University of Sussex

A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

NEAR REAL-TIME MONITORING OF BURIED OIL PIPELINE RIGHT-OF-WAY FOR THIRD-PARTY INCURSION

By

OLAWALE, BABATUNDE OLUMIDE

Submitted For the Degree Of
Doctor of Philosophy

ENGINEERING & DESIGN RESEARCH
SCHOOL OF ENGINEERING AND INFORMATICS
UNIVERSITY OF SUSSEX
BRIGHTON
UK

May, 2016

DEDICATION

To my late mother, Olori Olusola Elizabeth Adeyoola Olawale

Abstract

Many security systems employing different methods have been proposed to protect buried oil pipelines transporting petroleum products from the well head via the refinery to: depots and other receiving stations. Currently there is a security gap in the monitoring of these buried pipelines in real time and in keeping them protected from third party interference. This thesis addresses the problem of monitoring these systems by developing an automated image analysis system with the aid of a low-cost multisensory Unmanned Aerial Vehicle (UAV) for monitoring of buried pipeline right-of-way (ROW). The method used in this research is based on the identification of threat objects of interest from the video frame sequences of the pipeline right-of-way acquired by the UAV. This is achieved by training the system to recognise objects of interest using trained correlation filters. To determine the geographical location of detected objects, the Video frame sequences captured by the UAV platform were ortho-rectified to form ortho-images which were then mosaicked to form a seamless Digital Surface Model (DSM) covering the test area using a photogrammetry model. The DSM formed from the mosaicking of ortho-images is then emerged with a digital globe for geo-referencing of detected objects. Experiments were carried out on a test field located in United Kingdom and Nigeria, where video and telemetry data were collected, then processed using the techniques created in this research. The results demonstrated that the developed correlation filter was able to detect objects of interest despite the distortions that come with the object image, due to the fact that the expected distortion was compensated for using the training images. When compared with the 6 control points in the digital globe the accuracy of the two-dimension DSM gave a misalignment error of between 2 and 3 metres.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

ACRONYMS

ACE	Average Correlation Energy
AT	Aerial Triangulation
CAD	Computer Aided Design
CCD	Charge-Coupled Device
CPR	Correlation Pattern Recognition
CS	Control Station
DEM	Digital Elevation model
DFT	Discrete Fourier Transform
DJI	Djibouti Dow Jones
DLT	Direct Linear Transform
DoG	Difference of Gaussian
ETM+	Enhanced Thematic Mapper Plus
FFT	Fast Fourier Transform
GCPs	Ground Control Points
GCS	Ground Control Station
GPS	Global Positioning System
INS	Inertial Navigation System
LoG	Laplacian of Gaussian
LWIR	Long Wave Infrared Region
MACE	Minimum Average Correlation Energy
MACH	Maximum Average Correlation Height
MEO	Medium Earth Orbit

MVSDF	Minimum Synthetic Discriminant Function
ONV	Output Noise Variance
OT-MACH	Optimal Trade-off Maximum Average Correlation Height
PSDF	Projection Synthetic Discriminant Function filter
PSR	Peak-to-Side lobe Ratio
RF	Radio Frequency
ROW	Right-of-Way
RPV	Remotely Piloted Vehicle
SAR	Synthetic Aperture Radar
SDF	Synthetic Discriminant Function
SIFT	Scale Invariance Feature Transform
SURF	Speed-Up Robust Feature
SWIR	Short Wave Infrared Region
UAS	Unmanned Air System
UAV	Unmanned Aerial Vehicle

LIST OF PRINCIPAL SYMBOLS

u   Image plane coordinate in the x-axis direction.

v   Image plane coordinate in the y-axis direction.

w   Image space coordinate in the z-axis direction.

(X, Y, Z)   Ground coordinate system, where the Z value represents the altitude.

       above sea level for a given object point.

(u_0, v_0) Principal point coordinates.

(X_0, Y_0, Z_0) Exposure station coordinates.

f Camera focal length.

$(\Delta u, \Delta v)$ Distortion in the image point coordinates.

$L_1 - L_{11}$ Interior and exterior orientation parameters of an image.

Δu_r Radial distortion to the image plane in the x-axis direction.

Δv_r Radial distortion to the image plane in the y-axis direction.

Δv_d Decentering distortion to the image plane in the y-axis direction.

Δu_d Decentering distortion to the image plane in the x-axis direction.

Δu_u Out of plane unflattening distortion to the image plane in the x-axis

direction.

Δv_r Out of plane unflattening distortion to the image plane in the x-axis

direction.

(c, r)   Pixel column and row number respectively and is referenced as the pixel

Coordinates.

(u, v)   Image plane origin coordinates.

ω Rotational parameter of the camera system at the x-axis.

φ Rotational parameter of the camera system at the y-axis.

κ Rotational parameter of the camera system at the z-axis.

v_1 Vector of image coordinate residuals.

(P_1, P_2) Tangential distortion coefficients.

(K_1, K_2, K_3) Radial distortion coefficients.

f_u Focal length of video frame camera in the x direction.

f_v Focal length of video frame camera in the y direction.

X(a,b)   2D-Discrete Fourier Transform (DFT) of the query image (a,b).

H(a,b)   Spectrum of the reference template.

*   Complex conjugate of the filter spectrum.

Y(a,b)   DFT of the correlation output.

ρ Single Lagrange multiplier that forces ACE to a fixed value.

δ Standard deviation of the correlation values of some side lobes surrounding

the peak.

μ Mean peak value of the MACH filter.

(α, β, γ) Optimal trade-off parameters of the OT-MACH filter.

CHAPTER ONE

Introduction

Many Unmanned Aerial Vehicle (UAV) technologies have been developed, refined and used for military applications and in the public sector. These have led to useful applications in both the public and private sectors. Typically for all these applications, the key component is the on-board sensor systems (video camera and the GPS/INS). In the private sector, the sensor systems are used for different kinds of applications, among which are homeland security [1], forestry fire monitoring [2], intelligent surveillance and target acquisition [3]. Other researchers, such as [4], [5], [6], [7] and [8] have used their UAV for different kinds of application, which present different difficulties that require customized solutions.

In this research, a low-cost multi-sensor UAV is used for monitoring buried oil pipeline right-of-way (ROW). Interference with pipeline Right-of-Way (ROW) is, typically caused by a third party using construction equipment on the ROW, which can cause mechanical damage to the pipe or lead to problems ranging from pipe failure to explosion or environmental pollution. Since most of the reported damage to the pipeline ROWs are caused by third-party intrusions [9], the ability to detect third-party and construction equipment entering a pipeline ROW before it can contact the oil pipeline would greatly reduce damage caused by third party incursion.

Many approaches have been used for monitoring and reporting third-party contact or activities along the pipelines; among these are the use of wired and wireless fiber-optic sensors buried alongside the pipe, satellite technology, manned aircraft, as well as foot and car patrols. All these methods have one limitation or another, hence the need for alternative methods for monitoring the ROW.

1.1 Aims and Objective

Pipeline monitoring and threat detection are based on images taken from a low-cost multi-sensor UAV. The aim is to develop an automated image analysis system with the use of light UAVs for monitoring pipeline ROWs. This system should be able to identify potential hazards and vandals along the pipeline ROW automatically and send alerts to the pipeline response team in near real- time. The major task to be solved is the detection of:

- Human tampering and/or theft and any type of third-party intervention

- Unauthorized construction equipment

- Unauthorized vehicle on the right-of-way

1.2 Methodology

The technique used in this work is to lead the UAV to the pipeline ROW and arrange it to fly along the ROW with the aid of a Global Positioning System (GPS).The UAV auto-pilot system is then programmed with four way-points to enable the UAV to keep tracking the pipeline ROW.

For the automatic analysis of this work, the BPS video converter 1.4 software was used for de-compilation of the raw video streams acquired by the UAV into individual frames, trained Optimal Trade-off Maximum Average Correlation Height (OT-MACH) filter was used for object detection, and photogrammetry modelling, this uses algorithms based on the direct linear transform (DLT), to establish the relationship between the camera sensor used to capture the imagery, the imagery itself and the Earth's surface – the output from this was used for determination of the geographical location of the object detected from the video imagery. The system is made up of the aerial platform and the ground station. The aerial

platform consists of the UAV and sensors, which consist of a GPS, Inertial Navigation System (INS) and camera, all integrated into the UAV. The ground station consists of a portable PC computer and three servers, which comprise: an image processing application server, a threat database server and the base map server.

The UAV platform is responsible for data acquisition and the GPS and INS give the navigational position and attitude of the UAV, respectively. Data acquired are transmitted via a radio link to the ground station for processing.

At the ground station, the image processing application server, which contains the BPS software is responsible for de-compilation of the raw video streams captured under the UAV platform into frames. This module then hands over to the threat database server.

The threat database server module is responsible for detecting and identifying objects of interest on the pipeline ROW imagery that might represent a potential danger to the pipe. Here objects of interest are trained using the OT-MACH correlation filter and stored in a database. Objects detected look for a candidate match in the database by the degree of matching with each object of interest in the database. A match indicates a threat and the image is then handed over to the base map server.

The ortho-rectification and geo-referencing of video frames with the digital globe for determination of the geographical location of objects detected from the video frame takes place in the base map server.

Once a threat is detected and its location known, an alert will be sent by the pipeline operator to the response team in near real time. The architecture and workflow of the buried oil pipeline right-of-way monitoring system is shown in Fig.1.1
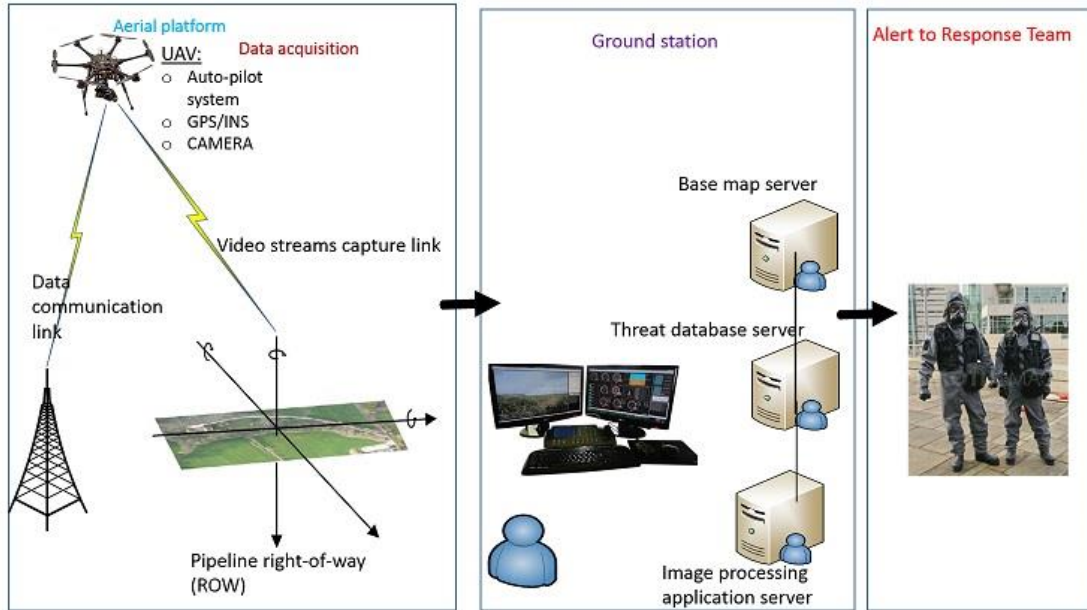
Fig.1.1 General System Architecture and Workflow Chain

1.3 **Achievements**

The major contributions of this research are:

- The OT-MACH correlation filter is trained to detect object images by cropping and training the same object of interest under different viewing angles and illumination conditions. With these, the problem of rotation and illumination variance of the correlation filter was solved. An experiment was carried out to test this method and the results are published in [10].

