2015) or key point matching approaches (Wagner, Nowak et al. 2003, Wu and Chen 2007, Attaullah, Ahmad et al. 2015). Then they exploited the redundancies among the correlated images to reduce

the transmitted data volume. The approaches could be broadly divided into two groups; transmission based and processing based methods.

2.3.2.1 Transmission based Methods

They are networking approaches that segment captured images and transmit overlapping (OVL) only once to minimize introduced redundancy throughout VSNs. They give the highest priority to overlapping regions and low priority to non-overlapping regions during data transmission. The reason is that OVL carries information of multi images and is a key part to reconstruct the monitoring area of interest which loss of those packet degrades the picture quality in the destination (Padwalkar and Ambawade 2013). For example, (Jayashree, Biradar et al. 2012) developed a prioritized multipath QoS routing that evaluates nodes residual energy, buffer size and hop count to score paths. OVL as a prioritized region will be sent only once and through the best path accordingly to get improved image quality, higher packet delivery and hence throughput with minimum packet loss. Similarly, a routing protocol proposed in (Padwalkar and Ambawade 2013) called ad-hoc on-demand multipath distance vector routing that rates different paths and delivers a single OVL from the most reliable path. The routes are rated based on the availability of resources like buffer size, power, transmission time, and packet loss. Moreover, (Wang, Peng et al. 2007) designed a network of image sensors in which collaborative sensors transmit images to their Cluster Head (CH) throughout single or multiple hops. An energy-aware routing strategy is proposed in which path with less average packet loss probability is dynamically chosen for OVL transmission. This protocol is improved in (Wang, Peng et al. 2010) in term of security where OVL are transmitted throughout path with higher security level as well. The work proposed in (Kenchannavar and Kulkarni 2011) introduced an energy aware approach based on camera nodes overlapping. Sensor node with maximum overlapping factor and energy is

chosen as CH to transmit images and other nodes go into sleep mode to save system energy.

However, transmission of whole overlapping and non-overlapping region even on a priority basis is still not efficient in terms of networking performance, cost, energy consumption and lifetime. These methods focus on the priority to "send" OVL and non-OVL regions but finally, they will send all information. Therefore, they do not take full advantage of the segmentation of these regions.

2.3.2.2 Processing based Methods

In order to tackle inter redundancy problem, a number of research efforts are currently underway to process captured images to remove or minimize spatial redundancy even before transmission. As it is explained before, the ordinary compression scheme is only capable of removing the redundancies within an image itself. Distributed Source Coding (DSC) emerged as a compression mechanism to exploit the spatial association of correlated sensors. DSC for VSN, referring to the image compression of multiple cooperative sensors while they do not communicate with each other. The compression rate is determined according to redundant information in the captured images (Wang, Li et al. 2007). In DSC, multiple source nodes send their compressed version of images to a sink. The sink node, in turn, performs joint decompression. Numerous schemes are proposed based on Slepian-Wolf (SW) and Wyner-Ziv (WZ) theorems which are survived and explained in details in (Ebrahim and Chong 2014). They apply loosless and lossy compression schemes respectively to encode interrelated data independently which will be decode jointly in a sink. For example, (Elzouki, Dumitrescu et al. 2018) proposed a distributed source coding model for the case of two correlated nodes based on lattices. The authors in (Wagner, Nowak et al. 2003) presented a distributed coding technique for VSN images exploring the mutual field of view between cooperative sensors. OVL region

is identified via the proposed registration method. After that, each individual sensor transmits a low-resolution version of OVL area toward the receiver. The super resolution methods reconstruct the high-resolution version of the OVL region in the station. This process distributes a load of coding among nodes and shifts the complexity to the decoder side considering resource constrained source nodes. The main merit of DSC in compared with traditional multimedia coding methods like JPEG 2000 and MPEG.x, is the ability to shift more computation complexity from the encoder to the decoder side (Li, Kim et al. 2019). However, given some difficulties like image correlation modeling and synchronization, DSC has major challenges, particularly in image transmission applications. In addition, the way the sensor correlation model is utilized for image transmissions is only determined by source image sensors consideration, and other network parameters such as routing patterns are not taken into account (Wang, Peng et al. 2007). Moreover, as redundancy increases, the performance of the DSC also increases, else the performance decreases. Another distributed coding model is compressive sensing in which the image can accurately be co-reconstructed in the decoder sink. This theory indicates that under-sampled measurements can be the basis of reconstructing a compressible signal. This can save transmission energy, but in the price of a big computational burden while decoding which is not applicable in real-time surveillance applications (Li, Duan et al. 2018). Another processing paradigm for removing inter redundancy is multimedia data fusion where typically an intermediate sink node gathers data from cooperative neighboring nodes and combines them to reduce data redundancy (Li, Kim et al. 2019). For example, in (Chow, Lui et al. 2006, Sastra, Widyantara et al. 2008) intermediate sensors combine the images captured by cooperative cameras to reduce images size. They combine and calculate the average of duplicated information in the OVL portions. Different JPEG compression rates are considered in cameras and intermediate nodes for the whole image. To keep tradeoff between image quality and

energy consumption, they demonstrated that image quality is sacrificed to save transmission energy. Same authors in (Chow, Lui et al. 2006, Chow, Lui et al. 2007), developed a protocol in which a mobile sink node collects images instead of transmission towards a far station in a large geographical area. Similarly, intermediate sensors aggregate images of overlapping regions collected by the mobile sink to reduce the data size. The physical GPS locations and orientations of camera nodes define OVL and non-OVL regions. Intermediate nodes in (Chia, Ang et al. 2009) stitch the images taken by neighboring sensors in conjunction with SPIHT coding to discard the overlap redundancy before transmission to a centralized decoder. The flexibility in coding stitched images is enhanced utilizing a tree structure adapted by SPIHT. They combine multiple images with overlapping fields of view to create a high-resolution image. Coşar and Çetin (2015) proposed an efficient tracking system in VSN in which sparse representations of the likelihoods of overlapping images are transmitted toward a fusion node. In this way, they avoid overloading network with redundant information. Temporal redundancy is removed in (Wu and Chen 2007) among image sequences. The proposed model subtracts and transmits the background image only once. Upon any triggered event, the target regions are detected and transmitted along with their spatial locations. The whole image will be reconstructed by fusing the background, the targets and spatial locations in the station. Although, all these methods reduce image size for transmission, the different importance levels between OVL and non-OVL are not taken into account and the whole images are compressed or fused. The methods in (Wang, Peng et al. 2007, Wang, Peng et al. 2010) considered this key point and used distortion operations with different rates for the overlapping and non-overlapping regions to reduce the size of the image. The idea is that given overlapping region is the key region, a distortion operation is applied to reduce the size. At the same time, these methods use different distortion levels for non-overlapping regions because they are not as vital as overlapping regions. Since the method involves a

Mean operation, which introduces blur during the operation, the images lose quality. In addition, since the window size of the operation is fixed, one cannot expect stable and consistent results.

It can be noted from the above discussion that, although these inter redundancy removal methods reduce image size to an extent, they ignore quality of the images. However, the quality is important for the proposed work.

2.4   Bit Plane Segmentation for Image Size Reduction

Every pixel of an image can be represented by a binary sequence with 8 bits. A bit-plane of an image is a set of bits in the same corresponding position in the sequence (Zhou, Yu et al. 2018). The use of significant bit planes and bit plane detection are not new concepts in the field of image processing and pattern recognition. However, use of bit planes for improving network performance is new and hence the related methods hardly could be found. This sub-section provides a review of the bit plane based approaches which can be classified into three classes; eight bit plane based methods, predefined bit plane selection and feature extraction based methods.

2.4.1   Eight Bit Plane based Methods

In the literature there are methods that use all eight bit planes for different purposes. For example, Yoo, Kim et al. (2017) employed bit plane decomposition for local binary pattern extraction for face recognition. The method extracts high dimensional features from the bit plane images by exploring local structure of the images and concatenates on them. In fact, all the bit planes are concatenated into a single vector to generate those features. Although, all planes in the vector can enhance the discriminative power of face recognition on a local server, it will be very expensive if they be transmitted throughout a network. Chen, Ma et al. (2017) proposed the method of attention region detection based on analysis of the bit planes. The method extracts color contrast information from all the

planes to find attention regions in the images. The method assumes the higher order planes often contain the majority of the visually significant data. However, they used information in all the planes for the detection purpose. This limits robustness of the method because other planes, unlike most significant bit plane, usually provide unwanted information. This unwanted information affects the performance of the method. Yang, Frater et al. (2008) designed a denoising methodology for digital images. The bit planes are divided into two groups, the three lowest order bit planes (1-3) containing more residual noise are subjected to a bitonic filtering for noise removal. The rest of five bit planes (4-8) including more meaningful features are kept unprocessed. The number of planes in each sub class is determined empirically. However, their division policy limited to the testing images and the result is not same for general images. In addition, there are studies utilizing image bit planes for security purposes. For instance, Zhou, Yu et al. (2018) proposed a bit plane based encryption method in which the four bit planes (5-8) and also (1-4) bit planes are combined and converted by the discrete fractional angular transform to yield the encrypted image. Although, combination of the planes reduces the image size significantly, it is not suitable solution in term of expected visual quality for surveillance applications. Another encryption example is (Sravanthi, Patro et al. 2019) where every bit plane is XORed with a key image to provide encryption complexity. However, all planes that are vital in their solution are not as important as in surveillance applications.

From the above review, it is noted that methods use characteristics connected component analysis in the images. The connected component analysis is good for the images of high contrast but not low contrast images.

2.4.2   Predefined Bit plane selection Methods

Different bit planes carried various information level of an image (Francis-Lothai and Bong 2014). There are methods that use bit plane information for their aims but assume

one or few predefined planes as the most informative plane and not all. For example, Felemban, Sheikh et al. (2014) proposed improving response time in critical visual sensor network applications. The method presents priority routing framework for image transmission. To reduce the size of the image and for smooth transmission, it uses the bit planes of the color images. However, rather than finding informative bit planes, it fixes 8th plane as the best plane number, which may not be effective always such as in the case of complex background. Similar assumption is followed in (Wang, Leng et al. 2007) for an face recognition proposed algorithm. But they showed this assumption for images after histogram equalization does not work. Çelebi, Yavuz et al. (2018) visually decided 7th bit plane as the most significant bit plane providing rough information and the 6th and 5th plane are considered to provide more details. Other bit-planes are not checked since they do not provide essential motion information according to their experimental analysis. However, this assumption may only work for motion detection and not for other applications. A bit-plane selection method for reducing complexity of motion estimation in (Goyal, Dogra et al. 2018) choses the plane number 8 and 7 to quantized a full-precision images into one-bit images. In addition, Bit-plane extraction is also one of the popular methods in biometric applications. For instance, in iris recognition system presented in (Basit and Javed 2007) only bit-plane (2-6), in face recognition system in (Ting, Bong et al. 2008) only combination of (5-8) planes and in palmprint recognition in (Lee and Bong 2013) only bit-plane number 7 are selected to process. Therefore, each of the bit-plane plays different role for different applications and cannot be generalized (Francis-Lothai and Bong 2017). Moreover, there are studies that define the most significant plane based on conducting a set of experiments. Etemad, Samavi et al. (2018) developed a watermarking model in which the best bit planes maintaining both robustness and transparency are selected to hide information. To figure out the best plane, they have tested the proposed method on 100 natural gray-scale images and the results show 5th bit

plane has the highest rate giving balance of transparency and robustness. However, the conclusion depends on the test images and cannot be generalized. In addition, the bit plane keeping this balance does not necessarily contain the highest information level among the all planes. In (Francis-Lothai and Bong 2017) the role of bit-plane extraction in fingerprint recognition is investigated to reduce required large storage capacity. The most useful bit-plane is selected as 8th plane based on conducted experiments on two fingerprint datasets since it gives highest fingerprint recognition in compare with all other bit planes. It is valid and expected, since finger print includes clear edge, but it is not true for other datasets. A face recognition approach based on the bit-plane binary images is designed in (Wei and Li 2015). Upon being equalized, gray images are divided into eight planes. Five bit-plane images containing more discrimination information are defined empirically. Experiments on two face standard datasets reveal that for their application planes number 1, 5, 6, 7, 8 lead to the higher recognition rate. However, the result of these experiments cannot be applicable for different image types and datasets and images. Li, Shao et al. (2016) presented a bit plane based object tracking in video surveillance system. To this aim significant bit planes representing the tracking object are exploited. They tested 20 common and noisy video sequences to find the most informative bit planes. They demonstrated that bit plane number 8 yields better results for common frames and bit planes number (6-8) are better for noisy frames. Therefore, they defined bit planes number (6-8) as the most informative planes for their tracing purpose. However, the experiments have been done on limited images with clear shapes according to their application.

In summary, from the above discussion, it can be inferred that the methods assume few fix planes as the most significant planes are just suitable for limited images and datasets which contains homogeneous background with clear object features. They may not be applicable for scene images considered in this work, where one can expect large variations in the background and foreground including contrast, image resolution etc.

2.4.3   Feature Extraction based Method

There are few studies that detect the most informative bit plane more carefully and based on extracting planes features. For instance, Raghunandan, Shivakumara et al. (2018) proposed a multi script oriented text detection and recognition approach exploring iterative nearest neighbor criteria for detecting the best bit plane from eight planes. The method uses convex hull based features for detecting candidate pixels in the plane images. However, the use of this method is limited to the images of text information but not general images because the features are extracted based on character components. Dutta, Mandal et al. (2007) proposed the method for bit plane extraction to a class of intensity-based corner detection algorithms. The method explores the information in the set of planes for corner detection. It finds a number of planes, which contribute significant information based on corners. The corners of all the significant planes are then combined to remove false corners. When the plane (especially lower bit plane) is too noisy, the corner detection may not work well and hence the method is not robust. Additionally, the idea does not work well for complex scene images with unexpected corner information in the foreground and background of the image. Duvar and URHAN (2018) presented a bit-planes selection method for motion estimation of screen content video. The planes are weighted according to the level of details and the number of 1s in a generated edge map. If the number of 1s in the 8th plane is higher than a fixed threshold so that be enough for the matching process, plane number 8,7 and 6 will be chosen as the most significant planes. Otherwise, the 5th bit-plane is also will be included. However, the method is limited to screen content and a predefined threshold and three or four planes are detected which is still very expensive for transmission.

In summary, the current methods exploit the most informative plane considering their specific applications such as text, biometric or face images which usually include expected fine features. Almost none of the bit plane feature extraction methods are

designed for scene image of surveillance applications. In addition, they selected a bunch of planes for their processing goals instead of detecting a single best plane keeping a balance between quality and efficiency in term of network constraints. Further, the methods are good for high quality images and for the low quality images, the methods may not give satisfactory results.

2.5 Detection of Safe Zone for UAV Landing

There are many studies that design different UAV landing models where UAVs detects a predefined location or visual marks to land in a normal circumstance (Lange, Sünderhauf et al. 2008, Patruno, Nitti et al. 2018, Polvara, Sharma et al. 2018, Hu and Mishra 2019). However, it is essential to maintain the UAV operation and control the risk in all surveillance missions even during the occurrence of extreme interior failures and/or environmental events to avoid any injuries or damages to the UAV, its surrounding environment and especially people. These emergencies, triggers safe landing zone (SLZ) procedure when "return home/ station" and "land at a predetermined alternative safe landing zone" options are not possible. Since the proposed SLZ method focuses on the segmentation of flat regions, this sub-section reviews the methods that perform zone detection for UAV landing.

2.5.1 Non-visual based Safe Land Detection Methods

There are studies that detect a safe surface for UAV landing loading the available datasets, exploiting GPS or other sensors providing environmental scalar data. For example, Scherer, Chamberlain et al. (2012) proposed autonomous landing at unprepared sites by full-scale helicopters. The method focusses on the factors such as minimum size of the site, the altitude, slope and roughness of the site, skid intersections, rotor, tail clearance, wind direction, and ground paths. In order to estimate the mentioned parameters and extract geometrical properties, the method uses a Lidar range scanner and

GPS information. In other words, the method studies reflections of GPS locations for safe landings without analyzing the surface of the ground. This is good when all departments and units are in good condition. However, the target of the proposed work is only for emergency situations where we can expect technical failures in the system and GPS information may not available. (Mayank Garg 2015) proposed terrain-based landing site selection and path planning for fixed-wing UAVs. This method uses digital elevation map to find the flatness of the landing site. The whole region is split into cells and height variance between cells are computed to merge the similar cells. If the variance is lower than a threshold, the site is chosen as a potential site. A list of potential sites along with Euclidean distance of their path to UAV is assessed to make the final decision. Basically, these method studies offline geometrical properties of chosen sites from the elevation map, not real-time visual information. During many disaster surveillances such as earthquake or flood, the defined geometrical properties change and therefore this method does not work to find appropriate landing locations. Carney, Castano et al. (2019) utilized openly available elevation, population, ruggedness, and land cover type datasets (LS, DEM, TRI, and LC) for calculation of site safety level. A weighting algorithm is used to merge all these static information and find the optimal UAV landing zone. For instance, LandScan or LS dataset provides the system with population density data to select a less populated landing location for both the UAV and the surrounding people. Similarly, a pre-loaded database provides safe locations proposed in (Lusk 2018). However, these datasets cannot guarantee safety in real dynamic circumstance and obviously are not included unexpected people or geometric changes. In (Maturana and Scherer 2015) a strategy for SLZ detection based on the point clouds captured from a LIDAR sensor is proposed. A volumetric density map generated from the original point cloud is the input of a 3D CNN to predict the probability of the evaluated area as being a safe landing zone.

In summary, these classes of existing methods are designed for known terrain and assume that the environment does not change and they totally trust the offline uploaded dataset which is not a realistic assumption. Although, scalar data providing more reliable and real-time data, they are not sufficient for decision making in complex real word environments.

2.5.2   Visual based Safe Land Detection Methods

To address the limitations of the previous subsection, there are prior knowledge-free studies capturing images that providing a rich source of information to detect a safe zone. The methods extracting different features that indicate safe zones for UAV landing such as flatness and roughness. For instance, (Mittal, Valada et al. 2018) presented a vision based SLZ detection on rubble piles of catastrophic environments. A depth map is generated from a deployed stereo camera on a drone. Then, a cost map based on flatness, steepness, depth, and energy consumption factors is built. The criteria for the selection of landing sites among potential lists are being reasonably flat and free of obstacles and having the acceptable slope range. However, considering image resolution limitations, it cannot detect a safe land in the area with less depth variance. In (Garg, Yang et al. 2018) another vision based landing is proposed combining a slope and homography features of the landing surface. A UAV is proposed to be equipped with a stereo camera generating surface depth information to find slope and roughness. In addition, a monocular camera provides dense optical flow to compute planar homography and detect the water's surface. The approach uses an internal measurement unit along with stereo geometric information which may not be available during emergencies. Plus, as in the previous method, since it only depends on the depth information, it cannot detect the safe land in all general images. Marcu, Costea et al. (2018) designed a CNN on RGB images for generating 3D virtual environment of urban areas. The synthetic data is employed to segment the areas into safe-landing, obstacle, and other regions. In addition, a synthetic dataset is obtained for

safe landing and improving Google Earth. However, the method only considers urban areas with static obstacles and does not include actual site surface analysis. Hinzmann, Stastny et al. (2018) proposed a visual landing site detection for autonomous planes. The method explores canny edge detector to avoid high contrast obstacles, then texture and color of the potential sites are evaluated. These geometric features are the input of a binary Random Forest classifier. The method considers hazards within landing regions such as terrain roughness and slope, surrounding obstacles, and wind vector estimated by onboard sensors. However, in the real world the flat regions may have arbitrary shapes and colors. Aydin and Kugu (2016) used altitude information from high resolution digital elevation models. The model uses shuttle radar topography mission maps for landing zone detection. Initially, the method segments regions of interest called blobs, which are analyzed for landing zone detection. Since the success of the method depends on the success of segmentation, therefore, the method may not perform well for complex images.

Moreover, there are existing methods exploiting more details within the captured aerial images. (Patterson, McClean et al. 2014) proposed timely autonomous identification of UAV safe landing zones. The method combines ordnance surveys and Canny edge images of the input image for potential region detection. The method estimates safety scores based on the properties of regions to eliminate obstacles in an image. Further, the method proposes fuzzy-based features for the classification of regions. This helps the method to classify regions with paths, trees, rivers, etc., and homogenous zones are determined based on the roughness of regions. The method is limited to regions where there are no complex backgrounds because a Canny edge operator is sensitive to it. In addition, sometimes, such a region may contain unexpected areas and surfaces for classification. Similarly, (Aziz, Faheem et al. 2016) proposed a UAV emergency landing site identification system using machine vision. The method explores Canny edge information and a histogram of the ground images for finding safe land. In addition, the method

extracts statistical features from the ground to verify the detected safe landing sites. However, it is not clear how the method works for complex background images because Canny is sensitive to complex backgrounds. In addition, histogram operations and thresholding may not work well for different situations. (Li 2013) proposed a software scheme for UAV safe landing area discovery. The method explores a similarity-based texture area for finding possible landing areas. Then the features are passed to an SVM classifier for confirming landing areas. As discussed above, textures alone are not sufficient to define flat regions or landing regions because they can have any texture and sometimes no texture. Moreover, the area can contain plants of the same size and surface, which appear like textures. Therefore, the method has inherent limitations when we compare it to the work considered. (Chieh-Ling Huang 2016) proposed geological segmentation for UAV aerial images using shape-based LSM with the dominant color. The method uses fuzzy c-means clustering for classification based on color information. It finds dominant colors and then estimates a similar distribution map. Further, a shape-based level set method is proposed to segment regions. The method works well for images where dominant colors represent flat regions. However, this is not true in many cases.

There are some methods detect a safe site based on deep learning methods. For example, a system for the detection of small and potentially obscured obstacles in vegetated terrain is proposed in (Maturana and Scherer 2015). A volumetric occupancy map with a 3D Convolutional Neural Network (CNN) detects the safe zone as it distinguishes between vegetation and solid objects that should be avoided. However, all safe sites have no greenery. The assumption is not always true. More deep learning methods for unmanned aerial vehicles are presented in (Carrio, Sampedro et al. 2017).

In summary, although, the proposed vision based methods can cope with real-time and dynamic nature of emergencies in surveillance applications. The methods perform well

based on specific visual clues, such as texture, edge pattern, color etc. However, the considered image may not provide specific visual clues. Therefore, there is an urgent need to generate some general features for any unexpected surface.

2.6   Summary

Overall, the approaches which are proposed for image classification based methods for surveillance applications, segmentation for image size reduction, bit plane segmentation for image size reduction, detection of safe zone for UAV landing are reviewed. Applying these approaches for surveillance applications raises several problems to tackle. By reviewing the existing methods, it is noted that there are still major challenges as discussed below for surveillance applications.

(1) It is noted from discussions on current classification methods that most of them are developed for a particular dataset with specific knowledge about that dataset. Besides, existing methods expect images that should contain objects with a clear shape. This is not true in this research case because when we consider a Soil image, as an example, one cannot necessarily expect objects and shapes. The same conclusions can be drawn for Air pollution and Garbage images. Despite deep learning-based methods having an advantage over conventional methods, they have inherent limitations, such as the need to label a large number of samples, optimizing the parameters, and setting a generalized framework (Zhu, 2017; Sharama & Sankar, 2016). Therefore, the classification of images of different types, which is considered in this study, is an unsolved problem so far.

(2) Lots of existing image size reduction methods for VSN usually work on networking parameter and ignore the potential of visual processing. The studies include the methods for addressing intra redundancy through compression and region of interest detection. These methods are not adequate for the images with

large variations. In the same way, there are methods for image size reduction by finding bit plane detection. Most methods assume bit plane for achieving the results.

(3) From the review of the existing systems on safe zone detection, it is noted that most systems use the feature such as texture, edge and other geometrical features for safe land detection. In addition, the methods focus on the classification of segmented regions based on the nature of surfaces for verification. In other words, existing systems need prior knowledge of the regions to define features. However, it is true that defining the features for unexpected regions is challenging, and hence existing systems are appropriate for a limited number of regions. Therefore, there is a need for developing generalized systems for safe zone detection without prior knowledge of the land and terrain.

The above shortcomings of the current approaches have motivated to propose new methods, which are able to classify and segment the scene images of surveillance applications, irrespective of specific features regarding the nature of considered images.

CHAPTER 3: EDGE BASED METHOD FOR SCENE IMAGE CLASSIFICATION

3.1 Background

It is noted from the review presented in the previous chapter on video, scene image and visual scene image classification that a robust method independent of specific features is required and essential for classification of visual scene image to achieve a multi-purpose surveillance application. Therefore, in this chapter, a novel classification method for scene images is presented.

The proposed approach is presented for categorizing five scene classes, namely: soil, flood, air pollution, plant, and garbage. As it is discussed in the introduction section, these emerging ubiquitous surveillance and monitoring systems require help from image processing. In general, such classes do not depend much on image content. As a result, generalized features are required to classify and identify the images of these classes. It is observed that the presence of unique objects in the images makes a difference from one image to another image. Moreover, it is true that the objects in the images usually represented by focused edges and objects. Therefore, rather than defining shape of the objects in the images, a method is presented to extract edge patterns, which represent the uniqueness of the images. This is evident from the illustration shown in Figure 3.1, which provides vital cues of edge patterns. This motivates to propose edge strength and sharpness for classification.

(a)Soil	Sobel image	Canny image
(b) Flood	Sobel image	Canny image
(c) Air pollution	Sobel image	Canny image
(d) Plant	Sobel image	Canny image
(e)Garbage	Sobel image	Canny image
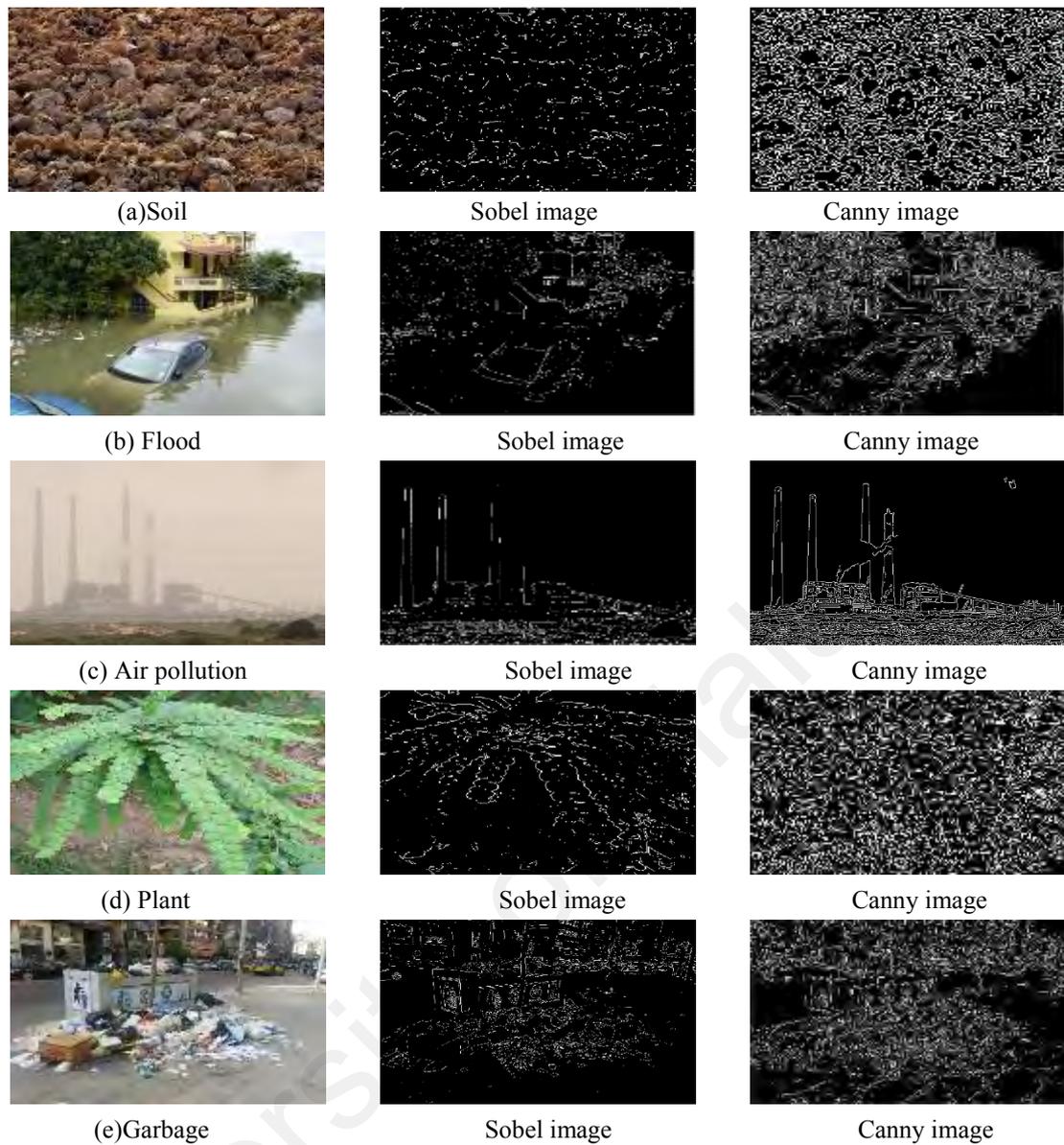
Figure 3.1: Basis for classification

As noted in (Dornaika 2016), hybrid systems enrich discrimination capacity; hence a technique is proposed to combine edge strength and sharpness features to solve the classification problem. For example, in the case of Plant, Garbage and Flood, we can see fine edges (focused edges) and different patterns, whilst for the images of Soil and Air Pollution, images lose fine edges and patterns. Edge strength extracts the following edge patterns: (1) local regularity/smoothness/continuity along a certain direction, (2) the local irregularity/oscillation/discontinuity along the orthogonal direction and (3) there exist anisotropic structures in the image and the edges always stretch out in multiple directions.

In the same way, sharpness extracts heavy-tailed distributions with the help of gradient magnitude. In other words, the components with a larger variance are believed to be responsible for the heavy-tailed property.

