STATUS OF THESIS

Title of thesis	Height Estimation based on Convolutional Neural Network and Sparse Representation Techniques using Aerial Stereo Imagery for Monitoring of Vegetation near Power Lines.

I ABDUL QAYYUM

hereby allow my thesis to be placed at the Information Resource Center (IRC) of Universiti Teknologi PETRONAS (UTP) with the following conditions:

1. The thesis becomes the property of UTP
2. The IRC of UTP may make copies of the thesis for academic purposes only.
3. This thesis is classified as

☐ Confidential

☑ Non-confidential

If this thesis is confidential, please state the reason:

____

____

____

The contents of the thesis will remain confidential for __________ years.

Remarks on disclosure:

____

____

____

Endorsed by

____________________	____________________
Signature of Author	Signature of Supervisor
Permanent address: D.K.505/m-3,	Name of Supervisor
near 6th road, Dhok Paracha,	Assoc. Prof. Dr. Aamir Saeed Malik
Rawalpindi, Pakistan	
Date : __________	Date : __________

UNIVERSITI TEKNOLOGI PETRONAS

HEIGHT ESTIMATION BASED ON CONVOLUTIONAL NEURAL NETWORK AND SPARSE REPRESENTATION TECHNIQUES USING AERIAL STEREO IMAGERY FOR MONITORING OF VEGETATION NEAR POWER LINES

by

ABDUL QAYYUM

The undersigned certify that they have read, and recommend to the Postgraduate Studies Programme for acceptance this thesis for the fulfillment of the requirements for the degree stated.

Signature: ____________________

Main Supervisor: Assoc. Prof. Dr. Aamir Saeed Malik

Signature: ____________________

Co-Supervisor: Assoc. Prof. Dr. Mohamad Naufal Bin Saad

Signature: ____________________

Head of Department: Assoc. Prof. Ir. Dr. Nursyarizal Mohd Nor

Date:

HEIGHT ESTIMATION BASED ON CONVOLUTIONAL NEURAL NETWORK AND SPARSE REPRESENTATION TECHNIQUES USING AERIAL STEREO IMAGERY FOR MONITORING OF VEGETATION NEAR POWER LINES

by

ABDUL QAYYUM

A Thesis

Submitted to the Postgraduate Studies Programme

as a Requirement for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND ELECTRONIC ENGINEERING

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SERI ISKANDAR,

PERAK

MARCH 2018

DECLARATION OF THESIS

Title of thesis	Height Estimation based on Convolutional Neural Network and Sparse Representation Techniques using Aerial Stereo Imagery for Monitoring of Vegetation near Power Lines.

I ABDUL QAYYUM

hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTP or other institutions.

Witnessed by

Signature of Author	Signature of Supervisor
Permanent address: D.K.505/m-3, near 6th road, Dhok Paracha, Rawalpindi, Pakistan.	Name of Supervisor Assoc. Prof. Dr. Aamir Saeed Malik
Date :	Date :

DEDICATION

This thesis is dedicated to: The sake of Allah, my Creator and my Master, My great teacher and messenger, Mohammed (P.B.U.H), who taught us the purpose of life, My great parents, who never stop giving of themselves in countless ways, My dearest wife, who leads me through the valley of darkness with light of hope and support, My teachers who helped and support me in this critical journey, My beloved brothers and sisters who stands by me when things look bleak, To all my family, the symbol of love and giving, My friends who encourage and support me.

All the people in my life who touch my heart, I dedicate this research.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Merciful, the Most Compassionate all praise be to Allah, the Lord of the worlds; and prayers and peace be upon Mohamed His servant and messenger. First and foremost, I must acknowledge my limitless thanks to Allah, the Ever-Magnificent; the Ever-Thankful, for His help and bless. I am totally sure that this work would have never become truth, without His guidance. I owe a deep debt of gratitude to our university for giving us an opportunity to complete this work. I am grateful to some people, who worked hard with me from the beginning till the completion of the present research particularly my supervisor Assoc. Prof. Dr. Aamir Saeed Malik, who has been always generous during all phases of the research, and I highly appreciate the efforts expended by Assoc. Prof. Dr. Assoc. Prof. Dr. Mohamad Naufal Bin Mohamad Saad, Prof. Dr. Tuan bin Rashid, Dr. Mehboob Iqbal and Assoc. Prof. Dr. Rosdiazli B Ibrahim. I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis. I also would like to express my wholehearted thanks to my family for their generous support they provided me throughout my entire life and particularly through the process of pursuing the Ph.D. degree. Because of their unconditional love and prayers, I have the chance to complete this thesis.

I owe profound gratitude to my wife, Moona, whose constant encouragement, limitless giving and great sacrifice, helped me accomplish my degree. I am very appreciative to my colleagues at Universiti Teknologi Petronas, who participated in this study. I am also grateful to the fellows at the Universiti Tenaga Nasional who let me implement my work with their support. Without their support, this study would not have been possible.

Last but not least, deepest thanks go to all people who took part in making this thesis real.

ABSTRACT

The dangerous overgrown vegetation / trees under high voltage (HV) transmission lines right-of-ways (ROWs) have caused severe blackouts / flashovers due to interference with the power lines which leads to short-circuiting among the conductors. Therefore, these dangerous encroachments are monitored periodically along the electrical distribution networks ROWs through visual inspection, or by the airborne system. Airborne LiDAR scanners, videography, and aerial multispectral images are now available for the monitoring of HV transmission lines from those unintended encroachments such as trees/vegetation. Each of these methods has their own attributes and limitations and have proved to be costly, time-consuming and not much accurate. This thesis proposes an innovative idea of utilizing aerial (UAV and satellite) stereo images for monitoring dangerous vegetation (trees, shrubs, and plants, etc.,) below and near HV lines ROWs. The main focus is to develop a system to monitor vegetation / trees near transmission lines poles using aerial based stereo images. After pre-processing and orthorectification of stereo images, the proposed algorithms based on Convolutional neural networks (CNN) and sparse representation (SR) has been used to estimate the disparity map which further used to estimate the height of vegetation / trees near power transmission lines / poles. The proposed system based on CNN and SR design are used to identify different threat levels by estimated the distance between vegetation / trees and power transmission lines and height of towers, vegetation, and trees in the ROWs. The proposed algorithms are compared with the state-of-the-art stereo matching algorithms based on performance metrics (accuracy, precision, and recall). The results show that proposed algorithm based on CNN model achieved highest performance metrics (91% accuracy) as compared to extant stereo matching algorithms. The performance evaluation of real-time developed proposed algorithms prove the feasibilities of integrating the method for HV transmission line maintenance.

ABSTRAK

Tumbuh-tumbuhan berbahaya yang terlalu besar / pokok-pokok di bawah garis transmisi voltan tinggi (HV) yang betul (ROW) telah menyebabkan terputus bekalan elektrik / lebihan kuasa akibat gangguan terhadap talian kuasa yang membawa litar pintas di antara konduktor. Oleh itu, pencerobohan berbahaya ini dipantau secara berkala sepanjang rangkaian pembahagian elektrik ROW melalui pemeriksaan visual, atau oleh sistem udara. Pengimbas udara LiDAR, videografi dan imej multispektral udara kini boleh didapati untuk pemantauan talian transmisi HV dari pencerobohan yang tidak diingini seperti pokok / tumbuh-tumbuhan. Setiap kaedah ini mempunyai ciri-ciri dan batasan tersendiri dan telah membuktikan kos yang tinggi, pengunaan masa yang panjang dan kurang tepat. Tesis ini mencadangkan idea yang inovatif dengan menggunakan imej stereo udara (UAV dan satelit) untuk memantau tumbuh-tumbuhan berbahaya (pokok, pokok renek, dan tumbuh-tumbuhan, dan lain-lain) di bawah dan berhampiran garisan HV ROW. Tumpuan utama adalah untuk membangunkan sistem untuk memantau tumbuh-tumbuhan / pokok berhampiran kutub talian transmisi menggunakan imej stereo berasaskan udara. Selepas pra-pemprosesan dan orthorectifikasi imej stereo, algoritma yang dicadangkan berdasarkan rangkaian neural convolutional (CNN) dan perwakilan jarang (SR) telah digunakan untuk menganggarkan peta perbezaan yang digunakan untuk menganggarkan ketinggian pokok / tumbuh-tumbuan berhampiran talian transmisi kuasa / tiang. Sistem yang dicadangkan berdasarkan reka bentuk CNN dan SR digunakan untuk mengenal pasti tahap ancaman yang berbeza dengan menganggarkan jarak antara tumbuh-tumbuhan / pokok-pokok dan talian transmisi kuasa dan ketinggian menara, tumbuh-tumbuhan, dan pokok-pokok di ROW. Algoritma yang dicadangkan dibandingkan dengan algoritma pemadanan stereos yang bersesuaian berdasarkan metrik prestasi (ketepatan, kejituan, dan ingatan). Keputusan menunjukkan bahawa algoritma yang dicadangkan berdasarkan model CNN mencapai metrik prestasi tertinggi (ketepatan 91%)

berbanding dengan algoritma pemadanan stereo yang ada. Penilaian prestasi algoritma yang dicadangkan pada masa nyata membuktikan kemungkinan mengintegrasikan kaedah untuk menyelenggarakan saluran transmisi HV.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HV	High voltage
ROWs	Right-of-ways
HT	High tension
NESC	National electrical safety code
TNB	Tenaga nasional berhad
DEM	Digital elevation model
LiDAR	Light detection and ranging
DGPS	Differential global positioning system
INS	Inertial navigation system
DP	Dynamic programming
GC	Graph-cut
CNN	Convolutional Neural Network
SR	Sparse Representation
SC	Sparse coding
SSD	Sum of squared difference
SAD	Sum of absolute difference
NCC	Normalized cross correlation
BMA	Block matching algorithm
DRT	Discrete ridgelet transform
DCT	Discrete cosine transform
DWT	Discrete wavelet transform
DTT	Discrete Tchebichef Transform
OMP	Orthogonal matching pursuit
BP	Basis pursuit
DP	Deep Learning
BPD	Basis pursuit de-noising
SAE	Sparse auto encoder
LASSO	Least absolute shrinkage and selection operator

LARS	Least angle regression
FOCUSS	Focal underdetermined system solver
SBL	Sparse bayesian learning
ORMP	Orthogonal regression matching pursuit
OCD	Over-complete dictionary
SGD	stochastic gradient descent
GSD	Ground sample distance
PS	Pixel size
TP	True positive
TN	True Negative
FN	False negative
FP	False positive
UAV	Unmanned aerial vehicle
ReLU	Rectified linear unit
BP	Belief propagation
CA	Cooperative algorithm
ROC	Receiver operating characteristic
GPS	Global positioning system

LIST OF SYMBOLS

O	The size of convolutional output feature map
I	Input image
F	Spatial filter
P	Zero padding
S	Stride in convolutional layer
α_1	Sparse coefficients
D	Dictionary
ε	Sparsity level
$\emptyset$	Empty set
τ	Small constant used in OMP algorithm
d	Dictionary element
H	Disparity map
f_1	Function used to compute Euclidian distance
k	Disparity range
B_l	Left image patch
B_r	Right image patch
T_0	Threshold value used in sparse representation
B_1	Recurrence coefficients
f_{n_1}	Tchebichef polynomials.
f_{n_2}	Tchebichef polynomials.
ϑ	Ricker wavelet function
a	Scale parameter used in discrete ridgelet transform
b	Location parameter used in discrete ridgelet transform
θ	Orientation parameter used in discrete ridgelet transform
T_1	Threshold set in DRT dictionary
f	Feature map
α	Activation function
w	Convolutional weight
K	Normalization constant
β	Variance used in normalization layer
p	Size of the local spatial region
c_1	Fully connected layer feature vector
λ	Wavelength of the sensor
U	Object height
F	Focal length
t	Distance between sensors
PS	Pixel size
$\mathrm{m}_{\mathrm{tline}}$	Slope of the transmission lines
θ_{tline}	Angle between two lines
$D_{l_1-t_1}$	Distance between tree and lines
D_{2-t_2}	Distance between line and power poles.

CHAPTER 1

INTRODUCTION

1.1 Overview

The overhead high voltage (HV) lines distribute the electrical energy from point of generation to the point of consumption. Highly elevated structures are needed to carry transmission lines and HV distribution lines to the consumer's edge as shown in Figure 1.1. Low voltage distribution lines are usually installed along the side path with the roads. In order to ensure safety and reliability of transmission system, it is necessary for electrical utilities (power companies) to maintain clearance from encroached vegetation and trees for the overhead lines. High tension (HT) overhead power lines are often installed with bare copper conductors that are prone to blackouts / flashovers [1]. A power line corridor describes the strip of land upon which utility companies construct their electrical infrastructure. Monitoring power line corridors are crucial for the reliability of electricity transmission. Trees and shrubs often create obstructions in corridors and pose risks to power lines, and therefore utility companies need to scrutinize where and how trees grow in or close to power line corridors [2]. Blackouts / flashovers are the grounding of large electrical power, or simply the short-circuiting of electricity to the ground [1]. When protection systems of transmission lines observe similar fault, the circuit protection elements such as circuit breakers and fuses etc. may trip the supply of electricity at that particular distribution section. The event of flashovers disrupts the distribution and transmission of electricity. For a single phase or a three-phase system, the fault occurs when they come in contact with the ground or any grounding object.

HV overhead line conductors come in contact with each other to cause flashovers. It occurs when the conductors come closer to one another by passing the minimum allowable distance as shown in Figure 1.2 owing to heavy wind/storm, the oscillation or sagging of HV conductors can also lead to blackouts / flashovers. However, one of the main reasons for flashovers among transmission lines is encroachments near to the lines [3]. Overgrown vegetation under the overhead lines may strike / interfere with the HV line circuitry to cause short-circuiting. Figure 1.3 shows flashovers due to the interference of excessed vegetation growth with power lines. In addition to economic / financial losses, the flashovers result in outages for the customers as well. Blackouts covering a large area are disastrous to industrial and business activities that rely on HV electrical power for their operations. This is especially true for plants and industries to which utility companies are bound to supply uninterrupted electrical power under the laws and ordinances of electricity.

Figure 1.1: a) 500 kV highly elevated poles carrying HV transmission lines. b) 275 kV transmission lines carrying electrical power to cities [1].

Figure 1.2: HV transmission line conductors closer to one another [2].

Figure 1.3: (a) Blackouts that occur due to the tree encroachments within the residential areas. (b) Outages due to the collision of Limbs [3].

It is, therefore, obligatory for the power suppliers to inspect and save their HV transmission lines against that hazardous vegetation (including trees, shrubs, or plants etc.). In the event of a bushfire, the vegetation growth close to the power lines may pose a serious threat of damaging the conductors. Such vegetation may also pave the way for animals to intrude the line. In urban areas, vegetation encroachment is less serious than in rural areas as the access is much easier and prompt maintenance can be achieved. Moreover, local councils and private landowners regularly maintain their trees facilitating the overall maintenance process. However, in rural areas, inspection maintenance becomes difficult due to limited access and large distances to cover. In these areas, traditional calendar-based tree trimming is often a strategy used by energy distributors. Other short-term strategies might be to identify and remove nearby objects (i.e. buildings and vegetation) found near power lines. More generally, the risk of manmade structures can be controlled through building regulations. However, vegetation grows naturally and particularly in rural areas, the growth of vegetation is unmanaged. Strong winds and storms can bring branches or even entire trees into contact with power lines. Unmanaged vegetation can also grow up into power lines and cause bushfires. Unfortunately, vegetation management over large powerline networks is cumbersome and expensive. To manage the tradeoff between safety and cost, utility companies often introduce a regular maintenance cycle (e.g. say once every 5 years). This strategy assumes that once inspected, properly maintained vegetation can be assumed to remain separated from the power line infrastructure until the maintenance cycle is repeated 5 years later.

However, trees often grow unexpectedly during the period between maintenance due to inaccurate vegetation growth modeling and climate changes. The subjective nature of conventional maintenance strategies often result in some zones being trimmed more frequently than required, or conversely, some zones not being trimmed often enough. Remote sensing technologies represent an attractive and potentially automated solution for power line corridor monitoring activities. Recent efforts toward remote sensing based methods include improved data collection using satellite sensors [4],[5], an airborne stereo vision system [6] and unmanned aerial vehicles (UAVs) [7],[8]. Recently airborne laser scanning has attracted particular attention in power line corridor monitoring problem due to the possibility of achieving three-dimensional models of infrastructure[2],[9]. Many utility companies and researchers, if not most, use commercial data and software but also generate algorithms tailored for their needs. However, there are some critical issues to be addressed regarding automated data collection and processing over complicated power line networks. For example, the quality of collected data has much to do with aircraft platform stability. The complex nature of navigating aircraft over extensive amounts of power line networks calls for an increased level of flight automation. To increase the reliability of the information extraction from remote sensing data, combining the complementary information derived from multi-source data can be very useful.

The vegetation-related outages could easily have been mitigated by early fault diagnosis using remote sensing technology. However, current acquisition costs and the related manually intensive process of reviewing the remotely sensed data make such a service impractical and commercially unviable. To overcome this limitation an automated process is urgently required to determine the potential impact on power-line safety by observing changes in clearances to vegetation and other objects, operating temperature, the detection of new buildings or structures alongside or between towers and their associated line spans, erosion-induced terrain changes, and tree health and the detection of physical damage or deterioration of structures, wires or other assets. Traditionally such corridor analysis has relied on labour intensive manual approaches that entail manual inspection or the capture and inspection of video footage captured on site by ground personnel or during airborne patrols.

1.2 Motivation

The regular monitoring of a power transmission network is of the utmost importance. Many methods are used to estimate the height of vegetation in vulnerable zones, including traditional manual line inspection; albeit, time-consuming and subject to human error [2]. Moreover, severe weather and dangerous terrain that can host wild animals increase risks for employees. Such limitations and risks invite efficient monitoring systems that simplify the process and reduce overall threats to power networks and human lives. Aerial remote sensing is useful because aerial data can effectively and quickly identify and isolate potential risks to a power system, which allows power company employees to selectively clear areas of encroachment. Aerial remote sensing utilizes satellites and aircraft to collect large-scale earth data. Moreover, airborne systems are upgraded at regular intervals; thus keep abreast as technological advances improve sensors that offer higher spectral and spatial resolutions such as LiDAR [10]. However, traditional piloted airborne systems have high operational costs, which is the major limitation that prevents wider use. Another drawback of piloted aircraft is its higher level of risk to personnel [11]. Furthermore, these techniques have their limitations in detecting and estimating the extent of vegetation encroachment of HV (high voltage) line networks, especially in areas of non-uniform terrains.

Unmanned aerial vehicles (UAV) with remote sensors for data collection address the limitations of conventional pilot-based airborne sensors [12]. Thus, a UAV-based system may provide a cost-effective and flexible solution for spatial data acquisition. Recent technological advances in UAV based systems have undoubtedly made UAV an attractive candidate for research in many domains. This includes the efficient and effective monitoring of power transmission lines because they provide an effective and low-cost solution that obtains detailed maps of vegetation groupings and tree levels near power lines. The aerial based stereo images such as UAV may provide the useful solution for monitoring the vegetation / trees near power transmission lines / poles. The image processing and pattern recognition methods can be used to monitor the vegetation / trees near power transmission lines by measuring the distance between heightened objects (vegetation, trees, and shrubs) and power transmission lines/poles. In order to monitor the objects (vegetation / trees, shrubs) near power lines, the accurate and cost-effective solution is needed. For a better estimation of the distance between objects and power transmission lines, the stereo matching algorithms are used to estimate the accurate disparity map which further used to estimate the height of objects. There are plenty of stereo matching algorithms used to estimate the disparity map [13]. However, these algorithms cannot provide the feasible disparity maps [15],[16].

Other than the existing stereo matching algorithms, sparse representation has been successfully used in other image areas such as image de-noising, image inpainting, image enhancement, image compression, image classification and many other applications in signal processing. The sparse representation has a capability to extract the prominent features from the stereo imagery in a compressed form that could make sparse representation a better approach in estimating the disparity map because the accurate disparity map required perfect matching features in stereo images.

Similar to sparse representation, the deep learning algorithms have been widely used in image classification, pattern recognition, and object detection. The Convolutional Neural Network (CNN) is the famous deep learning algorithm used in high-level image application such as classification, segmentation and object detection. Recently the CNN has been deployed on the low-level vision such as motion detection, segment labeling, and stereo matching. The CNN extracts prominent features using the stereo patches extracted from the stereo images which necessary for accurate disparity map estimation. Therefore, this thesis focuses on sparse representation and CNN based algorithms to provide better disparity map for monitoring the vegetation / trees near and under the power transmission power lines / poles using aerial stereo imagery.

1.3 Research Problem Statement

Traditional power transmission network monitoring involves foot patrolling which is time-consuming, subject to human errors and risky due to bad weather conditions and wild animals. To overcome the limitations of existing methods, there is a need for efficient monitoring to enhance the accuracy and to avoid the risks. In contrast to these foot patrolling approaches, aerial stereo imagery effectively and quickly identify and isolate the potential risks to a power system, which allows power company employees to selectively clear areas of encroachment.

In aerial stereo imagery, the stereo matching algorithms should be used to estimate the accurate disparity map which further used to estimate the height of objects for better estimation of the distance between objects and power transmission lines, There are plenty of stereo matching algorithms are used to estimate the disparity map [13]. There are many constraints have been found in stereo matching processes such as noise, textureless region [10], depth discontinuity and occlusion detection [11]. A highly textured region is needed to propagate data from left to right in stereo images. Without textureless regions, depth and accuracy are reduced due to point mismatching [12]. In-depth discontinuity, the data is improperly propagated or ceases at object boundaries. To compute depth discontinuity at a boundary, improved stereo matching techniques must be developed [13]. Moreover, occlusion occurs when some obstacles appear in both right and left views because of which some pixels cannot be matched within the reference image [14].

Existing stereo algorithms such as global matching algorithms; graph cut (GC) [15],[16],[17] and dynamic programming (DP) [14],[18],[19] and local matching algorithm that are based on area or windows can address the limitation of stereo matching process at some extent. However, the drawback of global matching algorithms is that they are computationally complex and consume more time as compared to the area based algorithms which are less accurate. The existing global stereo matching algorithms based on GC and DP could not provide the reliable disparity map that has been used in height estimation of objects (vegetation / trees) near power transmission lines / poles. The DP algorithm propagates a local error through all the search line between stereo images and it also requires uniqueness and ordering constraints in stereo images. The DP also provides the streaking effect in the stereo matching process for disparity estimation. The DP did no handle textureless and depth discontinuity in the stereo matching process for disparity map estimation using stereo images [14]. The GC can cause excess regularization (over smoothing) between each cut of the graph and disparity tends to blur using large displaced values. The GC may be produced better disparity map in occlusion regions of the stereo images but do not provide an optimal solution in textureless regions due to large displaced values in stereo matching [15]. However, these algorithms cannot provide the feasible disparity map and height estimations for vegetation / trees monitoring near power transmission lines.

In contrast to existing stereo matching algorithms, sparse representation algorithm has been successfully used in other image areas such as image de-nosing, image inpainting, image enhancement, image compression, image classification and many other applications of signal processing. The sparse representation has a capability to extract the prominent features from the stereo imagery in a compressed form which could make the sparse representation a better solution to estimate the disparity map because an accurate disparity map required perfect matching features in stereo images. Likewise, sparse representation, the Convolutional Neural Network (CNN) is the famous algorithm that has been used in high-level image application such as classification, segmentation and object detection. The CNN has outperformed in stereo matching as compared to existing stereo matching algorithms. In order to provide the better disparity map, the CNN extract prominent feature using the stereo patches extracted from the stereo images. To address the issues of existing vegetation/trees monitoring techniques of foot patrolling and aerial imagery, sparse representation, and CNN algorithms are proposed in this thesis to handle textureless regions and depth discontinuity.

1.4 Research Hypothesis

The conventional methods for monitoring the transmission lines / poles due to impending vegetation / trees right-of-ways have proven to be high-priced, time-consuming and less accurate. To overcome these issues, the proposed algorithms in this thesis that will use for height estimation of objects (vegetation / trees, power poles) using aerial stereo imagery. The proposed algorithms using aerial (UAV and satellite) stereo images based on a stereo matching techniques could be less time to consume and accurate. The stereo vision algorithms (CNN and sparse representation) are implemented that have the ability to measure the height of vegetation and trees near or under the transmission lines / poles and also have the ability to estimate the distance among them.

- The proposed algorithms based on CNN and sparse representation will estimate the better disparity map because they can extract prominent features in aerial stereo images.
- The proposed algorithms based on sparse representation and CNN have the ability to estimate the height of vegetation / trees near or under the transmission poles because they can estimate the better disparity maps.

1.5 Research Objectives

Nowadays, transmission power line monitoring is an on-site surveying process that urgently requires the improved capability to quickly and accurately detect, identify and monitor threatening objects within the corridor. These key corridor objects include terrain, vegetation, towers, power lines, buildings, roads, and waterways. Traditionally such corridor scene analysis has relied on labor-intensive manual approaches that entail investigation of video footage captured on site and other monitoring methods. Hence, today's corridor mapping practice still remains an expensive manual process that is not suitable for the large-scale and a rapid commercial compilation of corridor maps. The different aerial inspection methods are used to monitor these corridor objects (vegetation / trees) near power transmission lines. These aerial methods are expensive and time-consuming. The present research work mentions a new technique for continuous monitoring of HV (high voltage) overhead lines against the hazardous vegetation / trees encroachments by using aerial stereo matching algorithms. The stereo matching techniques are most elegant techniques that have been used to estimate the distance between heightened corridor objects (vegetation / trees) and power transmission lines The solution needs less time to cover more area of interest with low cost and better accuracy.

The existing stereo matching algorithms do not handle textureless regions and depth discontinuity in the stereo matching process and may not produce accurate disparity map that is used for estimation of the height of dangerous objects (vegetation / trees) near and under the power transmission lines / poles. In order to effectively handles the textureless regions and depth discontinuity in the stereo matching process for better disparity map estimation, the spare representation and CNN algorithms are proposed in this thesis. This thesis aims at developing an automated monitoring system using sparse representation and CNN algorithm for accurate disparity map estimation which further used for height and distance measurement of vegetation / trees near power lines / poles from aerial stereo images. In order to minimize the issues such as textureless regions and depth discontinuity for better disparity map and height estimation using remote sensing data for vegetation / trees monitoring near power transmission lines, the following number of objectives have been implemented in this thesis:

- To develop the algorithms based on CNN and sparse representation for disparity map estimation using aerial stereo imagery.
- To calculate the height of vegetation / trees in transmission lines corridor and estimate the distance between vegetation / trees and power transmission lines for managing the various number of threats among them (vegetation, trees, and transmission lines) using aerial stereo images.

Based on the height and distance estimation from the accurate disparity map using proposed algorithms (sparse representation and CNN), the different number of threats are computed for monitoring the dangerous objects (vegetation / trees) near power transmission lines and poles using aerial (satellite and UAV) stereo images. The performance analysis (accuracy, precision, recall) has been evaluated between the estimated height calculated by proposed and existing algorithms and ground-truth height of the area of interest.

1.6 Scope of Research

The height of trees and vegetation near power transmission lines is measured using aerial images based on sparse representation and CNN stereo matching technique. The proposed method provides the automated monitoring system that is more accurate and reliable for power management companies. It provides the automatic generated threat level information that could help to locate the danger zone area to overcome the future threats of blackouts. In this thesis, the only rectified aerial stereo images have been used and the dataset used in the proposed system is based on aerial stereo imagery. The key assumptions or scope of this research described that the proposed method used un-cloudy satellite stereo images as the height estimation is not possible where the satellite images have cloudy area. For height estimation, the images must be stereo and orthorectified. The satellite and UAV sensors based stereo images are cheap as compared to other sensors based on LIDAR, helicopter monitoring and piloted monitoring. The proposed algorithms are applied on the captured stereo images where the processing was offline unlike real time processing. The proposed algorithms used rectified stereo images (images are displaced in horizontal direction). The proposed algorithms used small displaced or disparity values to measure the disparity map, the larger disparity values can be used but it may affect the results accuracy.

1.7 Research Contribution

The most significant contributions of this thesis for power line monitoring using aerial stereo imagery (satellite and UAV) are: (i) disparity map estimation based on sparse representation algorithm using various proposed dictionaries on stereo aerial images, (ii) disparity map estimation based on convolutional neural network (CNN) algorithm which provides higher accuracy as compared to our proposed sparse based algorithm as well as existing disparity map estimation algorithms. The CNN based method captures the prominent or unique similar points in both stereo images in order to develop disparity map using feature vector as last fully connected layer, (iii) the height estimation system based on proposed sparse representation and CNN algorithms provided the information about different threat levels near and under the power poles and HV lines. These threat levels are quantified based on estimated height and distance between vegetation / trees and power transmission lines, these are denoted as high, medium and low threats based on standard height and distance specifications. The proposed monitoring system provides cost-effective and reliable solution for power lines companies.

1.8 Thesis Structure / Outline

This thesis is organized as follows: Chapter 2 discusses the related work on the vegetation management and monitoring practices for HV transmission lines. This chapter provides an overview of power line hazards due to vegetation encroachments and history of famous power outages occurred in the world and specifically in Malaysia due to vegetation encroachments, which caused huge financial and economic losses, and standards of vegetation management for HV transmission lines as prescribed by National Electrical Safety Code (NESC). The methods of vegetation monitoring near HV power lines have been discussed based on foot-patrolling, video surveillance using a helicopter, aerial inspection using LiDAR and satellite stereo images. The classification of state-of-the-art stereo matching algorithms has presented with an overview description of proposed height estimations based on sparse representation and deep learning based (CNN) techniques. The related work on the sparse representation and deep learning has been discussed and critically analyzed the limitations in the existing as well as the proposed techniques.

Chapter 3 discusses the disparity map computation based on the proposed algorithms and state-of-the-art stereo matching algorithms using UAV and satellite stereo images to identify endangering encroachments vegetation near power transmission lines and poles. The acquired datasets from the area of interest based on the Satellite and UAV stereo sensors have been discussed in detail. This chapter explains the complete methodology based on proposed and existing stereo matching use to measure the height of vegetation and trees near power transmission lines and power poles and estimate the distance between vegetation / trees and power transmission lines. This chapter proposes a novel method using CNN and sparse representation algorithms for depth estimation from aerial stereo images and derived mathematical models to identify actual metrics of the encroached vegetation including; absolute height of trees, the distance between HV lines / poles and vegetation / trees within and outside the power transmission poles.

Chapter 4 addressed the simulation results on disparity map estimation and the different tuning parameters have been discussed for optimization of proposed algorithms. This chapter is dedicated to present the results of proposed height estimation algorithms using different scenarios of UAV and satellite stereo images with a different number of power transmission poles and disparity values. In this chapter, the various number of threats have been identified using proposed and extant stereo matching algorithms based on estimated height and distance of vegetation / trees near and under power transmission lines and poles. Finally, chapter 5 provides concluding remarks on the proposed work and some key points of the future improvements are highlighted that can be made in this thesis.

CHAPTER 2

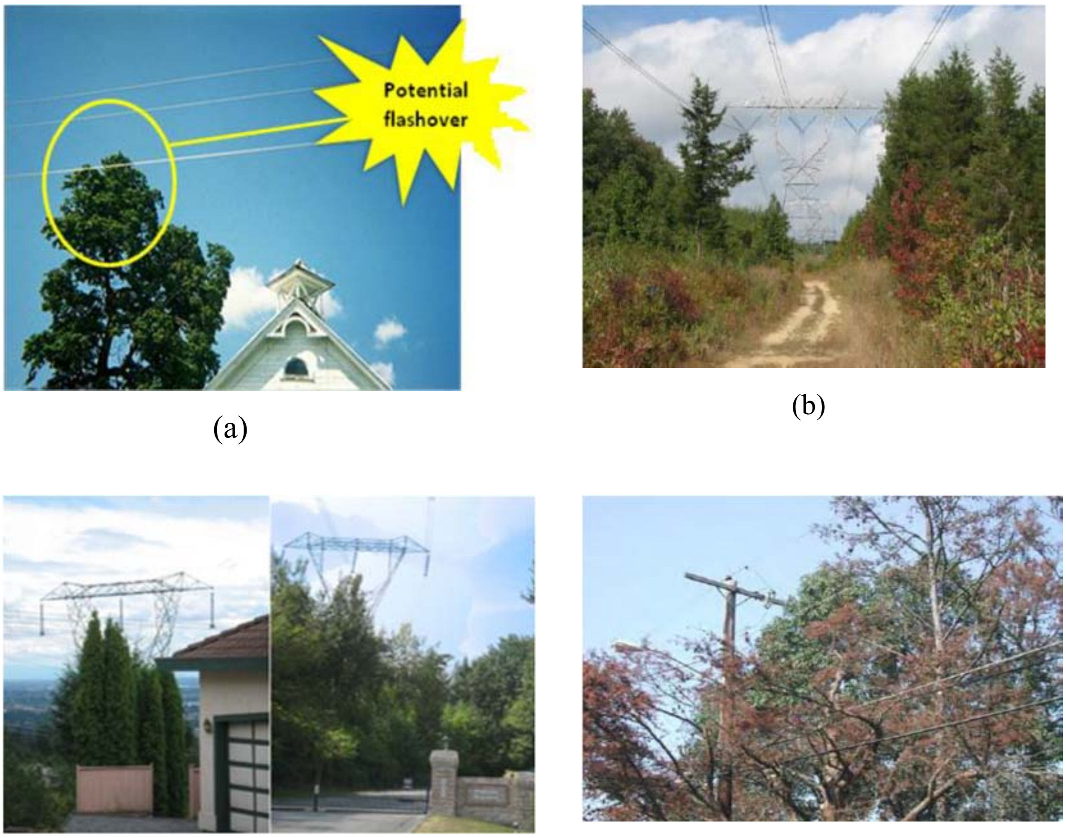
LITERATURE REVIEW

2.1 Introduction

The roving branches of vegetation under HV transmission lines or trees falling on them can cause serious interruptions of the power supply and damages to the electrical utilities and human beings living nearby those HV lines. This creates a chaotic condition for the utilities to restore the power supply by taking swift corrective measures involving high repair cost. In this chapter, a brief account of blackouts / flashovers due to vegetation encroachments, vegetation management standards for the maintenance of HV lines, and the traditional techniques of monitoring the HV lines has been described and also discussed the comparison between the existing monitoring techniques. Further, in this chapter also described the stereo matching algorithms and their applications in image and computer vision problems. The detail of each algorithm has been provided and also gave the limitation of stereo matching algorithms. The introduction of proposed algorithms has been described in this chapter and also discussed the related work of proposed algorithm.

2.2 Tree Encroachment and Vegetation Hazards

Tree encroachment is the largest factor to damage the power transmission lines due to their inference with the power line and it needs to be inspected near the power transmission poles and lines [16]. When the conductor strikes or contact with the branches or bushfire of the trees, it may cause the flashovers and due to striking with the conductor, it may damage psychical or done to weaken or even break the conductor [17],[18]. The mechanism and nature of vegetation-induced faults, including low-current and high-current event, are presented in [17]. Low-current events are those which do not trip protection devices whereas high-current events cause protection devices to trip. The vegetation or trees management is important near power transmission lines to save the electricity as well as expected harmful occur due to the contacting the trees with electrical infrastructure. It is necessary for the power distribution companies to manage the trees or vegetation near the transmission lines. The main causes of power interruption are the vegetation encroachment indicates that the power distribution companies spent most of the budget on the trimming of trees, which cause the major portion of the electrical supply interruption. Figure 2.1 shows the trees and shrubs striking or nearly to interfere the HV overhead lines. Due to the encroachment of tree, it contributes the major cost overhead and big financial losses for electrical supply companies.



(a) (b)

(c) (d)

Figure 2.1: (a) The potential interference of tree with power line due to its excess growth. (b) The trees nearly to strike 500 kV HV line in pinelands near New Jersey. (c) Trees encroaching near 275 kV overhead lines in Washington. (d) Shrubs clearly striking the power lines in Kansas [17].

There are many blackouts occur due to tree striking with heavy transmission lines in Europe as well as in Malaysia. Some of the main blackouts discussed in this section.

2.2.1 Western U.S. Blackout of August 1996

The northwestern U.S. power grid system in Western Electric Coordinating Council (WECC) area suffered two power outages in 1996. Among those, the blackout of August 10, 1996 [19],[20] was the most severe one. During the year 1996, heavy rain and hot weather in California, Washington, and Oregon resulted in the faster tree growth than normal weather and the sagging of 500 kV HV transmission lines into tree branches. The trees were not cleared ROWs accordingly which contributed to the blackouts from HV conductors and trees referred to as “tree faults” [20]. As a result of this mishap, about 7.5 million consumers were affected and it also resulted in the interruption of 30,000 megawatt load and generating a capacity of over 25,000 megawatts was also dropped from the system.

2.2.2 North American Blackouts of August 14, 2003

On August 14th, 2003, a supplier utility working in North America could not properly manage vegetation growth along its distribution network ROWs and it became one of the main causes that resulted in outages of three 345 kV lines and one 138 kV line [21]. Nearly, about 50 million people were affected by these blackouts. Roughly about 63 GW of the load was interrupted, which equivalent to about 11% of the total load is served in North U.S. system. Consequent upon the event, over 400 transmission lines and 531 generating units at 261 power plants tripped. These cascading outages affected most of the New York state and a part of Michigan, Ohio, Pennsylvania, and Canada as well. In the year 2004, a report [22] by U.S.-Canada Power System Task Force indicated tree encroachments as one of the major causes of those massive outages of August 14, 2003. There are many states were affected due to occurrence of the blackouts and vegetation encroachment is the biggest cause of blackouts in these states. The trees and other equipment’s faults also play key role in these blackouts in the Canada states.

2.2.3 Italian Blackouts of September 28, 2003

In the same year, a blackout occurred in Italy when tripping of a major tie-line between Italy and Switzerland was caused by a tree flashover [21], [23]. The delay in corrective measures resulted in tripping of a second 380 kV line on the same border (Italy-Switzerland). Owing to the cascading sequence, the power deficit was such that Italy lost synchronization with rest of the European countries within few seconds and HV lines on the interface between Italy and France tripped because of distance relays. The similar outage happened for the 220 kV and 380 kV interconnections at Italian interface with Austria and Slovenia. It was the worst blackout in Italian history which left great losses of about 6400 mega-watts with the Italian system. Nearly 60 million customers were affected by these outages. In view of the above-reported disastrous incidents, it is evident that vegetation encroachments especially trees cannot be taken lightly even though they may not always cause tripping of a power line.

2.2.4 Vegetation Hazards for HV Overhead Lines in Malaysia

There have been frequent flashovers happened due to the trees or vegetation encroachment in Malaysia, it has a lot of power outages due to this vegetation or tree encroachments. The first blackout occurred in 1985 due to the inference the excessive vegetation with HV power lines Peninsular (Western) Malaysia, causing load losses (tripping) of 600-megawatt power between Paka, Terengganu and Kampung Awah, Pahang [24]. The next blackouts occurred on August 3, 1996, in Paka, there are massive trees fallen on the HV transmission lines and cause power outages in many towns of Peninsular Malaysia. These towns are Putrajaya, Johor, Kuala Lumpur, Selangor, Melaka, and Negeri Sembilan etc. [24]. During 2005-2008, due to the excessive trees or vegetation contacting with HV power lines, there are many interruptions of power lines have been mentioned in Sarawak in East Malaysia [25]. The Figure 2.2 [26], shows that the third biggest cause of electricity supply interruption is due to tree encroachments, although the other causes of interruption have been reported i.e.; installation fault , quality of work and maintenance quality is very poor. These factors also contributing the losses of power supply energy to the consumer but the major portion is the tree encroachments as highlighted in Figure 2.2.

There are many natural disasters are occurring in Malaysia, these disasters including in flood, wind, landslides, and thunderstorms. Due to these disasters may damage the power supply interrupting more than the trees encroachments. The compulsory by ethically and law, the power supply companies to monitor the power installation for continuing supply to the customers. To find the fault, the Remote and on-line monitoring may help to provide better surveillance information to locate the fault before time. This is the concise and better facility for the continuation of the power supply to the customers with less wastage time.

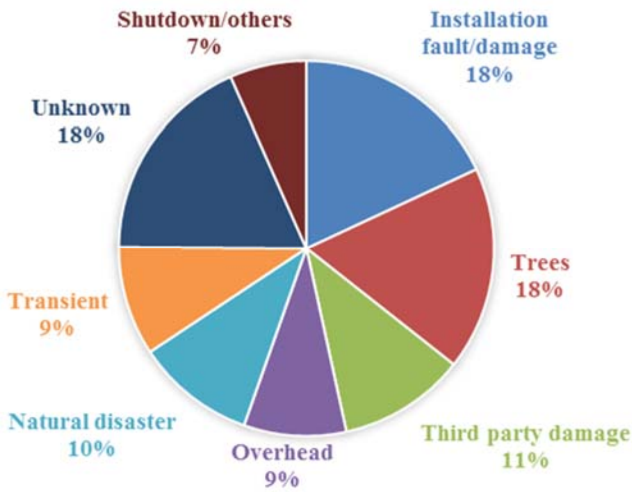


Figure 2.2: Causes of electricity supply interruptions in Sarawak (East Malaysia) [26].

The highest supply interruption factors are indicated in Figure 2.3, 70% interruption of total power supply is due to the overhead of power lines excluding high voltage supply interruptions. During 2005 to 2008, the power supply interruption increase due to an increase of time as shown in Figure 2.4. Therefore in order to ensure the reliability of power supply system for the protection of overhead line besides endangering encroachments. The vegetation encroachments involve disruption of power supply needs to be trimmed/cleared in the right of ways that may damage the power supply. The Figure 2.5 shows that taken snapshots of different location in Sarawak, East Malaysia. The trees or dense vegetation surrounding the HV transmission lines across the road, center of the juggles and non-uniform terrain. Therefore, it is important to ensure reliability in the power systems for the safety of overhead line against endangering encroachments.

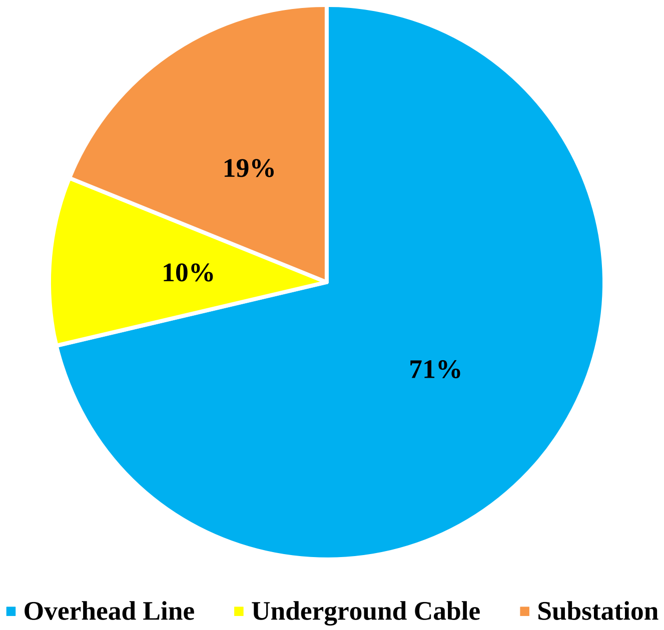


Figure 2.3: Percentage of a total number of supply interruptions in Peninsular Malaysia based on the component of the network [26].

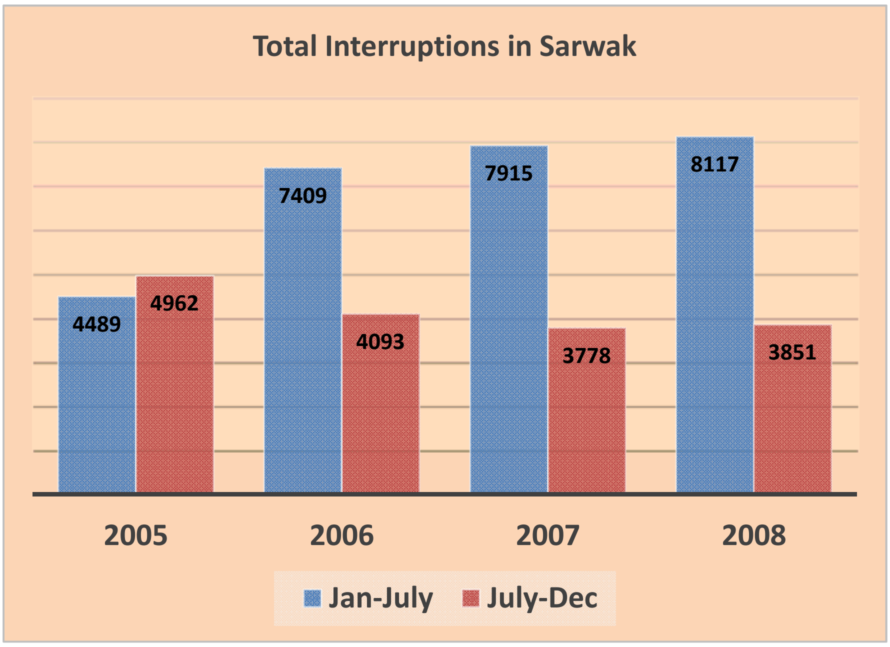


Figure 2.4: Total Interruptions in Sarawak in the year 2005, 2006, 2007 and 2008 due vegetation and tree encroachments near power lines and poles [25].

(a)

(b)

(c)

(d)

(e)

(f)

Figure 2.5: (a) Overhead lines passing through fields in Ipoh. (b) The encroachments due to uneven dispersion of power lines. (c) Overhead lines passing through a non-uniform terrain with dangerous trees ahead near Ipoh. (d) The encroachment due to uneven dispersion of power lines. (e) Overhead lines in the dense jungle outside Kuching in Sarawak. (f) Overhead lines in the dense residential area in Sarawak [25].

Vegetation maintenance involves trimming / removing of the trees that may come in detestable contact with overhead lines and cause service disruption. All trees and bushes that could potentially interfere with the conductors are trimmed to a specified clearance. Figure 2.5 highlights some of the snapshots, taken from various locations in Ipoh (West Malaysia) and Sarawak (East Malaysia), presenting dense vegetation surrounding the HV lines.

2.2.5 HV Overhead Line Clearance in Malaysia

In order to ensure the safety of the power system as well as the people living nearby HV transmission lines, Malaysian authorities enacted the Electricity Supply Act 1990 and the Electricity Regulations 1994. As per Malaysian regulations, a minimum clearance of the lowest conductor from the ground for HV supply lines is given in Table 2. 1 [1]. Tenaga Nasional Berhad (TNB), the electrical utility working in Malaysia has also outlined safety parameters [27]. The salient features of these are:

- The minimum distance between the two power lines having less rating (15 kV, 22 kV, and 33 kV) should be 90 m.

- The minimum distance between highly rated transmission lines (275 kV, 500 kV, and 800 kV) must be 400 m.

- The minimum clearance of any building or structure, or when using, transporting or moving any long object such as a pole, ladder etc. from high voltage supply must not be less than 4.57 m.

- Ensure that trees that grow in the vicinity of the HV lines to be less than 8 feet or 2.4 m in height.

- Ensure that objects or materials placed under the HV supply are less than 8 feet or 2.4 m.

These measurement provided the condition to measure the threats occur due to vegetation / trees near or under the power transmission lines and power poles.

Table 2. 1: Minimum clearance of the lowest conductor from the ground for HV supply lines in Malaysia [27].

Operating Voltage of the transmission lines	Clearance from ground (meter)		
	Over Roads	Other than roads	In position to inaccessible to vehicular
11 kV 22 kV 33 kV	5.7	5.4	4.8
66 kV	6.1	6.1	5.1
132 kV	6.7	6.7	5.7
275 kV	7	7	7
500 kV	7.3	7.3	7.3

According to the laws and regulations of TNB mentioned above, the standard distance between the HV transmission poles that carry heavy electrical power from power generation units to the power stations installed between the cities must be 400 m. Similarly, the power poles that are installed along the roads inside the cities that are responsible for the supply of electricity to residential areas must be 90 m apart from one another. The minimum distance of any residential structure from the high voltage supply lines passing within the cities must not be less than 4.57 min order to prevent the power lines from striking the tall building structures that are above the power lines, and the trees that grow under the power lines must not be above a certain level to avoid potential encroachment.

2.3 Methods for Inspecting HV Overhead Power Lines ROWs

As the HV transmission lines are connected in cascading manner, the outages in any transmission part of the ROWs can cause damage throughout the cascading network [28]. The events of cascading outages have increased over the past few years [7] and encroachment is one of the main reasons. It is, therefore, necessary for electric utilities to inspect their overhead transmission lines at regular intervals for both business and legal reasons. Following are the traditional methods of inspection of HV transmission lines:

1. Field Survey by using Vehicles.
2. Aerial Video Surveillance.
3. Aerial Multispectral Imaging.
4. Airborne LiDAR Scanning.
5. Satellite Stereo Images
6. Unmanned Aerial Vehicle (UAV)

2.3.1 Field Survey by using Vehicles

One of the conventional practices involves field survey or ground inspection whereby a team is deputed to visually inspect the conditions of the overhead power lines either by pole climbing, foot patrolling or inspection using a vehicle [29]. For pole-climbing, linemen climb halfway up a pole and hold a mirror along insulated stick to inspect installation parts. Foot patrols are carried out by a team of two who walk along a section of the network while making visual inspections. Linesmen are equipped with computers that contain information about the equipment at that particular pole. Information regarding the state of the line, summary and prioritized list of items for repair is keyed in [7]. The linemen, thus after the visual inspection of transmission line components (such as corridors, insulators, and Feeders etc.) and vegetation surrounding the lines, order the trimming of trees that endanger the HV lines ROWs as in Figure 2.6 (a) [30]. Unfortunately, this method is costly because a suitable amount of professional and a well-trained team (that should be able to identify hazards from HV transmission lines) equipped with PC laptops, geographical data, and maps etc., for inspection requires a healthy amount of budget along with the vehicle and trained peoples for vegetation encroachment inspection near power lines and poles as in Figure 2.6 (b).

(a) (b)

Figure 2.6: (a) Linemen climbing up on the HV lines to inspect the essential components such as corridors, insulators, and feeders. (b) Visual inspection of vegetation in the way of transmission lines ROWs using vehicles [30].

This method is also time consuming as it requires a suitable amount of time depending upon the length of transmission distribution network ROWs, and also less accurate due to fact that human being may make judgmental errors especially to encroachments that appear to be in safe clearances but dangerous during bad weather conditions e-g., heavy rain, wind, and storm etc. It is also not suitable in a hilly area where a human cannot go and cut the trees or vegetation.

2.3.2 Aerial Video Surveillance

Helicopters and airplanes having video surveillance cameras or imaging sensors mounted on them are useful for obtaining a video film or sequence of images from a wide range of area as shown in Figure 2.7 (a) [1]. Video surveillance or imaging by helicopters is mostly employed in the regions where it is difficult for a person to reach or where the area covering HV lines is too wide to inspect using vehicles. The traditional method for aerial video surveillance consists of an aircraft or helicopter that flies slowly or have frequently at 50~60 miles per hour (mph) with binoculars and surveillance cameras, and record data in the log book from mission start to end and stays 30~50m away from the HV lines and structures. Various methods of aerial inspection of overhead lines are reported by Lili Ma et al. in [31]. An image surveillance and inspection system (ISIS) has been developed by aerospace precision imaging at Utah State University Research Foundation to inspect power lines. Using ISIS, capturing and storing high-resolution images can be done while the helicopter flies up to 300 feet away from structures. Electric Power Research Institute developed a system called Airborne Inspection System (AIS) that can produce images and videos that reveal small objects, as small as a broken insulator while the helicopter is flying at 70 mph to 90 mph and depending upon the application, information collected during the airborne surveillance includes data regarding:

- The equipment (including feeders, insulators, and pole and tower conditions).
- Thermal properties of HV line conductors.
- Vegetation along with other potential intrusions ROWs, and other abnormal conditions.

A manned helicopter for the inspection of HV lines applications mostly consists of:

- Geo-map of the target locations.
- An active vision system (digital surveillance camera) for the acquisition of high-resolution videos or still images.
- Sensors for capturing temperature and pressure conditions.

In [32], the cost of aerial visual inspection of power lines using helicopters is discussed. A bigger aircraft will add 15% - 20% to the cost that is needed to inspect and record mechanical defects or dangerous trees near HV lines. It is also outlined that factors affecting the cost of aerial inspection depend on the type of conditions including the location of transmission and distribution lines, weather conditions, and other different problems. The aerial survey can cover a large area but it is not reliable for non-uniform terrains covering distribution networks as in Figure 2.7 (b). It is however also ambiguous due to distortion of surveillance data by the random movement of airborne. It is also not reliable for non-uniform terrain covering distribution network.

(a)

(b)

Figure 2.7: (a) Video Surveillance using a helicopter. (b) Non-uniform terrain [32].

2.3.3 Aerial Multispectral Imaging

Aerial multispectral imaging is a remote sensing technology that captures images at certain frequencies along with the Electromagnetic (EM) spectrum that is separable by filters or by the use of instruments that are sensitive to particular wavelengths. It is widely used for the agriculture problems and so for encroachment monitoring. The progress in this field took place in the 1980s where studies involved cameras that possess sensitivities to EM spectrum range regarding the near-infrared light (700 up to 1100 nm). The multispectral cameras are mostly used to take aerial images. Helicopters, aircraft, and balloons are used to take the aerial images of vegetation endangering the HV from highly elevated positions [4] as shown in Figure 2.8. In the past, aerial images were not being used for power line surveys because power lines were too thin in them. However, with the advancement in image resolution technology, it became possible to use aerial images for inspection of HV overhead power lines.

Digitized multispectral aerial photos by means of stereovision imaging [6], [14], [33] are used to estimate the height of the encroached trees. Stereovision can yield a depth of the scene similar to the human vision system. It takes images of a scene from two cameras, or two different locations, to produce the disparity map that can easily be converted into a respective 3D-depth map. Stereo vision is the ability to infer information on the 3D structure from two or more images captured from different field of view [34]. A stereo pair of images is an essential source for extracting depth. The similar stereo images are used to give vision to automated vehicles and 3D-Digital Elevation Model (DEM) from airborne aerial images.

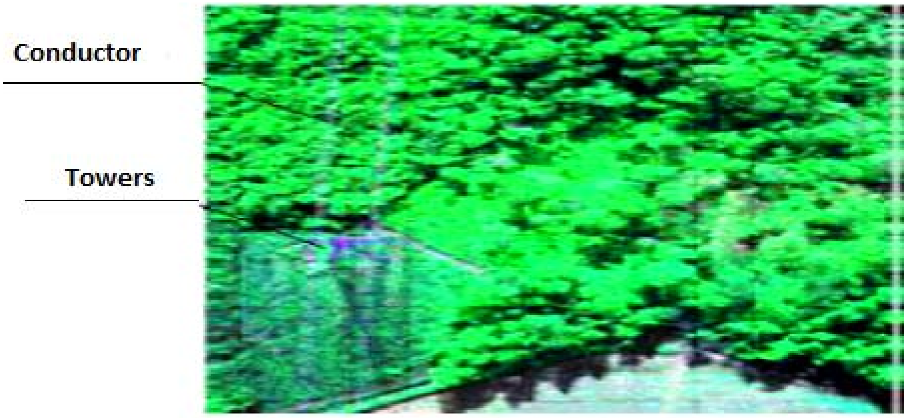


Figure 2.8: Image through Multispectral Imaging [4].

In general, the stereo matching is classified into Local and Global techniques [35]. Local techniques utilize the intensity or color values for each pixel to determine the disparity. Whereas, Global techniques determine the disparity by applying energy minimization techniques such as dynamic programming, graph-cuts, and scan-line optimization [36],[37]. Once the disparity map is evaluated it becomes easy to extract the DEM of a scene. C. Sun et al. in [6] have presented a technique that automatically measures the distance between the HV lines and trees. They have used an airborne equipped with stereo cameras pairs and flying at an altitude 80 m above the ground. The distance between the stereo pairs is 10 m. Thus, the steps followed to calculate the distance between the catenary envelope and trees are:

- Matching individual stereo pairs to obtain the disparity maps.
- Identification of HV poles and matching them in stereo pairs.
- Computing the 3D information of the power poles from top to bottom of the scene using disparity maps.
- Calculating the catenary envelope of HV lines using the 3D information of the poles within the span.

- Generating the Digital Elevation Model (DEM) model of each stereo pair, and mosaicking them into a 3D pole-to-pole image.

- Measure the approximate distance between the catenary envelope and the tree boundaries.

Similarly, [14] and [33] also recover DEM of the scene to identify vegetation endangering HV lines ROWs. However, aerial images do not show the accurate features on the earth due to the displacement caused by the tilt of the sensor and the terrain. To overcome this, aerial images should be orthorectified first to remove the spatial ambiguities. In order to orthorectify the acquired images, some transform models are required which takes into the account various sources of image distortions generated at the time of image acquisition. These include:

- Sensor orientation.

- Errors associated with the sensor.

- Topographic & Atmospheric corrections.

- Earth shape and rotation.

- Sensor altitude variations.

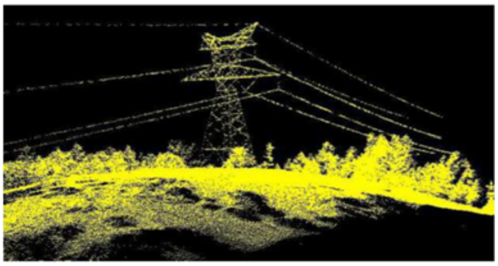
The ambiguities in aerial images are removed by passing the captured images from orthorectification software packages built-in with the above mentioned mathematical models. Therefore, aerial multispectral imaging has proved to be more accurate as compared to visual field inspection and aerial video surveillance as it allows the capturing of aerial images with an appropriate spatial resolution depending on the altitude of aircraft. A very comprehensive research conducted by S. J. Mills et al. in [33] has shown in detail that aerial stereoscopy is somehow not very much accurate due to sensor parameters, orthorectification issues, and atmospheric variations (weathers). Also, because of the lower flying altitude (< 500 m), it is difficult to maintain aircraft above the HV lines, which results in time consumption for a large area. For larger area, more time required to deploy the method for vegetation surveillance near power lines.

2.3.4 Airborne LiDAR Scanning

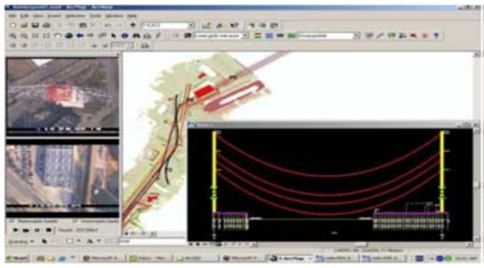
LiDAR (Light Detection And Ranging) is an optical remote sensing technique that measures the property of scattered light to find the range or other information of the distant target. It uses time delay between the transmitted and the reflected signal similar to the RADAR technology. LiDAR uses much shorter wavelength typically ranging from 10.6 μm to the Ultraviolet (250 nm). Light-weight, compact and low power LiDAR is mounted on the airborne vehicle to determine the distance between the HV power lines and trees [11]. The 3D urban environment is complex due to different heights and nature of buildings, trees, roads and commercial areas passing over one another. LiDAR data is increasingly used for three-dimensional (3D) urban feature detection [38] because of its accuracy in height and range measurements. This technique is, therefore, applied to monitor the dense vegetation encroachment under the HV transmission lines ranging from 33 kV to 500 kV. The data collected by the system is processed by the software running on the personal computer. Airborne mounted with LiDAR has four components [48] that are given below:

- Differential Global Positioning System (DGPS) to locate the aircraft in space.
- Inertial Navigation System (INS) to identify the orientation of aircraft in space.
- LiDAR to estimate the distance between aircraft and objects below.
- PC (laptop) interfaces with Global Information System (GIS) to obtain the accurate coordinates of the scenes and to visualize the 3D information.

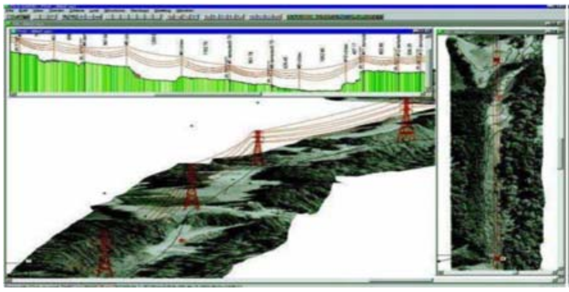
In general, by using the LiDAR, an average of about 100 km of data is collected per day and no permission from the landowners is needed. The advantage of LiDAR in comparison with photogrammetry is that in combination with topographical data, the 3D catenary model of the transmission lines as discussed above can be extracted from the data. Although millions of laser points can be collected within a single kilometer, the data is useless if they are not processed in a user-friendly format. In general, the data is processed in AutoCAD, CAD files or other software packages like PLS-CADD, ESRI, and Small World etc., which are available by different software vendor companies.



(a)



(b)



(c)

Figure 2.9: (a) Laser point cloud raw data (image courtesy of Fugro FLI-MAP). (b) Integrated LiDAR data on GIS. (c) LiDAR data with PLS-CADD (images courtesy of Fugro FLI-MAP and Powerline systems) [39].

Figure 2.9 [39] shows that after having laser information, the catenary model of the transmission line can be mapped accurately. Once the lines and its surroundings including the ground levels, vegetation and buildings have been modeled, analysis of different forms, such as vegetation and ground clearances, electromagnetic interference levels in housing and thermal rating studies, can be performed quickly. LiDAR survey has proved to be reliable for there is a minimum risk of missing information. A quantitative comparison [40] (based on performance parameters like RMSE, variance etc.,) showed LiDAR scanning to be more accurate than the aerial stereoscopy imaging for vegetation management of power line corridor ROWs. However, owing to occasional uneven hovering of air-borne, un-calibrated camera parameters and the curvature of the earth may cause ambiguities in the data recorded by LiDAR and consequently the software used for 3D tracking of the transmission line can produce an ambiguous model of the scene.

2.3.5 Satellite Stereo Images

Jardini [4] maintain the database or information system for the vegetation oversight of transmission power lines corridor and compare the performance of satellite images, aerophotogranometry, videography and airborne laser scanning. It has been proved that satellite images are cost-effective and more accurate as compared to other sensors [4]. Mattias S. Moeller [40] was used satellite stereo imagery to monitor power line corridors and he was used QuickBird has a resolution of 0.6 ground sample distance. Yoshihiro Kobayashi introduces an innovative concept for identification of vegetation near power lines using satellite stereo images but he did not provide any quantitative analysis. He generated Digital Elevation model for identifying and measuring the height of vegetation using satellite stereo images. DEM has been used cross-correlation method which belongs to local area stereo matching method. Using multispectral stereo images, the height of each pixel is determined and the distance from the conductor is calculated. When this distance is less than the endangerment zone surrounding the conductors, the determinate object/tree is pretended to menace the line [5].

Junaid Ahmad [1] suggest a concept of satellite stereo images to estimate the 3D depth of vegetation/trees that can hammer power lines to damage the lines and result may cause of blackouts. He has used the region based stereo algorithm to minimize the energy function, but he did not apply the stereo algorithm on a pair of images and show no experiment results [1]. A series of research projects by Yasuoka [41] shows techniques to estimate the number of trees from satellite images and LiDAR data. The concept of the utilization of satellite images for the cited task offers potential improvements over direct visual inspection and airplane-based (e.g., LiDAR) technologies:

- Wide area of coverage;
- Frequent overhead passes;

• Areas with restricted access may be viewed;

• Potentially lower cost because of processing a large geographical area at one time in an automated workflow.

An important issue relating to the use of satellite images for the identification of any ground features relates to the utilization of stereo pair images. Processing of stereo images allows the user to extract height information for the observed terrain. Because the satellite is effectively infinitely far from the ground features, it is necessary to access two separate images recorded utilizing slightly different angles. The Figure 2.10 shows satellite stereo scanning concept. Indirect visual observation of transmission rights of way, the relative position of trees and other objects to the transmission line is resolved by the interpreter's observation. However, it appears that the wide area coverage of satellite images (e.g., in comparison with direct visual observation) may more than compensate for the need for stereo images. The cost to achieve an acceptable level of resolution is an unresolved issue, because of the high cost of stereo, multispectral satellite images. However, the wide coverage, the fast processing of the satellite images and the expected decrease of image cost by the increase of demand offer the potential of lower overall cost especially in a large volume purchase of images [42].

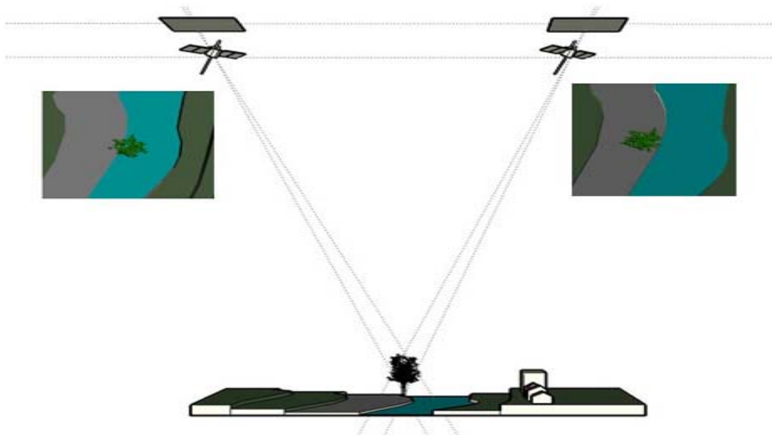


Figure 2.10: Pictorial of stereo satellite imagery [43].

At present, there are several main civilian sources and types of satellite images. For extracting ground features in remote sensing, the accuracy depends on the ground sample distance (GSD) value. GSD is the spatial resolution of imagery and represented as a distance of metres per pixel. According to [44], [45], the images of one meter GSD are suitable for extracting ground feature, and IKONOS, QuickBird and OrbView satellites produce images commercially. The IKONOS satellite is the world's first commercial satellite to collect panchromatic images with one meter GSD and multi-spectral imagery with 4 meter GSD. IKONOS was launched in September 1999 and started providing imagery in January 2000. The images are taken at the satellite altitude of 680 km. The accuracy of the orthorectified image is 1.75 m. It takes 11 days to return to the same location (revisit cycle). QuickBird is a high-resolution satellite owned and operated by DigitalGlobe. Using state-of-the-art technology, Ball's Global Imaging System 2000 sensor, QuickBird uses remote sensing to 0.61 m GSD resolution. QuickBird was launched in October 2001 and acquires images at the satellite altitude of 450 km. The satellite returns in 1.0 to 3.5 days. According to [43], [46], a stereo pair of QuickBird images is readily used to extract height of buildings. OrbView also offers one of the high-resolution satellite images. The original OrbView-1 was launched in 1995, and the satellite was placed in service in 2000. OrbView-2 and Orb-View-3 were launched in 1997 and in 2001. The OrbView-3 takes 1 m panchromatic and 4 m multispectral images. The next generation, OrbView-5 was launched in 2008, and it offers unprecedented special resolution by simultaneously acquiring 0.41 m panchromatic and 1.64 m multispectral imagery.

The OrbView-5 images are the highest resolution satellite images available to civilians. Interestingly the conductor is visible in many satellite images even though the diameter of the conductor is well below the image resolution [43]. This visibility is due to the optical effect of averaging and interpolation across the adjacent pixels. There are many formats for digital images such as JPG, GIF, TIFF, and RAW. Remote sensing techniques, especially for the detection of vegetation, commonly use TIFF format (Tagged Image File Format) because it is able to save four bands: red (R), green (G), blue (B), and near-infrared (NIR) bands in one file. The R, G, B, (RGB) and NIR bands are wavelengths of the electromagnetic spectrum/energy. The R band has a wavelength in the range of about 0.6 to 0.7 m; the G band from 0.5 to 0.6 m, the B band from 0.4 to 0.5 m, and the NIR band from 0.7 to 1.2 m. The RGB bands are visible to the human eye, but the NIR is invisible. The intensity of the spectrum at each point of the TIFF image is represented as a set of four band values. The values are usually digitized as an

8-bit integer value in the range 0 to 255, and this representation is denominated as the digital number (DN). The satellite image providers offer TIFF image file with R, G, B, and NIR bands as a multispectral satellite image [42].

The availability of repeatedly updated data makes it possible to develop a territory monitoring system that not only acts as a supporting design decision tool but also as a way to verify design strategies in case of unforeseeable events. IKONOS® Pro 1-meter and 4-meter products are perfect for projects requiring high-resolution imagery and positional accuracy when ground control may be costly, difficult, or impossible to acquire. Providing a strong base for three-dimensional feature recognition, extraction, and exploitation, the product provides two images with stereo geometry to support a wide range of stereo imagery applications such as DEM (Digital Elevation Model) creation and three-dimensional feature extraction. Stereo products in epipolar or map projections provide RPC (Rational Polynomial Coefficients) camera model data. The new very high spatial resolution satellite images, with a ground pixel size of 0.6 m or 1 m such as QuickBird and IKONOS, open new possibilities for cartographic applications [47]. These DSMs (Digital Surface Model) are generated by stereo matching from very high resolution (VHR) satellite stereo-pairs imagery, reconstructing the 3D surface corresponding to the first surface view of the earth containing both micro-relief (buildings, trees and so on) and bare terrain [42]. Existing very high-resolution optical satellite sensors (≤ 1m GSD for pan) [48] Pan: Panchromatic, ms: Multispectral, GSD: Ground sample distance are shown in Table 2. 2.