- For ortho-rectification and geo-referencing of video imagery the relationship between the camera sensor used to capture the imagery, the imagery itself and the Earth's surface using collinear equations was evaluated.

- An experiment using the algorithms developed was carried out in a test field where aerial imagery and telemetry data were collected. The data collected were tested with this method and the results are published in [11].

1.4 **Related Research**

There are a number of technologies used for buried oil pipeline monitoring and threat detection. Most of these technologies are based on using remote sensing [12], to detect and report potential hazards. These technologies depend on some types of communication networks, which collect data and send alerts from inside and outside the buried pipe to the control station. Different types of network architecture have been used to provide effective communication in pipeline monitoring systems. These architectures, which are either wired or wireless networks or a combination of both networks, rely on factors such as power supplies and physical network security to be effective [13], [14], [15].

Fiber optic cables [16] are used by wired networks for monitoring buried pipeline ROWs. These cables are usually connected to sensor devices that measure the flow rate, the pressure and the temperature of the oil in the pipe. The networks, which extend linearly along the pipeline, collect and send information from sensor nodes, these are spread over the pipeline and transmit their data to the control station [15], [16], [17]. Wired network based monitoring systems are faced with the following types of problems:

- If any wire in the network disconnects or is damaged, the whole pipeline monitoring system will be vulnerable to vandals.

- The physical security of the system is not guaranteed when the pipeline extends over large areas.

- Location and repairs of a faulty network can be very difficult since most of the pipelines are buried under-ground. Hence to maintain and repair a faulty network is a difficult task.

- Repeated and irrelevant signals may be transmitted on the network causing delay for other relevant signals.

[18], [19], and [20] addressed the short-comings of wired sensors by using wireless sensors to replace the wired sensors to monitor the buried pipeline system. In the wireless networks, the sensors, which are distributed inside the buried pipe along the pipeline are divided into network segments. Unlike the wired sensors, if a sensor in a segment of the pipeline fails, due to damage to one of the sensors or any destructive action, the network is not affected because other sensors in the same segment or other segments will quickly detect the faulty sensor. The damaged sensor can then be physically replaced and new ones installed. The new sensor will automatically connect to the network segment without the need to program it into the sensor network. In the wireless sensor network, each sensor node acts as a communication relay node, such that each sensor node collects information from the nearest sensor node to it. The sensor node filters the sensed data and transfers it from one communication relay node to another until it reaches a data dissemination node, which will then transfer it to the pipeline control station through another network. The problems with wireless sensors are:

- If any node in the network develops a fault, the connectivity of the segment where the node belongs will be lost and the network is partitioned.
- For pipelines that extend to large areas, a wide signal range will be needed for sensor nodes. This wireless range to stay connected will consume more energy from sensor batteries. This may lead to the requirement for frequent changes of sensor batteries.

- Since most of the pipelines are buried under the ground, the task of maintaining the sensor network is difficult.

[21] and [22] used a satellite based technology for oil pipeline monitoring and threat detection. This technique utilizes pipeline and satellite data for surveying or providing surveillance of the pipeline. The satellite data is integrated with the pipeline data to produce a current pipeline map by using change detection analysis, i.e.; the current pipeline map is compared with a previous map to determine whether the route of the pipeline or the surrounding environment of the pipeline has changed. The satellite makes use of high resolution imagery of the pipeline and the geographical location of the pipeline.

Satellite technology provides one of the most effective and efficient means for pipeline monitoring and threat detection. However, it is very expensive to build a satellite platform and sensor system, to launch it, to control it in its orbit and to recover data, as compared to operating a light aircraft with a good camera and scanner [23].

Also, for mapping to high accuracy over a relatively small area, data from sensors flown aboard an aircraft are much more useful than satellite data [21], [23]. Moreover, a satellite cannot take good quality images when the weather is cloudy. The fact that an unmanned aerial vehicle flies so much lower than satellites means that one can see more detail on the ground than can be obtained from commercial satellites.

The most widely used methods for pipeline monitoring include foot patrols along the pipeline ROW and aerial surveillance using helicopters [24]. These patrols check for unauthorized intrusion into the pipeline ROW and leakages from the pipeline.

A disadvantage of this method arises from its cost and concern for the safety of the pilot flying at low altitude, especially during bad weather. The cost of foot patrols is high in terms of personnel and their time. The use of UAVs for pipeline monitoring reduces

operational costs, speeds up the process of monitoring and can be used in situations where manned inspection is not possible.

1.5 **Thesis Outline**

The remainder of this thesis is organized as follows. Chapter 2 provides the background, which provides the pre-requisites for the rest of the Chapters in the thesis. The background contains an overview of: UAV system; remote sensing system; global position system; inertial navigation system; camera and imaging system; spatial and frequency domain filters; and image mosaicking. In Chapter 3, different matching techniques used for object image detection: feature based, area based and least square approximation are explained and compared. In Chapter 4, the procedure and results of a test experiment on object image detection using trained OT-MACH filter are discussed. Chapter 5 explains the photogrammetry model and the techniques used for ortho-rectification and geo-referencing of aerial imagery. In Chapter 6, the procedure and results of a test experiment conducted on ortho-rectification and geo-referencing of video streams captured at the UAV platform are discussed. Chapter 7 discusses the experimental procedures and results of the experiment conducted by integrating the object detection process with the ortho-rectification and geo-referencing processes. In Chapter 8, the thesis is concluded by summarizing the goals, contributions, results and the future work.
.

CHAPTER TWO

Background

2.1 Introduction

This chapter gives the basic literature review needed to understand the concepts discussed in the chapters of this thesis.

2.2 Overview of UAV

The concept of Unmanned Aerial Vehicle (UAV) started when balloons carrying explosive were used by Austria to attacked Venice in 1849 (Fig.2.1) [25]. Lawrence Sperry [26], also demonstrated his flying machine without a pilot by installing a gyrostabilizer on a Curtis Bi-plane. UAV technology used for military applications has led to useful applications in both public and private sectors. The acronym UAV has evolved to UAS (Unmanned Air System) because of the rapid growth and development of more sophisticated ground control systems, payload and other components [25], [26]. Unmanned Aerial Vehicle (UAV) is an integral part of the Unmanned Air System (UAS), but when combined with other sub − systems like control station (CS), communication systems and other support systems; it is then addressed as Unmanned Air System (UAS) [26].

In a simplistic view, an Unmanned Air System (UAS) is an aircraft with its flight crew removed and replaced by a computer system and a radio − link. However, in the real life scenario, the UAS is more complicated than that, the complete UAS comprises a number of sub − systems which includes:

- A ground control station (GCS), which houses the control systems and the operators.

- The aircraft, often referred to as Unmanned Aerial Vehicle (UAV), which comprises of payload of different types.

- The communication system, which transmits signals between the GCS and the aircraft usually through radio signals

Apart from missiles, there are three types of aircraft that can fly without pilots on board. This includes, Unmanned Aerial Vehicle (UAV), Remotely Piloted Vehicle (RPV) and Drones. They are of course, all unmanned, so the name "Unmanned Aerial Vehicle" or UAV can be thought of as the generic title [26], [27].
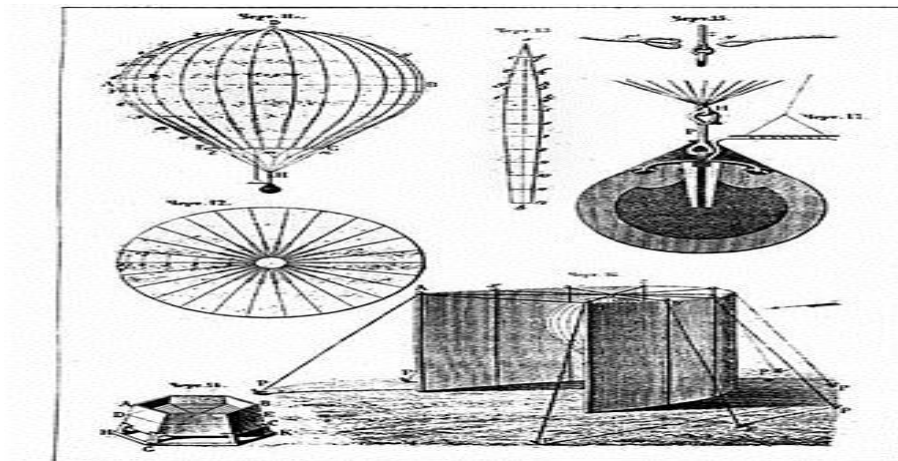
Fig 2.1 Concept of UAV: Balloon carrying explosive [25].

It is incorrect to use the terms RPV and UAV interchangeably as done by some people. The Remotely Piloted Vehicle is an aircraft without a pilot that is controlled from a remote location. Hence an RPV can always be referred to as a UAV, but a UAV which can be autonomous or pre-programmed to perform a mission need not always be an RPV [27].

Also Unmanned Aerial Vehicle (UAV) must not be confused with model aircraft or with drones. A model aircraft is also an aircraft without a pilot they are usually used for sport and are controlled by a radio. Model aircraft must stay within the sight of the operator. The operator cannot do any other thing than to instruct the aircraft to ascend or descend and to either turn left or right [26], [27].

A drone aircraft unlike the model aircraft can fly out of sight of the operator but has no intelligence. A drone is operated by launching it into a pre-programmed course or mission. It does not communicate with the control station, for example, data are usually obtained from the aircraft until it return to base. [27], [28].

A UAV on the other hand, has some level of intelligence. It communicates with the control station while on mission. It does not need to go back to base station before data can be recovered from it. For example, the UAV is able to communicate with the operator on the ground station and return payload data, such as imagery and video streams. The UAV also, transmits information about itself during flight, such as its position, altitude, component temperature and the conditions of its engine [26], [27], [28].