For an input image, the proposed technique extracts the edge strength and sharpness features, whereby subsequently, the features are fed to a K-means clustering approach to obtain the clusters, which contain high edge strength and sharpness values called Max clusters. To eliminate non-significant pixels in Max clusters, it is proposed to perform the intersection of the Max clusters for edge strength and sharpness features, which outputs the Candidate Pixels (CPs). Since the background of the images is complex, the candidate pixels are intersected with the Sobel edge image of the input image. This results in focused edge pixels, which represent high contrast values. To restore the structure of the focused edge pixels, the proposed method extracts edge components corresponding to focused edge pixels from the Sobel and Canny edge images of the input image, which results in focused edge components. For each focused edge component, the proposed method forms patches based on neighboring edge components with respect to Sobel and Canny edge images. These patches are used to extract local and global features with the help of Max cluster and Min clusters given by K-means clustering on edge strength and sharpness. Furthermore, the features are passed to an SVM classifier for classification. The whole framework of the proposed method can be seen in Figure 3.2.
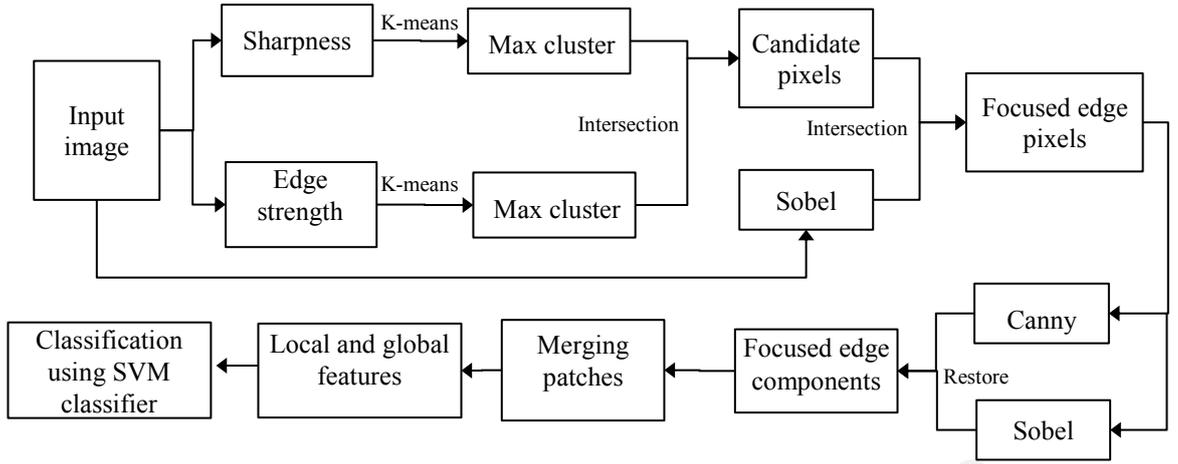
Figure 3.2: Logical flow of the proposed classification method

3.2 Candidate Pixel Detection

Inspired by the work proposed in (Zhang, Feng et al. 2013) for image quality assessment using edge strength similarity, the edge strength is proposed to explore for extracting the above-mentioned edge patterns. For the input image shown in Figure 3.1, which is a sample flood image, the proposed method obtains gradient images of four directions, namely, horizontal, diagonal-1, vertical and diagonal-2 as shown in Figure 3.3 (a-d), respectively. The reason to consider four directions is that the local regularity of the image along multiple directions is measured by directional derivatives. It is true that the regularity along with a direction and the irregularity along its orthogonal direction together imply the possibility of an edge. In other words, the difference between regularity/varying trends along two orthogonal directions gives the likelihood of the edge. With this notion, the edge strength is defined in the diagonal directions around the ith pixel P, ∂P_i^2 and ∂P_i^4 as in equation (3.1) and vertical or horizontal directions, ∂P_i^1 and ∂P_i^3 as in equation (3.2) (Zhang, Feng et al. 2013).

$$ES_i^{2,4}(P) = |\partial P_i^2 - \partial P_i^4| \tag{3.1}$$

$$ES_i^{1,3}(P) = |\partial P_i^1 - \partial P_i^3| \tag{3.2}$$

Furthermore, the total edge strength is defined as equation (3.3) and the sample result of the total edge strength is shown in Figure 3.3 (e) whereby we can see clear edge patterns where there are high contrast pixels, and no edges where there are low contrast pixels.

$$ES(P, i) = \max(ES_i^{1,3}(P), ES_i^{2,4}(P)) \tag{3.3}$$

In order to separate high contrast pixels that represent high edge strength values, K-means clustering with number of clusters $N^{cr} = 2$ is applied on the edge strength matrix, which gives two clusters. This is because the pixel, which has low contrast values, may not contribute much towards classification. The cluster is considered, which gives high mean values as the Max cluster (high-edge strength value cluster) as shown in Figure 3.3 (h).

| (a) Horizontal (∂^1) | (b) Diagonal-1 (∂^2) | (c) Vertical (∂^3) |
| (d) Diagonal-2 (∂^4) | (e) Edge strength | (h) Edge strength Max cluster |
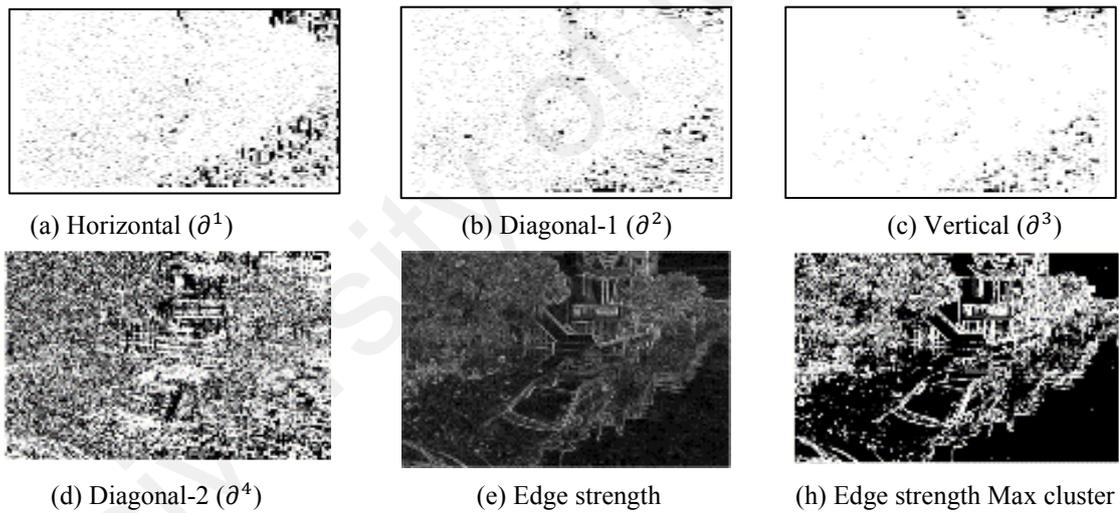
Figure 3.3: Edge strength estimation

Similarly, as discussed in the background section, edge strength alone may not be sufficient to solve the complex classification problem. Inspired by the work in (Yi and Eramian 2016) for segmentation of defocused blur using sharpness features, the proposed method explores the sharpness estimation for classification. It is noted that the gradient magnitude of sharp images exhibits a heavy tailed distribution. To extract such properties, the sharpness is estimated by computing the gradient magnitude difference as defined in equation (3.4) where $M(g)_{Sw}$ denotes the average of gradient values g in the window of

size $S^w = 5, (5 \times 5)$. In equation (3.4), the numerator denotes fourth moments around the mean divided by the square of the second moment around the mean. The local sharpness is estimated as defined in equation (3.5) where g_x and g_y are gradient magnitudes along the x and y directions, respectively (Yi and Eramian 2016).

$$D = \frac{E[(g - M(g)_{S^w})^4]}{E[(g - M(g)_{S^w})^2]} \tag{3.4}$$

$$SH = \min\left(Ln(D(g_x)), Ln(D(g_y))\right) \tag{3.5}$$

The effect of g_x and g_y for the input flood image as shown in Figure 3.1, is presented in Figure 3.4 (a) and Figure 3.4 (b), respectively. Similarly, $SH(g_x)$ and $SH(g_y)$ can be seen in Figure 3.4 (c) and Figure 3.4 (d), respectively. Figure 3.4 (c) and Figure 3.4 (d) provide horizontal and vertical details of the edges. As K-means clustering is used in the case of edge strength to obtain the pixels, which have high sharpness values, Max cluster is obtained for the sharpness matrix as shown in Figure 3.4 (e). Due to contrast variations in the image, the max clusters may contain low contrast pixels, which are not significant. To eliminate such pixels, an intersection between the Max cluster given by edge strength features and the Max cluster given by sharpness features is performed as shown in Figure 3.4 (f), whereby this results in Candidate Pixels (CP). In this way, the advantages of edge strength and sharpness features are combined.

| (a) x-gradient image | (b) y-gradient image | (c) $SH(g_x)$ |

| (d) $SH(g_y)$ | (e) Sharpness Max cluster | (f) Candidate Pixels |
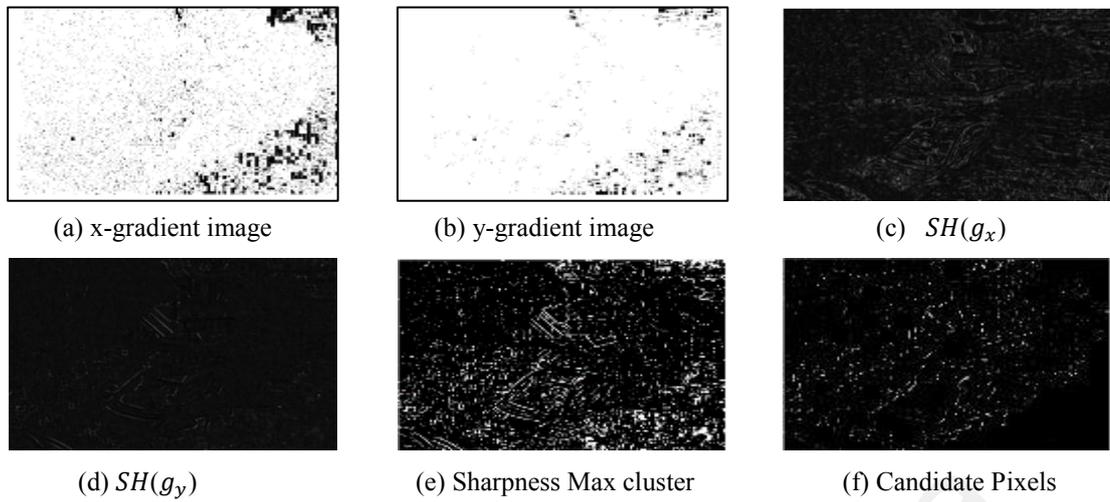
Figure 3.4: Edge sharpness estimation and candidate pixel detection

3.3 Focused Edge Component Detection

The algorithm presented in the previous section gives candidate pixels for the input image. Though expected, candidate pixel image should contain focused edge pixels, however sometimes, due to background complexity, there is a chance of classifying defocused pixels as focused pixels. It is known that Sobel gives fine edges for high contrast pixels in the image. It is also true that high contrast pixels usually represent focused edge pixels. Therefore, the proposed method intersects the CP pixel image shown in Figure 3.4 (f) with the Sobel edge image of the input image shown in Figure 3.5 (a), which results in a focused edge pixel image as shown in Figure 3.5 (b). In order to study the edge pattern as discussed before, the proposed method restores edge components corresponding to focused edge pixels in Figure 3.5 (b) from Canny edge images of the input image shown in Figure 3.5 (c). Here, the edge component is collection of connected pixels. This results in edge components associated with structures as shown in Figure 3.5 (d). Similarly, edge components with structures are restored from the Sobel edge image of the input image as shown in Figure 3.5 (e).

(a) Sobel image   (b) Focussed edge pixels   (c) Canny image

(d) Focused components- Canny   (e) Focused components- Sobel
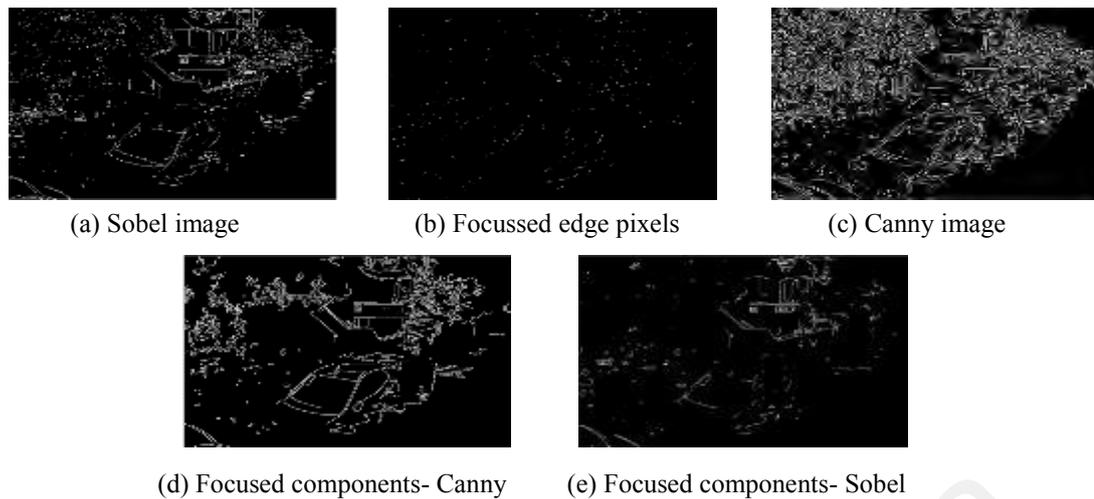
Figure 3.5: Focused edge components detection

3.4 Feature Extraction for Classification

Restoring structural edge components for each focused edge pixel may not be sufficient to extract regularity, discontinuity and other edge properties. Therefore, the proposed method merges nearest neighbor edge components in Canny and Sobel edge images as shown in

Figure 3.6 (a) and (b), respectively where we can see patches with different colors, which provides more information for feature extraction. The patch represents the area of the edge component.

(a) Patches-Canny (b) Patches-Sobel (c) Max-edge strength

(d) Min-edge strength (e) Max-Sharpness (f) Min-Sharpness
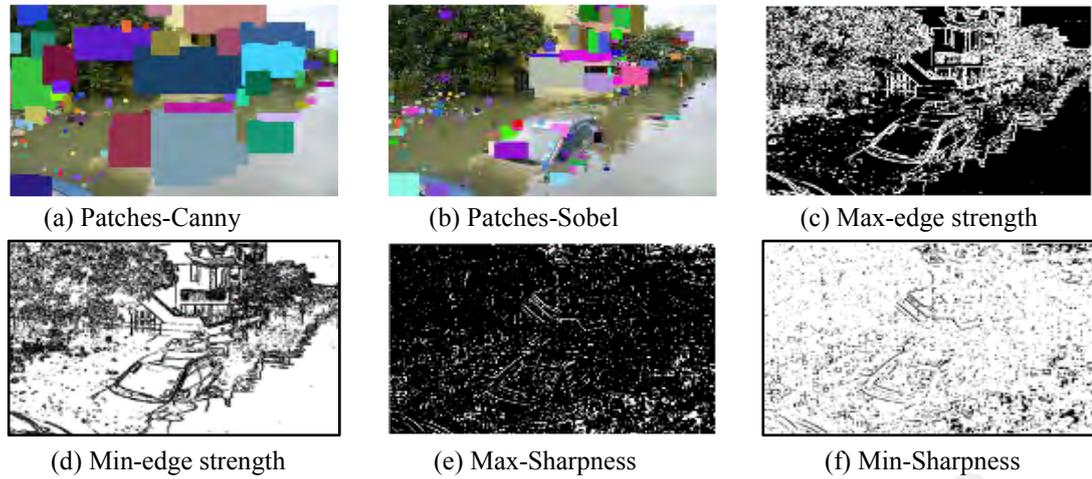
Figure 3.6: Patch formation and feature extraction using pixels in patches with four clusters

The nearest neighbor edge component is defined as a 25% overlapping region between seed edge components with a neighbor edge component. The value of 25% is determined experimentally and it will be presented in the experimental section. Due to the merger of the nearest neighbor edge components, the patch may contain both focused and defocused edges pixels. It is believed that the contribution of focused and defocused edge pixels of the edge components in the patch helps us to extract distinct features for classification of the images. Therefore, the contribution of the focused and defocused edge pixels are calculated by comparing the pixels in Max and Min clusters of the edge strength and sharpness features. It is true that the Max cluster contains focused edge pixels and the Min cluster contains defocused edge pixels as shown in Figure 3.6 (c-f), where white pixels in the Max clusters represent focused pixels, and white pixels in Min clusters represent defocused pixels. Note that the Min cluster results are complementary to those of the Max cluster. For each patch in Figure 3.6 (a) and Figure 3.6 (b), the proposed method calculates the percentage of pixels, which are defined as the number of common pixels in the patch and the clusters (say, Max cluster of edge strength) divided by the total number of pixels in the focused edge component image as shown in Figure 3.6 (d) and as defined in equation (3.6).Where N_{pt-cr}^{p} denotes the number of common pixels in a patch

and a cluster. There are four clusters, namely, Max and Min of the edge strength and sharpness features. Pg_{pt}^{cr} denotes percentage values with respect to the patch and the cluster. N_f^p is the total number of pixels in the focused edge components.

$$Pg_{pt}^{cr} = \frac{N_{pt-cr}^p}{N_f^p} \qquad (3.6)$$

For the same patch, the proposed method finds the percentage with Min cluster of edge strength features. This gives two percentage values. Similarly, two more percentage values for Max and Min clusters of sharpness features are obtained. As a result, each patch gets four percentage values. In addition, there are four more percentage values obtained for the patches in the Sobel edge image. In total, the proposed method extracts 8 percentage values. Finally, the standard deviation of the features of all the patches, as defined in equation (3.7) and equation (3.8), respectively considered as a feature vector, which contains 8 features. Where i varies from 1 to the number of patches N^h, j varies from 1 to 4 clusters and M_j^{cr} denotes the mean of the percentage values of each cluster. In equation (3.8), SD denotes the standard deviation of percentage values for each cluster.

$$M_j^{cr} = \frac{\sum_{i=1}^{N^h} Pg_i^j}{N^h} \qquad (3.7)$$

$$SD_j = \sqrt{\frac{\left(\sum_{i=1}^{N^h} Pg_i^j - M_j^{cr}\right)^2}{N^h}} \qquad (3.8)$$

Since the above features are extracted at the patch level, they represent only local information of the input image. To strengthen the features, the proposed method considers pixels in all the patches to compute percentage values with Max and Min clusters of the edge strength, and sharpness features with respect to Canny and Sobel images. As a result, this process gives eight more features and they are considered as global features. In total,

16 features are extracted for the input image as shown in Figure 3.7 using the Flood image in Figure 3.1, where the x-axis denotes the number of features and the y-axis denotes the value of the features. In the same way, the proposed method extracts 16 feature vectors for the other images for classification.
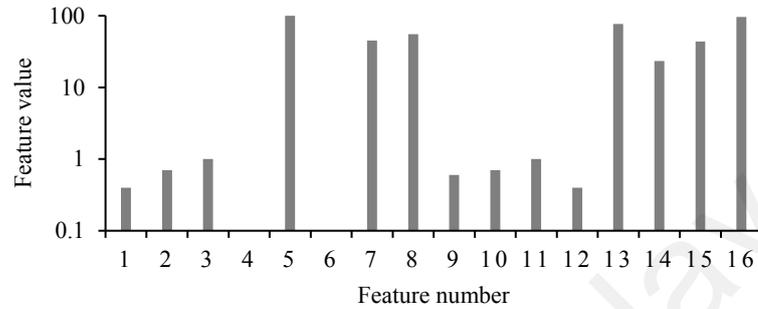
Figure 3.7: Features for the sample Flood image in Figure 3.1

It is observed from the feature vectors of all the classes that the feature vectors provide clear differences between the classes. As a result, the rules using 16 features are derived based on predefined samples. The samples are chosen according to 10-fold cross-validation. The rules for all the five classes can be seen from equation (3.9) to equation (3.13) where all thresholds ρ are determined empirically based on the training samples chosen from 10-fold cross-validation.

$$Soil = \begin{cases} 1, & if\ (FV_7 \geq FV_8) \wedge (FV_{13} < \rho_1)\ \wedge (FV_{14} < \rho_2) \\ 0, & else \end{cases} \quad (3.9)$$

$$Flood = \begin{cases} 1, & if\ (FV_{13} \geq FV_{16}) \wedge (FV_7 < FV_8) \\ 0, & else \end{cases} \quad (3.10)$$

$$Air\ pollution = \begin{cases} 1, & if\ |FV_9 - FV_{12}| \leq \rho_3 \wedge (FV_7 < FV_8) \\ 0, & else \end{cases} \quad (3.11)$$

$$Plant = \begin{cases} 1, & if\ F_{16} \geq \rho_4 \wedge (F_{13} \leq \rho_5) \wedge (F_8 \leq \rho_6) \\ 0, & else \end{cases} \quad (3.12)$$

$$Garbage = \begin{cases} 1, & if\ (FV_2 \geq \rho_7) \wedge (FV_8 \geq \rho_8) \wedge (FV_{12} \leq \rho_9) \wedge (FV_{13} \geq \rho_{10}) \\ & \qquad \wedge (FV_{13} \leq \rho_{11}) \wedge (FV_{16} \geq \rho_{12}) \\ 0, & else \end{cases} \qquad (3.13)$$

It is noted from the rules that some parameters and constant thresholds are required. For example, equation (3.10) may not be robust to classify the image shown in Figure 3.1. To overcome this issue, a templates using 16 feature vectors is created by averaging 50 images as defined in equation (3.14) of each class as shown in Figure 3.8 , where one can see each template having different distributions. Where N^I is the number of images, FV denotes feature vectors containing 16 features and M^{cs} denotes the mean of feature vector of each class.

$$M^{cs} = \frac{\sum_{i=1}^{N^I} FV_i}{N^I} \qquad (3.14)$$
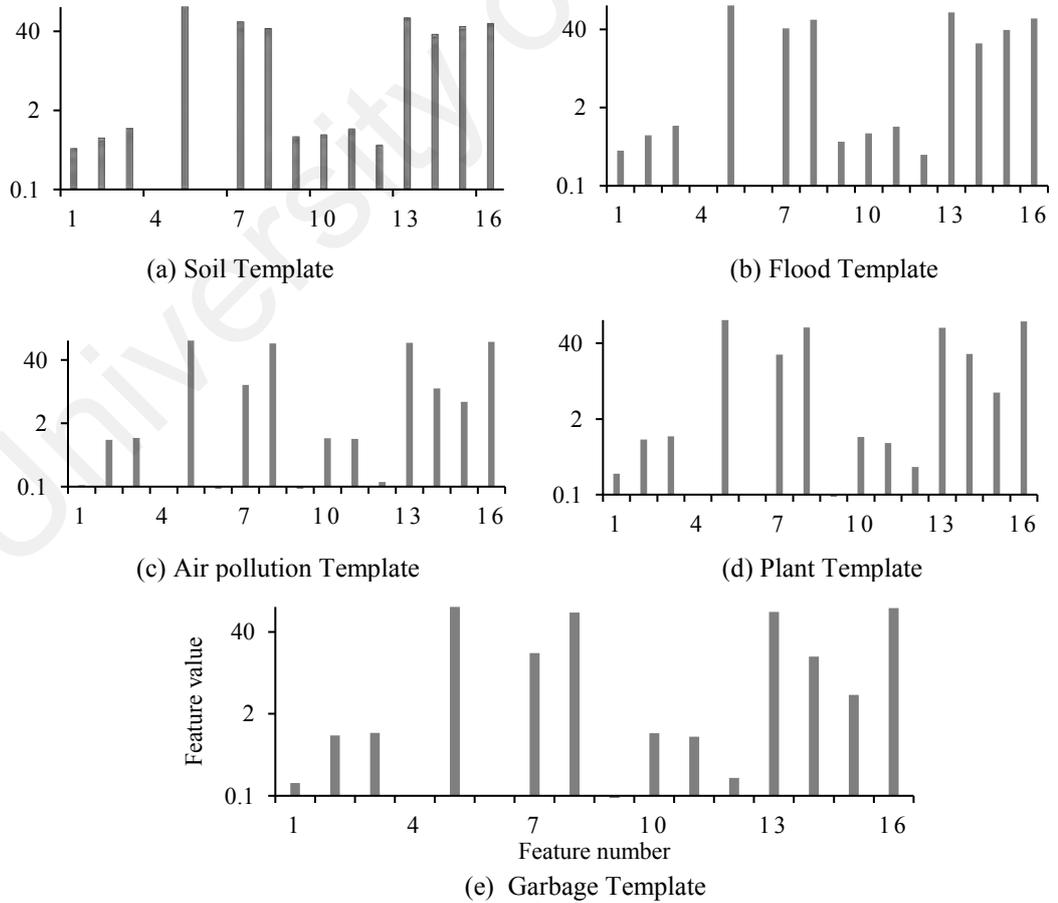
Figure 3.8: Template construction for classification

58

The proposed method estimates Euclidean distances d between feature vectors of unknown input images and the template as defined in equation (3.15) and equation (3.16) .The class, which gives a minimum distance for the input image I, is considered as the correct class, as defined in equation (3.16). d^{cs} denotes the distance between the image feature vector and template of each respective class. It is also observed from the templates shown in Figure 3.8 that the variations are not significant and hence the template-based method is sensitive to distances and is not robust and effective.

$$d^{cs} = FV^I - M^{cs} \tag{3.15}$$

$$Class_{image_I^{cs}} = \begin{cases} 1, & if\ min\big(d_I^1, d_I^2, d_I^3, d_I^4, d_I^5\big) = d_I^{cs} \\ 0, & else \end{cases} \tag{3.16}$$

Therefore, it is proposed to use a classifier (SVM), presented by (Hastie and Tibshirani 1998) to overcome the problems of rule and template-based methods. The proposed method sends the feature vector to the SVM classifier for final classification. The classifier is trained and tested according to 10-fold cross-validation. It uses the default parameters set as discussed in (Hastie and Tibshirani 1998) for multi classification.

There are many classifiers available in the literature, such as SVMs, KNNs and Random forests, etc. For this study, SVM is employed for classification. The software available in WEKA[1] is used for the SVM with default parameters and architecture. The way the features are extracted from the images ensures sufficient discriminative power. It is a fact that when the features are good, the classifier is not necessarily important in achieving good results. In addition, the SVM is a supervised machine-learning algorithm, which finds an optimal hyper-plane maximizing the distances or margin between nearest

---

[1] https://www.cs.waikato.ac.nz/ml/weka/downloading.html

data points. The SVM was originally designed for binary classification (two-class classification). Now, it is extended to multi-class classification using different ways (Hastie and Tibshirani 1998). Of the many ways, the most popular one is one-against-one or pair-wise. In this case, a multi-class, say, N^{cs} number of classes, classification problem is converted to a two-class problem with the following equation, $\frac{N^{cs}(N^{cs}-1)}{2}$. Further, the outputs of all the binary classes are combined for multi-class classification. In this study, the Kullback–Leibler distance criterion is used to find the best approximation of pairwise coupling probabilities to predict class membership as discussed in (Hastie and Tibshirani 1998). Based on this, the class is assigned to the input image according to the voting of pairwise comparisons with the other classes. On the other hand, other classifiers as mentioned above, are sensitive to parameters. For example, the main issue with the KNN is that it defines the value for N^k. That is, the performance of KNN depends on choosing the right value for k. Since the proposed work considers general scene images where we cannot find clear shaped objects, it is hard to define the values for k. Similarly, in the case of Random Forest classifiers, there are chances of omitting the feature which contributes more towards classification. The performance of classifiers with SVM will be discussed in the experimental section.

The proposed classification method involves two key steps that are edge strengths and sharpness estimation. For these features extraction, the method requires two loops because it considers image of two dimension as input. For this operation, the method process the whole image. Therefore, there is no best and worst case complexity for the proposed classification method. As a result, the time complexity of the proposed classification method is O(n^2) for both best and worst case. Hence, it is $\Theta(n2)$ time complexity.

3.5 Experimental Results

This Section is organized as follows. In Section 3.5.1 the datasets used for the approaches and evaluation metric have been explained. Section 3.5.2 illustrates the contribution of key steps of the proposed method. Section 3.5.3 shows the results of proposed method by different classifiers and datasets, and finally, section 3.5.4 gives the details of comparison of the proposed and existing classification methods.

3.5.1 Dataset and Evaluation

A dataset is created for evaluating the proposed method because there are no standard datasets available in the literature for classification of scene images, which are essential for ubiquitous surveillance applications. The dataset includes five classes, namely, (1) Soil (SL), which consists of images of different soils; (2) Plant (PT), which contains images of different plants; (3) Air Pollution (AP), which contains images of smoke, haze, fire; (4) Flood (FD), which contains images of different floods of rivers; and (5) Garbage (GE), which contains images of dirty items, waste and trash items. Sample of the collected image of the dataset are shown in Figure 3.9. To capture the images, a low cost camera is used with a resolution of 600×450 pixels. Image size is 300×500 which are taken in different lightning, weather conditions and daytime from Malaysia, Iran, India and Germany[2]. They are chosen for classification because according to the Government of Malaysia, these are the key issues and a part of the country's vision for research. Each class consists of 100 images, which are captured at different places in Kuala Lumpur, Malaysia with lots of intra- and inter-variations in terms of contrast, background, resolution, and views, to name a few. For example, a flood image may contain plants as greenery. Garbage images may contain plants and water as in flood images. In total, 500

---

[2] The data set will be shared upon request for supporting reproducibility.

images for five classes are considered for evaluating the performance of the proposed method. Therefore, the dataset is complex, challenging and useful. The size of the database is restricted to 500 images, though the intention was to collect a large number of images because the process of capturing the above-mentioned images requires the right time, situation, place and permission from authorities, etc. In addition, the deployed camera or camera on a Drone set up is expensive and it is not easily available for public and/or student use for capturing images. In order to test the objectiveness and robustness of the proposed method, we also test on the benchmark 8 scene category database (Oliva and Torralba 2001) , which consists of Highway (HY), Coast (CT), Tall Building (TB), Street (ST), Mountain (MN), Inside City (IC), Forest (FT), and Open Country (OC).

(a) Soil sample images

(b) Flood sample images

(c) Air pollution sample images

(d) Plant sample images

(e) Garbage sample images
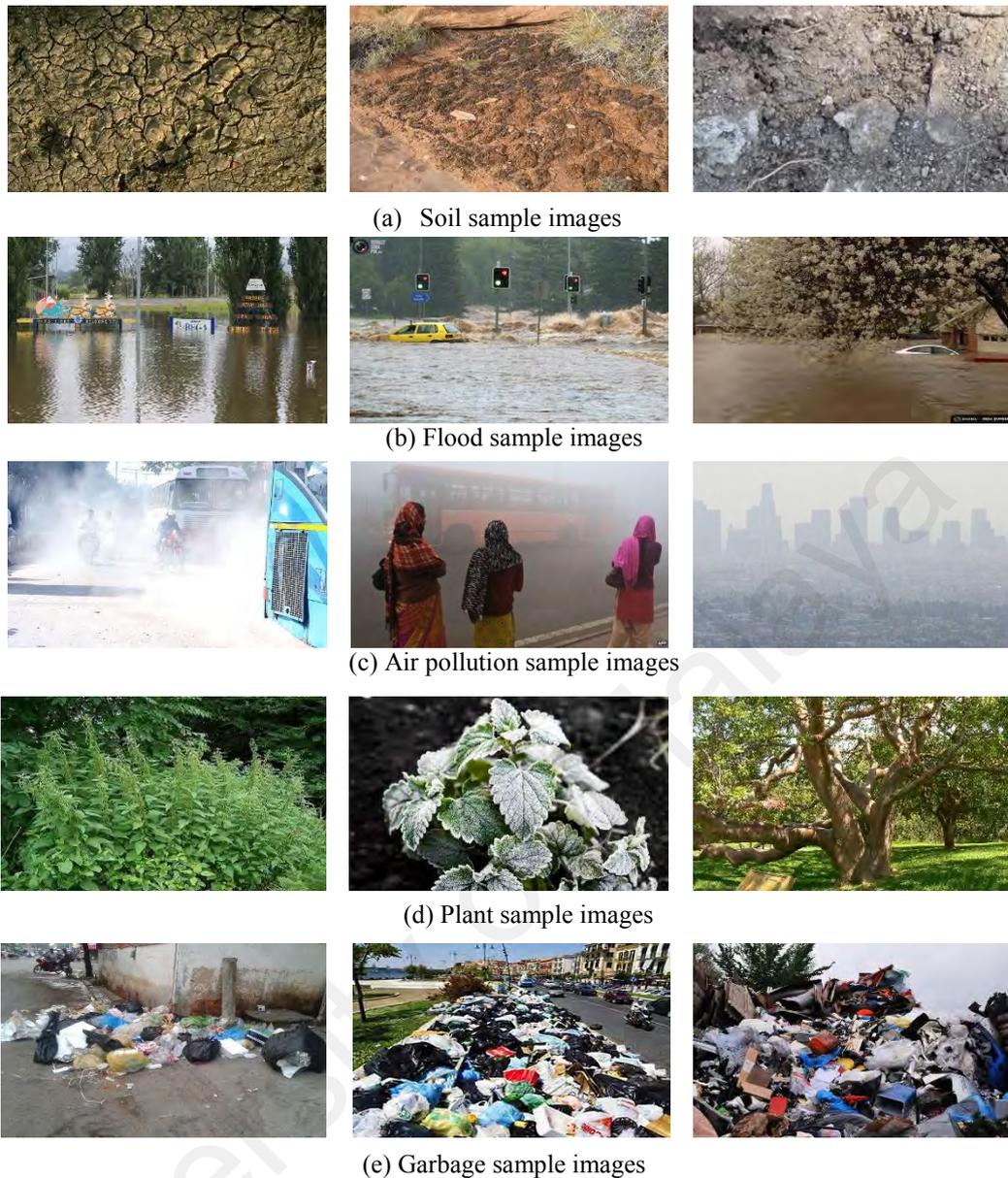
Figure 3.9: Sample image of collected dataset for classification

This dataset contains 8 classes, in total, 2688 images and is much more complex than our dataset because it includes indoor and outdoor images while the created dataset includes only outdoor images. The images of the benchmark dataset contain clear objects; for example, Tall Building and Street classes where we can find objects with specific shapes. When the dataset represents combined indoor and outdoor images, it is hard to extract the features, which facilitate good results in classification due to the presence of images with specific shaped objects as well images with no specific shaped objects. To measure the performance of the proposed method, standard measures are calculated,

namely, Average Classification Rate ($\overline{CLR}$) through a confusion matrix and the Average Processing Time ($\overline{T^P}$), which is the average time for processing each image. The classification rate is defined as the number of images classified correctly N^{cd} by the proposed method divided by the total number of images in a class N^I, as defined in equation (3.17).

$$\overline{CLR} = \frac{N^{cd}}{N^I} \qquad (3.17)$$

In addition, to evaluate the proposed and existing systems for the classification, well-known measures are used, namely, Recall, Precision and F-measure as defined in equation (3.18) till equation (3.20).

$$Recall = \frac{tp}{tp+fn} \qquad (3.18)$$

$$Precision = \frac{tp}{tp+fp} \qquad (3.19)$$

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (3.20)$$

To show the effectiveness of the proposed method, an existing method (Qin, Shivakumara et al. 2016) is implemented, which proposes statistical, structural and spatial features for video type classification to improve text detection and recognition performance, as a comparative study. Also state-of-the-art methods of scene image categorization are considered to compare with the proposed technique system, namely, (Bosch, Zisserman et al. 2008) which proposes color, texture, and SIFT features for the objects in the image for classification, and (Dunlop 2010) which extracts color, edge, line, texture and shape features for segmenting the region of interest for classification of scene images. The main reason to consider these methods for a comparative study is as follows. Qin, Shivakumara et al. (2016) focuses on video image classification based on edge features as in the proposed method, where we can expect low contrast to improve text

detection and recognition methods. (Bosch, Zisserman et al. 2008) and (Dunlop 2010) focus on scene images for categorization as is the case for the proposed method, where we can expect high contrast images and these methods are the state-of-the-art.