Pleiades satellite has been launched on 17 December 2011 and within one year another Pleiades satellite sensor has been launched on December 2, 2012 [49]. The Pleiades consists of high-resolution optical sensors with 0.7-meter panchromatic resolution and 2.8 multispectral resolution capability, which covers the swath about 20 km. It provides low-cost solution with minimum area scan 100 km and has low weight [50]. The main characteristics of Pleiades satellite are that stereo data can be attained during one pass using forward and backward sensor capability as compared to the other very high-resolution satellite sensors, namely IKONOS, QuickBird, GeoEye and Worldview [51],[39]. One of the exciting features of Pleiades is that it can obtain three stereo images, in a single pass, in the same orbit [52].

Table 2. 2: The satellite sensors with their properties

Sensor	**launch**	**Altitude [km]**	**GSD pan [m]**	**Swath in nadir view**	**Pan/ms channels**	**Imaging capacity [km²/day]**
IKONOS 2	1999	681	0.82	11.3 km	Pan, 4ms	150 000
QuickBird	2001	450	0.61	16.5 km	Pan, 4ms	135 000
KOMPSAT-2	2006	685	1.0	15 km	Pan, 4ms	----------
WorldView-1	2007	494	0.45	17.6 km	Pan	750 000
WorldView-2	2009	770	0.46	16.4 km	Pan, 8ms	975 000
GeoEye 1	2008	681	0.41	15.2 km	Pan, 4ms	700 000
Cartosat-2, 2A, 2B	2007-10	631	0.82	9.6 km	Pan	528 000
Pleiades 1	2011	694	0.50	20 km	Pan, 4ms	1000 000
Pleiades 2	2012	694	0.50	20 km	Pan, 4ms	1000 000
WorldView-3	2014	620	0.31	13.2 km	Pan, 8ms	676 000
DMC-(3satellites)	2014	630	1.0	22.6 km	Pan, 4ms	100 000

The satellite sensors of Pleiades, 1A, and 1B, are almost identical in their features, however, Pleiades 1A has more accuracy according to a geometric model based on rational polynomial coefficients [53][54]. In photogrammetry, one of the fundamental parameters, known as baseline ratio, is the ratio between the distance of flight height and two separate acquisitions. In order to measure a good 3D height, there is a need for low baseline ratio value, which incorporates low occlusion. The base to height ratio has a range between 0.7 to 1 for good 3D mapping applications [55]. On a theoretical point of view, the larger is the B/H ratio, better is the stereoscopy and the height estimation. On the other hand, images acquired from very different viewing angles contain certain disadvantages. First of all, they show occlusions in urban areas, occlusions due to buildings. Therefore, corridors between buildings cannot be modeled properly. In addition, large angles produce longer shadows, and therefore low-texture homogeneous areas where image matching does not perform well. The accurately dual-purpose optical observation system of Pleiades has large applications in agriculture, resource management, geology survey, forestry mapping, and risk management for commercial, scientific and institutional customers [1]. It also imparts a large field of view in a long track and across track scenarios [56]. In photogrammetry, one of the fundamental parameters, known as baseline ratio, is the ratio between the distance of flight height and two separate acquisitions. In order to measure a good 3D height, low baseline ratio value is needed, which incorporates low occlusion. The base to height ratio has a range between 0.7 to 1 for good 3D mapping applications [57]. It is decided to use two aerial technologies for data collection. First, the data has been collected through satellite and then from UAS (Unmanned Aircraft Systems).

2.3.6 Unmanned Aerial Vehicle (UAV)

Technological progress constantly offers significant reductions in size and costs of GPS devices, processing computers and inertial navigation sensors (INS) that open gateways for innovative remote sensing applications [58]. Currently, such commercially available devices include miniaturized models for unmanned aerial vehicles (UAVs) [5]. Civilian UAV applications have dramatically increased due to cost reductions and miniaturization of craft and GPS devices as well as sensors and processing hardware [59]. Furthermore, with the development of robust and smaller autonomous sensors [12], UAVs have quickly evolved as stand-alone systems that provide required data with higher temporal and spatial resolution. UAVs equipped with remote sensing devices can now acquire spatial geographic coverage, etc., with the advantage of high-resolution spatiotemporal images [12]. These benefits attract much attention in terms of research and manufacturing potential. UAVs can also monitor power stations, power transmission lines, and power grids. Effective mapping developments with improved precision also create vegetation maps for an important application that provides highly detailed topology [60]. In addition, even detailed types of vegetation groupings and tree level topography can now be generated.

Therefore, UAV systems equipped with remote sensing devices are now used in forest resource management as well as for vegetation, river and power line monitoring.

For example, acquired sub-decimeter resolution images are able to capture the finer details of ground objects for extremely precise urban vegetation mapping. Thus, compared to piloted airborne platforms [61], UAVs provide safer data acquisition systems with minimized overhead. Furthermore, UAVs can fly at much lower altitudes than piloted airborne systems, which results in exceptionally high spatial resolution imagery [62]. Thus, UAV has become the better alternative for monitoring vegetation and /or trees near high-voltage power transmission lines and poles. The Figure 2.11 shows the UAV equipment and software used for controlling the UAV.

(a)

(b)

Figure 2.11: UAV used for data collection of the area of interest: (a) UAV sensor, (b) the controlling software that handles the UAV [12].

2.3.7 Comparison of Traditional Vegetation Inspection Techniques

The distributed transmission lines located inside a city tend to be more difficult to inspect via air-borne methods and a good view is often blocked by the trees. The second most difficult is sub-transmission lines because they are mostly erected with many turns, unlike transmission lines that are built on a straight path and have fewer connections, switches etc [7]. A typical comparison of conventional methods of inspections in terms of accuracy, time consumption, and cost assessment is given in Table 2. 3. The data shown in the table is based on monitoring of a distribution network comprising one thousand HV transmission poles ROWs. It is evident from the Table 2.

3 that the time of monitoring by field survey is comparatively higher. The accuracy obtained by field inspection is lower due to judgmental errors by inspectors in locating the encroachments that appear to be safer but can interfere or strike the HV line during bad weather conditions. However, the cost of field inspection appears to be moderate. Airborne surveillance techniques are comparatively expensive due to the higher cost of equipment and its maintenance. Lower accuracy of airborne inspection is attributed to the distortion of surveillance data by uneven hovering of airborne, its inaccessibility to the scene and non-uniform terrains covering HV lines.

Table 2. 3: Methods for monitoring vegetation for transmission lines.

Method	**Approximate Time (days)**	**Accuracy (100%)**	**Approximate Cost (RM)**	**Monitoring Area (Km)**
Field Survey [1]	7	40	20000	100
Airborne [63]	2-3	35	25000	100
Aerial Multispectral Imaging [64]	5-6	40	25000	100
LiDAR [65][10]	2-3	88	90000	100
Satellite Stereo Images [35]	2-3	87	15000	100

The aerial multispectral imaging uses the airborne method but acquires multispectral stereoscopy data at lower altitudes which increases the accuracy to a moderate level due to airborne stability and orthorectification issues as discussed above. However, because of less area coverage due to low altitudes, this method of inspection consumes more time. Whereas, the cost would be similar to airborne inspection method as in Table 2. 3. LiDAR uses an airborne equipped with laser scanner components having much higher cost than rest of the methods. Presently, U.S. is spending about $2 billion annually on using LiDAR and image mapping surveys of power lines for vegetation management [66]. However, LiDAR data is modeled in 3D which distinguishes HV transmission lines from trees and buildings using various software packages which make its accuracy higher (about 80%) by utilizing the same amount of time as traditional airborne technique. Based on technologies issue, the satellite stereo images can provide cost-effective solution to monitoring the vegetation near power lines and it provides a less costly solution as compared to other airborne and foot patrolling methods [66].

2.4 Classification of Stereo Matching Methods

There are two broad classes of stereo methods first one is the Local matching methods [67] and second is the Global matching methods [68] based on dynamic programming [69], [70], graph cut [71],[15],[72], belief propagation [71],[35] and cooperative algorithm [73], [74] as shown in Figure 2.12. The local methods are further divided into the area based and feature based. The area-based methods accuracy depends upon the size of the windows like a fixed window, adaptive window, and shiftable window. Local methods aggregate the matching cost of all pixels lying in a support window surrounding a point q and assign the disparity to q those results in the lowest aggregated cost. These methods assume constant disparity over the support, which does not hold at disparity discontinuities and leads to boundary fattening artifacts. Explicit occlusion detection [73], multiple windowing [75], or adaptive local support weighting reduce this effect, but cannot avoid it completely [76]. The optimal choice of window size depends on the local amount of variation in texture and disparity [77],[78], [79]. Feature-based algorithms extract the features of interest from images (edge, corner, and line) and match them in two or more views. Usually, these methods only yield very sparse depth maps [80].

The global method uses the Markov Random Field (MRF) stereo which changes the specific probe lint energy function and design of the energy function based on the Bayes rule is the very active topic in computer vision [72]. The energy function is computational expensive that’s why the global method uses the energy functions are not real implement. The global method for dense stereo matching produced very good accuracy in discontinuity area as compared the local methods. There are many algorithms that lie under the umbrella of global optimization algorithms these are dynamic programming, cooperative algorithm, graph cut, and belief propagation. These algorithms can be applied to satellite stereo vision and try to find good disparity map for measuring the 3D depth of vegetation near power poles. This is the novel idea for calculation the 3D depth of vegetation using satellite.

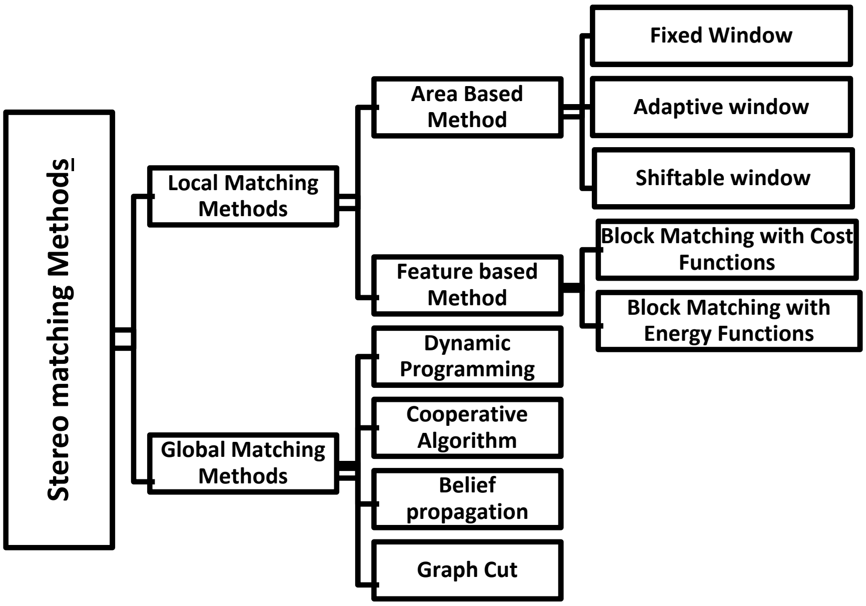


Figure 2.12: Existing Stereo Matching Algorithms.

2.4.1 Local Stereo Matching Algorithms

The most important and useful algorithms exist in literature are the local stereo matching algorithms. The local stereo matching algorithms computes the cost function based on fixed and adaptive windows pixel by pixel shifting of each window scan the complete image and store displaced values giving by each window indexed value. The matching of the window is based on block matching, or pixel by pixel matching to produce disparity map using stereo matching algorithms. The details of local methods with a cost function and local methods with energy minimization using block matching technique have been provided in Appendix A. The local stereo matching algorithms are very fast and provide efficient solution in real time implementation. However, these algorithms could not provide better accuracy for stereo matching process in aerial images and the solution may not feasible where better accuracy required.

2.4.2 Global Stereo Matching Algorithms

The global stereo matching algorithms provide the cost functions to compute the disparity map for stereo images locally and apply minimization energy functions globally to further enhance the accuracy of the disparity map. There are many global algorithms; dynamic programming, graph cut, belief propagation, cooperative message passing [76]. The detail explanation of each algorithm has been provided in this section. The dynamic programming and graph cut algorithms have produced more accuracy as compared to others global stereo matching algorithms

2.4.2.1 Dynamic Programming

It can produce best disparity maps using scanline search [81],[82],[69]. It gives the good correspondence between left and right images. The dynamic programming searches the best correspondence points between left and right stereo images and must enforce the ordering constraint between scan lines of both stereo images. The dynamic programming gives a smoothness disparity map due to strong correspondence between left and right image. Dynamic programming finds the minimum path from the top left corner of the 2D matrix to the right-bottom corner [83]. It finds the optimal path and matches the sequence in left scan-line in left image to the right scan-line in the right image optimally. If the ordering constraint is assumed, the best path can be computed to match the pixels belonging to the left image and the right image, so the dynamic programming provides the best path on the 2D grid that satisfies the ordering constraints [84]. The detailed analysis of dynamic programming algorithm with mathematical derivation has been given in Appendix A.

2.4.2.2 Stereo Matching based on Graph-Cut Algorithm

Stereo is a classical vision problem, where graph based energy minimization method has been successfully applied [85][86]. The three basic graph-based methods are used to solve stereo correspondence problems are pixel labeling with the Potts model, stereo with occlusion and multicamera scene reconstruction [86]. Boykov introduced similar work based on energy minimization using expansion move algorithm. This algorithm minimizes the energy function in an iterative manner [87]. It minimizes energy function by transforming it into a minimum cut problem on the graph and then cut the graph at each iteration to solve such problem. The algorithm is run until the convergence occurred and the result is a pretty strong local minimum of the energy function. The detailed analysis of graph cut algorithm with energy minimization technique using potts model is given in Appendix A.

2.4.2.3 Stereo Matching based on Belief Propagation (BP) Algorithm

The magic of the BP algorithm lies in its powerful message passing [59]. A message presents the probability that the receiver should be at a disparity according to all information from the sender up to the current iteration. Message passing has two important properties. First, it is asymmetric. The entropy of the messages from high-confidence nodes to low-confidence nodes is smaller than the entropy of the messages from low-confidence nodes to high confidence nodes. Second, it is adaptive. The influence of a message between a pair of nodes with larger divergence would be weakened more [71]. Therefore, BP's message passing provides a time-varying adaptive smoothing mechanism for stereo matching to deal with textureless regions and depth discontinuities naturally. The global methods based on belief propagation algorithm as the ones proposed [88] result in disparity maps which exhibit the best quality. In sparse stereo, distinct image points are usually extracted. Then, the corresponding pairs are matched using a local based approach. The main problem of such algorithms is that the smoothness constraint cannot be enforced in the computations which will lead to some unwanted disparity jumps along the surface of an object. In addition, such methods have tendencies for errors at the object boundaries.

2.4.2.4 Stereo Matching Based on Cooperative Algorithms

Marr and Poggio [89] presented two basic assumptions for a stereo vision algorithm. The first assumption states that at most a single unique match exists for each pixel that links each pixel to a single surface point. When using intensity values for matching this uniqueness assumption may be violated if surfaces are not opaque. A classic example is a pixel receiving a contribution from both a fish and a fish bowl. The second assumption states that disparity values are generally continuous, i.e., smooth within a local neighborhood. In most scenes, the continuity assumption is valid since surfaces are relatively smooth and discontinuities occur only at object boundaries.

2.5 Limitations in Stereo Matching Algorithms

There are some limitations on different stereo matching algorithms and their applications as shown in Table 2. 4. The limitation of each stereo matching algorithm depends on the orthorectification of stereo images and illumination and pre-processing of each stereo image. The area-based, feature-based and global stereo matching algorithms have some limitations as explained in Table 2. 4. The area-based methods are very slow and cannot deal the images with a sharp matching point in stereo images. They are very sensitive due to noise and varying illumination. The main advantage of these algorithms, they perform well with smooth and continuous surface stereo images planes. Feature-based matching: These features have suitable only for the small distance between two stereo cameras. The feature-based methods need ordering and uniqueness of pixels between left and right image. The feature may not match due noise in the image. They need high corresponding distinct points between two images. These limitations addressed to implement feature-based methods. Global based stereo matching methods have a main disadvantage, they don't perform well in textureless region and depth discontinues surface. They have a high computational cost especially graph cut algorithm. These methods provide high accuracy as compared to area-based method and well implemented in stereo matching literature for disparity map calculation [81]. The next section will provide the introduction of sparse representation and their applications in image and pattern recognition field. The most effective and important algorithm of deep learning is CNN (convolutional neural network). The introduction and related work based on stereo matching using CNN algorithm have been discussed in detail. The results of proposed algorithms (SC and CNN) will be compared with existing algorithms like local area based and global like Graph-Cut

(GC), dynamic programming (DP), Semi-Global Matching (SGM), Belief Propagation (BP) and cooperative algorithm (CA).

Table 2. 4: Limitation of Existing Stereo Matching Algorithms.

Stereo Matching methods	Limitations in stereo matching existing algorithms
Local methods (Area based)	• Slow processing speed and suitable only for images with little distortion • Cannot deal with smooth areas and high computational complexity • Sensitive to image intensity changes which are caused by noise, varying illumination and different sensors [90].
Local methods (Feature based)	• Suits only image with short baseline and the feature description must fulfill several • Conditions involving invariance, uniqueness, stability, and independence [91] • Need high dimensional non-rigid mapping and a large number of correspondences are needed for the refinement of mapping functions. • The features cannot be exactly matched because of noise affected by the existence of outliers [92].
Global Methods (DP)	• Give streaking effect and cannot handle textureless region and depth discontinuity [81].
Global Methods (GC)	• Lack of textureless region must need ordering constraint [87].
Global Methods (CA)	This is computationally expensive and could not handle textureless region and depth discontinuity [89].
Global Methods (BP)	This algorithm is computationally expensive and could not handle depth discontinuity [93].

2.6 Related Work

The sparse representation widely used in signal processing, image processing, and computer vision applications. In this thesis, the sparse representation is introduced in stereo matching using fixed and adaptive dictionaries. The introduction and related work presented by various authors of sparse representation have been explained in this section. Moreover, the deep learning techniques used in high level and low-level vision are well presented and most famous and excellent algorithm called convolutional neural network (CNN) have been discussed in detail. The CNN algorithm produced highly accurate results in the field of image classification, object detection, and various other tasks; the image segmentation, stereo matching, image inpainting and image rendering [94].

2.6.1 Sparse Representation Background

The sparse representation has a wide range of applications in image processing and computer vision such as image separation using sparse representation, Image in-painting, signal source separation, image classification, de-noising, signal blind source separation. In recent years sparse representation has much attention due to their wide variety of applications used in image processing [95]. In sparse representation, the problem can be solved for the most compressed representation of a signal in terms of a linear combination of atoms in an over-complete dictionary. The process of obtaining a sparse representation for a signal or image requires explicit knowledge of the synthesis dictionary. One crucial problem in a sparse representation based application is how to choose the dictionary [96]. Nowadays the representation of signals in multi-orientation and multiscale there are many transforms used such as ridgelet, curvelet, counter let, shearlet, bandlet [96]. These are the motivation for the research on the sparse representation. Sparse representation produces good performance as compared to other methods based on time domain processing and orthonormal basis transforms. Particularity, the focus of research based on sparse representation has three parts. First is the dictionary design and trained dictionary using well known KSVD algorithm, the second and most important is optimization algorithms based on pursuit methods like matching pursuits, orthogonal matching pursuit, and basis pursuits and the third area is the LARS/homotopy methods [97].

There are two types of dictionaries are used in sparse representation predetermined and adaptive dictionaries. The predetermined dictionaries are curvelet, discrete cosine, wavelet, ridgelet, and bandlet, and adaptive dictionaries are trained KSVD and method of directions (MOD) algorithms. The over-complete dictionary is generated by combining multiple standards transforms, including curvelet transform, ridgelet transform, and discrete cosine transform. The over-complete predetermined dictionary such as wavelet-based dictionary, a shape based and region based dictionary is used to represent the small image patch sparsely [98]. Currently, the most mostly used transforms for execution that tasks is the Discrete Cosine Transform (DCT) and Discrete Wavelet Transforms (DWT). An important reason for the attractiveness of both these transforms is the feasibility of their fast implementation. Wavelet and DCT transforms are broadly used in image processing applications. Many image processing tasks take advantage of sparse representations of image data where most information is packed into a small number of samples. Typically, these representations are achieved via invertible and non-redundant transforms. Despite the fact that wavelets have had a wide impact on image processing, they fail to efficiently represent objects with highly anisotropic elements such as lines or curvilinear structures (e.g. edges). The reason is that wavelets are non-geometrical and do not exploit the regularity of the edge curve. The success of wavelets is mainly due to the good performance of piecewise smooth functions in one dimension. Unfortunately, such is not the case in two dimensions. In essence, wavelets are good at catching zero-dimensional or point singularities, but two-dimensional piecewise smooth signals resembling images have one-dimensional singularities. That is, smooth regions are separated by edges, and while edges are discontinuous across, they are typically smooth curves. Intuitively, wavelets in two dimensions are obtained by a tensor-product of one-dimensional wavelet and they are thus good at isolating the discontinuity across an edge, but will not see the smoothness along the edge.

In order to use over-complete and sparse representations in applications, one needs to fix a dictionary D. Exact determination of sparsest representations proves to be an NP-hard problem [99]. Hence, approximate solutions are considered instead. In the last decade or so, several efficient pursuit algorithms have been proposed. The simplest ones are the Matching Pursuit (MP) [100] or the Orthogonal Matching Pursuit (OMP) algorithms [101]. These are greedy algorithms that select the dictionary atoms sequentially. A second well-known pursuit approach is the Basis Pursuit (BP) [102].

The algorithms range widely in empirical effectiveness, computational cost, and implementation complexity. Unfortunately, there is little guidance available on choosing a good technique for a given parameter regime [103]. Extensive study of these algorithms in recent years has established for the solution is sparse enough [98].

- Types of Dictionaries

Dictionary selection is based on the criterion that it can span whole signal space; it means that any set of atoms which are capable of capturing the signal structure. The sparse signal is an interesting example which can be represented by linear combinations of only a small number of atoms. Sparsity considered as the most important player in compression of signals and also different types of inverse problems. The main issue of sparsity is that it depends on how to design the dictionary and it is a challenging task [96]. Less research work has been done on learning dictionaries for stereo image representations. Therefore, learning such dictionaries could make significant improvements in for depth or height estimation. When the dictionary forms a basis, every signal is uniquely represented as the linear combination of the dictionary atoms. In the simplest case, the dictionary is orthogonal, and the representation coefficients can be computed as inner products of the signal and the atoms; in the non-orthogonal case, the coefficients are the inner products of the signal and the dictionary inverse, also referred to as the bi-orthogonal dictionary. For years, orthogonal and bi-orthogonal dictionaries were dominant due to their mathematical simplicity [104]. One of the dictionary class used in signal and image processing is called an over-complete dictionary. The details of the over-complete dictionary with the specific signal application is explained in next section.

- Over-complete Dictionary

A dictionary is called over-complete for sparse representation of image patch if the number of elements in the dictionary is more than the number of pixels in the image patch. This led to the development of newer over-complete dictionaries, having more atoms than the dimensions of the signal, which promised to represent a wider range of signal phenomena. The move to over-complete dictionaries was done vigilantly, in an attempt to minimize the loss of promising properties offered by orthogonal transforms. Many dictionaries formed tight frames, which guaranteed that the representation of the signal as a linear combination of the atoms could still be recognized with the inner products of the signal and the dictionary. Another approach, manifested by the best basis algorithm, utilized a specific dictionary structure which essentially allowed it to serve as a pool of atoms from which an orthogonal sub-dictionary could be efficiently selected. Research on general over-complete dictionaries commonly started over the past decade and is still extremely ongoing. Such dictionaries introduce an interesting ambiguity in the definition of a signal representation [105].

- Types of Over-complete Dictionary

At present, there are generally two ways to obtain an over-complete dictionary.

- Building a specifying dictionary based on a mathematical model of the data using fixed basis dictionary composed of over-complete basis vectors.
- Adaptive over-complete dictionary by learning a dictionary to perform better on a training set [106].

The history of modeling dictionary could be traced back to 1960s, such as the fast Fourier transform (FFT), an over-complete dictionary that can lead to sparse representation is usually achieved by exploiting a pre-specified set of transformations functions, e.g transform domain method or is devised based on learning [107]. Both of the transform domain and dictionary learning based methods transform image samples into other domains and the similarity of transformation coefficients is exploited. The difference between them is that the transform domain methods usually utilize a group of fixed transformation function to represent the image sample, whereas the dictionary learning methods apply sparse representation on an over-complete dictionary with redundant information. Moreover, exploiting the pre-specified transform matrix in transform domain methods is attractive because of its fast and simplicity. Specifically, the transform domain methods usually represent the image patches by using the orthonormal basis such as over-complete wavelets transform, super-wavelet transform, bandlets transform, curvelet transforms counter lets transform and steerable wavelet filters [108]. Sparse and redundant representation, modeling of data assumes an ability to describe signals as linear combinations of a few atoms from a pre-specified dictionary. This dictionary is simpler and leads to fast algorithms for the assessment of the sparse representation. This is definitely the case for over-complete wavelets, Discrete cosine transform, Haar Wavelet curvelets, ridgelet, quintuplets, steerable wavelet filters, short-time Fourier transforms. The flow chart of different dictionaries is shown Figure 2.13. The fixed dictionary, as well as adaptive dictionaries, are further classified into different bases functions. The fixed dictionaries are classified as DCT, DWT, HAAR, Ridgelet, Curvelet, Bandlet and Courterlet. The adaptive dictionaries are based on the method of optimal direction (MOD), maximum-likelihood (ML), maximum a posterior (MAP) and K-SVD algorithms as shown in Figure 2.13 [104]. Types of approaches used in dictionaries based on sparse representation as shown in Table 2. 5. The Model approach uses the fixed dictionaries based on pre-specified bases functions are easy to use, simple and provide faster response. The training approach is based on adaptive dictionaries and produces may be slightly more accuracy as compared to fixed, but this approach is computationally expensive and takes more time as compared to a model approach based on the fixed dictionary. Choosing an appropriate dictionary is a key step towards a good sparse decomposition. Thus, to represent efficiently isotropic structures, a qualifying choice is the wavelet transforms [108]. The curvelet system is a very good candidate for representing piecewise smooth images away from contours [96]. The ridgelet transform has been shown to be very effective for sparsely representing global lines, in an image. For locally oscillating textures, one can think of the local discrete cosine transform (DCT), wave atoms or brushlets [105]. These transforms are also computationally manageable, particularly in large-scale applications. The proposed system model used fixed dictionaries due to their fast response. The proposed fixed dictionaries are discrete cosine transform (DCT), discrete wavelet transform (DWT), and discrete ridgelet transform (DRT) and discrete Tchebichef Transform (DTT). The detail explanation of each dictionary based on fixed bases function is explained next section. The DCT is a variation of the Discrete Fourier Transform; replace complex analysis with real numbers by asymmetric signal extension. The DCT is an orthonormal transform; known to be well suited for first-order Markov stationary signals.

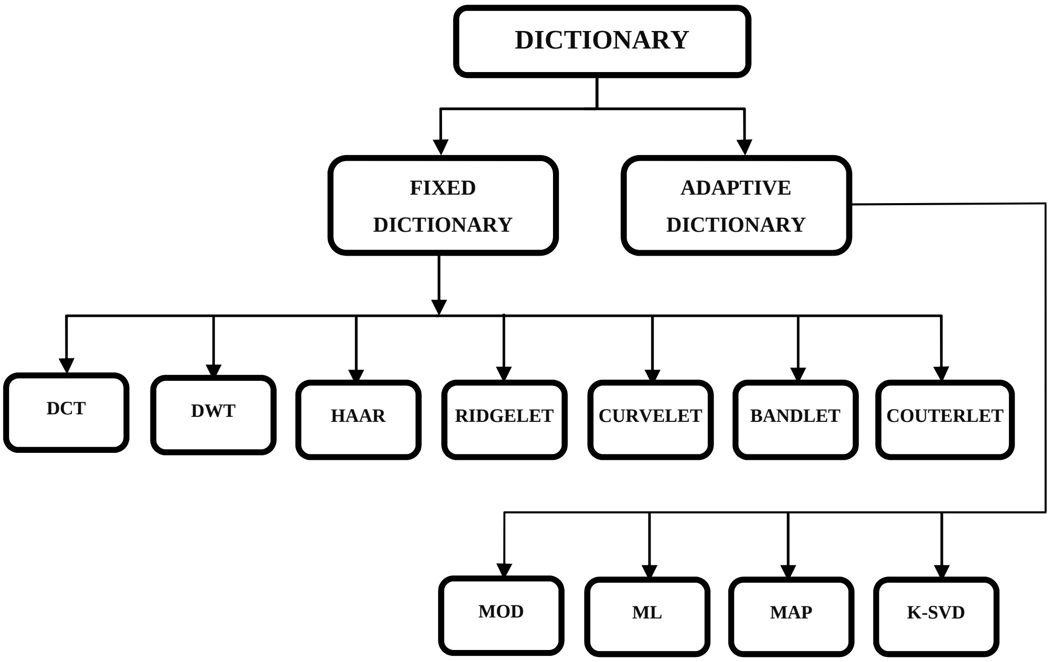


Figure 2.13: Types of dictionary based on fixed and adaptive dictionaries.

Table 2. 5: Types of approaches used in dictionaries based on sparse representation.

Types of approach	Characteristic
Model Approach	• Choose Predesigned transform • Simpler • Pre-specified • Fast transform • Maybe not feasible for small signals
Training Approach	• Adaptive training set • Fits the data • Design fits the true objectives • Can be adapted for small signal families. • No fast version • Slow training

A discrete cosine transform (DCT) expresses a finite sequence of data points in terms of the sum of cosine functions oscillating at different frequencies. DCTs are important for numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high-frequency components can be discarded) [75], to spectral methods for the numerical solution of partial differential equations. The use of cosine rather than sine functions is critical for compression since it turns out fewer cosine functions are needed to approximate a typical signal, whereas for differential equations the cosines express a particular choice of boundary conditions.

In particular, a DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT) but use only real numbers. DCTs are equivalent to DFTs of roughly twice the length, operating on real data with even symmetry (since the Fourier transform of a real and even function is real and even), wherein some variants the input and/or output data are shifted by half a sample. There are eight standard DCT variants, of which four are common. The multidimensional variety of the various DCT types follow straightforwardly from the one-dimensional definitions: they are simply a separable product of DCTs along each dimension. The pseudo code used in proposed method to build over-complete DCT dictionary is shown in Table 2. 6. The 2-D DCT is expressed in mathematical form as follows.

$$X_1 = cos\left[\frac{\pi \times (2n_1 + 1) \times T_1}{2M_1}\right] \tag{2.1}$$

$$X_2 = cos\left[\frac{\pi \times (2n_2 + 1) \times T_2}{2N_1}\right] \tag{2.2}$$

$$X_{T_1,T_2} = \sqrt{\frac{2}{M_1}} \times \sqrt{\frac{2}{N_1}} \sum_{n_1=0}^{M_2-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} X_1 X_2 \tag{2.3}$$

where x_{n_1,n_2} is input signal in 2-dimensional, X_{T_1,T_2} is the output data has size $T_1 \times T_2$, $\sqrt{\frac{2}{M_1}}$, $\sqrt{\frac{2}{N_1}}$ are the scaling factors. X_1, X_2 is the 2-D basis used in the discrete cosine transform. This simple idea captivated the signal processing and harmonic analysis communities, and in a series of influential works by Meyer, Daubechies, Mallat, and others an extensive wavelet theory was formalized [109]. The dictionary based on DCT bases function provides efficient solution for signal reconstruction in repetitive patterns.

Table 2. 6: The dictionary bases function based on proposed DCT bases function.

	Dictionary based on DCT Basis Functions
1	D is dictionary size, K_1, K_2 scale factors, n_1, n_2 are the indexing parameters, M_1number of dictionary atoms, N_1 size of dictionary, M_1, M are scaling factors.
2	**FOR** $K_1 = 1\!:\!M$ **do**
3	**FOR** $K_2 = 1\!:\!M_2$ **do**

4	**IF** $K_2 = 1; n = \sqrt{1/M_1}$
	ELSE
	$n = \sqrt{1/M_1}$
	END IF

5	**FOR each** $n_1 = 1\!:\!M_1$**do**
6	**FOR each** $n_2 = 1\!:\!N_1$ **do**
7	$Temp(n_1, n_2) = dct(n_1, n_2, K_1, K_2)$
8	$D = Temp(n_1, n_2)$
9	**END FOR End FOR**
10	$D(:, count) = Temp(:)$
11	**END FOR** **END FOR**

The theory was formulated for both the continuous and discrete domains, with a complete mathematical framework relating the two. A significant breakthrough came from Meyer's work, who found that unlike the Gabor transform (and contrary to common belief) the wavelet transform could be designed to be orthogonal while maintaining stability. It is Van is extremely appealing property to which much of the initial success of the wavelets can be attributed specifically to the interest of the signal processing community. Mallat and his colleagues [110] which established the wavelet decomposition as a multi-resolution expansion and put forth efficient algorithms for computing it. In Mallat's description, a multi-scale wavelet basis is constructed from a pair of localized functions referred to as the scaling function and the mother wavelet. The mother wavelet is a high-frequency signal with its various scales and translations spans the signal detail. In the orthogonal case, the wavelet basis functions at each scale are critically sampled, spanning precisely the new detail introduced by the finer level.

Non-linear approximation in the wavelet basis was shown to be optimal for piecewise-smooth 1-D signals with a finite number of discontinuities [104]. This was a striking finding at the time, realizing that this is achieved without prior detection of the discontinuity locations. Unfortunately, in higher dimensions the wavelet transform loses its optimality; the multi-dimensional transform is a simple separable extension of the 1-D transform, with atoms supported over rectangular regions of different sizes. This separability makes the transform simple to apply; however, the resulting dictionary is only effective for signals with point singularities, while most natural signals exhibit elongated edge singularities. The JPEG2000 image compression standard, based on the wavelet transform, is indeed known for its ringing (smoothing) artifacts near edges. The Morlet wavelet is the most popular complex wavelet used in practice, which mother wavelet is defined as

$$\varphi(t) = \frac{1}{\pi^{1/4}}\left(e^{j\omega_0 t} - e^{{\omega_0}^2/2}\right)e^{-t^2/2} \tag{2.4}$$

where ω_0 is the central frequency of the mother wavelet. Note that the term $e^{{\omega_0}^2/2}$ is used for correcting the non-zero mean of the complex sinusoid, and it can be negligible, $\omega_0 > 5$. Therefore, in some research the, mother wavelet definition of

The Morlet wavelet is:

$$\varphi(t) = \frac{1}{\pi^{1/4}} e^{j\omega_0 t} e^{-t^2/2} \tag{2.5}$$

where the central frequency$\omega_0 > 5$. There is mother wavelet function used in this experiment to make the dictionary function. This function is representing in 2-dimensinal wavelet function [111]. The complex mother Morlet wavelet in two-dimensional given in Equation (2.6). The pseudo code to implement DWT over-complete dictionary is shown in Table 2. 7.

$$F(n_1) = 2^{S_1/2}\sin(5(n_1 - \tau))e^{-t^2/2} \quad (2.6)$$

$$F(n_2) = 2^{S_2/2}\sin(5(n_2 - \tau))e^{-t^2/2} \quad (2.7)$$

$$F(n_1, n_2) = F(n_1) \times F(n_2) \quad (2.8)$$

.

The research in dictionary learning has followed three main directions that correspond to three categories of algorithms:

- The probabilistic learning methods.
- The learning methods based on clustering or vector quantization.
- The methods for learning dictionaries with a particular construction [108].

In this thesis, only fixed dictionaries are focused due to simplicity, easy to implement and provides faster response. Sparsity and overcompleteness have been successfully used for dynamic range compression in images, separation of texture and cartoon content in images, in painting, and more [105]. There are a lot of methods are exist in the literature for the extraction of the sparsest representation. Initially, it is assumed that all those methods used a fixed and known dictionary. Generally, the methods are used to solve the sparsest solution are further divided into three classes as shown the Figure 2.14. These classes or approaches are

- Convex Relaxation Approach
- Non-convex local optimization approach
- Greedy search methods

Commonly used strategies are often based on convex relaxation, non-convex (often gradient based) local optimization or greedy search strategies. Convex relaxation is used in algorithms such as Basis Pursuit (BP) and Basis Pursuit De-Noising (BPD), the least absolute shrinkage and selection operator (LASSO) and Least Angle Regression (LARS) [112]. Non-convex local optimization procedures include the Focal Underdetermined System Solver (FOCUSS) and Bayesian approaches such as the Relevance Vector Machine, also known as Sparse Bayesian Learning (SBL). The interest of the proposed work is in greedy methods, the most important of which are Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP) and Orthogonal Least

Squares (OLS), also often known as orthogonal regression matching pursuit (ORMP) in the regression literature, as forwarding selection [104].

Table 2. 7: Dictionary based on DWT basis function based on Morlet wavelet transform difference of Difference of Gaussian.

Dictionary based on Difference of Gaussian wavelet Basis Function	
1	D is dictionary size, S_1, S_2 scale factors, θ_1, θ_2 are the dilation parameters, M_1 is number of dictionary atoms, N_1 size of dictionary M, N are scaling factors.
2	**FOR each Scale** $S_1 = 1{:}\,M$ **do**
3	**FOR each Scale** $S_2 = 1{:}\,M$ **do**
4	**FOR each translation** $\theta_1 = 1{:}\,N$ **do**
5	**FOR each translation** $\theta_2 = 1{:}\,N$ **do**
6	**FOR each** $n_1 = 1{:}\,M_1$ **do**
7	**FOR** $n_2 = 1{:}\,M_1$ **do**
8	$temp(n_1, n_2) = DOG\_Wavelet(S_1, S_2, \theta_1, \theta_2)$
9	$D = temp(:)$
10	**ENDFOR END FOR**
11	$D(:, count) = temp(:)$

12	**END FOR**
	END FOR
	ENDFOR
	ENDFOR

The simple ones are the matching pursuit (MP) and the orthogonal matching pursuit (OMP) algorithms. These are greedy algorithms that select the dictionary atoms sequentially. These methods are very simple, involving the computation of inner products between the signal and dictionary columns, and possibly deploying some least squares solvers. Sparse signal representation has emerged as an extremely successful tool for analyzing a large class of signals. Many signals like audio, images, video, etc., can be efficiently represented with a linear superposition of only a small number of properly chosen basis functions.

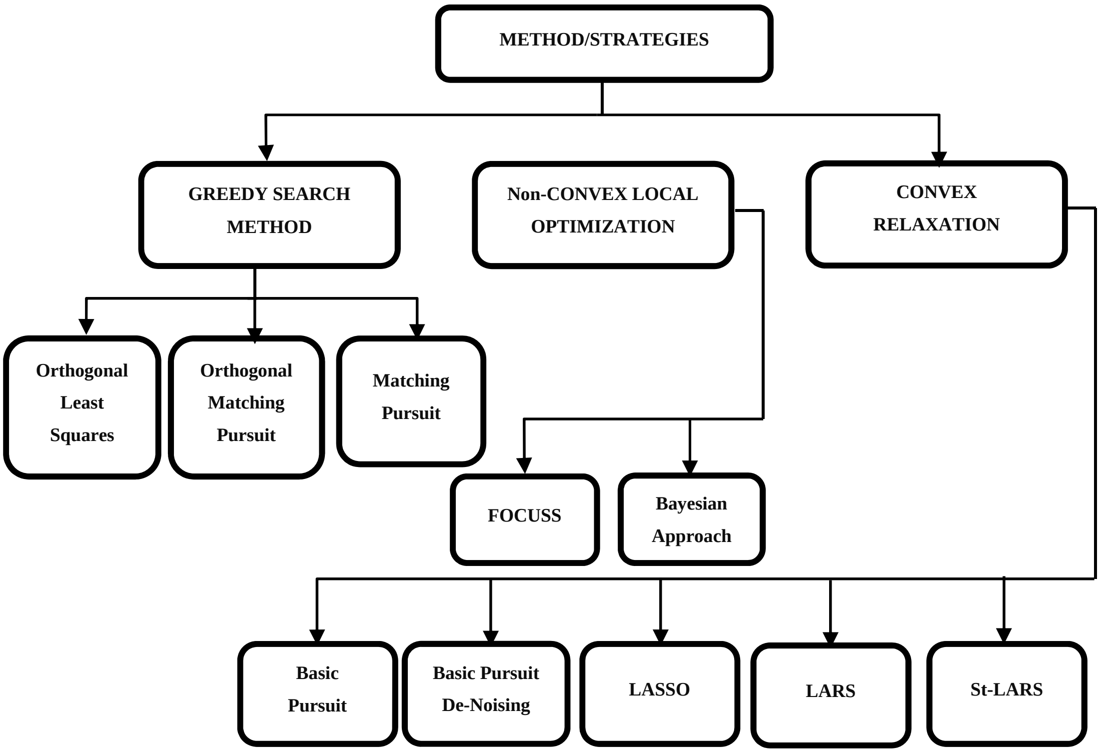


Figure 2.14: Methods or techniques for sparse exact solution.

Although the use of orthogonal bases like Fourier or Wavelets is widespread, the latest trend is to use over-complete basis where the number of basis vectors is greater than the dimensionality of the input vector. A set of over-complete basis (called a dictionary) can represent the essential information in a signal using a very small number of nonzero elements. This leads to higher sparsity in the transform domain as compared to that achieved by sinusoids or wavelets alone. Such compact representation of signals is desired in many applications involving efficient signal modeling. The sparse representation of an image means that an image can be represented by a linear combination of fewer elements of an over-complete dictionary (OCD) [105]. The dictionary elements could also be taken as features extracted from the image patches of similar characteristics. The dictionary representation based on image patch is represented as Figure 2.15. The dictionary construction using image sample is based on the noise free image patch in aerial images.

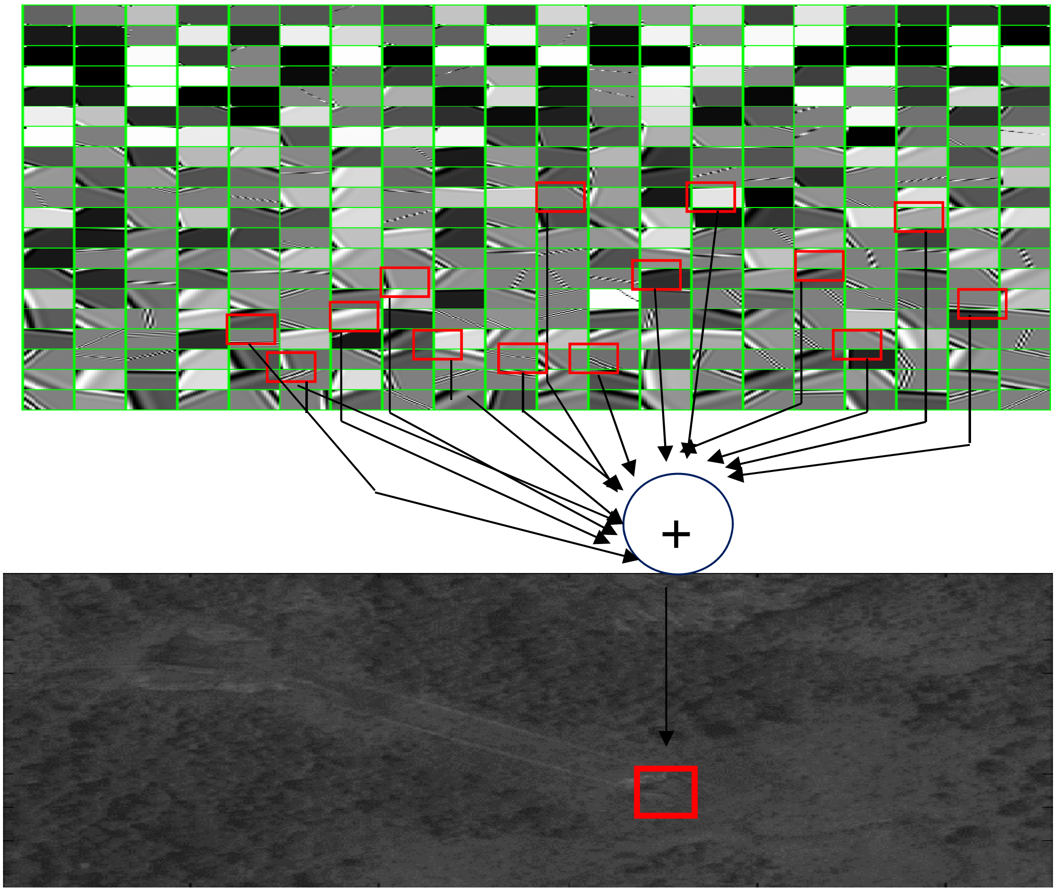


Figure 2.15: Sparse representation of an image patch using an over-complete dictionary.

The image sparse representation has been used for super-resolution, fusion, in-painting, compression and in a variety of inverse problems. However, sparsity largely depends on the design of the dictionary, which represents a difficult and challenging problem. A comparison of all state-of-the-art dictionary learning methods is made difficult by the fact that the efficiency of the algorithms differs with the dictionary size and the training data. The first proposed method used to calculate the height of vegetation near power transmission poles and lines based on spare representation algorithm. The disparity map can be computed based on proposed sparse method, which could be further used for height estimation. The overview of deep learning and convolutional neural network and it's all related components are mentioned with details in the next section. The deep learning algorithms could provide better solution for stereo matching algorithms and disparity estimation.

2.6.2 Deep Learning and CNN Algorithm Background

Nowadays, the deep learning methods [113],[114],[115]have prospered great success in many practical and classical problems, such as speech recognition, object recognition, detection, and natural language processing. These methods have tremendous improvement in the state-of-the-art records and have much attention in both the academic and industrial communities [114]. The deep learning algorithms rely on hierarchical features belongs to the different abstraction level. The most prominent and efficient technique in deep learning is the deep convolutional neural networks (CNN) [116]. The deep convolutional neural networks (CNN) [116], is one of the most attractive methods and has been used in scene classification, recognition, and detection tasks. It produced excellent results using state of the art number of benchmarks [117] [118],[119] and proven to be the best method for automatic feature extraction and classification. CNN is a biologically inspired architecture based on multiple stages and consists of sequentially convolutional, pooling and fully-connected layers.

It is trained based on the supervised way. Usually, it is difficult to train low-level dataset effectively. CNN used small image patch with a fixed size is called receptive field for extraction of deep features in order to intact spatial information. The Chen [120] proposed deep CNN (DCNN) to detect vehicles on the roads and results shows that CNN can be used for object detection for high-resolution aerial images. The Yue [121],[122] proposed deep CNN framework for classification of remote sensing images and explore spatial and spectral features in higher dimensional space which produced more computational cost and may deliver lower classification accuracies.

Moreover, recently the CNN have been used in low-level vision tasks such as stereo matching, and optical flow prediction [123], super-resolution [124], denoising [125], and single-view depth estimation [126]. In the perspective of stereo estimation, the author [127] proposed CNN to compute the matching cost between two image patches. Explicitly, a siamese network has been utilized to take the same sized left and right image patches using a few fully-connected layers to predict the matching cost. Also, they trained the model using binary cross-entropy loss function and explored diverse CNN [127] based architectures for comparing image patches. The CNN has been utilized to measure the matching cost between two image patches. It has also been subjugated to learn small patches in stereo images of the stereo matching task. The CCN has been used for stereo matching [128] in recent years and achieved good results comparatively with existing stereo matching methods. These methods trained the CNN network using a set of stereo image patches and provide the matching cost by using trained CNN networks. These methods have the main difference in architecture point of view. The proposed CNN in our design used to design the disparity map which further needs to measure the height of vegetation and trees near power transmission poles.

To compute disparity map based on convolutional neural networks, the authors [128] provides disparity map using pooling and subsampling layer for larger patch sizes and larger variation in viewpoint. The proposed methods based on UAV and satellite stereo images newly computed disparity map using CNN approach. Further, depth or height of the object can be measured and evaluated based on ground truth height of power transmission poles as a reference map. The CNN algorithm has been proposed as matching cost function as proposed algorithm and disparity map can be computed using fully connected feature vector from the left image and search feature vector from the right image. In our proposed work CNN has a different approach in architecture and optimization point of view.

The feature of convolutional layers would be reconstructed to images similar to the original image. The convolutional layer generates spatial features by activating the output values of previous layers with spatial features. The convolutional layer produced spatial features as output by convolving with spatial filters of previous layers. In this example, two filters are used to convolve with input image has size 6x6. The result of a convolution between a single filter mask and the input image is referred to as a map. A single map is a 2D matrix which is the result of applying a filter to the entire image. Each layer of a CNN will output multiple maps, one for each of the neurons, kernels, features of filters in that layer. The filter can be pre-trained by choosing random initialization with supervised fine-tuning or unsupervised feature learning algorithm. The filter parameters are trained by using stochastic gradient descent (SGD) algorithm with backpropagation trained network in the fully connected layer. The two output feature maps are produced by convolving the input with twos spatial filters. The size of feature map is less than the size of the input image. The formula to calculate the size of convolution output is given in equation (2.9).

$$O = \left(\frac{I - F + 2P}{S}\right) + 1 \tag{2.9}$$

where O is size of convolutional output feature map. The I is input image, F spatial filter, P is the zero padding and S is the stride. In this example, $P = 0, S = 1, F = 3 \times 3$ and $I = 6 \times 6$. Subsitute the values in the equation to get the size of O which is equal to 4.In this way , the size of output of the convolutional layer can be calculated. The two feature maps as shown in Figure 2.16 and Figure 2.17.

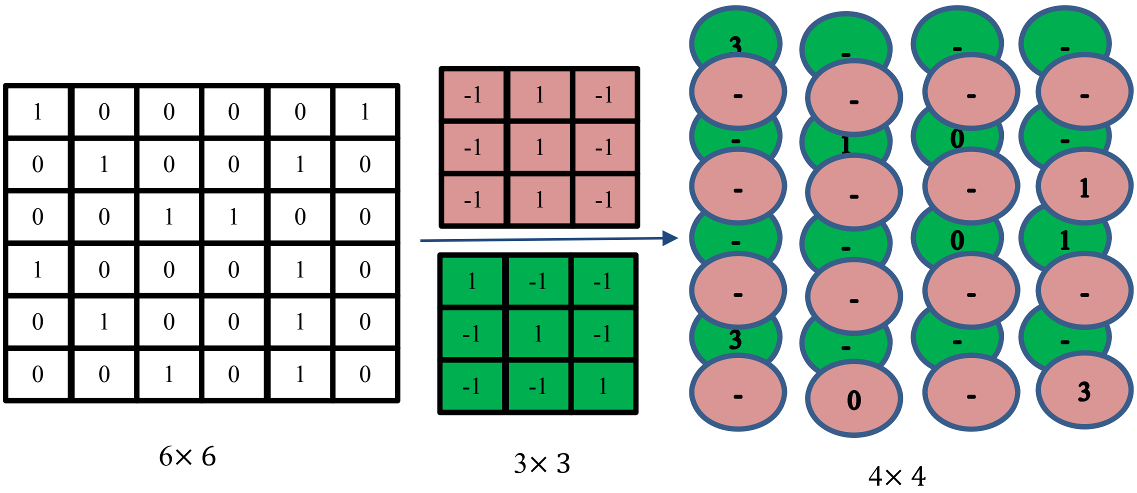


Figure 2.16: The convolutional operation using two filters produce two feature maps.

Pooling layer reorganizes the information from lower level layers to create the abstract representation. Pooling layer performed on feature maps by reducing the sample size based on the maximum or mean function over the local spatial region with the non-overlapping framework. Poling layer which preserves the transitional and rotational invariant features. A fully contended layer takes all neurons in the previous layer (be a fully connected, pooling or convolutional) and connects it to every single neuron it has. Fully connected layer are not spatially located anywhere (you can visualize then as one –dimensional), so they can be on convolutional layer after a fully connected layer.

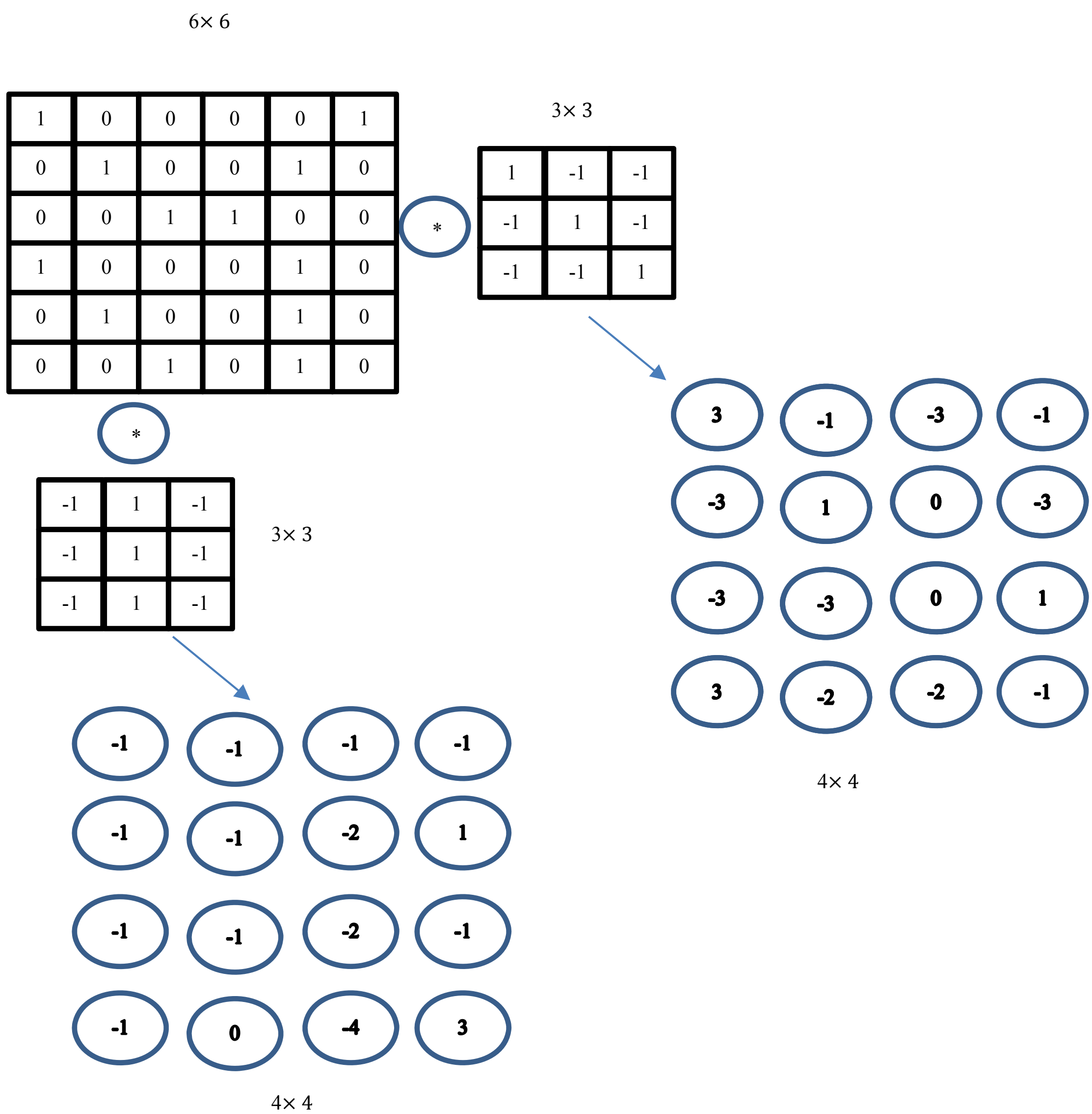


Figure 2.17: The convolutional layer performed on the input image (6x6) with filter size (3x3).

The rectified linear unit (ReLU) used in CNN as an activation function. It provides a faster response as compared to other activations such as a continuous trigger, hyperbolic tangent, sigmoid and absolute of hyperbolic tangent function. The example of ReLU function using a small input image is shown in Figure 2.18. The ReLU performed on the image only on positive values and insert zero on the negative values.

15	20	-10	35
18	-110	25	100
20	-15	25	-10
101	75	18	23

ReLU function →

15	20	0	35
18	0	25	100
20	0	25	0
101	75	18	23

Figure 2.18: The ReLU function operation on a small image (4x4).

2.6.3 Limitation of Existing Work

The distributed transmission lines located inside a city tend to be more difficult to inspect via air-borne methods and a good view is often blocked by the trees. The second most difficult is sub-transmission lines because they are mostly erected with many turns, unlike transmission lines that are built on a straight path and have fewer connections, switches. The accuracy, time and cost assessment of traditional methods have been discussed in the aforementioned section with a limited number of power transmission poles. The literature reviewed indicates that the traditional methods for monitoring the transmission lines can be costly, time-consuming and inaccurate. To overcome these circumstances, the satellite stereo imaging can be used for the protection of power lines from blackouts due to dangerous vegetation encroachment. The satellite stereo imaging provides the cost-effective solution for monitoring the vegetation or trees near power transmission lines. The challenges can be solved by using existing stereo algorithms based on area or windows and include Global Matching Algorithms (GMA), Graph-Cut (GC), Dynamic Programming (DP), and Local Matching (LM) [74].

However, GMAs are computationally complex and time consuming compared to area-based algorithms, which are, however, less accurate. Existing stereo matching algorithms that are based on GC and DP do not provide a reliable disparity map for use in height estimation of objects and vegetation near power transmission poles. The DP algorithm propagates a local error through all search lines between stereo images and also requires uniqueness and ordering constraints [81]. In addition, it produces a streaking effect during stereo matching. The GC approach can cause excess regularization (over smoothing) between each cut of a graph; hence, disparities tend to blur when using larger displaced values [37]. To solve the cited issues in stereo matching, and especially computation costs for disparity mapping, improved stereovision algorithms are needed. Contemporary deep learning algorithms have been applied for depth estimation with stereo, optical flow estimation, and stereo matching techniques.

Deep learning methods have prospered with much success in many practical and classical problems such as speech recognition, object recognition and detection, and natural language processing [94]. These approaches have tremendously improved state-of-the-art procedures and have garnered much attention academically and industrially. Such algorithms rely on hierarchical features that hold different levels of abstraction. The most prominent and efficient approach to deep learning is that of deep convolutional neural networks (CNN) that are used for recognition and detection tasks as well as high-level scene classification. Furthermore, CNNs have been employed in low-level vision tasks such as stereo vision and optical flow prediction. Within the stereo estimation framework, CNN is utilized to measure the matching cost between two image patches and has also been harnessed to learn small patches in stereo images for stereo matching tasks. The existing stereo matching algorithms based on CNN model did not handle the depth discontinuity in the stereo matching process and these are only dealt with the textureless region and some of them handled occlusion detection. The detail of each existing stereo matching algorithm with their limitations based on CNN model is shown in Table 2. 8. To overcome these limitations in existing and related stereo matching algorithms, the solution based on sparse representation and CNN algorithm is proposed to measure the disparity map which can be used for height estimation of an object near power transmission lines and power poles. Based on limitations in existing methods, a novel method is being based on sparse representation image stereo matching technique and also proposed the second method based on deep learning approach. These novel methods provide the reliable disparity map using aerial stereo imagery (satellite and UAV). However, these methods may be computationally expensive. The proposed system also estimate various number of threats near the power transmission lines and power poles. These methods may provide feasible solution for power management companies for monitoring of objects near power transmission lines and power poles.

Table 2. 8 The most related existing CNN models used to estimate disparity map based on stereo imagery.

No	Author with years	Limitations in stereo matching algorithms based on CNN models
1	Zbontar [94]	• Trained a CNN to compute the matching cost. • Used five fully connected layer • Used binary predictions on image patches • More hyperparameters and fine-tuning necessary. It deals the textureless regions in the stereo image based on camera stereo images.
2	Tyler Jordan[129]	• The CNN used four fully connected layers • Used cross-entropy function to compute the disparity map. This model produced better accuracy and well reconstructed textureless region. It did not provide the solution in stereo images in occlusion and depth discontinuity.
3	Zhuoyuan Chen [130]	• Used multiscale CNN with the ensemble for matching the cost. Automatically fuses features vectors learned at different scale-space. It handled only the textureless region and could not provide a better solution for depth discontinuity and occlusion in stereo matching process.
4	Zagoruyko [131]	• CNN layers and softmax classifier with spatial pyramid layer. Comparing small patches used in narrow baseline stereo. This method only produced better disparity map using narrow baseline stereo imagery.

2.7 Summary

In this chapter, the importance of vegetation monitoring and blackouts occurred due to vegetation near power lines internationally and particularly in Malaysia has been discussed. The detail of each blackout has been given in this chapter and the standards of vegetation monitoring in Malaysia has been discussed. The statistical figures of hazards occurred due to vegetation or tree interfering power transmission lines in different states of Malaysia has been explained. The literature on existing monitoring methods and their limitations are explained in this chapter. According to literature the graph cut algorithm, as well as dynamic programming algorithm, provides better accuracy as compared to area-based stereo matching algorithm. The introduction of proposed technique based on sparse representation and CNN algorithm has been provided in this chapter. The classification of stereo matching algorithms based on local and global methods has been explained. The main limitations in local and global stereo matching methods have been described. The existing work related to CNN algorithm has been explained and briefly discussed in architecture point of view.

CHAPTER 3

METHODOLOGY

3.1 Introduction

In this chapter, the data acquisition based on satellite and UAV images has been discussed. These images are segmented into smaller parts by using automatic cropping techniques due to the large size of aerial images. The proposed framework and the details of each step of the proposed framework for measuring the height and a various number of threats due to vegetation / trees imposed on the power lines / poles have been discussed in detail. An overview of different stereo matching algorithms has been given that exist in the literature for estimating disparity map. The two proposed algorithms based on the stereo matching process have been implemented for height and distance estimation. These algorithms are based on sparse representation and CNN model. The proposed sparse representation algorithm is first time introduced to compute the disparity map based on aerial (Satellite and UAV) stereo images. The sparse based stereo matching approach used adaptive and fixed dictionaries to store local information of stereo images. The construction of dictionaries is described briefly in this chapter. The proposed dictionaries such as discrete tchebichef transform (DTT), and discrete ridgelet transform (DRT) are discussed and explained with the mathematical formulation. The second proposed method based on CNN algorithm has been implemented to compute disparity map and this novel method provides excellent accuracy as compared to existing stereo matching algorithms. The proposed triangulation technique has been discussed to estimate the height and distance between vegetation / trees and power transmission lines and power poles. The performance metrics (accuracy, precision and recall) has been computed using the estimated height.

3.2 Data Acquisition

There are two types of datasets have been used in this work. The first dataset is consists of satellite stereo images and the second dataset comprises UAV stereo images of the area of interest. The details of each dataset are explained in next sections.

3.2.1 Satellite Stereo Images

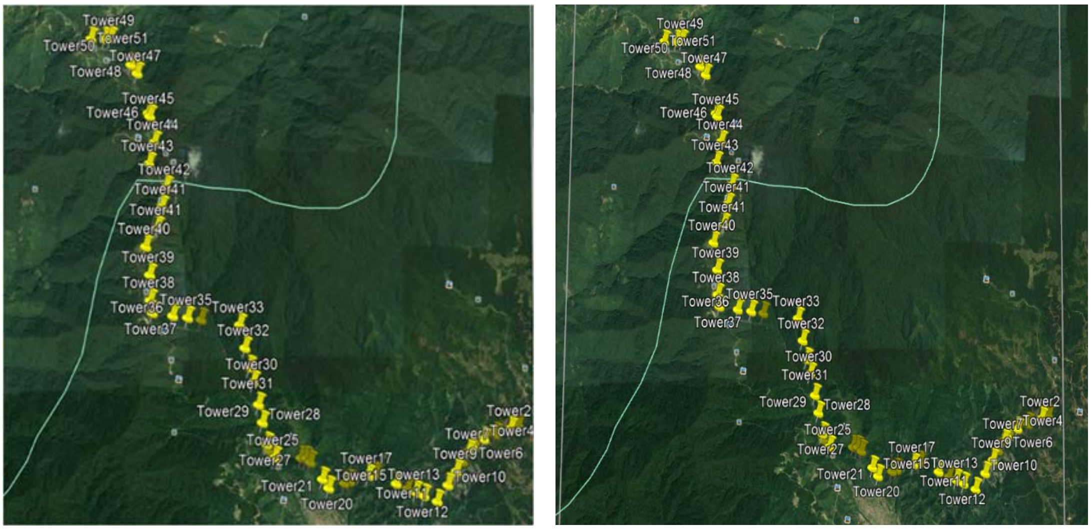
There are two stereo Pleiades satellite images: one is the left and one is the right image as shown in Figure 3.1. The ground sample distance of Pleiades satellite image is 0.5m. The area of interest scanned by the Pleiades is 100 square kilometer (Sqkm). The one image was collected on 13 March 2014 and the other image was collected on 16 May and 18 June respectively Figure 3.1 and Figure 3.2. The overlapping area of two stereo images is above 60 percent. There has been 71 to 123 power transmission poles covered from the area of Kota Kina Balu in Sabah East Malaysia. Total 52 poles are covering the area of 100 km from starting to ending.

The first satellite to be launched for the creation of Pleiades constellation was the Pleiades. It was launched in December 2011 from French Guiana. The Pleiades constellation provides very-high-resolution optical products in record time, offering daily revisits to any point on the globe and acquisition capabilities tailored to meet the full spectrum of civil and military requirements. Pleiades swath covers 20 km at nadir, the satellites' high agility allows acquiring in the same pass a mosaic of images covering a larger area (up to 120km*120km), or stereoscopic images of 300 km long. In all cases, minimum area is 25 sq.km. For archive imagery and 100 sq.km. For tasked imagery, with a width of at least 5km in any direction for tasked imagery (500m for archive imagery). The user can select acquisition according to the area; 5 acquisition modes are there Target, Strip Mapping, Tri-stereo, Corridor and Persistent Surveillance.

The area of interest of the proposed work is from Kota kina balu in East Malaysia where the vegetation or trees near power poles are monitored. This area of interest comprises 10 km square. The total power poles in this area are 72 started from pole number 23 to 128. The Pleiades satellite sensors (1A and 1B) are used to scan the area of interest with a different date. The first stereo image acquisition is obtained on 27 July 2014. The images are panchromatic and multispectral and second stereo images were acquired on 9 September 2014. The stereo images contain the power transmission poles and whole right of way contains fifty-one (51) poles in both stereo images as shown in Figure 3.2. The stereo images of the area of interest have some clouds, some urban area and most of the region based on the jungle. The characteristics of stereo images of the area of interest showed different parameters associated with the stereo images. The Table 3. 1 has information regarding a number of bands, incident angle, across track angle, along with track angle, solar evaluation, cloud coverage, acquisition date and time for panchromatic as well as multispectral images. The Table 3. 1 shows the data taken from the same area of interest with different date and time. The propose of acquisition of data with different dates is to evaluate the depth or height level using satellite stereo images to monitor vegetation or tree near or under the power transmission power poles.

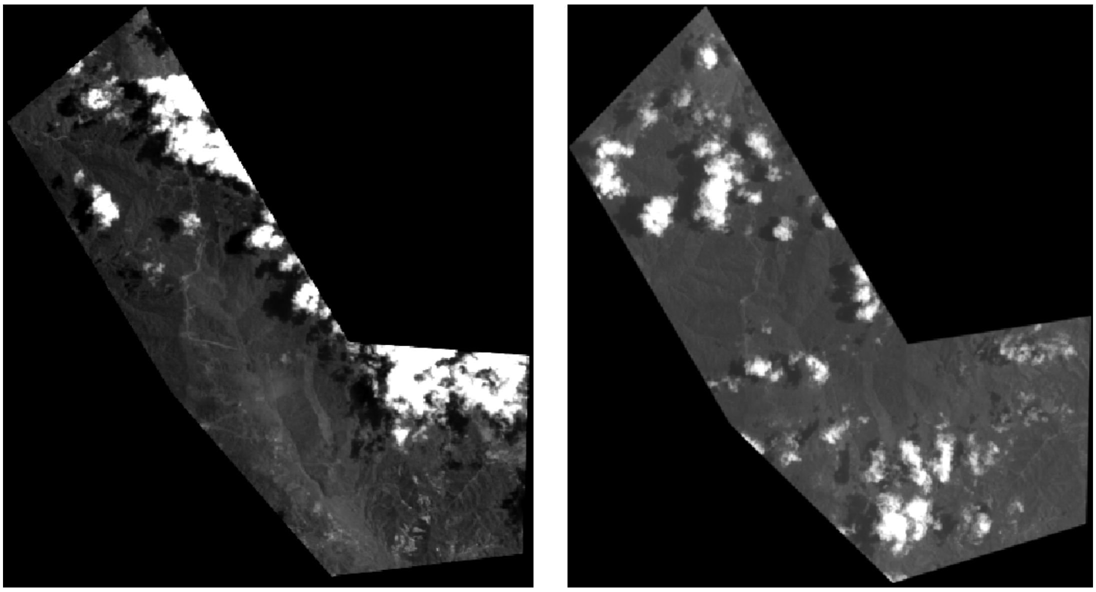
Table 3. 1: Dataset for the acquisition of panchromatic Pleiades satellite stereo images at date 3-9-2014 and 27-6-2014.