2.2.1 UAS System Composition

The Unmanned Air System (UAS) consist of some major sub-systems. These sub-systems form part of a total system and do not exist on their own. These sub-systems are:

2.2.1.1 **Control Station**

The control station (CS) can be based on the ground, or on a ship on shore or a parent aircraft [27], [28]. The UAV flight plan is done at the control station. Also, at the control station the operator can communicate and control the UAV via different types of communication up-links. Likewise, the UAV can send information and images to the control station through a communication down-link. These information includes, payloads data, UAV status and position information. The hovering, launching and recovery of the UAV can also take place at the control station. An example of a UAV control station is shown in Fig.2.2.

Fig 2.2 UAV Control Station (CS) [27].

2.2.1.2 **UAV Payload**

The main work of the aircraft is to carry the payload to its place of application. The type of payload a UAV carries depend on the mission the UAV was designed to carry out. These may range from a simple sub-system consisting of video camera with a fixed lens having a mass as little as 200g, to a very powerful radar having a mass up to 1000Kg [28]. Data or images obtained by this sub-system are transmitted through the communication down-link to the ground control station.

2.2.1.3 **UAV Communication Systems**

The UAV communication system provides data communication links between the aircraft and control station. The communication link could be either up-link (from the control station to the aircraft) or down-link (from the aircraft to the control station). The transmission medium mostly used is the radio frequency (RF) [29], an alternative to this medium is a light signal in the form of a laser beam or via optical fibers [28], [29].

The type of data communication links and their task is as follows:

- **Data up-link communication**: - This is the process of transferring information from the ground control station to the aircraft. The data communication up-link is responsible for the transmission of the flight way-points plan to the aircraft during autonomous flight control.

  Also, the data communication up-link transmits real-time flight control commands to the aircraft.

Moreover, data communication up-link transmits real-time commands to the payload mounted on the aircraft.

- **Data down-link communication**: - This is the process of transmitting information from the aircraft to the ground control station. It is responsible for the real-time transmission of aircraft altitude and attitude to the ground control station.

  The data communication down-link also handles the task of transmitting data captured by the payload mounted on the aircraft to the ground control station.

2.2.2 **Classification of UAV**

UAVs are classified into tiers or classes but there is no common term adopted for this classification. The Air force for an example define tier in terms of their operational altitude and endurance while the US Army define tier in terms of manufacturing company or brigade [30]. The Navy and Marines on the other hand have their own way of defining UAV tier classifications. Since there is no categorization standard, in this thesis UAVs are placed in a specific tier based on their size, weight, payload, endurance and range as listed in Table 2.1.

Tier	Weight (Kg)	Endurance	Flight Altitude (m)	Range (communication) (Km)
I (Micro) (Mini)	2 – 5	20-40 min	250	5
	25 – 150	40 min – 1hr	150 - 500	10
II (Close Range)	150 – 250	12 – 14hrs	500 - 2000	125 - 250
III (Medium Range)	200 – 250	21hrs	>3000	Global
IV (Long Range)	1000 -22000	34hrs	>15000	Global

Table 2.1 UAV Tier Structure

2.3 Remote Sensing

An image is a digital representation of an object. Remotely sensed image data is the acquisition of information about an object from a far distance [31]. Remote sensing of an object or scene can be achieved by examining radiation reflected or emitted from/by the object or scene. In practice, aerial imagery, satellite imagery and radar are all products of remotely sensed data. In order to capture this remotely sensed data, an imaging system is required. An imaging system comprises of sensors/camera, an imaging platform (airborne or satellite) and other supporting systems, such as a global positioning system (GPS), an inertial navigation system (INS) and a computer system [31], [32].

2.3.1 **Electromagnetic Spectrum**

Remote sensing system both in airborne and satellite platform, usually record electromagnetic radiation. Electromagnetic radiation is the energy that is transmitted in form of electric and magnetic waves [33]. Remote sensors consist of detectors that can record specific wavelengths of the electromagnetic spectrum. The range of electromagnetic radiation which extends from cosmic waves to radio waves is known as the electromagnetic spectrum [32], [33].

A portion of the electromagnetic spectrum is been absorbed by different kinds of land cover, such as rock, water bodies, etc. This leads to giving the electromagnetic radiation a distinguishable signature [33], [34]. Equipped with the knowledge of which wavelengths are absorbed and the intensity of the reflection of electromagnetic radiation of certain features, a remotely sensed image can be analysed and accurate assumptions can be made about the scene. The electromagnetic spectrum consists of regions, such as near-infrared and middle-infrared, which is also known as the Short Wave Infrared Region (SWIR), and the thermal or far infrared region, also known as the Long Wave Infrared Region (LWIR) [34]. This illustrated in Fig.2.3.
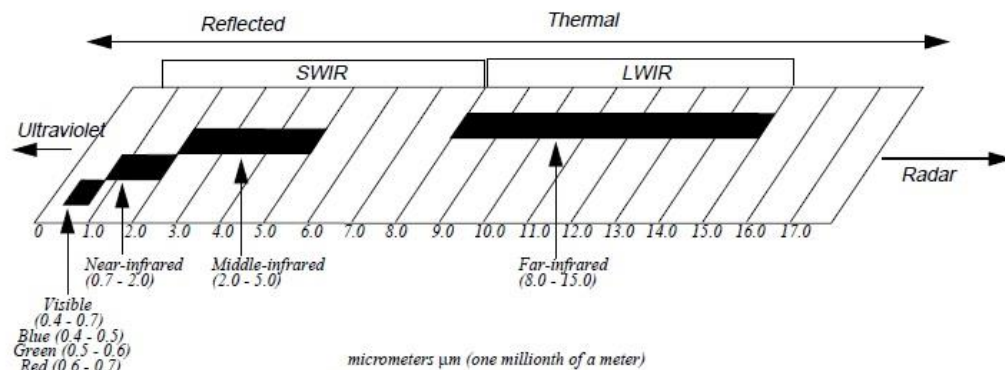
Fig.2.3 Electromagnetic Spectrum [34].

When there is an interaction between an electromagnetic radiation and objects, some wavelengths are absorbed and others reflected.

2.3.2 Remote Sensory System

It is possible to observe the interaction between an electromagnetic wave and a target material. When an electromagnetic wave strikes a target surface, there are three types of interaction that can take place, this includes, reflection, transmission and scattering. However, it is only the reflected radiation that is recorded by the remote sensor [35]. The remote sensed data comprises of reflectance values which translate into discrete values (numbers) that is recorded by the sensing device [34] [35]. These values are grey scale values (0 to 255), which fit within a certain bit range (e.g., 8 bits) depending on the properties of the sensor.

Sensors can be classified based on their sensitivity to a specific frequency of an electromagnetic radiation. A sensor that is sensitive to more than one electromagnetic radiation frequency is known as a multispectral system, this includes, Enhanced Thematic Mapper Plus (ETM+) [36]. The ETM+ is sensitive to electromagnetic radiation centered on 8 different frequencies.

Generally, multispectral system samples a small number of frequencies over a large bandwidth. A higher spectral system is the hyperspectral system, which samples a large number of frequencies over small bandwidths [35], [36]. Remote sensors can be classified as follows:

- **Infrared Sensors**: - These are type of remote sensors that sense radiation emitted by objects in a scene. The amount of radiation is highly connected with the temperature of the target object [35].

- **Optical Sensors**: - Optical sensors sense radiation at frequencies very close to human eye sensitivity. It senses the amount of sunlight that is reflected by the target object through the optical lens of the sensor. The sensor makes use of passive light from the sun and are therefore called a passive sensor. The optical sensors can only be used in daytime and can be affected by cloud cover [35], [36].

- **Radar Sensors**: - Radar sensors transmit microwave radiation to the target object which are then scattered back and stored by the sensor. The most commonly used radar sensor is the Synthetic Aperture Radar (SAR) [37]. The magnitude of the backscattered signal from the target object that is illuminated by the radar corresponds to some factors, such as, the surface roughness of the object, the electromagnetic properties of the object that was struck by the microwave signals and the frequency of the radar sensor [36], [37]. The resolution of the radar sensor is determined by the length of the synthetic aperture. Hence the resolution does not depend on the distance between the target object and the sensor.

  Radar sensors, unlike the optical sensors are not affected by weather and light conditions. However, SAR images consist of less detailed information and more noise, compared with optical sensors [37].

2.4 Global Positioning System

The global positioning system (GPS) is a space-based navigation system that uses satellite technology to pinpoint a precise location on the Earth. It does this, by using a minimum of three different satellites, which transmit radio signals to a receiver on Earth [38], [39]. The GPS was originally designed for military use, but nowadays, GPS use has extended to many applications, such as, land survey, automobile navigation, the GPS components in cell phones can be used to find and locate someone that is lost [38].

The GPS was created by the United States government. They maintain the system and make it accessible to anyone with a GPS receiver. The GPS system consists of three segments:

2.4.1 GPS Space Segment

The space segment of the GPS consists of a constellation of satellites, which transmit radio signal to the GPS receiver on the Earth [30]. The GPS satellites fly at an altitude of approximately 20,200km in medium Earth orbit (GEO), and each GPS satellite circles the Earth two times in a day [39].

The constellation of satellites in the GPS consists of a 24 slot arrangement, in which six equally-spaced orbital planes surrounds the Earth, with each plane containing four slots for baseline satellites. This slot arrangement enables the users to view a minimum of four satellites from any point on the Earth. In June 2011, three of the 24 slots were expanded, therefore increasing the constellation slot to 27-slots. This lead to improved satellite

coverage in most parts of the world [38], [39]. Fig.2.4 shows the illustration of slots arrangement in the GPS satellite constellation.
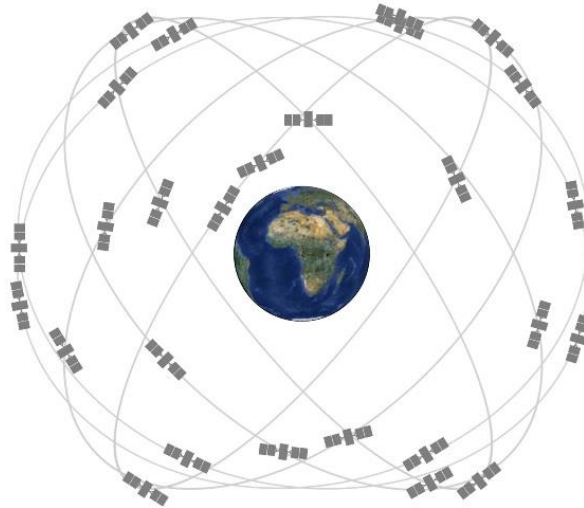
Fig.2.4 Slots satellite constellation [38]

2.4.2 **GPS Control Segment**

The GPS control segment consists of ground facilities. These ground facilities include, a master control station, monitor stations and the ground antennas [39].

The master control station sends command and control signals to the GPS constellation. It listens to the satellites to know their health state by checking their signal integrity and orbital stability.