In addition, since deep learning is widely used in a wide range of applications, the system called GOOGLE API is used, which is available publicly (Google API system) for annotating the images of the scene images. Furthermore, the system explores deep learning concepts and uses a large number of features for labeling query images. Since the website does not provide technical details of deep learning and feature extraction, a discussion is provided based on the experimental analysis on different input images. To test whether the proposed technique is competitive with such (Google API system), the results of the proposed technique is compared with (Google API system). The GOOGLE API system provides a confidence score for the input image with several labels in order. To calculate the classification rate for the input image, a cut-off point of an 85% threshold is applied to the confidence score and relevant labels are manually chosen for classification. For example, the system provides labels, such as water, river, and flood for the flood image with a confidence score. In this case, they are counted manually if the relevant labels score more than 85% to determine correct classification. Otherwise, the images are discarded. A threshold 85% for classification is determined because it is noted that if we increase the threshold, the system includes more irrelevant labels whereas, if we decrease the threshold value, the system misses some relevant labels.

In section 3.4, feature extraction has been presented for the patches where the proposed method uses 25% overlapping regions to find the nearest neighbor region to merge into one region. The 25% is determined empirically as shown in Figure 3.10 where different thresholds are tested for the overlapping region, and average classification rate are calculated accordingly.
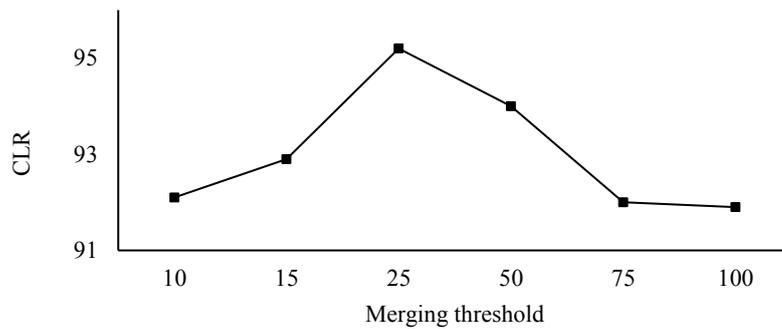
Figure 3.10: Determining degree of overlapping regions for the best classification

Figure 3.10 shows that the classification rate increases gradually until 25% and then the classification rate decreases. Therefore, it be can asserted that if we increase the percentage of the overlapping region, the proposed method loses its ability to classify the images due to redundant information. At the same time, if we decrease the percentage of overlapping regions, the proposed method loses its accuracy due to a loss of significant information. This is valid because if it is less than 25%, we lose shapes of the edge components in the image and if it is more than 25%, it covers more unwanted background edge components. Therefore, the experiments show that 25% overlapping is the ideal to achieve better results.

3.5.2   Evaluating Key Steps of the Proposed Method

The proposed method involves several key steps to achieve better classification results, such as edge strength with Canny, edge strength with Sobel and sharpness with Canny, as well as sharpness with Sobel. To analyze the contribution of each step, experiments are conducted on 500 images for each step as reported in Table 3.1. The average Classification Rate ($\overline{CLR}$) is calculated for each experiment, which is the result of determining the mean of the diagonal elements of the confusion matrix. Table 3.1 shows that each step contributes significantly to achieving better results. For example, Edge

strength with Canny and Sobel is better than Canny alone and Sobel alone; Sharpness with Canny and Sobel is better than Canny alone and Sobel alone.

Table 3.1: **Contribution analysis of the key steps of the proposed method**

Key steps of the proposed method	$\overline{CLR}$ (in%)
Proposed method	**95.2**
Edge strength With Canny	74
Edge strength with Sobel	21
Edge strength With Canny and Sobel	67.8
Sharpness with Canny	65.4
Sharpness with Sobel	86
Sharpness with Canny and Sobel	88.8

3.5.3 Evaluating the Proposed Classification Method

In this section, rule- as well as template- and classifier-based approaches are discussed for classification. To illustrate that the proposed features with a classifier have the ability to classify the images; experiments on the rule, template and classifier-based methods are conducted. The quantitative results in terms of the confusion matrix of classification rates and F-measure of the rule, template and the proposed method (classifier-based method) are reported in Table 3.2.

Table 3.2: **Confusion matrices of the rule-based, template and the proposed method**

Methods	Rules					Template					Proposed Method				
Classes	SL	FD	AP	PT	GE	SL	FD	AP	PT	GE	SL	FD	AP	PT	GE
Soil	81	7	1	2	0	92	4	0	3	1	100	0	0	0	0
Flood	0	68	3	3	5	19	63	6	3	8	0	92	2	3	5
Air Pollution	0	1	75	0	1	0	6	74	9	11	0	1	99	0	0
Plant	5	7	5	65	5	6	11	5	65	13	3	3	0	92	2
Garbage	0	3	12	4	53	1	3	11	3	82	0	4	3	3	93
$\overline{CLR}$ (in%)	68.4					75.2					**95.2**				
$\overline{T^P}$ (in s)	0.009					0.33					2.05				
F	81.2					85.6					97.5				

It is observed from Table 3.2 that the template-based method is better than the rule-based method and lower than the proposed method in terms of average classification rate and F measure. However, the rule-based method is the best at average processing time ($\overline{T^P}$), compared to the template-based and the proposed method. This is valid because the rule-based method does not involve a larger number of computations compared to the template-based and the proposed methods. Since it is sensitive to conditions and parameters, the rule-based method reports poor results compared to template-based and the proposed method. Similarly, the template-based method is insensitive to parameters and it does not involve constant thresholds, but it is sensitive to templates and distance measures. On the other hand, the proposed method does not involve any fixed parameters and it uses conditional probability for estimating the confidence score for classification. Therefore, the proposed method is better than the rule- and template-based methods in terms of the cost pertaining to processing time. When $\overline{CLR}$ and F-measure of the rule-based, template and classifier-based methods are compared, one can conclude that the contribution of feature extraction is more than the classifier because rule and template-based methods report more than 68% $\overline{CLR}$ without a classifier.

In order to show effectiveness of the SVM classifier for classification, the results of the SVM classifier with two other popular classifiers are compared, namely, KNN and Random forests using our dataset. The quantitative results of the SVM, KNN and Random forest classifiers are reported in Table 3.3 where it is noted that the SVM classifier gives better results than the other two classifiers. This is valid because the performance of the KNN depends on the value of N^k and the performance of the Random forest depends on features chosen randomly. On the other hand, the proposed SVM considers all the features for classification. In this chapter an SVM is selected with a polynomial kernel rather than an RBF kernel. The reason is that the polynomial is simple and it requires less computations compared to

RBF. In addition, since the proposed features are sufficiently strong, choosing different kernels and other parameters do not have a substantial effect on the overall performance of the proposed method.

Table 3.3: **Performance of different classifiers on the dataset**

Classifier Methods	KNN					Random Forest					SVM				
Class	SL	FD	AP	PT	GE	SL	FD	AP	PT	GE	SL	FD	AP	PT	GE
Soil	81	11	2	4	1	73	14	4	6	2	100	0	0	0	0
Flood	9	73	3	5	10	10	56	7	14	13	0	92	2	3	5
Air Pollution	0	4	80	4	12	1	4	54	18	23	0	1	99	0	0
Plant	6	12	6	60	15	9	13	6	55	16	3	3	0	92	2
Garbage	2	12	13	16	59	2	10	14	14	62	0	4	3	3	93
$\overline{CLR}$ (in%)	70.6					60					**95.2**				
F	82.7					75.0					**97.5**				

To test the expandability of the proposed technique, 50 more images are added for each class. In addition, three more classes are added, namely, Car, which contains images of different cars, Animal, which contains images of different wild animals and Pipe, which contains images of damaged pipes to the five classes. Similarly, to show the proposed technique is invariant to rotation and scaling, it is tested on different rotated and scaled up-down images in the class. In the same way, to show the proposed technique is robust to noise and blur compared to the proposed technique with rules and templates, the proposed techniques are tested on images with added Gaussian noise and blur at different levels as shown in Figure 3.11 (a) and Figure 3.11 (b), respectively. It is observed from Figure 3.11 (a) and (b) that as noise and blur level increase, the performance of rule-, template- and classifier-based techniques decreases gradually. The average classification rates of the rule-based, template and the proposed technique for the above experiments are reported in Table 3.4, where we can see for all the experiments, the rule-based method scores lower as compared to the template-based and the proposed technique. As a result, it can be concluded that the rule-based method is not robust compared to the template-

based method, while the template-based method is not as robust in comparison to the proposed technique. This is valid because the rule and template based methods are sensitive to pixel changes, while the proposed classifier based technique does not. It involves probability for predicting labels for the classes, which makes the proposed technique insensitive when compared to rule- and template- based methods. Some sample images which are classified successfully are shown in Figure 3.12.
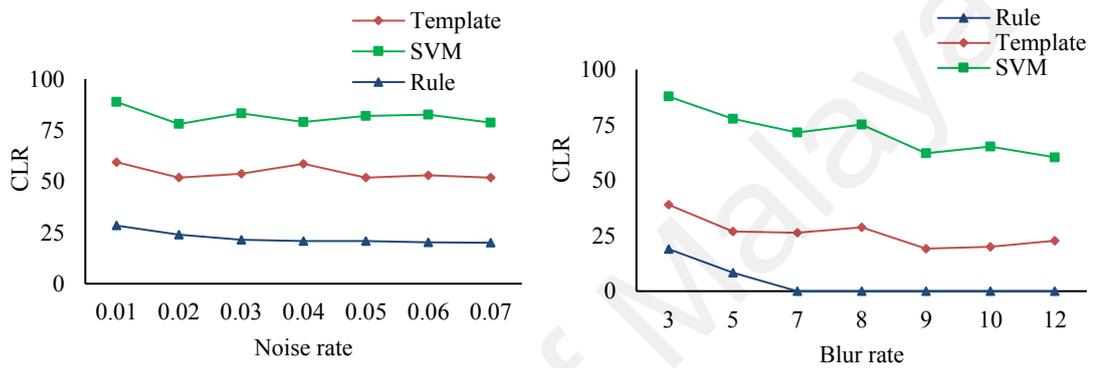
(a) Experiments on different scales of noise   (b) Experiments on different scales of blur

Figure 3.11: Robustness of the rule, template and the proposed method

Table 3.4: Performance of the proposed method on different cases, $\overline{CLR}$(in %)

Experiment	Rule	Template	Proposed Technique
Normal dataset	68.4	75.2	95.2
Adding more number of images	56.0	66.9	89.6
Adding more number of classes	-	47.5	85.25
Different rotations	6.2	28.2	80.4
Adding noise	-	-	88.8
Adding blur	-	-	87.8

Soil   Flood   Air pollution

Plant   Garbage

Figure 3.12: Sample successful results of the proposed method

3.5.4 Comparative study

The quantitative results of the proposed method and existing methods (Bosch, Zisserman et al. 2008, Dunlop 2010, Qin, Shivakumara et al. 2016, Google API system) are reported in Table 3.5 through a confusion matrix of the average classification rate and F-measure. It is observed from Table 3.5 that the proposed method is the best at $\overline{CLR}$ and F-measure compared to the other systems. Note that since the GOOGLE API system provides a confidence score as well as relevant labels and the number of labels is unknown only the classification rate is reported for each class instead of a confusion matrix.

The main reason for the existing method (Qin, Shivakumara et al. 2016) scoring poorly is that the proposed features are more inclined towards text properties rather than general features. However, the extracted data does not include text information and the contrast is not restricted to low or high levels. Other existing methods (Bosch, Zisserman et al. 2008, Dunlop 2010) require multiple objects with clear shapes in the images to achieve better results. This is not true in the case of the proposed model. Similarly, the GOOGLE API system extracts features based on the object shapes as per existing scene classification methods. Since the system does not use created dataset for training, it may report poor results for this problem compared to the proposed technique.

Table 3.5: Comparative study of the proposed and existing methods on classification

Methods	Proposed Method					Qin et al. (2016)					Dunlop (2008)					Bosch et al. (2008)					GOOGLE API
Classes	SL	FD	AP	PT	GB	SL	FD	AP	PT	GB	SL	FD	AP	PT	GB	SL	FD	AP	PT	GB	ACR
SL	100	0	0	0	0	80	7	5	4	4	22	27	17	8	26	70	20	0	10	0	82
FD	0	92	2	3	5	11	57	5	4	23	14	25	26	17	18	0	70	10	10	10	41
AP	0	1	99	0	0	10	6	80	2	2	22	19	28	15	16	0	0	100	0	0	31
PT	3	3	0	92	2	3	5	1	84	7	21	25	20	15	19	20	30	10	30	10	86
GB	0	4	3	3	93	4	19	0	5	72	19	23	11	8	39	0	20	20	10	50	60
$\overline{CLR}$ (in %)	95.2					74.6					25.8					64.0					60.0
F	97.54					85.45					41.01					78.04					75.0

The quantitative results of the proposed and existing methods are reported in Table 3.6 where one can see that the proposed technique gives better results than existing methods in terms of classification rate and F-measure except for the GOOGLE API system. This is valid because the GOOGLE API system considers huge labeled samples, which covers almost all variations for training and classification. However, when we compare the results on the created dataset and the benchmark dataset reported by the GOOGLE API, the system scores poorly for the collected dataset compared to the benchmark dataset. This is due to the lack of training samples, which resemble the created dataset. When the results of the proposed technique on the dataset and the benchmark dataset are compared, the proposed method produces low score results on the benchmark dataset. This is valid because the method is developed for scene image classification where we cannot find specific shapes of the objects, whilst the benchmark dataset contains the images with specific shapes of the objects. Therefore, the extracted features may lose discriminative power when the images are mixed with indoor and outdoor scenes. In addition, the benchmark dataset contains 8 classes whilst our dataset contains 5 classes. With these experimental results, it can be concluded that the proposed technique is expandable and generic with little modifications.

Table 3.6: Comparative study of the proposed and existing methods on benchmark 8 scene category database

Method	Proposed								Qin et al. (2016)								Dunlop (2008)								Bosch et al. (2008)								GOOGLE API
Class	HY	CT	TB	ST	M	IC	FT	OC	HY	CT	TB	ST	M	IC	FT	OC	HY	CT	TB	ST	M	IC	FT	OC	HY	CT	TB	ST	M	IC	FT	OC	ACR
HY	188	14	9	12	11	16	2	8	125	30	26	13	20	21	15	10	140	4	23	32	1	11	5	24	158	11	14	10	7	28	20	12	196
CT	23	228	22	16	18	23	9	21	29	175	34	12	29	16	22	43	16	230	29	19	5	5	5	41	16	181	9	25	12	14	39	64	257
TB	9	35	212	22	19	20	11	28	25	34	175	19	19	33	24	27	45	4	188	28	25	19	6	11	10	52	187	10	9	24	19	45	286
ST	14	12	25	161	27	14	15	24	16	12	30	150	26	20	17	21	46	4	36	140	10	16	1	29	3	34	6	155	23	14	7	50	230
MN	17	21	21	29	210	15	26	35	27	27	23	33	197	13	18	36	15	28	52	4	229	9	16	21	22	28	46	3	201	16	38	10	291
IC	7	30	19	8	10	207	4	13	28	14	35	23	18	150	18	22	23	4	39	14	15	161	11	42	9	10	11	8	45	195	4	26	219
FT	4	0	7	10	24	4	244	25	17	20	23	18	27	25	176	22	6	8	37	22	43	13	171	26	24	34	6	21	12	9	184	38	238
OC	15	23	33	26	34	16	26	237	17	43	37	30	39	17	40	187	12	7	16	9	49	36	13	269	7	39	11	24	27	10	12	280	321
$\overline{CLR}$ (in %)	62.23								49.73								56.21								57.26								75.68
F	76.71								66.42								71.96								72.82								86.15

3.6 Summary

A new technique for scene image classification applications to ubiquitous visual surveillance is proposed that integrates the merits of edge strength and sharpness features for classification of images. The proposed technique detects candidate pixels by intersecting the output of K-means clustering on edge strength and sharpness features. The candidate pixels are used for focused edge pixel detection with the help of the Sobel edge image of the input image. The edge components with shapes for each pixel in a focused edge pixel image are restored to extract the features. The unique features are extracted from the patches formed by the edge components and the neighbor edge component with clusters of edge strength and sharpness features. Furthermore, the features are fed to an SVM classifier for final classification. The experimental results on rule-based, template-based, classifier-based, and existing scene classification methods show that the proposed technique outperforms the existing state-of-the-art methods.

CHAPTER 4: REGION SEGMENTAITON AND BIT PLANE DETECTION METHODS FOR IMAGE SIZE REDUCTION

4.1   Background

In the previous chapter, a new method for scene image classification is presented. Due to inherent limitation of Visual Sensor Network (VSN), transmitting classified the whole image affect life of networking and performance. Therefore, this chapter focuses on image size reduction to improve the performance of a surveillance system. For this purpose, the proposed work presents a new method for ROI segmentation and Bit plane detection. It is observed that when a large area covered by multiple cameras, most of the camera focus on a particular area. This results in a common area and it is expected high contrast information. This leads to proposing a method for segmenting common region as ROI. In the same way, in case of image representation, the most significant bit provide important information compared to other bits. This observation motivates the proposed work to introduce bit detection for image size reduction. This chapter is divided into two sub-parts, namely the method for ROI segmentation and the method for bit plane detection, which are presented in subsequent sections.

4.1.1   ROI Segmentation

It is expected that input images can have any size and content. It is a true that when the system deploys multiple cameras, in general, all cameras focus on particular regions of the images, if the region contains any useful information (Wang, Peng et al. 2007, Wang, Peng et al. 2010). To choose the overlapping region that contains useful information, sharpness estimation for the overlapping regions is proposed. This stage finds the region, which gives the highest sharpness compared to other regions in the image

as Dominant Overlapping Regions (DORs). If the proposed method does not find overlapping regions, it considers the whole image as input for the image reduction process. Since different camera nodes have different resource capacities, to take advantage, the DOR region is further divided into sub-DORs according to the capability and facilities of the cameras, which is called a partition process. To reduce the size of the sub-DOR further, inspired by the method in (Wang, Peng et al. 2007, Wang, Peng et al. 2010) where the Mean operation is used to remove unwanted information from the images, a Median operation is proposed. This is due to the fact that the Median operation does not introduce blur during the process while the Mean operation introduces blur. This is valid because the Median operation chooses the middle value, which is neither noise (high value) nor non-significant pixels (low value) while Mean considers both. The output of the Median operation over sub-DOR is considered as a Compressed Significant Region (CSR). The non-DOR is not an important region compared to DOR. However, sometimes, the situation demands other region information along with sub-DORs. Instead of sending the color information, the proposed method sends Sobel edges of non-DORs to throughout VSN. The main reason to choose the edges is that that the Sobel operator gives fine edges for high contrast information in the image in binary form, which helps in reducing the data. Finally, in the proposed method, CSR and the Sobel of non-DORs are passed to the VSN. In this way, the proposed method achieves image size reduction for collaborating images, along with retaining the quality of the image.

4.1.2   Bit Plane Detection

When the system does not find overlapping region, the system considers the whole image as input for transmission. To reduce the size of such images, the method is proposed for bit plane detection. This chapter also works on intra redundancy and ROI segmentation based bit plane detection to reduce image size. (Harel, Koch et al. 2007), a new method is proposed based on saliency map segmentation for detecting such

informative bit planes automatically. However, due to background and foreground variations especially for the scene images shown in Figure 4.1, one cannot expect that the most significant bit plane (higher order bit plane) provides significant information for all the situations. It is evident from Figure 4.1 that bit planes B_3, B_4, and B_5 being lower orders, provide better information compared to the higher order bit planes, B_8 and B_7, which are expected to carry better information.

Input image	B_1	B₂
B₃	B₄	B₅
B₆	B₇	B₈
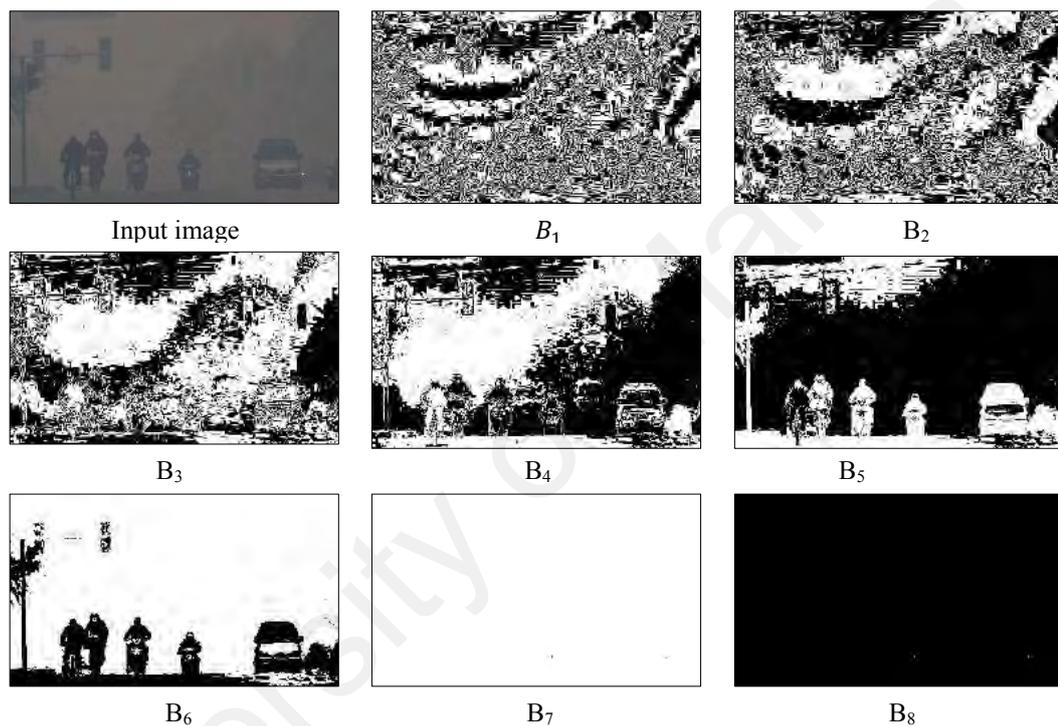
Figure 4.1: Bit planes for the input images

It is noted that informative content in the image usually draws the attention of the user (Harel, Koch et al. 2007). In other words, the important and significant part of the image has high contrast compared to its background. Inspired by this fact, the method is proposed for extracting saliency, which represents region of interest (significant information in the image). For input image, the proposed method obtains eight bit planes as shown in Figure 4.1, as in B₈, B₇, B₆, B₅, B₄, B₃, B₂ and B₁. For each bit plane, the proposed method estimates saliency based on variations in neighbor pixels. When the background and foreground is affected by large background and foreground variations, the intensity of grayscale background and foreground pixel changes. It is evident from bit

planes, B_3 and B_4 that the intensity of objects is large in B_4 and small in B_3. The same thing is true for background intensities. To reduce the computational burden and the impact of intensity changes, the proposed method obtains Canny edge map of salient segmented region from each plane. In order to study the distribution of saliency in the plane images, a ring-growing method is proposed.

The whole flow chart of the whole method is shown in **Error! Reference source not f ound.**.
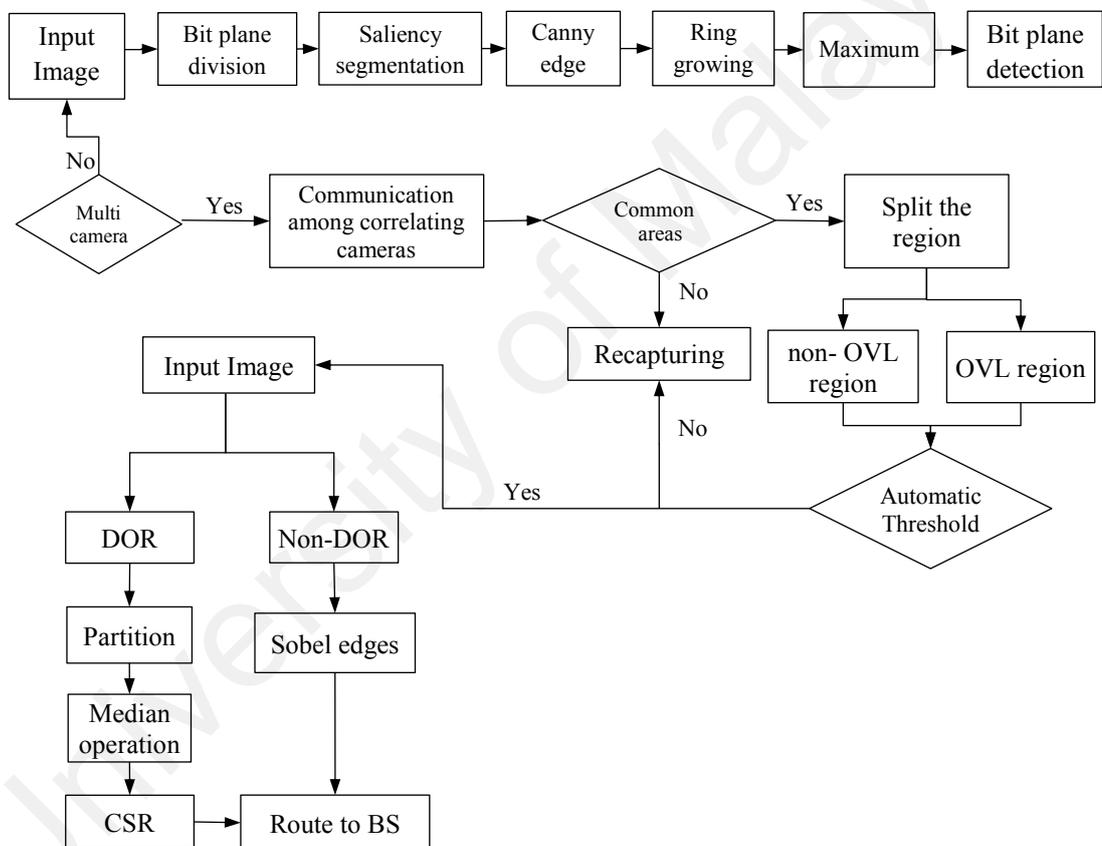
Figure 4.2: Block diagram of the proposed method

4.2 Proposed Method for ROI Segmentation

This section describes the proposed ROI segmentation including two subsections, namely, OVL and non-OVL region detection and partitioning and also size reduction model for optimal transmission.

4.2.1 OVL and non-OVL Region Detection and Partitioning

A sample VSN environment can be seen in Figure 4.3, where the camera nodes inside the clusters transmit images to a Cluster Head (CH). The network consists of N^c number of camera nodes, say, C_i, $i=[1,2,..,N^c]$ with a predefined initial energy E_i^{in}. When camera C_i detects an event and captures an image, it broadcasts a query packet to its neighboring camera nodes to find the overlapping region. The responses or communication between cameras indicates that there are overlapping regions. With this networking setup, the proposed model finds overlapping regions by performing an intersection between the areas captured by correlating cameras.
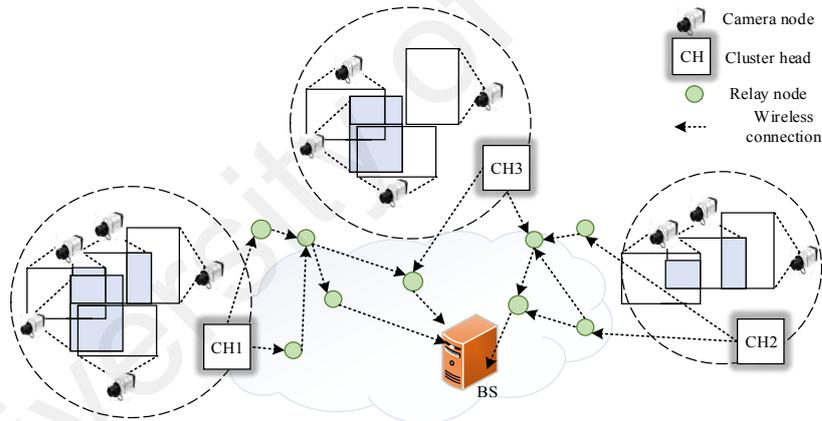
Figure 4.3: Sample network system for image transmission

The process of finding overlapping regions is illustrated in Figure 4.4 (a) where we can see the nodes are fixed to find overlapping regions from the view of two cameras. Sometimes, the systems give small overlapping regions or big overlapping regions due to different locations and scenarios. If the overlapping region is too small, the proposed method ignores it because it does not contribute. Therefore, lower and upper limits for overlapping regions are defined (Costa, Duran-Faundez et al. 2017). According to the literature and the experiments, if the size of overlapping regions is lower than 10% of the

image captured by a camera, it is considered as being too small and hence it is ignored. Similarly, if the size of the overlapping region is greater than 10% and less than 50% of the captured images, it is considered as an actual overlapping region. If the overlapping region is almost the same as the input image size, the proposed method considers the whole image as input for the image reduction process. However, there is no guarantee that the overlapping region provides useful information at all times because the content of overlapping regions vary from one location to another and camera to camera. Therefore, the sharpness for the overlapping region is estimated to highlight the vital information. When two cameras are focusing at the same point (overlapping region), one can expect high sharpness at the overlapping region compared to other regions of the captured images. Based on this observation, the sharpness for the overlapping region is computed as defined in equations (3.4) and (3.5). If the sharpness of the overlapping region is higher than other regions of the captured images by the cameras, the proposed method considers it as an actual overlapping region that provides dominant information. If the sharpness of the overlapping region is lower than the other region, the proposed method rejects the overlapping region and images are captured again with the cameras. In this process, the proposed method considers the overlapping region as a window and the same window is moved over captured images as shown in the steps in Figure 4.4, where (b) and (c) show images captured by camera C_1 and C_2, respectively, and (d) is an overlapping region detected by the camera locations. For visualization, the proposed method performs max-min clustering on the sharpness of the whole image, which classifies high sharpness values into one max cluster and low sharpness values into a min cluster. The output of max clusters of image-1 and image-2 are shown in Figure 4.4 (e) and Figure 4. 4 (f), respectively. It is evident from Figure 4.4 (e) and Figure 4.4 (f) that the window marked by an orange color has more details compared to the windows marked by a red color. This indicates that the overlapping region marked by the orange color

gives high sharpness compared to the other window marked by a red color as shown in Figure 4.5(a). Therefore, it can be asserted that the overlapping region contains useful or focused information, which is called a Dominant Overlapping Region (DOR). Sample overlapping regions that give dominant information for the different images are shown in Figure 4.5 (b).