Sensor Name	**Band info**	**Incidence angle**	**Across track Angle**	**Along-track angle**	**Solar azimuth**	**Solar elevation**	**Cloud Cover**	**Weight**	**Acquisition date and time**
PHR 1A	Panchromatic	24.6	-2.9	12.6	83.6	70.7	19.7	1449	2014-09-03 0257:59.9
PHR 1A	Panchromatic	25.6	-16.5	-19.5	83.6	70.8	17.3	1450	2014-09-03 02:5706
PHR 1B	Panchromatic	22.9	-16.2	-16.7	54.6	65.2	6.7	563	2014-07-26 02:57:25
PHR 1B	Panchromatic	24.9	-21.3	8.6	54.4	65.4	7.7	563	2014-07-26 2:56:43



(a) (b)

Figure 3.1: Left and right Satellite stereo images, (a) Left Satellite stereo image taken at 03-09-2014, (b) Right Satellite stereo image was taken at 03-09-2014.



(a) (b)

Figure 3.2: Original dataset obtained using Pleiades satellite stereo sensors of the area of interest, (a) Left satellite stereo image taken at 26-6-2014, (b) Right satellite stereo image was taken at 26-6-2014.

- Segmentation of Satellite data

The stitching of all segmented images as shown in Figure 3.3. The segmented images based on power transmission poles are shown in Figure 3.4. After cropping, the satellite images have been shown with an indication of their power poles labeling and locations. Each segment contains eleven power transmission poles of the area of interest. Purpose of cropping satellite images is to reduce the size of the big satellite image. The cropping has been done automatically by providing GPS coordinates and location of power transmission poles.

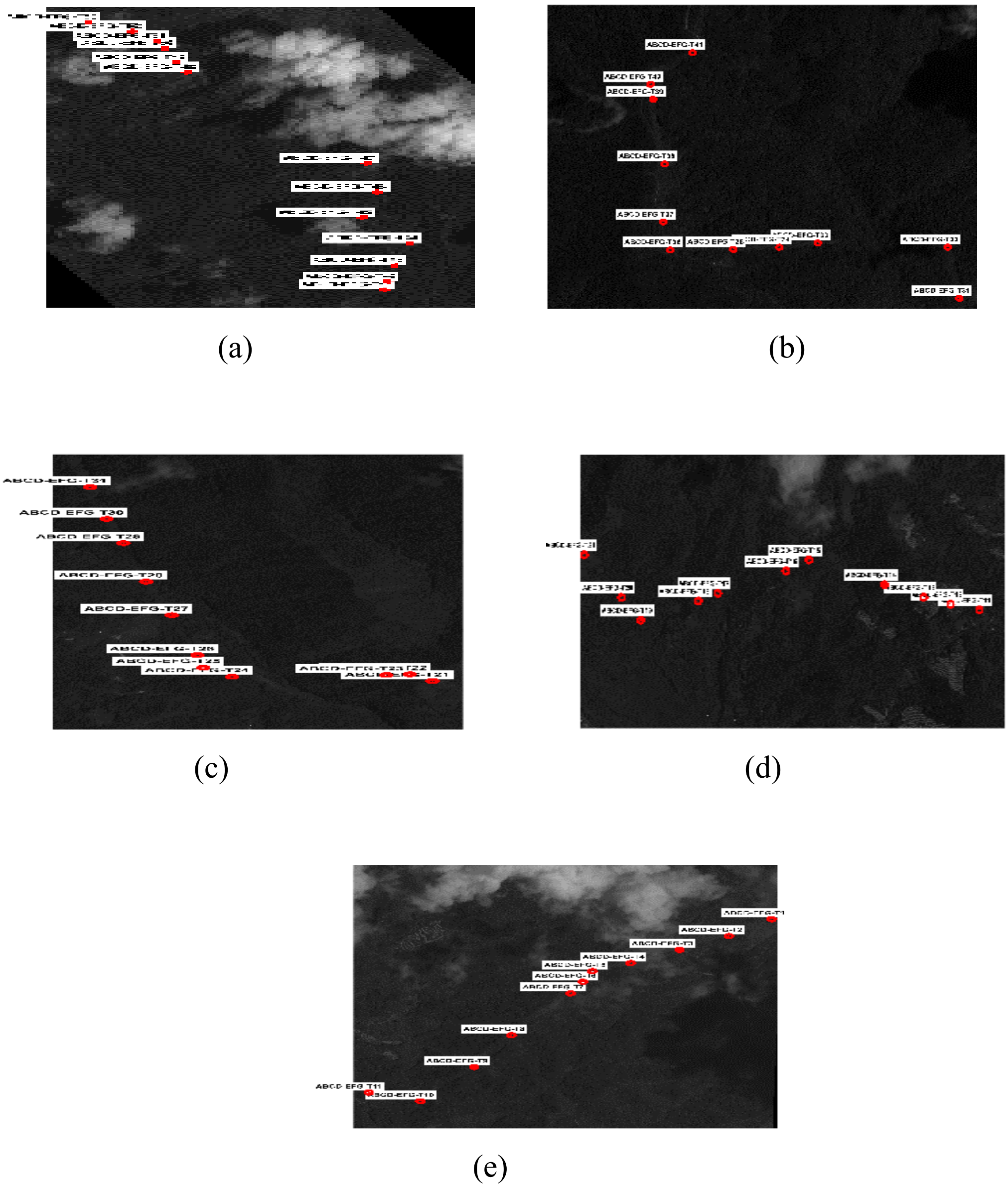
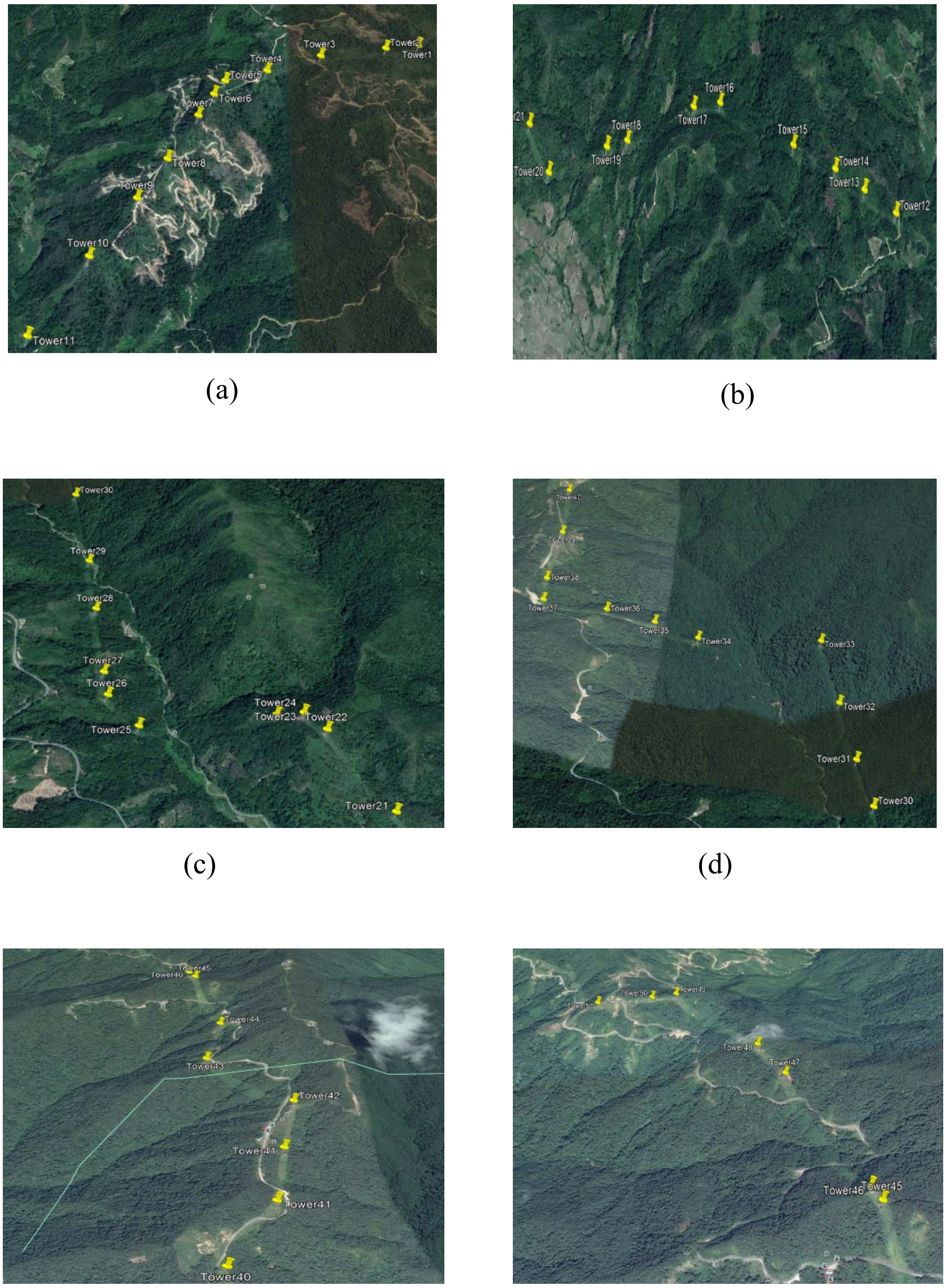


Figure 3.3: Stitching of all segment of the original dataset obtained using Pleiades satellite stereo sensors of the area of interest.



(a) (b)

(c) (d)

(e) (f)

Figure 3.4: A segment of the original dataset obtained using Pleiades satellite stereo sensors from the area of interest: (a) segment one contains (1-11) towers, (b) two contains (11-21), (c) three have (21-31), (d) four has (31-41) and (e) segment five consists of (41-53) number of towers.

3.2.2 UAV Stereo Images

The UAV can fly 72 km per hour and the battery timing is usually 3 hour to maintain the flight. The area of interest to monitor the vegetation or trees near power transmission poles is a village named Tambunan, Sabah East Malaysia. It's very challenging area due to bad weather conditions such as weather changes within 5 mints, always remain cloudy environment and most of the time is the rainy season. The vendor experienced the only best window time is 6.30 to 7.30 mornings and data collection should be based on the season. The best season is June to August in every year. The ground sample distance (GSD) is varied and depends on the height of the UAV. For the GSD value of 15 cm, the height will be of 700 meters. This is the normal range to collect the data. In our design experiment 700 meters height used for 15 km span area in a square kilometer. The whole area comprises 72 power poles, dense jungle, mountain, hilly area and non-uniform terrain in the right of way. The first UAV image was collected in June 2014 of the area of interest as shown in Figure 3.5 and the second image was collected in July 2015 as shown in Figure 3.6.

The UAV based system can acquire a large number of high-resolution images in a very short time. For example, in this case, 300 images were acquired at 0.15 cm resolution in only 30 minutes of flying time over a one-day and total round trip area was around 10 km. The stereo images taken from UAV have sufficient displacement and large disparity range, due to height constraints; these images have less resolution as compared to other sensors usage. The one advantage to use UAV is its less cost as compared to other airborne sensors used for monitoring vegetation and trees. The drawback is the more noise due to camera shakiness, different controlling attitude and fixed sensor orientation at the fixed view of angle. These constraints are major drawbacks for collecting data using UAV sensors. Nowadays, the cost of UAV remote sensing platforms suggests business-oriented applications such as smart farming in which user can pay per services based on the timeline provided and frequency of the services used. The original data of the complete area of interest are shown in using a different number of classes is shown in Figure 3.7. The training images based on various number of prominent objects (trees, power poles, roads and buildings, power lines, small grass) are shown in Figure 3.8.

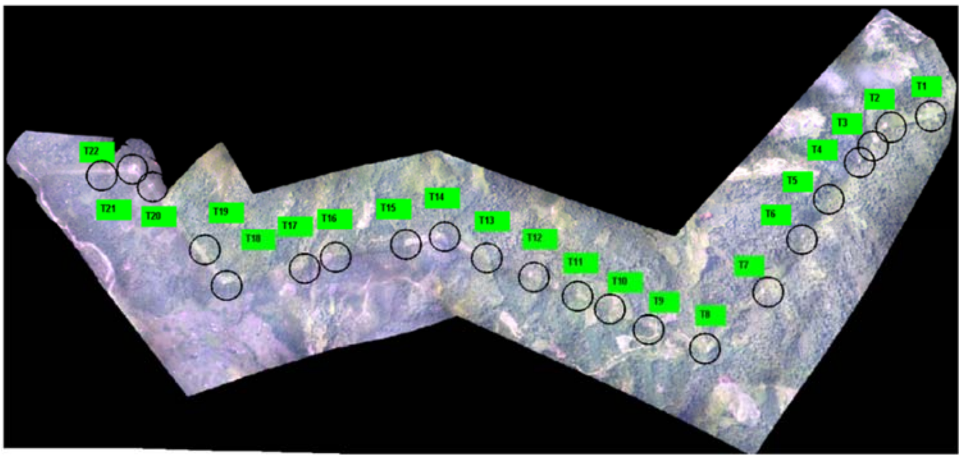


Figure 3.5: Original UAV data of area of interest

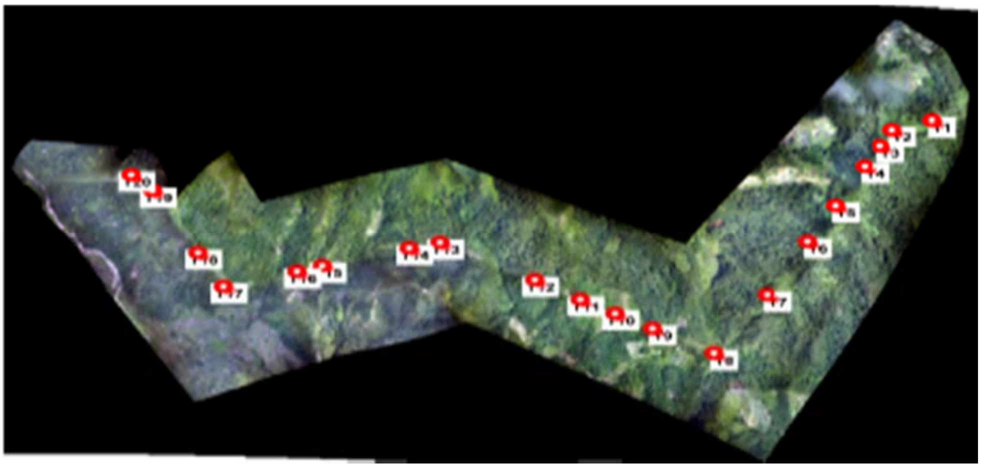


Figure 3.6: Original UAV data of area of interest

Figure 3.7: UAV dataset with three cropped cases (trees, power poles, buildings, power lines, small grass, vegetation).

➢ Segmentation of UAV Data

The number of samples for each class as shown in Figure 3.8. The data consists of four classes (trees, vegetation, roads, and building). The segmentation was selected due to prominent features.

Figure 3.8: The training samples of UAV images.

3.3 State-of-the-Art Stereo Matching Algorithms

The existing stereo matching algorithm has been used to estimate the disparity map. These algorithms can be classified into local and global stereo matching algorithms. The existing stereo matching algorithms have been used for the comparison with a proposed algorithm for disparity map estimation. These stereo matching algorithms are categorized into two main types: The local method and global methods. The local methods are further divided into the area based and feature-based stereo matching methods. The area-based methods estimate disparity using stereo images based on the fixed and adaptive window in a certain disparity range. The feature-based method is further divided into block matching with a cost function and block matching with energy minimization function. The block matching with cost function technique has been used traditional block matching cost functions to estimate disparity map using stereo images. These traditional cost functions are the sum of absolute difference (SAD), the sum of squared difference (SSD), normalization cross-correlation (NCC). The other feature-based technique used energy function to minimize the cost of stereo matching for disparity map estimation.

The global method used energy minimization function along with local methods to estimate the disparity map. The dynamic programming can have produced better disparity map using scanline search and estimate better disparity map due to strong corresponding between stereo images. It can find the minimum path cost from the top left corner of the 2D matrix to the right bottom corner for disparity estimation. The graph cut algorithm used three different methods in visual corresponding problems. The first is based on energy minimization using Potts model, the second method used stereo with occlusion and third method used multi-scene reconstruction. The most widely used method in graph cut is the energy minimization with Potts model. The graph cut algorithm estimate disparity map using the local method and further used energy function in scanline globally to optimize the disparity using stereo images. The cooperative and belief propagation stereo matching algorithms produced better disparity using the message passing algorithm between stereo pixels labeling in stereo images. The detailed comparison between existing state-of-the-art stereo matching algorithms and proposed algorithms on existing stereo images has been provided in Appendix B. These methods used for comparison with a proposed algorithm for disparity map estimation which further used in height estimation using aerial imagery.

3.4 Proposed Framework

The UAV images acquired from our area of interest were huge in size such as more than 1GB size. To reduce the processing time, the images have been divided into five equal numbers of segments. Each segment contains ten (10) number of power poles. The segmentation is done using automatically cropping based on pixels locations considering the latitude and longitude as well. The stereo matching algorithms needed orthorectified images for disparity map computation, the orthorectification has been performed on each small segmented image. The stereo matching algorithm has been applied to the orthorectified images to calculate the disparity maps using our proposed stereo matching algorithms. After calculation of disparity map, the height of the object is estimated based on the correct estimated disparity map. The ground truth height is compared with estimated height using performance metrics, accuracy, sensitivity, and recall. The flow diagram of our proposed method is shown in Figure 3.9.

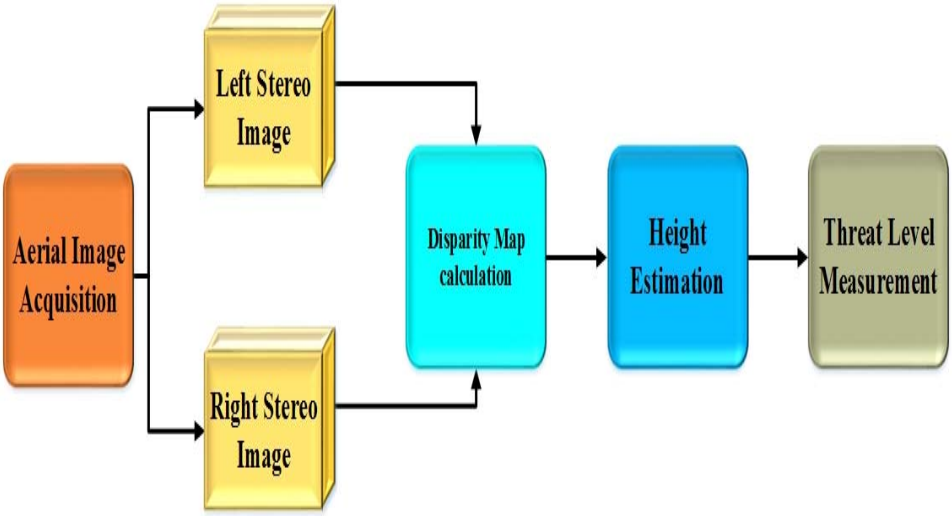


Figure 3.9: Proposed framework based on image aerial stereo images (satellite and UAV) for height estimation.

The flow of proposed method with step by step processing is shown in Figure 3.10. The following steps are also explained below.

Step1: Aerial image acquisition has been performed based on satellite and UAV sensors.

Step 2: After aerial image acquisition, the aerial stereo images have been received from the area of interest.

Step 3: Due to a large area and aerial stereo image size, the aerial stereo images are divided into a number of segments, there is fifty-one (51) number of power transmission poles in the selected area of interest. For the further computation, the complete area has been divided into five (5) number of segments and each segment contains ten (10) number of power transmission poles. After segmentation of stereo images, the ortho-rectification has been performed on each stereo pair for stereo matching estimation. The complete detail of orthorectification process with the concrete example is given in appendix D.

Step 4: The proposed stereo matching algorithms are applied on each segment to compute the disparity map. The designed disparity map is inversely proportional to depth or height of power transmission poles, vegetation, and trees. The proposed stereo matching algorithms are based on sparse representation and Convolutional Neural Network (CNN). The results are compared with the existing stereo matching algorithms (Graph-Cut, Dynamic Programming, Belief Propagation, Semi-Global Matching, area-based matching; the sum of absolute difference, the sum of squared difference, normalized cross-correlation).

Step 5: The depth or height of vegetation and trees near power transmission poles measured using designed disparity map based on triangulation method using proposed stereo matching algorithms.

Step 6: The estimated height and distance values are quantified into three different threat levels based on standard distance and height measurement values. These three threat levels are called high, medium, and low threat. The number of threats is found and monitor in such area which has high threat level in order to cut the trees or vegetation near power transmission poles.

The detail of each component explains in the next sections especially the proposed algorithm and their related functional components. After image rectification, the proposed stereo matching algorithms based on sparse representation and convolutional neural network (CNN) are implemented to measure the disparity map. These two algorithms may provide the feasible solution for disparity map computation. They could be used in real time applications due to their better accuracy and less computational complexity. The detailed discussion of each component used in proposed algorithms with their mathematical derivations is given in the below sections. The performance parameters also discussed with details in below section (Section 3.9).

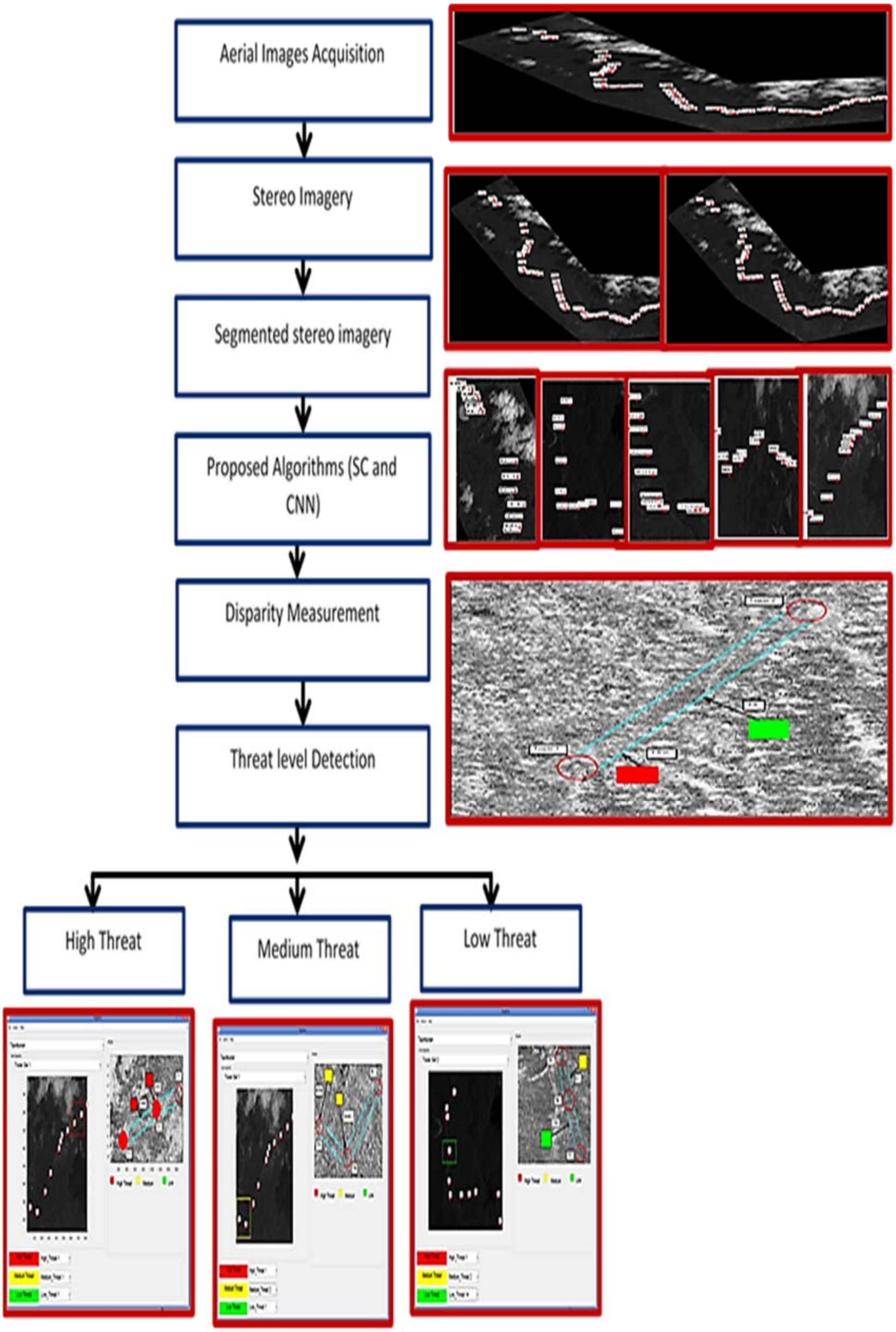


Figure 3.10: The step by step execution of proposed method based on proposed algorithms.

3.5 Proposed Disparity Map Estimation based on Sparse Representation

Recent years have observed an increasing interest in the search for sparse representations of signals [97]. A sparse signal expansion is a signal model that uses a linear combination of a small number of the elementary waveforms selected from a large collection to represent or approximate a signal. Such expansions are of increasing interest in signal processing with applications ranging from source coding [74] to de-noising [98], source separation [99] and signal acquisition [95], [100]. Using an over-complete dictionary matrix $D \in R^{N \times M}$ that contains an M prototype signal-atoms for columns,$\{c_j\}_{j=1}^{M}$, a signal b can be represented as a sparse linear combination of these atoms. The representation of b may be either exact $B_1 = D\alpha_1$, or approximate, $B_1 = D\alpha_1 + \varepsilon$, satisfying $\|B_1 - D\alpha_1\| \leq \varepsilon$.The vector $\alpha_1 \in R^M$ contains the representation coefficients of the signal B_1.

If $M < N$, and D is the full rank matrix, infinite number of b solution is available to search for a vector b is optimizing a certain sparsity measure. The solution with the fewest number of nonzero coefficients is certainly an appealing representation. This sparsest representation is the solution of either

$$(T_0) \qquad min_D \|\alpha_1\|_0 \quad \text{subject to} \quad B_1 = D\alpha_1 \tag{3.1}$$

$$(T_{0,\varepsilon}) \qquad min_D \|\alpha_1\|_0 \quad \text{subject to} \ \|B_1 - D\alpha_1\| \leq \varepsilon \tag{3.2}$$

The interest is to solve the exact solution as shown in equation (3.1 and 3.2). The proposed method is applied to stereo images for calculation of disparity maps. The Figure 3.11 shows the block diagram of the proposed framework. First, stereo images are preprocessed and extraction of similar patches was carried out by cropping smaller size images of the large satellite stereo images. The small patches are selected based on the sparse representation in both stereo images, further, the ridgelet transform dictionary is used in the sparse coefficient approximation. These coefficients were compared for two similar patches between stereo images. The sparse coefficient calculated from the first stereo image based on the small patch using an optimization algorithm like orthogonal matching pursuit (OMP). The sparse coefficients also computed from the second stereo image based on the sparse representation, after these two approximately closed coefficients were selected based on Euclidian norm distance. The disparity maps were constructed based on these selected coefficients using minimum Euclidian norm distance value. The depth map is the inverse of the disparity map.

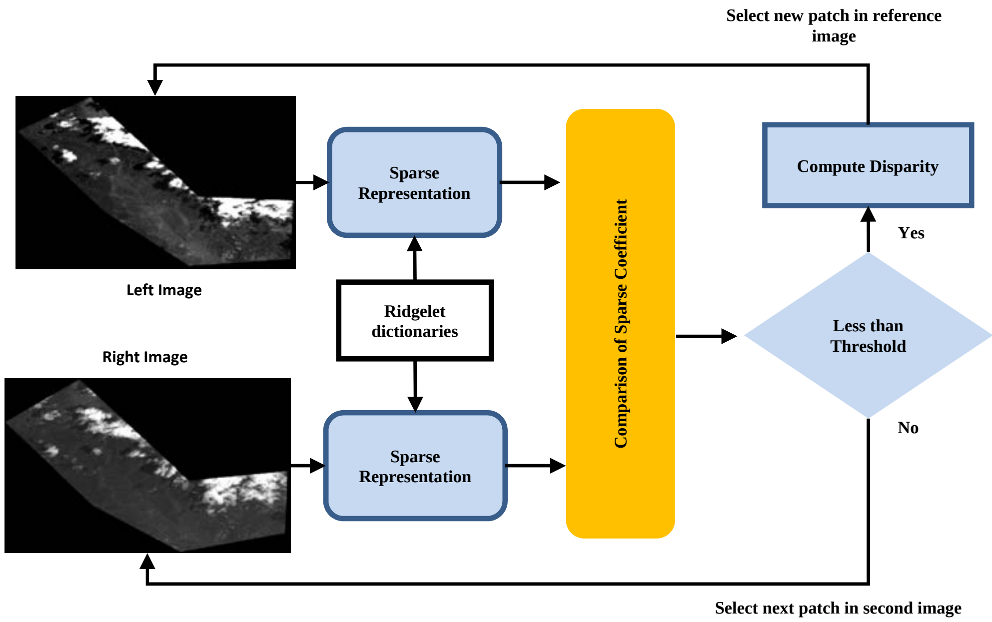


Figure 3.11: Flowchart of proposed Method.

In this work, there is proposed a new technique based on sparse coefficients for calculating the disparity map based on the over-complete dictionary. Normally different dictionaries are reported in the literature which is mostly fixed, e.g., Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Curvelets etc. The different analytical dictionaries have been computed, based on the experimental values. The dictionary based on the ridgelet base function is selected to compute the sparse coefficients due to its efficient capturing capability of unique points in the images. The ridgelet transforms stores the data distinct feature points of the reference image then it stores the distinct feature points of the second image, lastly it searches the similar coefficients based on prominent feature between two images. The dictionary size of proposed ridgelet transform is 1024. The patch size of 8x8 is extracted from the reference image. The extracted patch size of 8x8 is subtracted from the mean of the reference image as well as the mean of the second stereo image pair. Primarily, on the extracted patches from the two stereo images, the ridgelet transform is applied and then Orthogonal Matching Pursuit (OMP) algorithm is implemented for calculating the sparse coefficients. Next, the sparse coefficients of left image are compared with the sparse coefficient search in the right image. The extracted patch from the reference image based on dictionary coefficient is further compared with a similar patch of the same size in the right image within the disparity range (20-40).

The sparse coefficients are computed by the correspondence of fixed patch in the left stereo image and by shifting a fixed window in the right stereo image. The movement of the window in the right images scans on the horizontal line as the images have already been rectified. The similarity is achieved when the coefficient values are approximately similar. Further, the disparity map is generated based on sparse coefficient vectors generated after applying Euclidian norm distance technique. The distance between similar patches from two stereo images is estimated using Euclidian distance technique. At the end, the disparity maps based on this Euclidian distance has been computed. Further, depth maps are constructed based on the disparity maps. It is recommended to rectify the stereo images so that the correspondence point of both images lies on the same horizontal scan line. To find the corresponding point between the two images, the matching cost function can be used to measure the image similarity. Commonly used matching cost function simply compares the intensities of the images by calculating the absolute or squared differences.

Many other cost function matching methods are also available and reported in the literature. Generally, these cost functions matching methods can be classified into two main categories: local methods and global methods. Local methods assign the disparity to each position individually, whereas global methods are used to minimize the energy function of the entire scan line or even on the entire generated map. Local methods include feature-based and area-based methods. In the feature-based method, the features are extracted from the two images, e.g., corner, line and edge and then they are matched from the two images. The result of this method is a generation of varying sparse depth maps. On the other hand, area-based approach matches the neighboring pixel values, define the inside windows between the images and computes the correlation or sum of the squared differences (SSD). After finding complete correspondence points, a disparity map of the observed scene is generated. Disparity maps are further used to generate the 3D depths of the scenes, e.g., an application like robotic navigation, 3D television [133] etc. An example to compute the disparity map using sparse representation.

The initially the left satellite stereo image is cropped and a small image patch (8x8) is extracted from the cropped image. The cropped image has been orthorectified for further processing. Further, the orthogonal matching pursuit algorithm is applied based on proposed dictionary bases function to compute the sparse coefficients. The dictionary captures the optimal sparse coefficients. Next, the small patch from the right stereo image is extracted in the certain scanline within some range of displaced values. Similar to the left image, the cropped image patch from the right image is first orthorectified. Later, the OMP is applied to the right small image patch and the sparse coefficients are computed based on the fixed dictionary. This process is repeated several times in a certain scanline to compute the coefficients again and again. The loop will be stopped when two sparse coefficients are enough close to each other based on the minimum normative values of the two sparse coefficients. The process to compute disparity map based on proposed sparse representation technique is shown in Figure 3.12. The index of the difference of these pixel values is the displaced value. All displaced values are computed using other similar patches of each scanline from the left and right image until the whole images are scanned. The process is repeated and all displaced values are stored as disparity map.

The basic idea of the matching pursuit (MP) algorithm is to select the best matching atom from the over-complete dictionary to construct sparse approximation during each iteration to compute the signal representation residual and then to choose the best matching atom till the stopping criterion of iteration is satisfied. Many greedier algorithms such as the efficient orthogonal matching pursuit algorithm (OMP) subsequently have been proposed to improve the pursuit algorithm. The OMP algorithm is an improvement of the MP algorithm. The OMP employs the process of orthogonalization to generate the orthogonal direction of projection in each iteration. It has been verified that the OMP algorithm can be convergent in limited iterations. It is an excellent choice to imply the greedy strategy to approximate the solution of sparse representation with the lo-norm minimization. The algorithm used OPM optimization technique is shown in Table 3. 2 and OMP algorithm provides an efficient and fast solution for sparse coefficients extraction based on prominent feature capturing from the input image.

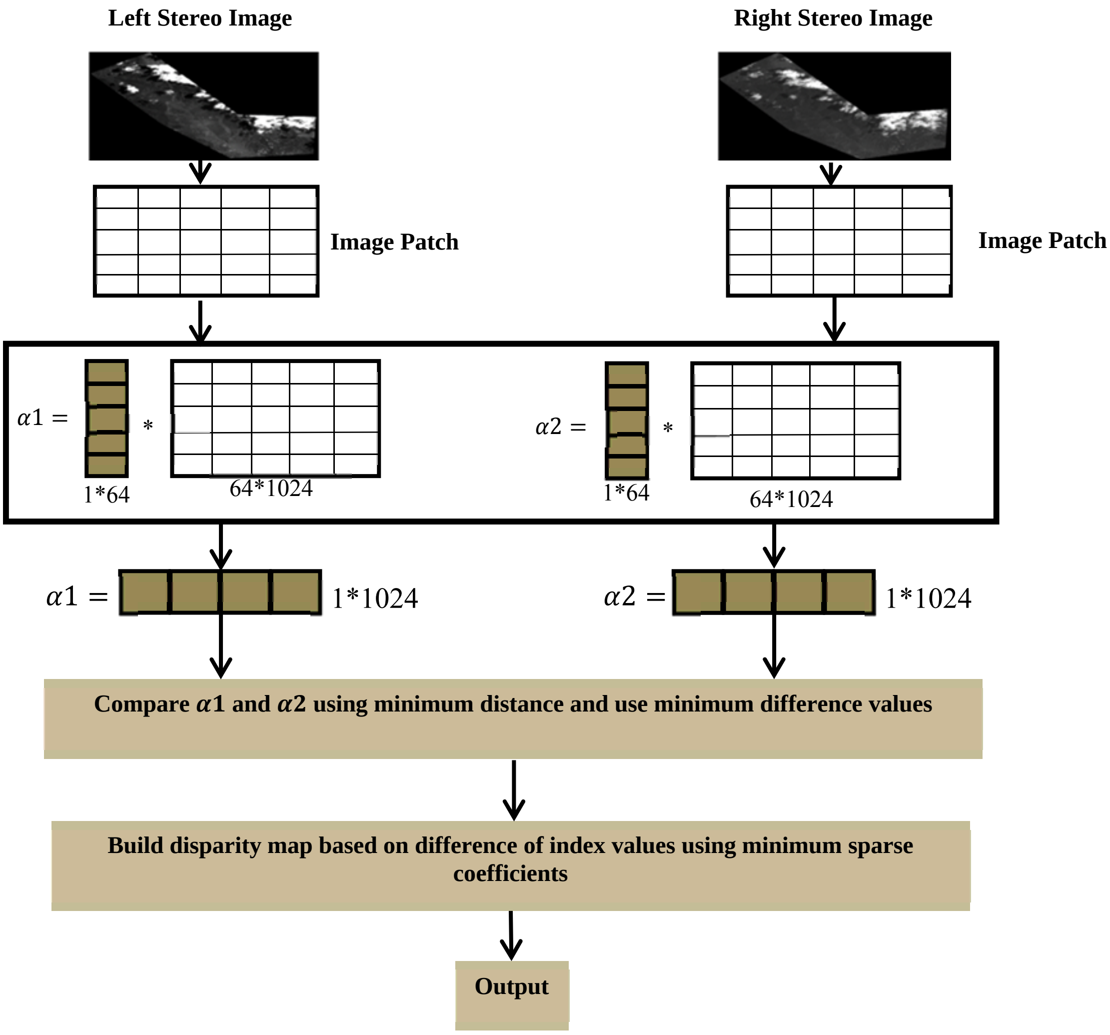


Figure 3.12: Disparity map using a simple example based on proposed method.

The proposed the ridgelet transform to model the sparse signal or images for measuring the sparse coefficients. The complexity of the sparse coefficients is based on the signal or image patch. The patch size is 8x8 for testing the image. The ridgelet transform is the efficient 2-D transform and can be used to store the multidimensional signal or image and it can be used for 1-D morelet wavelet function. The OMP used due to its simplicity and fast solution. So far it is seen that the obtained denoising algorithm based on OMP calls for sparse coding of small patches and averaging of their outcomes. The process can repeat the sparse coding technique and this time working on patches from the already denoised image once this is done, a new averaging should be calculated and so on. In this way, optimization is done for denoised and this method is totally based on the dictionary which is known.

Table 3. 2: Orthogonal Matching Pursuit (OMP) Algorithm

Algorithm. Orthogonal matching pursuit algorithm.

Task: Approximate the constraint problem:

$\alpha_1 = \dot{arg}\ min_{\alpha_1} \|\alpha_1\|_0$ s.t B_1 =Dα_1

Input: Input sample B_1 , Dictionary matrix D , sparse coefficients vector α_1.

Initialization: $t = 1, r_0 = B_1, \alpha_1 = 0, D_0 = D$, index set $\Lambda_0 = \emptyset$. Where $\emptyset$ denotes empty an set, τ is small constant. d_j is the dictionary elements stack as column vector.

While $\|r_t\| > \tau$ **do**

Step 1: Find the best matching sample, i.e. the biggest inner product between r_{t-1} and $d_j(j \notin \Lambda_{t-1})$ by exploiting

$$\lambda_t \dot{=} arg\ max_{j \notin \Lambda_{t-1}} |\langle r_{t-1}, d_j \rangle|.$$

Step 2:Update the index set $\Lambda_t = \Lambda_{t-1} \cup \lambda_t$ and reconstruct data set

$$\dot{D_t} = |D_{t-1}, d_{\lambda_t}|.$$

Step 3: Compute the sparse coefficients by using the least square algorithm

$$\breve{\alpha}_1 = \dot{arg}\ min \|B_1 - D_t \breve{\alpha}_1\|_2^2$$

Step 4: Update the representation residual using $r_t = B_1 - D_t \breve{\alpha}_1$

Step 5: $t = t + 1$

End

Output: D

This suggests that a library of many clean image patches and train a dictionary using DRT algorithm should formed. Numerous dictionaries have been trained by using DRT algorithm. An interesting property of dictionary learning algorithms is their robustness to noise, which means that learning from the noisy image leads to well behaved and noise free dictionary. The disparity map estimation using proposed algorithm is shown in Table 3. 3. This algorithm provided the disparity map based on sparse representation technique using aerial imagery. The sparse representation along with prespecified dictionary provided sparse coefficients feature vector using left image patch and similarly the sparse representation algorithm extract sparse coefficients based on right image patch, compute minimum value using both sparse coefficients vectors in certain range of displaced values and stored the index of the minimum corresponding between sparse coefficients. Repeat loop for all displaced values and stored all minimum values in this range. To compute the minimum values, save all minimum values and stored in 2D matrix, this 2D matrix is called disparity matrix or disparity map. The $\alpha_1(x_l, x_y)$ is the sparse coefficents vector extracted from left image patch and $\alpha_2(x_r, y_r)$ is the sparse coefficients vector extracted from right image patch having window of size $n \times n; n = 8$. Given a point $P_1 \equiv (x_l, x_y)$ in the left image I_1 and $P_2 \equiv (x_r, y_r)$ is the point in the right image I_2 located on the epipolar line. The window of size $n \times n$ is associated centered in (x_l, y_l) with the sparse vector coefficients $\alpha_1(x_l, y_l)$. Now given a potential match (x_r, x_r) in I_2, where $(x_r = x_l)$ and $(y_r = y_l + d)$ with $(d \in [d_{min}, d_{max}])$, it can construct another sparse coefficients vector $\alpha_2(x_r, y_r)$ for each d. Now, a function f is defined to measure the similarity between left and right patches to evaluate the correspondence is

$$f_1(d) = min_d \sum_j |\alpha_1 - \alpha_{2j}| \tag{3.3}$$

a similarity measure consisting of the minimum distance between the coefficients and $d = (y_r - y_l)$ and $x_l = x_r$ is called disparity and can be computed based on sparse representation algorithm is shown in the Table 3. 3.

The complete process used for disparity map estimation based on sparse representation technique is shown in Figure 3.13. In this example, the three (3) disparity ranges have been selected and disparity using proposed dictionaries is computed. The disparity is calculated on the similarity of two patches in stereo images based on

Euclidian distance technique. The technique based on disparity map has been optimized using sparse signal reconstruction algorithm.

Table 3. 3: Disparity computation based on sparse representation algorithm.

Description: Disparity map computation based on Sparse Representation Algorithm	
1	D dictionary matrix, α_1 , α_2, sparse coefficients vector from left image and right image, B_l and B_r are image patches from left and right images, ε is called sparsity coefficient. d is disparity, k is disparity range and H is disparity map. M, N are length of left and right images.
2	**FOR** $i = 1\!:M$ **do**
3	**For** $j = 1\!:N$ **do**
4	Extract sparse coefficients vector from the left and right image. (T_0) $min_D\|\alpha_1\|_0$ subject to $B_1 = D\alpha_1$ $(T_{0,\varepsilon})$ $min_D\|\alpha_1\|_0$ subject to $\|B_1 - D\alpha_1\| \le \varepsilon$ (T_0) $min_D\|\alpha_2\|_0$ subject to $B_2 = D\alpha_2$ $(T_{0,\varepsilon})$ $min_D\|\alpha_2\|_0$ subject to $\|B_2 - D\alpha_2\| \le \varepsilon$
5	**For** $k = 1\!:[d_{min}, d_{max}]$ **do**
6	Apply Euclidean distance technique to compute the similarity between sparse coefficients α_1 and α_2 vectors. $f_1(d) = min_d \sum_k \lvert\alpha_1 - \alpha_{2k}\rvert$
7	A similarity measure consisting of the minimum distance between sparse coefficients. Compute disparity based on the similarity between sparse coefficients vector **if** $d \ge d_{min}$ **do** $d = (y_r - y_l)$ and $x_l = x_r$
8	**END IF** **END FOR**
9	Disparity map based on disparity computation $H(i,j) = d$
10	**END FOR** **END FOR**

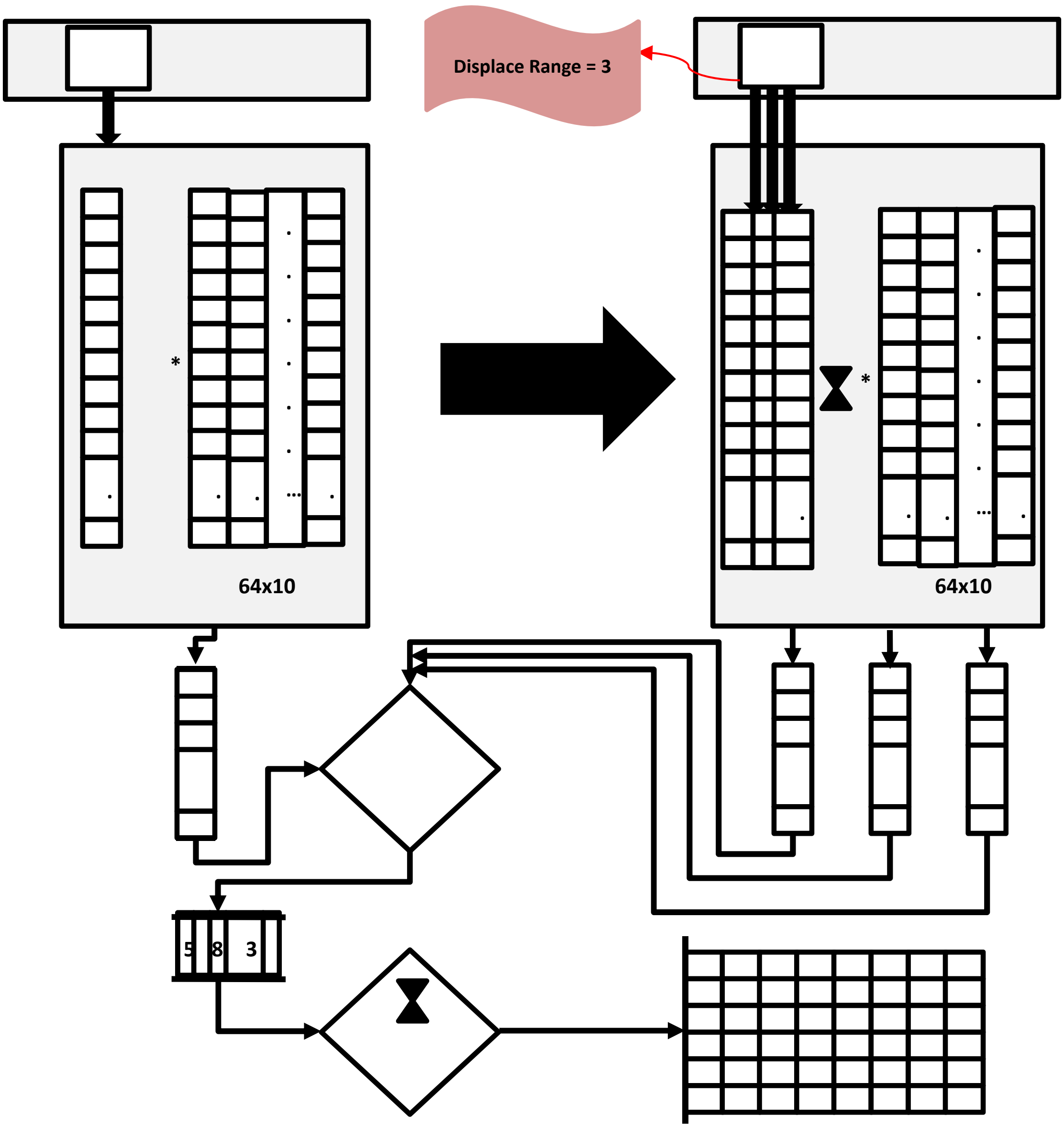


Figure 3.13: Disparity map computation based on sparse representation using proposed dictionaries.

3.5.1 Proposed Dictionaries

There is two dictionaries are proposed based on DTT and DRT. These dictionaries are pre-specified and provide efficient sparse signal representation in sparse space. The details of these proposed dictionaries are given in the following section below.

3.5.2 Discrete Tchebichef Transform (DTT)

As an alternative to the Discrete Cosine Transform (DCT), the Discrete T-Chebyshev Chef Transform (DTT) is chosen for the image compression algorithm (similar to JPEG). DTT has comparable properties to DCT, the only differences being lower computation time and higher energy compression rate based on the recurrence relation set. DTT derives from the popular discrete class of Tchebichef polynomials and represents a new version of the orthonormal transform and are widely used in applications for image compression and analysis [83]. The DTT is proposed as a fixed dictionary to optimize sparse coding based on Tchebichef moments that are produced in the basis matrix, which is similar to DCT's basis matrix as derived from trigonometric functions. For an image of size M × M, the forward discrete Tchebichef transform can be written in the following form:

$$F_{n_1 n_2} = \sum_{v=0}^{M-1} \sum_{u=0}^{M-1} f_{n_1}(v) f_{n_2}(u) f(v,u) \tag{3.4}$$

$$F(v,u) = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{M-1} F_{n_1 n_2}(v) f_{n_1}(v) f_{n_2}(u) \tag{3.5}$$

Where $n_1, n_2, v, u = 0, \dots \dots \dots, M-1$ parameters are used in forward as well as in reverse DTT function.

Now, the basis function of DTT is defined.

$$t(v, n_1) = f_{n_1}(v) \tag{3.6}$$

$$t(v, n_2) = f_{n_2}(u)$$

$f_{n_1}(v)$ is the v_{th} order of the Tchebichef moments. These can be defined using the following function over the discrete range$[0, M]$.

$$f_{n_1}(v) = n_1! \; \sum_{t=0}^{n_1} -1^{n_1-t} \binom{M-1-t}{n_1-t} \binom{n_1+t}{n_1} \binom{v}{t} \tag{3.7}$$

This expression can be solved by using the following recurrence relation

$$f_0(v) = \frac{1}{\sqrt{M}} \tag{3.8}$$

$$f_1(v) = (2v + 1 - M)\sqrt{\frac{3}{M(M^2 - 1)}} \tag{3.9}$$

For $f_{n_1}(v) n_1 = 2, \dots \dots \dots, M - 1$, this recurrence relation can be solved.

$$f_{n_1}(v) = (B_1 v + B_2) f_{n_1 - 1}(v) + B_3 f_{n_1 - 2}(v) \tag{3.10}$$

$$B_1 = \frac{2}{n_1}\sqrt{\frac{4{n_1}^2 - 1}{M^2 - {n_1}^2}} \tag{3.11}$$

$$B_2 = \frac{1 - M}{n_1}\sqrt{\frac{4{n_1}^2 - 1}{M^2 - {n_1}^2}} \tag{3.12}$$

$$B_3 = \frac{n_1 - 1}{n_1}\sqrt{\frac{2n_1 + 1}{2n_1 - 3}}\sqrt{\frac{M^2 - (n_1 - 1)^2}{M^2 - {n_1}^2}} \tag{3.13}$$

$$F_{n_1 n_2} = \sum_{v=0}^{M-1} f_{n_1}(v) \sum_{u=0}^{M-1} f_{n_2}(u) f(v, u) \tag{3.14}$$

$$g_{n_2}(u) = \sum_{u=0}^{M-1} f_{n_2}(u) \tag{3.15}$$

$$F_{n_1 n_2} = \sum_{v=0}^{M-1} f_{n_1}(v) f_{n_2}(v) \tag{3.16}$$

This expression is mathematically equal to the expression of discrete cosine transform. The pseudocode used in a proposed method to build over-complete DTT dictionary is shown in Table 3. 4. The DTT dictionary extracted prominent features from the stereo images and also minimized the noise due to compressed characteristics of that dictionary. These are the prominent features in this proposed DTT dictionary.

Table 3. 4: The dictionary bases function based on proposed DTT bases function.

	Algorithm. Dictionary based on DTT bases function
1	D is dictioanry size, a_1 , a_2 , a_3 are coefficients vectors, n_1 , n_2 are the index parameters, $i = n = 0,1,2,\ldots,N-1$, $j = m = 0,1,2,\ldots,M-1$. f_{n_1} and f_{n_2} are the Tchebichef polynomials.
2	**FOR** $each\ i = 1{:}n$ **do**
3	**FOR** $each\ j = 1{:}m$ **do**
4	Construct Tchebichef polynomials. $f_{n_1}(v) = (B_1\ v + B_2\ )f_{n_1-1}(v) + B_3\ f_{n_1-2}(v)$ $f_0(v) = 1/\sqrt{M}$ $f_1(v) = (2v + 1 - M)\sqrt{3/M(M^2-1)}$
5	**FOR** $each\ n_1\ = 1{:}n$ **do**
6	**FOR** $each\, n_2\ = 1{:}m$ **do**
7	Temp(n_1 , n_2)= $f_{n_1}(v).f_{n_2}(v)$
8	$D = temp(:)$
9	**ENDFOR** **END FOR**
10	$D(:,count) = temp(:)$
11	**END FOR** **END FOR**
12	Select size of dictionary $D_1\ (:,count) =\ D_1\ (:,size\ of\ dictioanry)$ $D = D_1$

3.5.3 Proposed Dictionary based on DRT

There has been much basis function or transform like discrete cosine, discrete wavelet, and ridgelet transform to compute the dictionary elements. The ridgelet transform has very high accuracy as compared to another transform like discrete transform, wavelet transform and etc. For instance, Fourier analysis and Wavelet analysis, ridgelet analysis is also being used for approximation of non-linear signals. A better approximation can be constructed by using a simple algorithm of N ridge functions [134]. By analyzing the continuous wavelet transform, its association is drawn with continuous wavelet transform.

In R_2 the continuous wavelet transform can be defined as

$$RI_f(a,b,\theta) = \int \vartheta_{a,b,\theta}(x) f(x)\, dx \tag{3.17}$$

Where the ridgelet $\vartheta_{a,b,\theta}(x)$ in 2-D are defined from a wavelet-typed function in 1-D $\vartheta(x)$ as

$$\vartheta_{a,b,\theta}(x) = a^{-1/2}\ \vartheta\left(\frac{(x_1 cos\theta + x_2 sin\theta - b)}{a}\right) \tag{3.18}$$

where ridgelet has the following parameters, scale(a), location parameter (b) and orientation (theta). While in R^2 continuous wavelet transform can be defined as

$$\vartheta_{a_1,a_2,b_1,b_2}(x) = \vartheta_{a_1,a_2}(x_1)\vartheta_{b_1,b_2}(x_2) \tag{3.19}$$

Of 1-D wavelet

$$\vartheta_{a,b}(t) = a^{-1/2}\, \vartheta_{a,b}(t - b/a) \tag{3.20}$$

$$t = x_1 cos\theta + x_2 sin\theta \tag{3.21}$$

The CRT and 2-D CWT emerges to be similar except the replacement of point parameters (b_1, b_2) with line parameters (b, θ).These transforms are associated by [135]

Wavelet: ϑ_{scale} Point--------------Point

Ridgelet: ϑ_{scale} Line --------------Position

The significances show that wavelet analysis for isolated point singularities is very efficient for object representation. Similarly, the ridgelet analysis for object representation is very effective with singularities along the lines by concatenation of 1-D wavelet transform. The inspiration for using ridgelets transforms in image processing is becoming very tempting because singularities are frequently joined together along edges in the image [136]. The ridgelet transform is an efficient 2D transform that can be used to store multi-dimensional signals or images. For instance, in Fourier and Wavelet analysis, ridgelet analysis is also used to approximate non-linear signals. Enhanced approximations can be constructed by using a simple algorithm for ‘N’ ridglet functions. Ridgelet analysis for purposes of object representation is extremely effective due to singularities along lines by means of the concatenation of 1-D wavelet transforms.

This inspired use of ridgelets transforms in image processing is very attractive because singularities are frequently joined together along edges in the image. Thus, the proposed hybrid dictionary became the better choice for the construction of an over-complete dictionary to provide better approximations for sparse representation. The algorithm used to construct the hybrid dictionary with ridgelet-based functions. Dictionaries comprising ridgelet-based functions are over-complete and constructed by different scaling factors and basis functions that employ Ricker wavelet functions. Each loop illustrates different parameters such as scaling, translation, and rotation to select the basis for DRT dictionary functions. The temp function was used to store the basis of the dictionary for each column, and the threshold for a variance was set at 0.05 to normalize and scale dictionary atoms. The pseudo code to implement DRT over-complete dictionary is shown in Table 3. 5. The DRT based dictionary provides the better discriminative capability and provides efficient reconstruction in sparse signal reconstruction. This dictionary produced feasible solution in stereo images for disparity map estimation and height estimation. Therefore, this dictionary can be used for vegetation monitoring near power transmission lines and poles. The DRT dictionary has efficient representation of prominent feature extracted from the stereo images for disparity estimation. This dictionary construct efficient feature space based on stereo imagery and reconstructs better disparity map which further used in height estimation and distance computation. The DRT also has better characterization to capture the feature from right and left stereo image for better disparity estimation. Due to this characteristics, it achieved high accuracy as compared to other predefined dictionary.

Table 3. 5: Dictionary based on DRT bases function.

	Algorithm. Dictionary based on hybrid Ricker Wavelet Basis Function
1	D is dictioanry size, S scale factor, T translation parameter , θ is the rotation parameter, M_1 number of dictionary atoms, N_1 size of dictionary
2	**FOR** $each\ Scale\ S = 1{:}\ M_1$ **do**
4	**FOR** $each\ translation\ T = 1{:}M$ **do**
5	**FOR** $each\ rotation\ \theta = -\pi{:}\ \pi$ **do**
6	**FOR** $each\ n_1 = 1{:}M_1$ **do**
7	**FOR** $each\ n_2 = 1{:}N_1$ **do**
8	$Temp(n_1\ , n_2\ ) = \sqrt{S} \times \sin[\,n_1 \times cos\theta + n_2 \times sin\theta - T]e^{-t/2}$
9	$D = temp(:)$;
10	**ENDFOR END FOR**
11	$storeD(:,count) = temp(:)$

12	**END FOR**
	END FOR
	ENDFOR

13	Select bases have greater variance than a certain threshold $T_1 = M^{2\times0.05}$ Select size of dictionary $D_1\ (:,count) = D_1\ (:,size\ of\ dictioanry)$ $D = D_1$ **END FOR**

3.5.4 Dictionary Construction based on Sparse Representation

A dictionary is called ‘over-complete’ if the number of pixels in an image patch is less than the number of elements in the dictionary for any sparse representation of an image patch. This dilemma caused us to improve different over-complete dictionaries that hold more atoms than a signal's dimensions, which then guarantees signification for a broader range of signal phenomena. Hence, the vigilant efforts are made to solve the over-complete dictionary impasse and minimize the loss of promising properties provided by orthogonal transforms. Sparse redundant representation and modeling of data assume the ability to describe signals as linear combinations of a few atoms from a pre-specified dictionary. Such a dictionary is simpler and should permit speedy algorithms to assess sparse representation. The most important step, therefore, was how to construct such a dictionary and afterward make a dictionary choice that provided optimized solutions for the calculation of sparse coefficients.

Dictionaries used in the present study are based on functions derived from DCT, DWT, DRT, and DTT. Since it is more attractive to learn the dictionary from a given training dataset, missing patterns within fixed dictionaries complicate the process due to the randomness of patch selections. Moreover, the size of a block and a dictionary's learning capability is critical. In addition, the only possible recourse for image reconstruction remains with block averaging of patches. Hence, the extraction of sparse coefficients with image patches can be reconstructed using a specified dictionary (See: Figure 3.14). Every patch from both image and dictionary atom were randomly selected to provide sparse coefficients vectors that represented compressed data taken from the input patch or signal. In this manner, all sparse coefficients were constructed based on an input image and dictionary elements. The DCT-based dictionary performed well and captured smooth and regularly structured image patterns. It also simultaneously captured multiple patterns, although this is not shown in Figure 3.14 (a). The DWT based-dictionary captured singular points with regular patterns. Its transform basis is deemed a 'structured dictionary'; i.e., one that captures structural atoms. DWT's dc values were scattered among different rows and columns compared to the initial cells of the dictionary [See: Figure 3.14 (b)]. Furthermore, the DRT-based dictionary stored singular points in a well-ordered format in addition to lines [See: Figure 3.14 (c)] and also captured regular patterns via small image patches. These constructed dictionaries captured the prominent features from the stereo images for better disparity map estimation. The DRT produced better atomic decompositions of stereo imagery for disparity map and stored effective features for stereo matching.

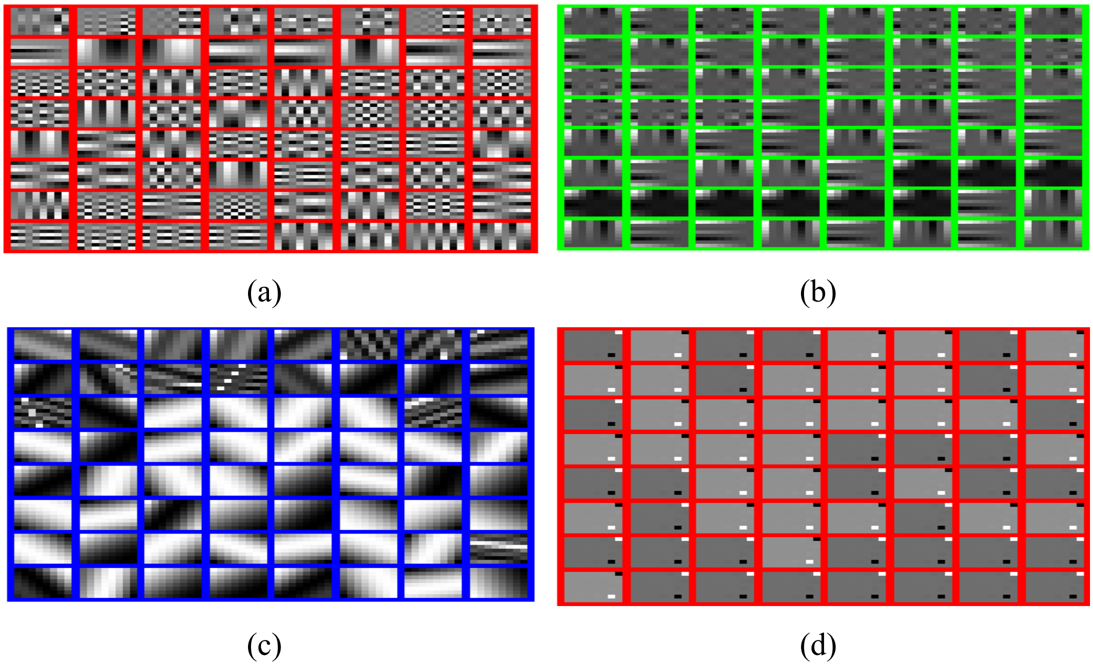


Figure 3.14: The bases function used in DCT, DWT, DRT, DTT using 8x8 image patch.

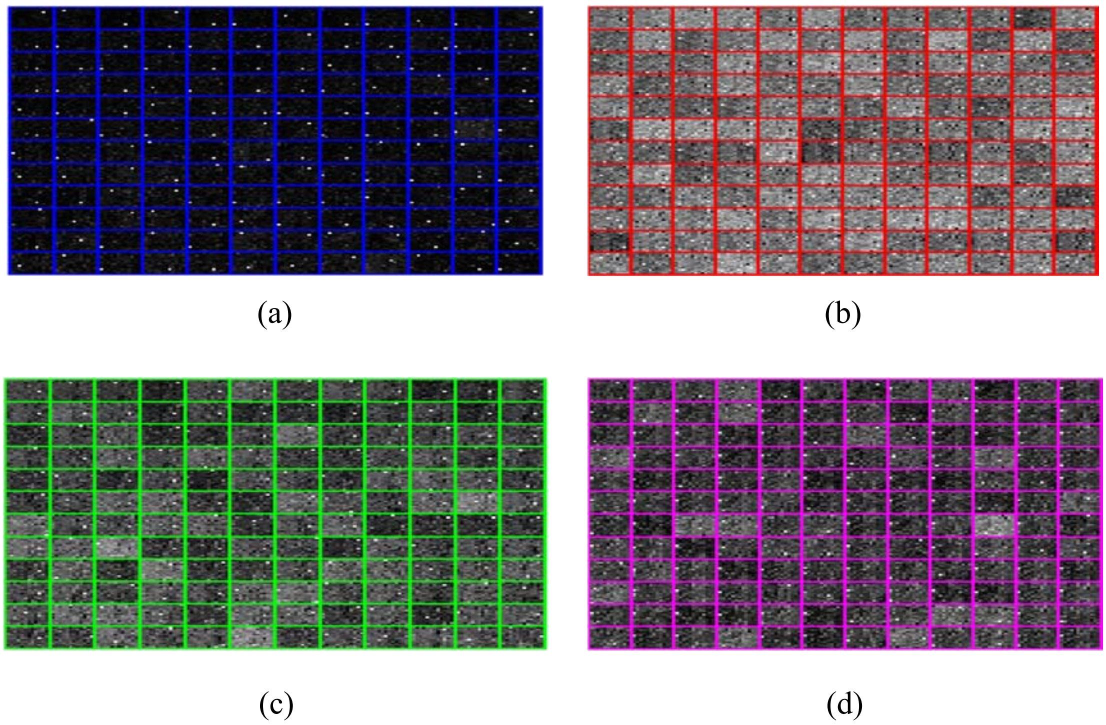


Figure 3.15: Dictionaries have size (16x16) image patch based on (a) (DCT), (b) DWT, (c) DRT, (d) DTT.

Different DRT-dictionary sizes were also implemented to assess accurate approximations of sparse coefficients for sparse representation. The DTT-based dictionary captured well-defined structural atoms in regular forms and also stored well-structured image features in regular patterns (See: Figure 3.14 (d)). The dictionary patches are represented using 16x16 DRT dictionaries as shown in Figure 3.15 for satellite images. The dictionary patches have the distinct pattern based the extracted patches. The sixteen (16) patterns are shown in the Figure 3.16 using left image and OMP optimization algorithm applied to store this pattern using dictionary bases function. The dictionary patches based on UAV images are shown in the following figure. The DRT dictionary used to construct these patterns and each pattern size is equal to 16x16. The dictionary over-complete elements show regular pattern based on DRT basis function. The visualization of sparse coefficients is shown in Figure 3.16.

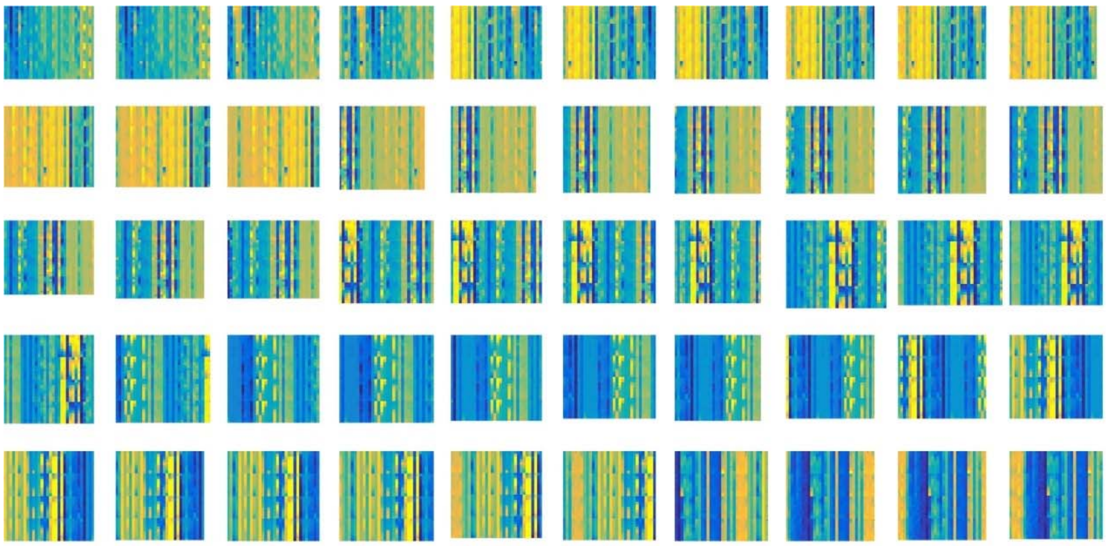


Figure 3.16: Dictionary elements used in DRT have size 16x16.There are 50 patterns shows in this and most of the dictionary patterns are well structured.