The monitor stations are responsible for tracking the GPS satellites as they pass overhead and send their observations back to the master control station. They make use of a sophisticated GPS receiver and are controlled by the master control station.

The ground antennas are used for communications between the ground stations and the GPS satellites for command and control purposes.

2.4.3 **GPS User Segment**

The user segment consists of the GPS receiver. The GPS receiver reads available satellite signals to compute the user's position and time. In order to get an accurate fix position, the receiver must receive a minimum of three satellite signals simultaneously. The user's position refers to the 2D coordinates (latitude and longitude) position on the Earth. Fig.2.5, illustrates how the position of a user is determined using the signals from different satellites. The area marked with 'red x' is the region at which the user can be located (the region where the signals of the 3 satellites intercept).
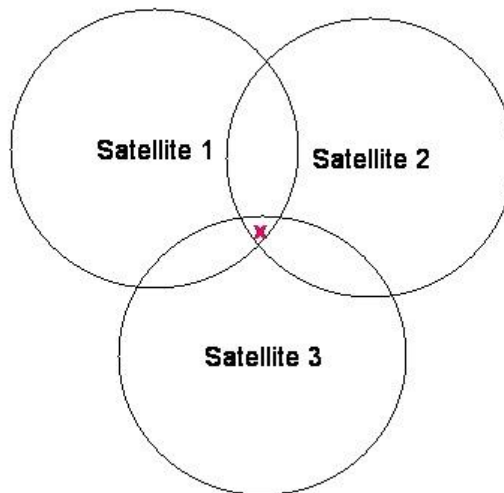
Fig.2.5 User's Location Computation using 3 Different GPS Satellites Signal [38].

2.5 **Inertial Navigation System**

A minimal inertial navigation system (INS) comprises: a computer and a module containing accelerometers, gyroscopes and other motion sensors [40]. Mostly, the INS is initially provided with its own position and velocity data from a GPS receiver. Afterwards, it then computes its own updated position and velocity by integrating information it received from the motion sensors. Once the INS has been initialized by the GPS or by any other source (e.g. human), it needs no external references in order to compute its position, velocity or orientation [41].

Most applications make use of both GPS and INS together by integration of the two for navigation purposes. The use of GPS with INS has the benefit of GPS signals calibrating the INS [4]. INS being able to provide position and angle updates at a faster rate than the GPS, it is able to fill in the gaps between GPS positions in a fast moving vehicle like missiles and aircraft [40], [41].

Another advantage of combining both the GPS and INS is that GPS may lose its signal, and during this period of loss of GPS signal, the INS can continue to compute the position and angle of the aircraft.

GPS/INS is usually used on aircraft for navigation. This allows for stable position and velocity estimates, which are provided at a sampling rate faster than when a GPS receiver is used [41]. The combined GPS and INS also allow for precise approximation of the aircraft attitude (roll, pitch and yaw) angles.

2.6 Image Acquisition using Digital Camera

Digital cameras consist of an array of CCD (charge-coupled device) sensors. CCD sensors are widely used in digital cameras and other light sensing equipment [42]. These sensors are usually arranged in two dimensional arrays. A typical CCD sensor array is manufactured with a broad range of sensing properties that can be packaged in arrays of 512 x 512 elements or more.

A complete image is captured by focusing the energy pattern onto the surface of the array. This is illustrated in Fig. 2.6, which shows show the energy from an illumination source is been reflected from a viewed scene object. After which, the imaging system collects the reflected signal and focusses it onto its focal plane. Then, the sensory array, which corresponds to the focal plane, produces outputs which is equivalent to the integral of the light signals received at each sensor and finally, the output signal received by the sensors are converted into a video signal, which is then digitized by another process of imaging discussed in the next section.
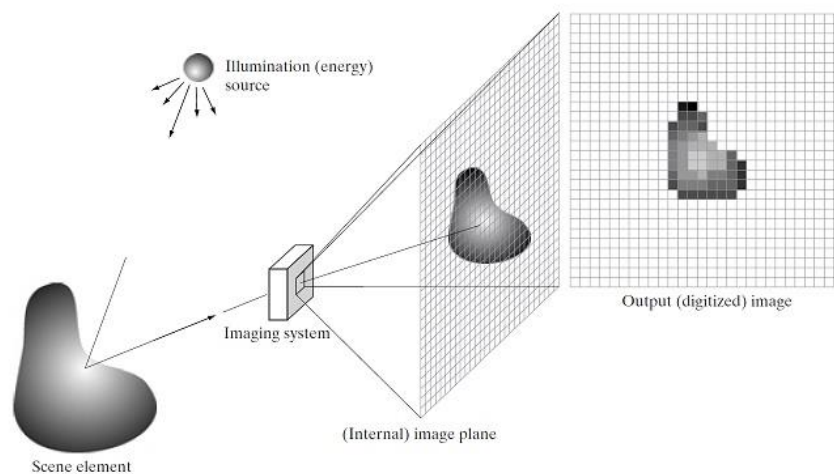
Fig. 2.6 Process of Capturing an Image on CCD Sensory Plane [43].

2.7 **Image Digitization**

As mentioned in section 2.6, the output of most sensors is a continuous voltage waveform, which represents the physical properties of the object or scene being sensed. In order to be able to analyze this continuous voltage waveform from the imaging sensors, the continuous waveform needs to be digitalized. This means, we need to convert the continuous sensed data into digital form. The two processes involved in image digitization are Sampling and Quantization [42], [43].

In a continuous image x and y coordinates as well as its amplitude need to be sampled in order to convert it to digital form. The process of digitizing the coordinate values of the image is known as sampling, while the process of digitizing the amplitude values of the analog image is known as quantization [43], [44]

Sampling includes choosing a finite number of points within an interval, while quantization involves assigning an amplitude value that lies within a range of possible finite values to each of those selected points. The outcome of the digitization process is a rectangular array of pixels, from image elements whose values correspond to their intensities for monochrome images or color components for color images. Fig. 2.7 shows the continuous image projected onto the CCD sensor array and the result of its digitalization.
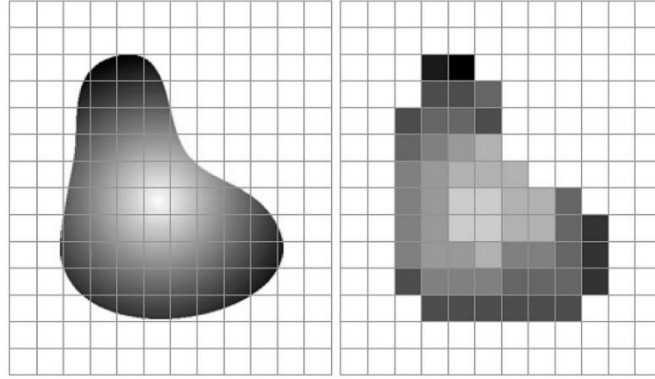
Fig. 2.7 Left: Sensed Image and Right: Result of Image Sampling and Quantization [43].

Sampling rate is the number of samples across the height and width of the image. Care must be taken when choosing a sampling rate during image sampling. Choosing inadequate values may affect the quality of the image, leading to what is known as aliasing. Aliasing or under sampling occurs, when the sampling rate is lower is lower than twice the highest frequency component of the signal [44]. This can lead to not having enough points to ensure proper reconstruction of the original signal.

2.7.1 Digital Image Representation

A digital image obtained as a result of sampling and quantization of an analogue image or one that is already in a digital form can be represented by the function of a two-dimensional array of real numbers [43]. A monochrome image $f(x, y)$ of size M x N implies x in the function $f(x, y)$ denotes the row number (between 0 and M -1) and y denotes the column number (between 0 and N -1) as shown in (2.1).

The values of a given pixel of coordinates $f(x_0, y_0)$ in the image $f(x, y)$ is called the intensity or gray level of the image at that pixel location.

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0, N-1) \\ f(1,0) & f(1,1) & \cdots & f(1, N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1, 0) & f(M-1, 1) & \cdots & f(M-1, N-1) \end{bmatrix} \tag{2.1}$$

The maximum and minimum pixel intensity value varies, depending on the data type. For example, for double data type the intensity range is between 0.0 (black) and 1.0 (white), and for unsigned 8 bits integer (unit 8), the range is between 0 (black) and 255 (white) [42], [43]. The two commonly use format for encoding the contents of a digital image is the raster (bitmap) and vector image representation formats. The raster representations make use of two-dimensional array of pixels, while the vector representations use drawing commands to represent a digital image. The raster representations have an advantage of high quality graphics and display speed. However, the raster format, occupies a large amount of memory and enlarging the raster image may lead to artifacts. The vector format on the other hand, requires less memory and is very robust to resizing and geometric manipulation without introducing artifacts [43], [44]. Selecting suitable image representation has to do with the type of application one is working on because both raster and vector formats have their pros and cons.

2.7.1.1 **Binary Images**

Binary images are two-dimensional array of pixels, containing one bit per pixel, where the pixel value 0 corresponds to black and a pixel value of a 1 corresponds to white. This type of image representation is suitable for images of very small size, such as graphics, text or line art [45]. Fig. 2.8, shows, a binary image and its corresponding pixel values in a 6 x 6 neighborhood.
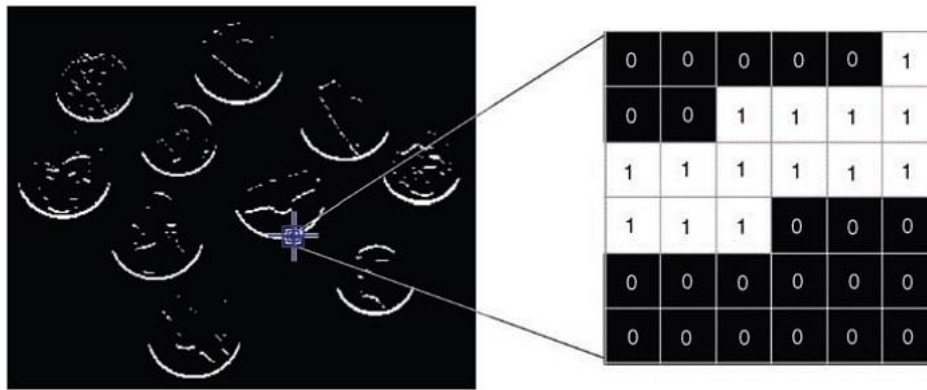
Fig. 2.8 Binary Image with its Corresponding Pixel Values [45]