(a) Example of finding an overlapping region using a two-camera set-up

(b) Image captured by C_1 (c) Image captured by C_2 (d) Dominant overlapping region

(e) Max cluster sharpness of C_1 (N11, N12, DOR)   (f) Max cluster sharpness of C_2 (DOR, N21, N22)

N11, N12: non-overlapping regions of the C_1 image; N21, N22 denote non-overlapping regions of C_2
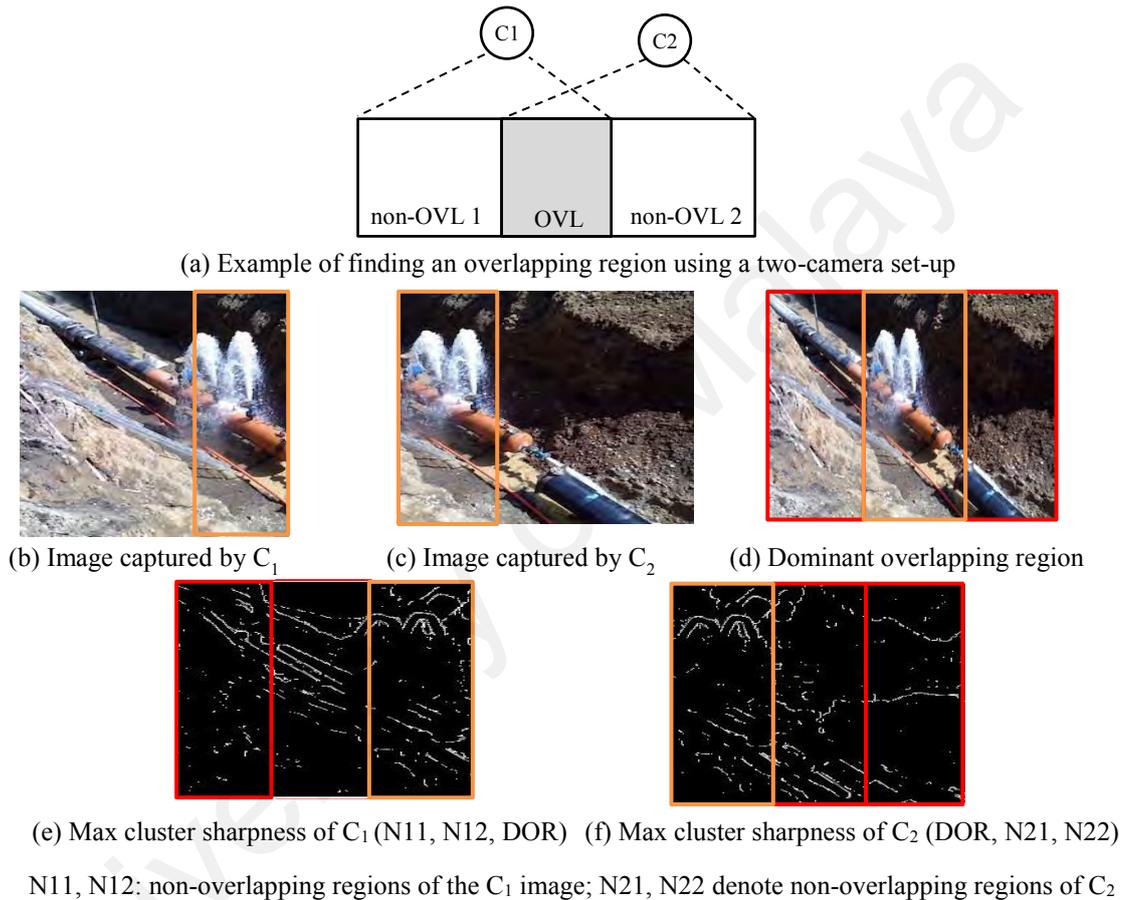
Figure 4.4: Finding an overlapping region that provides dominant information
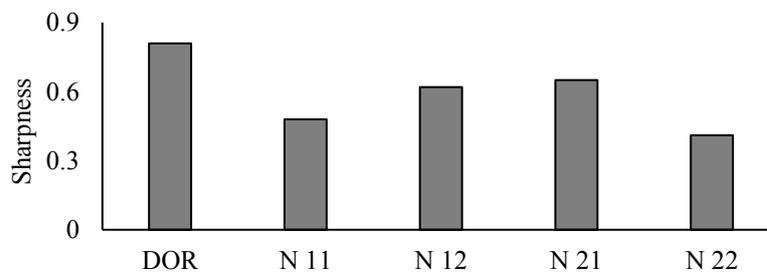
(a)   Sharpness of overlapping and non-overlapping windows of the captured images.

(b) Example of overlapping regions that provide dominant information

Figure 4.5: Dominant Overlapping Regions (DORs) based on sharpness

In this way Region of Interest (ROI) is segmented carefully and will be sent by the cooperative cameras only once to remove inter redundancy between images. In order to utilize the resources of each camera node, the proposed model partitions the DOR region further into sub-DORs according to energy levels at camera nodes, as shown in Figure 4.6 (a), where sub-DOR-2 is a small region requested by camera node C_2 and sub-DOR-1 being larger than sub-DOR-2 is requested by another camera node C_1. The energy share of camera node i is defined as E_i^s in equation (4.1). For example, a sample blue colored DOR region is shown in Figure 4.6 (a) where C_1 with residual energy $E_1^r = 45\ J$ and C_2 with $E_2^r = 15\ J$ divide the share of the DOR based on the energy proportion defined in equation (4.1) where N^{cc} is the number of correlating cameras. This shows that 75% of the DOR (sub-DOR-1) are transmitted through C_1 and 25% of DOR (sub-DOR-2) are transmitted through C_2 as shown in Figure 4.6 (a). The process of partitioning is illustrated on the sample image in Figure 4.6 (b).

$$E_i^s = \frac{E_i^r}{\sum_{j=1}^{N^{cc}} E_i^r} \tag{4.1}$$
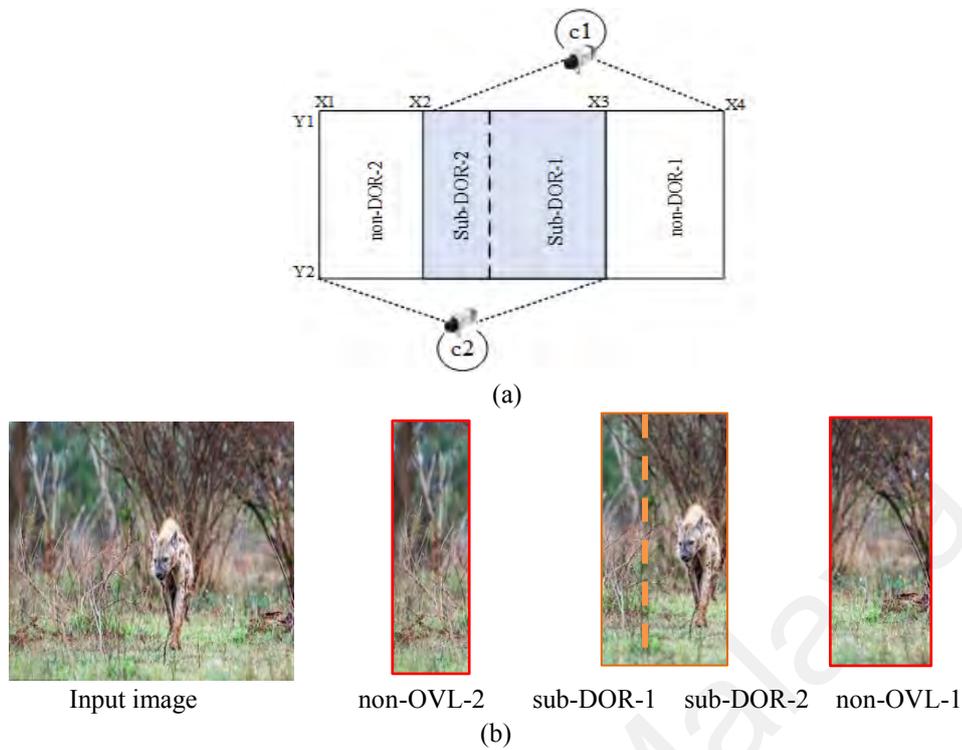
Figure 4.6: Sub-DOR and non-DOR after partition

4.2.2   Size Reduction Model for Optimal Transmission

Sometimes, segmented DORs may still contain intra redundancies and unwanted information due to some operations. The proposed method performs a Median operation, which eliminates unwanted information from the images without affecting vital information. In contrast to the Mean operation used in (Wang, Peng et al. 2007, Wang, Peng et al. 2010), the Median operation is better in terms of quality. Therefore, one can say that the Median operation can also reduce the size of the images, which enhances VSN performance further. It is evident from Figure 4.7 (a), where it can be observed that as the window size S^w increases, the blurriness increases, at the same time, the number of bytes N^b also decreases. On the other hand, in the case of the Median operation, though the number of bytes decreases, the quality does not decrease as for the Mean operation, as shown in Figure 4.7 (b). This shows that the Mean operation is too sensitive to content, but the Median is not. However, it is confirmed that the quality of the image and size of the image depend on the window size.

82

Sw:	2	3	4	5	6	7	8	9
N^b:	0.22	0.19	0.16	0.15	0.13	0.12	0.11	0.10

(a)

N^b:	0.23	0.20	0.18	0.16	0.15	0.13	0.13	0.11

(b)
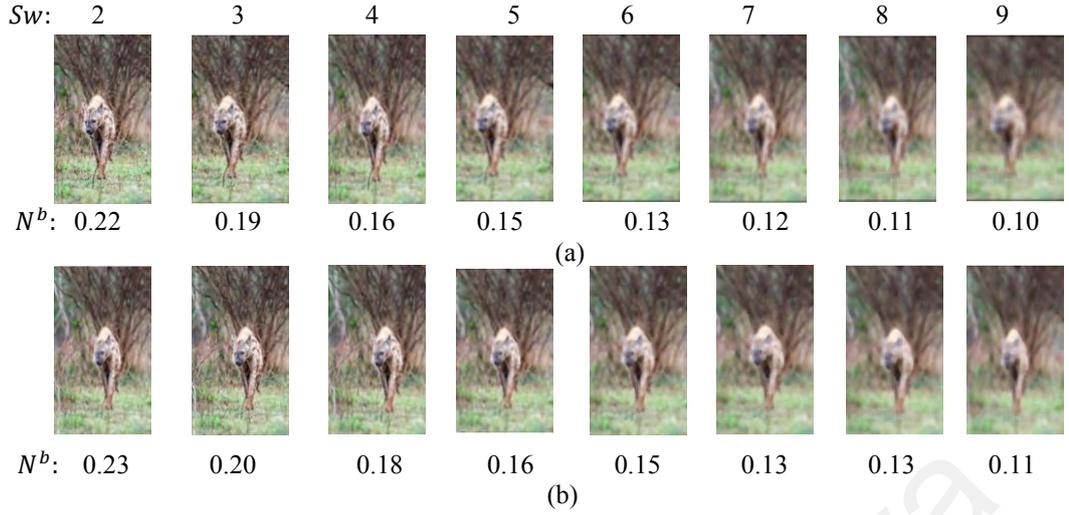
Figure 4.7: The results of Mean (a) and Median (b) operations on a sample DOR image for different windows (S^w), N^b denotes the number of bytes.

To determine window size automatically, a new idea is proposed which estimates quality of the DOR by varying the window size. For this, standard quality measures, namely $PSNR$ and $SSIM$ are used as defined in equation (4.2) and (4.3), respectively.

$$PSNR = 10 \times log_{10}(P_{v^{max}}^2 / MSE) \tag{4.2}$$

$$MSE = 1/S^I \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - I'(i,j)]^2$$

$$SSIM(x,y) = \frac{(2M_xM_y+c_1)\times(2CV_{xy}+c_2)}{(M_x^2+M_y^2+c_1)\times(V_x^2+V_y^2+c_2)} \tag{4.3}$$

In equation (4.2), $P_{v^{max}}^2$ is the maximum possible value of the image's pixels and MSE is the mean square error. MSE is the summation of all squared differences of image I and I', as the result of window operations divided by image size S^I, n and m are image height and width. In equation (4.3), $SSIM(x,y)$ is a similarity metric in which M_x, M_y and V_x^2, V_y^2 are the average and variance of windows of x and y with a defined window size, while CV_{xy} denotes their covariance. Note that to calculate these two measures, the original DOR is used as ground truth and the result of the window operation as processed images. Higher $PSNR$ and $SSIM$ values indicate high quality of the image. This is

illustrated in Figure 4.8 (a) and Figure 4.8 (b) where for window size S^w=3, both $PSNR$ and $SSIM$ score high values. This shows that window size S^w=3 is appropriate for the DOR. In this way, the proposed model determines window size automatically.
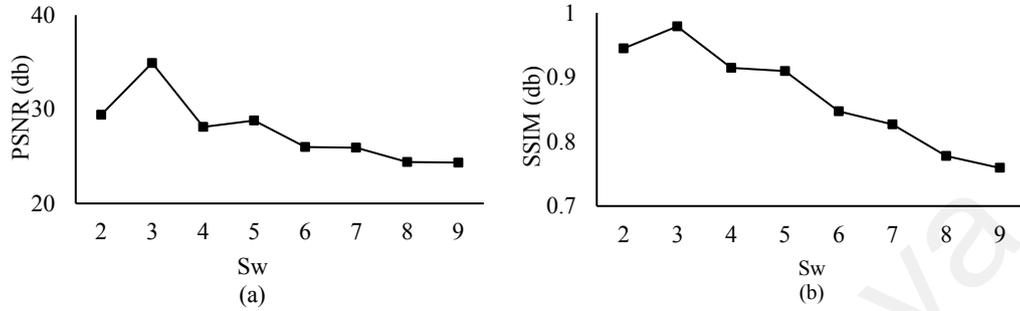
Figure 4.8: Illustration for choosing S^w for image size reduction

The proposed method sends the result of the Median operation of DOR to the VSN, which is called the Compressed-Significant Region (CSR). As mentioned in the proposed method section, the non-DOR is not important for the VSN, and thus Sobel edges of the non-DOR is sent along with the CSR, as shown in Figure 4.9. Where we can see for the sample input image in (a), the CSR given by Median operation with S^w=3 and Sobel edges for the non-DOR are the final data to be sent to the VSN, as shown in Figure 4.9 (b). When the total size of the final image is calculated according to equation (4.4), it gives 8.7 *kb*, which includes 5.4 *kb* of CSR and 3.3 *kb* for the Sobel of the non-DOR in contrast to the total size of the original image, which is 37.9 *kb*. This shows that the proposed model uses the resources of the camera nodes efficiently in terms of image size reduction and quality of the image. In equation (4.4), S^I_{tr} denotes the total size of the image that will be sent, S^{CSR} denotes the size of the processed overlapping segment and $S^{non-DOR}$ denotes the size of the non-overlapping region.

$$S^I_{tr} = S^{CSR} + S^{non-DOR} \tag{4.4}$$

(a) Sample input image　　　　(b) Final image sent to VSN (CSR+ Sobel of non-DOR)
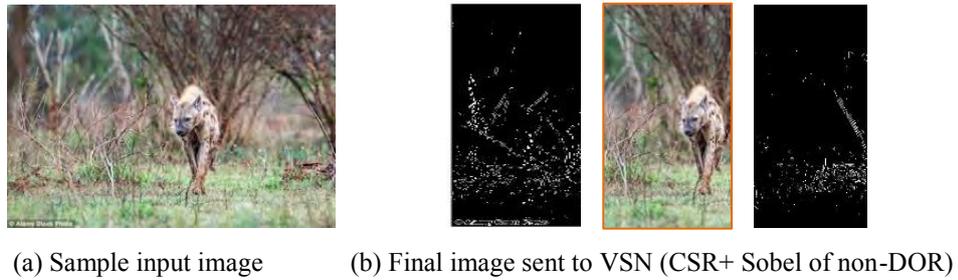
Figure 4.9: Compressed significant region with edge information of the non-DOR

4.3　Proposed Method for Bit Plane Detection

This section describes the proposed Bit plane detection including three subsections, namely, saliency detection for the planes, informative bit plane detection, overview of general TCP/IP network.

4.3.1　Saliency Detection for the Planes

The method presented in the previous section works well when system gets overlapping region from the images captured by multiple cameras. It is not true sometimes and therefore, there are chances of getting images without overlapping regions. In this case, to reduce size of the images, the proposed work aims at bit plane detection based on saliency detection in the images. (Harel, Koch et al. 2007). It consists of two steps: the forming of activation maps from feature channels and their normalization that highlights the salient regions. The effect of these steps are shown in Figure 4.10 for the respective planes shown in **Error! Reference source not found.** where SA_8, SA_7, SA_6, SA_5, SA_4, SA_3, SA_2 and SA_1 denote saliency. It is evident that the bit plane 3 gives better information compared to all other bit planes, since it clearly indicates objects in the input images while other bit planes do not.

SA$_8$ SA$_7$ SA$_6$

SA$_5$ SA$_4$ SA$_3$

SA$_2$ SA$_1$
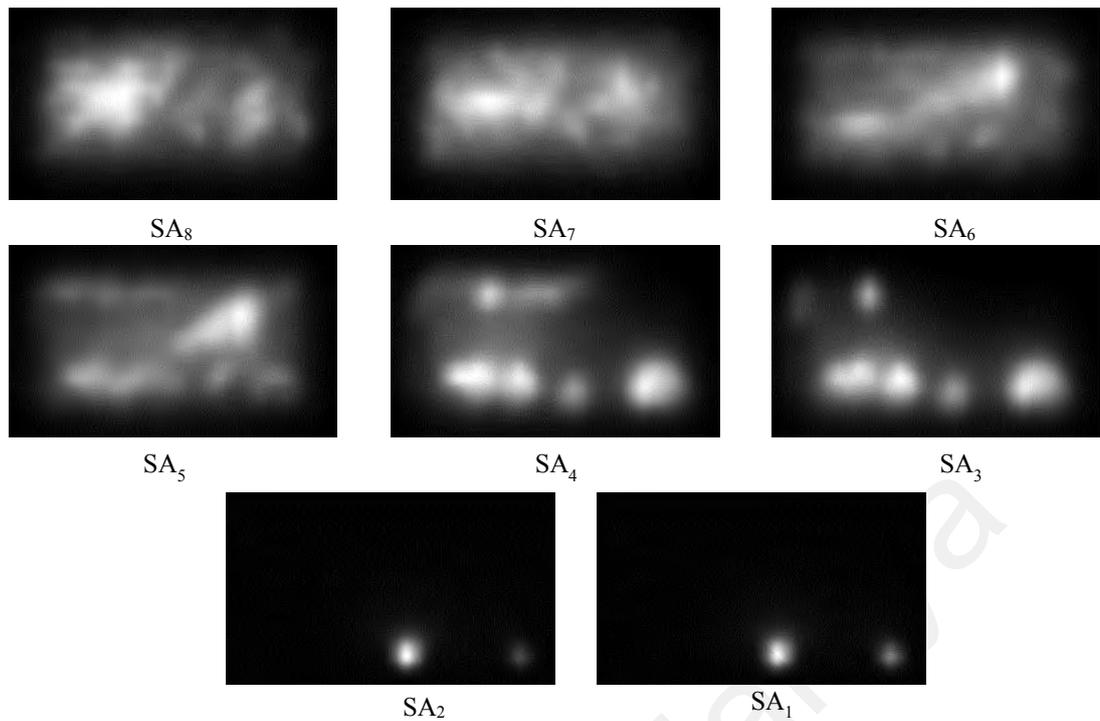
Figure 4.10: Saliency of respective planes in Figure 4.1

4.3.2 Informative Bit Plane Detection

To save number of computations, the proposed method obtains Canny edge maps as shown in Figure 4.11 for each saliency map in Figure 4.10 because Canny edge maps are in binary and the features are extracted at edge level rather than pixel level. The bit plane 3 and 4 give clear information about the objects in the input image compared to the other planes. When we compare bit plane 3 and 4, the bit plane 3 is better than the bit plane 4 because it gives clear indication of all the objects in the input image while the bit plane 4 misses some information especially the top-left corner information. Therefore, it is expected that the bit plane 3 should provide high saliency values where objects are located in the image compared to other planes.

<div align="center">

CA₈ CA₇ CA₆

CA₅ CA₄ CA₃

CA₂ CA₁
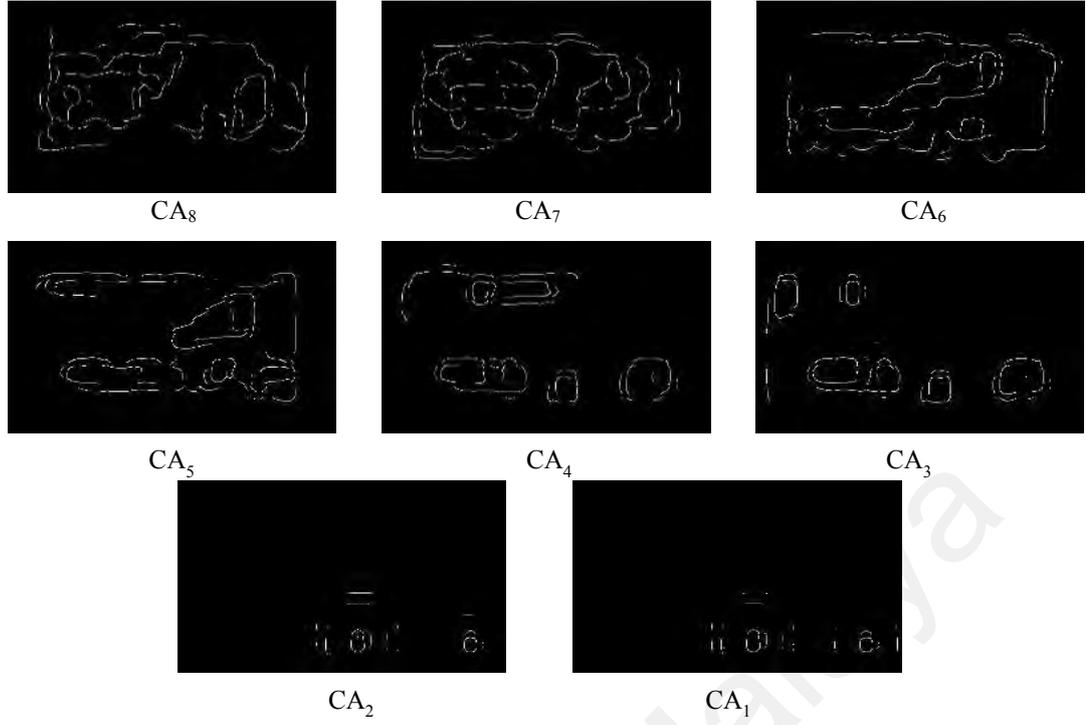
Figure 4.11: Canny edge maps of the saliency images in Figure 4.1

</div>

To extract the above observation, the proposed method draws rings over Canny edge maps from center with radius of 10 pixels. In other words, 10 pixels is the step size for drawing rings over the Canny maps from the center. Sampled bit plane 3 with centroid marked by red color and ring growing are shown in Figure 4.12 (a) and Figure 4.12 (b), respectively. Centroid is center of all existing canny pixel of each map image. The ring growing terminates when it reaches farthest pixel from the centroid. To study the spatial distribution of saliency, the proposed method calculates mean of saliency values $\overline{SA_r}$ of edge pixels for a ring as defined in equation (4.5).

$$\overline{SA_r} = \frac{\sum_{i=1}^{N_r^p} SA(P^i)}{N_r^p} \tag{4.5}$$

In this equation, $SA(P^i)$ is corresponding saliency value of ith pixel in a ring r, N_r^p shows the number of pixels in that ring and finally, $\overline{SA_r}$ is the average saliency value of rth ring.

(a) Centroids   (b) Ring growing
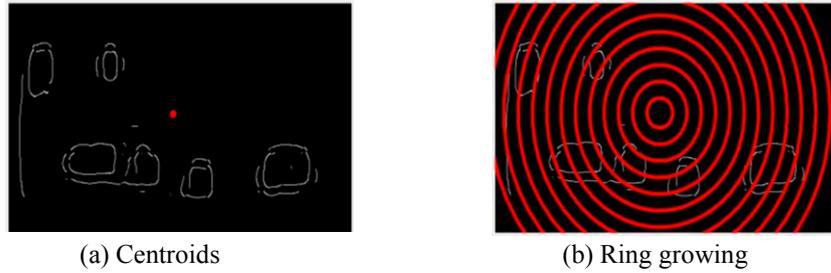
Figure 4.12: Ring growing for the Canny edge maps of the planes

Since saliency provides high values where there is object in the image, the mean gives high value for the ring, which covers object pixels. The histogram for the number of rings versus mean values are plotted as shown in Figure 4.13, where the mean values marked by red color reaches maximum compared to other bit planes. As a result, the bit plane 3 gives highest number of mean values, which reaches maximum as defined in equation (4.6), where P_i is a counter to determine i leading to the highest max values. Therefore, the bit plane 3 is considered as informative plane for the input image. The steps for bit plane detection are presented in the form of pseudo code in Algorithm-1.

$$B_i = \begin{cases} B_i + 1, & if\ i \in Max\ [\overline{SA_r}]\ , i = [1, 2, \dots, 8] \\ B_i, & else \end{cases} \tag{4.6}$$
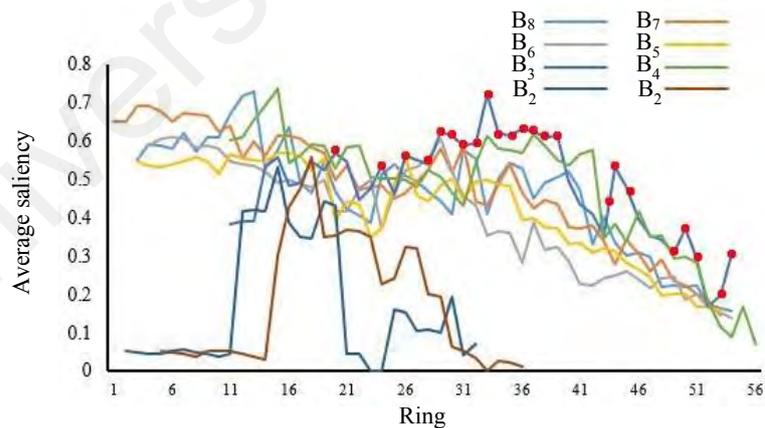
Figure 4.13: Mean of saliency vs Rings in Figure 4.12: Ring growing for the Canny edge maps of the planes

88

Algorithm 4.1: Plane detection

```
1:   Initialization: r =[1,2,…,N^r], i=[1,2,..,8]
2:   For CA_i  do
3:   Compute center coordination  (P_x , P_y)_center
4:   Compute Ring masks with center of
5:   For each r  do
6:   Compute SA_r
7:   End For
8:   End For
9:   For each r in all CAs do
10:  Update B_i=B_i+1 where i^th plane has maximum SA_r
11:  End For
12:  Choose Plane with the highest B_i
```

4.3.3 Overview of General TCP/IP Network

The main objective of this proposed method is to detect the significant bit plane out of eight planes such that image data can be sent through network smoothly, and efficiently without loss, of information and quality. Therefore, a general TCP/IP network is considered for testing whether the bit plane detection helps in improving the performance of network applications. The general TCP/IP network as shown in Figure 4.14 is the basic requirement for all the network applications, such as Wireless Sensor Network, Visual Sensor Networks, etc. If the bit plane detection improves performance at this stage, one can expect still higher improvement for other advanced applications. For transmitting data, widely protocols used in client-server computing paradigm are Transport Control Protocol (TCP) and the User Datagram Protocol (UDP). Since TCP is suitable as per our objective and it is popular, TCP is preferred for experimentation in this chapter. It is fact that the process of transmission from client to server involves all the layers of Open System Interconnection (OSI) models (Briscoe 2000). In general, this process works well for data of small size. When the data is huge (for example images), the network divides into several packets according to strengths and constraints of network. If any error exists during the division, it affects all other layers and processes. Therefore, it may results in

either loss or extra information after reaching destination as compared to data at the original point. Due to this effect, one can expect received data with poor quality.

Because of network delay, the network retransmits the data leading to network congestion at the transport layer, which in turn leads to loss of data (Fall and Stevens 2011). This is one key reason for the loss and/or extra information for the huge data like images. In the same way, one expect loss and extra data from other layers due to improper information. For instance, the TCP protocol contains a parameter known as the maximum segment size within TCP segment that specifies the maximum amount of data (in bytes), which can be received by a node in a single TCP segment. This is required to send packets smoothly over the network. In case the data sent is over or lower the threshold of MSS, it again results in buffering of incoming packets, resulting in packet loss or delay. The pseudo code for data transmission through TCP/IP is presented in Algorithm-2.

Figure 4.14: Data delivery in TCP/IP Network

The pseudo code of the approach is given here:

Algorithm 4.2: image transmission over TCP/IP
1: Initialization: server port-number, Server address
2: **If**   Server request denied
3:          Terminate connection
4: **Else**
5:          Establish connection
6:          Receive ("input ")
7: **End If**
8: Client establish connection
9: **For** (list of images) **do**
10:     Read image ("input")
11:     Send to server ("input")
12: **End For**

The ROI segmentation method requires two loops for finding a region and hence the time complexity of this method is O(n2) for the worst and best cases. Hence, it is a theta time complexity. The same thing is true for bit plane detection method because the method processes the whole image to detect the significant plane and therefore, the time complexity of the bit plane detection is theta time complexity.

4.4 Experimental results

In this section, the proposed method is evaluated in term of the image quality and network lifetime enhancement. The section is organized as follows. Section 4.4.1 describes the created dataset and evaluation method. Section 4.4.2 reports the experiments and comparative study on ROI segmentation and experiments on bit plane detection are explained in section 4.4.3.

4.4.1 Dataset and Evaluation

Since there is no standard or benchmark dataset for evaluating the performance of the proposed method, a dataset for experimentation in this chapter is created. Situations are considered, where networking is essential to monitor, namely, animal behavior, flood levels, plant growth, cracks in the soil or water pipes and cleaning garbage, vehicles, buildings, traffic and air pollution. In total, 500 scene images are considered for the experimentation. To capture the images, a low cost camera is used with a resolution of 600×450 pixels. Image size are 300×500 which are taken in different lightning, weather conditions and daytime from Malaysia, Iran, India and Germany[3]. Sample images are shown in Figure 4.15, where one can guess that the scene images are complex in nature due to variations in background and foreground. The reason to choose the above set of

---

[3] The data set will be shared upon request for supporting reproducibility.

images is that there are many systems with advanced networking installed for the purpose of monitoring and surveillance of these image types.