3.6 Proposed Disparity Map Estimation based on CNN Algorithm

CNN composed of convolutional layers followed by nonlinear and pooling layers for processing the input image patch with input weights. The convolutional layers are not computation efficient and fully connected layer add the system complexity by using more parameters. The second benefit of CNN used the pooling layer which perseveres the translational and rotational invariant features. A deep CNN composed of multiple layer architecture and extract the lower level features (edges, corners) further these features are combined to construct mid-level and high-level feature abstraction (objects). The last CNN layer is called fully connected layer and this layer further used for the classification task. The single CNN layer has a module which consists of convolution, non-linear activation, and pooling. Further, the convolutional layer composed of k feature maps f. Each feature map is measured by captivating the dot product between a local region x of size $m \times m$, with c number of channels, $x \in R^{m \times m \times c}$ and the kth filter w^k of size $n \times n$, $w \in R^{n \times n \times k}$. The feature map for the kth filter is $f \in R^{(m-n-1)\times(m-n-1)}$ calculated as in Eq. (3.30).

$$f^k{}_{ij} = \alpha(\sum_{c}\sum_{a=0}^{n-1}\sum_{b=0}^{n-1} w^k{}_{abc} x^c{}_{i+c,j+b}) \tag{3.22}$$

where σ is the non-linear activation function. The CNNs usually used activation functions based on a hyperbolic tangent [114] and rectified linear units (ReLU) [137]. Local Response Normalization (LRN) type of layer turns out to be useful when using neurons with unbounded activations (e.g. rectified linear neurons), because it permits the detection of high-frequency features with a big neuron response while damping responses that are uniformly large in a local neighborhood. It is a type of regularize that encourages "competition" for big activities among nearby groups of neurons. For every particular position (x, y) and kernel i that corresponds to a single 'pixel' output a 'filter' is applied, that incorporates information about outputs of other n kernels applied to the same position. This regularization is applied before activation function.

$$b^i_{x,y} = \frac{a^i_{x,y}}{(K + \alpha\,(\sum_j a^j_{x,y})^2)^\beta} \tag{3.23}$$

The filters w can be pre-trained by choosing random initialization with supervised fine-tuning or unsupervised feature learning algorithms. The k-means or auto-encoder models can be used in unsupervised feature learning algorithm. The pooling performed on feature maps by reducing the sample size based on the maximum or mean function over the local spatial region with the non-overlapping framework. The pooling layer for the kth filter $g \in R^{(m-n-1)/p \times (m-n-1)/p}$ is calculated as in equation (3.32)

$$g^{k}{}_{ij} = \max(f^{k}{}_{1+p(i-1),1+p(j-1)}, \dots, f^{k}{}_{pi,1+p(j-1)}, \dots, f^{k}{}_{pi,pj}) \quad (3.24)$$

where p is the size of the local spatial region and $1 \leq i, j \leq (m-n+1)/p$. The parameters and softmax classifier are trained using random initialization, stochastic gradient descent (SGD) algorithm with backpropagation trained network in fully connected layers. The stereo images based on CNN model shows the cascaded structure of each basic layers in Figure 3.18 and Figure 3.19. In this design, CNN consists of convolutional, non-linear, normalization, pooling and fully connected layers. The feature maps of each layer provide visualization capability in CNN design based on input UAV image. Similar, the feature maps of each layer based on CNN design as shown in Figure 3.17.The visualization capability of each layer with their feature maps shown clearly the discriminate powers of a feature in each layer. A method is proposed to compute disparity map using remote sensing images based on deep learning approach. The convolutional neural network is the main techniques used under the deep learning methods. The idea is to develop the feature space by extracting the small patches of the input image using moving windows. The proposed CNN based algorithm used to measure the disparity map as shown in Figure 3.20.

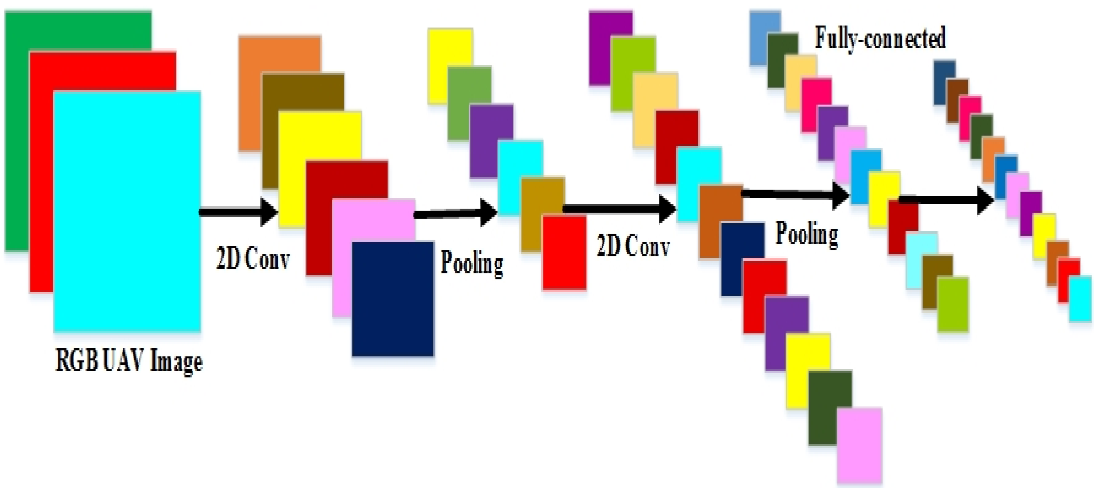


Figure 3.17: The CNN design layers for feature extraction

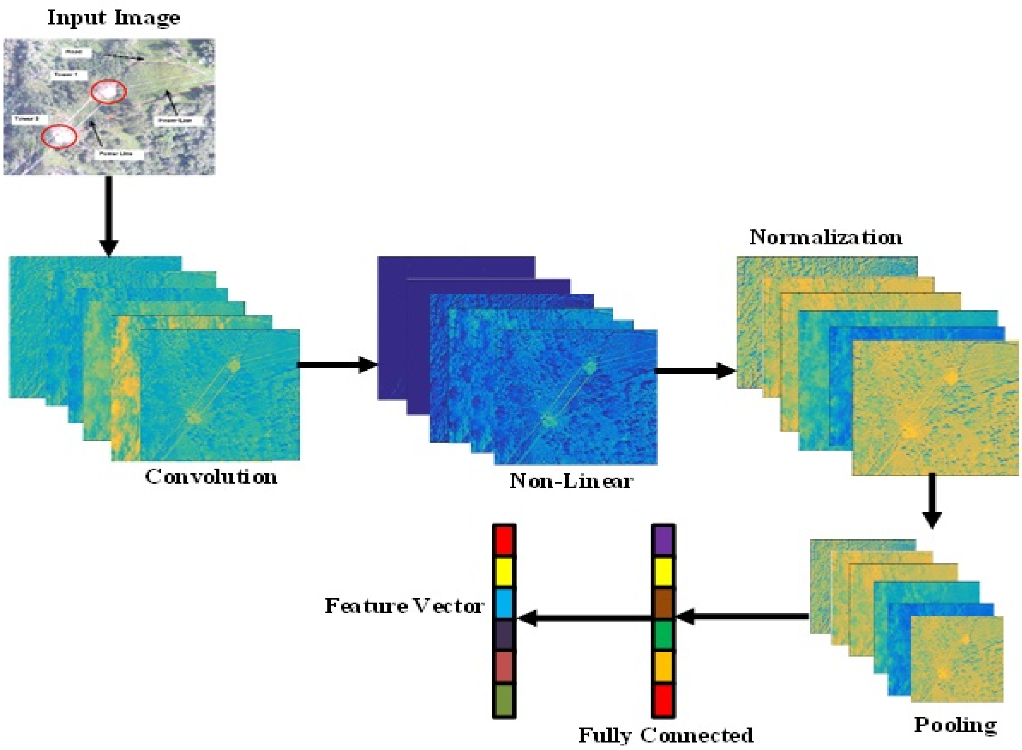


Figure 3.18: The CNN algorithm consists of a different number of layers based on the first input image.

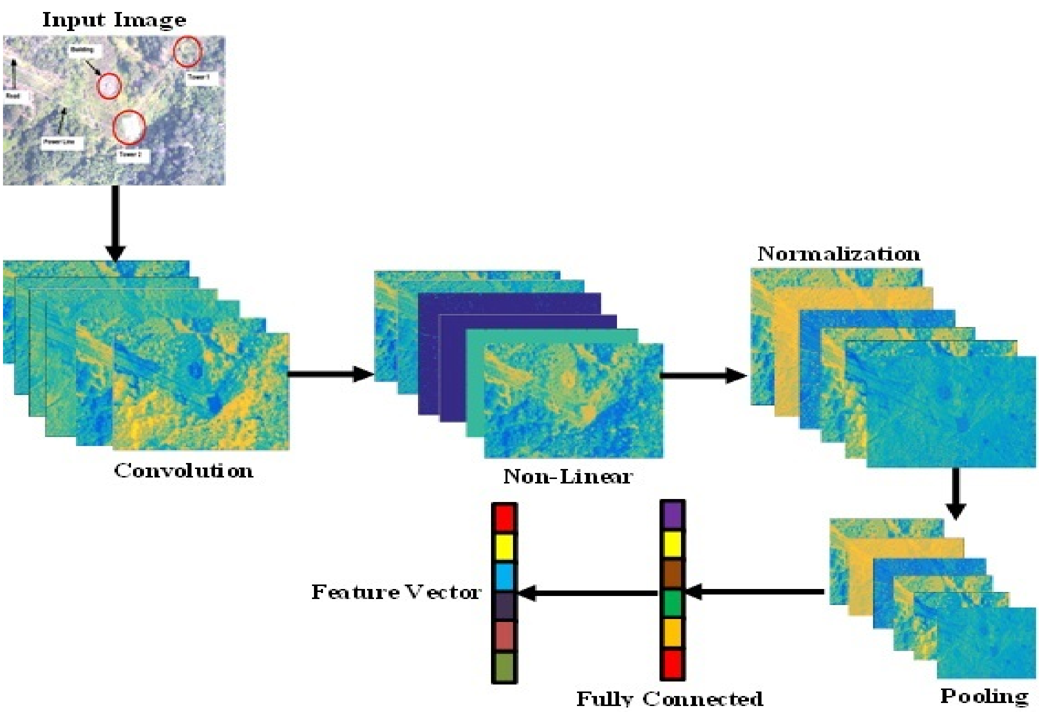


Figure 3.19: The CNN algorithm consists of a different number of layers based on the second input image.

Aerial stereo images are extremely useful for the computation of 3D height estimations. Our method proposes a deep learning CNN computation of a disparity map using remote sensing images that are processed by different data layers. A disparity map is used to compute the height of vegetation near power transmission poles by exploiting estimated vs. actual object height. The CNN algorithm was optimized with a number of hyper-parameters based on image patches. The Figure 3.20 outlines the complete process. Orthorectified stereo images were required to compute the disparity map by employing different layers within the proposed CNN algorithm. The resultant disparity map was then employed to measure the height of vegetation and trees near power transmission poles. The CNN algorithm's trainable base derives from small stereo patches used to extract spatial scale-invariant features of displaced values in a scanline based on orthorectified stereo images. An area of interest within stereo images was orthorectified solely to process these images for CNN training for the purpose of disparity map estimation. In this way, the disparity map is computed by scanning all the pixels values between left and right stereo images. The depth map is measured based on computed disparity using CNN algorithms.

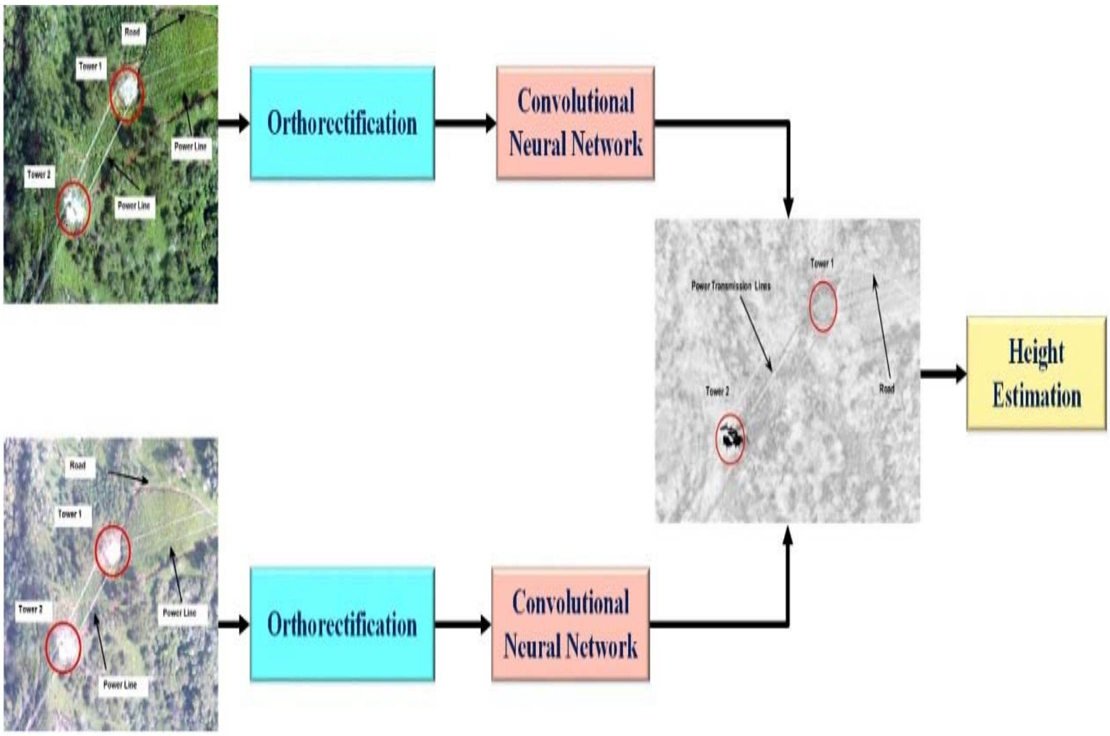


Figure 3.20: The block diagram to compute the height of towers using CNN algorithm.

Estimated height results were compared with actual tower height to measure accuracy, sensitivity, and recall. The first convolutional layer employs six random filters (receptive fields). After convolution of the input image, six feature maps are generated as output. This initial output then serves as input for the next step or pooling layer. The pooling layer down-samples the input using the max function. The next convolutional layer employs twelve (12) feature maps and the second pooling layer is then applied to all twelve and followed by a fully connected layer. At the network's final stage, fully connected feature vector lengths equal 192. The size of fully connected is 192 feature vector lengths as shown in Figure 3.21.

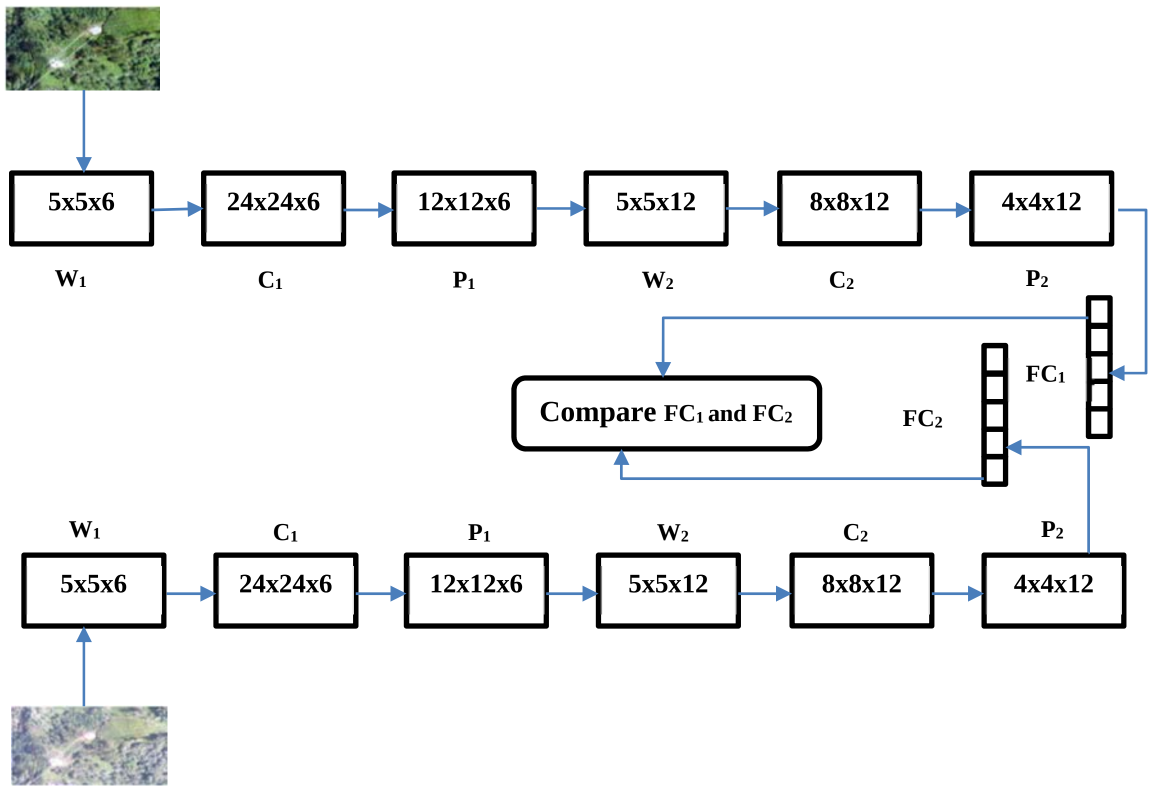


Figure 3.21: Convolutional layers using stereo images.

Figure 3.21 provides an outline of all layers (convolutional, two pooling and 1 fully connected) used to compute CNN feature vectors that are extracted from the fully connected layer for a specific patch representation taken from both stereo images. Similarly, an extracted patch from the right stereo image is applied to the CNN algorithm to compute the similarity value between both vectors using the minimum distance technique. Similarity values between fully connected feature vectors are computed using the Euclidean distance technique.

The concept here is to develop feature space by extracting small patches from the left stereo image and compare these with extracted patches from the right image by using a moving window. A 28x28 window size was used to extract patches from both images and then apply the CNN design to convert them into feature vectors. Initially, a window from the left stereo image is used as input for CNN measurement of feature vectors. A patch from the right image similarly follows. There is also need to select displacement values for a certain range of disparity that arises when moving the window pixel by pixel. For each pixel in right and left images, one right shift is performed to compute a feature vector. The similarity between extracted vectors is then compared using the Euclidian distance metric. If feature vectors are similar, the distance metric delivers a minimum value. All minimum values generated by both left and right images, within a certain range of disparity, are then stored as an 'index of disparity' between left and right images. Thus, the disparity map is computed by scanning all pixel values between left and right stereo images.

The disparity map's calculation can be explained as follows. Initially, orthorectified stereo images are extracted to compute feature space for objects based on small stereo image patches. The CNN algorithm is applied to each patch to extract feature vectors using the fully connected final layer of the algorithm. After feature vector extraction, minimum distance values are computed based on the similarity between two feature vectors by using the minimum distance technique. They are then stored in the 'minimum values index' as disparities produced by feature vectors. The Figure 3.22 illustrates the algorithm's configuration. Design specifications for different layers are shown in Table 3. **6**. The design for CNN1 used a 28x28 image patch with two convolutional, two pooling and one fully connected layer. Similarly, the CNN2 design had the same number of layers and parameters for the right image patch size (28x28). The small image patch used a small CNN network for feature extraction. The number of unique neurons in each convolutional and fully connected layer is shown. Note that the fully connected layer uses more neurons compared to convolutional layers. All layers configuration in CNN algorithm is shown in Figure 3.22. Table 3. 7 shows the proposed CNN algorithm's disparity map estimations. After patch extractions of equal size from left and right stereo images, the algorithm is applied to both. The convolutional layer then reconstructs images that are similar to original image features.

Table 3. 6: CNN Design specification used for height estimation.

CNN designs	Input Image	Number of convolutional layers	Number of pooling layers	Number of filters	FC Layer
CNN1	I_1=28x28	c_1=24x24x6	P_1=12x12x6	W_1=5x5x6	---
		c_2=8x8x12	P_2=4x4x12	W_2=5x5x6	1x192
CNN2	I_2=28x28	c_1=24x24x6	P_1=12x12x6	W_1=5x5x6	---
		c_2=8x8x12	P_2=4x4x12	W_2=5x5x6	1x192

This same layer generates spatial features with spatial filters by activating output values from previous layers. Feature vectors are then extracted from the final CNN layer by using these image patches (left and right) based on pooling layers. Pooling layers function progressively to reduce the spatial size of each representation, which lessens the number of parameters and computational requirements within the network, which also delimits overfitting. Pooling layers reorganize data taken from lower level layers to create abstract representations while also preserving translational and rotational invariant features. Each pooling layer operates independently on every depth slice input to spatially resize the data using the max operation. Pooling layer output is then converted to fully connected feature vectors by using the flatten technique to convert them into 1D feature vectors. These vectors are not spatially located and are visualized as one-dimensional. Fully connected 1D feature values from a particular patch in the receptive field of the input image are flattened vectors, devoid of spatial features. They provide the data for input pixels in feature space. These flattened feature vectors are shown in Table 1 as the fourth step of the algorithm. In step six, similarity depends on the input of flattened feature vectors for a certain range of disparity between left and right images.

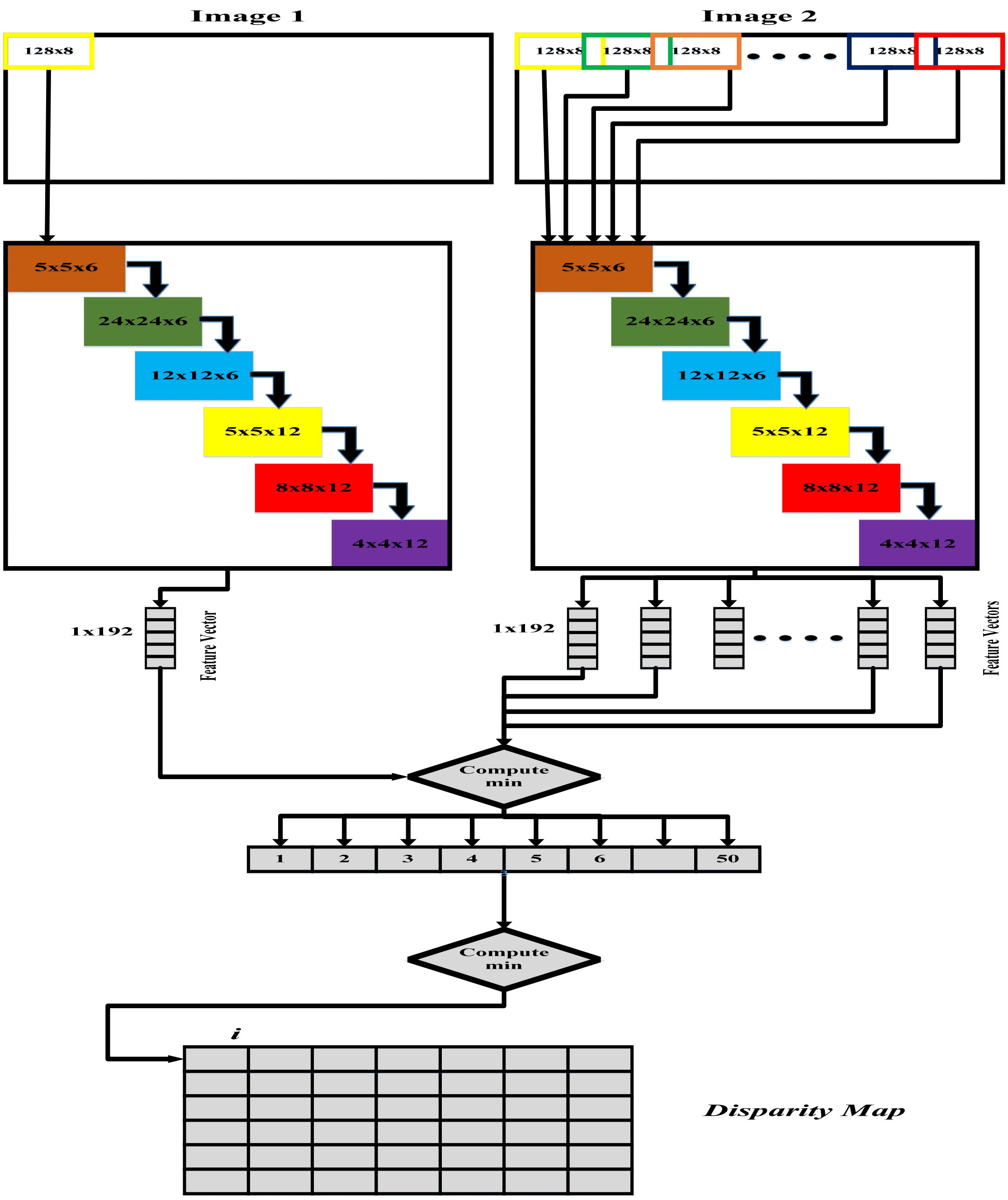


Figure 3.22: The flow diagram of disparity map estimation using CNN algorithm based on aerial stereo images.

Thus, CNN based feature vectors provide a strong correspondence between input stereo patches by capturing spatial feature patterns that were trained by the CNN algorithm's convolutional layer. The algorithm terminates at step seven with a certain disparity range $[d_{min}, d_{max}]$ when two feature vectors reach a minimum value using the scaline stereo image.

The Euclidian distance technique was employed to measure the similarity between feature vectors extracted by the final layer of the CNN algorithm, based on distinct image points between left and right stereo images. The minimum value between similar vectors demonstrates 'fill-up' similarity criteria. The index value corresponding to the minimum value and based on a specific disparity range is then stored. In step nine, all steps are repeated for calculations of all indices based on minimum values between scan lines (left and right) of stereo images to construct the disparity map.

Here, $c_1(x_l, y_l)$ is the CNN feature vector extracted from the left image patch as derived by the fully connected layer. Likewise, $c_2(x_r, y_r)$ is the CNN feature vector extracted from right image patch, of size $n \times n$ where $n = 28$. Given point $P_1 \equiv (x_l, x_y)$ from the left image (I_1), and $P_2 \equiv (x_r, y_r)$ from the right image (I_2), as located on the epipolar line, the associated window (size $n \times n$) centered in (x_l, y_l), with the CNN feature vector, $c_1(x_l, y_l)$. Now given a potential match (x_r, x_r) in (I_2), where $(x_r = x_l)$ and $(y_r = y_l + d)$ with $(d \in [d_{min}, d_{max}])$, the algorithm can then construct another CNN feature vector, $c_2(x_r, y_r)$, for each (d). The function (f) can be used to measure the similarity between left and right patches to evaluate correspondence as follows in equation (3.25):

$$f_1(d) = \min_d \sum_j |c_1 - c_{2j}| \tag{3.25}$$

Thus, a discreet similarity measurement consisting of the minimum distance between coefficients where $d = (y_r - y_l)$ and $(x_l = x_r)$, can be defined as the disparity, as computed and derived by the CNN algorithm. The CNN algorithm produced efficient solution for disparity map estimation and height estimation of vegetation and trees near power transmission lines and power transmission poles. It has a capability to apply in such type of application in real time environments. The proposed algorithm based on CNN provides excellent solution for threat estimation using images.

Table 3. 7: Proposed method based on CNN algorithm for stereo matching.

Description: Disparity map computation based on CNN Algorithm	
1	D dictionary matrix, c_1 , $c_{2,}$CNN feature vector from left image and right image, B_l and B_r are image patches from left and right images, d is disparity, k is disparity range and H is disparity map. M, N are length of left and right images.
2	**FOR** $i = 1{:}M$ **do**
3	**For** $j = 1{:}N$ **do**
4	Extract CNN feature vectors using last fully connected layer from left and right images. $$P_1 = \max({f_l}^k{}_{1+p(i-1),1+p(j-1)}, {f_l}^k{}_{pi,pj})$$ $$P_2 = \max({f_r}^k{}_{1+p(i-1),1+p(j-1)}, {f_r}^k{}_{pi,pj})$$ $c_1 = flatten(P_1),\ c_2 = flatten(P_2)$
5	**For** $k = 1{:}[d_{min}, d_{max}]$ **do**
6	Apply Euclidean distance technique to compute the similarity between sparse coefficients c_1 and c_2 vectors. $$f_1(d) = min_d \sum_k \lvert c_1 - c_{2k} \rvert$$
7	Compute disparity based on the similarity between feature vectors. **if** $d \geq d_{min}$ **do** $d = (y_r - y_l)$ and $x_l = x_r$
8	**END IF** **END FOR**
9	Disparity map based on disparity computation $$H(i, j) = d$$
10	**END FOR** **END FOR**

3.7 Disparity Map Generation

The depth map is computed from two stereo images by calculating the pixel-wise distance between the location of a feature in one image and its corresponding location in the second image, which generates a disparity map. Further, it gives a depth map because the pixels with higher disparities are closer to the camera, and those with smaller disparities are away from the camera.

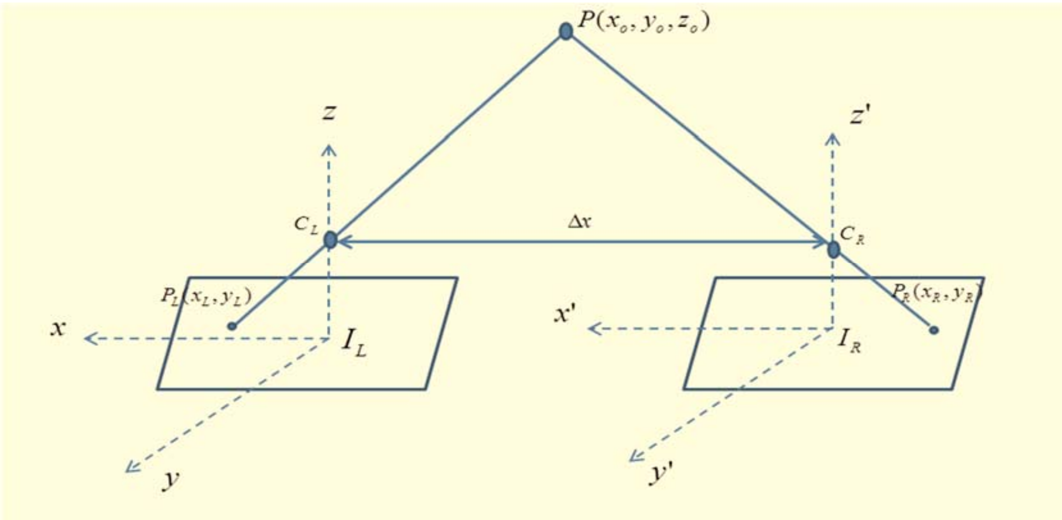


Figure 3.23: Stereo camera model [19].

In Figure 3.23, there are left and right camera images, where the left image has a center at 0 and right has a center at 0'. Therefore, 3D depth points can be calculated at coordinates (x_0, y_0, z_0). The following relation is from the above diagram [19]. By solving equation (3.26) and equation (3.27), the value of Zo is achieved. This value of Zo depends upon the value of the denominator factor which is called disparity value. Depth is inversely proportional to the disparity map as given in equation (3.28). In nadir using aircraft is the method for setting the camera position to acquire the image from the ground and position of the camera is set to the point straight down using the center of the field. The information is taken from the two stereo cameras displaced at a larger distance to ensure the good depth information of the objects. By using the available information of the camera separation, the height of the object is computed using the stereo overlapping region and can use displaced values for making an automatic map.

$$\frac{x_0}{x_L} = \frac{y_0}{y_L} = \frac{\lambda - Z_0}{\lambda} \tag{3.26}$$

$$\frac{x_0 + \Delta x}{x_R + \Delta x} = \frac{y_0}{y_R} = \frac{\lambda - Z_0}{\lambda} \tag{3.27}$$

Solving equation (3.26) and equation (3.27), the equation (3.28) is obtained.

$$Z_0 = \frac{\lambda + \lambda \Delta x}{x_L - (x_R + \Delta x)} \tag{3.28}$$

For height estimation, aerial stereo images are the better choice due to high resolution and low cost as compared to other remote sensors. The system model used to measure the height of vegetation and trees and distance between vegetation, trees and power lines explain in the next section with their analytical relation. The height or depth of vegetation or trees near power transmission lines based on aerial stereo images has been computed based on estimated disparity map calculated by proposed algorithms. The Pleiades satellite sensor was employed to take stereo photos of an area of interest in Sabah, East Malaysia. A small UAV was also used to obtain stereo images of the same area. Equation (3.29) was used to compute (estimate) the height of any object (vegetation / trees) as follows:

$$U = Ft/H \tag{3.29}$$

where U = object height; F = focal length; t = distance between camera / sensors; and H is the disparity map computed by the proposed algorithm. If the focal length and distance been cameras or sensors has been known, and also have the disparity map, the height or depth of a particular object has been estimated. The proposed datasets consists of UAV and satellite have different distances and distance between UAV sensors was

1 m and between satellite sensors was 10 meters. The focal length of the Pleiades satellite sensor was F = 12905 mm; pixel sizes were 0.13, 0.13 for panchromatic and 0.052, 0.052 for multispectral (provided by vendor). With pixel size, focal length and GSD, the altitude of the satellite sensor was easily determined by Equation (3.30), and likewise for the focal length for UAV images.

$$\frac{F}{H} = PS/GSD \tag{3.30}$$

where F = focal length; H = altitude of sensors; PS = pixel size; and GSD is ground sample distance for airborne or satellite dataset.

3.8 Distance between HV Lines and Trees outside ROWs

In real-time environment, some vegetation outside ROWs appear to be in safe zone with respect to HV overhead lines. However, due to bad weather, those dangerous trees can strike or even fall on high voltage (HV) lines. Therefore, it is necessary to maintain a suitable clearance distance from those encroachments. To overcome this problem, the distance between excess outside vegetation and HV transmission lines is determined. This is done firstly by monitoring the encroachments on the left and right side right of ways (ROWs). For this purpose, the algorithm has already determined absolute height of vegetation or tree as in section 3.8. After that, the algorithm is used to identify the zone (dangerous, medium and low) whether the encroached vegetation (outside ROWs) are at the safe distance from power lines by placing the predetermined threshold levels. After determining the zone of the encroached vegetation with respect to the HV lines and poles, a new analytical geometry is proposed as in Figure 3.24 that will utilize data from the above triangulation based technique to find out the accurate distance between the HV lines and the encroached vegetation outside ROWs.

The proposed system model based on three numbers of towers implemented in this work and this system model can be generalized for N (any length) number of towers. The system model for three number of power poles as shown in Figure 3.25.

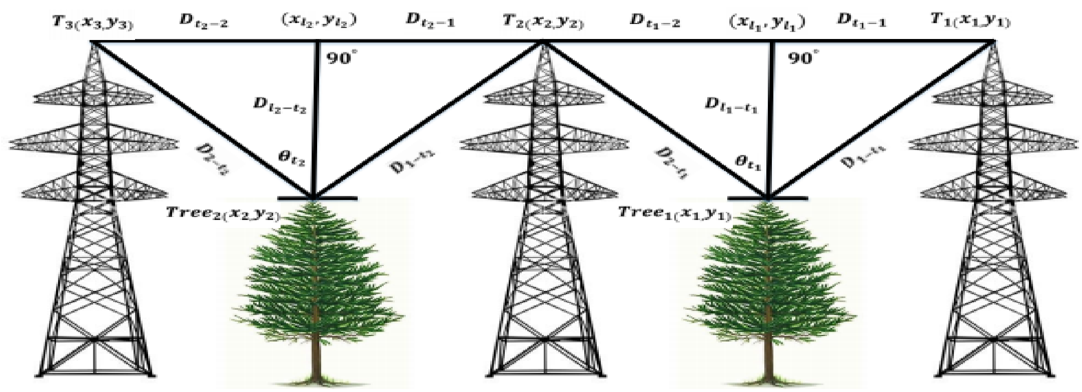


Figure 3.24: The block diagram to compute height and distance of vegetation, tree, and towers based on three towers configuration.

$T_1(x_1, y_1)$ = coordinates of Tower1

$T_2(x_2, y_2)$ = coordinates of Tower2

$T_3(x_3, y_3)$ = coordinates of Tower3

(x_{t_1}, y_{t_1})=middle point on the first transmission line

(x_{t_2}, y_{t_2})=middle point on the second transmission line

(x_{l_1}, y_{l_1})=Coordinates of the Tree1 outside right of way.

(x_{l_2}, y_{l_2})= Coordinates of the Tree2 outside right of way

D_{t_2-2}=Distance between intersection point on power transmission line2 and tower3

D_{t_2-1}= Distance between intersection point on power transmission line2 and tower2

D_{t_1-2}= Distance between intersection point on power transmission line1 and tower2

D_{t_1-1}= Distance between intersection point on power transmission line1 and tower1

$D_{l_2-t_2}$=Distance between tree2 and power transmission line2

$D_{l_1-t_1}$=Distance between tree1 and power transmission line1

D_{2-t_2}=Distance between tree1 and tower2

D_{1-t_2}=Distance between tower2 and tree2

D_{2-t_2}=Distance between tower3 and tree3

D_{1-t_2}= Distance between tower2 and tree2

As shown in Figure 3.25, the distance from the encroached vegetation to the HV transmission lines can be calculated by determining the coordinates of the intersection between the two lines. The slope of heavy transmission line as shown in equation (3.31).

The analytical expression provided the information of the line connecting the power poles.

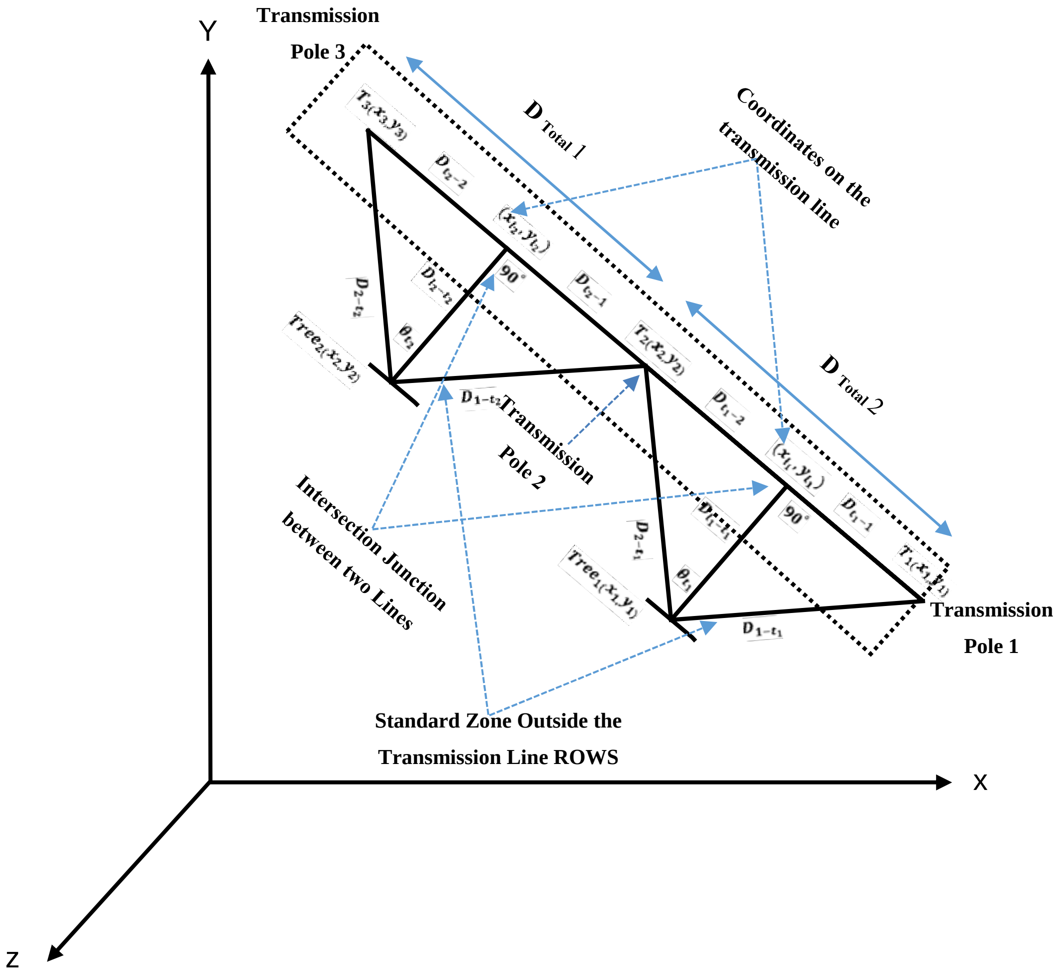


Figure 3.25: The distance between trees and transmission lines.

$$m_{tline} = \frac{y_2 - y_1}{x_2 - x_1} \tag{3.31}$$

$$\theta_{tline} = \tan^{-1}(m_{tline}) \tag{3.32}$$

The equation (3.31) for estimating perpendicular distance between the trees or vegetation (outside the HV transmission lines) and the HV line connecting the two transmission poles can be obtained using the equation (3.33). The slope and angle provided the information related to power lines for distance estimation.

$$\theta_{t_1} = \theta_{tline} - 90^{\circ} \tag{3.33}$$

$$\mathrm{m}_{t_1} = \tan^{-1}(\theta_{t_1}) \tag{3.34}$$

$$\mathrm{y} - y_{t_1} = \mathrm{m}_{t_1}(x - x_{t_1}) \tag{3.35}$$

The intersection of the two lines is given by

$$y_{l_1} - y_1 = \mathrm{m}_{\mathrm{tline}}(x_{l_1} - x_1) \tag{3.36}$$

$$y_{l_1} - y_{t_1} = \mathrm{m}_{t_1}(x_{l_1} - x_{t_1}) \tag{3.37}$$

$$\mathrm{y} - y_{t_1} = \mathrm{m}_{t_1}(x - x_{t_1}) \tag{3.38}$$

The equation (3.38) can be written as

$$y_{l_1} = \mathrm{m}_{\mathrm{tline}}(x_{l_1} - x_1) + y_1 \tag{3.39}$$

Put equation (3.39) into equation (3.37)

$$\mathrm{m}_{\mathrm{tline}}(x_{l_1} - x_1) + y_1 - y_{t_1} = \mathrm{m}_{t_1}(x_{l_1} - x_{t_1}) \tag{3.40}$$

$$\mathrm{m}_{\mathrm{tline}}x_{l_1} - \mathrm{m}_{\mathrm{tline}}x_1 + y_1 - y_{t_1} = \mathrm{m}_{t_1}x_{l_1} - \mathrm{m}_{t_1}x_{t_1} \tag{3.41}$$

$$\mathrm{m}_{\mathrm{tline}}x_{l_1} - \mathrm{m}_{t_1}x_{l_1} = y_1 - y_{t_1} - (\mathrm{m}_{t_1}x_{t_1} - \mathrm{m}_{\mathrm{tline}}x_1) \tag{3.42}$$

$$x_{l_1}(\mathrm{m}_{\mathrm{tline}} - \mathrm{m}_{t_1}) = y_1 - y_{t_1} - (\mathrm{m}_{t_1}x_{t_1} - \mathrm{m}_{\mathrm{tline}}x_1) \tag{3.43}$$

The coordinates (x_{l_1}, y_{l_1})

$$x_{l_1} = \frac{y_1 - y_{t_1} - (\mathrm{m}_{t_1}x_{t_1} - \mathrm{m}_{\mathrm{tline}}x_1)}{(\mathrm{m}_{\mathrm{tline}} - \mathrm{m}_{t_1})} \tag{3.44}$$

$$y_{l_1} = \mathrm{m}_{\mathrm{tline}}(x_{l_1} - x_1) + y_1 \tag{3.45}$$

The distance between the outside vegetation or trees and the HV overhead lines $D_{l_1-t_1}$is

$$D_{l_1-t_1} = \sqrt{\left(x_{t_1} - x_{l_1}\right)^2 + \left(y_{t_1} - y_{l_1}\right)^2} \tag{3.46}$$

The distance between the outside vegetation and the far away pole is given below.

$$D_{2-t_1} = \sqrt{{D_{l_1-t_1}}^2 + {D_{t_1-2}}^2} \tag{3.47}$$

$$D_{2-t_1} = \sqrt{{D_{l_1-t_1}}^2 + (D_{total} - D_{t_1-2})^2} \tag{3.48}$$

Where $D_{total} - D_{t_{1-1}} = D_{t_1-2}$

The equation of second transmission line is given by

$$\mathrm{m}_{tline1} = \frac{\mathrm{y}_3 - \mathrm{y}_2}{\mathrm{x}_3 - \mathrm{x}_2} \tag{3.49}$$

$$\theta_{tline1} = \tan^{-1}(\mathrm{m}_{\mathrm{tline1}}) \tag{3.50}$$

$$m_{t_2} = \tan^{-1}\left(\theta_{t_2}\right) \tag{3.51}$$

$$\mathrm{y} - y_{t_2} = \mathrm{m}_{t_2}\left(x - x_{t_2}\right) \tag{3.52}$$

The intersection of the two lines is

$$y_{l_2} - y_2 = \mathrm{m}_{tline1}\left(x_{l_2} - x_2\right) \tag{3.53}$$

$$y_{l_2} - y_{t_2} = \mathrm{m}_{t_2}\left(x_{l_2} - x_{t_2}\right) \tag{3.54}$$

Rearrange equation (3.53)

$$y_{l_2} = \mathrm{m}_{tline1}\left(x_{l_2} - x_2\right) + y_2 \tag{3.55}$$

Put equation (3.55) into equation (3.54)

$$\mathrm{m}_{tline1}\left(x_{l_2} - x_2\right) + y_2 - y_{t_2} = \mathrm{m}_{t_2}\left(x_{l_2} - x_{t_2}\right) \tag{3.56}$$

$$\mathrm{m}_{tline1} x_{l_2} - \mathrm{m}_{tline1} x_2 + y_2 - y_{t_2} = \mathrm{m}_{t_2} x_{l_2} - \mathrm{m}_{t_2} x_{t_2} \tag{3.57}$$

$$\mathrm{m}_{tline1}x_{l_2} - \mathrm{m}_{t_2}x_{l_2} = y_{t_2} - y_2 - (\mathrm{m}_{tline1}x_2 - \mathrm{m}_{t_2}x_{t_2}) \tag{3.58}$$

$$x_{l_2} = \frac{y_{t_2} - y_2 - (\mathrm{m}_{tline1}x_2 - \mathrm{m}_{t_2}x_{t_2})}{\mathrm{m}_{tline1} - \mathrm{m}_{t_2}} \tag{3.59}$$

$$y_{l_2} = \mathrm{m}_{t_2}(x_{l_2} - x_{t_2}) + y_{t_2} \tag{3.60}$$

The distance between the outside tree and the HV overhead lines 2 $D_{l_2-t_2}$ is given by

$$D_{l_2-t_2} = \sqrt{(x_{t_2} - x_{l_2})^2 + (y_{t_2} - y_{l_2})^2} \tag{3.61}$$

The distance between the outside tree and the far away pole is given below

$$D_{2-t_1} = \sqrt{{D_{l_2-t_2}}^2 + {D_{t_2-2}}^2} \tag{3.62}$$

$$D_{t_2-2} = \sqrt{{D_{l_2-t_2}}^2 + (D_{total} - D_{t_2-2})^2} \tag{3.63}$$

$$D_{t_2-2} = D_{total} - D_{t_2-1}$$

$$D_{1-t_2} = \sqrt{(x_{t_2} - x_2)^2 + (y_{t_2} - y_2)^2} \tag{3.64}$$

$$D_{2-t_2} = \sqrt{{D_{l_2-2}}^2 + {D_{l_2-t_2}}^2} \tag{3.65}$$

The dangerous zone width and height is estimated by using the formulas given in equation (3.66 and 3.67). The coordinates ($\mathrm{x}_{\mathrm{l}_2}$), ($\mathrm{y}_{\mathrm{l}_2}$) and ($\mathrm{D}_{\mathrm{l}_2-\mathrm{t}_2}$) can be found by using above Triangulation method. The overall status of the vegetation is found by comparing the level (height) of vegetation, and the distance from HV lines. For example, if the level of encroached vegetation outside the HV lines is in danger zone and distance of vegetation from the HV transmission lines is also in danger zone then the overall status of the vegetation encroachments shown by the algorithm would be dangerous and vice-versa. The sag depends on the distance from tower (x). The width and height of dangerous zone depends on sagmax, span, clearance width $D_{\text{clearance}}$ and $\emptyset_{\text{swing}}$ as shown in the following equations(3.66 and 3.67). This model is used to calculate the cross arms height using estimated height of the tower. This model is used to calculate the cross arms height using estimated height of the tower. This tower model has three cross arms.

$$W_{dang}(x) = D_{AC} + 2.D_{clearance} + 2\left[\frac{4sagmax}{span^2}\left(\frac{span}{2} - x\right)^2 + L_{chain}\right]\sin(\emptyset_{swing}) \quad (3.66)$$

$$H_{dang}(x) = H_{tower} - D_{clearance} - \left[\frac{4sagmax}{span^2}\left(\frac{span}{2} - x\right)^2 + L_{chain}\right]\cos(\emptyset_{swing}) \quad (3.67)$$

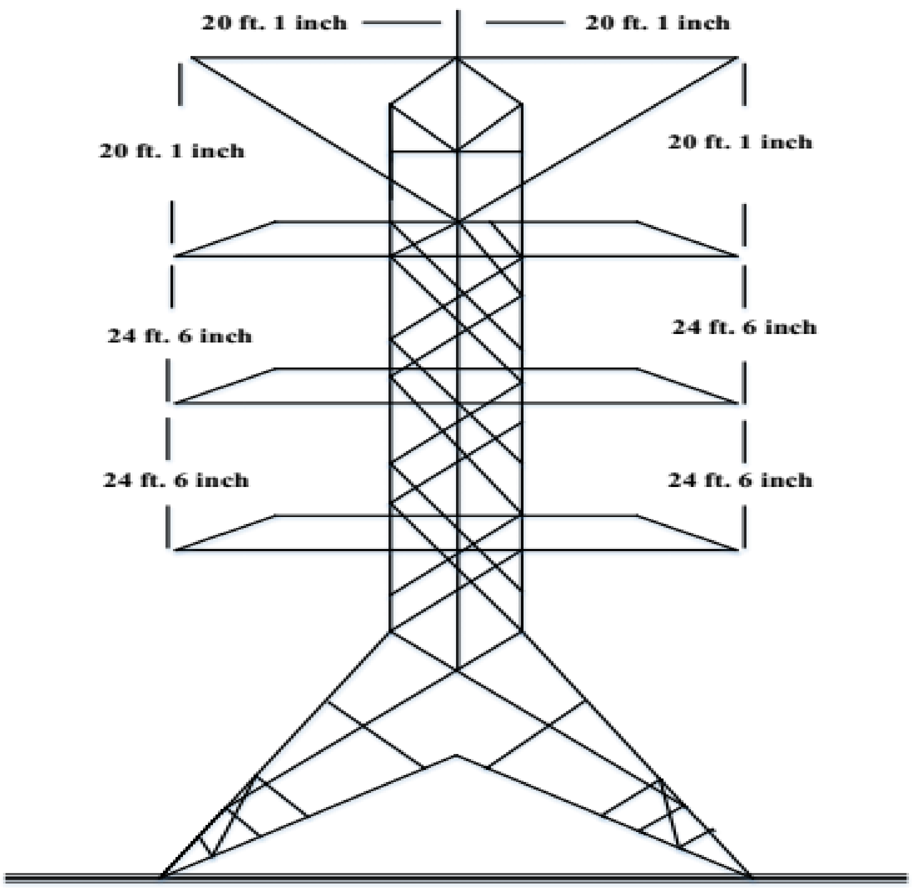


Figure 3.26: The model used to calculate cross-arms height based on estimated height of tower [63].

The each cross arm has a different vertical height from top of the tower. The First cross arm has a height from top of the tower is 20 ft and 1 inch and second cross arm have a height from top of the tower is accumulative (45 ft. and 7 inches). The third cross arm height is (68 ft. and 13 inches). If the total height of towers from the top is known, the height of each cross arms can be find by subtracting height from top of the tower. The configuration of each cross arm with tower height is shown in Figure 3.26.

3.9 Performance Matrices

In this section, the performance matrices are evaluated based on the ground truth and estimated height of power transmission poles in five different cases. These segmented images have four numbers of power transmission poles, by using stereo matching algorithm the height of power transmission poles, trees and vegetation has been measured. Then different performance matrices are required to compute the error or accuracy based on measured height or ground truth values. The objective is to estimate the accuracy, precision and recall performance matrices based on our proposed algorithm and existing algorithms.

Precision and recall are one of the most commonly used evaluation metrics in pattern recognition and information retrieval. Precision-recall depends on the relevance classification criteria. The precision can be defined as the ratio of a number of retrieved elements to the total number of relevant elements in an instance as shown in equation (3.67). The ratio depends on the true positive and false positive values calculated based on the ground truth and estimated height value. Precision and recall are calculated based on estimated height by the system and a certain threshold level is set for different height level approximation. The comparison of proposed with the existing algorithms like GC, DP, BP, and area based matching such as SAD, and SSD has been investigated and the performance of disparity map is compared in terms of height of vegetation near power transmission poles.

$$Precision = TP/(TP + FP) \tag{3.67}$$

$$Recall = TN/(TN + FP) \tag{3.68}$$

where TP is true positive, TN is true negative, FP false positive and FN false negative. The recall is defined as the ratio between the numbers of relevant elements on an instance to the number of retrieved elements as shown in equation (3.68). This ratio is calculated on the true height and the estimated height or computed height and comparison has drawn with the existing height estimation algorithms. According to the definition of precision, the algorithm returns the true height as higher as the ground height of the power pole. High recall means the algorithm produces more true height value as compared to actual highest value. The result interpretation of the accuracy shows that the proposed algorithm provides high height values as compared to the true actual height of power poles as shown in equation (3. 69). The detected height of power poles could be the further threshold to ten (10) percent to improve the sensitivity and recall. The threshold set ten (10) percent of estimation height value.

$$Accuracy = TP + TN/(TP + TN + FN + FP) \qquad (3.69)$$

3.10 Summary

In this chapter, the proposed methods based on sparse representation and CNN has been explained to measure the height of vegetation and trees near power transmission poles. The introduction of sparse, dictionaries used in sparse and construction of different dictionaries has been discussed in this chapter. The proposed disparity map estimation based on sparse representation algorithm has been discussed in detail with mathematical formulation. The sparse representation algorithm castoff to measure the disparity map which further used for height estimation of power poles, vegetation and trees near or under the power transmission lines and also compute the distance between them. The introduction of proposed method based on CNN algorithm and its parameters are explained. The CNN is introduced to measure the height based on stereo matching aerial data set. The mathematical model has been explained well. For simplicity, currently three number of power poles have been used for height estimation. The mathematical model consists of three consecutive number of power poles using triangulation method and derived model can be extended for any number of power poles. The different performance metrics explained in this chapter, these performance metrics are computed based on estimated height and existing height of power transmission poles.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the results have been evaluated based on existing and proposed stereo matching algorithms. The disparity map is estimated using proposed sparse representation and CNN based techniques. Based on performance metrics (accuracy, precision and recall), height of vegetation /trees has been calculated using satellite and UAV stereo images dataset. Further, the different hyper parameter used in CNN algorithm have described for optimization of CNN algorithm. The performance analysis shows that the proposed algorithms have a capability to measure height based on aerial images. On the basis of estimated height and estimated distance between vegetation / trees and power transmission lines, the different number of threats have been identified and evaluated using proposed algorithms (CNN and SR) for satellite and UAV stereo images. The discussion and analysis of proposed techniques has been analyzed in this chapter.

4.2 Experimental Results based on Sparse Representation

In this section, the experiments results using sparse representation algorithm have been discussed in detail. The important and complex aspects of occluded, low-textured and noisy inputs for disparity measurement in stereo imaging, for which the successfully designed and applied the sparse representation algorithm. The sparse representation (SR) algorithm introduced ordering constraints to enhance disparity map measurement that produced successful results, even for occluded areas. Modelled occluded regions were also much clearer compared to spurious mismatches and were visually appreciated in disparity maps as well as verified by error percentage calculations. Disparity maps demonstrated slight visual differences, but error measurements were very similar for the existing algorithms. Thus, the introduction of ordering constraints accommodated

SR algorithm processing of occluded areas and minimized noise in aerial (UAV and satellite) images, as demonstrated by disparity maps. The satellite stereo images were very low in texture yet good in resolution power so that differences in textureless regions were minimized by using the SR algorithm, even though another solution for textureless regions remains wanting, which adds a very important point to our discussion.

Based on results presented in this study, the sparse representation technique capturing the prominent features with ordering constraints to a universally accepted method appears to be a promising solution to problems that present when measuring disparity maps. Although results are less than astounding, they clearly indicate the importance of proposed SR algorithm when processing images containing larger disparity ranges and excessive noise. Large disparities present in stereo images taken by aerial sensors hold spurious foci. The introduction of ordering constraints insufficiently solved the problem of textureless areas due to the large number of displaced points. This was expected since the problem for such regions is a low signal-to-noise ratio. The only way to increase this ratio is to include more pixels when calculating the matching cost. Therefore, it is of great interest to combine robust area-based correlation matching with meticulous global SR-based matching into one algorithm. Another challenge is the conversion of matching problems to SR problems, which necessitates viewing the matching problem as a labelling problem. Labelling requires a less capable discrete disparity map to capture object shapes. But for this purpose, fronto-parallel surfaces in aerial images need orthorectification to provide a reliable solution. Nonetheless, it is now known that aerial stereo images reliably measure the height of vegetation and trees near power transmission lines when using global stereo matching methods such as SR to provide optimized disparity mapping. The UAV and satellite stereo images are used to measure the height of objects using stereo matching algorithms and disparity map has been compute based on the proposed sparse representation algorithm using proposed over-complete dictionaries. The results below belong from the highest accuracy based dictionary (DRT). The patch size of 8x8 is selected in left stereo image and applied dictionary on the selected patch to compute spare coefficients, which further used as feature vector in left and right images to compute the disparity map. The ridgelet transform basis function provides better and stable sparse coefficients for disparity computation. This transform is preserving more edges or corners or lines of the selected patch in the both stereo images as compared to other fixed size dictionaries such as the DCT and DWT etc. Our proposed algorithm performed better where image preserved more prominent features, the image rectification was required for searching image in one direction either in vertical or horizontal direction. This search was 1-D not in 2D. In this way, All disparity values are calculated in the complete image by scanning row-wise and column-wise. The different locations have been selected from area of interest using both stereo datasets (UAV and Satellite). The four different stereo segments have different locations were marked in both stereo images for UAV and satellite and calculated disparity map for each located stereo image pair. The area of interest images were extremely challenging due to occlusions and weather conditions that caused some false disparity outcomes. Hence, uniqueness constraints were imposed (integrated) to ensure correct disparity computations by the proposed sparse representation algorithm.

The first area segment contains the stereo image based on UAV and satellite sensors are shown in Figure 4.1 (a, b) and Figure 4.2 (a, b). In this case two power poles, power lines and roads objects has been marked with red circle on stereo images as shown in Figure 4.1 (a, b) for UAV images and Figure 4.2 (a, b) for satellite. For second area segment, three number of power poles has been selected and marked red circle on image as shown in Figure 4.1 (c, d) for UAV stereo images and Figure 4.2 (c, d) for satellite. The disparity map computed by proposed sparse algorithm has shown in Figure 4.3 (a, b) for UAV images and Figure 4.3 (c, d) for satellite images. Our area of interest images was very challenging due to occlusion and weather condition, may be some false disparity computed by our proposed algorithm. The uniqueness constraints were imposed to insure the correct disparity computation. The disparity map was computed for both cases have been utilized to measures the height of power poles. The results clearly shows that the disparity map was more accurate with distinct points particularly in second case for UAV images, the matching points were more as compared to satellite based patches. In second segment, due to some occlusion, change of texture in some region and depth discontinuities at certain points may change the disparity values and calculated the false disparity as shown in Figure 4.3 (b). The occluded region should be excluded for correct disparity estimation in stereo matching problem. The technology constraints based on UAV sensors and satellite sensors may hider to compute the best disparity estimation due to instability of vehicle movement and camera shaking. The disparity map computed based on satellite images using our proposed method for both cases as shown in Figure 4.3 (c, d). In satellite case, It can be seen visually, the disparity map has less accuracy due to there is textureless area and might be some points did not match due to smoothness of the surface. Disparity maps computed from those segments show some occlusions and texturless region in stereo regions where depth discontinuities obtained disparity values that were falsely matched at specific points based on existing stereo matching technique.

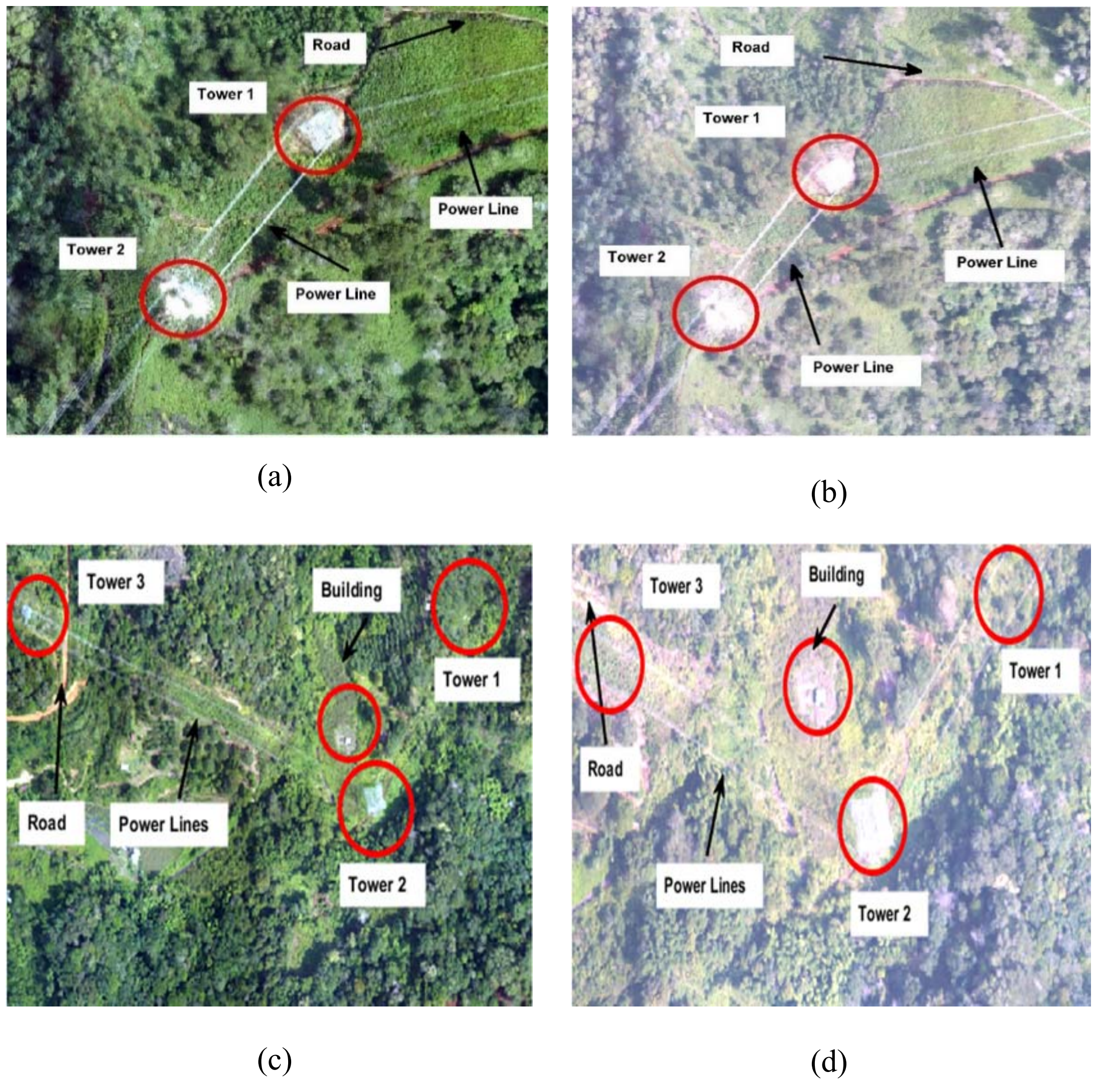
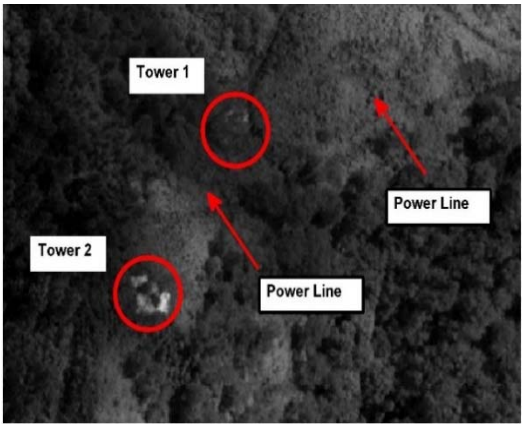
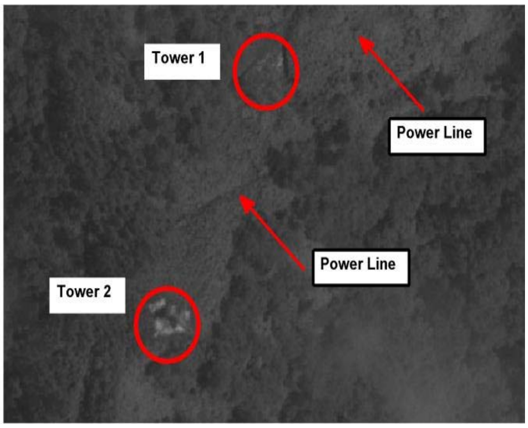
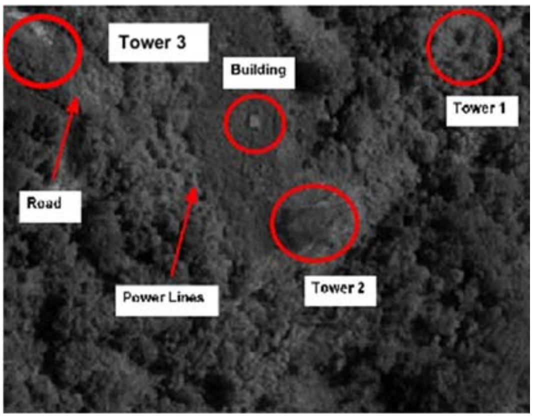


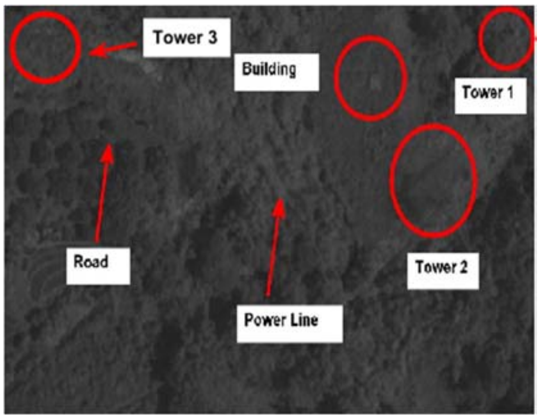
Figure 4.1: UAV stereo images: (a, b) Left and right stereo images for UAV case 1 (c, d) Left and right stereo images for UAV case 2.





(a) (b)





(c) (b)

Figure 4.2: Satellite stereo images: (a, b) Left and right stereo images for satellite case 1 (c, d) Left and right stereo images for satellite case 2.

Now, the third segment is discussed which contains two power transmission poles using UAV and satellite of same area of interest with same location as shown in Figure 4.4. The stereo images composed of two power poles, power lines and these stereo images become changed in texture and had some occlusion as shown in Figure 4.4 (a, b) for UAV images. Similarly the satellite stereo images of same selected area are shown in Figure 4.4 (c, d). The UAV images had some occlusion, depth discontinuity and some regions had less texture values. Due to these constraints, the disparity map computed by proposed sparse algorithm had less accuracy and difficult to reconstruct the some part of power lines as shown in Figure 4.5 (a). In satellite images, the disparity map is shown in Figure 4.5 (b) produce less accuracy as compared to UAV image based disparity.