2.7.1.2 **Gray-Level Images**

Gray-level images are also known as monochrome images. They are a two-dimensional array of pixels, containing 8 bits per pixel, where the pixel value of 0 corresponds to black, while pixel value of 255 corresponds to white and intermediate pixel values correspond to varying shades of gray. Fig. 2.9 shows, a 6 x 6 grayscale image, where brighter pixels correspond to larger pixel values.
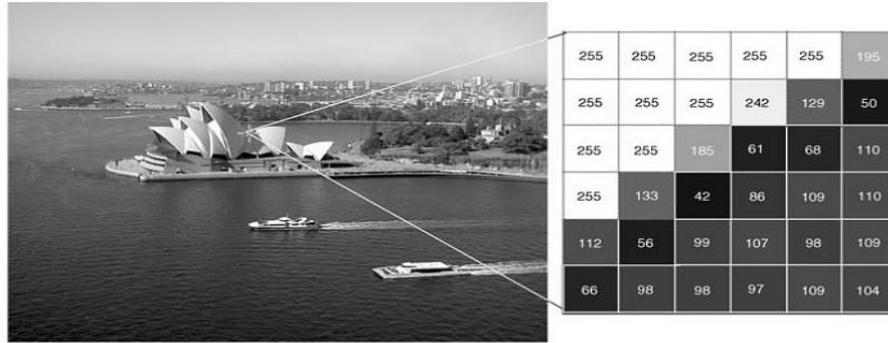
Fig. 2.9 Grayscale Image with its Corresponding Pixel Values in a 6 x 6 Neighborhood [45]

2.7.1.3 **Color Images**

Color images are stored in a RGB representation, where each pixel is represented by a 24 bits number containing the amount of red ®, green (G) and blue (B) components, and an indexed representation to a two-dimensional array containing indices to a color palette or lookup table [45], [46]. The indexing of color image and the indices is illustrated in Fig. 2.10.
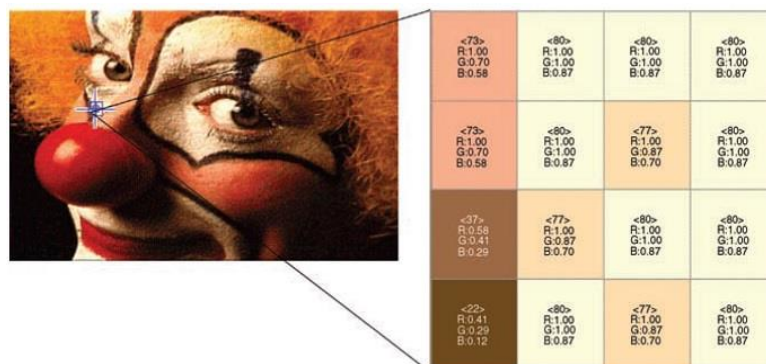
Fig. 2.10 An Indexed Color Image with a Pointer to a Color Palette [45].

2.8 **Spatial Domain Filters**

The name filter comes from the frequency domain processing. While filtering denotes to accept (pass) or reject certain frequency components [47]. A low-pass filter for example, passes low frequencies, this is normally used to blur (smooth an image. A similar smoothing can be done directly on the image by using spatial filters (also known as kernels, templates, spatial mask and windows). There is a one-to-one relationship between the linear spatial filters and filters in the frequency domain, but the spatial filters are more flexible than the frequency domain filters [47], [48].

A spatial filter consists of a small rectangular window called the neighborhood and a predefined operation that is performed on the image pixels neighbors [48]. The result of a filtering operation, is a new pixel with coordinates equal to the coordinates of the pixel at the center of the neighborhood, which have a new value as a result of the filtering operation. As the center pixel of the filter visits each pixel in the input image, a filtered image is generated.

The two major operations in linear spatial filtering are the correlation and convolution processes [49]. Correlation is the process of moving a window filter over the image and determining the sum of products at each location. While a convolution process is the same as the correlation process, except that the filter is first rotated by 180^0 [45], [46]. The best way to illustrate the difference between correlation and convolution process is with the aid of examples, which are given as follows:

Given an image;

$$A = \begin{bmatrix} 15 & 20 & 2 & 1 & 17 \\ 21 & 6 & 8 & 11 & 14 \\ 3 & 4 & 10 & 18 & 22 \\ 12 & 14 & 15 & 23 & 5 \\ 13 & 16 & 24 & 4 & 7 \end{bmatrix} \qquad (2.2)$$

and a correlation kernel;

$$K = \begin{bmatrix} 6 & 9 & 4 \\ 1 & 3 & 5 \\ 8 & 7 & 2 \end{bmatrix} \qquad (2.3)$$

The following steps can be used to determine the output pixel at location (2, 3) in the image

A:

- Slide the center element of the correlation filter, in such a way that it lies on top of the element at location (2, 3) of A.

- Multiply each weight in the correlation filter by the pixel value of A below it.

- Sum up the individual products from the output of multiplication of filter weights and the pixel values.

The correlation output of the pixel in location (2, 3) in A is:

$(6\text{x}20) + (9\text{x}2) + (4\text{x}1) + (1\text{x}6) + (3\text{x}8) + (5\text{x}11) + (8\text{x}4) + (7\text{x}10) + (2\text{x}18) = 365$

The correlation process is also illustrated in Fig. 2.11. Where the colored numbers are the values of correlation kernel, while the numbers in black are the image pixel values.

15	6x20	9x2	4x1	17
21	1x6	3x8	5x11	14
3	8x4	7x10	2x18	22
12	14	15	23	5
13	16	14	4	7

Fig. 2.11 Computation of the (2, 3) Output of Correlation.

The convolution process can also be explained using the same image A and correlation filter K as illustrated in the example for the correlation process. The convolution process is similar to the correlation process, the only difference is that the correlation filter is rotated 180^0 about its center to create a convolution filter. The output pixel using a convolution filter on pixel at location (2, 3) of A is:

$$(2x20) + (7x2) + (8x1) + (5x6) + (3x8) + (1x11) + (4x4) + (9x10) + (6x18) = 341$$

This is also illustrated in Fig. 2.12:

15	2x20	7x2	8x1	17
21	5x6	3x8	1x11	14
3	4x4	9x10	6x18	22
12	14	15	23	5
13	16	14	4	7

Center kernel

Fig. 2.12 Computation of the (2, 3) Output of Convolution

2.9 **Frequency Domain Filters**

Image processing filters that process images in the frequency domain are known as frequency filters. In this domain, the image is first of all Fourier transformed, then multiplied with the filter function and then transformed back into the spatial domain (Fig. 2.13).
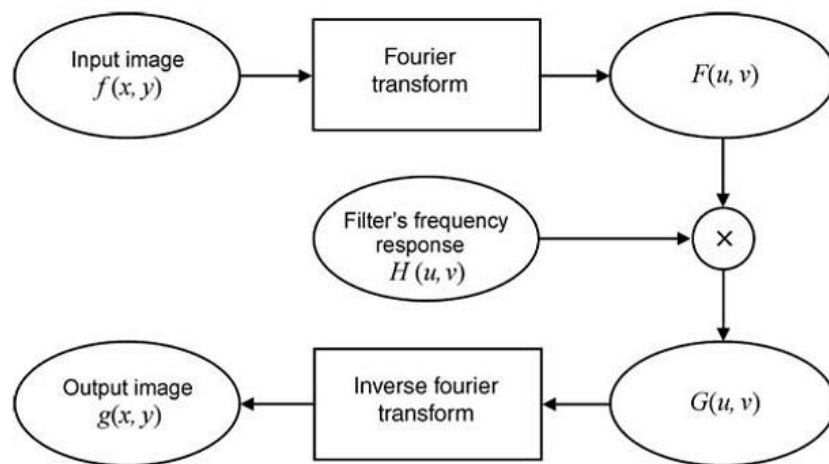
Fig. 2.13 Frequency domain filtering process

34

Filtering in the frequency domain is generally based on the Fourier transform and it has the advantage of computational speed when compared with spatial domain filtering, especially when dealing with filters of large sizes. In other words, it is computationally faster to perform a two 2D Fourier transforms and filter multiplication than to perform a convolution in the spatial domain [49], [50].

2.9.1 **Fourier Transform**

The Fourier transform is an image processing tool, which is used to convert an image into its sine and cosine components. The output of the Fourier transform is the image in the Fourier or frequency domain while the input image is usually in the spatial domain. Each point on the image contained in the spatial domain is represented by a particular frequency in the Fourier domain image. When the Fourier transform is applied to a digital image, it is known as Discrete Fourier Transform (DFT) [50].

The DFT is a sampled Fourier transform, it only contains a set of frequencies that is large enough to fully describe the image in the spatial domain, and it does not contain all the frequencies that form an image. However, the number of frequencies corresponds to the number of pixels in the spatial domain image. This implies that both images in the spatial and Fourier domains are of the same size.

Considering a square image of size N x N, the 2D DFT is given by:

$$F(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-i2\pi(\frac{xi}{N}+\frac{yj}{N})} \tag{2.4}$$

where $f(x,y)$ is the image in the spatial domain and the exponential term is the basis function corresponding to each point $F(u,v)$ in the Fourier space. Equation (2.4), can be interpreted as: the value of each point $F(u,v)$ and can be obtained by multiplying the spatial image with the corresponding base function and adding the result.

The basis functions are the sine and the cosine waves with their frequencies increasing from $F(0,0)$, which represents the dc-component of the image that corresponds to average brightness, and to $F(N-1,N-1)$ which represents the highest frequency.

In a similar way as the forward transform, the Fourier image can also be reverse transformed to the spatial domain. The inverse Fourier transform is given by:

$$f(x,y) = \frac{1}{N^2} \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v) e^{i2\pi(\frac{ux}{N}+\frac{iy}{N})} \tag{2.5}$$

The $\frac{1}{N^2}$ normalization term in the inverse transformation is sometimes applied to the forward transform, instead of the inverse transform, but it cannot be used for both transformations [50], [51].

In order to obtain the results for (2.4) and (2.5), a double sum needs to be computed for each image point. Since the Fourier transform can be separated, it can be expressed as:

$$F(u,v) = \frac{1}{N} \sum_{y=0}^{N-1} P(u,y) e^{-i2\pi\frac{iy}{N}} \tag{2.6}$$

36

where,

$$P(u, y) = \frac{1}{N} \sum_{x=0}^{N-1} f(x, y) e^{-i2\pi \frac{ix}{N}}$$ (2.7)

Using (2.6) and (2.7), the spatial domain image is first of all transformed into an intermediate image using N, one-dimensional Fourier transform. Afterwards, the intermediate image is transformed into the final image, also using an N one-dimensional Fourier transform. Stating the two-dimensional Fourier transform in terms of a sequence of 2N one-dimensional transforms reduced the number of required computations [49], [50], [51].

The one-dimensional DFT, despite the reduction of the number of required computations still has the N^2 complexity. But by introduction of the Fast Fourier Transform (FFT) [50], [51], to compute the one-dimensional DFT, the N^2 complexity is reduced.

The Fourier transform process produces two image outputs, either the real and imaginary parts or magnitude and phase. However in image processing, only the magnitude of the Fourier transform is usually displayed, because it contains most of the geometric information of the structure of the spatial domain image. Both the magnitude and phase of the Fourier image need to be preserved for use during inverse transformation of the Fourier image into the spatial domain [51].