The initial energy for each node i is set to $E_i^{in}= 50\ J$. Each camera node has an ability to compress the images using a compression algorithm until the energy reaches a minimum threshold. The distance between the node and the *CH* is $d = 10$ meters.

| Animal | Flood | Plant |

| Pipeline | Garbage |

Figure 4.15: Sample images of collected dataset for the experimentation

The proposed ROI segmentation method comprises two key steps, namely, finding the Compressed Significant Region (CSR) and the use of resources of the camera nodes in the VSN. Therefore, standard measures are used, namely, *PSNR* and *SSIM* as mentioned in the previous section for evaluating the quality of the CSR. In the same way, to assess the performance of the networking system, the energy model is defined as used for communication in existing methods (Nandhini, Sankararajan et al. 2015, Nandhini and Radha 2016), as defined in equation (4.7), where E^{bit} is the transmission energy per bit, T^{tr} denotes the transmission time on a 250 kbits/s IEEE 802.15.4 link, u and v denote the current and voltage of the *TelosB* node. Accordingly, the computed transmission energy per bit is 0.23 μJ.

$$E^{bit} = \frac{T^{tr} \times u \times v}{1024} \tag{4.7}$$

Average Residual Energy $\overline{E^r}$ is defined in equation (4.8) as the average residual energy of N^c nodes in the nth iteration. E_i^r denotes energy at the ith node and the nth iteration. This measure helps us to evaluate the efficiency in terms of energy during image transmission. In addition, Network Lifetime (T^{nl}) is defined in equation (4.8) where N^{it} is the number of iterations that the whole cluster can capture, process and transmit images before reaching a minimum energy threshold, T^{itr} is the estimated time for each iteration and e and h denote minutes and hours respectively (Nandhini, Sankararajan et al. 2015, Nandhini and Radha 2016).

$$\overline{E^r} = \frac{\sum_{i=1}^{N^c} E_i^r}{N^c} \tag{4.8}$$

$$T^{nl} = \frac{N^{it} \times T^{itr}}{e \times h} \tag{4.9}$$

Moreover, the proposed bit plane detection consists of two stages, namely, informative bit plane detection and plane data transmission through TCPT/IP network. To evaluate the performance of each stage, average Detection Rate ($\overline{DR}$) is used as defined in equation (4.10) where N^{cd} denotes correctly detected plane and N^I denotes the number of images and F-measure as defined in equation (3.20) for bit plane detection. Error estimation also is defined in equation (4.11) for image data transmission as the absolute difference between the image in byte at a transmitter and a receiving end. S_{tr}^I and S_{re}^I are image size at transmitter and receiver stations, respectively.

$$\overline{DR} = \frac{N^{cd}}{N^I} \tag{4.10}$$

$$\overline{ER} = \frac{\sum_{i=1}^{NI} |S_{tr}^I - S_{re}^I|}{N^I} \tag{4.11}$$

In addition, to measure the quality of the image data at the destination, Entropy is estimated as defined in equation (4.12) where H is normalized histogram counts of input image. It is a fact that when the image contains high contrast information, entropy gives

high values compared to low contrast information as it is statistical measure for randomness and texture of the images (Gonzalez, Woods et al. 2011). The high value of entropy indicates good quality. In this chapter, it is expected that the image data at received end must have high quality and minimum error compared to input image.

$$ET = -sum(H \times \log_2 H)\qquad(4.12)$$

A state-of-the-art method (Wang, Peng et al. 2010) is implemented to compare with the proposed ROI segmentation method in terms of image quality and other networking parameters. The reason to choose this method is that the objective of this method is same as the proposed method. It uses a Mean operation for image size reduction for the overlapping region as in the proposed method. For image size reduction, there are Distributed Source Coding (DSC) systems, which aim at improving the performance of the visual sensor network without losing quality of the images when the image data is huge. In order to show that the proposed method is superior to DSC, DSC is implemented as proposed in (Wang, Li et al. 2007).The DSC usually captures images by different sensors and cameras and then use different level of compression according to redundant information in the captured images. As redundancy increases, the performance of the DSC increases, else the performance decreases. However, in this chapter, 100 images captured by two cameras are passed to DSC for the experimentation. The level of compression is determined empirically by experimenting on pre-defined cases (scenarios) chosen randomly for the images in the database. It is expected that as the size of the input images increases, the performance of the DSC decreases in terms of quality, efficiency, energy etc. This is due to more operations involved in compression while distributing image data of a huge size. On the other hand, the proposed method finds CSR, which passes through the visual sensor networks, and is independent of input size. Therefore, the proposed method works well irrespective of the size of the images.

To show that the bit plane detection proposed steps are useful and effective, the method is compared with methods in (Dutta, Mandal et al. 2007, Felemban, Sheikh et al. 2014, Raghunandan, Shivakumara et al. 2018) .The reason to consider these methods is that (Dutta, Mandal et al. 2007, Raghunandan, Shivakumara et al. 2018) methods are proposed for general images and the images contains text, respectively. The study in (Felemban, Sheikh et al. 2014) is proposed for improving network performance. Since the considered images suffer from distortion, poor quality and complex backgrounds, these images are complex compared to general images and video text images. For comparative study on network part, two methods are implemented, namely (Felemban, Sheikh et al. 2014) which sends set of fixed bit planes (the bit planes, 6-8) for transmission and (Junaid and Shunmuganathan 2017) which sends the full images without bit plane detection. It is expected when detected plane is transmitted through network, it should conserve quality and no loss of the information compared to the original images, all the eight planes and set of planes.

4.4.2   Experiments for ROI Segmentation

To evaluate the proposed ROI segmentation two set of experiments are conducted in this section, namely, experiments for measuring image quality and experiments for networking system.

4.4.2.1 Experiments for Measuring Image Quality

To validate the image quality of CSR given by the proposed ROI segmentation method and existing methods, *PSNR* and *SSIM* are calculated for the output of the methods as shown in Figure 4.16 (a) and Figure 4.16 (b), respectively. It is observed from Figure 4.16 (a) and Figure 4.16 (b) that the proposed method reports high *PSNR* and *SSIM* for all the images compared to the existing methods. This is valid because the method (Wang, Peng et al. 2010) use Mean operations with a fixed sized window for removing unwanted

information, which introduces blur during the operation. As a result, the existing method loses quality of the image and hence leads to poor *PSNR* and *SSIM*. Since the performance of DSC depends on size of the image, as size varies the method loses the quality of the image. Therefore, the method reports low *PSNR* and *SSIM* compared to the proposed method. However, when the results of the two existing methods are compared, (Wang, Li et al. 2007) is better than the method in (Wang, Peng et al. 2010). On the other hand, the proposed method performs a Median operation for the sub-DORs given by a common region, sharpness and partition with a dynamic window size, that is why the proposed method reports the best score compared to the existing methods.

(a)

(b)
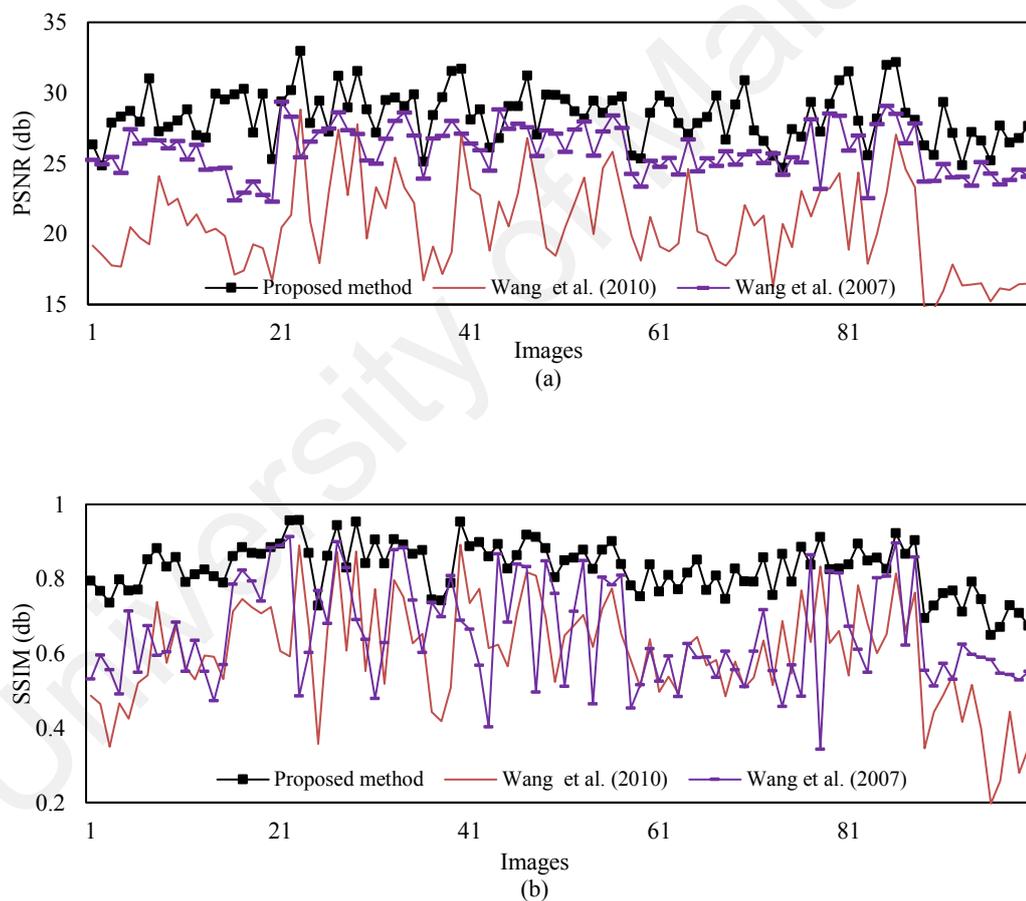
Figure 4.16: The quantitative comparison of the proposed and existing methods

4.4.2.2 Experiments for Networking System

In addition, for this set of experiments, 100 different scenarios are set for evaluating the proposed ROI segmentation and existing methods in terms of $\overline{E^r}$ and T^{nl}. To calculate

$\overline{E^r}$, 100 random scenarios are considered. $\overline{E^r}$ is then estimated for the 200ᵗʰ iteration of 100 scenarios to evaluate the proposed and the existing methods. (Wang, Peng et al. 2010) is used for this experiment because this method involves the same steps as the proposed method. The proposed method involves three key steps, namely, finding Overlapping Regions (OVL), finding Sub-dominant Overlapping Region (Sub-DOR) with the partition step, and the last Compressed Significant Region (CSR). To show these steps consume less energy and give a higher lifetime for the network compared to the whole image, $\overline{E^r}$ and T^{nl} are calculated by passing the whole image, OVL, Sub-DOR and CSR to the VSN as reported in Figure 4.17 (a) and Figure 4.17 (b), respectively. It is observed from Figure 4.17 (a) and Figure 4.17 (b) that the proposed and existing method report high $\overline{E^r}$ and T^{nl} for the steps compared to the whole image. This shows that steps are effective for enhancing the VSN performance. It is also noted from Figure 4. 18(a) and Figure 4. 1 8(b) that the proposed method scores the best $\overline{E^r}$ and T^{nl} for the sub-DOR and CSR compared to the existing method. However, the proposed and existing methods score the same for the whole image and OVL. This is valid because the methods use the same criteria for finding an overlapping region.
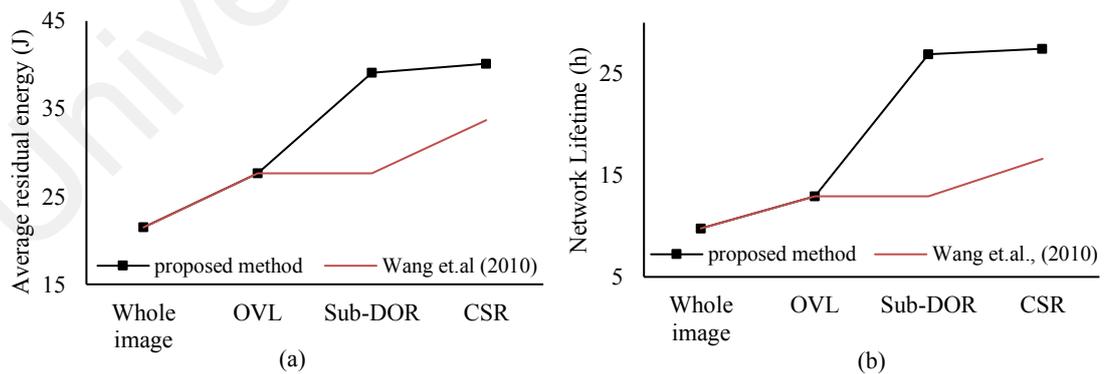
Figure 4.17: Comparing $\overline{E^r}$ and T^{nl} of the proposed system with the existing system with N^c=5

In addition, T^{nl} is calculated for the proposed method and existing methods including DSC by varying the number of cameras (N^c= 2, 3, 4, 5), as shown in Figure 4.18 (a)

where it can be noted that T^{nl} is higher for the proposed method compared to both the existing methods. The T^{nl} of (Wang, Peng et al. 2010) is higher than the DSC (Wang, Li et al. 2007), which is poorer than the proposed method. This is due to complex operations for finding compression in the case of DSC compared to a simple Mean operation. It was also found from Figure 4.18 (a) that the results of the proposed and existing methods increase as the number of cameras increase. This is because in the case of networking, as the number of choices increase, the network systems choose the best one for transmission. The main reasons that the existing method demonstrates poorer performance compared to the proposed method are the same as discussed above. In this experiments, five low cost cameras are used with $d = 10$ metre distance for simulation. In realtiy, one can expect different setups for the different situations. It is true that determination of the exact number of cameras for every situation is hard. In order to decide the number of cameras, we should have some knowledge of the location, situation, and condition, etc. In the same way, the distance between the camera is also an improtant parameter that affects the performance of the VSN as it depends on the area, capacity of the camera nodes and cost. One such illustration is shown in Figure 4.18 (b), where we can see that as the distance d increases, the nubmer of possible iterations decreases.
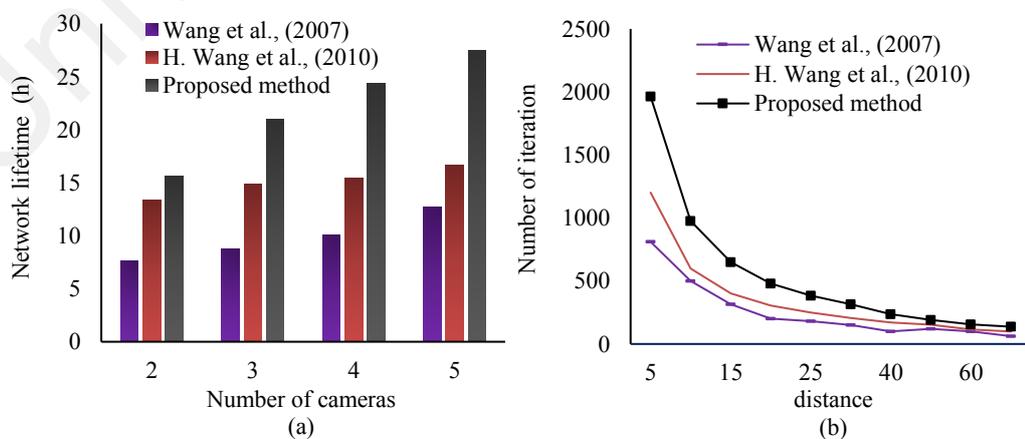
Figure 4.18: Comparing performance of the proposed and existing methods in terms of T^{nl} with N^c=2, 3, 4 and 5 (a) and N^{itr} with different d (b)

4.4.3 Experiments for Bit Plane Detection

To evaluate the proposed bit plane detection method, two set of experiments are conducted, namely, evaluation on bit plane detection and validating bit plane transmission through TCP/IP Network.

4.4.3.1 Evaluation on Bit Plane Detection

Sample quantitative results of the proposed and existing methods for the bit plane detection are shown in Figure 4.19 where we can see for different input images affected by poor quality and complex background, the proposed method detects informative bit plane correctly while existing methods fail. The reason for the existing methods producing poor results is that corner detection is sensitive to background noise in the case of (Dutta, Mandal et al. 2007) method. Moreover, the iterative nearest neighbor clustering based on convex hull deficiency proposed in (Raghunandan, Shivakumara et al. 2018) is only suitable for images containing text. Furthermore, the method developed in (Felemban, Sheikh et al. 2014) assuming that the most significant bit provides best information, does not work for the complex images. On the other hand, the proposed method involves saliency detection and the ring-growing feature for studying spatial distributions of saliency is shown to be effective. Detection rate and F-measure of the proposed and existing methods are reported in Table 4.1 where it is noted that the proposed method outperforms the existing methods. The values are counted manually as there is no ground truth for these images. Therefore, it can be concluded that the proposed method is effective and generic for informative bit plane detection irrespective of the content and image type, which is essential for networking applications.

(a) Input image

(b) (Dutta, Mandal et al. 2007)

(c) (Raghunandan, Shivakumara et al. 2018)

(d) (Felemban, Sheikh et al. 2014)

(e) Proposed method

Figure 4.19: Sample results of proposed and existing methods for bit plane detection

Table 4.1: Quantitative analysis of the proposed bit plane detection and existing methods in (%).

Method	Proposed method	(Dutta, Mandal et al. 2007)	(Raghunandan, Shivakumara et al. 2018)	(Felemban, Sheikh et al. 2014)
$\overline{DR}$	84.0	38.0	61.0	13.0
F	91.30	55.07	75.77	23.00

4.4.3.2   Validating Bit Plane Transmission Through TCP/IP Network

Since the bit plane image detected by the proposed bit plane detection method is in binary form, network experiments are conducted by sending detected plane, Canny edges of the input image and input image directly through TCP/IP network. In addition, the proposed method calculates error as defined in equation (4.11) for the size at starting and receiving ends. The same images also are sent with different formats, namely JPG and PNG to study the role of compression methods during transmission. Sampled results for the input and output of JPG and PNG formats are shown in Figure 4.20 (a) and Figure 4.20 (b), respectively, where we can see the size of the output image for color image is almost the same in case of JPG as shown in Figure 4.20 (a), while the size of the output increases compared to input for PNG as shown in Figure 4.20 (b). This shows that the JPG does not add extra bit for the output, while PNG does. This is mainly because of transmission error while sending large data images. However, the size of the output of Canny edge map and the bit plane increases compared to the size of the input for JPG, while it is almost the same for PNG format. This shows that JPG is good for compressing color images but not binary images, at the same time, PNG is good for binary images and not for color images. However, it is noted from Figure 4.20 (a) and Figure 4.20 (b) that the error as per the equation (4.11) is low for the bit plane compared to Canny edge map for JPG while it is almost same for PNG. Since PNG is good for binary images, it gives almost same error for Canny edge map and the bit plane but JPG is good for color images. There is a big difference in error for Canny edge map and the bit plane.

Input image	Canny edge map	Detected plane

| 28.748 | 35.102 | 35.102 |

Input images

| 28.753 | 70.936 | 63.407 |

Output images
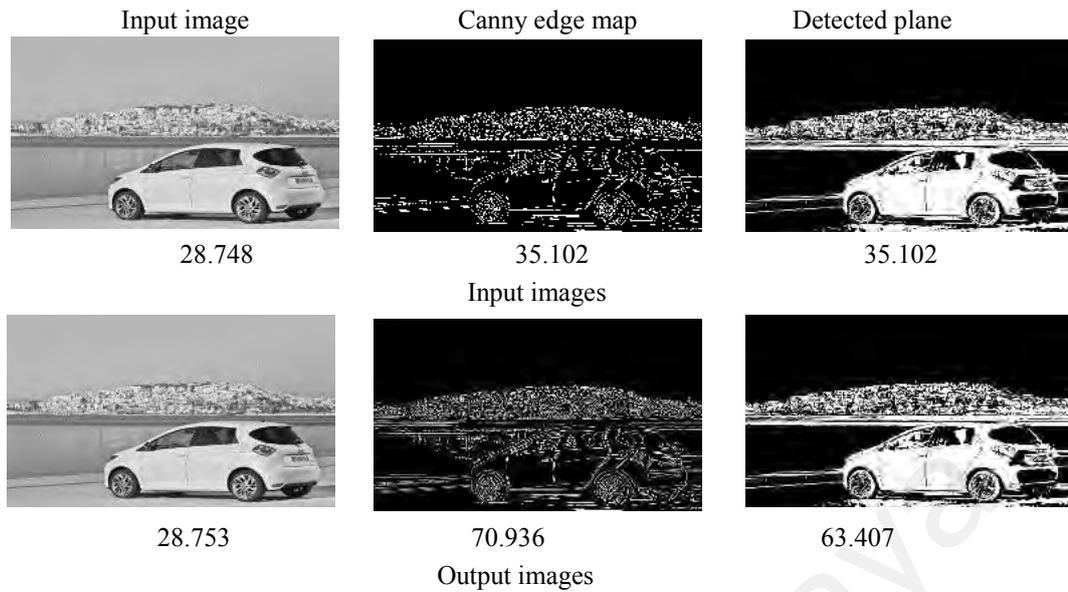
(a) Transmission error for JPG format data in in terms of byte

| 107.396 | 10.35 | 14.3 |

Input image

| 137.796 | 10.096 | 11.120 |

Output images
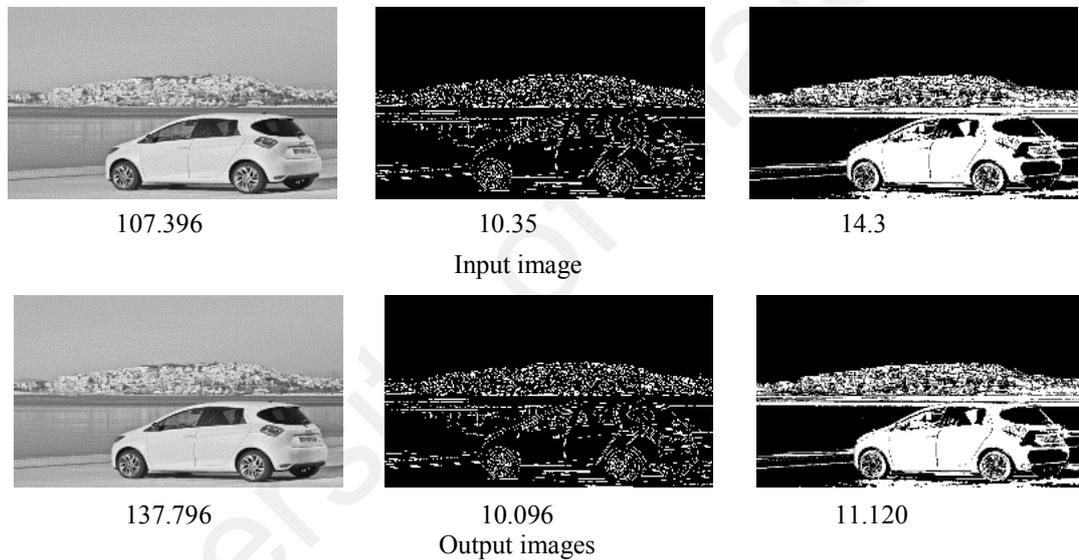
(b) Transmission error for PNG format in terms of byte

Figure 4.20: Error analysis during data transmission in TCP/IP network

With this analysis, it can be concluded that sending the bit plane through network is effective and enhances the performance of the system. In order to verify the above conclusion, experiments are also conducted for different procedures using JPG and PNG formats as shown in Figure 4.21. In this analysis, error is estimated by sending the input image, all eight bit planes, single plane detected by the proposed method, eight Canny edge maps of the plane images, Canny edge map of the input image, and Canny edge map of single plane detected by the proposed method. It is noted from Figure 4.21 (a) that error is low for single plane compared to all other cases but it is higher than the input

image. Similarly, when the results are observed in Figure 4.21 (b), error is low for the Canny edge maps compared to single plane, eight planes and input image. When we compare error analysis of JPG and PNG reported in Figure 4.21 (a) and Figure 4.21 (b), respectively, error estimation is inconsistent. It is observed from the quality analysis reported in Figure 4.21 (c) that the quality of the single plane is higher than all the Canny edge map combinations and eight planes but it is lower than input image. However, according to Figure 4.21 (b) error is high for eight planes and input image compared to single plane. Therefore, when we compare the quality and error analysis, single plane is better than all other combinations for transmitting through TCP/IP network because for single plane, error is low according to JPG, error is almost same compared to Canny edge maps according to PNG and quality is high according to the results reported in Figure 4.21 (c). Hence, detecting the informative bit plane for the input image irrespective of format or compression is useful and effective in improving network performance. The results of comparative study with the existing methods in terms of error analysis are reported in Figure 4.21 (d) where we can see the proposed method outperforms the existing methods. When we compare the results of existing methods, the method (Felemban, Sheikh et al. 2014) which transmits set of the bit planes, reports low error compared to the method (Junaid and Shunmuganathan 2017), which transmit the full image. This result is expected because when we send a large color image, there are high chances of getting high error compared to small size of the content. Thus, it can be asserted that the single plane detected by the proposed method is superior to existing methods in terms of error analysis.
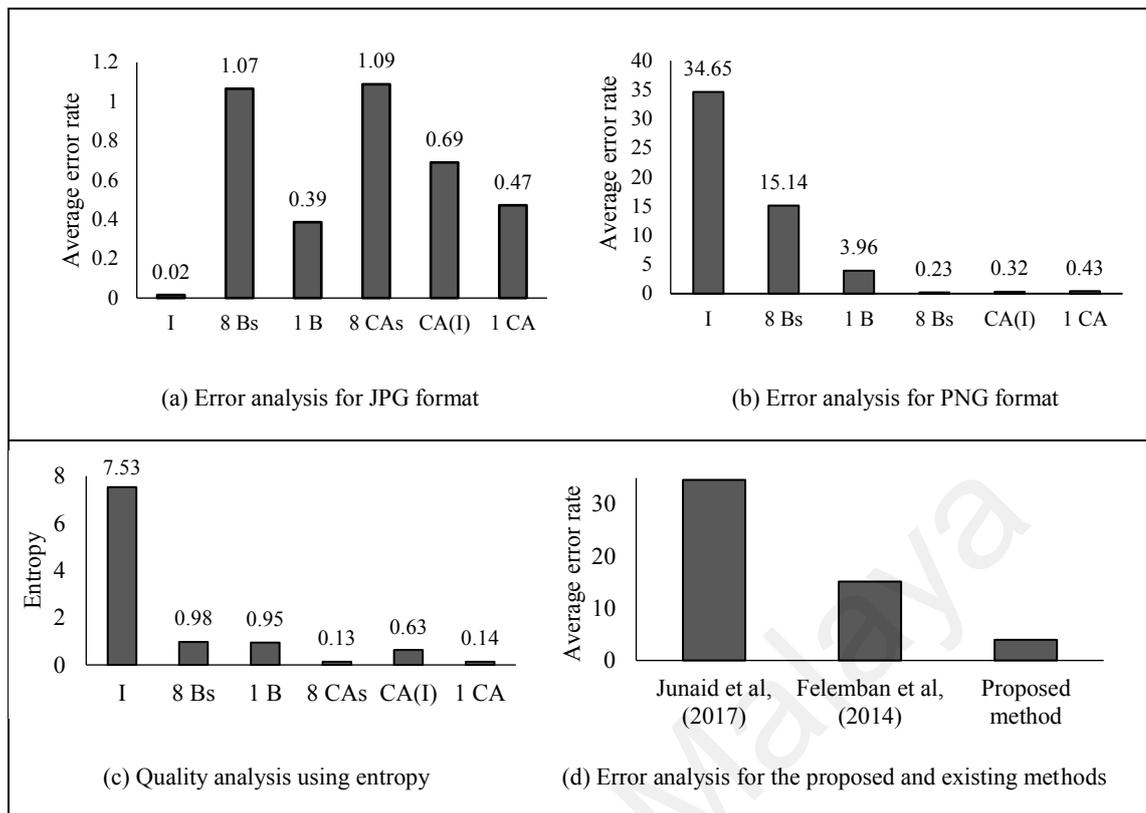
Figure 4.21: Entropy and error analysis of the proposed and existing methods
I denote Input color image, 8Bs denote eight planes, 1B dente single plane (the proposed method), 8CA denote eight Canny edge images of planes, CA(I) denotes Canny of input image and 1CA denotes Canny of single detected plane

4.5 Summary

A new image size reduction model for enhancing the performance of a VSN networking system is proposed in this section. The proposed model segment overlapping and non-overlapping regions based on responses and locations. In order to remove inter redundancies, the proposed model introduces a Dominant Overlapping Region (DOR) by analyzing the sharpness of the overlapping region, which confirms the useful information in the overlapping region. The proposed method uses the camera capacities to partition the DOR into sub-DORs to enhance the VSN performance further. Further, the proposed method performs a Median operation over the sub-DOR to remove any unwanted information, which results in a Compressed Significant Region (CSR) as the final segmented Region of Interest (ROI). However, as we cannot always expect cooperative

cameras, the idea is extended and a new method for detecting the informative bit plane from eight bit planes of the input image is proposed to improve system in terms of image size, error and quality analysis. This proposed method explores segmentation of saliency region of interest for finding dominant information in the planes. The ring-growing method is proposed to study the spatial distribution of saliency in all the eight planes. The plane which hit maximum times (mean values corresponding to rings) by reaching peak compared to other planes is considered as the significant bit plane.

The experimental results on plane detection with the existing methods show that the proposed method is better than the existing methods in terms of $PSNR, SSIM$ and detection rate. Further, experiments on a networking system show that the proposed model is also better than the existing methods in terms of iterations, energy consumption and network lifetime, error and quality analysis.

CHAPTER 5: AN AUTOMATIC ZONE DETECTION FOR SAFE UAV LANDING

5.1 Background

In the previous chapters, the role of segmentation in image size reduction is discussed. Since the objective of the proposed work is to address the challenges of flying drones, the segmentation is extended to detect particular region which provides a flat surface for a safe landing. This chapter presents the method for segmenting a safe landing zone from the segmented region of interests for safe landing.

As technology advances, we can see significant improvements in UAV landing. As a result, UAVs require flat areas irrespective of color, size even to some extent to poor weather conditions. With this notion, the fact that landing zones should be flat or homogenous/plain areas for safe landing is explored. For example, it could be a terrace, roof, ground, grass region, road, field, etc. This observation leads to propose a new method for safe land zone detection based directional features instead of magnitude values of the pixels. To extract such observations, motivated by the method proposed in (Saxena and Chaurasiya 2018) where the Gabor response of different orientations is used for finger print identification, Gabor filters of different orientations are explored for safe zone detection, which reflect flat regions. It is the fact that Gabor filters give low degrees of response for homogeneous areas and a high degree of variances for non-flat areas. For input images captured by UAVs, the proposed system obtains Gabor orientation responses of eight directions. The number of orientations (eight) was determined based on experimental analysis, and it is valid because those directions generally provide prominent information about objects in images. It is expected that flat regions provide uniform Gabor responses for their respective directions, and the proposed system

performs histogram operations over Gabor response images. Then, it finds the responses that contribute to the highest peak in the histogram as candidate pixels of the respective eight response images. As noted, Markov Chain Codes (MCCs) estimates the probability of using previous states to predict new states (Mulatu 2017). Inspired by this property, the probability of using eight neighbors from each pixel in the candidate pixel images of the respective eight responses are estimated, which predicts whether the center pixels are light or dark. This process results in clusters of pixels that share the same property of the center pixels, which are called Candidate Regions (CR).