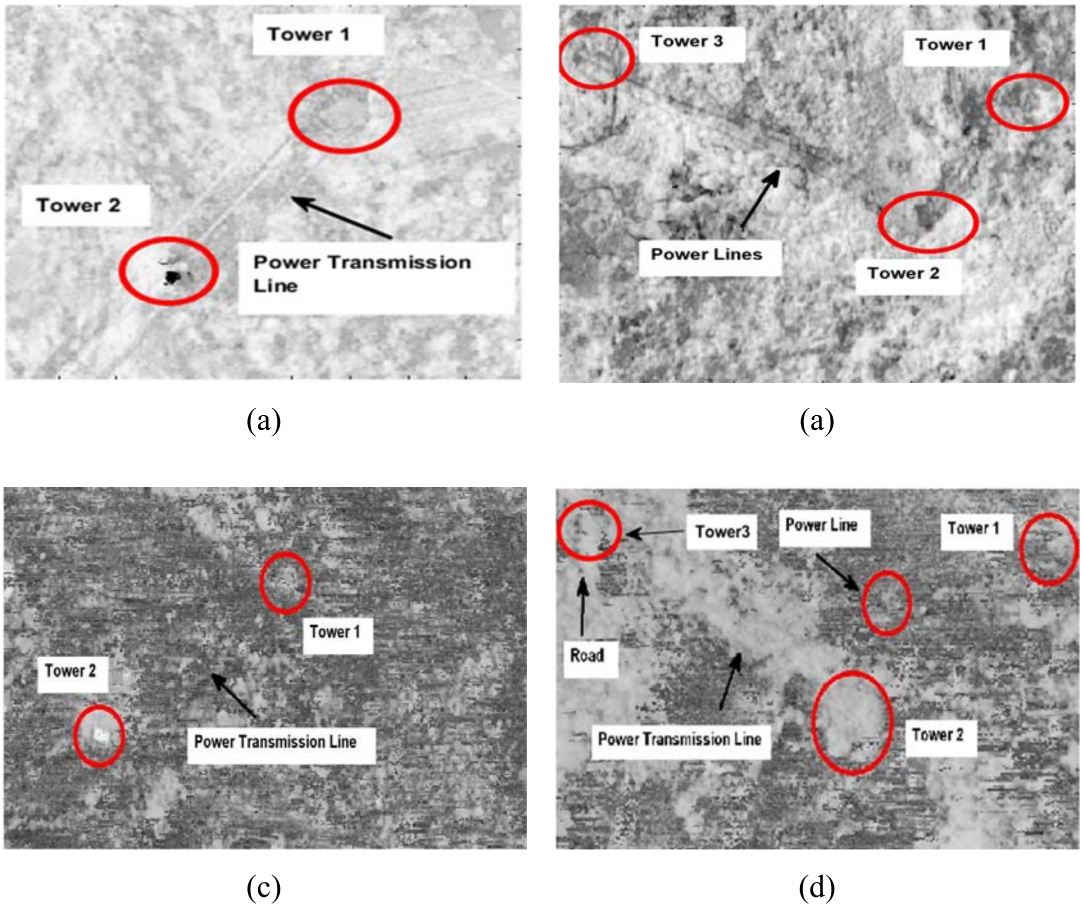
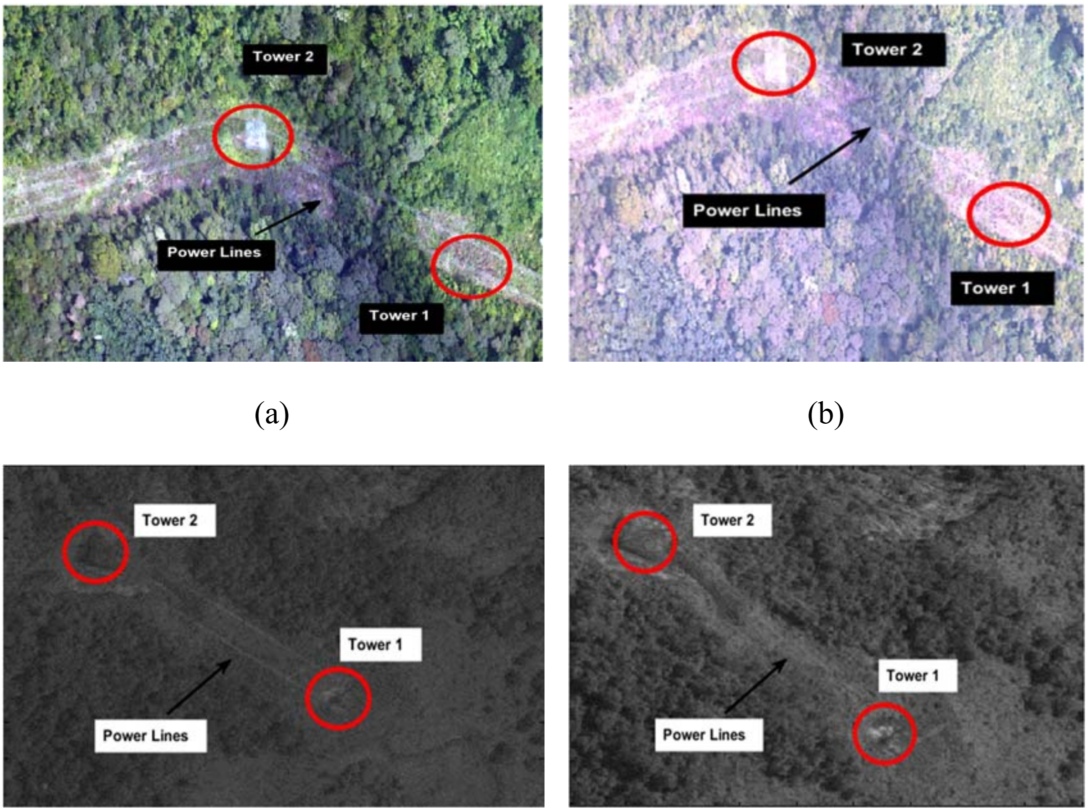


Figure 4.3: Estimated disparity map based on sparse representation algorithm (a, b) UAV image for case 1 and case 2 (c, d) Satellite image for case 1 and case2.

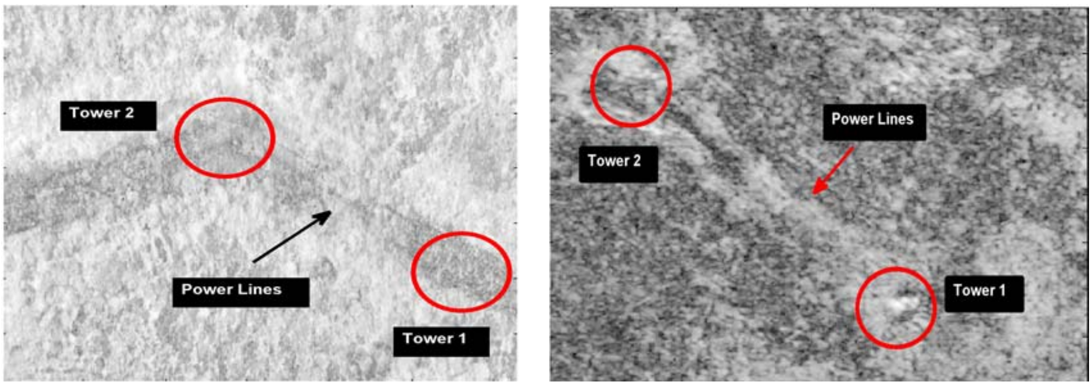
The stereo images based on UAV and satellite for 4th segment has been shown in Figure 4.6. In this segment, the number of power lines is prominent and can be seen clearly on the stereo images for both UAV and satellite. The disparity map computed by proposed algorithm for UAV image produced more accuracy and power lines can be seen on the images as shown in Figure 4.7 (b). Similarly, disparity map produced by satellite stereo images is shown in Figure 4.7 (a). The proposed algorithm reconstruct well disparity map in both case (satellite and UAV) for height estimation. The performance of proposed algorithm using UAV is better for disparity map estimation as compared to the satellite stereo images. The height estimation depends on the better disparity estimation. The proposed SR has a capability to reconstruct the disparity map.



(a) (b)

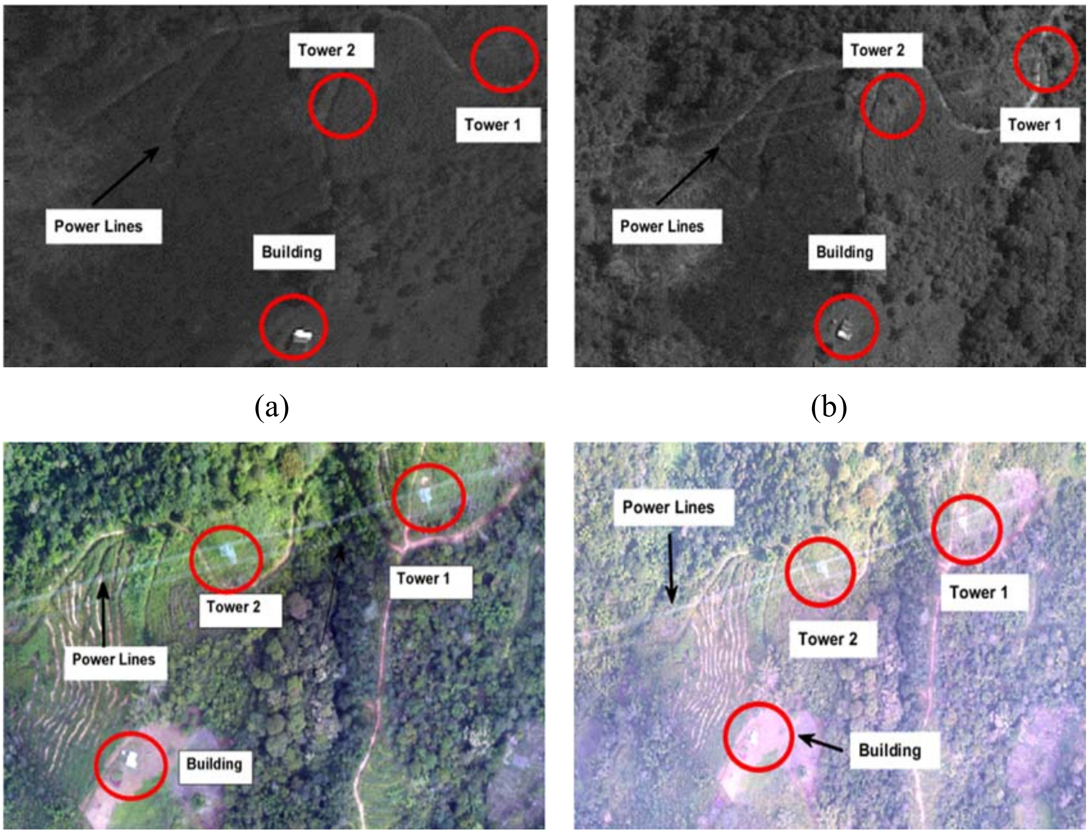
(c) (d)

Figure 4.4: Stereo images based on UAV and satellite (a, b) UAV based stereo images (c, d) Satellite based stereo images.



(a) (b)

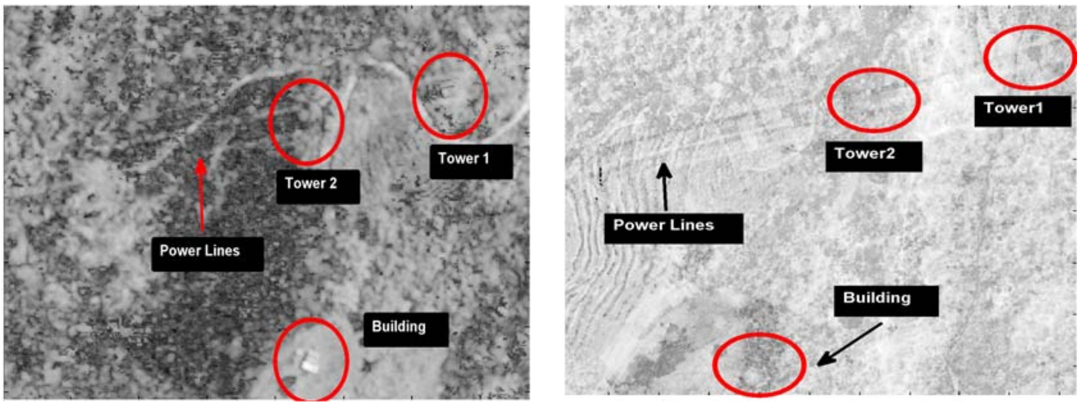
Figure 4.5: (a) Estimated disparity map using sparse representation algorithm based on UAV (b) Estimated disparity map using sparse representation algorithm based on satellite.



(a) (b)

(c) (d)

Figure 4.6: Stereo images based on UAV and satellite (a, b) UAV based stereo images (c, d) Satellite based stereo images.



(a) (b)

Figure 4.7: (a) Estimated disparity map using sparse representation algorithm based on satellite (b) Estimated disparity map using sparse representation algorithm based on UAV.

The proposed SR reconstructs disparity map efficiently using the proposed discrete ridgelet transform (DRT) dictionary based on UAV and satellite stereo imagery as explained in a simple scenario with two power poles, power lines and a small building. The proposed SR algorithm well reconstructs small building and power poles on the disparity map by estimated pixel information propagated from the object boundaries whichcannot not be handled by existing stereo matching algorithms. The proposed SR easily reconstructd the information on the object boundaries due to estimating the pixel similarity captured by the proposed dictionary. The proposed algorithm also estimated the disparity in textureless regions by estimating the missing pixels in different textured regions on the disparity map as shown in Figure 4.3 (a) using UAV stereo image. Similarly, the proposed sparse representation algorithm using proposed dictionary (DRT) handled the depth discontinuities on the disparity map at object boundaries due to efficient similarity matching capability in the proposed dictionary. The SR algorithm reconstructed tow (2) power poles and power lines on the disparity map on the basis of similarity matching in UAV stereo images by capturing the similar features in the stereo patches in sparse feature space.

The SR proposed algorithm could not handle the occlusion detection that's the reason some part of disparity was misleading the reconstruction area and showed less accuracy by mismatching the pixels during the sparse coefficient reconstruction process in proposed algorithm as shown in Figure 4.3 (b) for UAV stereo images. Similar results were obtained in satellite stereo images using proposed SR algorithm and visualization as well as analytical accuracy degraded due to low spatial resolution of satellite stereo images. The disparity map estimated by proposed SR in satellite stereo cases hadreconstructed the information at the object boundaries due to mismatching the pixels on the object boundaries. The SR have distinguished the pixel characteristics at the object boundaries and easily estimate the depth discontinuity although the resolution of the satellite images is very low. The estimated disparity map at object boundaries is shown in Figure 4.3 (c, d) for both satellite cases. The analysis has been reported for other cases (third and fourth) based on UAV and satellite stereo images. The SR estimated the disparity map effectively by estimating the pixels at object boundaries and that is how the proposed algorithm handled the object discontinuity in stereo matching process as shown in Figure 4.5 and Figure 4.7.

Learned representations of disparity are also important priors in depth estimation from stereo images, which are still a challenging problem in computer vision and robotics. As it is a highly ill-posed problem, stereo correspondence significantly depends on prior information about the depth structure in the scene. The proposed sparse representation algorithm efficiently resolved the problems exist in stereo matching points for accurate disparity map estimation using fixed proposed dictionaries. These proposed dictionaries (DRT and DTT) used prior information by extracting the patches from stereo images.

The various cases have been discussed based on UAV and satellite stereo images for disparity map computation using proposed sparse representation algorithm. The sparse representation algorithm used proposed dictionaries to capture the prominent features from stereo images on the scanline and estimate the disparity map based on patch similarity between stereo images using Euclidian distance technique. Two assumptions about the matching constrains were explicitly stated: uniqueness and depth continuity. Uniqueness: a given pixel or feature of an image can match no more than a pixel or feature in another image. Continuity: the cohesiveness of matters suggests that the matches disparity should smoothly vary everywhere in the image. The disparity map estimated based on sparse representation accommodated efficiently the aforementioned assumptions (uniqueness, depth continuity) along with textureless regions in the stereo matching aerial imagery. The sparse representation efficiently reconstructed the disparity map and resolved the depth discontinuity issue successfully based on patch based similarity using proposed dictionaries construction and sparse coefficients estimation.

4.3 Experimental Results based on CCN based Approach

Matching cost computation a fundamental step shared by both local and global stereo algorithms, plays an important role in establishing visual correspondences. Typically, the reconstruction accuracy of a stereo method largely depends on the dissimilarity measurement of image patches. However, the visual correspondences search problem is difficult due to matching ambiguities, which generally results from sensor noise, image sampling, lighting variation, textureless variation or repetitive region and occlusion. To compute a disparity map, network input is taken from a pair of small images and the output is a measure of their similarity. The proposed network's design provides a trainable feature extractor that represents each image patch as a feature vector. The similarity between the pair of patches is measured rather than between raw image intensity values. Thus, the CNN algorithm produces a disparity map that is smoother, one that lacks the jarring holes in maps produced by other techniques. Moreover, regions with smooth texture have been traditionally combined; thus, they have challenged any attempt to produce depth values. It is here that the proposed CNN approach obtains superior performance in addition to producing better disparity estimates between near textureless surfaces.

Specifically, the CNN offers the potential to describe the structural characteristics in high levels according to the hierarchical feature extraction procedure. At the same time, the fact is also exploit that the deep spatial features of the same class lie in a low-dimensional subspace or manifold and can be expressed by linear sparse regression. Thus, it would be worthwhile to used CNN with high dimensional deep features, which may provide better representation in terms of the characterization of spatial and spectral features and for better discrimination between different feature capabilities. Now the CNN proposed in order to design the disparity map based on aerial stereo imagery for height estimation.

4.3.1 Design of Disparity Map based on CNN Technique

This deep embedding model leverages appeasing data to learn visual dissimilarity between image patches, by explicitly mapping raw intensity into a rich embedding space. In order to compute disparity map, the input to the network is a pair of small image patches and the output is a measure of similarity between them. The design network contains a trainable feature extractor that represents each image patch with a feature vector. The similarity between patches is measured on the feature vectors instead of the raw image intensity values.

These four segments belong to same area of interest in both UAV and satellite. In first segment, the stereo images consists of two towers, trees, grass, power transmission lines and road as shown in Figure 4.8 (a, b) for UAV images and Figure 4.9 (a, b) for satellite images. The second segment consists of building, road, vegetation, trees, three towers and power transmission lines as shown in Figure 4.8 (c, d) for UAV and Figure 4.9 (c, d) for satellite . In satellite stereo images, the first segment consists of two power poles as indicated by red circle and second segment contains three power poles as shown by red circle marked on stereo images.

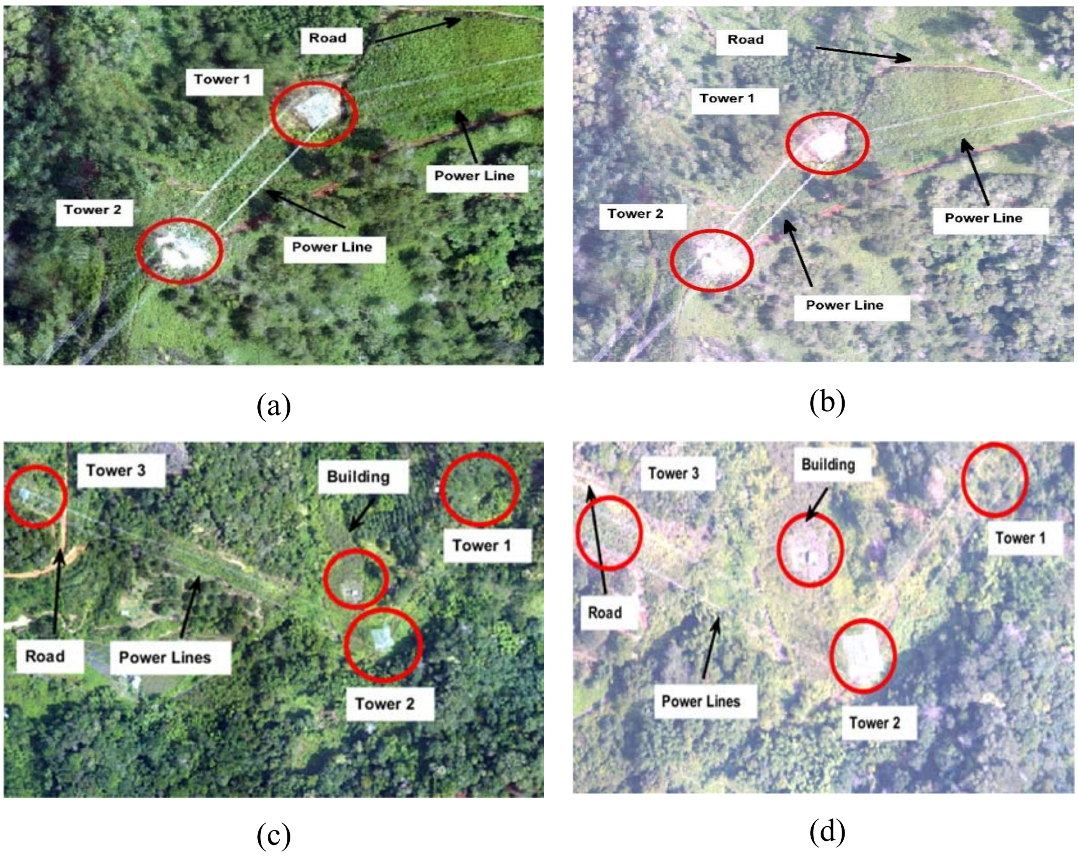
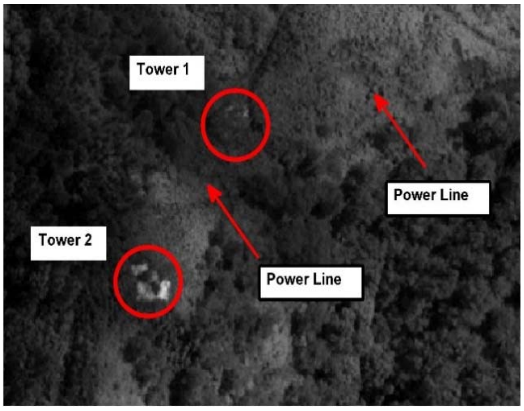
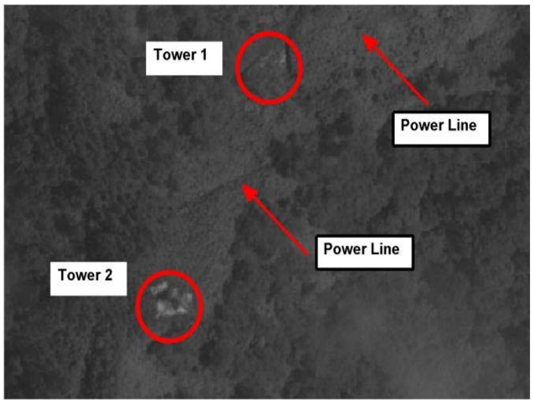


Figure 4.8: The stereo images based on UAV (a, b) Left and right stereo image for UAV case 1 (c, d) Left and right stereo image for UAV case 2.

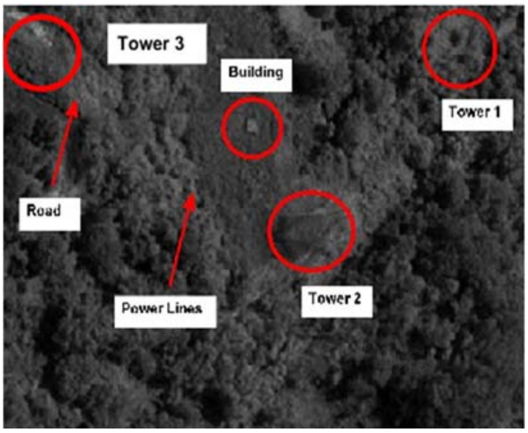
Figure 4.10 (a, b) shows the disparity map for UAV images for both segments (1, 2). The red circle marked on disparity map shows the estimated towers. The road and power transmission lines were also estimated using CNN based approach. The disparity map produced more accuracy by visual inspection in second segment, where all power poles were correctly estimated using the proposed design. The results show that proposed technique produced less error rate as compared to extant method for height estimation of trees and vegetation.



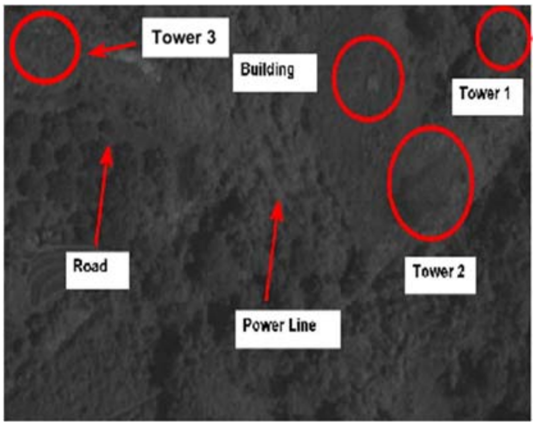
(a)



(b)



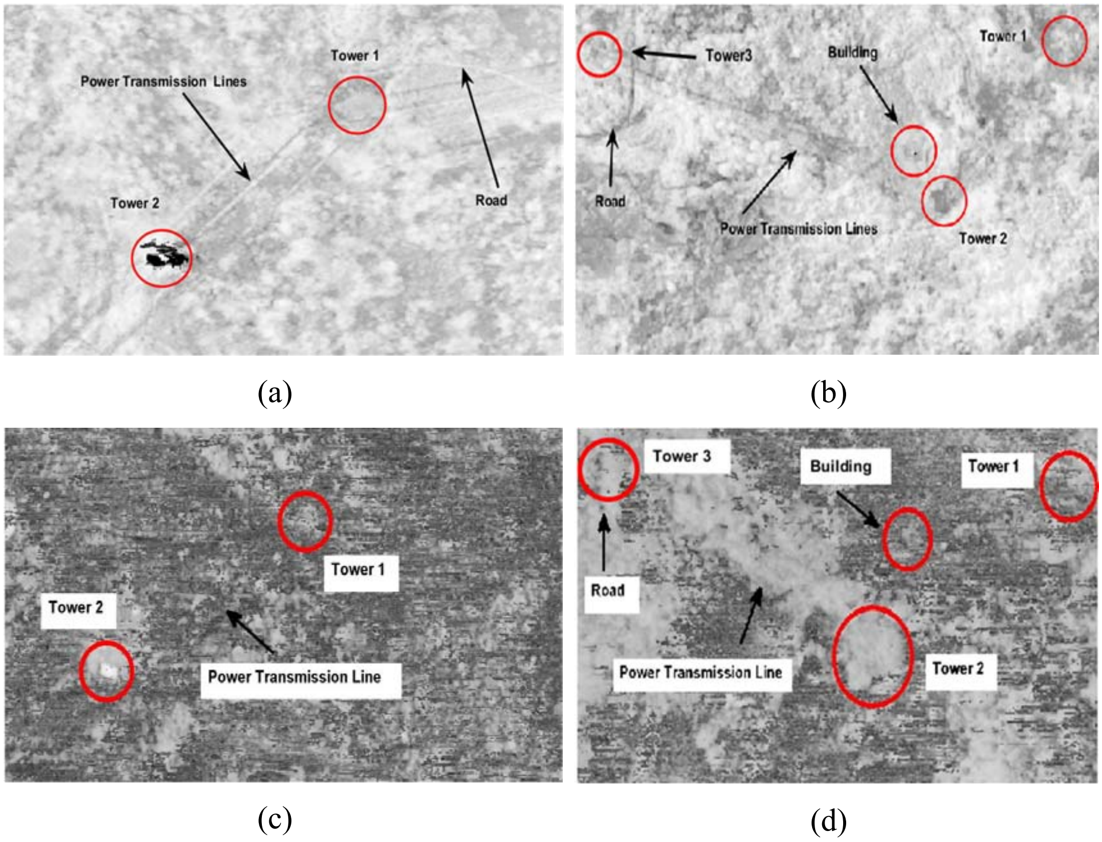
(c)



(d)

Figure 4.9: The stereo images based on satellite (a, b) Left and right stereo images for satellite case 1 (c, d) Left and right stereo images for satellite case 2.

Similarly, the disparity map estimated using stereo images for satellite is shown in Figure 4.10 (c, d) for segments (1, 2). The same area of interest has been shown using satellite stereo images. The power poles and power lines have been marked on the image using the red circle in both cases of stereo images. The second stereo image segment contains also building and three power poles as compared to first segment which have two power poles as shown in Figure 4.10 (d). The 1st segment based disparity map produced comparatively good visualization accuracy as compared to the disparity map of 2nd segment and it has estimated tower 2 well as marked on the disparity image. The power lines and power poles indicated by a red circle on the disparity map.

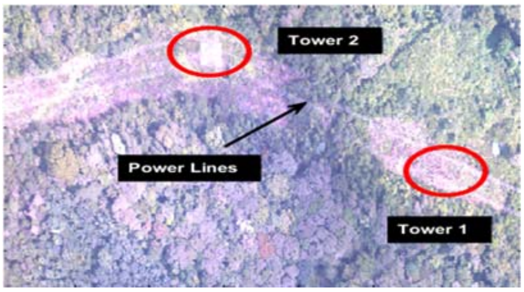


(a) (b) (c) (d)

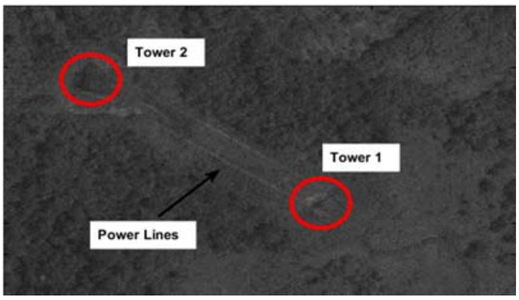
Figure 4.10: Disparity Map for UAV and satellite images: (a, b) Disparity map based on UAV case 1 and case 2 (c, d) Disparity map based on satellite case 1 and case 2.

Further, the other area has been chosen to test and compute the disparity map based on satellite and UAV stereo images and marked power poles as red circle on the stereo images is shown in Figure 4.11 (a). The purpose of chosen this area is that the power lines has clearly identified on the image and also one of power clearly shown especially in satellite images is shown in Figure 4.11 (b) . Now the third segment has been considered to measure the disparity for both UAV and satellite. The disparity map estimated based on CNN algorithm used two power poles of different area with power lines based on both satellite and UAV sensors. In UAV image, the proposed CNN algorithm estimated one of the power poles and marked this power poles as tower 2 shown in Figure 4.12 (a). The disparity map based on satellite images visually show less accuracy due to high noise and low resolution. The estimated disparity map based on satellite image estimate tower 1 very well as shown in Figure 4.12 (b). However, the proposed CNN algorithm did not reconstruct the power lines in case of satellite, the power lines reconstructed comparatively well in UAV case as shown in Figure 4.12 (a).

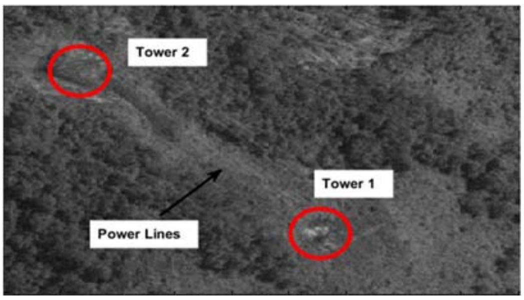
(a)



(b)

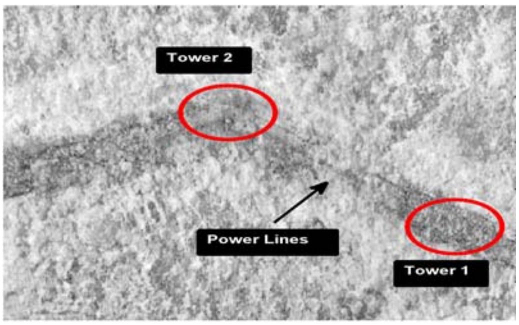


(c)

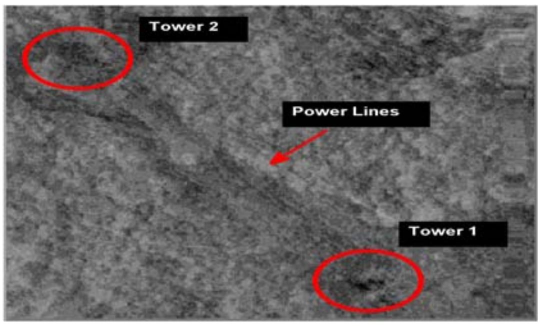


(d)

Figure 4.11: Stereo images using satellite and UAV sensors: (a, b) left and right stereo images using UAV (c, d) left and right stereo images using satellite.



(a)



(b)

Figure 4.12: Estimated disparity map based on proposed CNN algorithm (a) Disparity map based on UAV (b) Disparity map based on satellite.

The fourth segment has been considered in order to compute the disparity map for satellite and UAV. In this segment both stereo images consisted of two power poles marked as red circle on images and clearly shown power lines as shown in Figure 4.13. In this segment, the satellite based disparity provides more accuracy as compared to UAV. The disparity map generated by our proposed CNN algorithm provides higher accuracy and clearly reconstruct the disparity map in both case (satellite and UAV) as shown in Figure 4.14. Based on disparity map estimation results, it can be concluded that CNN based disparity map estimation delivers excellent results by using satellite as well as UAV image. The CNN extracted shift invariant feature in both stereo images and can handle the stereo matching constraints as comparison with sparse based disparity map algorithm.

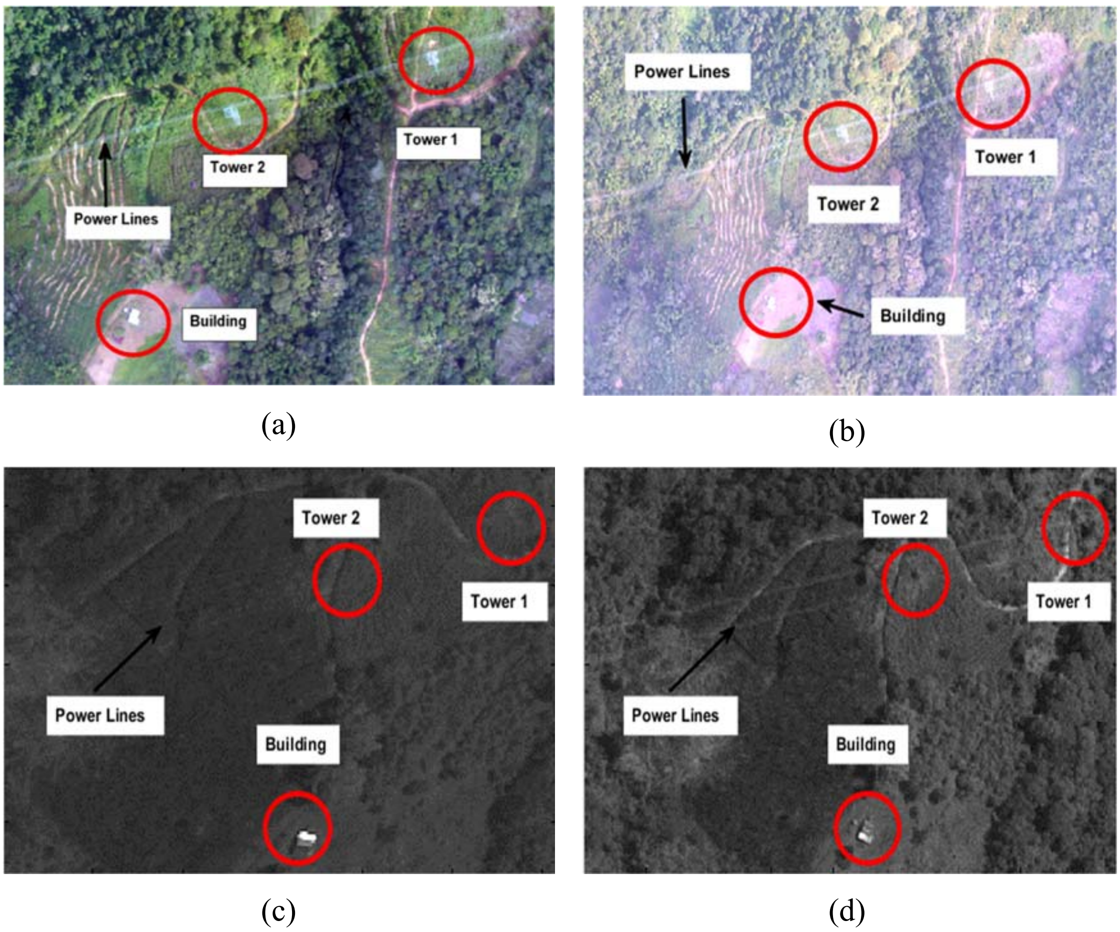
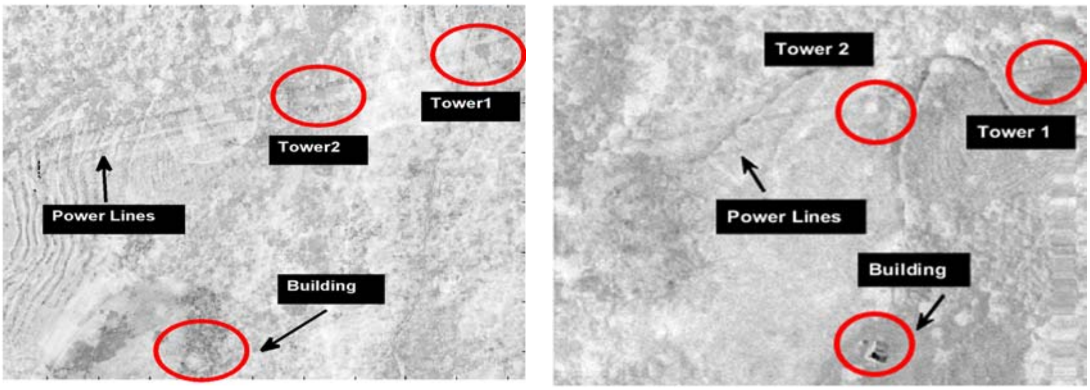


Figure 4.13: Stereo images based on UAV and satellite: (a, b) left and right stereo images for UAV (c, d) Left and right stereo images based on satellite.

The CNN algorithm produced disparity map that were smoother, and without the jarring holes in the maps that results excellent from this approach. Also, regions with smooth textures that are traditionally combined difficult to produce depth values are surprisingly well.



(a) (b)

Figure 4.14: Disparity map based on UAV and satellite stereo images: (a) Disparity map estimate based on UAV image (b) disparity map based on satellite image.

The CNN approach does a much better job at estimated disparity of near textureless surface and depth discontinuity. In this work, the disparity map was calculated using the convolutional neural network and the height of vegetation or trees was measured using disparity map. The disparity map estimation using CNN algorithm based on stereo aerial images (UAV and satellite) had been discussed using various scenarios. The CNN proposed algorithm extracted the prominent features based on spatial property from stereo patches and measured the similarity between these stereo images. The similarity between patches gave the index value which is called disparity. The disparity map produced by proposed CNN algorithm measured similarity at object boundaries and produced the accurate displaced values based on pixel difference at object boundaries, in this way CNN had handled the depth discontinuity in stereo matching images. The smoothness showed on the disparity map had been estimated by proposed CNN algorithm at textureless regions in the stereo patches. The CNN has a strong capability to extract the spatial characteristics in the stereo pairs and extracts deep features from the last fully connected layer to estimate the index at similar patches.

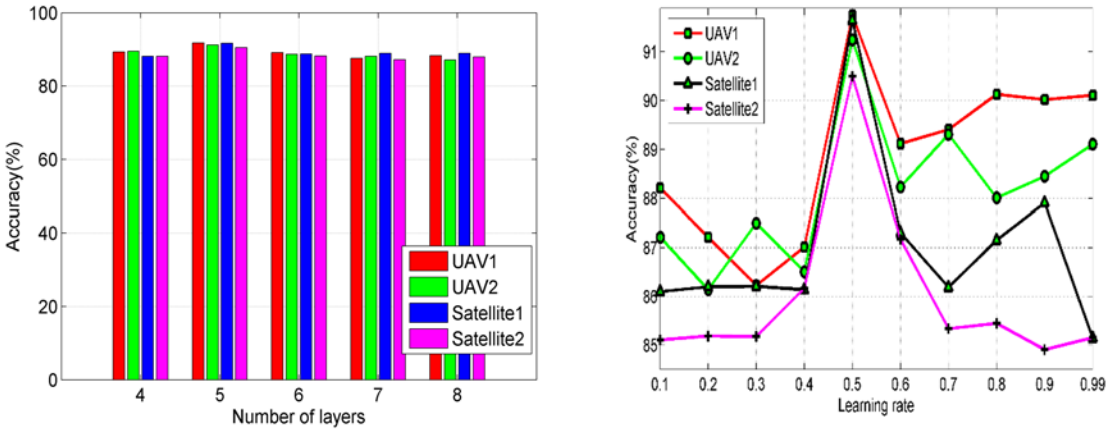
If the spatial structure is missing on the image pixels, the CNN reconstructs the pixels located on the stereo images with rich spatial characteristics and match the pixels in both stereo images for efficient disparity map estimation as shown in Figure 4.10 (a, b) for UAV and satellite stereo images. The proposed CNN reconstructed and estimated the depth discontinuity in stereo matching process while existing stereo matching algorithms did not handled the depth discontinuity at object boundaries in stereo images process. Likewise, the efficiently reconstructed disparity map using the proposed CNN in other cases (3rd and 4th) for UAV and satellite stereo images and proposed algorithm has a strong capability to suppress the noise and reconstructed power poles on the disparity map by mapping the pixels at the object boundaries as shown in Figure 4.12 (a) for UAV and Figure 4.12 (b) for satellite stereo images. In fourth case, the proposed CNN algorithm reconstructed excellent disparity map in UAV and satellite stereo images and well handled at object boundaries to estimate the depth discontinuity which could not handle by other stereo matching approaches. The pixels propagates at object boundaries well measured in stereo matching process based on feature similarity between stereo images. The textureless regions indicate the pixel variations in the stereo images. The proposed CNN well handled the pixel variation in stereo images and produced excellent disparity map as shown in Figure 4.14 (a) and Figure 4.14 (b) for UAV and satellite stereo images.

4.3.2 Hyper-parameters used in CNN Design

Searching out a good set of hyper-parameters is daunting, especially when search space grows exponentially along with the number of hyperparametrs minus a reliable gradient for guidance. To better understand the effect of each hyper-parameter on validation errors, we conducted a series of experiments in which we varied the value of one hyper-parameter while fixing others at default values. By increasing network size, capability for performance generalization improved. The fact that a relatively simple convolutional neural network outperformed all previous approaches to the well-studied problem of stereo imaging is thus an important demonstration of the power of modern machine learning approaches. In the present study, different parameters were explored using CNN-based methods. The most important and most fundamental parameters tuned by the CNN technique were the filter bank (k), and weights for filtering and bias, term (b). As such, these parameters were trained and kept current by using a gradient descent method based on the mini batch-processing unit. This step required no pre-processing since only parameter selection was needed to ease the challenge. However, CNN depth played a substantial role in error rates due to controlling the quality of spatial features based on the abstraction level.

A series of experiments was therefore conducted on the proposed dataset to measure the effectiveness of numerous depth configurations. To affirm the efficiency of depth parameters, five different depth sets were tested, ranging from 4 to 6 to 8 features in each layer. Parameters with deep CNNs provided superior performance compared to shallow CNNs. Spatial features taken from higher levels were far more effectively informed for more accurate image representations. Extracted features from lower CNN layers held a number of edges, lines and corners; hence, their features seemed more complex on the abstracted level. The feature map at layer five was specific within each class and retained semantic meanings. Eventually, eight layers provided optimal accuracy for the proposed study: five convolutional plus three pooling layers. Although an increased number of layers produced better accuracy, additional layers consumed more time due to the increased number of parameters that required training, which is computationally more expensive (See: Figure 4.15(a)).

In Figure 4.15 (a), the parameters of deep CNNs provide better performance as compared to the shallow ones. The spatial features based on high levels are much more effective and intelligent for image representation. The extracted features from lower CNN layer has a number of edges, lines and corners and the features seem to be more complex on an abstract level. The feature map at layer five is specific within the class and with semantic meanings. The optimal number of the layers provides the best accuracy. The increased number of layers could produce better accuracy but more layers consumed more time due to many numbers of parameters trained, which could be computationally more expensive. The stochastic gradient descent required proper learning rate in order to achieve best accuracy. In case of CNN model, the learning rate 0.5 attained good accuracy. The empirical values of the learning rate provide effective results for UAV dataset. The learning rates (0.5, 0.6) may succeed comparable results. The all learning rates acceptable for CCN design except (0.01) in this experiment. The learning rate has been shown in Figure 4.15 (b) and the different learning can be used to optimize the CNN models based on proposed technique. The learning rate is very important for optimization of deep CNN model and it plays an important role during optimization of CNN models.



(a) (b)

Figure 4.15: (a) The number of layers used in CNN algorithm, (b) the learning rate based on CNN algorithm

The number of output feature maps and kernel sizes are traditionally pre-defined parameters in deep CNNs. However, no particular method has been devised for parameter selection. One author previously proposed 7x7, 5x5 or 3x3 kernel sizes for the filter [138] A 5x5-sized kernel set was used in this experiment. The filter size affects speed in terms of computational complexity and further determines optimized accuracy. A trial and error approach was employed to select a number of features. Each trial was repeated ten times to obtain an average accuracy rate for each number of features. We analyzed several filter sizes to produce feature vectors in feature space. The 5x5 window filters produced superior accuracy rates when compared with other filters. Accuracy rates based on the number of features are shown Figure 4.17 (b). Different numbers of feature maps (features) in each layer were used to extract spatial properties by the deep CNN algorithm using different numbers of convolutional and pooling layers. Different numbers of feature maps (features) in each layer were used to extract spatial properties by the deep CNN algorithm using different numbers of convolutional and pooling layers. Six feature maps were used to extract spatial features due to small stereo image patches (See: Figure 4.16). The various number of feature maps have been used to estimate the better accuracy. The tested various number of feature maps are 6, 8, 10, 12, 14, 16, 18 and 20. The six (6) feature maps produced better accuracy for both satellite and UAV stereo images in proposed CNN model, the reason is that the proposed CNN comprises less number of convolutional layers and a very small network, if feature maps size will increase, the true accuracy will decrease due to small CNN model size. If the CNN model size is big, then may be CNN model achieve high accuracy at higher number of feature maps. The small CNN model may produce high accuracy by using high number of feature maps and could produce overfitting due to huge number of neurons used in last fully connected layer. The actual accuracy could be different with estimated accuracy in overfitting process. The proposed CNN model has no overfitting and produced true accuracy using small CNN network with small number of data. The dropout layer with thirty (30) percent probability reduction used to minimize the overfitting in proposed design before last fully connected layer.

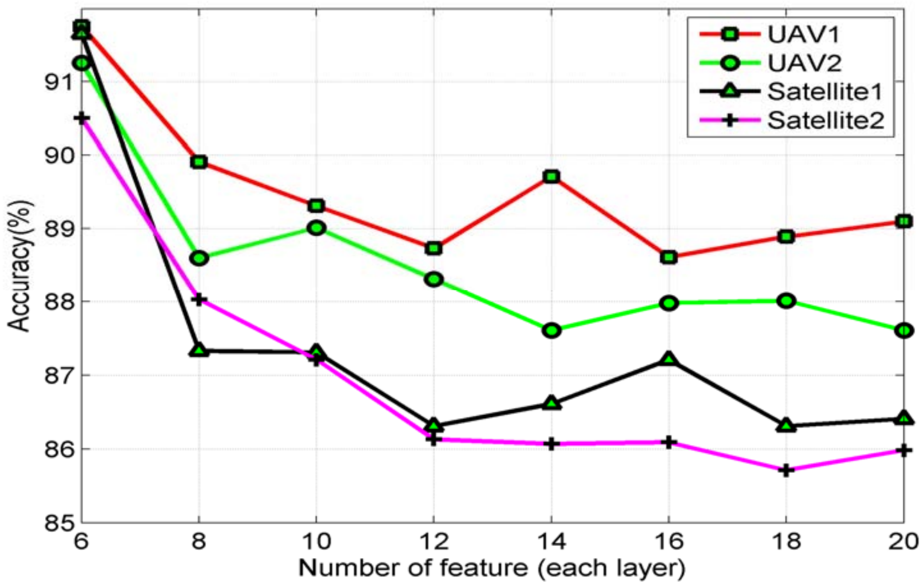
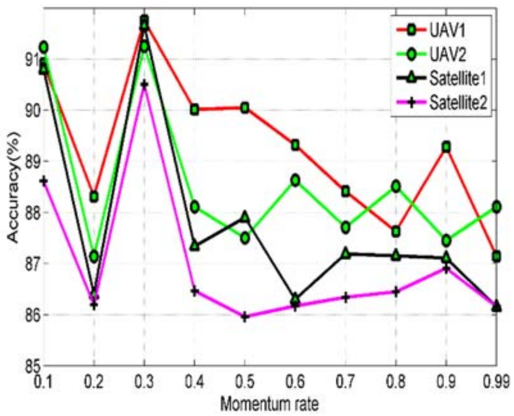


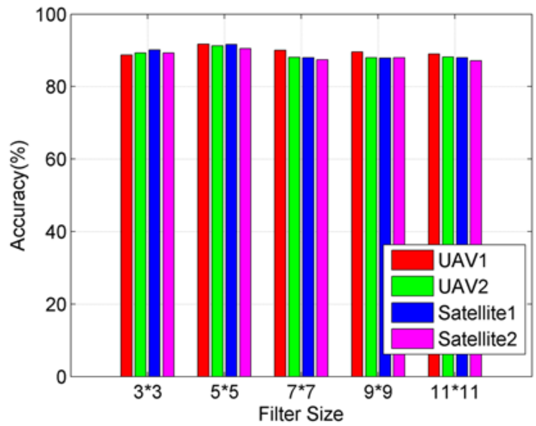
Figure 4.16: The number of features used in each number of layers

The momentum techniques can be used for searching the global minimum during optimization of stochastic gradient descent algorithm as soon as possible. These techniques tried to inhibit the model from accomplished to stuck in local minima. The different momentum values evaluated to measure the accuracy as shown in Figure 4.17.

The best accuracy achieved at 0.3 momentum value. The results of other momentum values are unbalanced due to misguiding the direction during optimization procedure. The higher momentum values slow the training parameters to reach global minimal point with minor increasing the accuracy. The different momentum value is chosen in this experiment to calculate the highest accuracy based on CNN design as shown in Figure 4.17 (a).



(a)



(b)

Figure 4.17: (a) Momentum rate used for optimization of CNN algorithm (b) Filter size used in CNN algorithm

4.3.3 Feature Visualization in CNN Design

The CNN power lies in the weights and in the starting point, the weights are selected randomly set with standard deviation 0.001. The weights of convolution kernel in the first convolutional layer showed in Figure 4.18 for UAV images. The visualization of each layer using satellite stereo images as shown in Figure 4.22. Two cases are considered in this example. First case contains two power poles and lines and second case also comprises two poles. The first column belongs to convolutional layer, second column consists of ReLU layer, the third column represents normalization layer and the last column has pooling layer for both cases as shown in Figure 4.22 and Figure 4.23. The visualization shows that the reconstruction using convolution layer provides useful feature of image and may be produce good visualization feature. The ReLU and pooling layer also produced more prominent features points in stereo images.

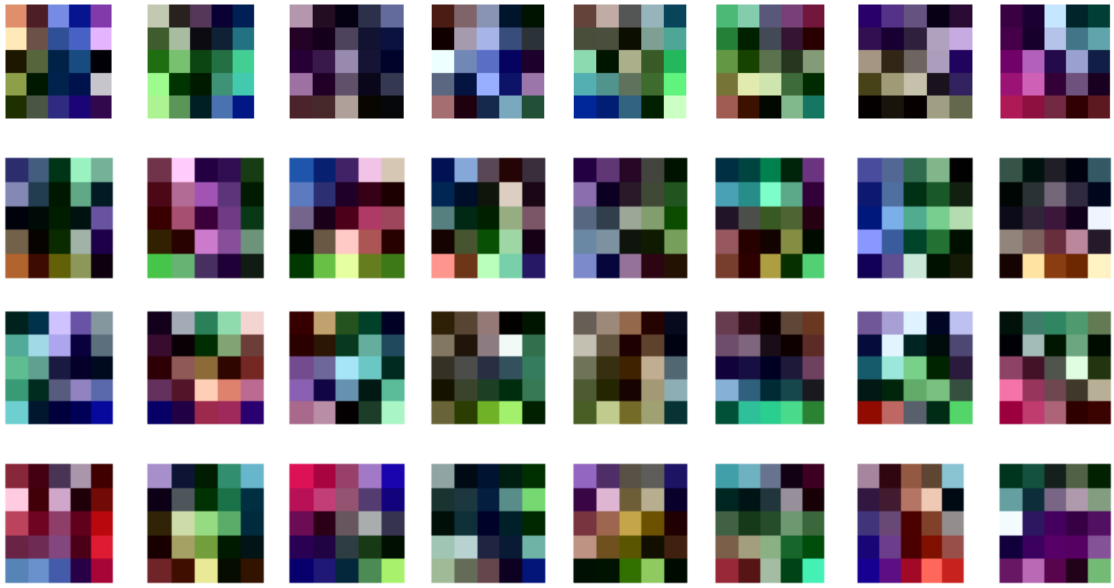


Figure 4.18: The filter weights used in first convolutional layer has size 4x4 due to 32 feature maps used in first convolutional layer.

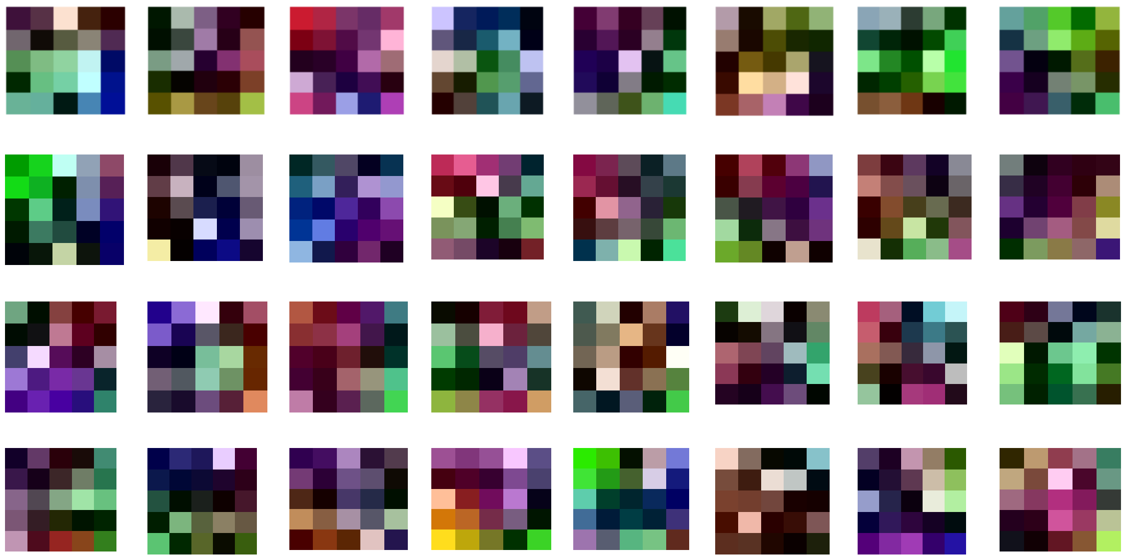
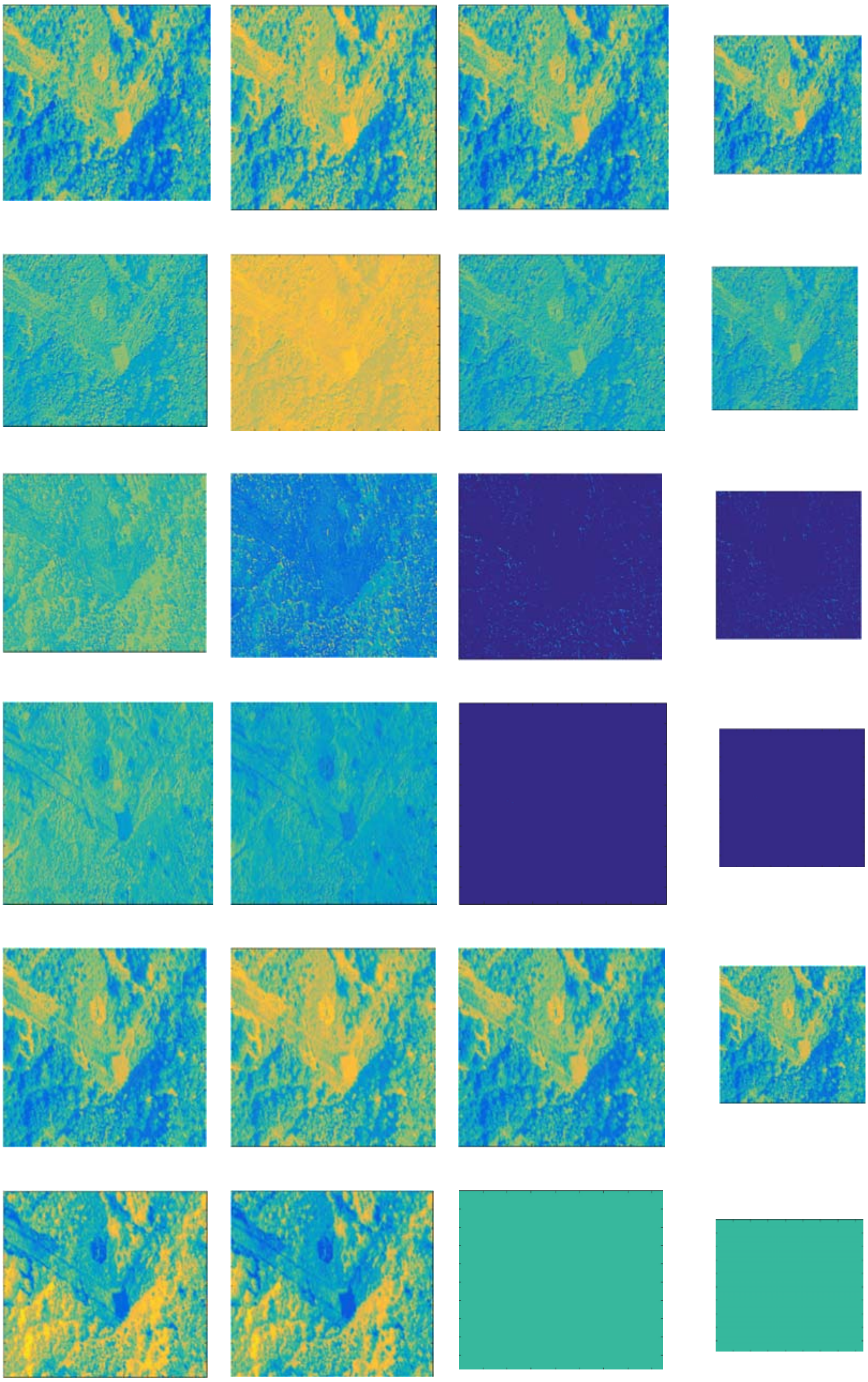


Figure 4.19: The first convolutional layer weights. Each image has size (4x4) for convolutional kernel. The 32 kernels use in first convolutional layer.

Figure 4.20: The feature maps based on different number of layers used in CNN algorithm for UAV images.



The weights and biases were used to estimate the training parameters of the proposed CNN models. The randomly generated weights provided the spatial information for the particular patch of the input image. The weights were optimized using stochastics gradient descent (SGD) algorithms based on mini-batch size. In order to determine the robust accuracy, the optimization of weights at each iteration is very important. The proposed mini-batch SGD was trained using various number of input images for optimization of the weights for initial CNN layers. For satellite images, the first convolutional layer in CNN algorithm has weights as shown in Figure 4.19.

The Feature used in CNN algorithm has been shown in Figure 4.20. The convolutional, pooling and normalization layer features have been computed. The features layer by layer has been shown in Figure 4.20 for convolutional, ReLU layer , normalization and pooling layer for stereo UAV image case1. Similarly, the second case of stereo image based on UAV has shown in Figure 4.21.

The visualization of each layer using satellite stereo images as shown in Figure 4.22. Two cases are considered in this example. First case contains two power poles and lines and second case also comprises two poles. The first column belongs to convolutional layer, second column consists of ReLU layer, the third column represents normalization layer and the last column has pooling layer for both cases as shown in Figure 4.22 and Figure 4.23. The visualization shows that the reconstruction using convolution layer provides useful feature of image and may be produce good visualization feature. The ReLU and pooling layer also produced more prominent features points in stereo images. These layers produced visualization capability of spatial feature in aerial images. These layers have a capability to extract prominent features in disparity map estimation and these layers also produced better estimation of feature mapping in stereo imagery for height estimation of objects near the power transmission line and power transmission poles. The visualization of stereo images based on a convolutional and pooling layer using CNNs model is shown in Figure 4.24. The representations of each CNN's layers have been visualized by using an image inversion technique, proposed for understanding the CNN activations intuitively. By using deeper layers, the images show more blurred responses. The CNN reconstruct deep features based on the strong visualization capability of CNN from stereo images.

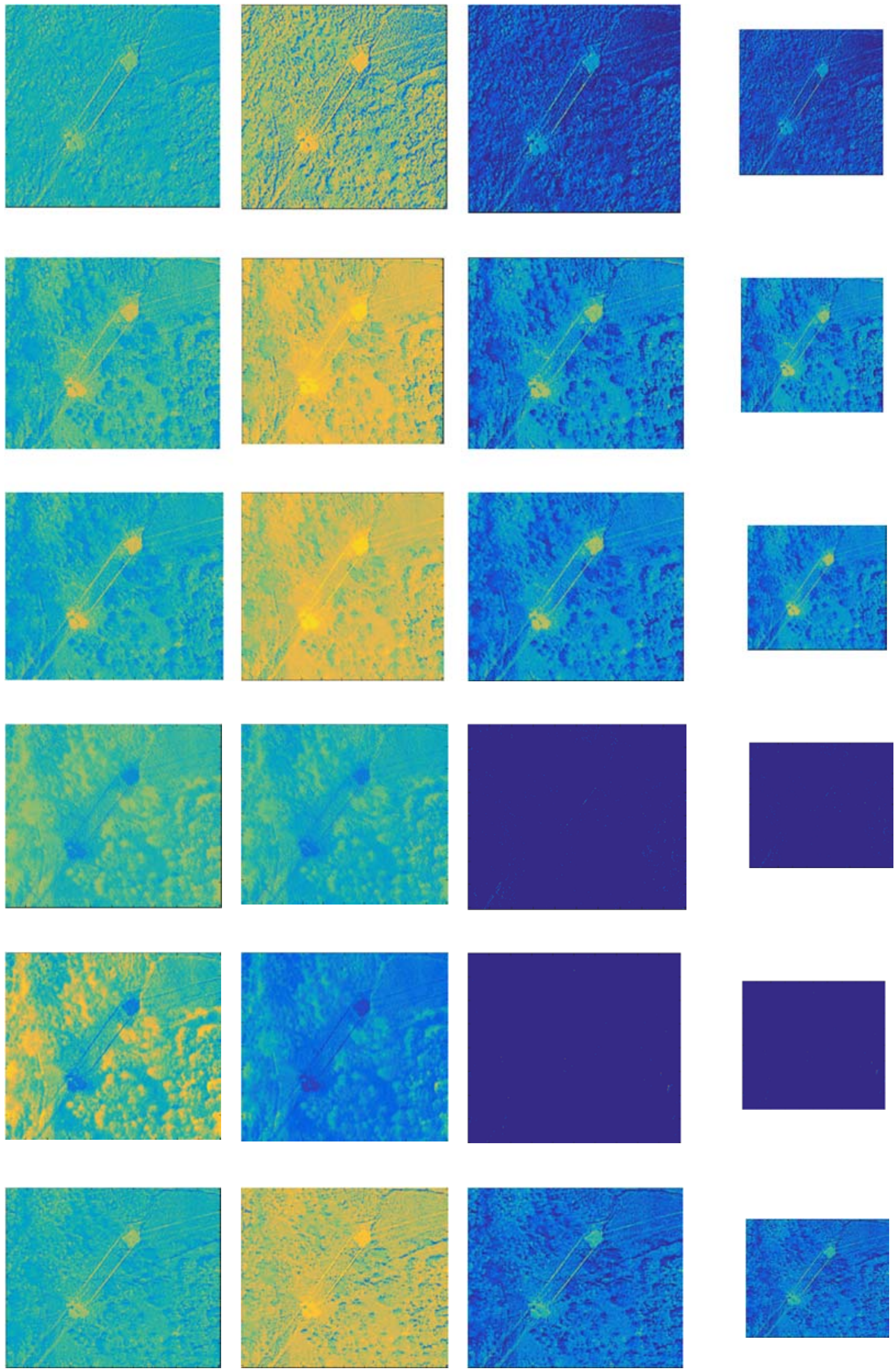


Figure 4.21: The feature maps based on different number of layers used in CNN algorithm for UAV images.

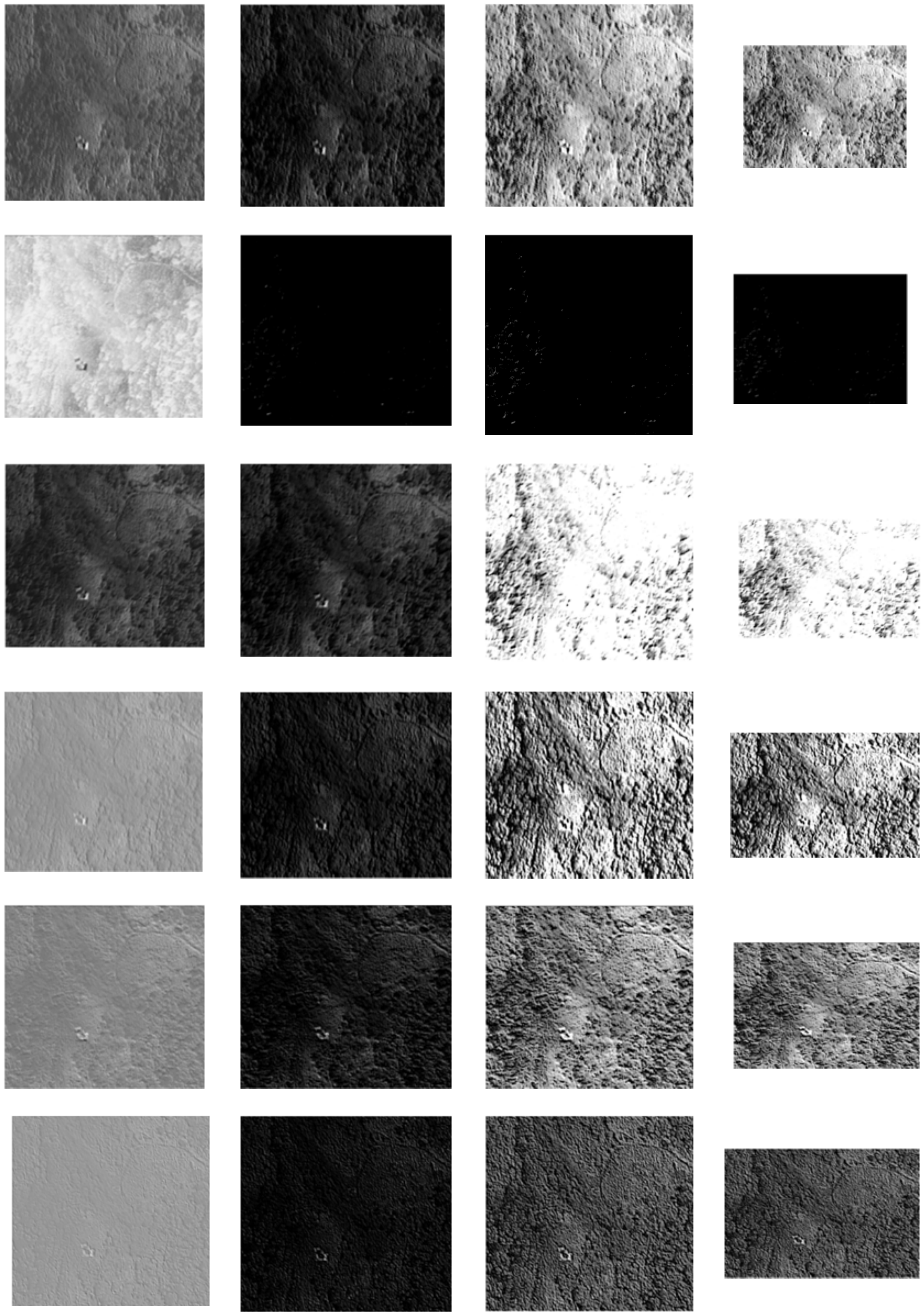


Figure 4.22: The feature maps based on different number of layers used in CNN algorithm for satellite images.