2.10 **Image Mosaicking**

Image mosaicking is the process of merging two or more images of a scene to obtain a wider field-of-view of the scene. It is frequently used for remotely sensed images, to provide a wider view of an entire geographical area, when sensors are not able to capture the entire scene at one time [52], [53]. Image mosaicking involves four steps of image processing: registration, projection, stitching and blending.

Registration involves the establishment of geometric correspondence between a pair of images showing the same scene [53]. In order to register a set of images, the geometric transformations, which align the images with respect to a reference image within that set that must be estimated. This set of images may include, two or more images of the same scene taken at different times, from different viewing positions, and/or by different sensors. The above-mentioned correspondence between images during image registration can be established either by matching templates between images or by matching features extracted from images [53], [54].

The next step after the registration process is the re-projection step. The re-projection process involves, the alignment of the images into the same coordinate system using the computed geometric transformations.

After the re-projection step, is the stitching step, where the aligned images are joined together to form a larger image. This is done by merging pixel values of the overlapping areas and retaining pixels where no overlap takes place.

The final step in the image mosaicking process is the blending step. Errors caused due to misalignment, which often results in undesirable object discontinuities are usually visible at

the boundary between the two images during stitching process. Hence, at the blending step, a blending algorithm is used during or after the stitching of images to reduce the discontinuities in the final appearance of the mosaic image. Fig. 2.14, illustrate the four steps involved in image mosaicking.
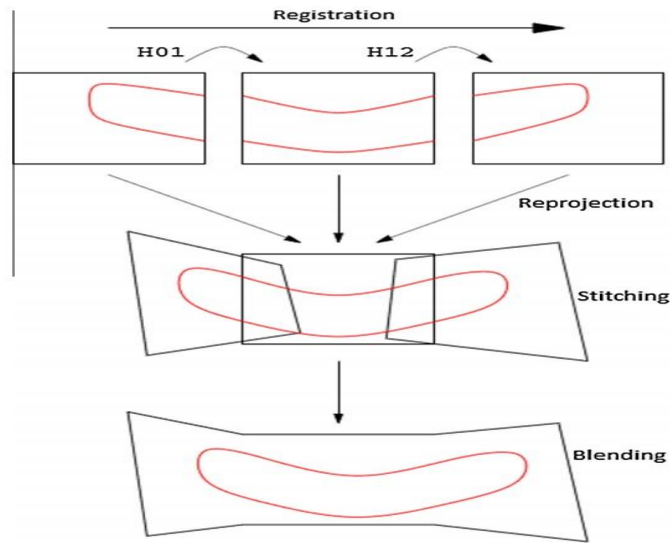
Fig. 2.14 Processes of Image Mosaicking [53].

In order to solve the problems of multiple color bands during the formation of mosaic image, the RGB images are first converted to grayscale before transformation parameters are obtained. However, all the color bands are processed and combined together at the re-projection stage to produce a color mosaic [54]. A detailed literature on image mosaicking can be found in [52], [53], [54] and [55].

2.11 **Chapter Summary**

This chapter presents the required background knowledge of digital image processing and also acts as a perquisite for topics in the remaining chapters of the thesis. The chapter consists of an overview of: UAV system, remote sensing system, global position system, inertial navigation system, camera and imaging system, spatial and frequency domain filters and image mosaicking.

CHAPTER THREE

State-of-Art for Image Matching and

Object Detection

3.1 **Introduction**

Image matching aims at matching two images to establish a degree of similarity between them. Image matching is an approach to object detection and it forms one of the major tasks in photogrammetry and computer vision, whose applications include, point transfer in relative orientation and image sequences or bundle blocks adjustment [56], which involves many points being transferred. Image matching, rather than feature points only, can also be applied to lines or segment transfer between images. Other applications of image matching are object and scene recognition, motion tracking, and texture classification [57], [58]. Image matching is a very rigorous task because of images taken under different lighting conditions, as well as image deformations and occlusions caused by different a view point of the imaging system. The process of making the pixels on one image to geometrically match with the pixels on another image, with both images corresponding to the same physical region of the scene being imaged is known as image matching [59].

Based on the aforementioned definition of image matching by Dai and Lu [59], it can be seen that image matching problems can be solved by applying the affine (i.e., translation, rotation and scaling) transformation, to one of the images in order to draw similarity with the other image as much as possible. This becomes very difficult to achieve due to the nature of the three dimensional world and more importantly because images could be taken under different lighting conditions and from different viewing points.

The common image matching methods can be divided into two major categories: the Feature based and Area based matching techniques [60], [61].

3.2 **Feature Based Matching Techniques**

Feature based matching techniques are usually applied when the information obtained from the local structure of the image is more important than the information obtained from the image intensities.

In feature based techniques, features in the referenced image are tested with the features of the sensed image. These features can be points, lines or regions [61]. However, in this research, point based features are used for image matching. The search for distinct image points in matching images can be divided into three major stages:

- **Interest Points Selection**: This is the first step of computational search, in which interest points are selected on locations on the image where there can be repeatability of the points. This means points found on a location in the image, where the points remain unchanged, when viewed under different camera positions. Such points are mostly found at locations, such as corners, T-junction and blobs.

- **Descriptor Vector**: At this stage, very interest point is represented by a descriptor vector. This descriptor must be unique and invariant to geometric and image deformations.

- **Description Matching**: This is the final stage, in which the descriptor vectors are matched between nearest images based on the Euclidean distance between their vectors. Matching time between descriptors depend on the dimension of the descriptor vector, smaller dimension descriptor vector takes less time for interest points matching compared with descriptor vector with a larger dimension. However, a small size dimension descriptor vector produces less distinct interest points than large dimension descriptor vectors.

3.2.1 **Harris Corner Detection**

The earliest use of feature based matching techniques can be found in the work of Moravec [62] in 1981, in which stereo matching was done using a corner detector. Chris Harris and Mike Stephens [63], improve on the work of Moravec by making their work more repeatable under minor image alterations and close to the edges. This was shown in 1988 in their paper, "A combined Corner and Edge Detector", which is today known as Harris Corner Detector. They expressed their idea in a mathematical form as:

$$D(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I(x,y)]^2 \qquad (3.1)$$

where:

      D = the difference between the tested window and the window that was shifted.

      u = the window's shift along x-direction.

      v = the window's shift along y-direction.

  w(x, y) = the window's location at (x, y). This ensures that merely the required window is allowed. The window function w(x, y) acts like a mask and can either be a rectangular window or Gaussian window.

  I(x, y) = the intensity of the original window.

I(x+u, y+v) = the intensity of the shifted window.

For Corner Detection, we need to look for a window that gives large value for D(u, v). This means, we have to maximize the value in the square bracket in (3.1), which is the second term in the equation.

Thus, we maximize the term:

$$\sum_{x,y}\left[I(x+u,y+v)-I(x,y)\right]^2 \tag{3.2}$$

We then, expand (3.2) using Taylor series, this gives:

$$D(u,v)\approx\sum\left[I(x,y)+uI_x+vI_y-I(x,y)\right]^2 \tag{3.3}$$

Expanding the square in (3.3), we have:

$$D(u,v)\approx\sum u^2 I_x^2 + 2uvI_X + v^2 I_y^2 \tag{3.4}$$

(3.4) in matrix form becomes:

$$D(u,v)\approx\begin{bmatrix}u & v\end{bmatrix}\left(\sum\begin{bmatrix}I_X I_x & I_x I_y \\ I_x I_y & I_y I_y\end{bmatrix}\begin{bmatrix}u \\ v\end{bmatrix}\right) \tag{3.5}$$

Hence,

$$D(u,v)\approx\begin{bmatrix}u & v\end{bmatrix}N\begin{bmatrix}u \\ v\end{bmatrix} \tag{3.6}$$

where,

$$N = \Sigma \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \qquad (3.7)$$

After D(u, v) is obtained, a score was created to help decide if a window contained a corner or not. Hence a score S was calculated for each window as:

$$S = \det(N) - \kappa(trace(N))^2 \qquad (3.8)$$

where:

$\det(N) = \lambda_1 \lambda_2$.

$trace(N) = \lambda_1 + \lambda_2$.

λ_1 and λ_2 = the Eigen values of N.

The value of these Eigen (λ_1 and λ_2) determines whether a region is flat, edge or a corner.

- If S has a small value, which implies the values of λ_1 and λ_2 are also small, then, the image area is referred to as "Flat".

- If the value of S is less than zero, which implies that either $\lambda_1 > \lambda_2$ or $\lambda_2 > \lambda_1$, then area is an "Edge".

- If the value of S is large, which implies that the values of both λ_1 and λ_2 are also large, then the area is referred to as a "Corner".

The above conditions are represented in Fig.3.1. With a suitable threshold, the outcome of Harris Corner Detector is a grayscale image that fulfilled the aforementioned score. Although feature detectors are generally referred to as corner detector, they are not based on corner picking only but rather finding image position with large gradients in all directions at a fixed scale [63], [64].

Early applications of the Harris corner detector were in stereoscopic and short range motion tracking. But later in the year 1995, Torr [65] used Harris corner for long-range motion matching, where geometric control were used on firm objects moving within an image.
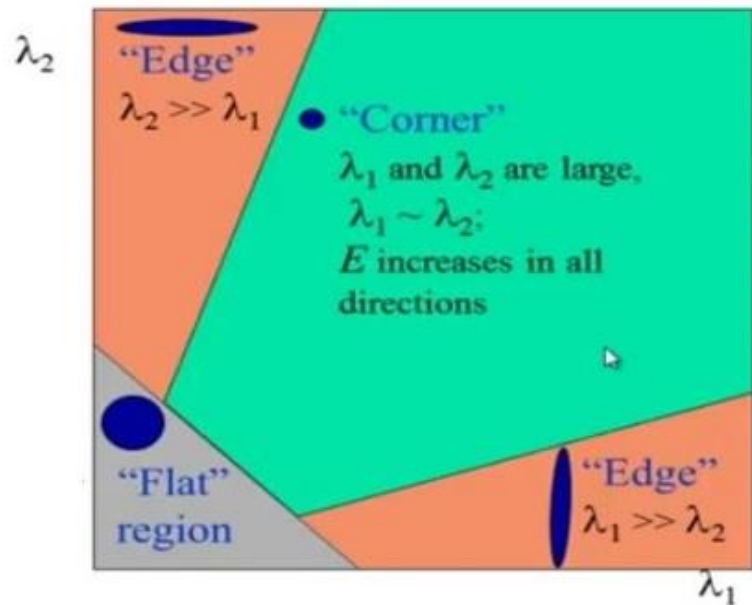

Fig.3.1. Conditions for Harris Corner Detector [64].