One can expect each response may contain several CRs due to variations in the Gabor orientations and the nature of the images. Therefore, the proposed system finds CRs, which give the highest areas compared to their corresponding CRs of all the response images, which are considered as reference CRs. It is noted that CRs in the response images have different areas. As a result, it is hard to ensure that the reference CR represents expected homogenous safe landing zones. To deal with this issue, for each CR in all the responses, the proposed system performs histogram operations over probability values given by the MCCs for the pixels in the CR as discussed above. It then estimates degrees of similarity between the histogram distributions through Chi square distance measures. Motivated by the method of (Ghosh, Subudhi et al. 2012) where Chi square has been used for estimating degrees of similarity between two histograms of the regions for object detection, the same is used in this work. The proposed system finds the most similar CR to the reference CR in each response image. This process eliminates non-CR regions as well as providing correct CR regions. With this advantage, the proposed system fuses the reference CR with the most similar region out of other regions in the response images. This results in the expected homogenous regions for safe landing of UAVs. The workflow of the proposed method is shown in Figure. 5.1.
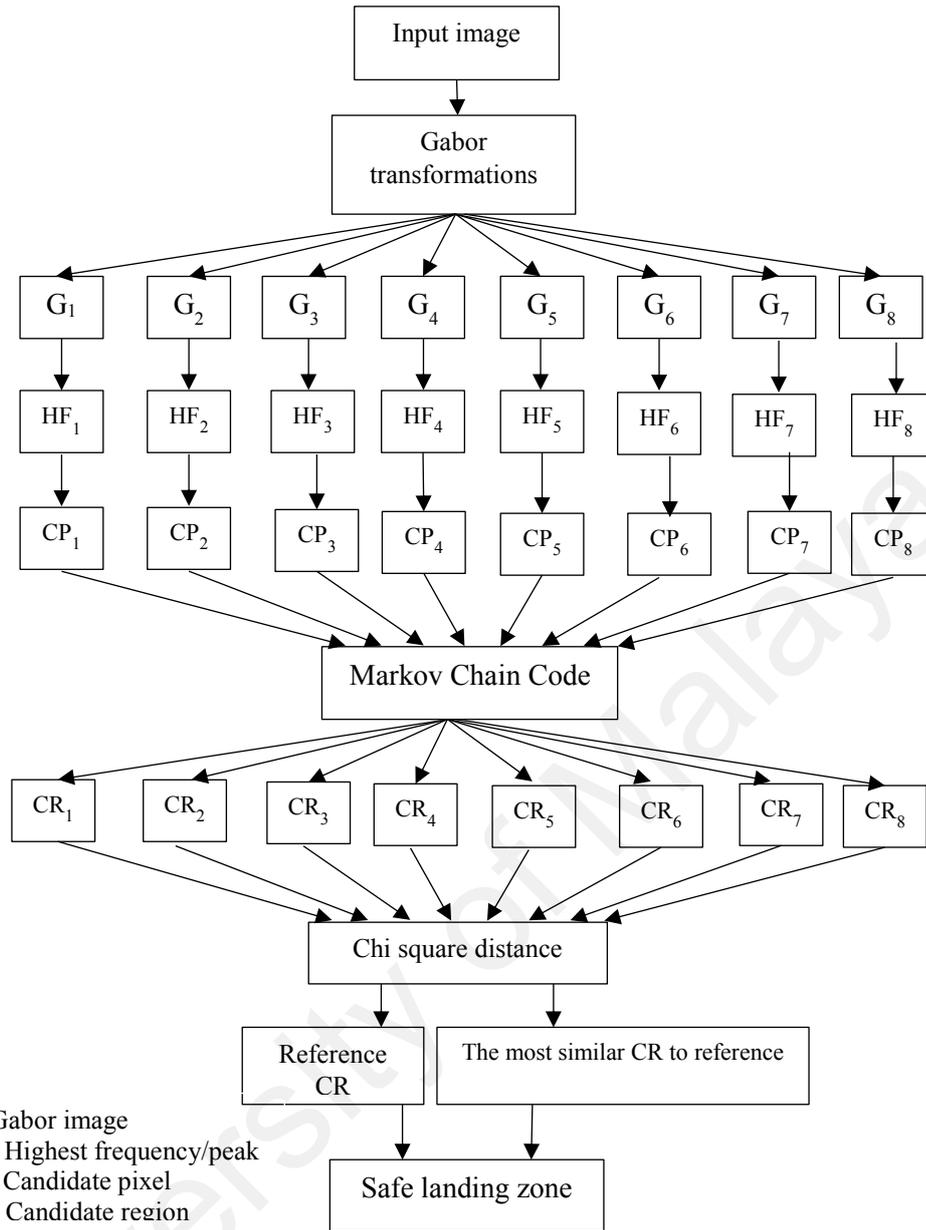
Figure 5.1: Block diagram of the proposed method

5.2 Gabor Transform for Candidate Pixel Detection

For each input image, the proposed system obtains eight Gabor orientations as defined in equation (5.1). Where, l is the frequency of the wave propagating in the direction of α from x-axis, while δ_x and δ_y define the Gaussian envelope along the respective axes. More details of the steps and implementation can be found in (Saxena and Chaurasiya 2018).

$$G(x, y; l, \alpha) = \exp\left\{\frac{-1}{2}\left[\frac{x'}{\delta_x^2} + \frac{y'}{\delta_y^2}\right]\right\}\cos(2\pi l x'), \tag{5.1}$$

$$x' = xsin\alpha + ycos\,\alpha, \qquad y' = xcos\,\alpha - ysin\,\alpha$$

The results of this computation are shown in Figure 5.2, where we can see clearly that Gabor responses give almost zero for the regions, which are homogenous and plain. This is valid because of the fact that the objects in the image contain prominent edge information that spread in 8 directions (Saxena and Chaurasiya 2018). This observation leads to determine eight equal angles from the range of [0-180] degree, resulting in 8 Gabor responses (alpha). It ensures that the considered homogenous region is the candidate region for finding safe landing zone, where the Gabor orientations give zero response in all prominent eight directions. The number of the orientations are determined experimentally. It is evident from Figure. 5.2 that the Gabor orientations provide almost uniform values for the homogeneous regions.

$G_1, \alpha=0$	$G_2, \alpha=22.5$	$G_3, \alpha=45$
$G_4, \alpha=67.5$	$G_5, \alpha=90$	$G_6, \alpha=112.5$
$G_7, \alpha=135$	$G_8, \alpha=157.5$	
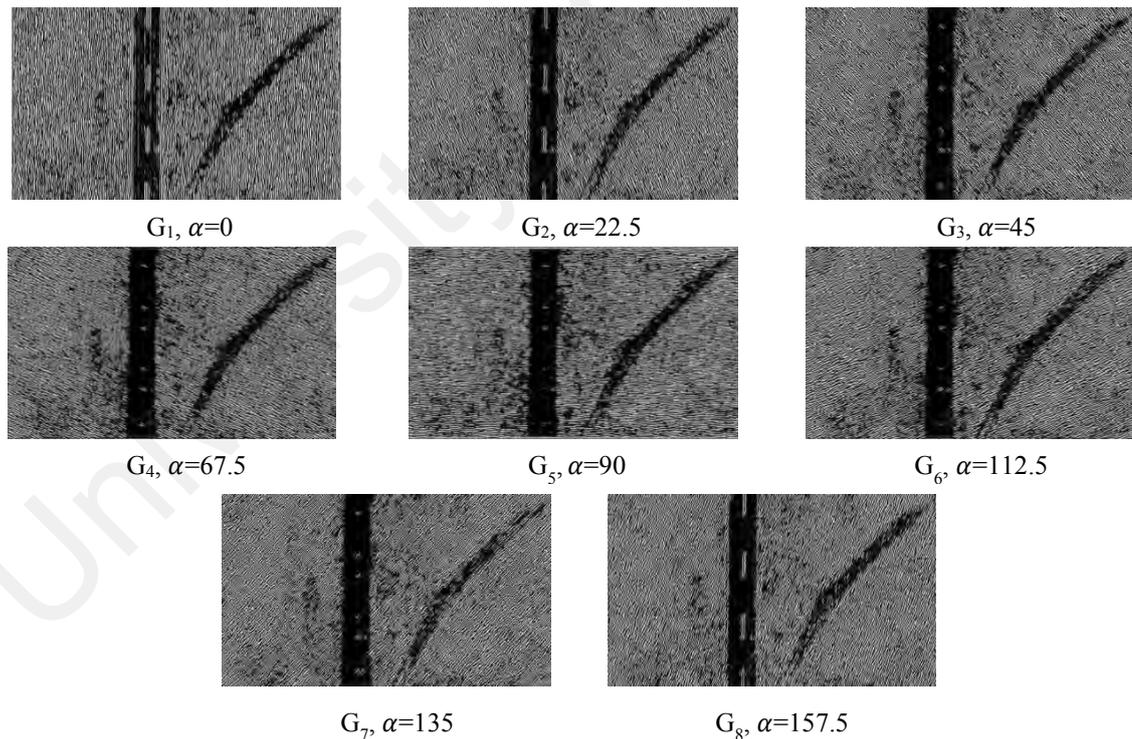
Figure 5.2: The response of different Gabor orientations for the first image in Figure 1. 11 (a)

The proposed system generates a histogram for the response of each pixel in images as shown in Figure 5.3. To extract such regions, the proposed system selects the pixels

that contribute to the highest peak in the histograms as Candidate Pixels (CPs) for the response images as shown in Figure 5.4, where the white pixels represent CPs. Before performing the histogram operation, the proposed system normalizes the values to the range of zero and one as defined in equation (6.2). Where $P_v max$ and $P_v min$ are the maximum and minimum values in a Gabor image.

$$N(G) = \frac{G}{P_v max - P_v min} \tag{5.2}$$
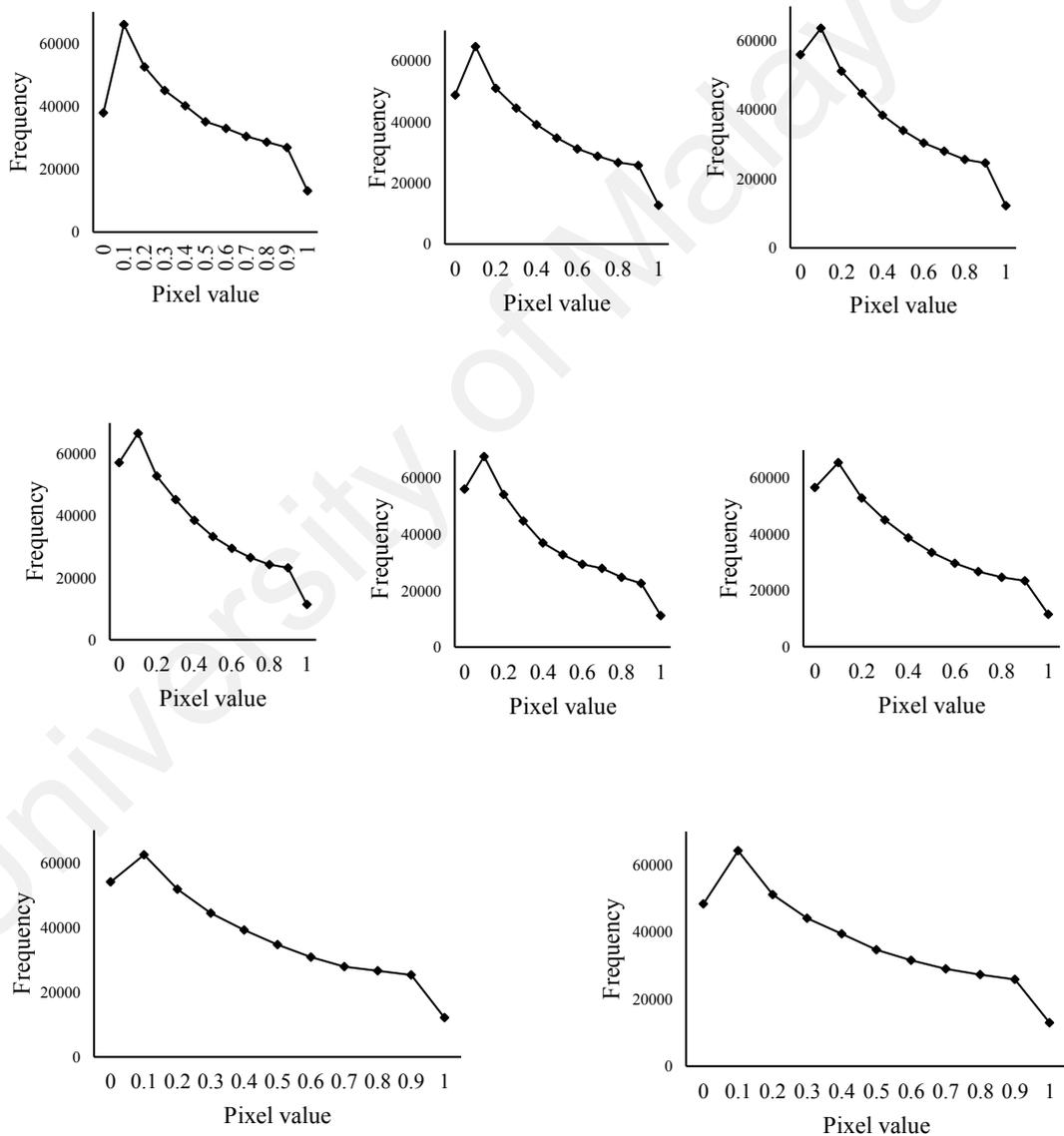
Figure 5.3: Histogram for pixel values vs their frequencies to find candidate pixel from respective Gabor orientations

<div align="center">CP$_1$, α=0   CP$_2$, α=22.5   CP$_3$, α=45</div>

<div align="center">CP$_4$, α=67.5   CP$_5$, α=90   CP$_6$, α=112.5</div>

<div align="center">CP$_7$, α=135   CP$_8$, α=157.5</div>
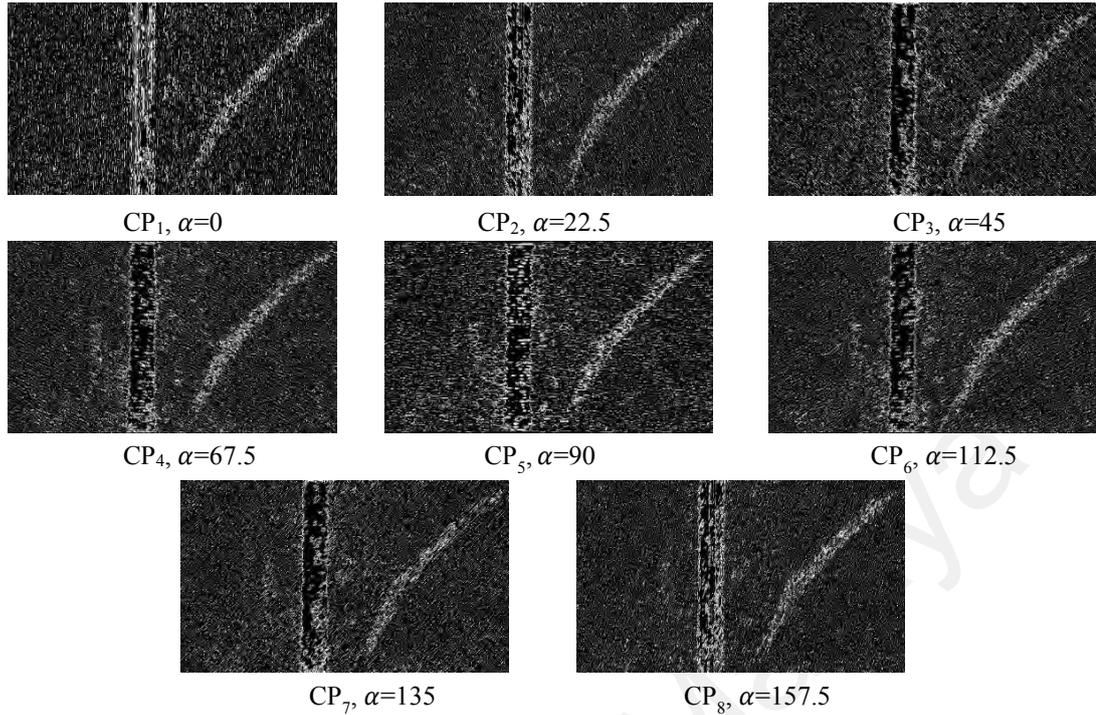
Figure 5.4: Candidate pixel (white colored pixels) detection for the respective Gabor orientations in Figure 6.3

It is observed from Figure 5.4 that the histogram operation detects other background pixels as CPs due to complex background and contrast variations. Therefore, in order to remove false CPs, the Markov Chain Code (MCC) is proposed, which will be discussed in the subsequent section.

5.3 Markov Chain Code Process for Candidate Region Detection

As mentioned, a Markov Chain Code (MCC) process helps us to group the pixels that share similar values. Therefore, for each pixel and its eight neighbors, MCC is employed as defined in equation (5.3) to (5.6) for the respective response images given in the previous section. The MCC works based on the fact that the states of the pixels should be defined based on the high probabilities of neighboring pixels. The process of MCC is presented in the following. Let, A denote the system state at time t as $A_t = [P_t, L_t]$ where $t = [1,2]$ represents the assigned time for the current and improved pixel, and $P_t \in [0,1]$ represents the pixel value at time t; $L_t \in [0,1,2,..,z^2]$ equals the summation of the

pixel and its neighboring pixels in an window size $S^w = z$ (since eight neighbors of each pixel are considered, z is 3). Each pixel i of the response image for time $t = 1$ is checked (see Figure. 5.4) to find $A_1 = [P_1, L_1]$, and then according to equation (5.3), the next state of pixel $A_2 = [P_2, L_2]$ is determined (see Figure 6.5). A probability is assigned to each pixel i as defined in equation (5.3) to (5.6). Since in MCC the summation of arrows exiting from a state is 1, $Pr(L_1) + Pr(L_2) = 1$.

$$A_2 = \begin{cases} 0, L_1 & \text{if } P_1 = 0 \text{ and } Pr(P_1) \geq Pr(L_2) \\ 1, L_1 + 1 & \text{if } P_1 = 0 \text{ and } Pr(P_1) \leq Pr(L_2) \\ 1, L_1 & \text{if } P_1 = 1 \text{ and } Pr(P_1) \geq Pr(L_2) \\ 0, L_1 - 1 & \text{if } P_1 = 1 \text{ and } Pr(P_1) \leq Pr(L_2) \end{cases} \tag{5.3}$$

When $P_1 = 0$, as the number of '0's increases and therefore L_1 decreases, the probability that in the next state A_2, $P_2 = 0$ increases. On the other hand, when $P_1 = 1$, as the number of '1's and therefore L_1 increase, the probability that in the next state A_2, $P_2 = 1$ increases.

$$Pr(L_1) = \begin{cases} 1 - \beta \times L_1 & \text{if } P_1 = 0 \\ 1 - \beta \times ((z \times z) - L_1) & \text{if } P_1 = 1 \end{cases} \tag{5.4}$$

$$Pr(L_2) = \begin{cases} \beta \times L_1 & \text{if } P_1 = 0 \\ \beta \times ((z \times z) - L_1) & \text{if } P_1 = 1 \end{cases} \tag{5.5}$$

$$\beta = 1/(z \times z) \tag{5.6}$$

Since it is a two-class problem, i.e. either a pixel is to be zero or one, two states are considered for grouping the pixels as the process shows in Figure 5.5 (a) and Figure 5.5 (b), respectively, where we can see step by step the process of MCC for classifying pixels as zero or one based on probability.

(a) Markov Chain Code for $P_1 = 0$

(b) Markov Chain Code for $P_1 = 1$
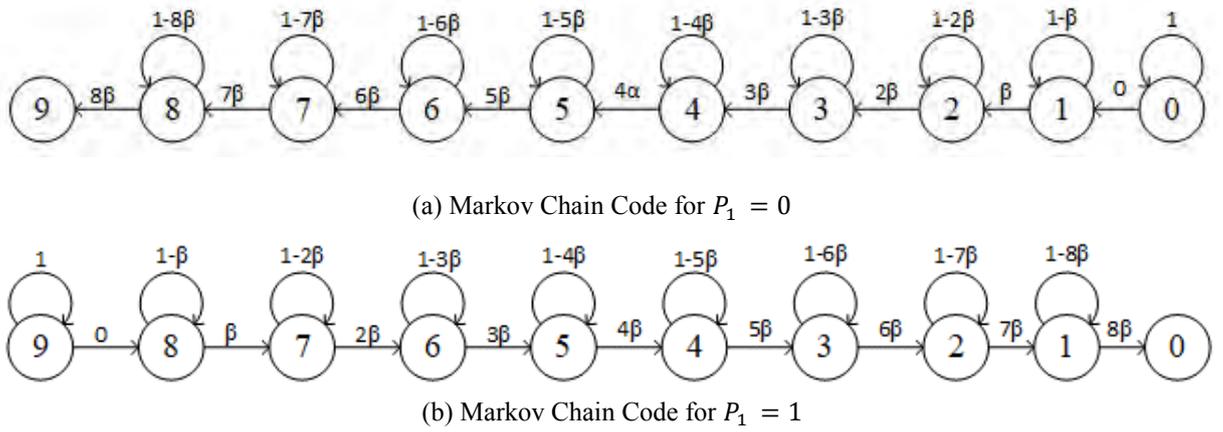
Figure 5.5: Markov Chain Code for grouping pixels to detect candidate regions

The effect of MCC for the respective response can be seen in Figure 5.6, where it is clear that MCC helps us to cluster the pixels that share the same values based on the probability of neighbors, which results in Candidate Regions (CRs).

CR_1, $\alpha=0$

CR_2, $\alpha=22.5$

CR_3, $\alpha=45$

CR_4, $\alpha=67.5$

CR_5, $\alpha=90$

CR_6, $\alpha=112.5$

CR_7, $\alpha=135$

CR_8, $\alpha=157.5$
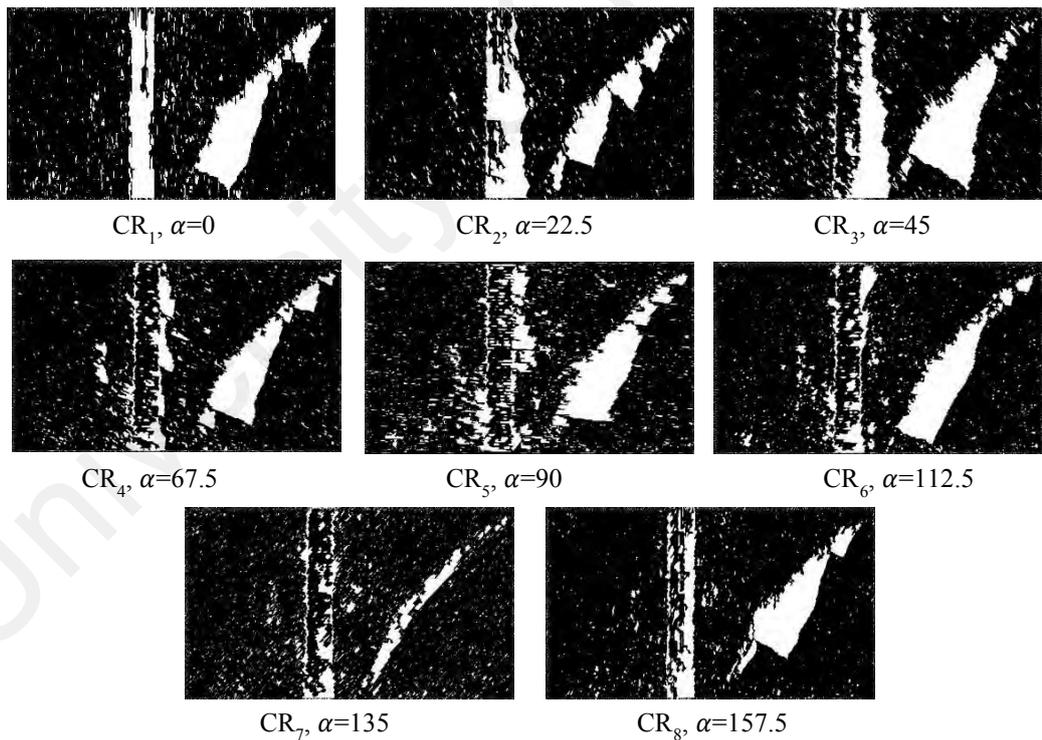
Figure 5.6: Candidate regions (white colored pixels) detection using a Markov Chain Code for respective Gabor orientations

It is noted from Figure 5.6 that there are unwanted CRs due to the complexity of the problem. To remove such noisy components from the results of the MCC, it is proposed to find the largest CR by comparing areas of the CR for the respective response images, which is considered as the reference CR for each response image. Since there are different sized reference CRs for the response image, an automatic criteria is proposed to calculate the threshold to classify unwanted small regions, which is independent of the CR of different response images. As a result, the area of the reference CR is compared with other CRs of the same response image based on a fix small threshold as defined in equation (5.7), where ρ denotes the threshold and its value is 0.03. The value of the threshold is determined empirically, which is shown in the Experimental Section. The steps to determine the threshold are as follows. Let each response image i, $i=[1,2,..,8]$, have j regions $j=[1,\ldots, N^{cg}]$ called CR_i^j. They are labeled and counted the number of their pixels $N^p_{CR_i^j}$ to find the biggest region of each image called CR_i^{max}. The next step is removing small regions to compare with the largest region. Equation (5.7) removes CR_i^j when $Y(CR_i^j) = 0$ and keeps it when it equals 1.

$$Y(CR_i^j) = \begin{cases} 0 & if\ N^p_{CR_i^j} \leq \rho \times N^p_{CR_i^{max}} \\ 1 & if\ N^p_{CR_i^j} \geq \rho \times N^p_{CR_i^{max}} \end{cases} \tag{5.7}$$

The effect of equation (5.7) is shown in Figure. 6.7, where we can see it removes almost all the unwanted CRs. To fill the small gaps between CRs, the proposed system performs the morphological operation over the result of equation (5.7) as shown in Figure 6.8, where it can be seen that the gaps are filled resulting in completely connected CRs. It is observed from Figure 6.8 that one response gives a good CR and another response loses information. This is evident because of the effect of different Gabor orientations for the unexpected nature of the image. To overcome this issue and to get a complete and

final CR, the degree of similarity to fuse the corresponding CRs in the response images are estimated.

CR$_1$, α=0   CR$_2$, α=22.5   CR$_3$, α=45

CR$_4$, α=67.5   CR$_5$, α=90   CR$_6$, α=112.5

CR$_7$, α=135   CR$_8$, α=157.5
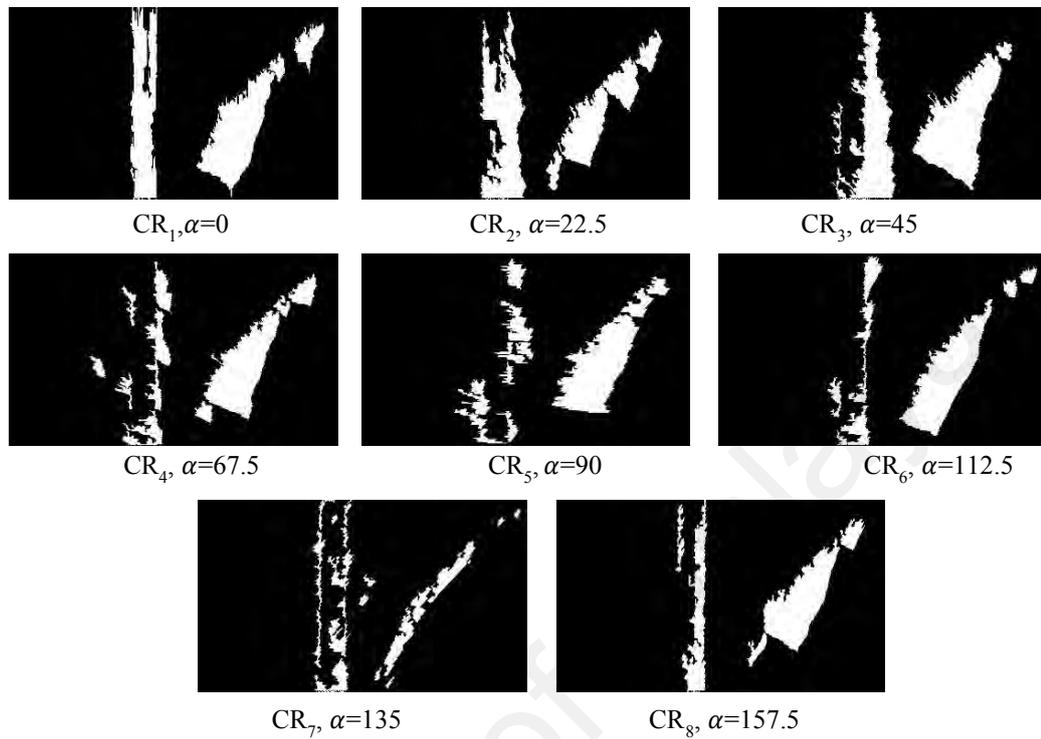
Figure 5.7: Candidate regions detection (white colored pixels) after removing small regions for the respective Gabor orientations.

CR$_1$, α=0   CR$_2$, α=22.5   CR$_3$, α=45

CR$_4$, α=67.5   CR$_5$, α =90   CR$_6$, α=112.5

CR$_7$, α=135   CR$_8$, α=157.5
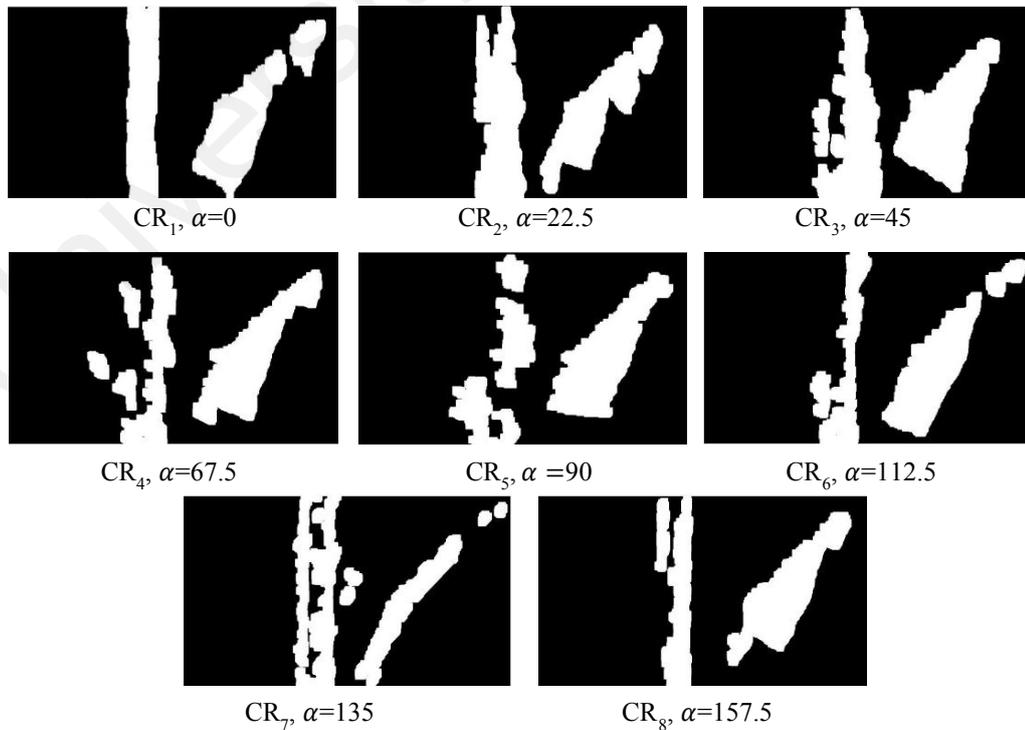
Figure 5.8: Dilating candidate regions (white colored regions) to fill small gaps between the components

5.4   Chi Square Distance Measure for Safe Landing Zone Detection

As discussed in the previous section, Chi square are explored to estimate the degree of similarity between the reference CR and corresponding CRs of the respective response images. For each CR, including the reference CR of the respective response images as shown in Figure 5.9 (a), which is a reference CR for the input image, the proposed system extracts the probability given by MCC for all the pixels of the CR. Then it generates a histogram for the probability values of the CR, which gives a distribution of pixels for each CR. Motivated by the method (Ghosh, Subudhi et al. 2012), the degree of similarity is estimated between the reference CR and the corresponding CRs of respective responses by considering shapes of the histograms (distribution) as defined in equation (5.8). Out of many CRs in the respective response images, the proposed system chooses the one which is the most similar to the reference CR as shown in Figure 5.9 (b). The proposed system performs a union operation for fusing the reference CR with most similar CR as shown in Figure 5.9 (c), where we can see the complete CR region. In the same way, the process continues for other CRs corresponding with the respective response images. The final result for the input image is shown in Figure 5.9 (d), where we can see two CRs which can be used for safe landing of UAVs for the input image. The formal steps for the above process are presented as follows.