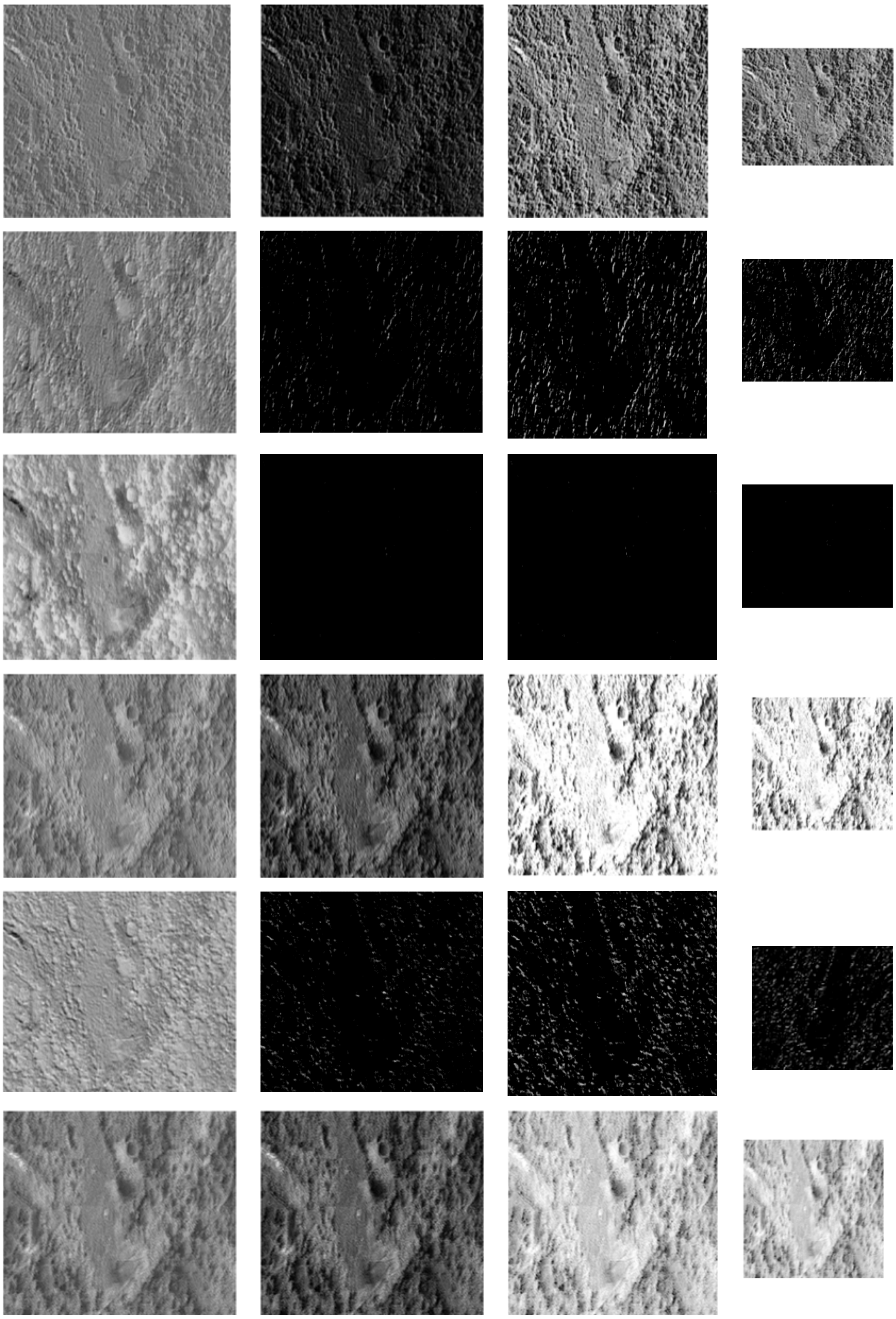


Figure 4.23: The feature maps based on different number of layers used in CNN algorithm for satellite images.

It is an excitement to see the representation of convolutional layers at each stage where images look more similar to the original image. The convolutional layer 1 and 2 has been reconstructed and shown based extracted spatial features and the 4th pooling layer also used to show the subsampling effect on the image. The power transmission poles are clearly reconstructed in the first stereo image by applying convolutional and pooling layer. Similarly, reconstructed images have been shown in the second case of stereo pair by using same convolutional and pooling layer.

Figure 4.24 shows stereo image visualizations reconstructed from CNN convolutional and pooling layers. Representations from each layer were visualized by image inversion to aid our intuitive understanding of CNN activity. At deeper layers, images were more blurred and it was an exhilarating experience to witness results from each level, as the photos increasingly resembled the original image. Based on extracted spatial features, examples from convolutional layers (c_1) one and two (c_2) were reconstructed that are shown in Figure 4.24, where power transmission poles are clearly reconstructed from the first stereo pair after convolutional and pooling application. Similarly, reconstructed images are shown for the second stereo pair. The second pooling layer showed a sub-sampling effect on the image. The pooling layers are denoted as (p_1) and (p_2). The input sample images are denoted as I_1, I_2, I_3 and I_4.

A reconstruction of the convolutional layers (c_1, c_2) and the pooling layers (p_1, p_2) are shown in Figure 4.25. The second convolution layer and rectified linear unit of the first stereo image clearly show visual accuracy compared to other images. Moreover, power transmission poles were identified by all layers for all images. Note that reconstructed area increased when the number of convolution and rectified layers were increased. Also, we observed a larger receptive field size in deeper layers compared to initial layers of the input image. The visual representations of each layer is created for each CNN in the proposed technique using image inversion to aid our intuitive understanding of CNN activation (See, Figure 4.26). Deeper layer images were more blurred but representations for each stage in convolutional layer processing progressively showed images looking more similar to the original. Features based on fully connected layers cannot be properly recognized as these features comprise similar but numerous meaningful components that are randomly distributed.

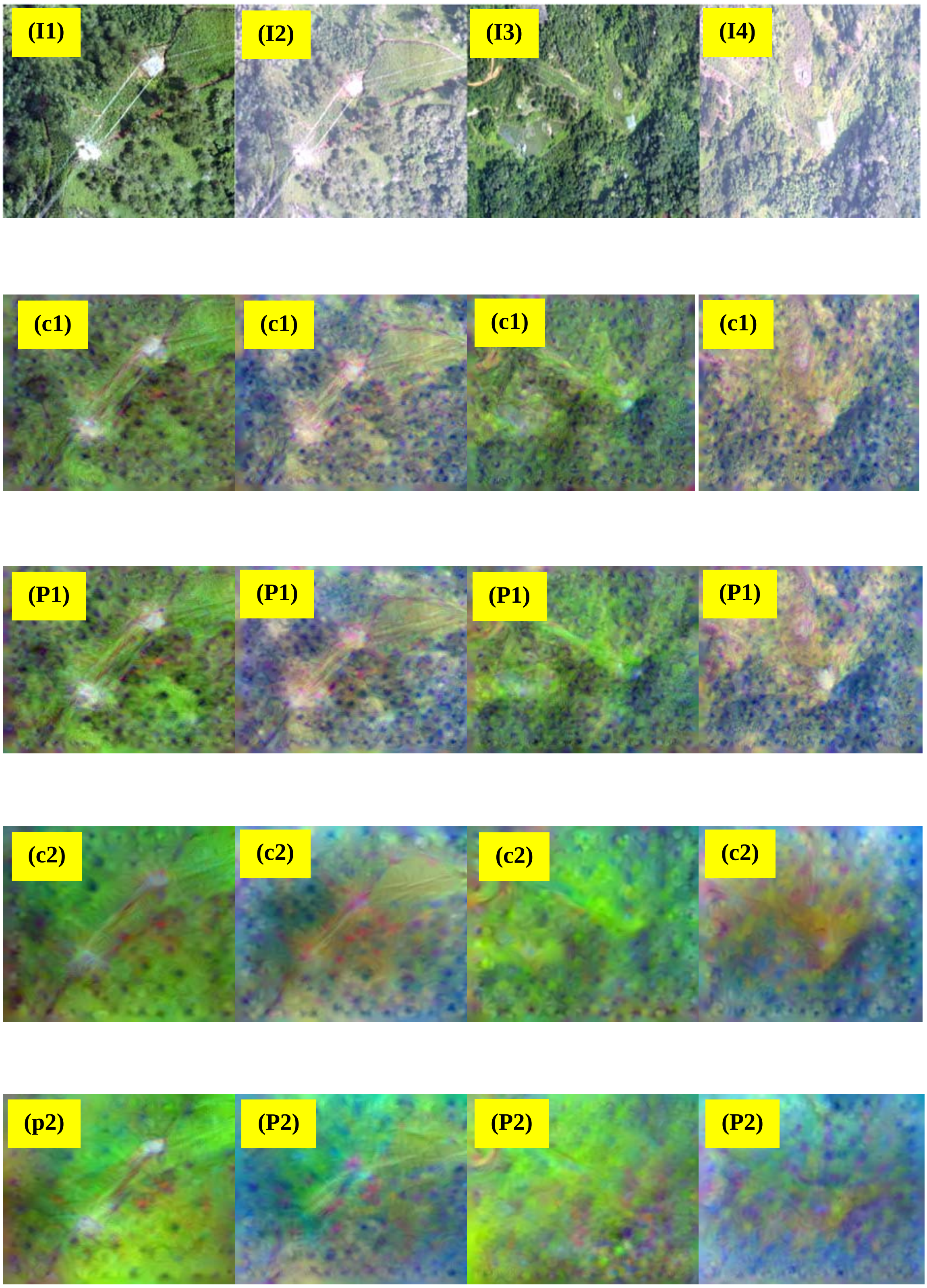


Figure 4.24: Visualization of stereo images using CNN based on convolutional and pooling layer. The Convolutional layer (c1, c2) and pooling layer (P1, P2).

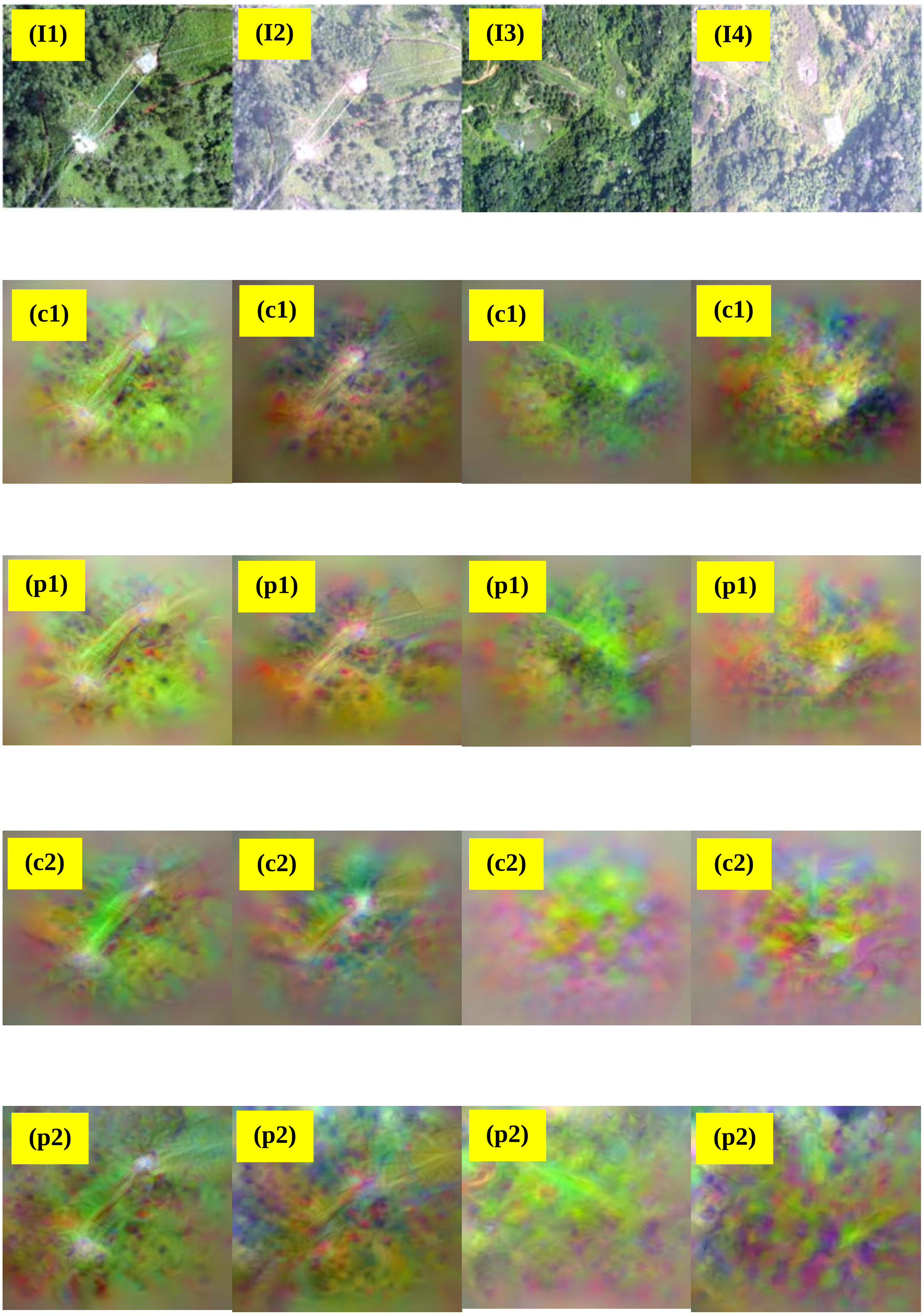


Figure 4.25: Visualization of stereo images using CNN based on convolutional and pooling layers. The Convolutional layer (c1, c2) and pooling layer (p1, p2).

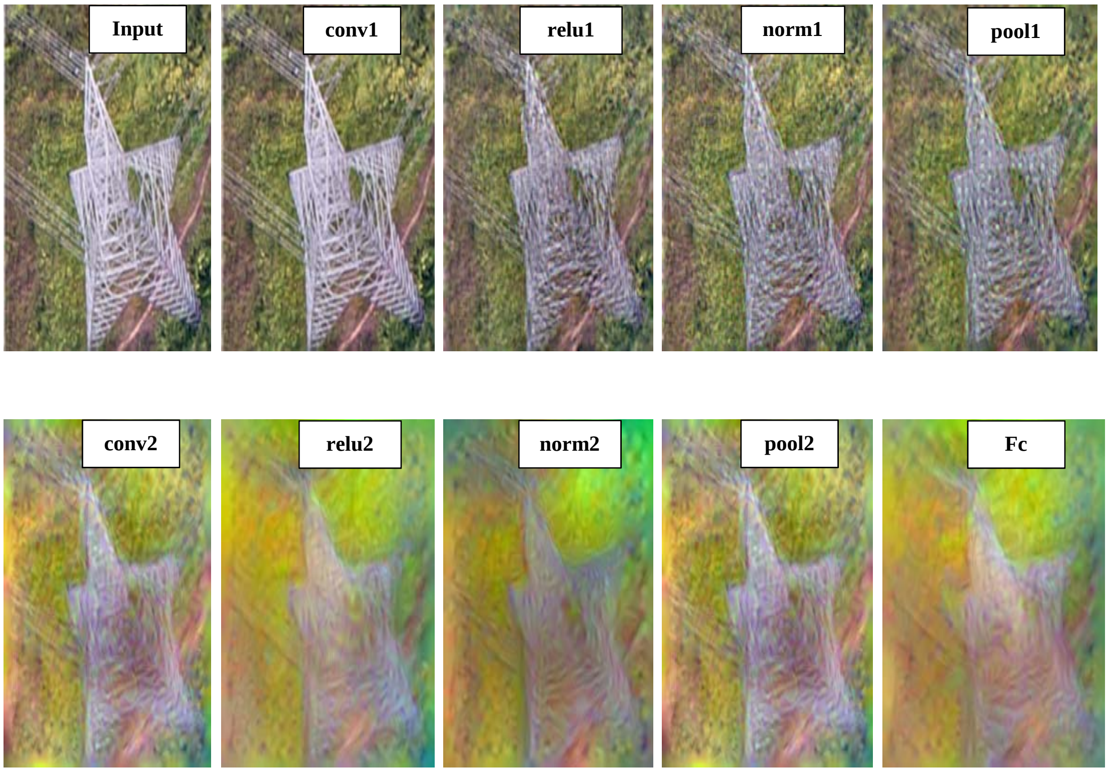


Figure 4.26. Reconstruction of images using proposed CNNs model. The Convolutional layer (conv1, conv2), rectified linear unit (relu1, relu2), normalization (norm1, norm2), pooling layer (pool1, pool2) and fully connected layer (FC).

Using only the last convolutional layer, Figure 4.26 shows reconstructions of different image samples with clear representations using various layers in CNN algorithm. These layers are based on convolutional $(conv1, conv2)$, rectified linear units $(Relu1, Relu2)$, normalization $(norm1, norm2)$ at multiple levels and last fully connected layer (FC). The proposed CNN algorithm extract spatial properties of the aerial images and extracted the prominent features in aerial imagery. The proposed CNN well reconstructed the prominent objects from the aerial images using various number of layers (convolutional, rectified, normalization and pooling layer). Consequently, it appears that each fully connected layer shows abstracted representations of the original image by using low-level layer data based on random rearrangements. These layers reconstructs spatial features from the stereo images and also reconstructs prominent features from the image for stereo matching process using aerial stereo imagery.

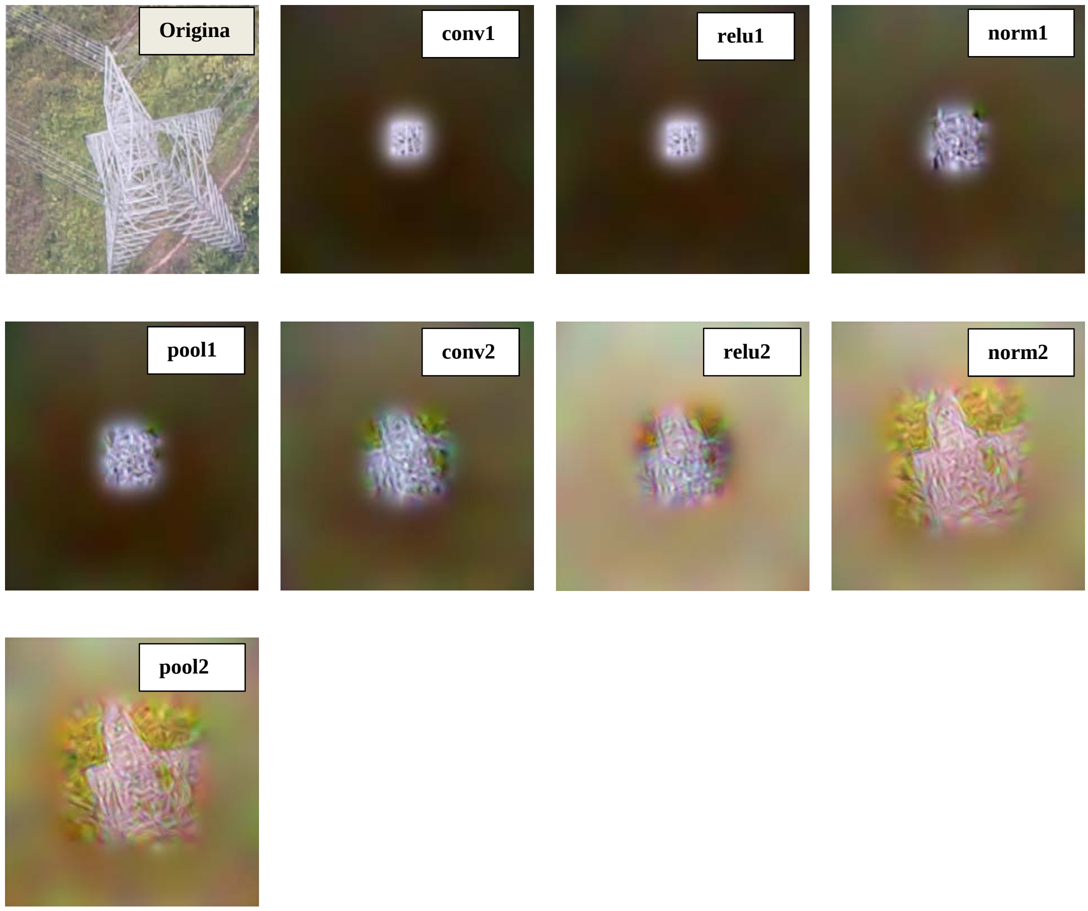


Figure 4.27. Input images; Last convolutional and pooling layers. The Convolutional layer (conv1, conv2), rectified linear unit (relu1, relu2), normalization (norm1, norm2), pooling layer (pool1, pool2) and fully connected layer (FC).

Figure 4.27 shows reconstructions based on local regions using different convolutional layers with different feature maps. The size of receptive fields increased when applying these results to deeper layers, becoming larger compared to input images. The four number of samples based on tower, building, trees and roads have been used to see the visualization capability of CNN algorithm based on convolutional, ReLU, normalization and pooling layers is shown in Figure 4.28. It can be concluded that the spatial deep properties of feature maps based on different convolutional and pooling layer has strong visualization capability to extract the spatial features. Similarly, the second convolutional and 2nd pooling layers produced accurate activation function and efficiently reconstruct the visualization features in four selected number of image samples is shown in Figure 4.28.

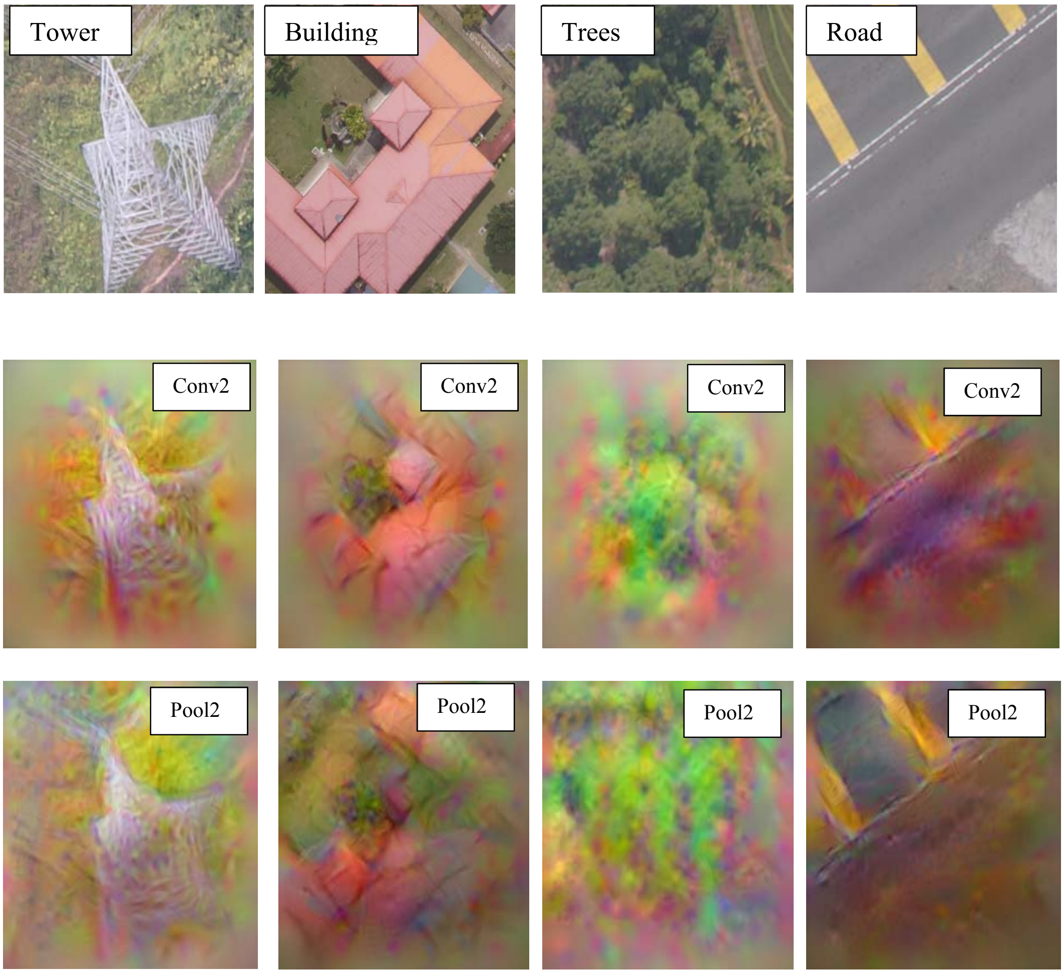


Figure 4.28. Convolutional and pooling layers with different receptive fields. The convolutional (Conv2) and pooling layer (Pool2).

4.3.4 Training and Testing Errors Rate

The main emphasis in this work is to analyze the over-fitting problems in the proposed CNNs algorithm. The training and testing losses are the indications that whether the deep CNN networks have over-fitting problems or not. The training and testing losses has been evaluated using proposed CNN algorithm based on stereo image patches. The over-fitting occurs due to hyper-parameters settings, for some cases, huge hyperparameters values show that over-fitting happens. By increasing the number of hidden units increases the capacity of the system which shows that the learning rate is within the allowable range and in this situation no under or over-fitting occur. For some cases, allowable weight decay coefficient zeros produced greatest effective system capacity and less error rate due over-fitting occur for small hyper-parameters sitting. Gradient descent algorithm unintentionally increases the training error when the learning rate is very high. In another case, on small learning rate, the training is slower and the system becomes stuck due to a high training error. The network model is either over-fitting or under-fitting diagnose by monitoring both training and testing error by tuning the parameters instead of learning rate. Higher training error determines the system is in over-fitting and requires regularization with appropriate hyper-parameters. Without using regularization, the optimization algorithm can work better by using more number of layers or adding more hidden units. Unfortunately, the significantly computational cost increases related to the deep CNN model. The deep CNNs model performs better by minimizing training and test error gap, the gap between test and train error should be minimized. The gap between train and test error can be reduced by changing hyper-parameters regularization such as weight decay or dropout. The best performance of deep CNNs models can be achieved by using the larger network and properly sitting regularization hyper-parameters such as dropout. In our designs networks provides better performance due to the small gap between test and train error. It is difficult to recognize that whether the system is underfitting or any other defect if both test and train errors are high.

The hyper-parameters (dropout and batch normalization) could be used as an alternative by improving the model capacity (less gap between training and testing error) for the small dataset. If the training and testing gap in proposed CNN algorithm is not in acceptable range, even by improving the regularization hyper-parameters then more data should be required for achievable optimal performance. By developing the machine learning models day by day, researchers try to use different set of datasets with different training sample size (increase or decrease), adjust regularization parameters effect on deep models capacity in terms of error rates and computational resources(memory, runtime), change hyperparameters of training optimization algorithms, need to improve optimization techniques related to deep CNN algorithms. The training and testing losses have been evaluated with or without dropout layer as shown in Figure 4.29. Four numbers of stereo images have been used to optimize CNN algorithm with and without dropout layer to evaluate the over-fitting layer. The small number of patches in both stereo images has been used in order to calculate the disparity map. Due to small dataset, the over-fitting problem could be occurring, in order to avoid over-fitting, the training and testing error can be calculated. The training and testing error with and without dropout has been evaluated. The higher test error without dropout layer shows over-fitting and lower test error with dropout technique indicates the CNN have no over-fitting problem. Similarly, the other stereo image has the same pattern for training and testing losses.

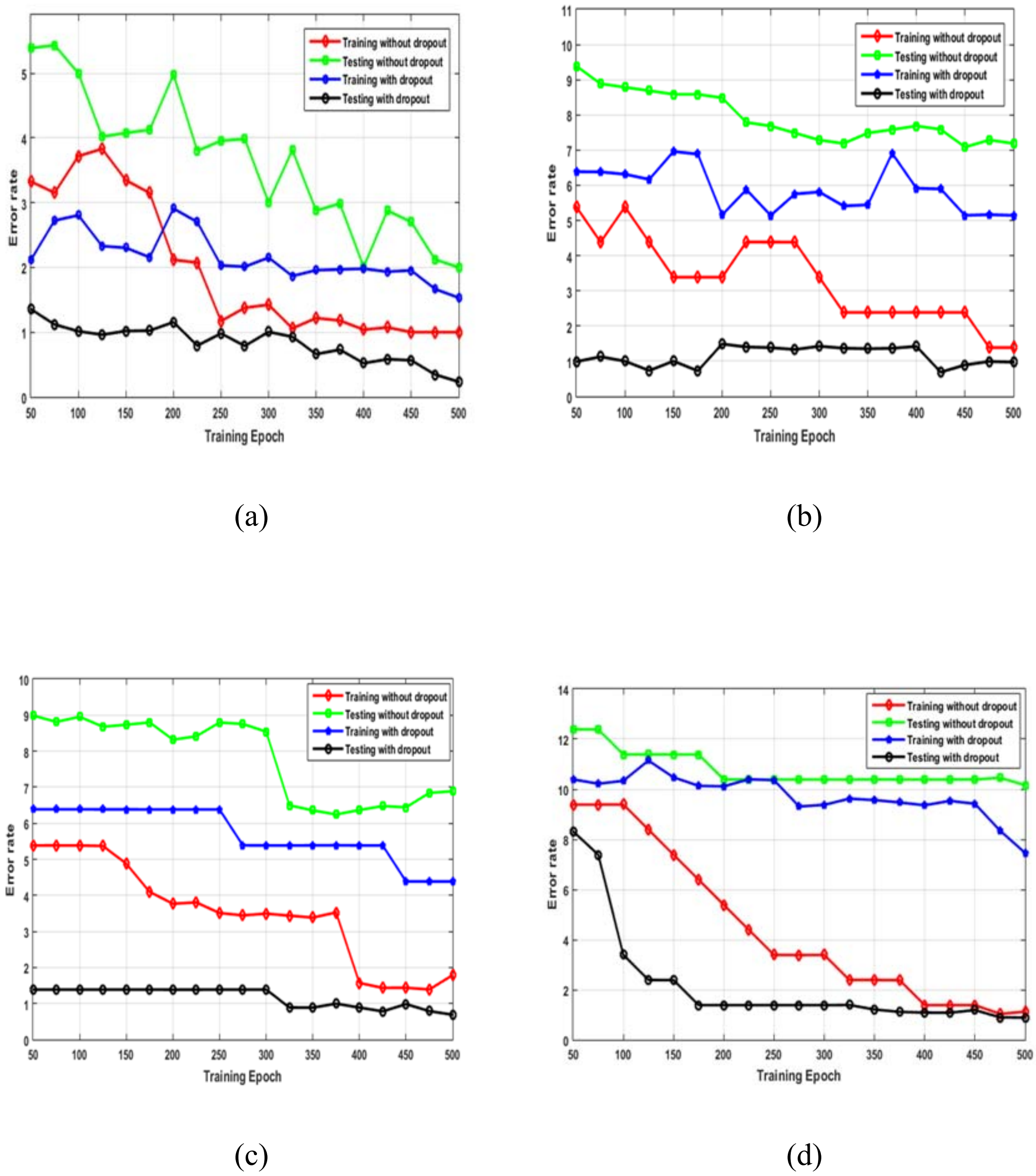


Figure 4.29: The training and testing error based on proposed CNN models using stereo images: (a) Image1 (b) Image2 (c) Image3 (d) Image4.

4.4 Height Estimation based on Disparity Map Algorithms

The performance matrices based on accuracy, precision and recall has been evaluated based on existing as well as proposed algorithms (Sparse Representation and CNN based algorithms).

4.4.1 Accuracy Comparison between Proposed and Existing Algorithms

The height of the vegetation and trees near power transmission poles and lines can be estimated using our proposed methods (CNN and sparse representation). After the power poles height estimation, the height of vegetation and trees near these transmission poles is estimated. The height of these objects (vegetation and trees) is estimated on the right of ways (near or around the power poles and power lines), and their height estimation is based on the standard distance. The most important parameters related to our area of interest are the bamboo trees; these trees grow up very fast and may touch the power transmission lines that may cause the blackouts or flashovers. The disparity maps are estimated in aforementioned sections for the selected images by using the proposed algorithms. In these experimental results, the number of disparity pixels was provided to the proposed algorithms for the height estimation of a particular object. The precision, recall and accuracy of approximated height of vegetation and trees has been measured from the computed disparity map. The number of pixels has been extracted and converted into meters for measuring the height or depth of a particular object using the proposed disparity map algorithms. Based on GCPs, the extracted area has certain vegetation and tree blocks near power transmission poles. In this way, the height of any object can be measured based on GCPs of that object.

The actual height of power transmission poles used as a ground truth height and estimated height has been evaluated based on proposed CNN based algorithm and existing algorithms. In this section, the performance matrices are evaluated based on the ground truth and estimated height of power transmission poles using ten (10) number of power transmission poles and 50 number of iterations performed for calculating the performance metrics (accuracy, precision and recall). By using stereo matching algorithm the height of power transmission pole, trees and vegetation can be estimated. To compute the error or accuracy based on estimated height or ground truth, different performance metrics are required. The objective is to estimate the accuracy, precision and recall performance metrics based on proposed and existing algorithms. Comparisons of the proposed algorithm with existing algorithms (GC, DP, BP, SSD, NCC, and BMA) revealed that our CNN algorithm achieved higher accuracy (90%). UAV image accuracy, precision and recall values (proposed technique vs. existing algorithms) are shown in in Figure 4.30 and Figure 4.31 using disparity estimation based on UAV stereo imagery. The proposed approach obtained higher accuracy compared to extant algorithms due to its high spatial feature extraction capability. The CNN-based height estimations also garnered better precision and recall values when compared with existing height estimation techniques.In this experiment, the different number of parameters has been explored using CNN based methods. The most important and fundamental parameters have been tuned in the CNN technique are the filter bank, k, weights of filter and bias term b. These parameters have been trained and updated using gradient descent method based on the mini-batch processing unit. In this step no pre-processing step is required, only parameters selection is needed to alleviate. CNN can effectively extract spatial-related high-level features in remote sensing imagery. The CNN based method learned the spatial-related parameters layer by layer from the image, while traditional spatial feature extraction methods use manual setting parameters. However, the configuration of CNN can greatly affect the accuracies in terms of spatial feature extraction.

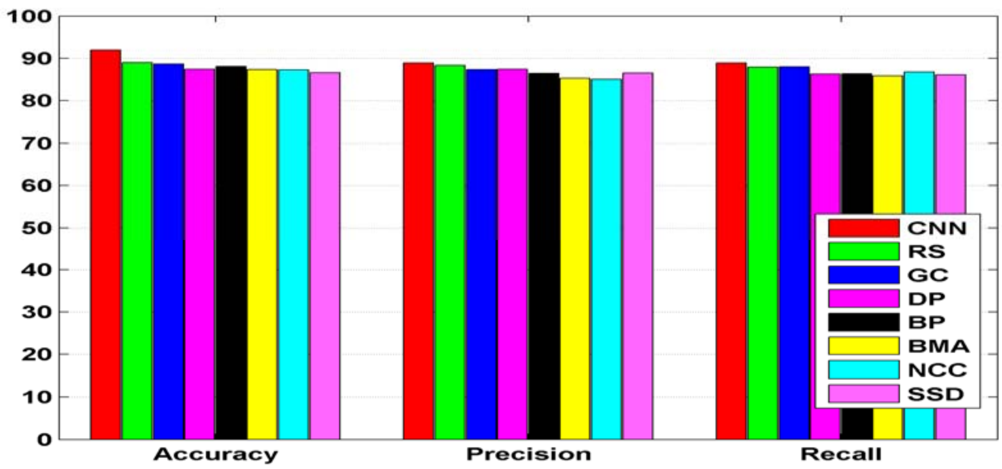


Figure 4.30: Comparison of performance metric: accuracy, precision, recall based on our proposed and existing algorithms for case 1 UAV stereo images.

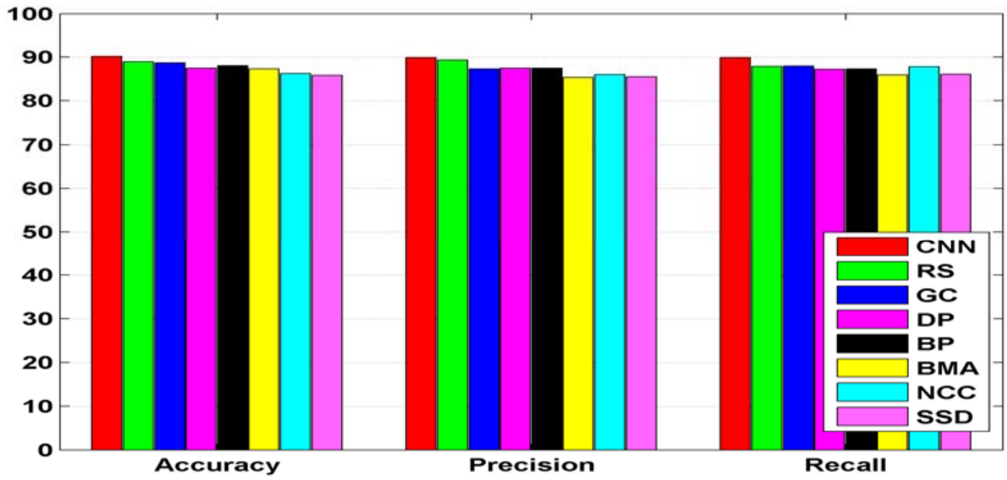


Figure 4.31: Comparison of performance metric: accuracy, precision, recall based on our proposed and existing algorithms for case 2 UAV stereo images.

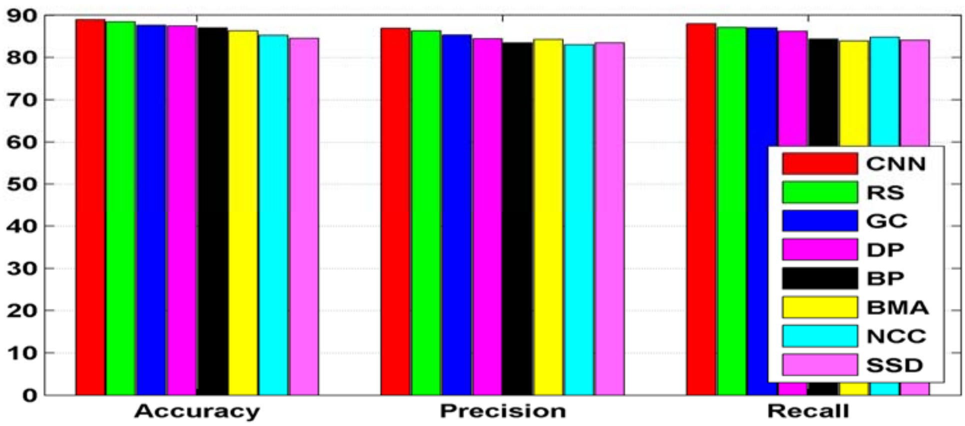


Figure 4.32: Comparison of performance metric: accuracy, precision, recall based on our proposed and existing algorithms for case 1 satellite stereo images.

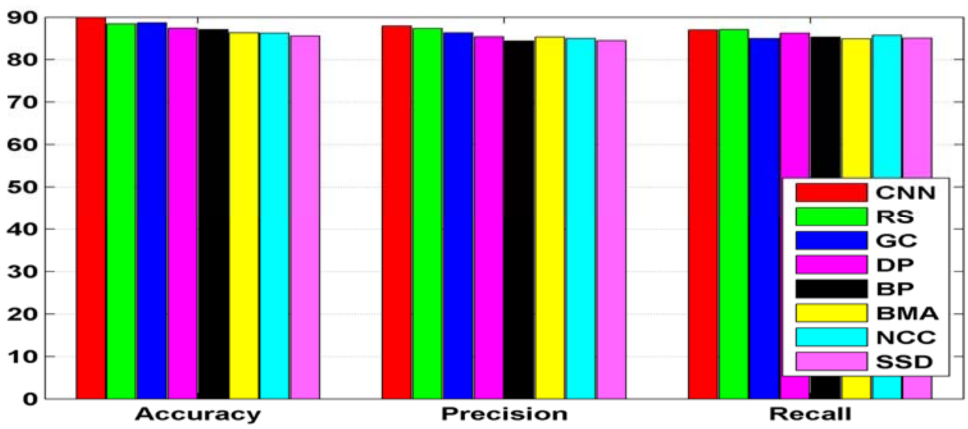


Figure 4.33: Comparison of performance metric: accuracy, precision, recall based on our proposed and existing algorithms for case 2 satellite stereo images.

The accuracy, precision and recall based on CNN proposed technique and existing algorithms are shown in Figure 4.32 and Figure 4.33 using satellite images. The CNN based approach achieved high accuracy as compared to existing algorithms due to high spatial feature extraction capability. The CNN based height estimation provides better precision and recalls values as compared to existing height estimation techniques. Different disparity values were used to compute object height by the proposed algorithm (See: Figure 4.34). The highest accuracy rate; albeit, with increased computational time, was obtained for a disparity of 40 for both UAV and satellite images. We noted that higher disparity values did not produce higher accuracy for either UAV or satellite stereo image pairs. The highest disparity value used for this experiment was 100.

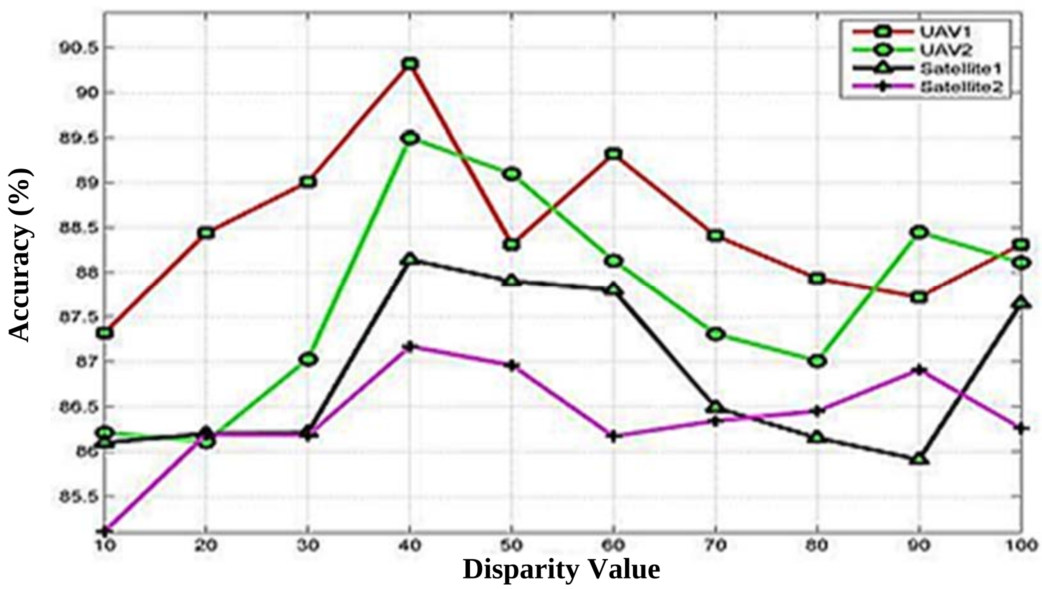


Figure 4.34: The different disparity value has been used to compute the accuracy using CNN algorithm.

Three cases have been considered to understand the process of height estimation of vegetation and trees based on aerial stereo images. The disparity values of 10, 20 and 30 have been selected in chronological order where 10 is the smaller disparity value, whereas 20 represents the greater disparity value and 30 is the higher disparity value. It is observed by visual inspection and quantitative analysis that disparity map based on displaced value (i.e. 20) has calculated slightly more height in comparison to the disparity value (i.e. 10). The disparity value 30 could produce higher accuracy as compared to previous cases (disparities=10, 20). The comparison of height estimation in terms of accuracy, precision and recall is shown in the following sections. The results show that proposed methods (CNN and sparse representation) are reliable and pertinent for power line inspection system and can detect obstacles (vegetation and trees) within the power transmission line corridor with practicable accuracy. The experimental results show that the height estimation of trees and vegetation has lower approximation values as compared to power transmission poles. The proposed matching methods are suitable for measuring vegetation and tree heights near power lines and poles based on UAV and satellite stereo imagery to avoid the future threats of electrical hazards. Comparison of precision, recall and accuracy based on the disparity map has been evaluated suing UAV and satellite stereo images. The twenty (20) numbers of power poles and disparity value 20 was used in the first experiment. The accuracy, precision and recall have been shown in the Table 4.1, Table 4.2 and Table 4.3 for UAV stereo images. The fifty (50) numbers of iterations have been selected to measure accuracy, precision and recall for this experiment.

Table 4.1: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=20, Total towers=20**) for UAV images.

Algorithms	**Tower Height (Accuracy %)**
Graph-Cut Algorithm	83.45±1.7
Sparse Representation Algorithm (proposed)	**88.75**±1.3
CNN based Algorithm (proposed)	**91.25**±1.2

Table 4.2: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=20, total number of towers=20**).

Algorithms	**Precision (%)**
Graph-Cut Algorithm	83.56±1.3
Sparse Representation Algorithm (proposed)	**88.25**±1.5
CNN based Algorithm (proposed)	**90.97**±1.7

Table 4.3: Comparison of height in terms of recall between existing and proposed algorithms for UAV stereo images (**disparity=20, total number of towers=20**).

Algorithms	Recall (%)
Graph-Cut Algorithm	85.03$\pm$1.8
Sparse Representation Algorithm (proposed)	**88.15**$\pm$1.4
CNN based Algorithm(proposed)	**89.49**$\pm$1.6

The accuracy, precision and recall using our proposed and existing stereo matching algorithms for satellite images is shown in Table 4.4, Table 4.5 and Table 4.6. The height estimated using ground truth height and estimated height produced by disparity map evaluated by our proposed algorithms for 50 number of iterations. Similarly, the precision and recall have been computed for the same specification used for measurement of accuracy. These performance metrics has been evaluated based on proposed and existing stereo matching algorithm.

Table 4.4: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=20, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Graph-Cut Algorithm	81.46$\pm$1.9
Sparse Representation Algorithm (proposed)	**87.50**$\pm$1.3
CNN based Algorithm (proposed)	**90.50**$\pm$1.5

Table 4.5: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=20, total number of towers=20**).

Algorithms	Precision (%)
Graph-Cut Algorithm	80.66$\pm$1.9
Sparse Representation Algorithm (proposed)	**86.59**$\pm$1.7
CNN based Algorithm (proposed)	**88.79**$\pm$1.5

Table 4.6: Comparison of height in terms of Recall between existing and proposed algorithms for satellite stereo images (**disparity=20, total number of towers=20**).

Algorithms	Tower Height
Graph-Cut Algorithm	81.23±1.8
Sparse Representation Algorithm (proposed)	**87.15**±1.9
CNN based Algorithm (proposed)	**89.49**±1.3

The accuracy, precision and recall using our proposed and existing stereo matching algorithms for UAV images with ten (10) disparity value and 20 number of power poles is shown in Table 4.7, Table 4.8 and Table 4.9. The accuracy may reduce due to lower disparity value range between two stereo images and it could be increase for higher disparity value. In this analysis, the disparity value is set optimal by using different number of disparity range for aerial images (UAV and satellite). The number of iteration and number of power poles for calculating accuracy, precision and recall has been same (50 iterations, 20 number of power poles) for this experiments. The various number of disparity values has been tested to check the feasibility of the proposed algorithms (CNN and sparse representation).

Table 4.7: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=10, Total towers=20**) for UAV images.

Algorithms	Tower Height (Accuracy %)
Graph-Cut Algorithm	82.05±1.7
Sparse Representation Algorithm (proposed)	**87.15**±1.5
CNN based Algorithm (proposed)	**90.11**±1.9

Table 4.8: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=10, total number of towers=20**).

Algorithms	Precision (%)
Graph-Cut Algorithm	82.16±1.4
Sparse Representation Algorithm (proposed)	**87.75**±1.6
CNN based Algorithm (proposed)	**90.17**±1.4

Table 4.9: Comparison of height in terms of Recall between existing and proposed algorithms for UAV stereo images (**disparity=10, total number of towers=20**).

Algorithms	Recall (%)
Graph-Cut Algorithm	84.73±1.4
Sparse Representation Algorithm (proposed)	**87.05**±1.8
CNN based Algorithm(proposed)	**89.11**±1.6

Table 4.10: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=10, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Graph-Cut Algorithm	80.36±1.7
Sparse Representation Algorithm (proposed)	**86.70**±1.6
CNN based Algorithm (proposed)	**89.90**±1.3

The proposed algorithm produced comparatively better performance parameters using disparity ten (10) in stereo matching algorithms and provides better solution as compared to the extant stereo matching algorithms for height estimation.

Table 4.11: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=10, total number of towers=20**).

Algorithms	Precision (%)
Graph-Cut Algorithm	79.26±1.6
Sparse Representation Algorithm (proposed)	**86.13**±1.4
CNN based Algorithm (proposed)	**88.19**±1.1

By using satellite images, the acquired results almost generate similar pattern for accuracy, precision and recall for the same specifications (disparity and number of power poles) as shown in Table 4.10, Table 4.11 and Table 4.12. Further, we consider third scenario where the disparity values were set high and same number of power poles were used.

Table 4.12: Comparison of height in terms of recall between existing and proposed algorithms for satellite stereo images (**disparity=10, total number of towers=20**)

Algorithms	Tower Height
Graph-Cut Algorithm	80.23±1.9
Sparse Representation Algorithm (proposed)	**86.95**±1.8
CNN based Algorithm (proposed)	**88.01**±1.4

The disparity was set to 30 and 20 number of power poles were selected. The accuracy is shown in Table 4.13 while precision and recall are shown in Table 4.14 and Table 4.15 for UAV stereo images.

Table 4.13: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=30, Total towers=20**) for UAV images.

Algorithms	Tower Height (Accuracy %)
Graph-Cut Algorithm	83.65±1.7
Sparse Representation Algorithm (proposed)	**88.95±1.6**
CNN based Algorithm (proposed)	**91.75±1.4**

Table 4.14: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=30, total number of towers=20**).

Algorithms	Precision (%)
Graph-Cut Algorithm	83.96±1.6
Sparse Representation Algorithm (proposed)	**88.95±1.7**
CNN based Algorithm (proposed)	**91.07±1.9**

The accuracy comparison using satellite stereo images based on disparity (30) and total number of power poles (20) is shown in Table 4.16. The precision and recall for the same number of specification (30 disparity value and 20 number power poles) is shown in Table 4.17 and Table 4.18.

Table 4.15: Comparison of height in terms of Recall between existing and proposed algorithms for UAV stereo images (**disparity=30, total number of towers=20**).

Algorithms	Recall (%)
Graph-Cut Algorithm	85.71±1.7
Sparse Representation Algorithm (proposed)	**88.75±1.5**
CNN based Algorithm(proposed)	**89.99±1.4**

Table 4.16: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=30, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Graph-Cut Algorithm	83.65±1.6
Sparse Representation Algorithm (proposed)	**88.95±1.3**
CNN based Algorithm (proposed)	**91.65±1.2**

Table 4.17: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=30, total number of towers=20**).

Algorithms	Precision (%)
Graph-Cut Algorithm	83.76±1.8
Sparse Representation Algorithm (proposed)	**88.89±1.8**
CNN based Algorithm (proposed)	**90.98±1.3**

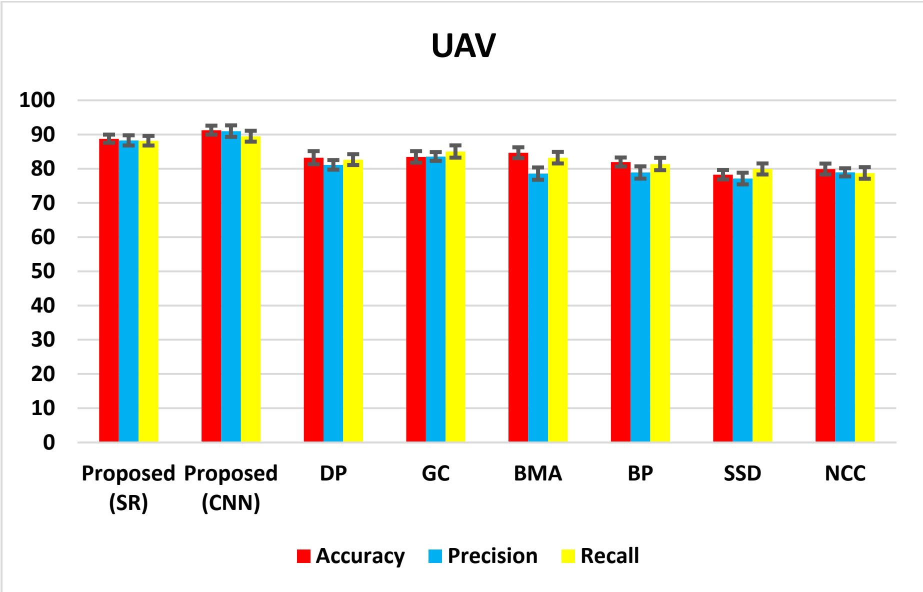
The system's aerial images can be reliably used to monitor the height of power poles likewise the vegetation and trees near these transmission poles has been detected and estimated based on the proposed CNN algorithm in digital terrain. The CNN produced highest performance by using the disparity (30) using UAV. The height estimation of power transmission poles has been measured based on proposed and existing algorithms. Different number of power transmission poles and different number of disparity value used in this experiment to estimate the height of objects.

Table 4.18: Comparison of height in terms of Recall between existing and proposed algorithms for satellite stereo images (**disparity=30, total number of towers=20**).

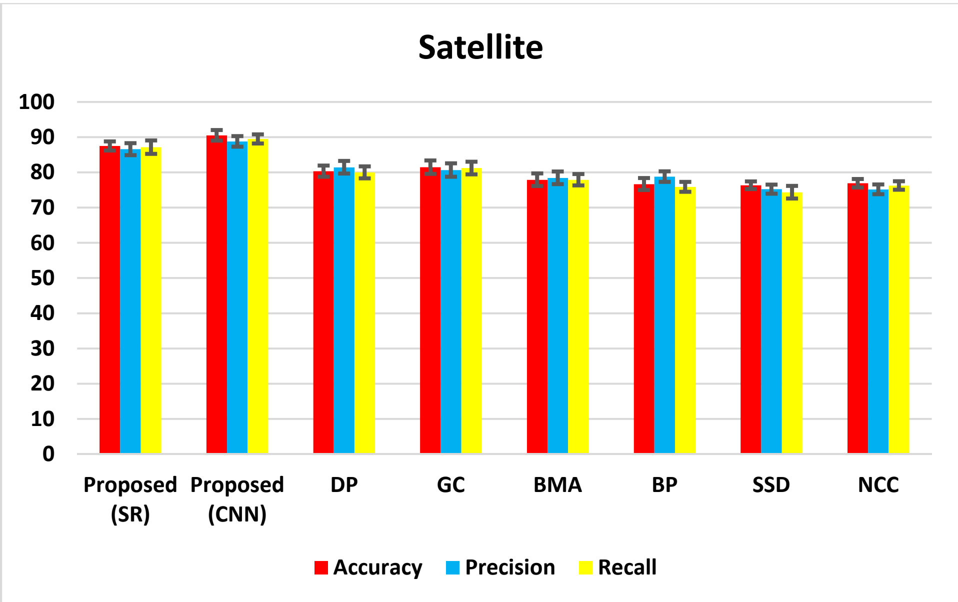
Algorithms	Recall (%)
Graph-Cut Algorithm	85.83±1.4
Sparse Representation Algorithm (proposed)	**88.75±1.3**
CNN based Algorithm(proposed)	**90.01±1.6**

The number of iterations remained constant throughout the experiment for calculating accuracy, precision and recall. The results show that by increasing the disparity value, the accuracy, precision and recall increased in a small fraction. The comparison between all existing state-of-the-art stereo matching algorithms and proposed algorithms in terms of accuracy, precision and recall has been discussed in details in appendix C.

The accuracy precision and recall for UAV and satellite stereo images based on proposed and existing algorithms is shown in Figure 4.35 (a, b) for disparity 20 and 20 number of power transmission poles. In case of displaced value (d=20), the accuracy is comparatively better in our proposed algorithms in comparison with existing algorithms. The proposed CNN produced high performance as compared to extant stereo matching algorithms as well as proposed sparse representation using UAV stereo images. In case of satellite stereo images, the proposed algorithm also achieved better results. The satellite stereo images have a capability to measure the performance using proposed technique and can be used for height and distance estimation. The proposed algorithms estimate height and distance of objects near power transmission poles and lines. These parameters (distance and estimated height) further used for performance analysis (accuracy, precision and recall) using UAV and satellite stereo images. Similarly for disparity 10 and total number of 20 power poles, the comparison between performance parameters is shown in Figure 4.36 (a, b) for both UAV and satellite datasets.



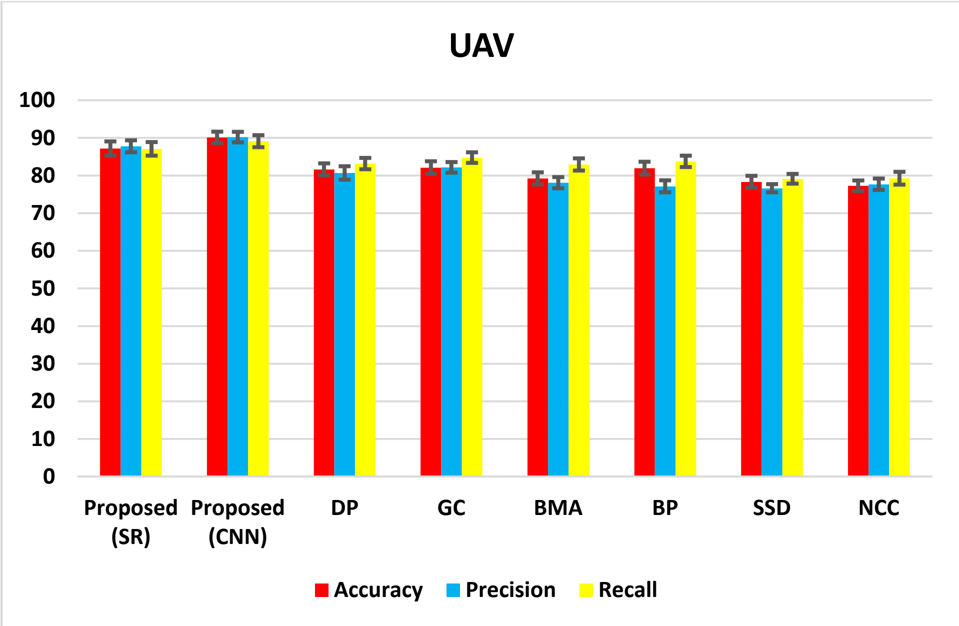
(a)



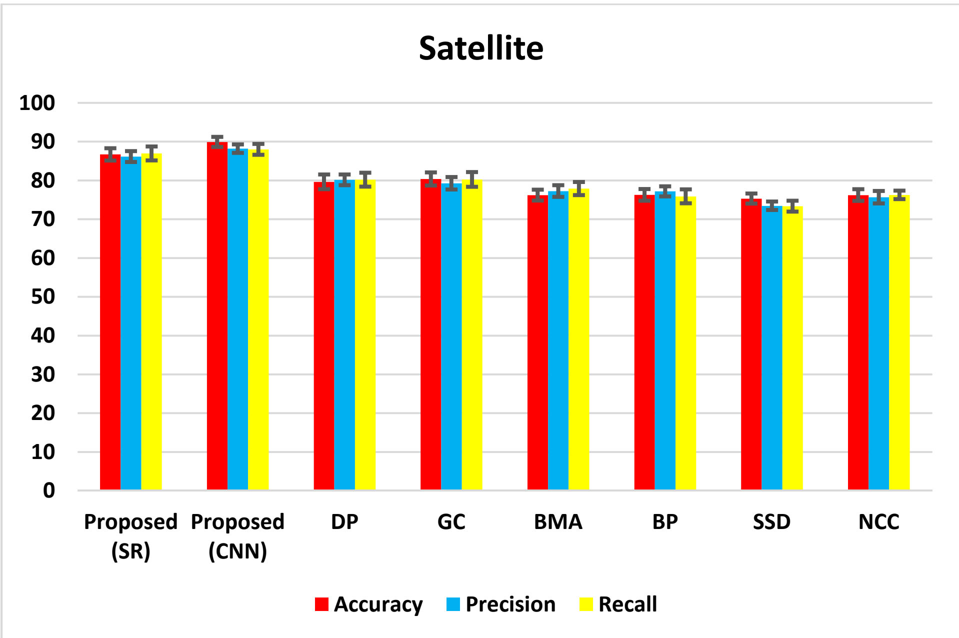
(b)

Figure 4.35: The accuracy, precision and recall (a) UAV images (d=20, Tower=20) (b) satellite images (d=20, Tower=20).

The system performance has been increased by increasing the disparity and twenty disparity map tends to produce better performance as compared to ten disparity. The performance in accuracy, precision and recall increased using the disparity twenty (20) for UAV stereo images. The proposed CNN algorithm produced highest performance as compared to extant stereo matching as well as proposed sparse representation algorithm. The results based on proposed CNN clearly indicate the performance.



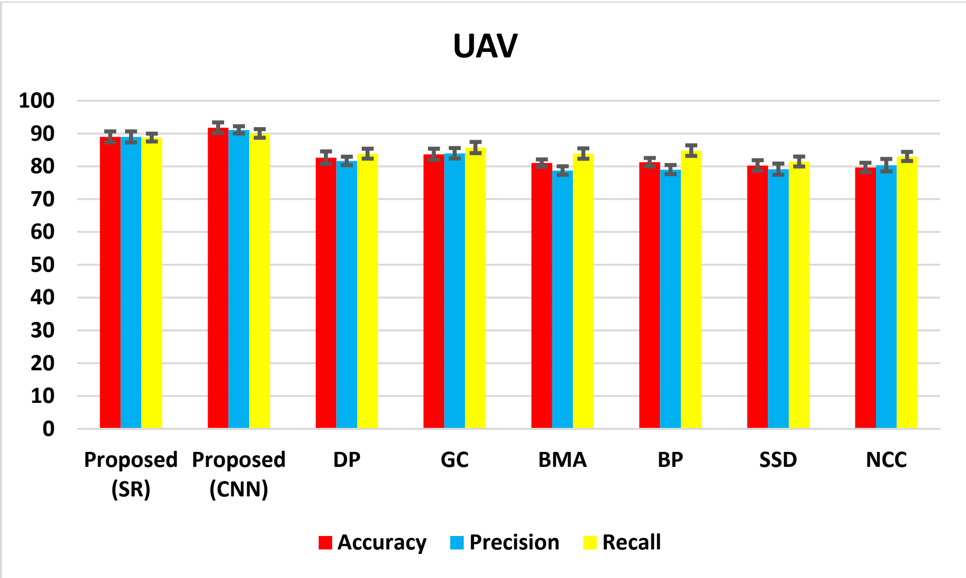
(a)



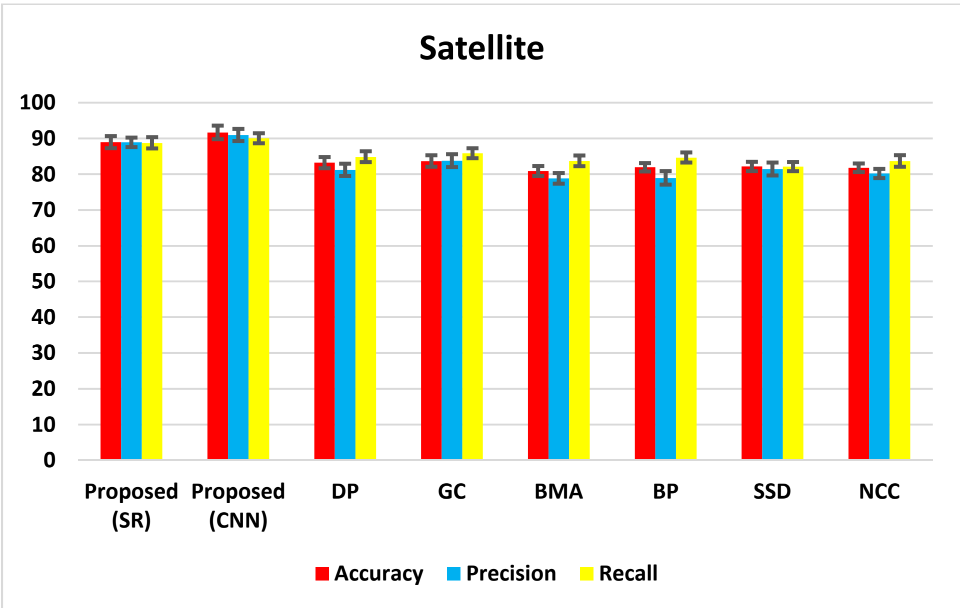
(b)

Figure 4.36: The accuracy, precision and recall (a) UAV images (d=10, Tower=20) (b) satellite images (d=10, Tower=20)

The proposed algorithms produced efficient solution to estimate the better disparity map and estimate better height as compared to existing stereo matching algorithms. The disparity twenty (20) is a good approximation for height estimation and provides better performance metrics as compared to existing algorithms. The proposed CNN algorithm achieved highest performance as compared to other stereo matching algorithms.



(a)



(b)

Figure 4.37: The accuracy, precision and recall (a) UAV images (d=30, Tower=20) (b) satellite images (d=30, Tower=20).

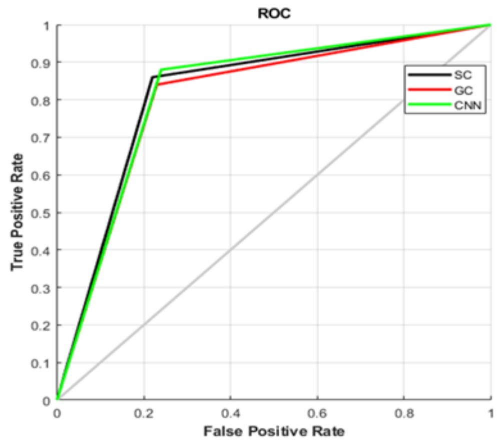
The accuracy is slightly less in case of small displaced value (d=10) using our proposed algorithms and existing algorithms. The accuracy precision and recall for disparity 30 and power transmission poles 20 is shown in Figure 4.37 (a, b) for UAV and satellite. The proposed algorithms provided higher accuracy using disparity value (d=30) for both UAV and satellite datasets as compared to aforementioned disparity values (d=10, 20). The standard deviation also measured based on the accuracy, precision and recall performance metrics. These standard deviation values indicated on the accuracy, precision and recall graph for bot UAV and satellite datasets with disparity values (10, 20 and 30). It is evidence from experimental results, the higher will be the disparity value the higher will be the performance as compared to lower disparity values. However, with increasing the disparity value the computational complexity will increase by searching the correspondences points between stereo images in each iteration. It can be concluded that the proposed algorithms for height estimation provided the more accurate height as compared to extent stereo matching algorithms for all disparity values.

4.4.2 ROC curves based on Precision and Recall Parameters

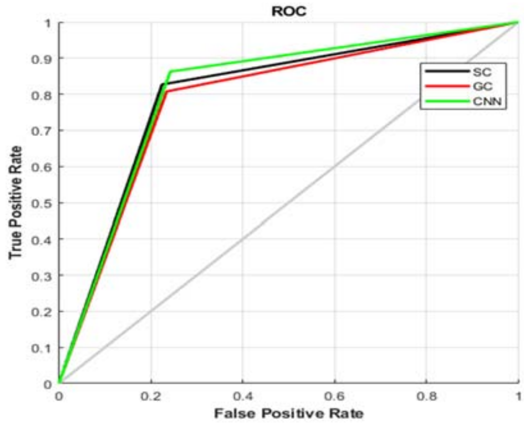
In statistics, a receiver operating characteristic curve, i.e. ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is a fundamental tool for diagnostic test evaluation. In a ROC curve the true positive rate (precision) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter. The true-positive rate is also known as sensitivity, recall or probability of detection in machine learning. The false-positive rate is also known as the fall-out or probability of false alarm and can be calculated as (1 − specificity). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups. This curve shows the sensitivities and specificities associated with different possible values of the classification threshold. Determination of the classification threshold requires a compromise between sensitivity and specificity, because both cannot be maximized simultaneously. Thus, a highly sensitive classification threshold (i.e., with a very low omission error) is associated with moderate specificity, and will lead to overestimation of certain vegetation monitoring area. An underestimation of vegetation monitoring areas would be obtained using a high specific threshold.

Thus, a model with a high specificity (0.900) is selected to minimize commission errors (false positives) in case of CNN algorithm accuracy comparison as compared to sparse representation and existing stereo matching algorithms. The region of convergence based on precision and recall values has been shown in Figure 4.38. It shows the different number of disparity values are used to measure the height of power poles using proposed algorithms. The estimated height and actual height can be used for precision and recall values estimation. The ROC curve is estimated based on UAV and satellite dataset using proposed CNN and sparse representation algorithm. These ROC curves are compared with the curves produced by existing GC (graph-cut) stereo matching algorithm. The ROC curve shows that the proposed algorithm produced higher performance curve as compared to existing algorithms. The curves produced by proposed CNN algorithm have higher values as compared to sparse representation and existing algorithm for both UAV and satellite datasets. The details of ROC map produced by proposed and all existing stereo matching algorithms has been provided in appendix C.