Two years later, Schmid and Mohr [66], solved image object detection problems by matching a test feature against a database of images. In their work, interest points were picked using a Harris corner detector and a descriptor vector was used instead of a

correlation window for image matching. This descriptor is distinct and invariant to rotation. This means, features matching is not affected by the change in orientation between matching images.

The Harris corner detector is not invariant to scale. Hence it does not yield accurate results if the matching images are not of the same sizes [65], [66].

3.2.2 Scale Invariance Feature Transform (SIFT)

In the last section, we saw that Harris Corner Detector is rotationally invariant but not scale invariant. This means, if the image is rotated, the same corner in the image can be found, this is obvious because a corner remains unchanged even if the image is rotated. However, a corner no more remains a corner if the scale of the image is changed. For example, a corner containing a small window becomes flat when zoomed in the same window. To solve the problem of scale invariance, David Lowe [67], in 1999 developed an algorithm known as Scale Invariance Feature Transform (SIFT), in which keypoints were extracted from images and are used to compute its descriptors. When the SIFT algorithm is used for image matching and object detection six processes involved are:

3.2.2.1 Scale Space Construction

The first process in the SIFT algorithm is to detect keypoints by identifying locations and scales that are invariant to scale change of the image and this is followed by searching for distinct features across all likely scales, by means of a continuous function of scale known as Scale Space.

Scale space can be constructed by taking the original image and generating a progressively blurred out images. Afterwards, the original image is resized to half its original size. Then, blurred out images are generated from it again. This process is repeated for each octave. As shown in Fig.3.2, images of the same size arrange vertically to form an octave, while images within the same octave are increasingly blurred from the top image to the bottom image. The number of octaves and blur levels used to construct a scale space depends on the size of the image. In Fig.3.3, for example, the scale space consists of three octaves and five blur levels.
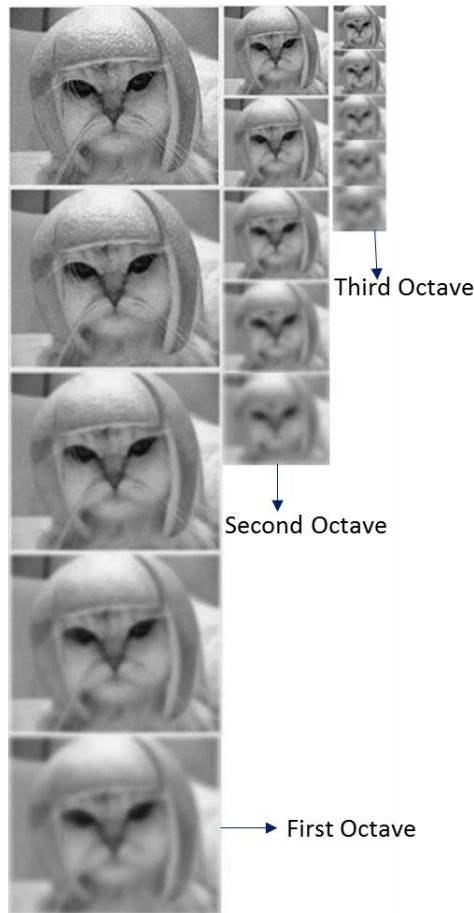

Fig.3.2. Octaves in Scale Space Construction [68].

49

Blurring referred to as the mathematical convolution of Gaussian's variable scale $G(x, y, \delta)$, with an input image $I(x, y)$. This can be expressed as:

$$L(x, y, \delta) = G(x, y, \delta) * I(x, y) \qquad (3.9)$$

where:

- L is the blurred image

- G is the Gaussian function

- I is the image

- x, y are the image position coordinates

- δ is the amount of blur, also known as scale parameter. The higher the value of δ the higher the blur.

- * is the convolution operation in x and y, and

$$G(x, y, \delta) = \frac{1}{2\Pi\delta^2} \exp\left[-(x^2 + y^2) \middle/ 2\delta^2 \right] \qquad (3.10)$$

In image blurring, a constant k is chosen, such that if the amount of blur in a particular image is δ , then, the amount of blur in the next image within an octave is k multiplied by δ (i.e., $k\delta$). Lowe [46], in his algorithm suggested a value of $\sqrt{2}$ for the constant k.

3.2.2.2 Difference of Gaussian Approximations (DoG)

In the previous section, a scale space was constructed by progressively blurring the original image, reducing the image size, blur the reduced image progressively and the process is repeated for each octave. The octaves formed from the scale space are then used to generate the Difference of Gaussian (DoG), in which adjacent images within an octave are subtracted from one another. This difference of Gaussian, which is equivalent to the Laplacian of Gaussian (LoG) was proposed by Lowe, to solve the problem of intensive computation involved in calculating Laplacian of Gaussian which contains second order derivatives that are very sensitive to noise.

The difference of Gaussian function when convolved with an image, $M(x, y, \delta)$ which is obtained from the difference between two adjacent scales separated by a constant k, can be expressed as:

$$M(x, y, \delta) = (G(x, y, k\delta) - G(x, y, \delta)) * I(x, y)$$

$$= L(x, y, k\delta) - L(x, y, \delta) \qquad (3.11)$$

The DoG produces images that are an approximation of a Laplacian of Gaussian and is very effective in detecting stable key points in scale space. The DoG is illustrated in Fig 3.3.
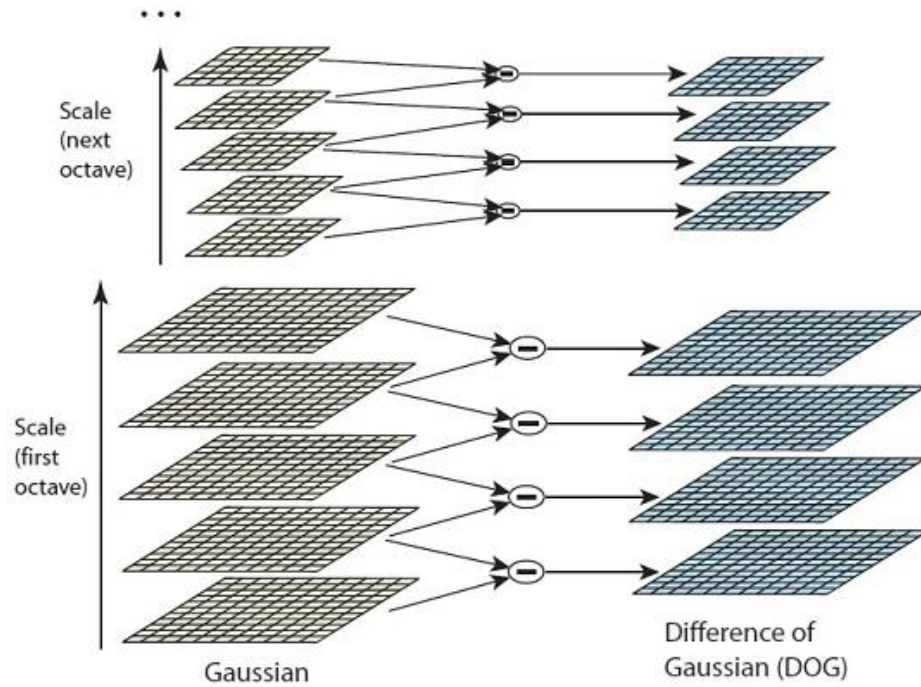
Fig.3.3. Adjacent Gaussian images are subtracted to generate DoG [67].

3.2.2.3 Finding Keypoints on the Image

After the construction of the scale space and determination of the DoG. The next step is to find the keypoints in the image. Finding keypoints in an image consists of three processes:

(a) **Location of Maxima and Minima in DoG Images**:

The process here is to locate coarsely the maxima and minima in the DoG images. This is done by iterating through each pixel and checking the entire pixel in its neighbourhood. As illustrated in Fig.3.4, the checks are done within the pixels surrounding the sample pixel.
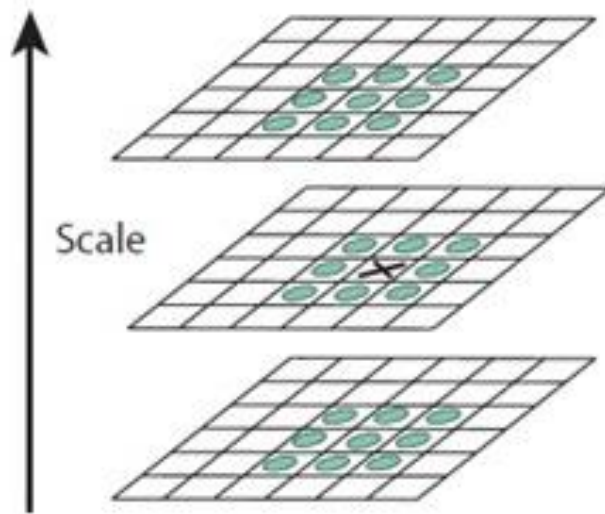
Fig.3.4 Checks for Maxima/Minima in DoG Images [67]

The sample pixel is denoted by X while the neighbouring pixels are denoted with green circles. X is chosen as keypoint if it has the largest value when compared with all its 26 neighbours. Note that, during the checks for maxima and minima, the lowermost and topmost scales are discarded because the region does not have enough neighbours to do the comparison. Hence, keypoints sample are not used in the lowermost and topmost region of the image.

All points marked as maxima or minima are actually the approximate maxima and minima because points found as maxima or minima lies somewhere in between the pixel not exactly on the pixel. Since data between pixels cannot be accessed, a mathematical model is used to find the subpixel location.

(b) **Location of Maxima and Minima in Subpixel**:

Once a candidate keypoint has been found by comparing a pixel to its neighbours, the next step is to localize keypoint by generating the subpixel values using a Taylor's expansion of the pixel surrounding the approximate keypoint.

The Taylor's expansion [69], of the scale space function, $D(x, y, \delta)$ is given by:

$$D(x) = D + \frac{\delta D^T}{\delta x} + \frac{1}{2} x^T \frac{\delta^2 D}{\delta x^2} x \qquad (3.12)$$

The subpixel keypoint location $\hat{x}$, can be determined by differentiating (3.12) and equating it to zero, given as:

$$\hat{x} = \frac{-\delta D^{-1}}{\delta x^2} \frac{\delta D}{\delta x} \qquad (3.13)$$

[67] and [69], suggested generating two extrema images, which means four DoG images are needed to achieve this, hence, five level of blurs are needed in each octave.

(c) **Removing Subpixel Keypoints with Low Contrast and the Ones Lying on the Edges**:

Most of the keypoints generated in the last section lie along the edge or have low contrast. Keypoints found in both cases are not good as features and need to be removed.

In order to remove and check for the contrast value of a subpixel keypoint, Taylor's expansion is again employed. Here, the function value of the subpixel keypoint, $D(\hat{x})$ is

used discarding subpixel keypoints with low contrast, if its intensity value is less than a user's specified threshold.