Let the jth candidate region of the ith response image be considered as a reference region, called $CR_{i=r}^j$, for example, CR_3^1, as shown in Figure 5.9 (a). For all CR_i^1 i=[1,2,..,8], a histogram operation is performed, $H_{CR_i^j}$ with N^{bin} bins for the probability values given by MCC. In order to study the shape of the histogram distribution, the proposed system finds the absolute difference between frequency values of the successive bins in the histogram, which outputs a new histogram H' as defined in equation (5.8), where

bin=[1,2,3,..,10]. In this work, a histogram is considered with 10 bins as it gives promising results.

$$\left(H'_{CR_i^j}\right)_{N^{bin}} = \left|\left(H_{CR_i^j}\right)_{N^{bin}} - \left(H_{CR_i^j}\right)_{N^{bin}+1}\right| \qquad (5.8)$$

Equation (5.9) is the X²-test that is a statistical method of determining the similarity between two distributions of regions. This equation calculates the difference between the histograms of the reference candidate regions $H'_{CR_r^j}$ with other candidate regions in the respective Gabor response images $H'_{CR_{i\ i\neq r}^j}$. According to this, the calculation CR_1^1 is determined as the most similar region to the reference as shown in Figure 5.9 (b).

$$X^2\left(H'_{CR_r^j}, H'_{CR_{i\ i\neq r}^j}\right) = \sum_{N^{bin}} \frac{\left(\left(H'_{CR_r^j}\right)_{N^{bin}} - \left(H'_{CR_i^j}\right)_{N^{bin}}\right)^2}{\left(\left(H'_{CR_r^j}\right)_{N^{bin}} + \left(H'_{CR_i^j}\right)_{N^{bin}}\right)} \qquad (5.9)$$

The same process continues for the second candidate region of reference CR_r^2 and the process terminates when it has visited all the CRs in the image.

(a) Reference CR   (b) The most similar CR among 8 CRs

(c) Fusing (a) and (b)   (d) Safe land
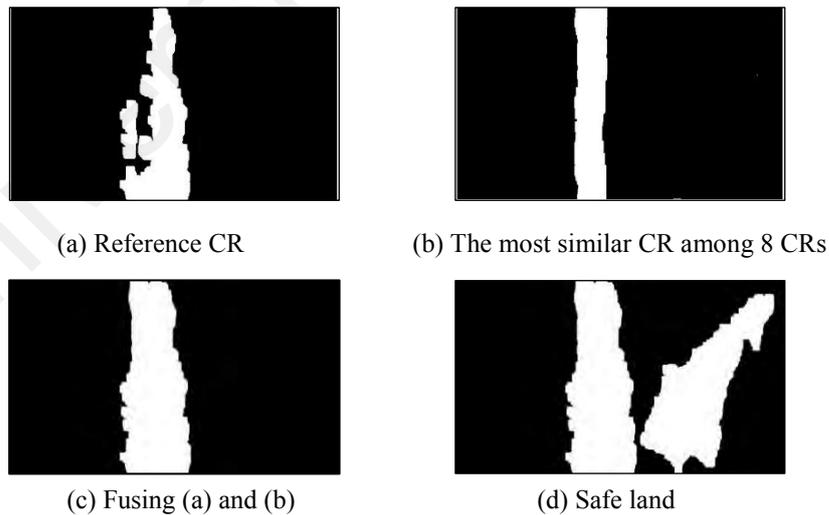
**Figure 5.9: Fusion process for finding a safe land region
(white colored region) from eight Gabor oriented images**

The proposed system consists of two parts, namely, for the given input image, the method detects candidate regions and then chooses relevant regions for safe landing of a UAV according to emergency scenarios. In this chapter, three common and sensitive dangerous conditions are considered, which may cause emergency landings, namely, communication, GPS and software-hardware/energy errors (Patterson, McClean et al. 2014). The details for each cause are as follows.

Case-1: A communication link enables UAVs to be controlled via sending and receiving essential commands from a human-in-the-loop. The loss of this wireless link would make the UAV out of control and would result in the occurrence of undesirable critical events. In this situation, in addition to general safety needs, the most important criterion is to choose a safe landing zone, which is near to a Transmitter Station (BS) to increase the chance of reconnection.

Case-2: A vital part of a UAV sensing system is the GPS. It provides the UAV with essential information. Although, studies mostly assume that it is always available and accurate, there are still limitations in real situations, such as the unavailability of satellite services, errors and uncertainties related to atmospheric layers, poor space weather conditions, physical obstacles and blocks in outdoor or indoor environments and very high latitude. In this case, the most important landing issue is being far from tall obstacles such as mountains, buildings or threats that may cause collisions.

Case-3: Sometimes, one can encounter unexpected software and hardware errors, for example, actuator failure or software module crashes that force the UAV to land as soon as possible to the nearest available region. In addition, UAVs may face very low power availability because of uncertainties in real environments, which would prevent it from returning to its base station and forcing a landing at the nearest opportunity. In this

situation, the main concern would be for navigation towards the nearest detected (safe) zone.

The proposed safe land detection method involves two key steps that are Gabor filter and Markov chain phase. Gabor filter requires $O(n)$ time complexity as the operation performs in a linear fashion. Markov chain phase requires two loops because of the method process the whole image. Therefore, the time complexity of this step is $O(n2)$. The overall time complexity of the proposed method is $O(n2)$ and hence it is $\Theta(n2)$ time complexity since the worst and best case require the same complexity.

5.5   Experimental Results

This section is divided into four sub-sections, namely, description of dataset and evaluation criteria for measuring the performance of the proposed and existing methods in Section 5.5.1. The empirical analysis for determining the value of the key parameters used in the proposed method is presented in Section 5.5.2. Section 5.5.3 presents evaluation of safe landing zone detection step and Section 5.5.4 provides validation for the method of safe landing zone detection.

5.5.1   Dataset and Evaluation

There is no standard dataset, which can be used for evaluating the proposed system. Therefore, a dataset is created by which includes images captured by a custom-build quadcopter and also Google images. In total, 50 images are collected for experimentation and evaluation in this work. To capture the images, a low cost camera is used with a resolution of 600×450 pixels. Image size are 300×500 which are taken in different lightning, weather conditions and daytime. The dataset consists of images of different situations, such as roads, trees, lawn, grass, forest and buildings. In addition, the dataset includes images captured by drones, which may fly in the range of 30 to 2000 meters with

reference to ground station[4]. Note: in this study, small or medium UAV is considered for the experiments. For example, multi-rotor drone quadcopters with four rotors (DJI™Phantom). The created dataset comprises the possible situations for emergency landing of a UAV.

In order to show the superiority to existing systems, state-of-the-art methods are implemented, namely, (Patterson, McClean et al. 2014), which explores edge-based features for potential zone detection. It then proposes a fuzzy-based method for classification of potential zones into categories of risky, safe and unsafe. In addition, (Aziz, Faheem et al. 2016)is considered which explores edge information and histogram operations to classify the given images into good and bad classes. The reason to choose these two methods for a comparative study with the proposed system is that the objective of those existing systems is the same as the proposed system, which segments regions and then shows the usefulness of segmented regions for the safe landing of UAVs. Note: since , (Patterson, McClean et al. 2014) method outputs three classes, namely, risky, unsafe and safe, for each given input image, the images that were classified are considered as safe for calculating measures in this work. Similarly, since, (Aziz, Faheem et al. 2016) considers the whole image as the input for the classification of good and bad for UAV landing, the regions segmented are passed by the proposed system as the inputs to the existing system to calculate relative success and unsuccessful rates. In other words, for each image, the proposed system obtains Candidate Regions (CR) and non-Candidate Regions (non-CR). For these two regions, if the existing system classifies candidate regions as good and non-candidate regions as bad, it will be considered as a successful segmentation, otherwise it is an unsuccessful segmentation. To evaluate the proposed and

---

[4] The data set will be shared upon request for supporting reproducibility.

existing systems for the detection of safe zones for UAV landing, well-known measures are used, namely, Recall, Precision and F-measure as defined in equation (3.18) till equation (3.20).

As noted in the proposed System Section, one can expect several safe lands for one image, which can be used for landing a UAV safely. As a result, for a large number of input images, the proposed system outputs a large number of safe landing regions. This leads to select the possible relevant safe landing sites according to the above three cases. Hence, in this chapter, the cost is estimated by assigning weights for different factors, such as distance from UAV and distance from terrain threat situations based on pre-defined experiments as defined in equation (5.13) and equation (5.14).

The overall cost function assigned for each safe region is $CF(CR_j)$. The objective is to find the jth region that gives the minimum cost according to equation (6.13), where τ_1, τ_2, τ_3 and τ_4 are constant and the sum of those values should satisfy $\tau_1 + \tau_2 + \tau_3 + \tau_4 = 1$. Let $d_{CR_j}^{uav}$, $d_{CR_j}^{BS}$ and $d_{CR_j}^{O}$ represent the Euclidean distance between the UAV and the jth safe region, transmitter station and tall obstacle, respectively. d_{max} is the farthest distance in the image to the UAV in order to normalize values. Finally, TH_{CR_j} is a parameter, which represents the changes according to the influence of terrain threat on the jth region. It is a summation of N^{TH} threat sources calculated as defined in equation (5.14) (Huang, Qu et al. 2016), where TH^i represents the threat level of the ith threat, k is a coefficient, and $d_{CR_j}^{TH}$ is the Euclidean distance between that threat and the region. It is divided by a maximum threat influence where threat level TH_{max} is the maximum threat level and the distance of the region and a threat is at a minimum called d_{min}.

$$CF(CR_j) = \tau_1 \times \frac{d_{CR_j}^{uav}}{d_{max}^{uav}} + \tau_2 \times \frac{d_{CR_j}^{BS}}{d_{max}^{TS}} + \tau_3 \times \frac{1}{d_{CR_j}^{O} \big/ d_{max}^{O}} + \tau_4 \times \frac{TH_{CR_j}}{\sum_{i=1}^{N^{TH}} TH_{max}^i \times e^{-kd_{min}^i}} \quad (5.10)$$

$$TH_{CR_j} = \sum_{i=1}^{N^{TH}} TH^i \times e^{-kd_{CR_j}^i} \qquad (5.11)$$

The above equations indicate that when the system gets different emergencies, higher values are assigned to the corresponding constant. Finally, j^{th} region which minimizes the cost is considered as a safe landing zone out of many candidate regions given by the segmentation step. Since there is no ground truth for evaluating the proposed system, the outputs of the proposed system are counted manually to calculate relative successful and unsuccessful rates as defined in equation (5.15), where $\overline{DR}$ is the number of successful safe regions chosen according to various situations, and N^{emr} is the number of all emergency cases.

$$\overline{DR} = \frac{N^{cd}}{N^{emr}} \qquad (5.12)$$

5.5.2 Empirical Analysis for Determining the Values of the Key parameters

The proposed system consists of key parameters, namely, different orientations for Gabor responses, the number of neighbors for the Markov Chain Code process to group the pixels, and the value of the threshold for removing noise candidate regions automatically by denoting areas of reference candidate regions. To determine the values for the above three parameters, predefined samples are chosen randomly from the database for experimentation by calculating the F-measure (F) by varying parameter values as shown in Figure 5.10 until Figure 5.13, respectively. It is noted from Figure 5.10 that as the number of the angle changes on the X axis, the F-measure on the Y axis for Candidate Region segmentation changes. However, for the number of angle (8), the proposed segmentation score is the highest and hence 8 is considered as the best number in this work. It is valid because for a smaller number of angles, there is a chance of losing salient information and for a higher number of angles, there is the potential of introducing redundancy.
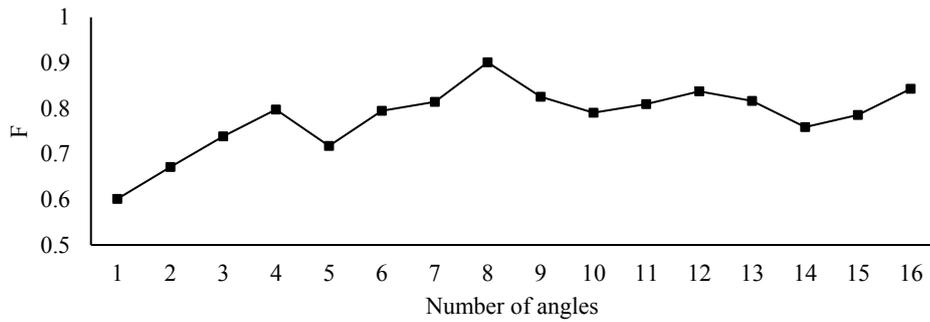
Figure 5.10: F-measure for different Gabor orientations to determine the best Gabor responses

Similarly, experiments have been conducted for the same predefined data for determining the best number of neighbors to group the pixels using the Markov Chain Code as shown in Figure 5.11, where we can see that as the window size increases, the performance of segmentation decreases. However, for the neighbors of 8, the proposed segmentation scores the highest F-score and thus 8 neighbors ($S^w=3$) is considered to be the best configuration. It is valid because when the size of the neighbors is increased, there is a chance of losing pertinent information. In addition, it will be a computationally expensive process.
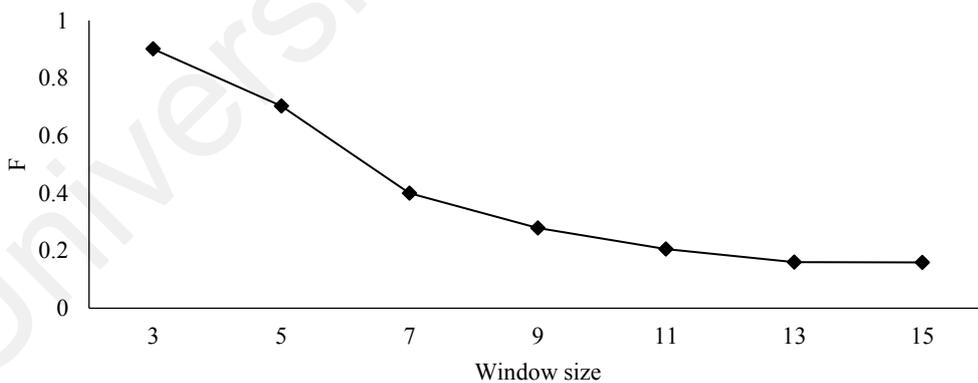
Figure 5.11: F-measure for different window sizes to determine the best candidate regions

In the same way, the value of the threshold is determined for removing unwanted candidate regions, which is independent of the candidate regions in the respective response images, as noted in the experiments shown in Figure 5.12, where it can be seen

that as the threshold value increases the performance of segmentation decreases. At the 0.03 value, the performance is the highest and therefore, 0.03 is considered as the optimum for achieving better results. This is because when we increase the threshold value, there is a chance of losing significant candidate regions in different response images. If it is less than 0.03, there is the possibility of increasing false candidate regions.
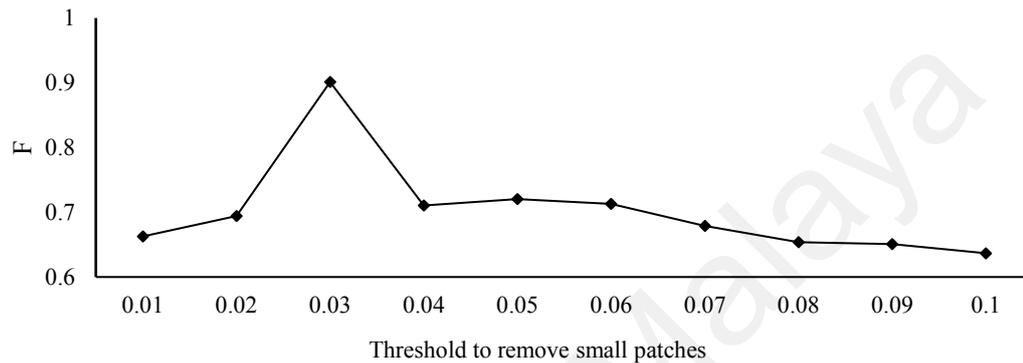
Figure 5.12: F-measure for different threshold values to determine the optimum value for removing small patches

5.5.3 Experiments for Evaluating Safe Landing Zone Detection

Qualitative results of the proposed and existing systems for safe landing zone detection of UAVs are shown in Figure 5.13, where we can see input images of different situatons in Figure. 5.13 (a), the result of the proposed system, (Patterson, McClean et al. 2014) and (Aziz, Faheem et al. 2016), respetively shown in Figure. 5. 13 (b) and Figure 5. 13 (c). It is observed from the results in Figure 5.13 that the proposed system detects safe landing regions marked with white color successfully for all the input images, while the existing systems fail to detect these accurately. For the last image, the proposed system does not detect the road as a safe landing zone due to vehicles which appear on it. Therefore, it can be concluded that the proposoed system detects safe landing regions well for complex situations. The reason that the existing systems report poor results is that they are not robust and consider limited situations. It is evident that the existing systems define specific features to segment regions, such as edge-based and geometrical-based features, while the proposoed sytem considers features which represent

124

homogenous and flat surfaces by exploring Gabor orientations. The same thing is true for the quantitative results reported in Table 5.1, where the proposoed system achieves the best at Recall, Precision and F-measure compared to both existing systems. This is due to generalised features of the proposed system without many constraints and assumptions. The reason for the poor results of the existing systems is the same as mentioned above. When we compare the results among existing systems, (Aziz, Faheem et al. 2016) gives better results in terms of Recall and F-measure. This is because (Aziz, Faheem et al. 2016) involves histogram operations to find homogenous regions, which is robust compared to Canny edges used by (Patterson, McClean et al. 2014) as Canny edges are sensitive to complex backgrounds.

(a) Input images of different situations.

(b) Proposed system

(c) (Patterson, McClean et al. 2014)
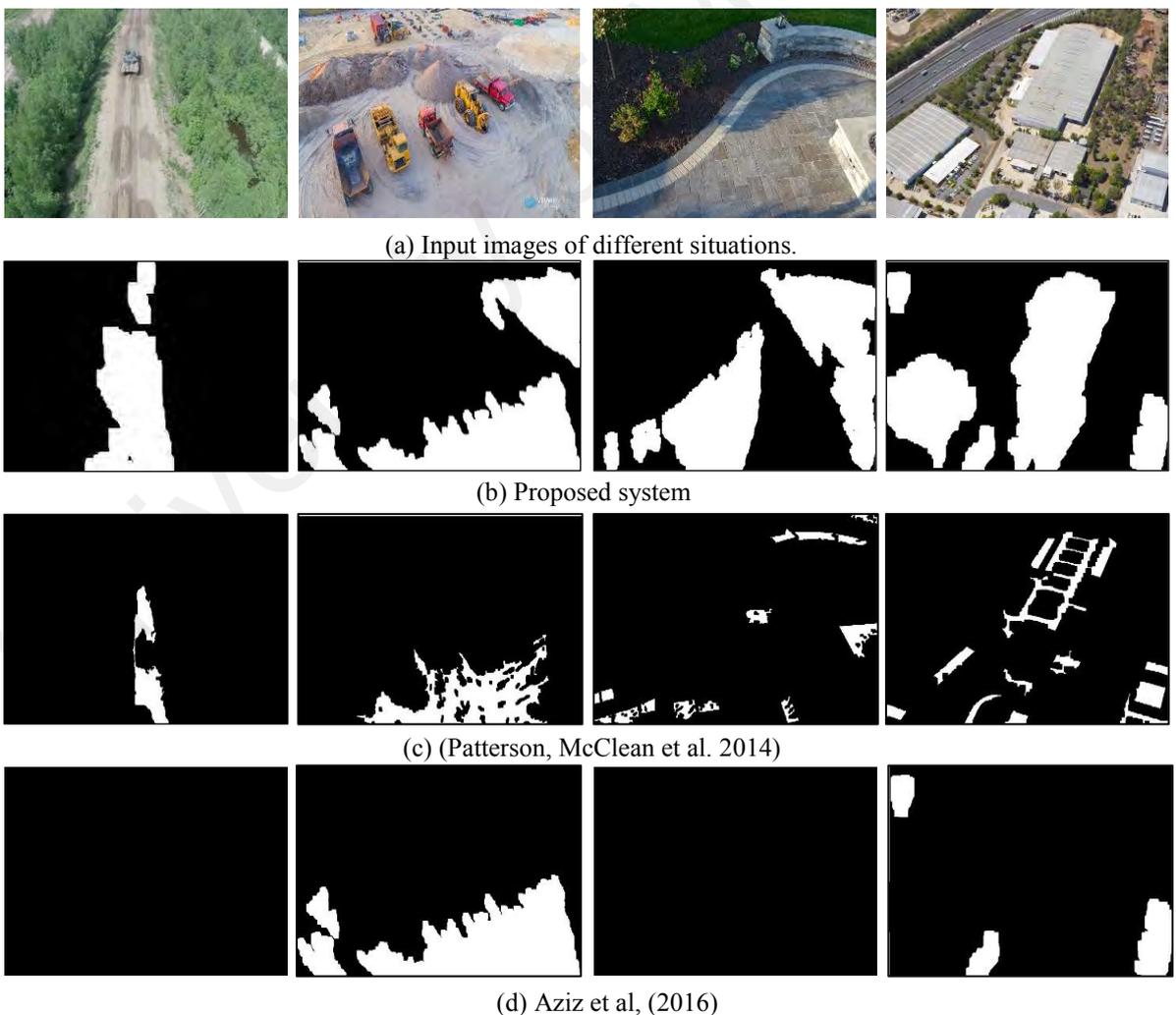
(d) Aziz et al, (2016)

Figure 5.13: Qualitative results for safe land detection (white colored regions) of the proposed and existing systems

Table 5.1. Performance of the proposed and existing systems for safe landing zone detection

	Patterson et al, (2014)	Aziz et al, (2016)	Proposed System
Recall	0.46	0.61	**0.87**
Precision	0.93	0.90	**0.93**
F-Measure	0.62	0.72	**0.90**

5.5.4 Experiments for Validating Safe Landing Zone Detection

The proposed segmentation step gives several regions to enable UAVs to land safely. According to the emergency situations it is mentioned earlier, the proposed system has to choose the correct region for landing from many safe landing spots. The experimental setup includes 500 random scenarios where artificial threats and segments (CRs) are created, such that the created scenario reflects a real networked environment (Huang, Qu et al. 2016). The proposed system calculates relative success rates as defined in equation (5.15) based on cost estimation defined by equation (5.13). It is true that the number of segments and the number of threats are key parameters, which affect success rates of the proposed system. Therefore, a threat level of 0.1 initially is set for conducting experiments to determine the number of segments as shown in Figure 5. 14, whereas when the number of segments increases, the success rate on the Y-axis also increases until 20 segments, after which the relative success rate reaches a steady state. This shows that after 20 segments, there is not much improvement in the relative success rate. It is valid because usually in the case of a real environment, the number of segments represents safe lands given the proposed step increase, thus the system gets more chances for selecting the best one, and hence the relative success rate either increases or it would tend toward a steady state. Therefore, the number of segments are considered as 20 for evaluating the proposed system.

Figure 5.14: Choosing an optimal number of safe lands for assessing safe landing performance

Similarly, to determine the relative threat level, the number of segment candidate regions are fixed as 20 and then experiments are conducted for different threat levels as shown in Figure 5. 15, where we can see that as the relative threat level increases, the performance of the system decreases. Note threat indicates obstacle, which prevent UAV for landing at the zone in spite of it is detected by the proposed system. For instance, when power cable or mountain or towers or tall buildings is near to detected safe zone, UAV cannot choose it for landing. This is because it may lead to collision. Therefore, when the number of such relative threats increases, the proposed system misses safe zone for landing and hence there is a poor relative success rate for a higher number of threats. Therefore, in this work, 0.1 threat level and 20 as the number of segments (safe lands) are considered for calculating the final relative success rates of the proposed and existing systems.
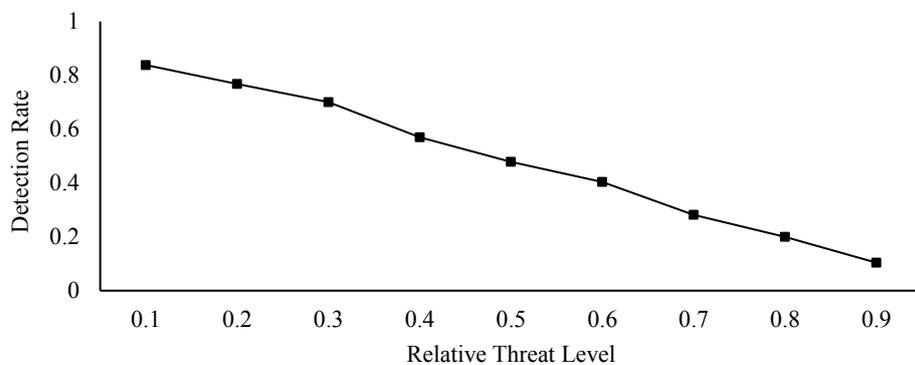
Figure 5.15: Safe landing performance as relative threat level increases after fixing the number of safe landing regions

The simulation results of the proposed and existing systems on 500 random scenarios by fixing the number of segments (safe regions) as 20 and threat level as 0.1 are presented in Figure 5.16. The proposed system has the highest relative success rate compared to the existing systems. In this experiment, the cost is estimated as defined in equation (6.13) for calculating relative success rates. Since the proposed system segments regions as the successful candidates for landing a UAV safely, there is no need for classification as in the existing systems (Patterson, McClean et al. 2014, Aziz, Faheem et al. 2016). Therefore, the cost is estimated according to emergency, namely, communication error (case-1), GPS error (case-2) and software-hardware error/energy constraints (case-3) for calculating relative success rates separately as shown in Figure 5.16 for the proposed and existing systems. The key reason of the existing systems having poor relative success rates compared to the proposed system is that the existing systems are not capable of classifying complex situations as they do not propose generalized features for segmentation. In addition, the success of classification depends heavily on the number of samples and classifier training. On the other hand, the proposed system considers generalized features for candidate region detection without classification, which contributes to its success. Note: when we compared the segmentation and classification results of the existing systems, for segmentation, (Aziz, Faheem et al. 2016) scores the best compared to (Patterson, McClean et al. 2014), but it is vice versa for the classification results. (Aziz, Faheem et al. 2016) does not have robust classification results as compared to (Patterson, McClean et al. 2014), which considers fuzzy-based features for classification. The fuzzy-based features are good when there is ambiguity in classification, due to the large number of CRs.
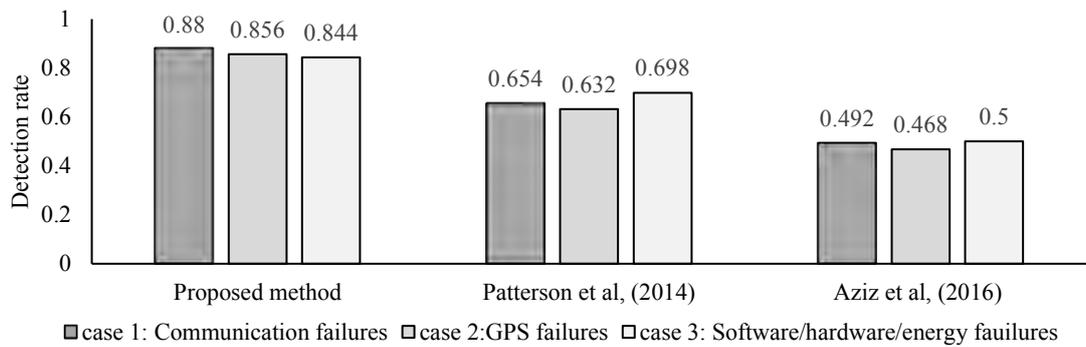
Figure 5.16: Detection rate of the proposed and existing methods for safe landing

5.6 Summary

In this chapter, a new system has proposed for candidate region detection, which can be used for safe UAV landing when there is an emergency. The proposed system explores different Gabor orientations based on the fact that the degree of Gabor orientation responses is high for variations and low for smooth/flat regions in images. The same observation has been explored for finding homogeneous regions in this chapter, which are called flat regions for landing UAVs safely. The proposed system performs histogram operations for Gabor response images to classify pixels which contribute to homogenous regions. This results in candidate pixel detection. The proposed system introduces a Markov Chain Code for grouping neighboring pixels that share the same values based on probabilities, which result in candidate regions. Due to complex backgrounds and contrast variations, there is a potential of detecting false candidate regions. To overcome this issue, the proposed system finds candidate regions using areas, which are the largest among other candidate regions corresponding to respective Gabor response images. The system compares the areas of other candidate regions to eliminate noise and small regions, which results in reference candidate regions. Further, the proposed system uses Chi-square to estimate the degree of similarity between the reference candidate region and other candidate regions in the respective Gabor response images. It finds the most similar candidate region among all the other corresponding regions, and then it performs a union

operation for fusing the reference candidate region and the most similar one, which denotes flat safe regions for UAV landing.

The experimental results on finding regions for safe UAV landing with existing systems in the literature show that the proposed system outperforms the existing systems in terms of recall, precision and F-measure. In addition, experiments are also simulated to choose the best safe landing areas out of many safe landing zones given by the segmentation steps according to emergency-situations by creating 500 random scenarios artificially, which includes a number of relative threats and more complex situations. Experimental results on classification through cost estimation show that the proposed segmentation approach is useful and works well for complex situations.

CHAPTER 6: CONCLUSION

6.1   Contributions of the Proposed work

This thesis investigates efficient classification and segmentation methods for surveillance applications. The contribution of this thesis are listed below and address the problem discussed in Sections 1.4 and achieve the objectives in 1.6.

Firstly, the power of both edge strength and sharpness features are integrated in a novel way so that the focused edges pattern can be extracted for classification purpose. Edge strength provides high gradient magnitude values for the focused edge pixels while edge sharpness provides precise gradient magnitude with respect to thickness of the edges. In addition, the combined features are robust to complex backgrounds of scene images, noise, because the proposed features do not depend on the objects' shapes in the images, rather they extract unique edge patterns of different images of classes. This contribution is presented in Chapter-3 and it is achieved as per the first objective.

Secondly, the dominant overlapping regions is detected so that it contains useful information by estimating sharpness of the pixels in the images. Then, the dominant overlapping regions are partitioned, according to capacity of camera nodes to utilize the resources efficiently. Finally, a new operation for reducing image size further without losing quality is proposed. Another contribution of this research work is in the use of saliency and ring-growing concepts to extract features so as to detect plane, which provide significant information. In addition, a general network for validating the bit plane detection is used in terms of error estimation and quality analysis. This contribution is described in Chapter-4 according to the second objective.

Finally, the third contribution is safe region segmentation that it does not require region specific features, and it does not need classification as in existing methods. An intelligent system is proposed to segment a safe landing region for UAV. It works based on the fact

that Gabor orientation responses are low for flat or homogeneous regions, whilst they are high for the other regions where some obstacles exist. This achievement is discussed in Chapter-5 and it is stated in the third objective.

6.2   Limitation of the Proposed Work and Future Work

The works presented in this thesis are able to classify and segment scene images in different surveillance applications. However, it is observed from the experimental results that there some limitations in aforementioned proposed approaches.

For example, for classification work, it is noted from experiments as shows in **Error! Reference source not found.** there are some misclassification of the images by the proposed method, where the air pollution image is misclassified as a flood and vice versa, and the garbage image is misclassified as a plant. Interestingly, one can notice from the misclassified images that when an image contains multiple regions, which represent different classes, the proposed method misclassifies. The solution is to segment such multiple regions before classification. This is part of future target to reduce complexity of ubiquitous surveillance applications further.

In case of image size reduction, as discussed in the experimental section, there is scope for improving the performance of the proposed method, especially improving performance of the whole system when the number of cameras increases and distance between the camera and ground increases. The proposed and existing methods show a sudden drop at a 70 metre distance. This shows that the resources used in the proposed and existing methods are not enough to handle the situations beyond a 70 metre distance. Therefore, developing a model which is independent of these parameters is interesting and challenging for future research. In addition, to improve the plane detection results further, the proposed idea will be extended by extracting new features with deep learning.

Finally, though the proposed safe landing zone segmentation works well for complex situations, sometimes, it detects regions as candidate ones for images containing water as shown in Figure. 7.2. This is because the surface of the water appears as homogenous, and homogeneity is the key feature for the proposed system. In addition, sometimes, if an image contains vehicles and is captured from a long distance (height), the proposed system does not work correctly. Therefore, there is a plan to introduce new features to integrate with the Gabor responses to improve the results under these scenarios in the near future. There are high chances of capturing blurred images due to adverse climate effect and variation in heights. In this case, the performance of the proposed system degrades. One possible way to overcome this situation is that we can choose the option of going back home if necessary. Otherwise, there is a need to investigate new features, which can be robust to low contrast and poor quality images. These are all in the scope of improvements of the proposed work in the future.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Published Journals:

[1] Kaljahi, M. A., Shivakumara, P., Anisi, M. H., Idris, M. Y. I., Blumenstein, M., & Khan, M. K. (2018). "A scene image classification technique for a ubiquitous visual surveillance system." Multimedia Tools and Applications, 1-28.