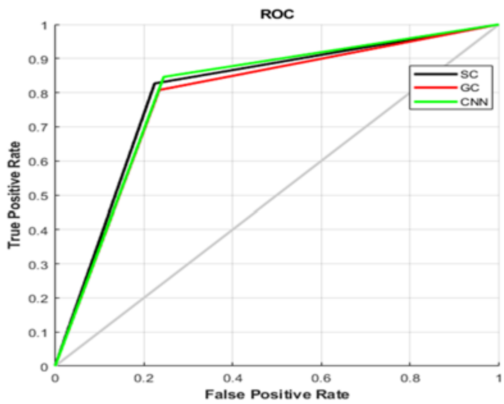
Effective monitoring of vegetation near power lines helps avoids losses and damages that interfere with power transmission and is an essential responsibility of power company management. We proposed an aerial-based computer vision algorithm to monitor vegetation near power lines / poles, one that provides a superior solution to existing stereo matching algorithms. We optimized the proposed CNN algorithm by using patches from left and right stereo input images. Different numbers of hyper-parameters were employed to train the algorithm based on scratch input patches. Our CNN technique captures prominent features in stereo images due to its strong spatial feature capability, and also identifies these features, as presented in right stereo images, to obtain reliable disparity map estimations. Thus, the CNN addresses stereo issues efficiently due to its robust feature extraction capability, and without any need for uniqueness constraints such as GC and DP.



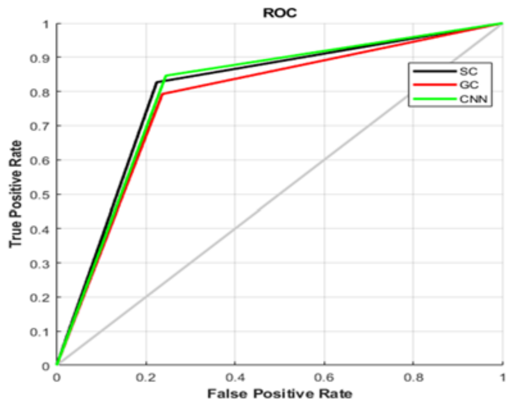
(c)



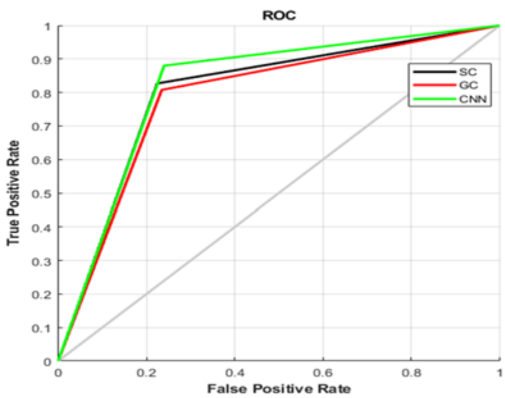
(b)



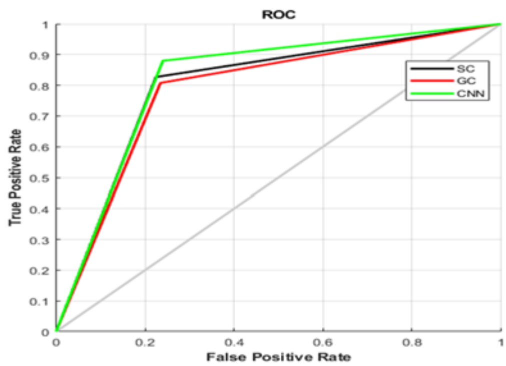
(c)



(d)



(e)



(f)

Figure 4.38: The ROC curves based on precision and recall; first column represented using UAV and second column represented Satellite based on disparity (d=10, 20, 30).

The proposed CNN algorithm overcomes this problem and thus provides a feasible disparity map for height estimates due to its ability to robustly store cascaded spatial features. However, the proposed program consumes additional time compared to extant algorithms. This is a disadvantage that needs to be tackled by minimizing computation time when using small stereo patches. This application is, however, offline, which is why the time component held little importance during this investigation. Several ROW threats were accurately identified based on height and distance estimations that were computed and based on the proposed algorithm's disparity map. Moreover, they and were also validated by performance metrics for accuracy, precision and recall via comparison with actual ground truth. The proposed computer vision based CNN algorithm (CNN) therefore provides an optimal solution for the monitoring of threats from vegetation and trees proximal to power transmission poles and lines by detecting hazards using aerial imagery.

4.5 Threat Levels Measurement based on Proposed Height Estimation Algorithm

There are three threat levels exist according to standard calculations. These three threat levels are called high, low and medium threat. The area of interest consists of fifty one (51) power transmission poles for satellite and 19 number of power transmission poles for UAV stereo images. Due to memory constraints and fast processing of satellite and UAV images, the area has been divided into number of segments. Each segment has eleven power poles (11). These segments have been used to measure the height of dangerous objects (vegetation / tress) near the power transmission lines / power poles. Due to very big size of satellite and UAV stereo images, the disparity map is calculated on small patches of stereo images for both satellite and UAV. Each patch contains three power poles and the disparity is measured automatically on cropped small images based on proposed stereo matching algorithm. The height is measured based on disparity map using proposed and existing algorithms. According to proposed algorithm, the height of each tower and distance of vegetation / trees near power transmission lines is measured in each segment. The accuracy, precision and recall values are calculated using estimated height and ground truth height of each tower in each segment. There are different number of threats found based on estimated height and distance of each tower outside in the right-of-the-way of trees and vegetation. According to standard distance measurement between objects and power transmission poles, the threat levels are divided into three levels. These threat levels are called high threat, medium threat and low threat. The threat levels have been marked by using different color on the disparity map. The red shows that high threat level occurs, yellow color shows medium and green color shows low threat level.

In this way the different number of threats are quantified and based on estimated distance between vegetation / trees and power lines, the number of threats are identified. The two dataset has been used to measure the distance between vegetation, power poles and lines, the first dataset based on satellite stereo images and second dataset based on UAV stereo images. The visualization of few number of threats (high, medium and low) has been shown in Figure 4.39. The green color shows the low number of threats, red color shows the high number of threats and yellow color represents the medium color of threats on the disparity map using three number of power transmission poles and power lines. Threats identified in UAV images by the proposed CNN are shown in Figure 4.40. Estimated distances were quantified according to accepted standards and classed as high, medium or as discussed previously. Several ROW threats were identified.

The standard clearance distance between a power pole and any object is 4 meters. Hence, any object within this range is considered a high threat. Between 4–6 m is considered a medium threat and ≥ 7 m is considered low threat. Thus, we quantified different threats based on estimated distances and two datasets were used to measure distances between vegetation and power poles / lines. The first was taken from satellite stereo images and second from UAV stereo images. The details of each threat based on proposed algorithms have been given in appendix E. The visualization of various threats using satellite stereo images based on proposed algorithms has been highlighted on the disparity maps and denote the various number of threats (high, medium, low) using various colors (red, yellow and green). The better distance and height estimation provided accurate estimation of different number of threats near power transmission lines and power transmission poles. These threats are determined based on the standard height and distance provided by the user and estimated the estimated height.

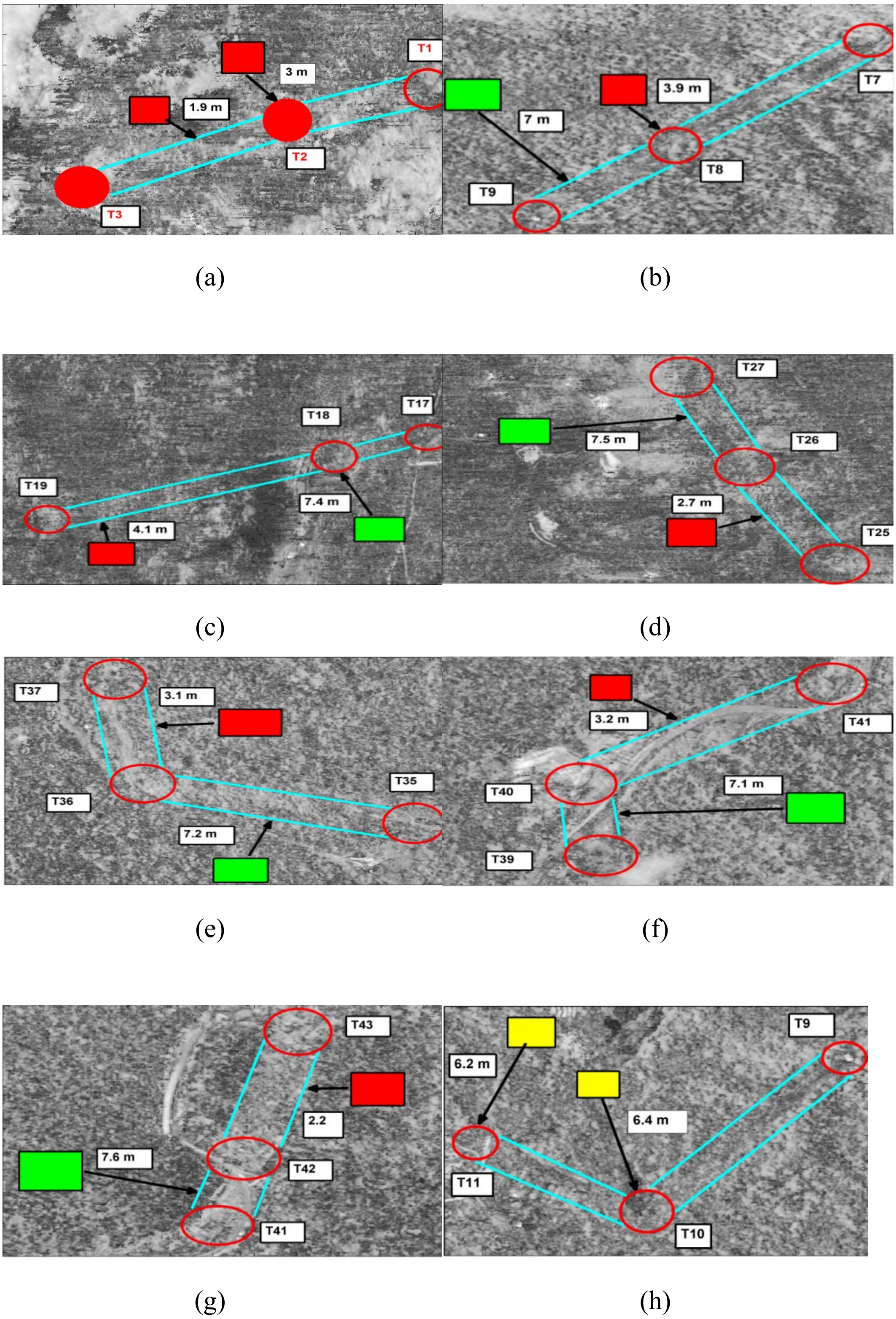


Figure 4.39: Number of high threats computed by the proposed algorithm based on estimated height. The red boxes shows the high threat level, yellow boxes show the medium threat level and green boxes show low threat level.

The different number of threat levels are quantified using sparse coding and CNN algorithms based on satellite and UAV stereo images. The number of high, low and medium threats estimated by CNN algorithm using satellite stereo images are shown in Figure 4.40. For sparse coding technique, the number of low, medium and high threats are found in Figure 4.41using satellite stereo images. The number of threats found based on UAV images using proposed sparse coding algorithm are shown in Figure 4.41. The threats are quantified based on standard distance between vegetation or trees near power transmission poles and power lines with comparison of estimated distance.

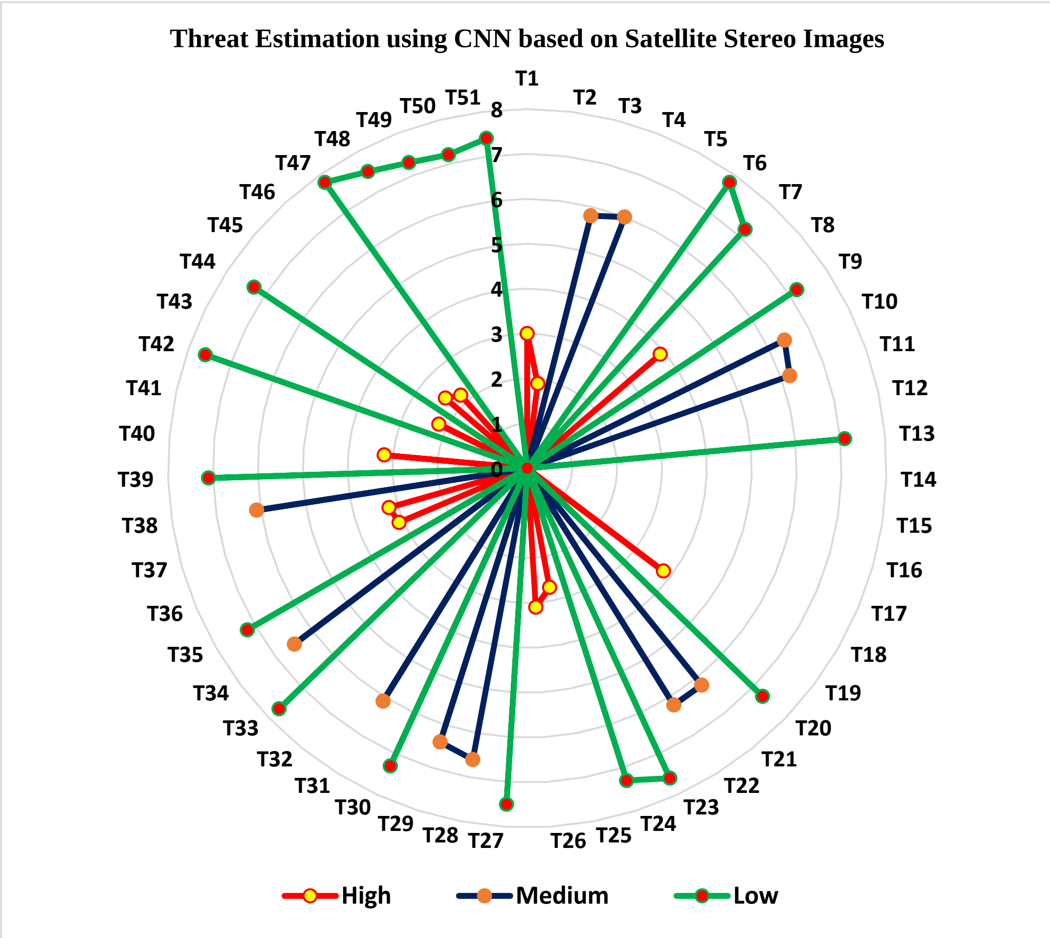


Figure 4.40: The number of threats found based on satellite stereo images using proposed CNN algorithm.

The various number of threats has been identified based on proposed algorithms using UAV and satellite stereo images. The proposed CNN based on UAV produced highest performance and estimate better threats (high, medium and low) level as compared to existing stereo matching algorithms.

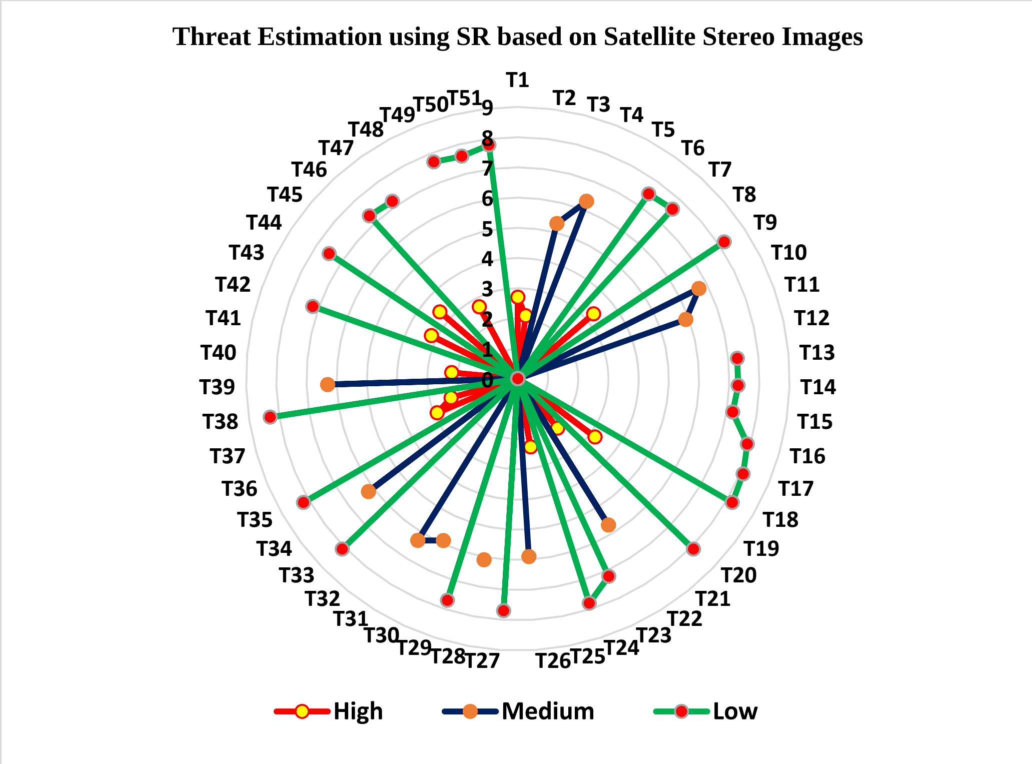


Figure 4.41: The number of threats found based on satellite stereo images using proposed sparse representation (SR) algorithm.

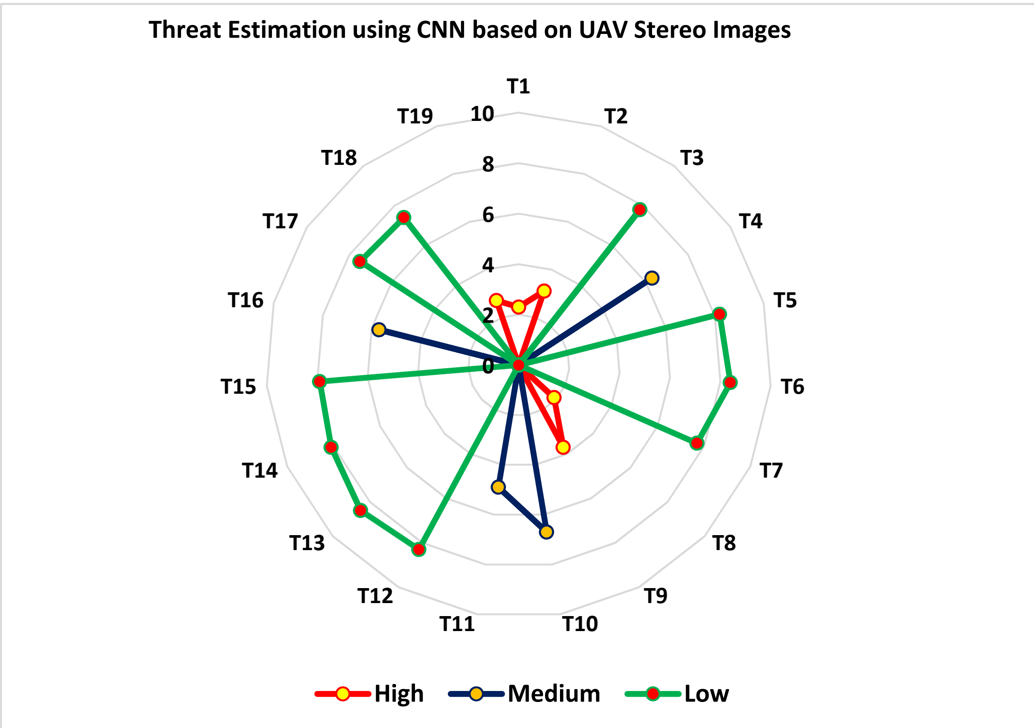


Figure 4.42: The number of threats found based on satellite stereo images using proposed CNN algorithm.

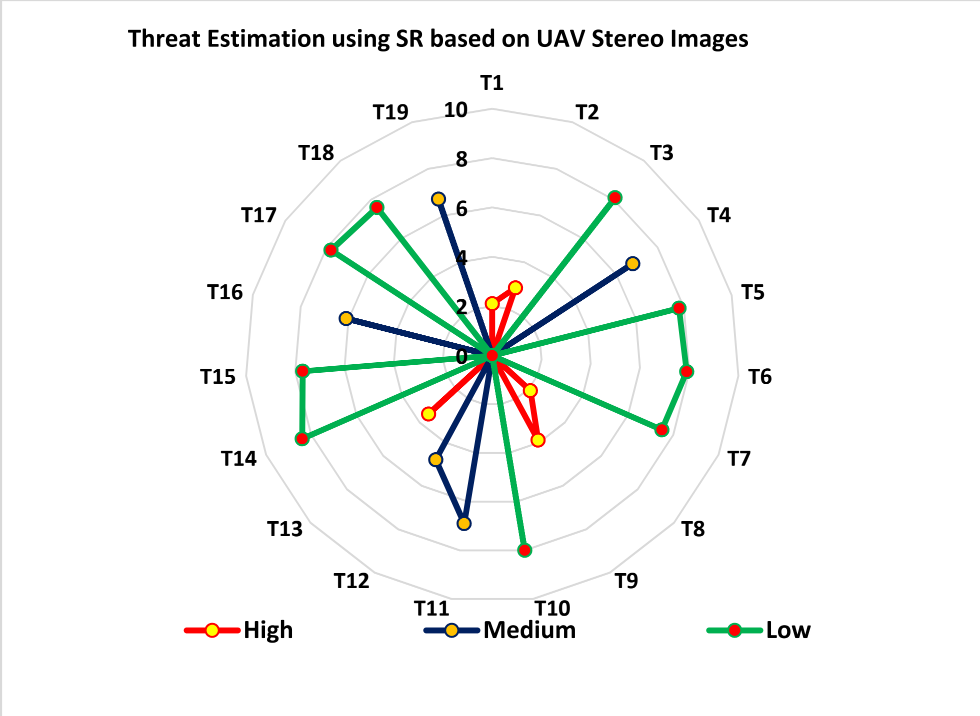


Figure 4.43: The number of threats found based on satellite stereo images using proposed sparse representation (SR) algorithm.

The disparity map is used to estimate the depth or height of vegetation / trees near or under the power transmission lines and poles. The estimated height and distance of objects (vegetation / trees) between centers of power transmission lines have been used for measuring the various number of threats. These threats are denoted as high, medium and low. These threats were estimated based on estimated distance and reference distance between objects and power transmission lines in the right -of -way (ROW). The threat is called danger, if the distance between vegetation / trees and power transmission lines is less than 4 meters and height of that object also greater than 2.4 meters than it will consider a threat.

In this experiment, the fifty-one (51) number of transmission poles have been used in satellite stereo images and twenty (20) number power poles has been used for UAV stereo images of area of interest. The three power poles have been used with power transmission lines in each stereo image pair for both UAV and satellite imagery to compute the distance between objects and power transmission lines. The 24 number of low, 12 number of medium and 11 number of high threats were found in this experiment using CNN approach based on satellite stereo image patches as shown in Figure 4.40.

By using the satellite stereo images, the SR produced 11 number of high, 12 number of medium and 24 number of low threats as shown in Figure 4.41. Similarly, in UAV stereo images, the CNN had found 5 high, 4 medium and 10 low number of threats as shown in Figure 4.42. The SR algorithm produced 5 number of high, 5 number of medium and 9 number of low threats using UAV stereo images as shown in Figure 4.43. It can be concluded that the proposed methods have a capability to estimate the various number of threats by estimating the height and distance between vegetation / trees and power transmission lines and power poles.

4.6 Discussion

Images that are acquired by UAV and satellite are stereo based. For power line monitoring, the height and distance parameters for vegetation and objects proximal to high voltage transmission lines has been required for threat estimation. A reliable disparity map based on stereo imagery is thus required to estimate the height. Hence, the present work's major contribution is the use of SR and CNN algorithms to derive a reliable disparity map from UAV and satellite stereo datasets. The disparity map computed by the proposed algorithms obtained accurate estimates for power poles and tree heights. Ground truth height results in the areas of interest were compared with estimations provided by proposed methods. Precision and recall values were computed and compared for estimations vs. actual reported heights for vegetation / trees near power lines/poles. Results indicated that the proposed method can be reliably used to monitor vegetation and trees proximal to power poles and under transmission lines. The sparse representation has been used for estimation of disparity map of aerial stereo images (satellite and UAV). The sparse representation algorithm used proposed hybrid dictionary (DRT dictionary) which extracts the prominent features from stereo images to construct the disparity map efficiently. The DRT based dictionary extracts the compressed and structured features from the stereo images for matching the certain patches of the images in a certain disparity range and compared these features using Euclidian distance approach for similarity measurement between image patches. The SR algorithm constructs the smooth disparity map based on the accurate estimation of the displaced points from stereo images. The main limitation in existing stereo matching method is that they did not handle the depth discontinuity effectively without some regularization parameters.

To overcome this problem, the SR technique better reconstructs smooth disparity at the edges of the objects. The second most important issue in stereo matching is the high noise in the aerial images. The existing stereo matching algorithms are failed to remove the noise from the aerial images. The proposed SR algorithm has the property to extracts the image patch in sparse domain and minimize the noise inheritably without any prior knowledge. The existing stereo matching algorithms could not provide the accurate disparity map due to depth discontinuity in the disparity map. The DP algorithm provided the streaking effect on the discontinuity region and failed to detect the depth or height at the object boundary. The smooth regularization factor can be imposed to handle the depth discontinuity at the edges of the object. The smooth regularization produced extra computational cost and may not provide the smooth disparity map. The GC algorithm produced over smoothness due to regularization parameter and tends to generate blur disparity at the edges of the objects on the disparity map. It also produced large disparity range due to over smoothing.

To address the depth discontinuity in the disparity map, the CNN algorithm has been proposed to estimates the depth discontinuity with different approach. The CNN algorithm used to extracts the feature vector from last fully connected layer and measured the similarity between these feature vectors in a certain disparity range in correspondence to the right and left stereo patches. The CNN worked on the basis of convolutional and pooling layer. The pooling layer extracts the translational invariant features from the image patches and convolutional layer provides the deep spatial features to efficiently reconstruct the prominent features from stereo patches. These special properties make the CNN more accurate patch extraction technique as compared to existing stereo matching techniques. For proposed work in this thesis, the CNN extracted the prominent features from the stereo images and produced smooth disparity map at the edges of the object and well handled the depth discontinuity in the stereo images. The mini-batched gradient descent algorithm has been used to train the CNN model and it also reduced the noise by using mini-batches samples of stereo image patches. The convolutional layer in the CNN model extracts the spatial features and measures the similarity between patches based on the spatial prominent features. These similar features preserved the displaced values of stereo patches, termed as disparity value. In this way, the CNN extracts the texture less regions efficiently based on the feature similarity index capability.

Comparisons with existing stereo matching algorithms for disparity calculations in terms of processing speed, accuracy, precision and recall revealed that the proposed algorithms detection rate and association of pixels for disparity calculations were more accurate than extant systems. Moreover, by computing height estimations via the proposed system generated disparity map, the proposed approaches identified real threats. Threat levels were further classified as high, medium or low by standardized measurement values. A threat was considered 'high' when the distance between transmission lines /poles to trees and /or vegetation was < 4 meters. A medium threat measured between 4–6 m, and a low threat ranged between 7–8 m in accordance with accepted legal standards for 'right of way' (ROW). The proposed stereo matching algorithms thus provided reliable disparity maps that can aid a power company's assessment of hazardous threats.

4.7 Summary

In this chapter the results of area of interest has been discussed based on proposed and existing stereo matching algorithms. The height estimation is the inverse of disparity map which is calculated using proposed as well as existing algorithms. The accuracy, precision and recall have been evaluated based on actual and estimated height of power poles. The proposed techniques using CNN and sparse representation algorithms perform excellent as compared to existing stereo matching algorithms based on the performance metrics. The different number of parameters such as disparity value, number of power poles and number of iterations has been used to measure accuracy, precision and recall. The different number of threats has been calculated, these threats indicate that how much is the striking risk of vegetation / trees to the power poles and power lines. The results show that the CNN algorithm produced more accuracy as compared to existing algorithms.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions

An attempt to fill the gap in shared knowledge was made through the research performed and workflow created in this dissertation. Hopefully this kind of research will be carried on further in an effort to find better ways to protect the precious electricity our society so heavily relies on every day. To follow the metaphor, it appears just as self-evident that similar preventative and diagnostic efforts that avert a stroke or heart attack in humans need to be applied to preventing power blackouts and their devastating impact on human lives and economic activity. Advancing technology such as aerial stereo imagery (UAV and satellite) moves this effort along, and more research is needed. In this dissertation, a novel method for vegetation monitoring under or near HV overhead lines is described using proposed sparse representation and CNN algorithms. The CNN proposed technique provides the best results among the proposed sparse representation method and existing approaches for height estimation of power poles / lines surrounding objects (vegetation, trees).

The proposed method uses the stereo matching algorithms to compute the disparity map then the height of vegetation / trees near or under the power transmission lines is estimated based on the computed disparity map. The proposed system also calculate the distance of vegetation/trees from transmission power lines and poles outside right of ways using triangulation method by providing the location on the image. The proposed system for vegetation monitoring has more advantages over the traditional vegetation monitoring practices (including field survey, airborne surveillance, aerial stereoscopy and LiDAR scanning etc.), it processes stereo images from satellite and UAV, where image processing and pattern recognition techniques are used to identify dangerous encroachments within a limited period of time with high accuracy. The image acquisition using satellite and UAV sensors has been used in proposed and existing algorithms. The Global positioning System (GPS) used to estimate the locations of towers and standard values of right of way; the captured stereo images were cropped automatically based on conversion of GPS points into coordinates on the image as they are very immense in size. Further, on this stage orthorectification was performed on the cropped stereo images of area of interest.

The disparity map has been estimated based on our proposed stereo matching algorithm and also compared the disparity map computed by existing stereo matching algorithms on cropped stereo aerial stereo images. The height of towers, vegetation / trees is estimated using disparity map as determined by the propose algorithms (sparse representation and CNN) based on triangulation method. The height of objects (vegetation, trees and towers) depends on the accurate disparity map of the system. The analytical geometrical model is developed using triangular relationship to measure the height of overgrown objects (vegetation, tress and towers) by giving their image coordinates and also estimate the distance between vegetation / trees near power transmission lines / poles.

The proposed mathematical models by utilizing the data from the above height estimation algorithm to compute the absolute distance between transmission poles and the overgrown vegetation, height of the trees, height of the towers, and distance between the HV lines and the overgrown vegetation outside right of ways. This adds more metrics to the proposed system as well as more accuracy. The system also estimates different cross arms height of towers using total estimated height of towers and also estimate different threat levels by giving calculated distance of objects from towers and power transmission lines using standard right of ways values. The threats levels are defined as dangerous, medium and low. The dangerous threat level means objects interfere the power transmission lines and towers, medium threat level means it may or may not interference with towers and transmission lines while low threat means no interference of objects. Using estimated metrics based on proposed algorithm, the operators at base station will order the trimmers to cut-off the vegetation wherever it is required. This method provides more accuracy, less cost and less human indolent as compared to other aerial vegetation monitoring methods. Whereas, the traditional vegetation inspection methods require a huge amount of budget for every inspection (in Malaysia, once every two months).

Thus proposed method as compared to the existing techniques will be cost-effective, less time consuming and more accurate. The common limitation in proposed algorithm based on CNN is the more computational time for disparity map estimation as compared to existing area based stereo matching methods. The small stereo patches can be used to minimize the computational time using CNN algorithm for disparity map estimation. The limitation in sparse representation algorithm is more unique points are needed in stereo images for matching proposes. In order to minimize this limitation, the adaptive patch extraction from stereo images using adaptive dictionaries could be used for smooth disparity map estimation.

5.2 Recommendations and Future Works

In order to make the system more reliable, there are given some recommendations that can be applied to the proposed system as the future extension of proposed work.

- To minimize the computational complexity, the different number of stereo matching algorithm can be revealed based on block matching, area matching and template matching.
- The proposed algorithms (SR and CNN) can be combine for estimation of the disparity map which may produce high accuracy.
- The various latest sensors can be used in stereo image acquisition to minimize the clouding effect in the stereo images.

REFERENCES

[1] J. Ahmad, A. S. Malik, L. Xia, and N. Ashikin, “Vegetation encroachment monitoring for transmission lines right-of-ways: A survey,” Electr. Power Syst. Res., vol. 95, pp. 339–352, Feb. 2013.

[2] Y. Jwa and G. Sohn, “A Piecewise Catenary Curve Model Growing for 3D Power Line Reconstruction,” Photogramm. Eng. Remote Sens., vol. 78, no. 12, pp. 1227–1240, Dec. 2012.

[3] S. D. Guikema, R. A. Davidson, and H. Liu, “Statistical Models of the Effects of Tree Trimming on Power System Outages,” IEEE Trans. Power Deliv., vol. 21, no. 3, pp. 1549–1557, Jul. 2006.

[4] M. G. M. Jardini et al., “Information System for the Vegetation Control of Transmission Lines Right-of-way,” in 2007 IEEE Lausanne Power Tech, 2007, pp. 28–33.

[5] Y. Kobayashi, G. G. Karady, G. T. Heydt, and R. G. Olsen, “The Utilization of Satellite Images to Identify Trees Endangering Transmission Lines,” IEEE Trans. Power Deliv., vol. 24, no. 3, pp. 1703–1709, Jul. 2009.

[6] C. Sun et al., “Measuring the distance of vegetation from powerlines using stereo vision,” ISPRS J. Photogramm. Remote Sens., vol. 60, no. 4, pp. 269–283, Jun. 2006.

[7] D. I. Jones and G. K. Earp, “Camera sightline pointing requirements for aerial inspection of overhead power lines,” Electr. Power Syst. Res., vol. 57, no. 2, pp. 73–82, Mar. 2001.

[8] Z. Li, R. Walker, R. Hayward, and L. Mejias, “Advances in vegetation

management for power line corridor monitoring using aerial remote sensing techniques," in 2010 1st International Conference on Applied Robotics for the Power Industry (CARPI 2010), 2010, pp. 1–6.

[9] Z. Li, T. S. Bruggemann, J. J. Ford, L. Mejias, and Y. Liu, "Toward automated power line corridor monitoring using advanced aircraft control and multisource feature fusion," J. F. Robot., vol. 29, no. 1, pp. 4–24, Jan. 2012.

[10] J. C. Suárez, C. Ontiveros, S. Smith, and S. Snape, "Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry," Comput. Geosci., vol. 31, no. 2, pp. 253–262, Mar. 2005.

[11] Y. Liu, Z. Li, R. Hayward, R. Walker, and H. Jin, "Classification of Airborne LIDAR Intensity Data Using Statistical Analysis and Hough Transform with Application to Power Line Corridors," in 2009 Digital Image Computing: Techniques and Applications, 2009, pp. 462–467.

[12] F. Nex and F. Remondino, "UAV for 3D mapping applications: a review," Appl. Geomatics, vol. 6, no. 1, pp. 1–15, Mar. 2014.

[13] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," Int. J. Comput. Vis., 2002.

[14] J. Cai and R. Walker, "Height estimation from monocular image sequences using dynamic programming with explicit occlusions," IET Comput. Vis., vol. 4, no. 3, p. 149, 2010.

[15] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 2, pp. 147–59, Mar. 2004.

[16] D. Louit, R. Pascual, and D. Banjevic, "Optimal interval for major maintenance actions in electricity distribution networks," Int. J. Electr. Power Energy Syst., vol. 31, no. 7–8, pp. 396–401, Sep. 2009.

[17] J. A. Wischkaemper, C. L. Benner, and B. D. Russell, "Electrical

characterization of vegetation contacts with distribution conductors - investigation of progressive fault behavior,” in 2008 IEEE/PES Transmission and Distribution Conference and Exposition, 2008, pp. 1–8.

[18] A. D. Filomena, M. Resener, R. H. Salim, and A. S. Bretas, “Distribution systems fault analysis considering fault resistance estimation,” Int. J. Electr. Power Energy Syst., vol. 33, no. 7, pp. 1326–1335, Sep. 2011.

[19] V. M. V. and Y. Li, “Analysis of 1996 Western American Electric Blackouts,” Bulk Power Syst. Dyn. Control, pp. 685–721.

[20] S. Guggenmoos and T.E. Sullivan, “Outside Right-of-Way Tree Risk Along Electrical Transmission Lines,” Util. Arborist Assoc., vol. 8, pp. 1–6, 1993.

[21] G. Andersson et al., “Causes of the 2003 Major Grid Blackouts in North America and Europe, and Recommended Means to Improve System Dynamic Performance,” IEEE Trans. Power Syst., vol. 20, no. 4, pp. 1922–1928, Nov. 2005.

[22] W. Liscouski, B., and Elliot, “Blackout in the United States and Canada: Causes and Recommendations,” A Rep. to US Dep. Energy, 2004.

[23] A. Berizzi, “The Italian 2003 blackout,” in IEEE Power Engineering Society General Meeting, 2004., vol. 2, pp. 1673–1679.

[24] “Online Available: http://thestar.com.my.”

[25] “Online Available: http://www.sarawakenergy.com.my.”

[26] “Energy Commission. (2009). Interim report on the performance of the electricity supply in Malaysia for the first half year of 2009. Energy Commission: Putrajaya, Malaysia.”

[27] “Online Available: http://www.tnb.com.my.”

[28] J. Chen, J. S. Thorp, and I. Dobson, “Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model,” Int. J.

Electr. Power Energy Syst., vol. 27, no. 4, pp. 318–326, May 2005.

[29] C. Mu, Q. Yan, Y. Feng, J. Cai, and J. Yu, “Overview of powerlines extraction and surveillance using remote sensing technology,” 2009, p. 74981M.

[30] P. A. Sousa, R. J. Santos, A. Matos Andr, and R. Bernardo, “Disruptive solution to mitigate risk and increase efficiency in EDP distribuiao, overhead lines inspections,” in 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013), 2013, pp. 0529–0529.

[31] M. Williams, D. I. Jones, and G. K. Earp, “Obstacle avoidance during aerial inspection of power lines,” Aircr. Eng. Aerosp. Technol., vol. 73, no. 5, pp. 472–479, Oct. 2001.

[32] H. M. M. W. Straatsma, “Airborne laser scanning as a tool for lowland floodplain vegetation monitoring,” Trends Challenges Sci. Manag., vol. 187, pp. 87–103, Jul. 2006.

[33] S. J. Mills et al., “Evaluation of Aerial Remote Sensing Techniques for Vegetation Management in Power-Line Corridors,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 9, pp. 3379–3390, Sep. 2010.

[34] R. Ghaffar, N. Jafri, and S. A. Khan, “Depth Extraction System Using Stereo Pairs,” 2004, pp. 512–519.

[35] A. Klaus, M. Sormann, and K. Karner, “Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure,” in 18th International Conference on Pattern Recognition (ICPR’06), 2006, vol. 3, pp. 15–18.

[36] B. B. Alagoz, “Obtaining Depth Maps From Color Images By Region Based Stereo Matching Algorithms,” Dec. 2008.

[37] Daolei Wang and Kah Bin Lim, “A new segment-based stereo matching using graph cuts,” in 2010 3rd International Conference on Computer Science and Information Technology, 2010, pp. 410–416.

[38] P. Gamba and B. Houshmand, “Digital surface models and building extraction: a comparison of IFSAR and LIDAR data,” IEEE Trans. Geosci. Remote Sens., vol. 38, no. 4, pp. 1959–1968, Jul. 2000.

[39] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, “Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images,” Inf. Fusion, vol. 6, no. 3, pp. 193–212, Sep. 2005.

[40] T. B. Moeller, Matthias S., “A new index for the differentiation of vegetation fractions in urban neighborhoods based on satellite imagery,” ASPRS Annu. Conf., Reno., 2006.

[41] Y. Hirano, Y. Yasuoka, and T. Ichinose, “Urban climate simulation by incorporating satellite-derived vegetation cover distribution into a mesoscale meteorological model,” Theor. Appl. Climatol., vol. 79, no. 3–4, pp. 175–184, Dec. 2004.

[42] K. Deilami and M. Hashim, “Very high resolution optical satellites for DEM generation : a review,” European Journal of Scientific Research. Eurojournals, Oct-2011.

[43] D. Kocaman, S., Zhang, L., Gruen, A., & Poli, “3D city modeling from high-resolution satellite images,” Proc. ISPRS Work. Topogr. Mapp. from Space, Ankara, Turkey., 2006.

[44] Ki In Bang, Soo Jeong, Kyung-Ok Kim, and Woosug Cho, “Automatic DEM generation using IKONOS stereo imagery,” in IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), vol. 7, pp. 4289–4291.

[45] T. Toutin, “Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-a, IKONOS-II, and QuickBird,” IEEE Trans. Geosci. Remote Sens., vol. 42, no. 10, pp. 2121–2129, Oct. 2004.

[46] M. Krauß, T., Reinartz, P., and Lehner, “Comparison of DSM generation

methods of urban areas from IKONOS images," Rev. Fr. Photogramm. Teledection, vol. 4, pp. 101–106, 2007.

[47] M. Ehlers, "Remote sensing for GIS applications: new sensors and analysis methods," in Proceedings of SPIE, 2004, vol. 5239, pp. 1–13.

[48] K. Jacobsen, "Airborne or Spaceborne Images for Topographic Mapping?," ipi.uni-hannover.de.

[49] L. Lebègue et al., "PLEIADES-HR IMAGE QUALITY COMMISSIONING," ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. XXXIX-B1, pp. 561–566, Jul. 2012.

[50] M. B. Boissin, A. Gleyzes, and C. Tinel, "The pléiades system and data distribution," in 2012 IEEE International Geoscience and Remote Sensing Symposium, 2012, pp. 7098–7101.

[51] D. Poli and T. Toutin, "Review of developments in geometric modelling for high resolution satellite pushbroom sensors," Photogramm. Rec., vol. 27, no. 137, pp. 58–73, Mar. 2012.

[52] P. Damilano, "Pleiades High Resolution Satellite : A Solution for Military and Civilian Needs in Metric-Class Optical Observation," AIAA/USU Conference on Small Satellites. 2001.

[53] "Delvit, J. M., Greslou, D., Amberg, V., Dechoz, C., de Lussy, F., Lebegue, L., ... & Bernard, L. (2012). Attitude assessment using Pleiades-HR capabilities. Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Inform."

[54] C. Latry, S. Fourest, and C. Thiebaut, "Restoration technique for Pleiades-HR panchromatic images," Proc. Int. Arch., 2012.

[55] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "An Automatic Building Reconstruction Method : A Structural Approach using High Resolution Satellite Images," in 2006 International Conference on Image

Processing, 2006, pp. 1205–1208.

[56] M. Straatsma and H. Middelkoop, “Airborne laser scanning as a tool for lowland floodplain vegetation monitoring,” Living rivers Trends challenges …, 2006.

[57] C. Siart, O. Bubenzer, and B. Eitel, “Combining digital elevation data (SRTM/ASTER), high resolution satellite imagery (Quickbird) and GIS for geomorphological mapping: A multi-component case study on Mediterranean karst in Central Crete,” Geomorphology, vol. 112, no. 1–2, pp. 106–121, Nov. 2009.

[58] J. A. J. Berni, P. J. Zarco-Tejada, G. Sepulcre-Cantó, E. Fereres, and F. Villalobos, “Mapping canopy conductance and CWSI in olive orchards using high resolution thermal remote sensing imagery,” Remote Sens. Environ., vol. 113, no. 11, pp. 2380–2388, Nov. 2009.

[59] A. S. Laliberte, M. A. Goforth, C. M. Steele, and A. Rango, “Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments,” Remote Sens., vol. 3, no. 12, pp. 2529–2551, Nov. 2011.

[60] Q. Feng, J. Liu, and J. Gong, “UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis,” Remote Sens., vol. 7, no. 1, pp. 1074–1094, Jan. 2015.

[61] G. Pajares, “Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs),” Photogramm. Eng. Remote Sens., vol. 81, no. 4, pp. 281–330, Apr. 2015.

[62] F. S. Leira, T. A. Johansen, and T. I. Fossen, “Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera,” in 2015 IEEE Aerospace Conference, 2015, pp. 1–10.

[63] L. Matikainen et al., “Remote sensing methods for power line corridor surveys,” ISPRS J. Photogramm. Remote Sens., vol. 119, pp. 10–31, Sep. 2016.

[64] S. Ashidate, S. Murashima, and N. Fujii, "Development of a helicopter-mounted eye-safe laser radar system for distance measurement between power transmission lines and nearby trees," IEEE Trans. Power Deliv., vol. 17, no. 2, pp. 644–648, Apr. 2002.

[65] M. Frank, Z. Pan, B. Raber, and C. Lenart, "Vegetation management of utility corridors using high-resolution hyperspectral imaging and LiDAR," in 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2010, pp. 1–4.

[66] E. W. Bork and J. G. Su, "Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis," Remote Sens. Environ., vol. 111, no. 1, pp. 11–24, Nov. 2007.

[67] L. Di Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," Image Vis. Comput., vol. 22, no. 12, pp. 983–1005, Oct. 2004.

[68] H. Hirschmuller and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[69] E. F. Sawires, A. M. Hamdy, F. Z. Amer, and E. M. Bakr, "Disparity map using suboptimal cost with dynamic programming," in The 10th IEEE International Symposium on Signal Processing and Information Technology, 2010, pp. 209–214.

[70] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nister, "High-Quality Real-Time Stereo Using Adaptive Cost Aggregation and Dynamic Programming," in Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), 2006, pp. 798–805.

[71] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in Proceedings Ninth IEEE International Conference on Computer Vision, 2003, pp. 900–906 vol.2.

[72] R. Szeliski et al., “A comparative study of energy minimization methods for Markov random fields with smoothness-based priors.,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 6, pp. 1068–80, Jun. 2008.

[73] C. L. Zitnick and T. Kanade, “A cooperative algorithm for stereo matching and occlusion detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 7, pp. 675–684, Jul. 2000.

[74] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” Int. J. Comput. Vis., vol. 47, no. 1–3, pp. 7–42, 2002.

[75] T. Kanade and M. Okutomi, “A stereo matching algorithm with an adaptive window: theory and experiment,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 9, pp. 920–932, 1994.

[76] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, “A Maximum Likelihood Stereo Algorithm,” Comput. Vis. Image Underst., vol. 63, no. 3, pp. 542–567, May 1996.

[77] P. Fua, “A parallel stereo algorithm that produces dense depth maps and preserves image features,” Mach. Vis. Appl., vol. 6, no. 1, pp. 35–49, Dec. 1993.

[78] D. Marr and T. Poggio, “A Computational Theory of Human Stereo Vision,” Proc. R. Soc. B Biol. Sci., vol. 204, no. 1156, pp. 301–328, May 1979.

[79] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 4, pp. 353–363, Apr. 1993.

[80] G. Medioni and R. Nevatia, “Segment-based stereo matching,” Comput. Vision, Graph. Image Process., vol. 31, no. 1, pp. 2–18, Jul. 1985.

[81] M. Gong and Y.-H. Yang, “Real-Time Stereo Matching Using Orthogonal Reliability-Based Dynamic Programming,” IEEE Trans. Image Process., vol. 16, no. 3, pp. 879–884, Mar. 2007.

[82] A. F. Bobick and S. S. Intille, “Large Occlusion Stereo,” Int. J. Comput. Vis.,

vol. 33, no. 3, pp. 181–200.

[83] S. Roy, “Stereo Without Epipolar Lines: A Maximum-Flow Formulation,” Int. J. Comput. Vis., vol. 34, no. 2–3, pp. 147–161.

[84] M. Gong and Y.-H. Yang, “Fast stereo matching using reliability-based dynamic programming and consistency constraints,” in Proceedings Ninth IEEE International Conference on Computer Vision, 2003, pp. 610–617 vol.1.

[85] M. Tappen and W. Freeman, “Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters,” Comput. Vision, 2003. …, 2003.

[86] V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts,” in Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 2001, vol. 2, pp. 508–515.

[87] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp. 1222–1239, 2001.

[88] J. Sun, N. Zheng, and H.-Y. Shum, “Stereo matching using belief propagation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 7, pp. 787–800, Jul. 2003.

[89] D. Marr, G. Palm, and T. Poggio, “Analysis of a cooperative stereo algorithm,” Biol. Cybern., vol. 28, no. 4, pp. 223–239, 1978.

[90] B. Zitová and J. Flusser, “Image registration methods: a survey,” Image Vis. Comput., vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[91] H. Chui and A. Rangarajan, “A new point matching algorithm for non-rigid registration,” Comput. Vis. Image Underst., vol. 89, no. 2–3, pp. 114–141, Feb. 2003.

[92] W. E. L. Grimson, “Computational Experiments with a Feature Based Stereo Algorithm,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-7, no. 1, pp. 17–34, Jan. 1985.

[93] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient Belief Propagation for Early Vision,” Int. J. Comput. Vis., vol. 70, no. 1, pp. 41–54, May 2006.

[94] and Y. L. Zbontar, Jure, “Stereo matching by training a convolutional neural network to compare image patches,” J. Mach. Learn. Res., vol. 17, no. 32, p. 2, 2016.

[95] S. Li, H. Yin, and L. Fang, “Remote Sensing Image Fusion via Sparse Representations Over Learned Dictionaries,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[96] M. Elad and M. Aharon, “Image Denoising Via Learned Dictionaries and Sparse representation,” in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR’06), 2006, vol. 1, pp. 895–900.

[97] A. Buades, B. Coll, and J. M. Morel, “A Review of Image Denoising Algorithms, with a New One,” Multiscale Model. Simul., vol. 4, no. 2, pp. 490–530, Jan. 2005.

[98] D. L. DONOHO and J. M. JOHNSTONE, “Ideal spatial adaptation by wavelet shrinkage,” Biometrika, vol. 81, no. 3, pp. 425–455, 1994.

[99] M. Elad and M. Aharon, “Image denoising via sparse and redundant representation over learned dictionaries,” IEEE Transations Image Process., vol. 15, no. 12, pp. 3736–3745, 2006.

[100] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis Pursuit,” SIAM J. Sci. Comput., vol. 20, no. 1, pp. 33–61, Jan. 1998.

[101] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, pp. 40–44.

[102] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis

Pursuit," SIAM Rev., vol. 43, no. 1, pp. 129–159, Jan. 2001.

[103] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," IEEE Trans. Image Process., vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[104] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for Sparse Representation Modeling," Proc. IEEE, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[105] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[106] W. Jin, L. Wang, X. Zeng, Z. Liu, and R. Fu, "Classification of clouds in satellite imagery using over-complete dictionary via sparse representation," Pattern Recognit. Lett., vol. 49, pp. 193–200, 2014.

[107] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust Face Recognition via Sparse Representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[108] I. Tosic and P. Frossard, "Dictionary Learning, What is the right representation for my signal?," IEEE Signal Process. Mag., vol. 28, num 2, no. MARCH, pp. 27–38, 2011.

[109] S. G. Mallat, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, 1993.

[110] S. G. Mallat, "Adaptive time-frequency decompositions," Opt. Eng., vol. 33, no. 7, p. 2183, Jul. 1994.

[111] A. Grossmann and J. Morlet, "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape," SIAM J. Math. Anal., vol. 15, no. 4, pp. 723–736, Jul. 1984.

[112] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," SIAM Rev.,

vol. 51, no. 1, pp. 34–81, Feb. 2009.

[113] G. E. Hinton, “Reducing the Dimensionality of Data with Neural Networks,” Science (80-.)., vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[114] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[115] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[116] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[117] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” 2014, pp. 818–833.

[118] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.

[119] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[120] C. H. Chen, X., Xiang, S., Liu, C. L., & Pan, “Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks,” IEEE Geosci. Remote Sens. Lett., vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[121] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, “On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery,” Int. J. Remote Sens., vol. 36, no. 13, pp. 3368–3379, Jul. 2015.

[122] J. Yue, W. Zhao, S. Mao, and H. Liu, “Spectral–spatial classification of hyperspectral images using deep convolutional neural networks,” Remote Sens. Lett., vol. 6, no. 6, pp. 468–477, Jun. 2015.

[123] A. Dosovitskiy et al., “FlowNet: Learning Optical Flow with Convolutional Networks,” in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2758–2766.

[124] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a Deep Convolutional Network for Image Super-Resolution,” 2014, pp. 184–199.

[125] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2392–2399.

[126] F. Liu, C. Shen, G. Lin, and I. Reid, “Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 10, pp. 2024–2039, Oct. 2016.

[127] C. Osendorfer, J. Bayer, S. Urban, and P. van der Smagt, “Convolutional Neural Networks Learn Compact Local Image Descriptors,” 2013, pp. 624–630.

[128] J. Chen and C. Yuan, “Convolutional neural network using multi-scale information for stereo matching cost computation,” in 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 3424–3428.

[129] and J. T. Jordan, Tyler S., Skanda Shridhar, “Usings CNNs to Estimate Depth from Stereo Imagery,” in IEEE proceeding, 2015.

[130] et al. Chen, Zhuoyuan, “A deep visual correspondence embedding model for stereo matching costs,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 972–980.

[131] and N. K. Zagoruyko, Sergey, “Learning to compare image patches via convolutional neural networks,” in Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4353–4361.

[132] and R. U. Luo, Wenjie, Alexander G. Schwing, “Efficient deep learning for stereo matching,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5695–5703.

[133] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka, “A stereo machine for video-rate dense depth mapping and its new applications,” in Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 196–202.

[134] M. N. Do and M. Vetterli, “The finite ridgelet transform for image representation,” IEEE Trans. Image Process., vol. 12, no. 1, pp. 16–28, 2003.

[135] J. Fadili and J.-L. Starck, “Curvelets and Ridgelets.,” Encycl. Complex. Syst. Sci., vol. 3, pp. 1718–1738, 2009.

[136] P. Carre and E. Andres, “Discrete analytical Ridgelet transform,” Signal Processing, vol. 84, no. 11 SPEC. ISS., pp. 2165–2173, 2004.

[137] W. Zhao and S. Du, “Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[138] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning Hierarchical Features for Scene Labeling,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[139] J. Banks, M. Bennamoun, and P. Corke, “Non-parametric techniques for fast and robust stereo matching,” in TENCON ’97 Brisbane - Australia. Proceedings of IEEE TENCON ’97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications (Cat. No.97CH36162), 1997, vol. 1, pp. 365–368.

[140] T. Tao, J. C. Koo, and H. R. Choi, “A fast block matching algorthim for stereo correspondence,” in 2008 IEEE Conference on Cybernetics and Intelligent Systems, 2008, pp. 38–41.

[141] D. Oram, “Rectification for any epipolar geometry,” BMVC, vol. 1, 2001.

[142] A. Fusiello and L. Irsara, “Quasi-Euclidean uncalibrated epipolar rectification,” in 2008 19th International Conference on Pattern Recognition, 2008, pp. 1–4.

[143] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," Mach. Vis. Appl., vol. 12, no. 1, pp. 16–22, Jul. 2000.

[144] C. Loop and Zhengyou Zhang, "Computing rectifying homographies for stereo vision," in Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), pp. 125–131.

[145] "Hartley, Richard, and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003."

[146] "Szeliski, Richard. Computer vision: algorithms and applications. Springer Science & Business Media, 2010."

PUBLICATION RECORD

Journal Papers:

1. **Qayyum, A.**, Malik, A.S., Saad, N.M., Iqbal, M., Faris Abdullah, M., Rasheed, W., Rashid Abdullah, T.A. and Bin Jafaar, M.Y., 2017. Scene classification for aerial images based on CNN using sparse coding technique. International Journal of Remote Sensing, pp.1-24. **(IF. 2.547) (Q2)**
2. **Qayyum, A.,** Malik, A.S., M. Saad, N., bin Abdullah, M.F., Iqbal, M., Rasheed, W., Bin Ab Abdullah, A.R. and Hj Jaafar, M.Y., 2017. Image Classification based on Sparse Coded Features using Sparse Coding Technique for Aerial Imagery: A hybrid Dictionary Approach. Neural Computing and Application (Accepted). **(IF. 1.763) (Q2)**
3. **Qayyum, A.,** Malik, A.S., M. Saad, N., bin Abdullah, M.F., Iqbal, M., Rasheed, W., Bin Ab Abdullah, A.R. and Hj Jaafar, M.Y., 2017. Measuring height of high-voltage transmission poles using unmanned aerial vehicle (UAV) imagery. The Imaging Science Journal, pp.1-14. **(IF. 0.497) (Q4)**
4. **Qayyum, A.**, Malik, A.S., Saad, M.N.M., Abdullah, M.F. and Iqbal, M., 2016. EVALUATION OF OVERCOMPLETE DICTIONARIES FOR IMAGE INPAINTING. Science International, 28(4) (ISI Index).
5. **Qayyum, A.,** Malik, A.S., Saad, M. and Iqbal, M., 2015. DESIGNING OF ADAPTIVE DICTIONARIES BASED ON MEXICAN HAT WAVELET FOR SPARSE REPRESENTATION. Science International, 27(6) (ISI Index).

Conference Papers:

1. **Abdul Qayyum,** Aamir Saeed Malik, Mohamad Naufal, Mohamad Saad, Mahboob Iqbal, "Dynamic programming based comparison including QuickBird and IKONOS satellite stereo images for monitoring vegetation near

power poles." In Control System, Computing and Engineering (ICCSCE), 2014 IEEE International Conference on, pp. 576-581. IEEE, 2014.

2. **Abdul Qayyum**, Aamir Saeed Malik, Mohamad Saad, Mohamad Naufal, Mahboob Iqbal,"Monitoring of vegetation near power lines based on dynamic programming using satellite stereo images." In Smart Instrumentation, Measurement and Applications (ICSIMA), 2014 IEEE International Conference on, pp. 1-6. IEEE, 2014.
3. **Qayyum, Abdul**, Aamir Saeed Malik, Mohamad Saad, Mohamad Naufal, Mahboob Iqbal, M. F. Abdullah, "Power Lines Vegetation encroachment monitoring based on Satellite Stereo images using stereo matching." In Smart Instrumentation, Measurement and Applications (ICSIMA), 2014 IEEE International Conference on, pp. 1-5. IEEE, 2014.
4. **Abdul Qayyum**, et al. "Identification of power poles based on satellite stereo images using graph-cut algorithm." Signal Processing & Its Applications (CSPA), 2015 IEEE 11th International Colloquium on. IEEE, 2015.
5. **Abdul Qayyum,** A. S. Malik, M. N. M. Saad, M. Iqbal, F. Abdullah, "VEGETATION HEIGHT ESTIMATION NEAR POWER TRANSMISSION POLES VIA SATELLITE STEREO IMAGES USING 3D DEPTH ESTIMATION ALGORITHMS." International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (2015) Germany.
6. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, "Segementation of satellite imagery using Pulse Coupled Neural Netwrok," The 2015 International Conference on Space Science and Communication (IconSpace2015).
7. **Qayyum, A**., Malik, A.S. and muhammad Saad, M.N.B., 2015, August. Comparison of digital elevation models based on high resolution satellite stereo imagery. In Space Science and Communication (IconSpace), 2015 International Conference on (pp. 203-208). IEEE.
8. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, "Digital Elevation Model Based on Rational Polynomial Function," 2015 IEEE International Conference on Signal and Image Processing Applications,19-21 October, Malaysia

9. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, “Disparity Map Estimation Based on Optimization Algorithms using Satellite Stereo Imagery” 2015 IEEE International Conference on Signal and Image Processing Applications,19-21 October, Malaysia
10. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, “Design of Dictionary Based on Discrete Tchebichef Transform” 9th international conference on signal processing and communication systems 14-16 December 2015, 11-14 December, Cairns (Barrier Reef), Australia.
11. **Abdul Qayyum**, A.S Malik, N.M.Saad and A.Rashid, “Design of Digital Elevation Model Based on Orthorectified Satellite Stereo Images” 9th international conference on signal processing and communication systems 14-16 December 2015, 11-14 December, Cairns (Barrier Reef), Australia.
12. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, “Designing of Disparity Map Based on Hierarchical Dynamic Programming Using Satellite Stereo Imagery” 2015 IEEE Student Symposium in Biomedical Engineering and Sciences (ISSBES 2015),3-4 November 2015, Shah Alam, Selangor , Malaysia
13. **Abdul Qayyum,** A.S Malik, N.M.Saad and A.Rashid, “Evalution of Optimization Algorithms for Sparse and Redundant Dictionaries” 2015 IEEE Student Symposium in Biomedical Engineering and Sciences (ISSBES 2015),3-4 November 2015, Shah Alam, Selangor , Malaysia.
14. **Abdul Qayyum**, A.S Malik, N.M.Saad and A.Rashid, “Designing of Overcomplete Dictionaries Based on DCT and DWT” 2015 IEEE Student Symposium in Biomedical Engineering and Sciences (ISSBES 2015),3-4 November 2015, Shah Alam, Selangor , Malaysia.

Patent File:

Submitted patent on proposed stereo matching algorithm.

Copyright:

Submitted copy right on automatic vegetation monitoring system prototype.

Gold Medal:

Won Gold-medal in ITEX (International Innovation and Exhibition), Malaysia 2016.

Prototype:

1. Automatic vegetation monitoring system prototype design for measuring height of vegetation near power transmission lines.

APPENDIX A

Stereo Matching Algorithms

- Local Matching Algorithm

  The local methods with block matching using cost function and block matching with energy minimization technique with their mathematical formulation has been explained is given in this section.

- Block Matching With Cost Function

The conventional block based method of calculating disparity map employs a window of pixels in reference image and look for matching window in the other image of the pair. The similarity cost function between left and right images of certain block can be found. This process can be done using basic cost functions such as sum of absolute difference (SAD), sum of squared difference (SSD) and normalized cross-correlation (NCC). In this technique, a block of pixels is selected from the left image and a most similar block in the right image is searched [139]. In gray scale images, absolute differences between intensities of corresponding pixels in the selected template of reference (right) and the block in the left image are summed up as a single value. The similar patches will have the lowest sum of absolute difference [140]. The different cost functions with their mathematical formulas are given below [67] in equations from (2.1) to (2.3)

$$SAD = \sum_{x,y} |I_1(x,y) - I_2(x,y+d)| \tag{2.1}$$

$$SSD = \sum_{x,y} \left(I_1(x,y) - I_2(x,y+d)\right)^2 \tag{2.2}$$

$$NCC = \frac{\sum_{x,y}\left(I_1(x,y)) \times I_2(x,y+d)\right)}{\sqrt[2]{\sum_{x,y} {I_1}^2(x,y) \times \sum_{x,y} {I_1}^2(x,y+d)}} \tag{2.3}$$

The disparity estimate comes from block matching has integer number, so this disparity does not produce good depth because of different disparity regions have no smooth transition. The sub-pixel estimation calculates and store the minimum cost of current and neighboring cost values of the disparity. Pseudo code for block matching and identification of vegetation near power poles using satellite stereo imagery as shown in Figure 5.1. The Figure 5.2 explained the different cost functions used in the stereo matching estimation techniques. The pseudocode for computing disparity map based on cost function of local stereo algorithm as shown in Table 5.1.

Table 5.1: The computation of cost functions using local cost metrics

```
Define n=no of rows scan
        m=no of column scans
            range=30
For i=1: m
    Set min and max row wise

    For j=1: n
    Set min and max disparity range column wise
            For d=dmin: dmax
            Compute disparity cost function
                    dcost (n,d+range)= sum(abs(Il(i,j)-IR (i,j+d)))
            End
    End
End
```

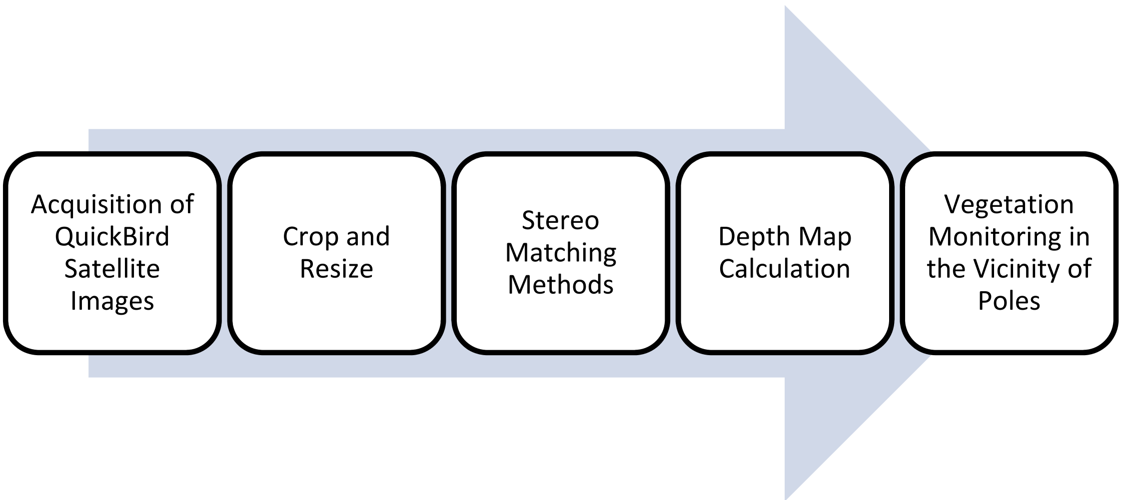


Figure 5.1: Flow diagram for identification of vegetation encroachment using stereo matching by left and right satellite images.

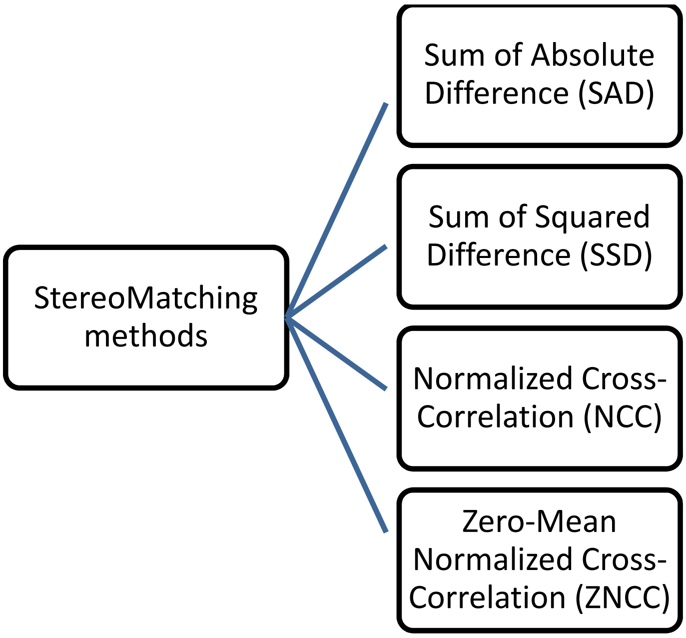


Figure 5.2: Commonly used stereo image matching estimation techniques.

- Block Matching with Energy Function

The block matching with energy minimization algorithm can enhance the accuracy of the disparity map [36]. First energy matrix for every disparity is constructed and error energy can be calculated by equation (2.4).

$$E(u,v,d) = \frac{1}{3mn}\sum_{x=u}^{u+n}\sum_{y=v}^{v+m}\sum_{t=1}^{3}\left(I_{mL}(x,y+d,t) - I_{mR}(x,y,t)\right)^2 \tag{2.4}$$

where t represents RGB components of images and takes value of $\{1,2,3\}$ corresponding to red, blue and green and I_{mL} and I_{mR} are left and right images and d is the disparity and $E(u,v,d)$ is the error energy. For mxn window size, the average filter is defined by equation (2.5).

$$\bar{E}(u,v,d) = \frac{1}{3mn}\sum_{x=u}^{u+n}\sum_{y=v}^{v+m} E(x,y,d) \tag{2.5}$$

In this relation d is the disparity and the filter can be defined for the certain disparity range from left to right image. The average filter applied to smooth the above energy function over window size $(m \times n)$. The average filters remove the noise at the edges and smooth the result of incorrect matching values. For every pixel (u, v) , the minimum $\bar{E}(u, v, d)$ can be choose and assign its disparity index to the $d(u, v)$ which is also disparity.

$$d(u, v) = Min\{\bar{E}(u, v, d)\} \tag{2.6}$$

From minimum energy function, the reliable disparity values will be achieved. The reliability $(\mathrm{R_d})$ of the disparity map is defined by taking the mean value of error energy of disparity maps c . It can be represented by equation (2.7).

$$R_d = \frac{1}{Mean(E_d(u, v) - \{p_\in\})} = \frac{1}{W_d}\left(\sum_{(u,v)-\{E_d(u,v)=p_\in\}}^{n,m} E_d(u, v)\right)^{-1} \tag{2.7}$$

Error energy of disparity map E_d can be found by the equation (2.8).

$$E_d(u, v) = \frac{1}{3mn}\sum_{x=u}^{u+n}\sum_{y=v}^{v+m}\sum_{t=1}^{3}\left(I_{mL}(x, y + d(u, v), t) - I_{mR}(x, y, t)\right)^2 \tag{2.8}$$

Due to object occulation in some images, disparity maps can have some inaccurate disparity estimations for a few points around the object boundaries. These unreliable disparities can be observed by seeing the high error energy present in E_d. The reliability in disparity map $d(u, v)$ can be increased by introducing the threshold mechanism as given in equation (2.9). This will filter some unreliability in disparity estimation.

$$\bar{d}(u, v) = \begin{Bmatrix} d(u, v) & E_d(u, v) \le p_\in \\ 0 & E_d(u, v) > p_\in \end{Bmatrix} \tag{2.9}$$

$$\overline{E_d}(u, v) = \begin{Bmatrix} E_d(u, v) & E_d(u, v) \le p_\in \\ n_e & E_d(u, v) > p_\in \end{Bmatrix} \tag{2.10}$$

$\bar{d}(u,v)$ will be the more reliable version of $d(u,v)$ by filtering some unreliable disparity estimations. Setting disparity to n_e in equation (2.10) refers "no-estimated" state and $E_d(u,v)$ values that have n_e state is excluded in calculation of R_d. W_d parameter in the equation (2.7) represents the number of points in error energy which are not n_e. $\overline{E_d}(u,v)$ is error energy for $\bar{d}(u,v)$ and $p_{\in}$ is error energy threshold for deciding disparity estimation to be unreliable one. The equation (2.11) can be used for determine $p_{\in}$.

$$p_{\in} = \alpha Mean(E_d) \tag{2.11}$$

In the equation (2.11), α is tolerance coefficient to adjust reliability of the filtering process. Decreasing the α leads the more reliable. Unfortunately, decreasing the α will erode disparity map because of eliminating more disparity points in the map. By applying this algorithm the depth of vegetation can estimate more accurately as compared to the simple block matching method.