[2] Kaljahi, M. A., Shivakumara, P., Idris, M. Y. I., Anisi, M. H., Lu, T., Blumenstein, M., & Noor, N. M. (2019). "An automatic zone detection system for safe landing of UAVs." Expert Systems with Applications, 122, 319-333.

[3] Kaljahi, M. A., Shivakumara, P., Anisi, M. H., Idris, M. Y. I., A New Image size reduction model to reduce transmission power for visual sensor networks. Visual Communication and Image Representation (VCIR), 63, 102573.

Under Review:

[4] Kaljahi, M. A., Shivakumara, P., Hakak. S, Idris. M. Y. I, Anisi. M. H and Rajan. D., "Saliency based bit plane detection for TCP/IP network", Multimedia Tools and Applications, submitted 14 Dec 2018.

Under preparation:

[5] Kaljahi. M.A, Parr. G, Anisi. M. H, Shivakumara. P, Idna Idris. M. Y, Luo. Ch,. "A survey on the safety management of autonomous unmanned aerial vehicles."

REFERENCE

(GOOGLE API). Google API System: https://cloud.google.com/vision/.

Afsharinejad, A., et al. (2016). "Performance analysis of plant monitoring nanosensor networks at THz frequencies." IEEE Internet of Things Journal **3**(1): 59-69.

Ahmad, J., et al. (2017). "Efficient object-based surveillance image search using spatial pooling of convolutional features." Journal of Visual Communication and Image Representation **45**: 62-76.

Al-Ariki, H. D. E. and M. S. Swamy (2017). "A survey and analysis of multipath routing protocols in wireless multimedia sensor networks." Wireless Networks **23**(6): 1823-1835.

Al-Kaff, A., et al. (2018). "Survey of computer vision algorithms and applications for unmanned aerial vehicles." Expert Systems With Applications **92**: 447-463.

Attaullah, H., et al. (2015)." Extracting and Encoding Mutually Sensed Region in Wireless Multimedia Sensor Networks to Remove Redundancy." Intelligent Systems, Modelling and Simulation (ISMS), 6th International Conference on, IEEE: 231-236.

Aurangzeb, K., et al. (2018). "Data Reduction using Change Coding for Remote Applications of Wireless Visual Sensor Networks." IEEE Access **6**: 37738- 37747.

Awad, F. H., et al. (2015). "Correlation model for randomly deployed heterogeneous cameras in wireless multimedia sensor networks." EUROCON International Conference on Computer as a Tool (EUROCON), IEEE.

Aydin, M. and E. Kugu (2016). "Finding smoothness area on the topographic maps for the unmanned aerial vehicle's landing site estimation." Digital Information and Communication Technology and its Applications (DICTAP), 6th International Conference on, IEEE: 183-186.

Aziz, S., et al. (2016). "Unmanned aerial vehicle emergency landing site identification system using machine vision." Journal of Image and Graphics **4**(1): 36-41.

Bai, S. (2017). "Growing random forest on deep convolutional neural networks for scene categorization." Expert Systems With Applications **71**: 279-287.

Balla-Arabé, S., et al. (2014). "Multi-kernel implicit curve evolution for selected texture region segmentation in VHR satellite images." IEEE Transactions on Geoscience and Remote Sensing **52**(8): 5183-5192.

Basit, A. and M. Javed (2007). "Iris localization via intensity gradient and recognition through bit planes." Machine Vision, International Conference on, IEEE: 23-28.

Bhandary, V., et al. (2016). "Routing in wireless multimedia sensor networks: a survey of existing protocols and open research issues." Journal of Engineering **2016**.

Bosch, A., et al. (2008). "Scene classification using a hybrid generative/discriminative approach." IEEE transactions on pattern analysis and machine intelligence **30**(4): 712-727.

Briscoe, N. (2000). "Understanding the OSI 7-layer model." PC Network Advisor **120**(2).

Carney, E., et al. (2019). "Determination of Safe Landing Zones for an Autonomous UAS using Elevation and Population Density Data." AIAA Scitech 2019 Forum.

Carrio, A., et al. (2017). "A review of deep learning methods and applications for unmanned aerial vehicles." Journal of Sensors **2017**.

Çelebi, A. T., et al. (2018). "Selective Gray-Coded Bit-Plane-Based Two-Bit Transform and Its Efficient Hardware Architecture for Low-Complexity Motion Estimation." IEEE Transactions on Consumer Electronics **64**(3): 259-266.

Chandra, S., et al. (2018). "Deep spatio-temporal random fields for efficient video segmentation." Computer Vision and Pattern Recognition, Conference on: IEEE: 8915-8924.

Chen, J., et al. (2017). "Attention region detection based on closure prior in layered bit Planes." Neurocomputing **251**: 16-25.

Chia, W. C., et al. (2009). "Multiview image compression for wireless multimedia sensor network using image stitching and SPIHT coding with EZW tree structure. " Intelligent Human-Machine Systems and Cybernetics. IHMSC'09. International Conference on, IEEE: 298-301.

Chieh-Ling Huang, J.-J. C., Ching-Ju Chen,Yung-Gi Wu (2016). "Geological segmentation on UAV aerial image using shape-based LSM with dominant color " Advanced Information Networking and Applications Workshops, International Conference on, IEEE: 928-933.

Chow, K.-Y., et al. (2006). "Balancing image quality and energy consumption in visual sensor networks." Wireless Pervasive Computing, 2006 1st International Symposium on, IEEE: 1-5.

Chow, K.-Y., et al. (2006). "Efficient on-demand image transmission in visual sensor networks." EURASIP Journal on Advances in Signal Processing **2007**(1): 095076.

Chow, K.-Y., et al. (2007). "Efficient selective image transmission in visual sensor networks." Vehicular Technology (VTC2007)- Spring. 65th Conference, IEEE:1-5.

Cohen, G., et al. (2018). "Spatial and Temporal Downsampling in Event-Based Visual Classification." IEEE Transactions on Neural Networks and Learning Systems(99): 1-15.

Coşar, S. and M. Çetin (2014). "Feature compression: A framework for multi-view multi-person tracking in visual sensor networks." Journal of Visual Communication and Image Representation **25**(5): 864-873.

Coşar, S. and M. Çetin (2015). "Sparsity-driven bandwidth-efficient decentralized tracking in visual sensor networks." Computer Vision and Image Understanding **139**: 40-58.

Costa, D. G., et al. (2017). "Availability issues for relevant area coverage in wireless visual sensor networks." Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Conference on , IEEE: 1-6.

D'Addabbo, A., et al. (2016). "A Bayesian network for flood detection combining SAR imagery and ancillary data." IEEE Transactions on Geoscience and Remote Sensing **54**(6): 3612-3625.

Davis, T. W., et al. (2012). "Analysis of power characteristics for sap flow, soil moisture, and soil water potential sensors in wireless sensor networking systems." IEEE Sensors Journal **12**(6): 1933-1945.

Dornaika, F., Moujahid, A., Merabt, Y. E. & Ruichek, Y. (2016). "Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors." Expert System with Applications **58**: 130-142.

Du, L. and H. Ling (2016). "Dynamic scene classification using redundant spatial scenelets." IEEE transactions on cybernetics **46**(9): 2156-2165.

Dunlop, H. (2010). "Scene classification of images and video via semantic segmentation." Computer Vision and Pattern Recognition Workshops (CVPRW), Computer Society Conference on, IEEE: 72-79.

Dutta, A., et al. (2007). "Bit-plane extension to a class of intensity-based corner detection algorithms." 15th European Signal Processing Conference on, IEEE: 267-271.

Duvar, R. and O. URHAN (2018). "Adaptive bit-plane selection-based low complexity motion estimation for screen content coding." Turkish Journal of Electrical Engineering & Computer Sciences **26**(5): 2206-2217.

Ebrahim, M. and C. W. Chong (2014). "A comprehensive review of distributed coding algorithms for visual sensor network (VSN)." International Journal of Communication Networks and Information Security (IJCNIS) **6**(2).

El Bastawesy, M., et al. (2012). "Detection and assessment of the waterlogging in the dryland drainage basins using remote sensing and GIS techniques." IEEE journal of selected topics in applied earth observations and remote sensing **5**(5): 1564-1571.

Elzouki, D., et al. (2018). "Lattice-based Robust Distributed Source Coding." Transactions on Information Theory, IEEE: 1764-1781.

Etemad, E., et al. (2018). "Robust image watermarking scheme using bit-plane of hadamard coefficients." Multimedia Tools and Applications **77**(2): 2033-2055.

Fall, K. R. and W. R. Stevens (2011). TCP/IP illustrated, volume 1: The protocols, addison-Wesley.

Fan, C., et al. (2019). "SUR-Net: Predicting the Satisfied User Ratio Curve for Image Compression with Deep Learning." Quality of Multimedia Experience (QoMEX), 11th International Conference on , IEEE: 1-6.
.

Fan, Z., et al. (2018). "Automatic tobacco plant detection in UAV images via deep neural networks." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **11**(3): 876-887.

Felemban, E., et al. (2014). "Improving response time in time critical Visual Sensor Network applications." Ad Hoc Networks **23**: 65-79.

Francis-Lothai, F. and D. B. Bong (2014). "An analysis of the effects of bit plane extraction in fingerprint recognition. " Systems, Process and Control (ICSPC), Conference on , IEEE: 132-136.

Francis-Lothai, F. and D. B. Bong (2017). "A fingerprint matching algorithm using bit-plane extraction method with phase-only correlation." International Journal of Biometrics **9**(1): 44-66.

Garcia-Pulido, J., et al. (2017). "Recognition of a landing platform for unmanned aerial vehicles by using computer vision-based techniques." Expert Systems With Applications **76**: 152-165.

Garg, R., et al. (2018). "Monocular and Stereo Cues for Landing Zone Evaluation for Micro UAVs." arXiv preprint arXiv:1812.03539.

Ghosh, A., et al. (2012). "Object detection from videos captured by moving camera by fuzzy edge incorporated Markov random field and local histogram matching." IEEE Transactions on Circuits and Systems for Video Technology **22**(8): 1127-1135.

Gonzalez, R., et al. (2011). Digital Image Processing Using MATLAB, 4. repr, Tata McGraw Hill Education, New Delhi.

Google (API system). https://cloud.google.com/vision/.

Goswami, P. P., et al. (2016). "Detecting moving objects in traffic surveillance video." International Journal of Control Theory and Applications **9**(17): 8423-8430.

Goyal, B., et al. (2018). "A three stage integrated denoising approach for grey scale images." Journal of Ambient Intelligence and Humanized Computing: 1-16.

Guerrero, J. M., et al. (2013). "Automatic expert system based on images for accuracy crop row detection in maize fields." Expert Systems with Applications **40**(2): 656-664.

Güngör, V. Ç. and G. P. Hancke (2013). "Industrial wireless sensor networks: Applications, protocols, and standards," Crc Press.

Hampapur, A., et al. (2003). "Smart surveillance: applications, technologies and implications. " Information, Communications and Signal Processing, 4th Pacific Rim Multimedia International Conference on, IEEE: 1133-1138.

Harel, J., et al. (2007). "Graph-based visual saliency." Advances in neural information processing systems: 545-552.

Hastie, T. and R. Tibshirani (1998). "Classification by pairwise coupling." Advances in neural information processing systems, 10. In: MIT Press Cambridge, MA.

Hayat, M., et al. (2016). "A spatial layout and scale invariant feature representation for indoor scene classification." IEEE Transactions on Image Processing 25(10): 4829-4841.

Hinzmann, T., et al. (2018). "Free LSD: prior-free visual landing site detection for autonomous planes." IEEE Robotics and Automation Letters 3(3): 2545-2552.

Hu, B. and S. Mishra (2019). "Time-Optimal Trajectory Generation for Landing a Quadrotor Onto a Moving Platform. IEEE/ASME Transactions on Mechatronics, 24(2), 585-596.

Huang, C.-L., et al. (2016). "Geological segmentation on UAV aerial image using shape-based LSM with dominant color. "Advanced Information Networking and Applications Workshops (WAINA), 30th International Conference on, IEEE: 928-933.

Huang, L., et al. (2016). "A novel coordinated path planning method using k-degree smoothing for multi-UAVs." Applied Soft Computing 48: 182-192.

Imran, M., et al. (2012). "Architecture of wireless visual sensor node with region of interest coding." Networked Embedded Systems for Every Application (NESEA), 3rd International Conference on, IEEE: 1-6.

Imran, M., et al. (2013). "Architecture of Wireless Visual Sensor Node with Region of Interest Coding." International Journal of Design, Analysis & Tools for Integrated Circuits & Systems 4(1): 1-6.

Imran, M., et al. (2010). "Exploration of target architecture for a wireless camera based sensor node." NORCHIP 2010, IEEE" 1-4.

Imran, M., et al. (2010). "Exploration of Target Architecture for aWireless Camera Based Sensor Node. " NORCHIP; Conference proceedings on: IEEE: 1-4.

Jang, S.-W. and S.-H. Cha (2014). "An approach to segmenting initial object movement in visual sensor networks." International Journal of Distributed Sensor Networks **10**(4): 928583.

Jayashree, A., et al. (2012). "Energy efficient prioritized multipath QoS routing over WMSN." International Journal of Computer Applications **46**(17): 33-39.

Jayshree, A., et al. (2012). "Energy efficient prioritized multipath qos routing over wmsn." International Journal of Computer Applications (0975-8887 **46**(17).

Jiang, B., et al. (2019). "Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues." Artificial Intelligence in Agriculture **1**: 1-8.

Junaid, K. M. and K. Shunmuganathan (2017). "A Peer to Peer Architecture for Optimized Energy Efficient Routing in Visual Sensor Networks." Wireless Personal Communications **94**(4): 2325-2330.

Karpathy, A., et al. (2014). "Large-scale video classification with convolutional neural networks." Computer Vision and Pattern Recognition, Proceedings conference on: IEEE: 1725-1732.

Kenchannavar, H. H. and U. P. Kulkarni (2011). "Optimized overlapping based energy-aware node collaboration in visual sensor networks." Wireless Technologies for Humanitarian Relief, ACM, Proceedings of the 1st International Conference on: 81-87.

Khedo, K. K., et al. (2010). "A wireless sensor network air pollution monitoring system." wireless and mobile networks: **2**(2) 31-45.

Khursheed, K., et al. (2012). "Detecting and coding region of interests in bi-level images for data reduction in Wireless Visual Sensor Network. " Wireless and Mobile Computing, Networking and Communications (WiMob), 8th International Conference on, IEEE: 705-712.

Kim, C. and J.-N. Hwang (2002). "Fast and automatic video object segmentation and tracking for content-based applications." IEEE Transactions on Circuits and Systems for Video Technology **12**(2): 122-129.

Kumar, M. S. and B. J. Priya (2018). "Complexity Reduction in Inter Layer Inter Prediction in Scalable High Efficiency Video Coding."Master thesis in University of Texas.

Lang, H., et al. (2014). "Scene classification by feature co-occurrence matrix." Computer Vision, Asian Conference on, Springer: 501-510.

Lange, S., et al. (2008). "Autonomous landing for a multirotor UAV using vision." Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR 2008), International Conference on: 482-491.

Lee, J., et al. (2017). "Monitoring and Mitigation of Ionospheric Anomalies for GNSS-Based Safety Critical Systems: A review of up-to-date signal processing techniques." IEEE Signal Processing Magazine **34**(5): 96-110.

Lee, T. and D. Bong (2013). "Palmprint recognition based on bit-plane extraction." technology **3**: 4.

Li, M., et al. (2018). "Learning convolutional networks for content-weighted image compression. " Computer Vision and Pattern Recognition, Proceedings Conference on, Proceedings Conference on, IEEE: 3214-3223.

Li, N., et al. (2016). "Object tracking based on bit-planes." Journal of Electronic Imaging **25**(1): 013032.

Li, R., et al. (2018). "An Energy-Efficient Compressive Image Coding for Green Internet of Things (IoT)." Sensors **18**(4): 1231.

Li, S., et al. (2019). "A Survey of Energy-Efficient Communication Protocols with QoS Guarantees in Wireless Multimedia Sensor Networks." Sensors **19**(1): 199.

Li, X. (2013). "A Software Scheme for UAV's Safe Landing Area Discovery." AASRI Conference on Intelligent Systems and Control **4**: 230-235

Lin, K., et al. (2011). "Energy efficiency QoS assurance routing in wireless multimedia sensor networks." IEEE Systems Journal **5**(4): 495-505.

Liu, G., et al. (2019). "Region-of-Interest Detection Based on Statistical Distinctiveness for Panchromatic Remote Sensing Images." IEEE Geoscience and Remote Sensing Letters **16**(2): 271-275.

Liu, J., et al. (2016). "Video classification via weakly supervised sequence modeling." Computer Vision and Image Understanding **152**: 79-87.

Liu, J., et al. (2016). "Recent advances in camera planning for large area surveillance: A comprehensive review." ACM Computing Surveys (CSUR) **49**(1): 6.

Liu, W., et al. (2016). "Robust individual and holistic features for crowd scene classification." Pattern Recognition **58**: 110-120.

Liu, W., et al. (2017). "A survey of deep neural network architectures and their applications." Neurocomputing **234**: 11-26.

Longhi, S., et al. (2012). "Solid waste management architecture using wireless sensor network technology. " New Technologies, Mobility and Security (NTMS), 5th International Conference on, IEEE: 1-5.

Lopez, C. V. and U. Stilla (2018). "Object-based SAR change detection for security and surveillance applications using density based clustering." Synthetic Aperture Radar, VDE EUSAR; 12th European Conference on: 1-6.

Lu, X. and R. Manduchi (2011). "Fast image motion segmentation for surveillance applications." Image and Vision Computing **29**(2-3): 104-116.

Lusk, P. C. (2018). "Vision-Based Emergency Landing of Small Unmanned Aircraft Systems." Master thesis at Birgham Young University.

Mammeri, A., et al. (2012). "A survey of image compression algorithms for visual sensor networks." ISRN Sensor Networks **2012**.

Marcu, A., et al. (2018). "SafeUAV: learning to estimate depth and safe landing areas for UAVs from synthetic data. " Computer Vision, European Conference on, Springer: 1-16.

Maturana, D. and S. Scherer (2015). "3d convolutional neural networks for landing zone detection from lidar." Robotics and Automation (ICRA), International Conference on , IEEE: 3471-3478.

Mayank Garg, A. K., and P.B. Sujit (2015). "Terrain-based landing site selection and path planning for fixed-wing UAVs." Unmanned Aircraft Systems (ICUAS), Denver, Colorado, USA, International Conference on: 246-251.

Memos, V. A., et al. (2018). "An efficient algorithm for media-based surveillance system (EAMSuS) in IoT smart city framework." Future Generation Computer Systems **83**: 619-628.

Messer, H. and O. Sendik (2015). "A new approach to precipitation monitoring: A critical survey of existing technologies and challenges." IEEE Signal Processing Magazine **32**(3): 110-122.

Mittal, M., et al. (2018). "Vision-based Autonomous Landing in Catastrophe-Struck Environments." arXiv preprint arXiv:1809.05700.

Mora, J. E. G., et al. (2018). "Automatic Visual Classification of Parking Lot Spaces: A Comparison Between BoF and CNN Approaches. " Experimental and Efficient Algorithms, International Workshop on, Springer: 160-170.

Mostafaei, H., et al. (2018). "Border surveillance with WSN systems in a distributed manner." IEEE Systems Journal (99): 1-10.

Mulatu, M. A. (2017). "Energy cooperation in communication of energy harvesting tags." AEU-International Journal of Electronics and Communications **71**: 145-151.

Nandhini, S. A. and S. Radha (2016). "Compressed sensing based object detection and tracking system using measurement selection process for wireless visual sensor networks." Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, IEEE: 1117-1122.

Nandhini, S. A., et al. (2015). "Video compressed sensing framework for wireless multimedia sensor networks using a combination of multiple matrices." Computers & Electrical Engineering **44**: 51-66.

Nirmala, D. E., et al. (2013). "Multimodal image fusion in visual sensor networks." Electronics, Computing and Communication Technologies (CONECCT), International Conference on, IEEE: 1-6.

Nogueira, K., et al. (2017). "Towards better exploiting convolutional neural networks for remote sensing scene classification." Pattern Recognition **61**: 539-556.

Obraczka, K., et al. (2002). "Managing the information flow in visual sensor networks." Wireless Personal Multimedia Communications, 5th International Symposium on, IEEE **3**: 1177-1181.

Ohn-Bar, E. and M. M. Trivedi (2017). "Are all objects equal? Deep spatio-temporal importance prediction in driving videos." Pattern Recognition **64**: 425-436.

Oliva, A. and A. Torralba (2001). "Modeling the shape of the scene: A holistic representation of the spatial envelope." International journal of computer vision **42**(3): 145-175.

Ozger, M., et al. (2016). "Event-to-sink spectrum-aware clustering in mobile cognitive radio sensor networks." IEEE Transactions on Mobile Computing **15**(9): 2221-2233.

Padwalkar, U. and D. Ambawade (2013). "Multipath based Energy Efficient (MEE) Routing Protocol for WMSNs." International Journal of Computer Applications **975**: 8887.

Paek, J. and J. Ko (2017). "$ K $-Means Clustering-Based Data Compression Scheme for Wireless Imaging Sensor Networks." IEEE Systems Journal **11**(4): 2652-2662.

Patruno, C., et al. (2018). "A Vision-Based Approach for Unmanned Aerial Vehicle Landing." Journal of Intelligent & Robotic Systems: 1-20.

Patterson, T., et al. (2014). "Timely autonomous identification of UAV safe landing zones." Image and Vision Computing **32**(9): 568-578.

Patterson, T., et al. (2010). "Integration of terrain image sensing with UAV safety management protocols. " International Conference on Sensor Systems and Software, Springer: 36-51.

Peixoto, J. P. J. and D. G. Costa (2017). "Wireless visual sensor networks for smart city applications: A relevance-based approach for multiple sinks mobility." Future Generation Computer Systems **76**: 51-62.

Polvara, R., et al. (2018). "Vision-based autonomous landing of a quadrotor on the perturbed deck of an unmanned surface vehicle." Drones **2**(2): 15.

Qin, L., et al. (2016). "Video scene text frames categorization for text detection and recognition. " Pattern Recognition (ICPR), 23rd International Conference on, IEEE: 3886-3891.

Raghunandan, K., et al. (2018). "Multi-Script-Oriented Text Detection and Recognition in Video/Scene/Born Digital Images. " IEEE Transactions on Circuits and Systems for Video Technology, **29**(4), 1145-1162.

Rao, V. S. H. and M. N. Kumar (2012). "A new intelligence-based approach for computer-aided diagnosis of dengue fever." IEEE transactions on information technology in biomedicine **16**(1): 112-118.

Redondi, A., et al. (2013). "Compress-then-analyze vs. analyze-then-compress: Two paradigms for image analysis in visual sensor networks. " Multimedia Signal Processing (MMSP),15th International Workshop on, IEEE: 278-282.

Rein, S. and M. Reisslein (2016). "Scalable line-based wavelet image coding in wireless sensor networks." Journal of Visual Communication and Image Representation **40**: 418-431.

Romero-González, C., et al. (2017). "On robot indoor scene classification based on descriptor quality and efficiency." Expert Systems with Applications **79**: 181-193.

Sastra, N., et al. (2008). "Modeling Wireless Visual Sensor Network with a Low Energy Consumption for Desired Image Quality and View Point." 6th Kumamoto University Forum, Surabaya-Indonesia, 1: 94-95.

Sastra, N. P. and G. Hendrantoro (2015). "Energy efficiency of image compression for virtual view image over wireless visual sensor network." Journal of Networks **10**(6): 385.

Sastra, N. P. and G. Hendrantoro (2015). "Energy efficiency of image compression for virtual view image over wireless visual sensor network." Journal of Networks **10**(6): 385-394.

Saxena, A. K. and V. K. Chaurasiya (2018). "Multi-resolution texture analysis for fingerprint based age-group estimation." Multimedia Tools and Applications **77**(5): 6051-6077.

Scherer, S., et al. (2012). "Autonomous landing at unprepared sites by a full-scale helicopter." Robotics and Autonomous Systems **60**: 1545–1562.

Shaban, K. B., et al. (2016). "Urban air pollution monitoring system with forecasting models." IEEE Sensors Journal **16**(8): 2598-2606.

Sharma, A. (2015). "Adapting off-the-shelf CNNs for word spotting & recognition.", Document Analysis and Recognition (ICDAR), 13th International Conference on, IEEE: 986-990.

Shaukat, A., et al. (2016). "Visual classification of waste material for nuclear decommissioning." Robotics and Autonomous Systems **75**: 365-378.

Shen, H. and G. Bai (2016). "Routing in wireless multimedia sensor networks: A survey and challenges ahead." Journal of Network and Computer Applications **71**: 30-49.

Shou, Z., et al. (2016). "Temporal action localization in untrimmed videos via multi-stage cnns.", Computer Vision and Pattern Recognition, Proceedings Conference on, IEEE: 986-990.

Shrivastava, P., et al. (2017). "Bridging the semantic gap with human perception based features for scene categorization." International Journal of Intelligent Computing and Cybernetics **10**(3): 387-406.

Sofi, S. A. and R. Naaz (2016). "Data compression in Wireless visual Sensor networks using wavelets. " Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, IEEE: 1286-1290.

Sofi, S. A. and R. Naaz (2016). "Data compression in Wireless visual Sensor networks using wavelets. " Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on, IEEE: 1286-1290.

Soro, S. and W. Heinzelman (2009). "A survey of visual sensor networks." Advances in multimedia **2009**.

Sravanthi, D., et al. (2019). "A Secure Chaotic Image Encryption Based on Bit-Plane Operation. " Soft Computing in Data Analytics, Springer**:** 717-726.

Sukumaran, A. N., et al. (2017). "Compressed sensing based foreground detection vector for object detection in wireless visual sensor networks." AEU-International Journal of Electronics and Communications **72**: 216-224.

Sun, X., et al. (2018). "Perceptual multi-channel visual feature fusion for scene categorization." Information Sciences **429**: 37-48.

Suseela, G. and Y. A. V. Phamila (2018). "Energy efficient image coding techniques for low power sensor nodes: A review." Ain Shams Engineering Journal **9**(4): 2961-2972.

Teschioni, A., et al. (1999). "A neural network approach for moving objects recognition in color image sequences for surveillance applications." NSIP: 28-32.

Tian, D., et al. (2016). "Keypoint trajectory coding on compact descriptor for video analysis." Image Processing (ICIP), International Conference on, IEEE: 171-175.

Ting, K., et al. (2008). "Performance analysis of single and combined bit-planes feature extraction for recognition in face expression database." computer and communication engineering, International conference on, IEEE: 792-795.

Valera, M. and S. A. Velastin (2005). "Intelligent distributed surveillance systems: a review." IEE Proceedings-Vision, Image and Signal Processing **152**(2): 192-204.

Van Opdenbosch, D. and E. Steinbach (2018). "Flexible Rate Allocation for Local Binary Feature Compression. " Image Processing (ICIP), 25th International Conference on, IEEE: 3259-3263.

Velastin, S. A. and P. Remagnino (2006). "Intelligent distributed video surveillance systems," Iet **3**.

Verma, A., et al. (2018). "RNN-LSTM Based Indoor Scene Classification with HoG Features. " Advanced Informatics for Computing Research, International Conference on , Springer: 149-159.

Verma, G., et al. (2018). "Implementation of Smart Video Surveillance System Using Motion Detection Technique. " Sensors and Image Processing, Springer**:** 65-72.

Wagner, R., et al. (2003). "Distributed image compression for sensor networks using correspondence analysis and super-resolution. " Image processing, ICIP. Proceedings. international conference on, IEEE **1**: I-597.

Wald, J., et al. (2018). "Real-Time Fully Incremental Scene Understanding on Mobile Platforms." IEEE Robotics and Automation Letters **3**(4): 3402-3409.

Wang, H., et al. (2007). "Application of image correction and bit-plane fusion in generalized PCA based face recognition." Pattern Recognition Letters **28**(16): 2352-2358.

Wang, H., et al. (2007). "Collaborative image transmissions based on region and path diversity in wireless sensor network. " GLOBECOM Global Telecommunications Conference, IEEE: 971-975.

Wang, H., et al. (2010). "Image transmissions with security enhancement based on region and path diversity in wireless sensor networks." IEEE Transactions on Wireless Communications **8**(2): 757-765.

Wang, P., et al. (2007). "Distributed data aggregation using clustered slepian-wolf coding in wireless sensor networks. " Communications, International Conference on, IEEE: 3616-3622.

Wei, D. and T. Li (2015). "Robust face recognition using voting by bit-plane images based on sparse representation. " Mobile Multimedia Communications, 8th International Conference on: 153-156,

Wu, M. and C. W. Chen (2007). "Collaborative image coding and transmission over wireless sensor networks." EURASIP Journal on Applied Signal Processing **2007**(1): 223-223.

Wu, Z., et al. (2016). "Harnessing object and scene semantics for large-scale video understanding. " Computer Vision and Pattern Recognition, International Conference on, IEEE: 3112-3121.

Xie, L., et al. (2016). "Simple techniques make sense: Feature pooling and normalization for image classification." IEEE Transactions on Circuits and Systems for Video Technology **26**(7): 1251-1264.

Yan, C., et al. (2018). "Supervised hash coding with deep neural network for environment perception of intelligent vehicles." IEEE transactions on intelligent transportation systems **19**(1): 284-295.

Yan, C., et al. (2014). "A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors." IEEE Signal Processing Letters **21**(5): 573-576.

Yan, T., et al. (2003)." Differentiated surveillance for sensor networks. " Embedded networked sensor systems, 1st international conference on, ACM: 51-62.

Yang, D. and H. Park (2016). "A new shape feature for vehicle classification in thermal video sequences." IEEE Transactions on Circuits and Systems for Video Technology **26**(7): 1363-1375.

Yang, K., et al. (2008). "Generalized framework for reduced precision global motion estimation between digital images. " Multimedia Signal Processing, 10th Workshop on, IEEE: 76-81.

Yap, F. G. and H.-H. Yen (2017). "Novel visual sensor deployment algorithm in occluded wireless visual sensor networks." IEEE Systems Journal **11**(4): 2512-2523.

Yi, X. and M. Eramian (2016). "LBP-based segmentation of defocus blur." IEEE Transactions on Image Processing **25**(4): 1626-1638.

Yoo, C.-H., et al. (2017). "High-dimensional feature extraction using bit-plane decomposition of local binary patterns for robust face recognition." Journal of Visual Communication and Image Representation **45**: 11-19.

Yuan, L., et al. (2015). "Improve scene classification by using feature and kernel combination." Neurocomputing **170**: 213-220.

ZainEldin, H., et al. (2015). "Image compression algorithms in wireless multimedia sensor networks: A survey." Ain Shams Engineering Journal **6**(2): 481-490.

Zhang, B., et al. (2018). "One-two-one networks for compression artifacts reduction in remote sensing." ISPRS Journal of Photogrammetry and Remote Sensing: 184-196.

Zhang, J., et al. (2018). "From Text to Video: Exploiting Mid-Level Semantics for Large-Scale Video Classification. " Pattern Recognition (ICPR), 24th International Conference on, IEEE: 1695-1700.

Zhang, X., et al. (2013). "Edge strength similarity for image quality assessment." IEEE Signal processing letters **20**(4): 319-322.

Zhou, Q., et al. (2016). "Multi-scale context for scene labeling via flexible segmentation graph." Pattern Recognition **59**: 312-324.

Zhou, Z., et al. (2018). "Image encryption combining discrete fractional angular transform with Arnold transform in image bit planes." Optica Applicata **48**(2).

Zhu, J., et al. (2016). "A reconfigurable tangram model for scene representation and categorization." IEEE Transactions on Image Processing **25**(1): 150-166.

Zuo, Z., et al. (2015). "Exemplar based deep discriminative and shareable feature learning for scene image classification." Pattern Recognition **48**(10): 3004-3015.