- Global Stereo matching Algorithms

The two global algorithms with their mathematical formulation has been discussed in detail.

- Dynamic Programming algorithm

The simple dynamic programming cannot detect occlusion, the proposed algorithm has introduced by Cai et al. for occlusion detection using dynamic programming is shown in Table 5.2 [14]. The proposed algorithm employs fixed occlusion parameter the proposed algorithm not only produces high accuracy, but also it detects occlusion in the disparity map based on dynamic programming. The following steps are incorporated the to calculate the minimum cost value for accurate disparity map and introduce some fixed occlusion cost from the original source algorithm steps. The four steps to calculate the minimum cost path and minimum disparity value using dynamic programming are as follows. Where n is the pixel index of the second scan line, $\beta(n,d)$ is the accumulated matching cost at nth pixel. The disparity range is from 0 to 30, $C(n,d)$ is the matching cost, C_1 and C_2 are matching costs of the left and right occlusion respectively. This disparity produces good smoothness and detects objects at depth discontinuous. Dynamic programming produces fine results as compared to block matching on satellite stereo images.

Table 5.2: Set the initial value to each path cost and accumulative cost.

Step 1) Initialization $\beta(u, v) = C(n, d)$ $p(n, d) = 0$ Calculate the minimum cost and occlusion detection in inter-scanline using dynamic programming $n = 0, 0 \leq d \leq 30$ **Step 2) Recursion from n=1 to N-1** $\beta(n, d) = Min\begin{Bmatrix} \beta(n-1, t) + 2C(n, d) + C_1 & if\ t < d \\ \beta(n-n, t) + 2C(n, d) & if\ t = d \\ \beta(n-t-d-1, t) + 2C(n, d) + C_2 & if\ t > d \end{Bmatrix}$ **Step 3) Number of paths to determine** $\beta(n, d) = t$ Save minimum Cost path and track the index of the minimum path in 2D matrix $d_{min} = min\{\beta(N-1, d)\}$ **Step 4) Path backtracking from time n=N-2 to 0** $d_{min} = C(n+1, d_{n+1})$ $n = n - (d_{n+1} - d_{min}) - 1$ $n = n - (d_{n+1} - d_{min}) - 1\ d_n = d_{min}$

The block diagram based on dynamic programming algorithm as shown in Figure 5.3.

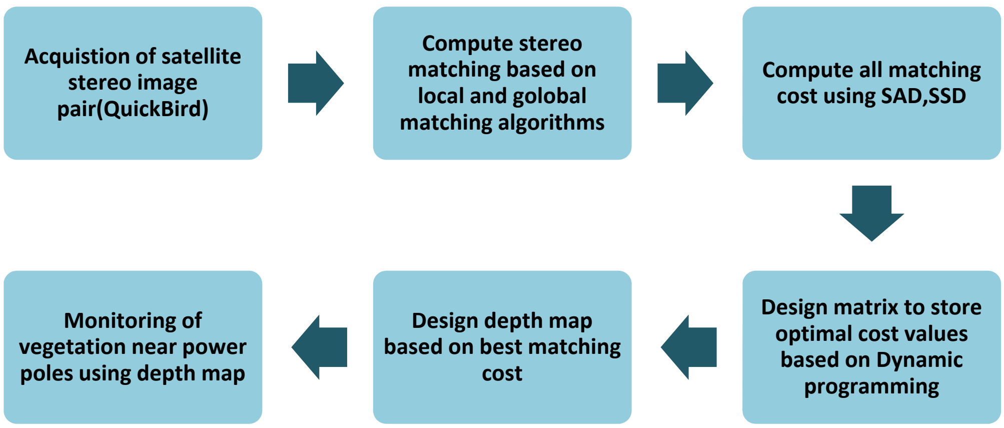


Figure 5.3: Flow chart to compute the depth map using dynamic programming Algorithms.

➢ Graph-Cut Algorithms

The stereo correspondence algorithms based on graph cuts that were presented here practiced the base from which innovative algorithms have evolved. The expansion-move algorithm has the following characteristics [71]. The large number of pixels can change their labels simultaneously. Finding an optimal move is computationally interactive. It takes almost less than one minute for complete execution as compared to other energy minimization algorithm like simulated annealing and iterated-conditional model they take 19 hours to complete execution in early days. Finds local minimum of energy with respect to small "one-pixel" moves. Initialization is an important practice. Theoretically, solution reaches the global minimum. Kolmogorov & Zabih [86] introduced the energy function which comprises three terms: a data term, an occlusion term magnificent a penalty for making a pixel occluded, and a smoothness term penalizing neighboring pixels pairs for having different labels. Based on energy function f of Kolmogorov and Zabih, different energy function can be defined as

$$\bar{E}(u,v)E(f) = E_{data}(f) + E_{occ}(f) + E_{smooth}(f) + E_{unique}(f) + E_{order}(f) \tag{2.10}$$

These energy terms can be difined one by one as the following

$E_{data}(f)$ define the matching cost of corresponding pixel and this matching cost can be calculated using four matching cost function given as:

- Sum of absolute difference(SAD)
- Sum of squared difference(SSD)
- Normalized cross-correlation(NCC)
- Zero-mean normalized cross-correlation (ZNCC)

The kolmogrov and zabih discussed the squared difference of intensity values. The Sum of absolute difference is used, which is easy and cost effective. The formula of the data cost function is given below.

$$E_{data}(f) = \sum_{\langle p,q \in B(f)\rangle} \left| I_{leftintensity}(p) - I_{righttintensity}(p) \right|^{a} \tag{2.11}$$

Where a is may be 1 for SAD and 2 for SSD. In the stereo cross ponding of the stereo pair $E_{occ}(f)$ adds a constant value to the total energy function for each occluded pixel

$$E_{occ}(f) = \sum_{p \in P} K_p F(|U_p(f) = 0|) \tag{2.12}$$

Where F evaluates 1 if its argument is true otherwise zero and the number of occluded pixels is an a□ne function of the number of inactive assignments, any inactive assignment is penalized by penalty K_p. The optimal value of K_p is fifteen (15) used in this experiment.

$E_{smooth}(f)$ if the neighbouring pixels have different disparity this smooth energy function imposes the penalty and can be defined as

$$E_{smooth}(f) = \sum_{\{b_1, b_2 \in N_1\}} U_{b_1,b_2} . F(f(b_1) \neq f(b_2)) \tag{2.13}$$

Where U_{b_1,b_2} is defined as

$$U_{b_1,b_2} = \begin{cases} a_1 = 3a & if \max(|I_1(p_1) - I_1(p_2)|, |I_2(q_1) - I_2(q_2)|) < 8 \\ a_2 = a & otherwise \end{cases} \tag{2.14}$$

If p_1 and p_2 have the same disparity the smoothness term for (b_1, b_2) is zero. Otherwise the penalty for not having the same disparity is small (a) if there is a significant contrast between the adjacent pixels, otherwise it is bigger (3 a). The smoothness term will be zero if the assignment b_1 and b_2 have the same disparity in the N_1 neighbourhood system for 4-neighbours in the input images otherwise it imposes penalty for different disparity of the neighbouring pixels. $E_{unique}(f)$ confines the possible solutions of the optimization problem to unique solutions. If pixel is containing more than one value in the corresponding image in stereo pair then it assign penalty for infinite value otherwise null value assign. This can be defined as

$$E_{unique}(f) = \sum_{P \in p} F(|N_p(f) > 1|).\infty \tag{2.14}$$

The ordering term in the above total energy function is introduced for calculating stereo matching. $E_{order}(f)$ can be written as

$$E_{order}(f) = \sum_{\{b_1, b_2\} \in N_2} F(f(b_1) = f(b_2) = 1).\infty \tag{2.15}$$

where N_2 is a neighbourhood system and can be explain as in such a way that $b_1 = \langle p_x, q_x \rangle$ and $b_2 = \langle p_x^-, q_x^- \rangle$ are neighbours pixels. They must fill full the order as if $(p_x > p_x^-)$ and $(q_x > q_x^-)$ is true.

The final energy function can be written as

$$E(f) = E_{data}(f) + E_{occ}(f) + E_{smooth}(f) + E_{unique}(f) + E_{order}(f) \tag{2.16}$$

The energy function can be minimized using Graph-Cut algorithm is the overall solution of the correspondence between stereo images.

APPENDIX B

Disparity Map Estimation using State-of-the-Art Work

- Disparity map based on existing state-of-the-art stereo images.

The disparity map estimate based on our proposed algorithms and existing algorithms using existing stereo dataset. The three existing datasets are used for this experiment to measure disparity map based on proposed and existing algorithms. The Tuskuba stereo, teddy stereo and art stereo images are used to estimate disparity map. The stereo images with their ground truth are shown in the Figure 5.4. In case of Tuskuba stereo, the disparity map produced by proposed CNN and sparse representation algorithms and existing algorithms based on graph-cut and dynamic programming is shown in Figure 1. The disparity map produced by our proposed CNN algorithm produced less error as compared to existing algorithms. The results obtained from CCN shows that some pixels are occluded in the disparity map. The result of disparity almost similar as ground truth. The occluded pixels are shown with cyan color on the disparity map. The disparity map produced the existing algorithms have more occluded pixels as shown on the disparity. The DP produced highest occluded pixels showed on disparity map. The image disparity produced by our proposed algorithm is smoother as compared to existing algorithms as shown in Figure 5.5 (a). The Figures produced less smooth region based on existing algorithm. The disparity map produced by CNN algorithms has less error as compared to existing algorithms in case of teddy stereo image. The disparity map produced by GC algorithm has more error as shown in the cyan color on disparity map. The DP produced even worsted error in disparity map is shown in Figure 5.6. The proposed CNN algorithm produced excellent disparity map with less error rate as compared to existing DP and GC algorithms and comparatively estimate better disparity map as compared to proposed sparse representation algorithm. The comparison for existing stereo datasets has been made. The bench mark existing dataset has been used in this experiment.

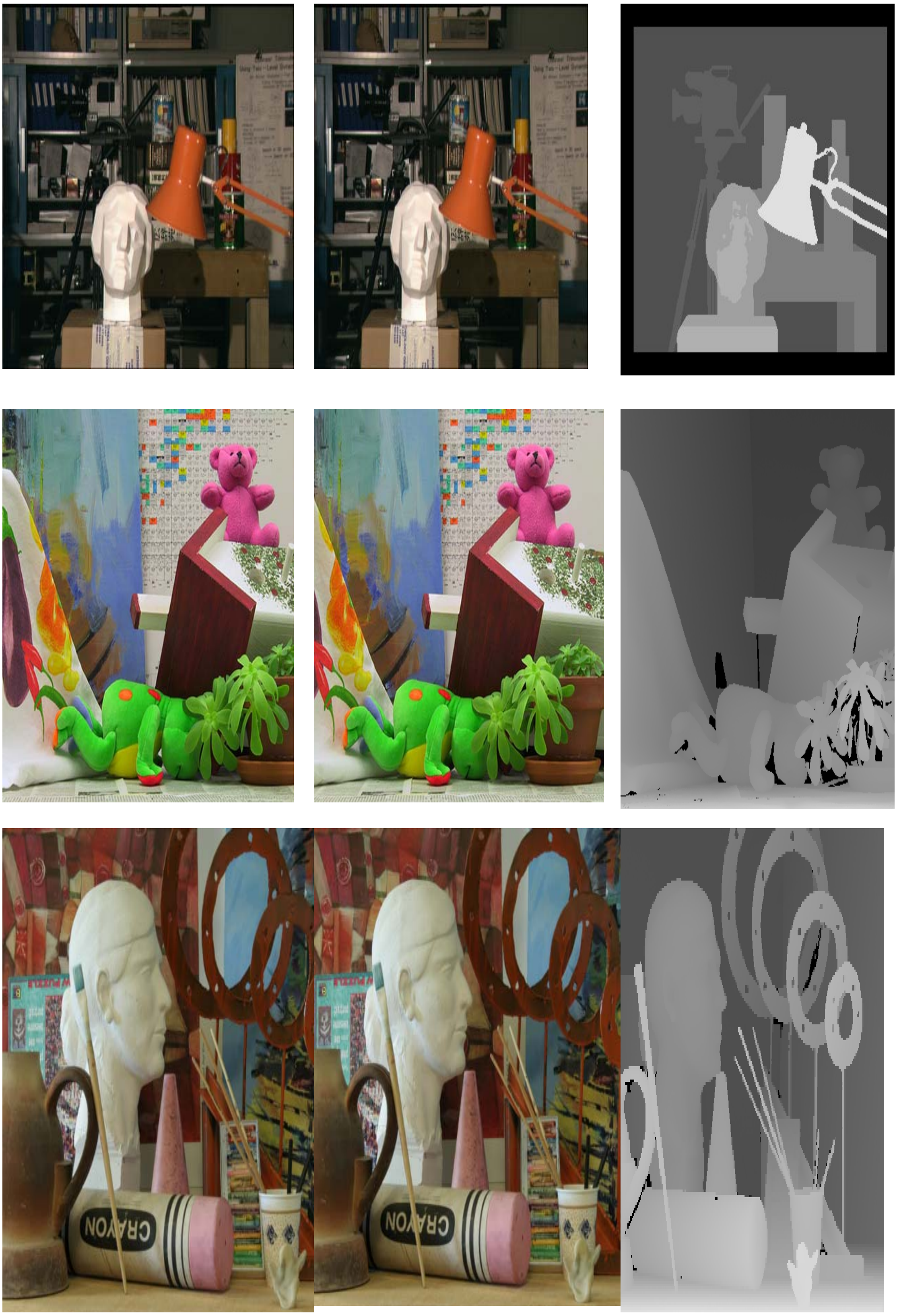
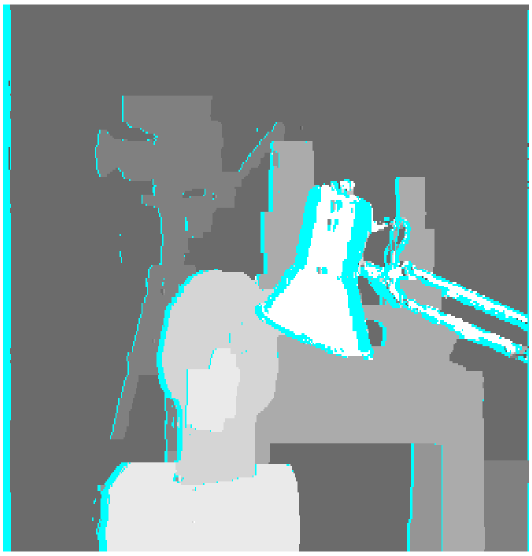
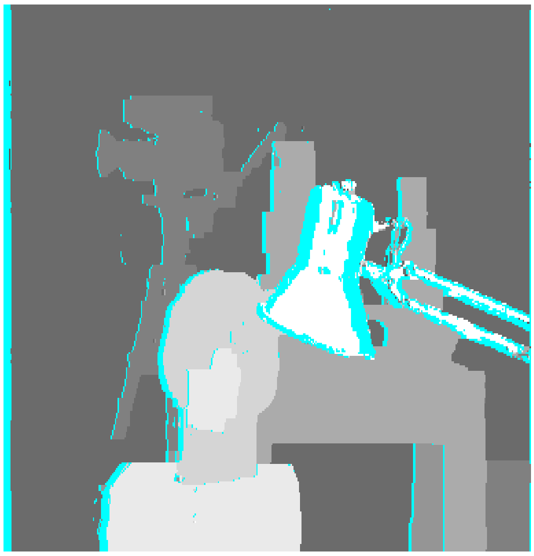


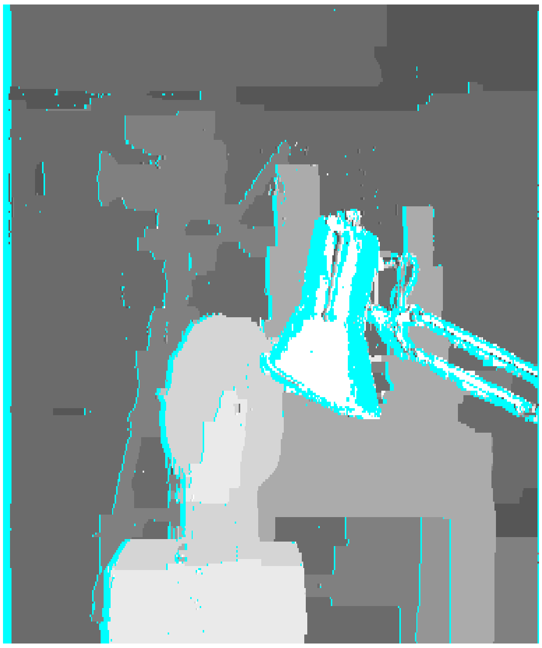
Figure 5.4: The stereo images and their ground truth using for disparity map computation.



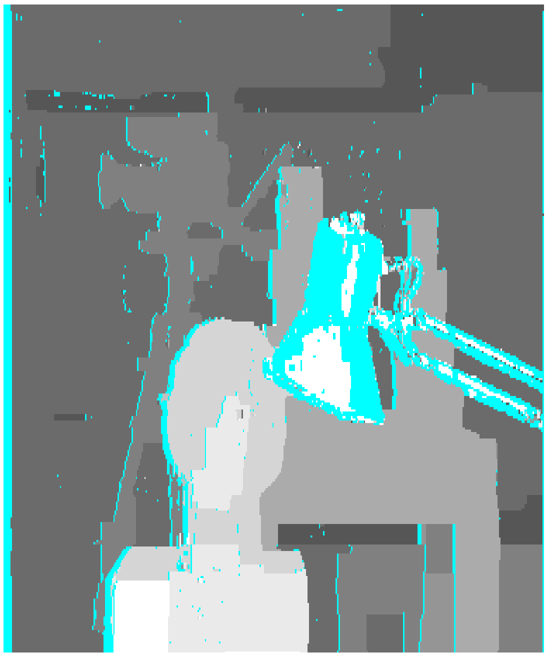
(a)



(b)



(c)



(d)

Figure 5.5: Disparity map: (a) Proposed CNN algorithm, (b) Proposed sparse Representation, (c) The Graph cut algorithm, (d) The dynamic Programming algorithm using Tuskuba stereo.

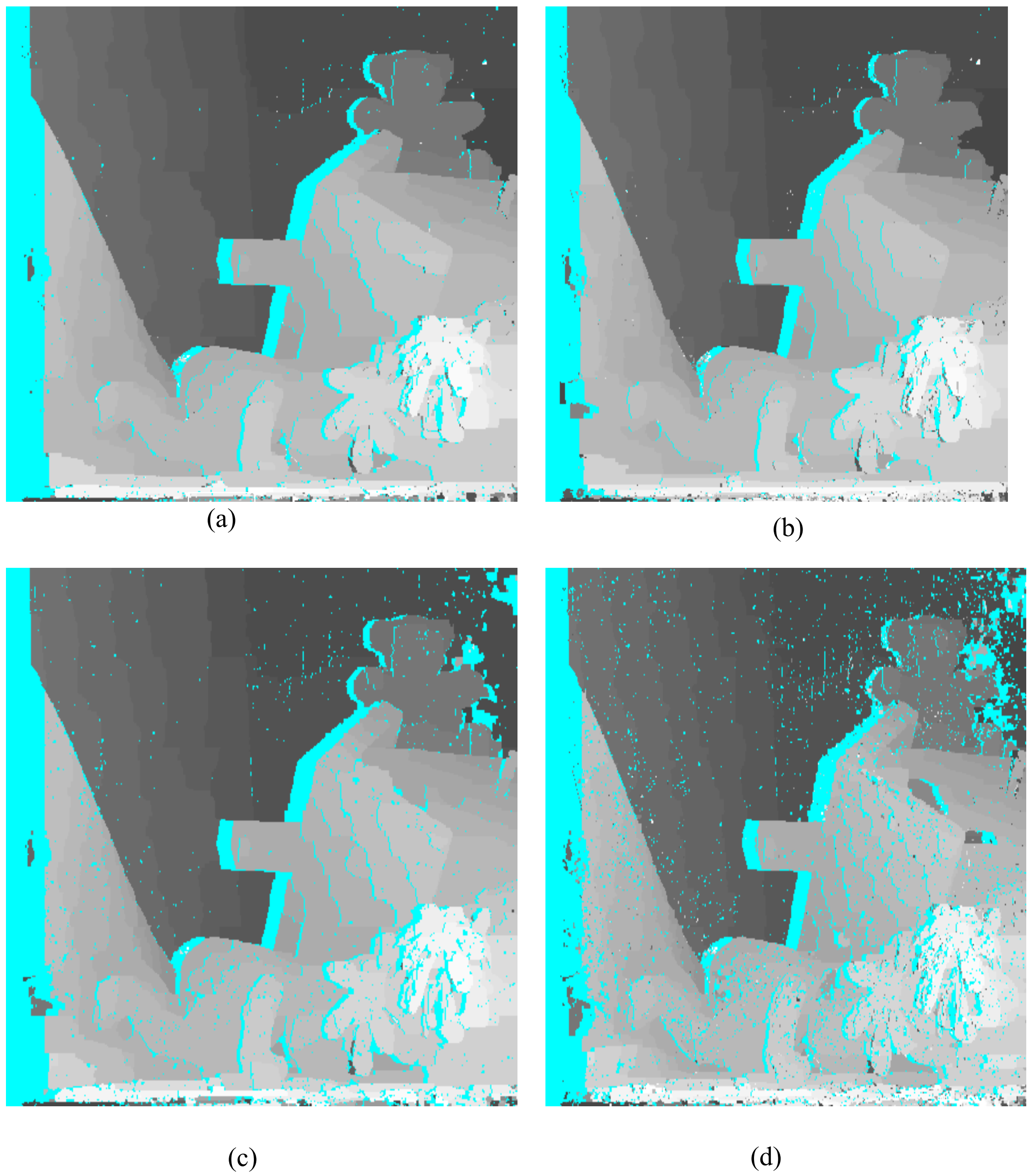


Figure 5.6: Disparity map: (a) Proposed CNN algorithm, (b) Proposed sparse Representation, (c) The Graph cut algorithm, (d) The dynamic Programming algorithm using teddy stereo.

In case of ART stereo images, the disparity map produced by proposed CNN algorithm shows less error. The error is represented by cyan color on the disparity map. The existing algorithm provided more error as compared to proposed algorithms is shown in Figure 5.7.

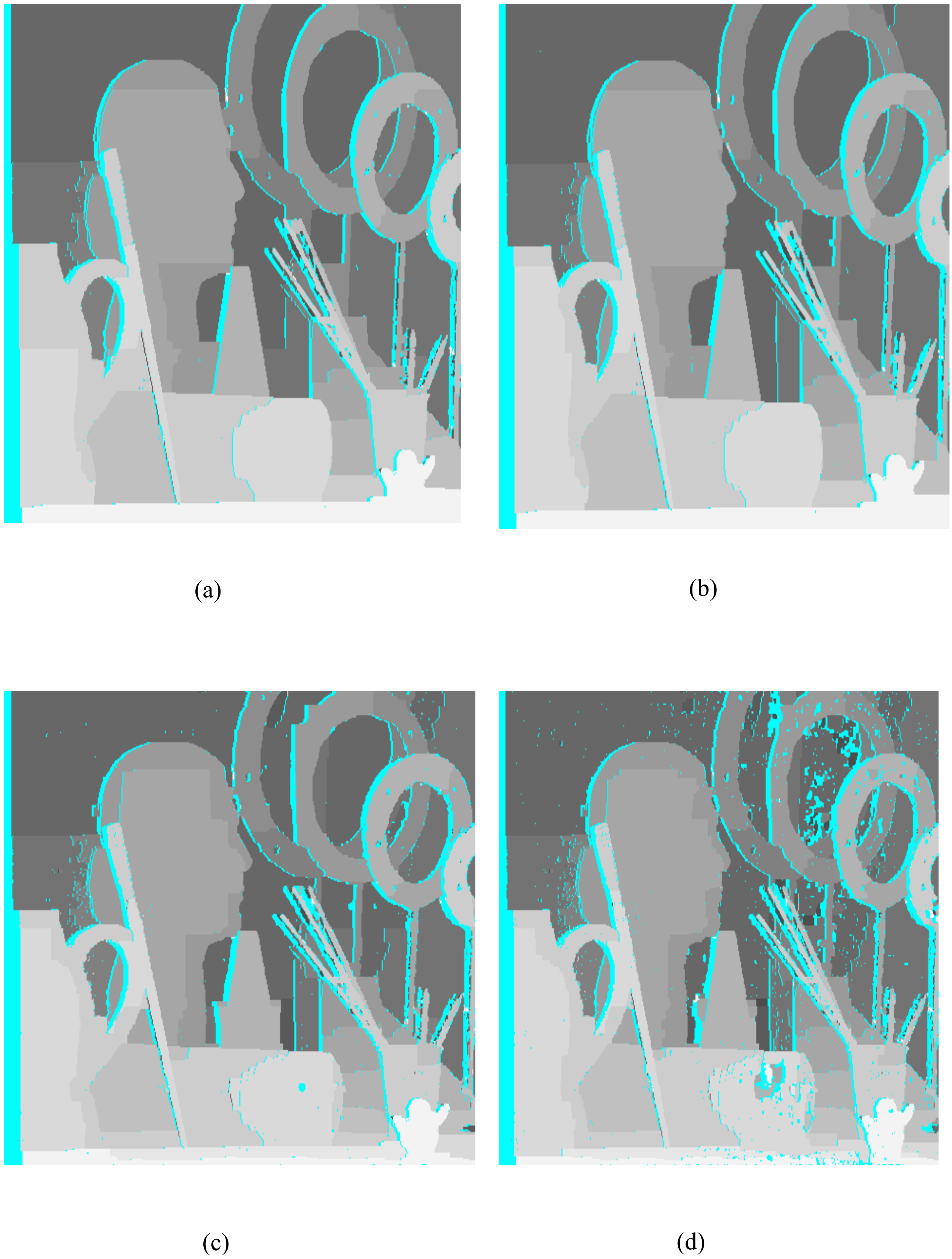


Figure 5.7: Disparity map: (a) Proposed CNN algorithm, (b) Proposed sparse Representation, (c) The Graph cut algorithm, (d) The dynamic Programming algorithm using ART stereo.

APPENDIX C

Height Estimation using Proposed Algorithms

- Height estimation of stereo images:

The height estimation between existing state-of-the-art and proposed stereo matching algorithms has been shown in this appendix. The three cases based on various disparity map has been discussed with constant number of iteration.

Table 5.3: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=20, Total towers=20**) for UAV images.

Algorithms	**Tower Height (Accuracy %)**
Area Based Method(SSD)	78.23±1.4
Area Based Method(NCC)	79.9±1.6
Dynamic Programming Algorithm	82.21±1.9
Graph-Cut Algorithm	83.45±1.7
Block Matching Algorithm	80.67±1.6
Belief Propagation Algorithm	81.95±1.3
Sparse Representation Algorithm (proposed)	**88.75**±1.3
CNN based Algorithm (proposed)	**91.25**±1.2

Table 5.4: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=20, total number of towers=20**).

Algorithms	**Precision (%)**
Area Based Method(SSD)	77.12±1.7
Area Based Method(NCC)	78.93±1.2
Dynamic Programming Algorithm	81.12±1.4
Graph-Cut Algorithm	83.56±1.3
Block Matching Algorithm	78.56±1.8
Belief Propagation Algorithm	78.90±1.8
Sparse Representation Algorithm (proposed)	**88.25**±1.5
CNN based Algorithm (proposed)	**90.97**±1.7

Table 5.5: Comparison of height in terms of recall between existing and proposed algorithms for UAV stereo images (**disparity=20, total number of towers=20**).

Algorithms	Recall (%)
Area Based Method(SSD)	79.92±1.6
Area Based Method(NCC)	78.73±1.7
Dynamic Programming Algorithm	84.66±1.6
Graph-Cut Algorithm	85.03±1.8
Block Matching Algorithm	83.21±1.7
Belief Propagation Algorithm	84.34±1.8
Sparse Representation Algorithm (proposed)	**88.15**±1.4
CNN based Algorithm(proposed)	**89.49**±1.6

The accuracy, precision and recall using our proposed and existing stereo matching algorithms for satellite images is shown in Table 5.4,Table 5.5 and Table 5.6 . The height estimated using ground truth height and estimated height produced by disparity map evaluated by our proposed algorithms for 50 number of iterations. Similarly, the precision and recall have been computed for the same specification used for measurement of accuracy.

Table 5.6: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=20, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Area Based Method(SSD)	76.34±1.1
Area Based Method(NCC)	77.91±1.2
Dynamic Programming Algorithm	80.33±1.6
Graph-Cut Algorithm	81.46±1.9
Block Matching Algorithm	77.89±1.8
Belief Propagation Algorithm	76.67±1.7
Sparse Representation Algorithm (proposed)	**87.50**±1.3
CNN based Algorithm (proposed)	**90.50**±1.5

Table 5.7: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=20, total number of towers=20**).

Algorithms	Precision (%)
Area Based Method(SSD)	75.22±1.3
Area Based Method(NCC)	75.17±1.4
Dynamic Programming Algorithm	81.44±1.8
Graph-Cut Algorithm	80.66±1.9
Block Matching Algorithm	78.44±1.8
Belief Propagation Algorithm	78.78±1.5
Sparse Representation Algorithm (proposed)	**86.59**±1.7
CNN based Algorithm (proposed)	**88.79**±1.5

Table 5.8: Comparison of height in terms of Recall between existing and proposed algorithms for satellite stereo images (**disparity=20, total number of towers=20**).

Algorithms	Tower Height
Area Based Method(SSD)	74.35±1.8
Area Based Method(NCC)	76.26±1.2
Dynamic Programming Algorithm	79.99±1.7
Graph-Cut Algorithm	81.23±1.8
Block Matching Algorithm	77.89±1.6
Belief Propagation Algorithm	75.88±1.4
Sparse Representation Algorithm (proposed)	**87.15**±1.9
CNN based Algorithm (proposed)	**89.49**±1.3

The accuracy, precision and recall using our proposed and existing stereo matching algorithms for UAV images with ten (10) disparity value and 20 number of power poles is shown in Table 5.7, Table 5.8 and Table 5.9. The accuracy may reduce due to lower disparity value range between two stereo images and it could be increase for higher disparity value. In this analysis, the disparity value is set optimal by using different number of disparity range for aerial images (UAV and satellite). The number of iteration and number of power poles for calculating accuracy, precision and recall has been same (50 iterations, 20 number of power poles) for this experiments.

Table 5.9: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=10, Total towers=20**) for UAV images.

Algorithms	Tower Height (Accuracy %)
Area Based Method(SSD)	78.30±1.6
Area Based Method(NCC)	77.26±1.4
Dynamic Programming Algorithm	81.61±1.5
Graph-Cut Algorithm	82.05±1.7
Block Matching Algorithm	79.22±1.6
Belief Propagation Algorithm	81.95±1.7
Sparse Representation Algorithm (proposed)	**87.15**±1.5
CNN based Algorithm (proposed)	**90.11**±1.9

Table 5.10: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=10, total number of towers=20**).

Algorithms	Precision (%)
Area Based Method(SSD)	76.60± 1.1
Area Based Method(NCC)	77.66±1.5
Dynamic Programming Algorithm	80.02±1.8
Graph-Cut Algorithm	82.16±1.4
Block Matching Algorithm	78.06±1.5
Belief Propagation Algorithm	77.10±1.6
Sparse Representation Algorithm (proposed)	**87.75**±1.6
CNN based Algorithm (proposed)	**90.17**±1.4

By using satellite images, the acquired results almost generate similar pattern for accuracy, precision and recall for the same specifications (disparity and number of power poles) as shown in Table 5.10. The recall, accuracy, precision and recall based on disparity ten (10) and twenty number (20) of power poles has been shown in Table 5.11, Table 5.12 and Table 5.13.

Table 5.11: Comparison of height in terms of Recall between existing and proposed algorithms for UAV stereo images (**disparity=10, total number of towers=20**).

Algorithms	Recall (%)
Area Based Method(SSD)	79.10$\pm$1.3
Area Based Method(NCC)	79.26$\pm$1.7
Dynamic Programming Algorithm	83.16$\pm$1.5
Graph-Cut Algorithm	84.73$\pm$1.4
Block Matching Algorithm	82.91$\pm$1.6
Belief Propagation Algorithm	83.74$\pm$1.5
Sparse Representation Algorithm (proposed)	**87.05**$\pm$1.8
CNN based Algorithm(proposed)	**89.11**$\pm$1.6

Table 5.12: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=10, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Area Based Method(SSD)	75.31$\pm$1.3
Area Based Method(NCC)	76.21$\pm$1.5
Dynamic Programming Algorithm	79.63$\pm$1.9
Graph-Cut Algorithm	80.36$\pm$1.7
Block Matching Algorithm	76.19$\pm$1.4
Belief Propagation Algorithm	76.27$\pm$1.5
Sparse Representation Algorithm (proposed)	**86.70**$\pm$1.6
CNN based Algorithm (proposed)	**89.90**$\pm$1.3

Further, the third scenario is considered, where the disparity values were set high and same number of power poles were used. The disparity was set to 30 and 20 number of power poles were selected. The recall is shown in Table 5.14 while accuracy and precision are shown in Table 5.15 and Table 5.16 for UAV stereo images. The performance for disparity map produced by proposed algorithm is better as compared to existing stereo matching algorithm for disparity (20) and total power poles (20).

Table 5.13: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=10, total number of towers=20**).

Algorithms	Precision (%)
Area Based Method(SSD)	73.45±1.1
Area Based Method(NCC)	75.67±1.6
Dynamic Programming Algorithm	80.14±1.4
Graph-Cut Algorithm	79.26±1.6
Block Matching Algorithm	77.24±1.5
Belief Propagation Algorithm	77.18±1.3
Sparse Representation Algorithm (proposed)	**86.13**±1.4
CNN based Algorithm (proposed)	**88.19**±1.1

Table 5.14: Comparison of height in terms of recall between existing and proposed algorithms for satellite stereo images (**disparity=10, total number of towers=20**)

Algorithms	Tower Height
Area Based Method(SSD)	77.57±1.4
Area Based Method(NCC)	78.14±1.1
Dynamic Programming Algorithm	80.19±1.8
Graph-Cut Algorithm	80.23±1.9
Block Matching Algorithm	75.88±1.7
Belief Propagation Algorithm	73.35±1.8
Sparse Representation Algorithm (proposed)	**86.95**±1.8
CNN based Algorithm (proposed)	**88.01**±1.4

The proposed algorithm produced better performance metrics using UAV stereo images by estimated the height and distance between vegetation/trees and power transmission poles and power transmission lines. The number of poles are fixed and disparity can vary to check the system performance in terms of accuracy, precision and recall using UAV and satellite stereo imagery.

Table 5.15: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=30, Total towers=20**) for UAV images.

Algorithms	Tower Height (Accuracy %)
Area Based Method(SSD)	80.23±1.7
Area Based Method(NCC)	79.64±1.6
Dynamic Programming Algorithm	82.61±1.9
Graph-Cut Algorithm	83.65±1.7
Block Matching Algorithm	80.99±1.1
Belief Propagation Algorithm	81.23±1.3
Sparse Representation Algorithm (proposed)	**88.95±1.6**
CNN based Algorithm (proposed)	**91.75±1.4**

Table 5.16: Comparison of height in terms of precision between existing and proposed algorithms for UAV stereo images (**disparity=30, total number of towers=20**).

Algorithms	Precision (%)
Area Based Method(SSD)	79.10±1.7
Area Based Method(NCC)	80.31±1.1
Dynamic Programming Algorithm	81.62±1.3
Graph-Cut Algorithm	83.96±1.6
Block Matching Algorithm	78.69±1.3
Belief Propagation Algorithm	78.97±1.4
Sparse Representation Algorithm (proposed)	**88.95±1.7**
CNN based Algorithm (proposed)	**91.07±1.9**

The accuracy comparison using satellite stereo images based on disparity (30) and total number of power poles (20) is shown in Table 5.16. The precision and recall for the same number of specification (30 disparity value and 20 number power poles) is shown in Table 5.17 and Table 5.18 using satellite stereo images. The accuracy, precision and recall based on satellite stereo images has been shown in Table 5.18, Table 5.19 and Table 5.20.

Table 5.17: Comparison of height in terms of Recall between existing and proposed algorithms for UAV stereo images (**disparity=30, total number of towers=20**).

Algorithms	Recall (%)
Area Based Method(SSD)	81.45±1.2
Area Based Method(NCC)	82.63±1.3
Dynamic Programming Algorithm	83.83±1.5
Graph-Cut Algorithm	85.71±1.7
Block Matching Algorithm	83.87±1.6
Belief Propagation Algorithm	84.78±1.6
Sparse Representation Algorithm (proposed)	**88.75±1.5**
CNN based Algorithm(proposed)	**89.99±1.4**

Table 5.18: Comparison of height in terms of accuracy between existing and proposed algorithms (**disparity=30, Total towers=20**) for satellite images.

Algorithms	Tower Height (Accuracy %)
Area Based Method(SSD)	82.15±1.7
Area Based Method(NCC)	81.78±1.9
Dynamic Programming Algorithm	83.21±1.6
Graph-Cut Algorithm	83.65±1.6
Block Matching Algorithm	80.91±1.4
Belief Propagation Algorithm	81.92±1.2
Sparse Representation Algorithm (proposed)	**88.95±1.3**
CNN based Algorithm (proposed)	**91.65±1.2**

The height estimation of power transmission poles has been measured based on proposed and existing algorithms. Different number of power transmission poles and different number of disparity value used in this experiment to estimate the height of objects. The highest performance has been achieved by using the proposed CNN algorithm as compared to extant stereo matching algorithms. The disparity (30) produced highest performance in this experiment using UAV ad satellite stereo images. These performance metrics indicate the capability of the proposed system. The higher disparity produced computational complexity in the proposed solution.

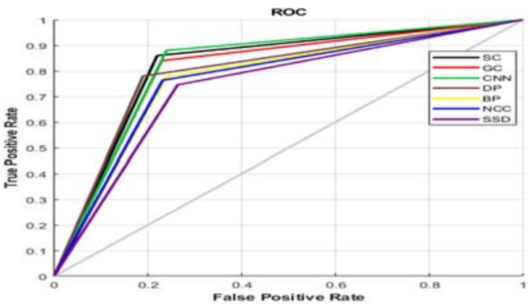
Table 5.19: Comparison of height in terms of precision between existing and proposed algorithms for satellite stereo images (**disparity=30, total number of towers=20**).

Algorithms	Precision (%)
Area Based Method(SSD)	81.43±1.3
Area Based Method(NCC)	80.18±1.7
Dynamic Programming Algorithm	81.22±1.7
Graph-Cut Algorithm	83.76±1.8
Block Matching Algorithm	78.82±1.5
Belief Propagation Algorithm	78.96±1.9
Sparse Representation Algorithm (proposed)	**88.89±1.8**
CNN based Algorithm (proposed)	**90.98±1.3**

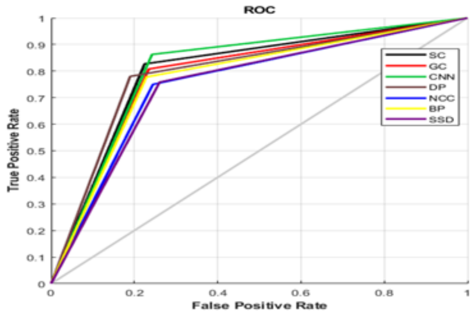
Table 5.20: Comparison of height in terms of Recall between existing and proposed algorithms for satellite stereo images (**disparity=30, total number of towers=20**).

Algorithms	Recall (%)
Area Based Method(SSD)	82.11±1.6
Area Based Method(NCC)	83.68±1.4
Dynamic Programming Algorithm	84.86±1.5
Graph-Cut Algorithm	85.83±1.4
Block Matching Algorithm	83.71±1.5
Belief Propagation Algorithm	84.64±1.4
Sparse Representation Algorithm (proposed)	**88.75±1.3**
CNN based Algorithm(proposed)	**90.01±1.6**

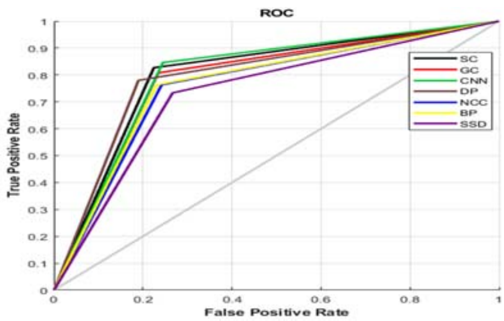
The region of convergence based on precision and recall values has been shown in Figure 5.8. It shows the different number of disparity values are used to measure the height of power poles using proposed algorithms. The estimated height and actual height can be used for precision and recall values estimation. The ROC curve is estimated based on UAV and satellite dataset using proposed CNN and sparse representation algorithm. These ROC curves are compared with the curves produced by existing state-of-the-art stereo matching algorithms. The ROC curve shows that the proposed algorithms produced higher performance curve as compared to existing algorithms. The curves produced by proposed CNN algorithm have higher values as compared to sparse representation and existing algorithm for both UAV and satellite datasets.



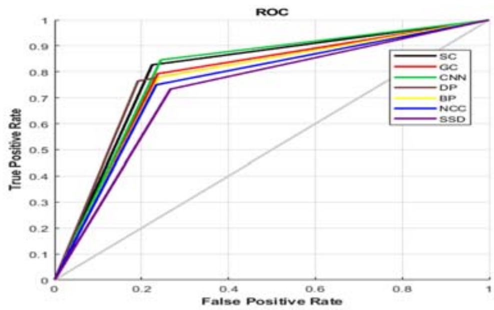
(a)



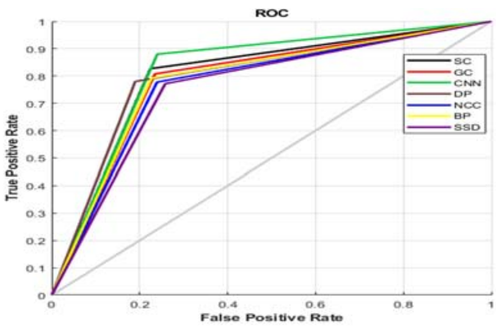
(b)



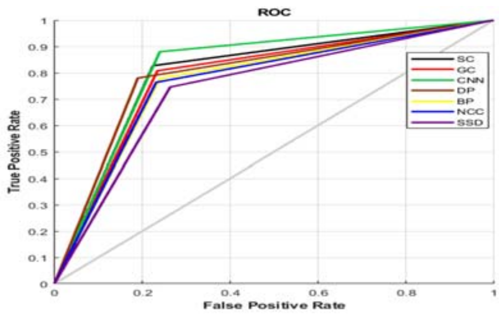
(c)



(d)



(e)



(f)

Figure 5.8: The ROC curves based on precision and recall; first column represented using UAV and second column represented Satellite based on disparity (d=10, 20, 30).

APPENDIX D

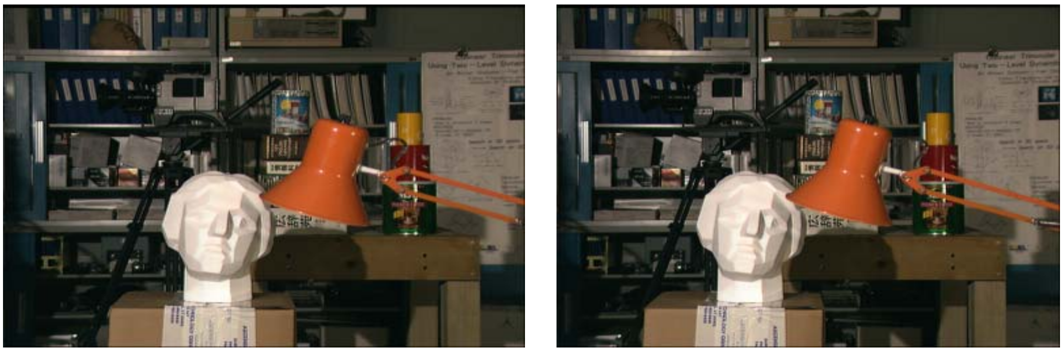
Orthorectification of Aerial Images

- Orthrectification of Aerial Images

The stereo vision fundamentals and orthorectification process used in stereo corresponding problems is discussed. The detailed interpretation of the orthorectification process using a concrete example.

- Stereovision

Binocular stereovision estimates the 3D model of a scene given two images taken from different points of view. Such images are called a stereo pair as shown in Figure 5.9. The simplest configuration that allows the estimation of a 3D map is the rectified epipolar geometry. In this case every 3D-point projected on one image is projected on the same horizontal line in the other image. These lines are epipolar lines. Basically, it corresponds to the configuration where the x-axes of the cameras are parallel to the line joining their centers and the principal (z-) axes are parallel (e.g. human eyes). When the images are not in this geometry, a rectification step can be used to warp the images. In epipolar geometry the motion of each point from one image to the other (called disparity) is horizontal shown in Figure 5.9. Moreover, Thales's theorem proves that each pixel disparity is inversely proportional to its distance from the observer. Thus, estimating the disparity map is su□cient to know the relative depth of a scene. Hence, binocular stereo algorithms usually only consist of matching every pixel of one image (the reference image) to a pixel in the other image. Image rectification is a transformation process used to project two-or-more images onto a common image plane. This process has several degrees of freedom and there are many strategies for transforming images to the common plane. It is used in computer stereo vision to simplify the problem of finding matching points between images (i.e. the correspondence problem). It is used in geographic information systems to merge images taken from multiple perspectives into a common map coordinate system. Stereo vision uses triangulation based on epipolar geometry to determine distance to an object.



(a) (b)

Figure 5.9: The Tusukaba rectified camera stereo images; (a) Left stereo image, (b) right stereo image.

More specifically, binocular disparity is the process of relating the depth of an object to its change in position when viewed from a different camera, given the relative position of each camera is known. With multiple cameras it can be difficult to find a corresponding point viewed by one camera in the image of the other camera (known as the correspondence problem). In most camera configurations, finding correspondences requires a search in two-dimensions. However, if the two cameras are aligned correctly to be coplanar, the search is simplified to one dimension - a horizontal line parallel to the line between the cameras. Furthermore, if the location of a point in the left image is known, it can be searched for in the right image by searching left of this location along the line, and vice versa (see binocular disparity). Image rectification is an equivalent (and more often used [141]) alternative to perfect camera alignment. Even with high-precision equipment, image rectification is usually performed because it may be impractical to maintain perfect alignment between cameras.

➢ Transformation Methods

If the images to be rectified are taken from camera pairs without geometric distortion, this calculation can easily be made with a linear transformation. X & Y rotation puts the images on the same plane, scaling makes the image frames be the same size and Z rotation & skew adjustments make the image pixel rows directly line up. The rigid alignment of the cameras needs to be known (by calibration) and the calibration coefficients are used by the transform [142].

In performing the transformation, if the cameras themselves are calibrated for internal parameters, an essential matrix provides the relationship between the cameras. The more general case (without camera calibration) is represented by the fundamental matrix. If the fundamental matrix is not known, it is necessary to find preliminary point correspondences between stereo images to facilitate its extraction [142]. There are three main categories for image rectification algorithms, planar rectification, [143] cylindrical rectification [141] and polar rectification. Now planar rectification using epipolar geometry is discussed. All rectified images satisfy the following two properties: [144]

- All epipolar lines are parallel to the horizontal axis.
- Corresponding points have identical vertical coordinates.

In order to transform the original image pair into a rectified image pair, it is necessary to find a projective transformation H. Constraints are placed on H to satisfy the two properties above. For example, constraining the epipolar lines to be parallel with the horizontal axis means that epipoles must be mapped to the infinite point [1, 0, 0] T in homogeneous coordinates. Even with these constraints, H still has four degrees of freedom. [145] It is also necessary to find a matching H' to rectify the second image of an image pair. Poor choices of H and H' can result in rectified images that are dramatically changed in scale or severely distorted.

There are many different strategies for choosing a projective transform H for each image from all possible solutions. One advanced method is minimizing the disparity or least-square difference of corresponding points on the horizontal axis of the rectified image pair [145]. Another method is separating H into a specialized projective transform, similarity transform, and shearing transform to minimize image distortion [144]. One simple method is to rotate both images to look perpendicular to the line joining their collective optical centers, twist the optical axes so the horizontal axis of each image points in the direction of the other image's optical center, and finally scale the smaller image to match for line-to-line correspondence [146].

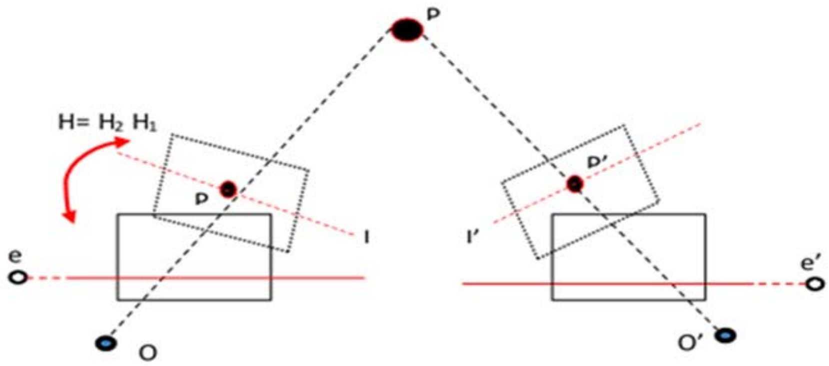


Figure 5.10: Model used for image rectification

The model used in rectification process is shown in Figure 5.10. The rectification of stereo images have the following steps explain in the Figure 5.11. The left and right stereo images are captured using stereo camera with baseline distance and wavelength. The distortion removed from the images using geometric transformation techniques. The stereo images are rectified after geometric alignment correction using image rectification techniques. The complete process of rectification of stereo images using a fundamental example is explained in the appendices with the help of mathematical formulation.

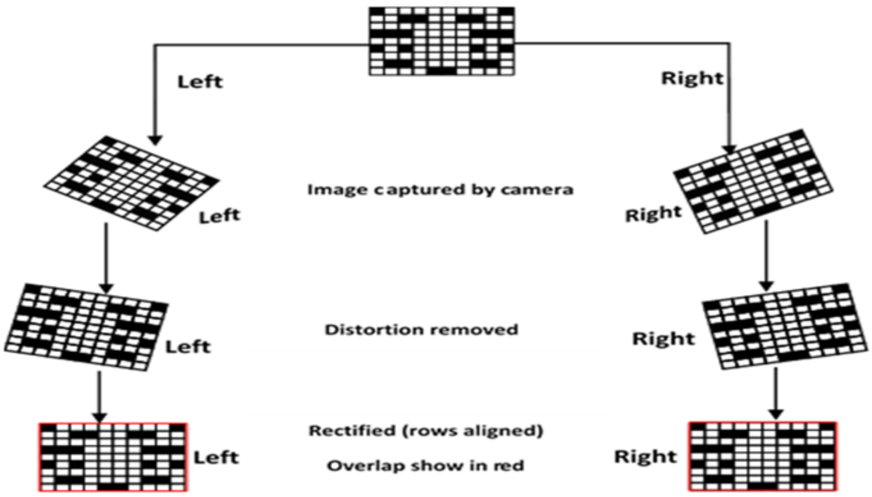


Figure 5.11: Rectification process of stereo images.

The orthorectification is the process used in stereo images to orthorectified the images in a scanline for finding corresponding points on the scanline. In this process, the first step is to extract stereo images, next select matching points between two stereo images using control points tool, After selection the GCPs points, save these points and apply transformation technique to orthorectified the stereo images. The orthorectification of stereo images have the following steps:

Step1: Read the input images

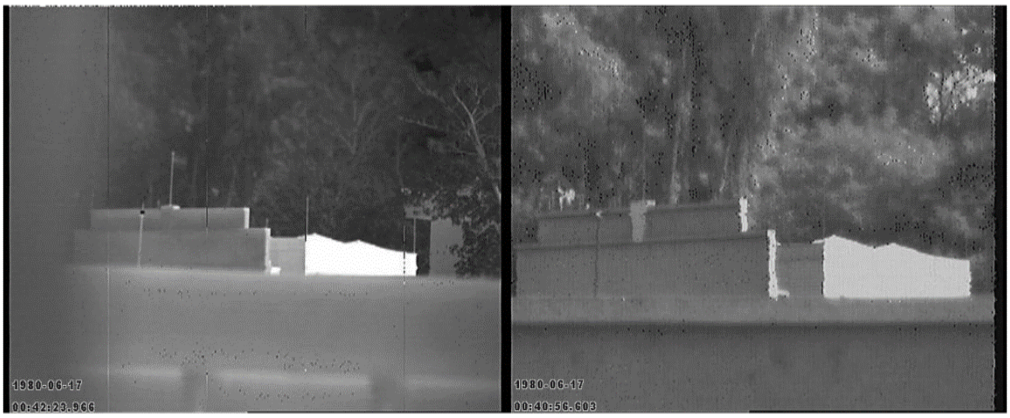


Figure 5.12: The left and right stereo pair for orthorectification process.

Step2: Choose control points in the images. In our example, the eight (8) control points in both images are chosen is shown in the following equations. The fixed points are belonging to input reference image and moving points are belonging to other image to be orthorectified. The fixed and moving points are shown in Table 5.21 and Table 5.22.

Table 5.21: Fixed Points used in orthorectification process.

310.750000000000	228.750000000000
347.750000000000	267.250000000000
190.250000000000	234.250000000000
414.750000000000	226.750000000000
405.750000000000	279.750000000000
181.750000000000	168.750000000000
429.250000000000	271.750000000000
505.250000000000	282.250000000000

Table 5.22: Moving Points used in orthorectification process.

316.250000000000	239.750000000000
343.250000000000	265.750000000000
222.250000000000	238.250000000000
393.250000000000	226.750000000000
388.750000000000	274.750000000000
221.250000000000	174.750000000000
405.750000000000	269.750000000000
461.250000000000	277.750000000000

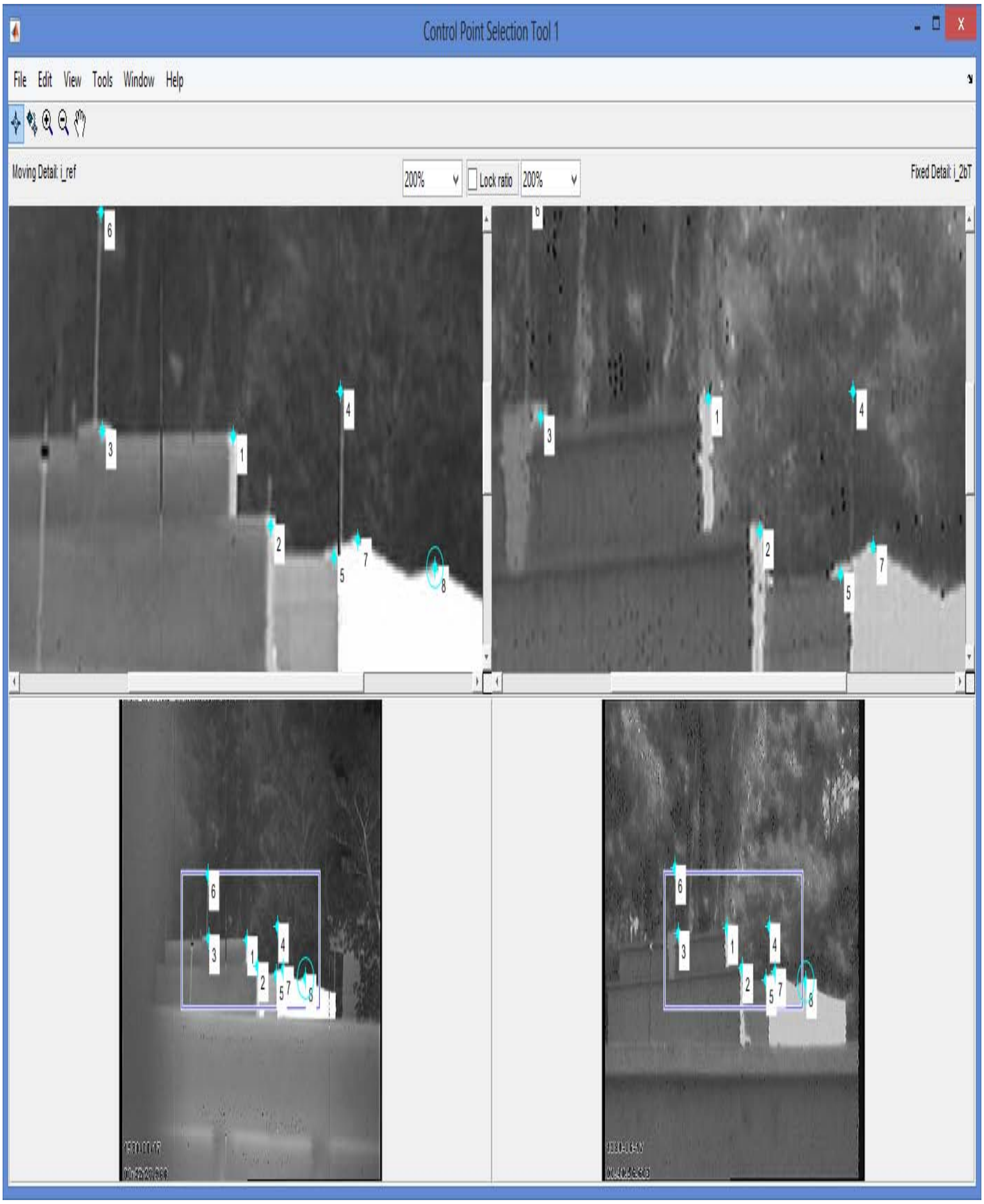


Figure 5.13: The controlling points chosen on left and right stereo images using orthorectifcation tool.

Step3: Save these fixed and moving points in the MATLAB workspace for further processing.

Step4: Fine-tune the control point pair placement. In this step, the refinement will be happen in orthorectification process.

Step5: Specify the type of transformation and infer its parameters. The orthorectified with reference image is shown below. The orthorectification process is shown in Figure 5.13. The output of the orthorectification process has been shown in Figure 5.14.

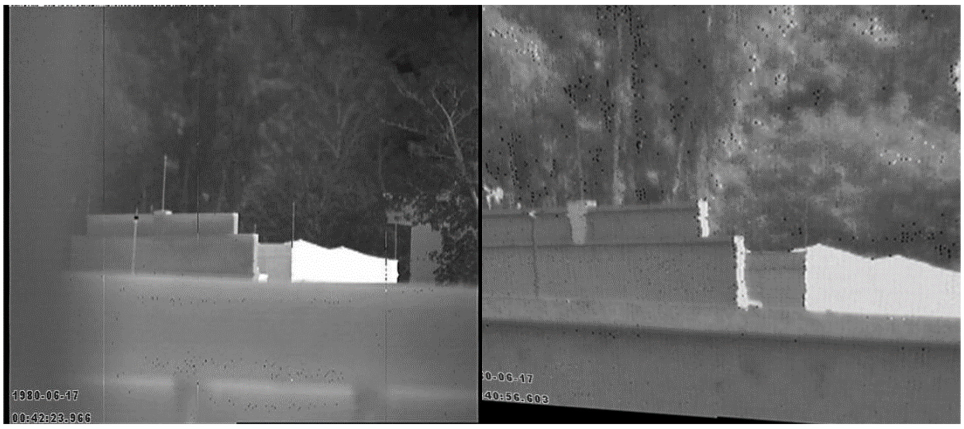


Figure 5.14: The output in orthorectified image pair form.

The mathematical explanation of transformation matrix is shown with the two point example. Let point P be defined in camera 1 coordinate system in equation (3.1d)

$$\widetilde{x_1} = K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \\ 1 \end{bmatrix} = K \begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \end{bmatrix} \tag{3.1d}$$

Then also the camera coordinate can be written as in equation (3.2d).

$$\begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \\ 1 \end{bmatrix} = K^{-1}\widetilde{x_1} \tag{3.2d}$$

If there is only rotation from camera # 1 to camera # 2, then in equation (3.3d)

$$\begin{bmatrix} X^{c_2} \\ Y^{c_2} \\ Z^{c_2} \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{c_1}_{c_2}R & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \\ 1 \end{bmatrix}, or \begin{bmatrix} X^{c_2} \\ Y^{c_2} \\ Z^{c_2} \end{bmatrix} = {}^{c_1}_{c_2}R \begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \end{bmatrix}$$

(3.3d)

$$\widetilde{x_2} = K\begin{bmatrix} X^{c_1} \\ Y^{c_1} \\ Z^{c_1} \end{bmatrix} = K_{c_2}^{c_1} R K^{-1} \widetilde{x_1}$$

The 3x3 matrix $K_{c_2}^{c_1} R K^{-1}$ is a projective transform (homograph) from image1 to image2. The transformation and inverse transformation matrix in our proposed example as calculated below. T= [1.3449 -0.0631 -0.0004 0.3359 1.4125 0.0010 -159.6865 -53.6208 0.9988]. InvT= [0.7617 0.0433 0.0002 -0.2558 0.6681 -0.0007 108.0461 42.7848 1.0000].

APPENDIX E

Threats Estimation using Proposed Algorithms

Threats in UAV images by the proposed CNN has been successfully identified. Estimated distances were quantified according to accepted standards and classed as high, medium or as discussed previously. Several ROW threats were identified. The standard clearance distance between a power pole and any object is 4 meters. Hence, any object within this range is considered a high threat. Between 4–6 m is considered a medium threat and ≥7 m is considered low threat. Thus, we quantified different threats based on estimated distances and two datasets were used to measure distances between vegetation and power poles/lines. The first was taken from satellite stereo images and second from UAV stereo images. Based on estimated vs. actual reported height of power transmission poles, the proposed method detected five 'high threats' to power poles (T1-T2, T8-T9, and T19) (See: Table 5.24). Four 'medium threats' were identified for power poles (T4, T10-T11, and T16) (See: Table 5.26). Ten 'low threats' were noted for power poles (T3-T7, T12-T15, T17-T18) (See: Table 5.33).

The different number of threat levels are quantified using sparse coding and CNN algorithm based on satellite stereo images. The number of high, low and medium threats estimated by CNN algorithm using satellite stereo images are shown in Table 5.23, Table 5.25, and Table 5.27. For sparse coding technique, the numbers of low, medium and high threats are found in Table 5.28, Table 5.29 and Table 5.31 using satellite stereo images. The number of threats found based on UAV images using proposed sparse coding algorithm are shown in Table 5.30, Table 5.32 and Table 5.34. The threats are quantified based on standard distance between vegetation or trees near power transmission poles and power lines with comparison of estimated distance. The various number of threats levels depend upon the estimated height ad distance between vegetation / trees and power transmission lines / poles and the dangerous threat indicates that the power management companies need to cut the trees near power transmission lines in order to avoid the blackouts of flashover in these areas. The proposed system may provide the feasible solution for power management companies.

Table 5.23: The number of high threats based on CNN using satellite stereo images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T1	3 m	High
2	T2	1.9 m	High
3	T8	3.9 m	High
4	T19	3.8 m	High
5	T25	2.7 m	High
6	T26	2.7 m	High
7	T36	3.1 m	High
8	T37	3.2 m	High
9	T40	3.2 m	High
10	T43	2.2 m	High
11	T45	2.4 m	High
12	T46	2.2 m	High

Table 5.24: The high threat levels based on CNN algorithm using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T1	2.3 m	High
T2	3.1 m	High
T8	1.9 m	High
T9	3.7 m	High
T19	2.7 m	High

Table 5.25: The number of medium threats based on CNN using satellite stereo images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T3	5.8 m	Medium
2	T4	6 m	Medium
3	T10	6.4 m	Medium
4	T11	6.2 m	Medium
5	T21	6.2 m	Medium
6	T22	6.2 m	Medium
7	T28	6.6 m	Medium
8	T29	6.4 m	Medium
9	T31	6.1 m	Medium
10	T34	6.5 m	Medium
11	T38	6.1 m	Medium

Table 5.26: The medium threats level based on CNN algorithm using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T4	6.3 m	Medium
T10	6.7 m	Medium
T11	4.9 m	Medium
T16	5.7 m	Medium

Table 5.27: The number of low threats based on CNN using satellite stereo images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T6	7.8 m	Low
2	T7	7.2 m	Low
3	T9	7.2 m	Low
4	T13	7.1 m	Low
5	T14	7.3 m	Low
6	T15	7.8 m	Low
7	T16	7.6 m	Low
8	T17	7.4 m	Low
9	T18	7.4 m	Low
10	T20	7.3 m	Low
11	T23	7.6 m	Low
12	T24	7.3 m	Low
13	T27	7.5 m	Low
14	T30	7.3 m	Low
15	T33	7.7 m	Low
16	T35	7.2 m	Low
17	T39	7.1 m	Low
18	T42	7.6 m	Low
19	T44	7.3 m	Low
20	T47	7.8 m	Low
21	T48	7.5 m	Low
22	T49	7.3 m	Low
23	T50	7.2 m	Low

Table 5.28: The low threat based on sparse coding algorithm using satellite images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T6	7.5 m	Low
2	T7	7.6 m	Low
3	T9	8.2 m	Low
4	T13	7.3 m	Low
5	T14	7.3 m	Low
6	T15	7.2 m	Low
7	T16	7.9 m	Low
8	T17	8.1 m	Low
9	T18	8.2 m	Low
10	T20	8.1 m	Low
11	T23	7.2 m	Low
12	T24	7.8 m	Low
13	T27	7.7 m	Low
14	T29	7.7 m	Low
15	T33	8.1 m	Low
16	T35	8.2 m	Low
17	T38	8.3 m	Low
18	T42	7.2 m	Low
19	T44	7.5 m	Low
20	T46	7.3 m	Low
21	T47	7.2 m	Low
22	T49	7.7 m	Low
23	T50	7.6 m	Low

Table 5.29: The medium threat based on sparse coding algorithm using satellite stereo images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T3	5.3 m	Medium
2	T4	6.3 m	Medium
3	T10	6.7 m	Medium
4	T11	5.9 m	Medium
5	T22	5.7 m	Medium
6	T26	5.9 m	Medium
7	T28	6.1 m	Medium
8	T30	5.9 m	Medium
9	T31	6.3 m	Medium
10	T34	6.2 m	Medium
11	T39	6.3 m	Medium

Table 5.30: The medium threats based on sparse coding using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T4	6.8 m	Medium
T11	6.9 m	Medium
T12	4.8 m	Medium
T16	6.1 m	Medium
T19	6.7 m	medium

Table 5.31: The number of high threats based on sparse representation using satellite stereo images.

No	Tower ID	Distance of vegetation near power poles and lines	Threat Status
1	T1	2.7 m	High
2	T2	2.1 m	High
3	T8	3.3 m	High
4	T19	3.2 m	High
5	T21	2.1 m	High
6	T25	2.3 m	High
7	T36	2.9 m	High
8	T37	2.3 m	High
9	T40	2.2 m	High
10	T43	3.2 m	High
11	T45	3.4 m	High
12	T48	2.7 m	High

Table 5.32: The high threats based on sparse coding the number of high threats using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T1	2.1 m	High
T2	2.9 m	High
T8	2.1 m	High
T9	3.9 m	High
T13	3.5 m	High

Table 5.33: The low threats level based on CNN algorithm using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T3	7.8 m	Low
T5	8.2 m	Low
T6	8.4 m	Low
T7	7.7 m	Low
T12	8.3 m	Low
T13	8.5 m	Low
T14	8.1 m	Low
T15	7.9 m	Low
T17	7.5 m	Low
T18	7.4 m	Low

Table 5.34: The low threats levels based on sparse coding using UAV stereo images.

Tower	Distance of vegetation near power poles and lines	Threats
T3	8.1 m	Low
T5	7.8 m	Low
T6	7.9 m	Low
T7	7.5 m	Low
T10	8 m	Low
T14	8.4 m	Low
T15	7.7 m	Low
T17	7.8 m	Low
T18	7.6 m	Low