The function $D(\hat{x})$ can be obtained by substituting (3.13) in (3.12) and given as:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\delta D^T}{\delta x} \hat{x} \qquad (3.14)$$

To remove keypoints that lie along the edge of the image of DoG in a subpixel, an approach similar to the Harris corner detector method for removing edge features is used. The idea here is to calculate two gradients at the subpixel keypoint location that are perpendicular to each other. The Hessian matrix [69], is used for this purpose. Hence, the Hessian matrix is used to check weather a keypoint is a corner or not. The Hessian matrix H can be computed at the location and scale of the keypoint as:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad (3.15)$$

By making α to be the bigger eigenvalue and β to be the smaller eigenvalue, then, the sum of the two eigenvalues ($\alpha + \beta$) can be computed from the trace of H and their products ($\alpha\beta$) from the determinant of H:

$$T_r(H) = D_{xx} + D_{yy} = \alpha + \beta \qquad (3.16)$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \qquad (3.17)$$

If r is the ratio between the two eigenvalues ($r = \alpha/\beta$), then $\alpha = r\beta$ and thus,

$$\frac{T_r(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)}{r\beta^2} = \frac{(r+1)^2}{r} \qquad (3.18)$$

Based on the pixel surrounding the keypoint, the following conditions can exist:

- Flat, if the two gradients have small values

- An edge, if one gradient is greater than the other.

- A corner, if the two gradients have large values.

The corner consists the most stable keypoints. Hence, if the two gradients value are big enough it will be accepted as a keypoint, otherwise it will be rejected.

3.2.2.4 Keypoint Orientations

In the previous sections, scale invariance and stable keypoints were generated. Next is to assign orientations to each of the keypoints, in order to make it rotation invariant.

Orientation can be assigned to each keypoint by checking the gradient magnitudes and directions of each pixel surrounding each keypoint. Then the most noticeable orientation in that area is selected and this orientation is then assigned to the keypoint. Any other calculations are done with respect to this orientation. Hence, for an image sample, $L(x, y)$ the gradient magnitude, $m(x, y)$ and the orientation, $\theta(x, y)$ is computed using the pixel differences, which is given as:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$(3.19)$$

$$\theta(x, y) = \tan^{-1}((L(X, Y+1) - l(X, Y-1))/L(x+1, y) - L(x-1, y)))$$ $$(3.20)$$

Once the gradient magnitude and orientation are determined for all pixels surrounding the keypoints, then an orientation histogram is created. The histogram created consists of 360 degrees orientation which is broken into 36 bins (i.e., each is 10 degrees). Each sample of the gradient magnitude and the window size of orientation collection area that is added to the histogram is blurred by the amount of δ that is 1.5 times the scale of the keypoint [69].

For example, if the gradient direction at a certain point in the orientation collection region is 13.645 degrees, then, this will go into the 10-19 degree bin as shown in Fig.3.5.
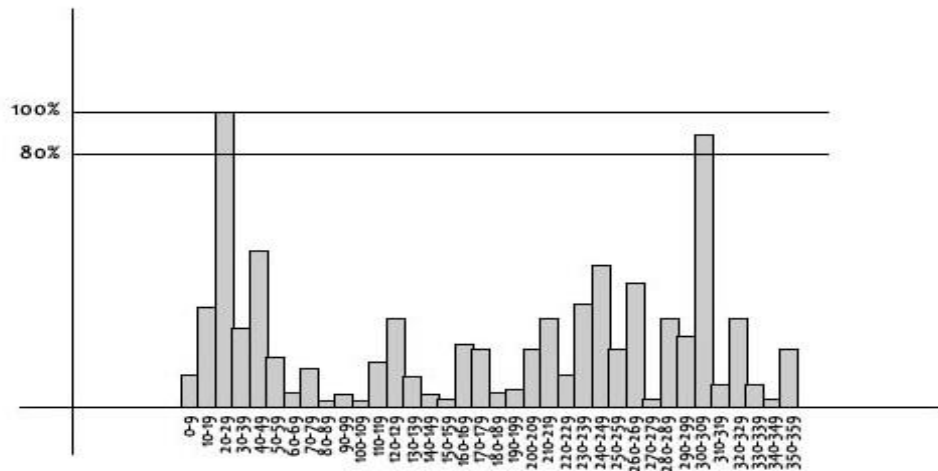
Fig.3.5.Orientation Histogram [68].

The highest peak in the histogram is at 20-29 degrees. Hence the keypoint is assigned orientation 3 in the third bin. Any other local peak that is within the range of 80% of the highest peak is used to create a new keypoint with that orientation. This means for location with many peaks of the same magnitude, multiple keypoints will be created at the same location and scale but with different orientations.

3.2.2.5 **Feature Vector Generation**

In the previous section, a stable, scale and rotation invariance keypoint is generated. The next step of SIFT algorithm process is to create a unique identity for each keypoint. For example, if an eye is a keypoint, a unique identifier that is created for the eye keypoint, will be to distinguish it from other identifier keypoints, such as ears, mouth, noses and so on.

In order to create this unique identifier for a keypoint, a 16 x 16 window is created around the keypoint location. Then the 16 x 16 window is broken into sixteen 4 x4 windows as shown in Fig.3.6.

(a) (b)
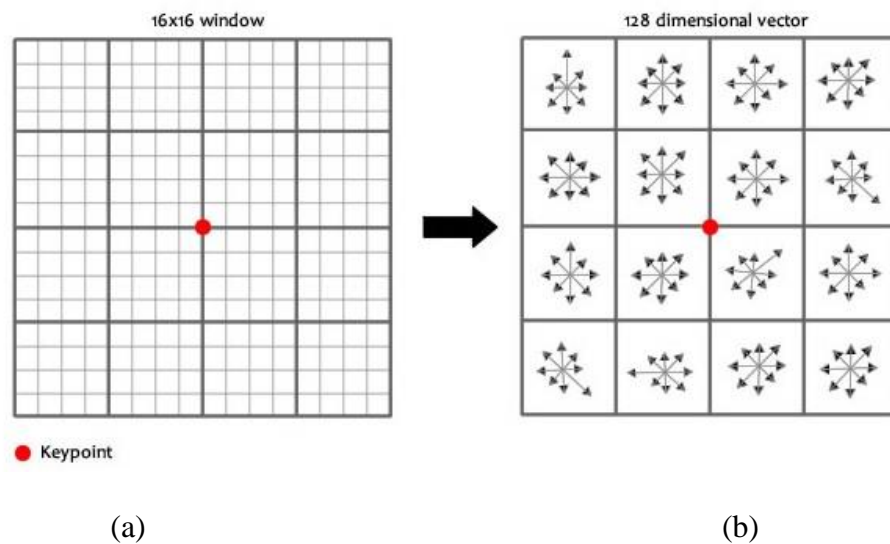
Fig.3.6 (a) 16 x16 Window Broken into (b) Sixteen 4 x 4 Window [68].

Within each 4 x 4 window, the pixel gradient magnitudes and orientations are calculated and the obtained orientations are then put into an 8 bin histogram. This is illustrated in Fig.3.7.

Added to the first bin of the histogram is any gradient orientation that falls in the $0 - 44$ degrees, and the second bin consists of gradient orientation in the range $45 - 89$ and so on. The magnitude of the gradient and its distance to the keypoints determines the amount added to the bin. This means, pixel gradient that is far away from the keypoint location will add smaller values to the histogram.

(a)   (b)
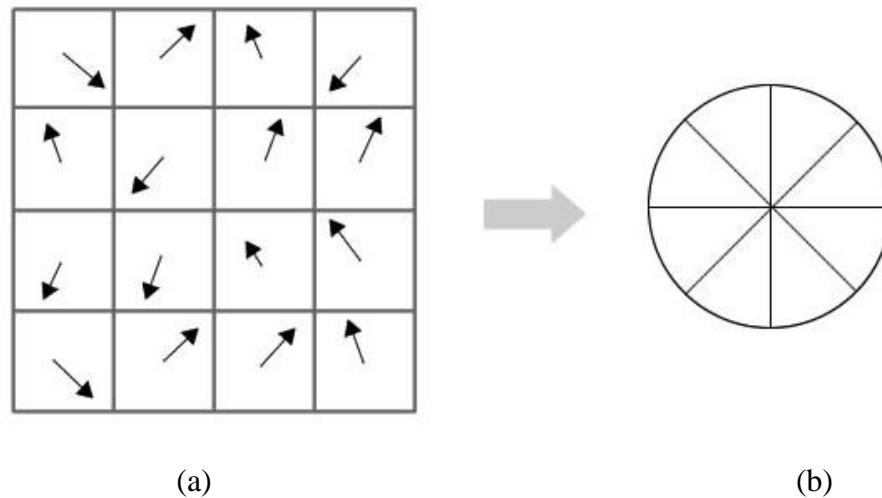
Fig.3.7. (a) Precomputed Gradients and (b) 8 Bin Histogram [68].

This is done with the aid of the Gaussian weighting function. The gradient generated by the Gaussian weighting function is multiplied by the magnitude of the orientation to get the weight of the magnitude. The further the gradient orientation to the keypoint location the lowerr the magnitude.

The gradient magnitudes and orientations are calculated for the 16 pixels as explained above and placed into 8 predetermined bins. This is done for all the sixteen 4 x 4 regions and this ends up with 128 numbers (i.e., 4*4*8 = 128). Once these 128 numbers are obtained, they are normalized and hence they form the feature vector. This feature vector is a unique keypoint identifier.

Once the feature vector is determined, it is further modified to reduce the effects of rotation and illumination change. The effect of these changes can be eliminated as follows:

- **Rotation Change**: Since the feature vector uses gradient orientation, if the image is rotated, all gradient orientation will change. To solve this problem of rotation variance, the angle at which the keypoint is rotated is subtracted from each keypoint's gradient orientation. Hence, each gradient orientation is relative to the keypoint's rotation [50].

- **Illumination change**: Illumination invariance can be achieved by the threshold of the value of each element of the feature vector to 0.2 and normalizing the resultant feature vector [68], [69].

3.2.2.6 Keypoint Feature Vector Matching

After the generation of stable keypoint descriptor (feature vector), this can be used for matching stereoscopic images, as well as object recognition.

For matching of two images, the SIFT feature vector are first extracted from each image as explained in previous sections and similar features are used to match the two images together in a process known as image mosaicking.

60

For image recognition, SIFT features are extracted from a referenced image and stored on a database. A test image to be detected has its SIFT features compared with each of the features stored in the database, in order to find a candidate match feature based on the Euclidean distance between their feature vectors.

A lot of algorithms has been used for the computational matching in the database. Among them are the Nearest-neighbour, Best-Bin-First, and Hough Transform algorithm [68].

Lowe [69], in his experiment, concluded that the Hough transform algorithm is the most efficient, especially in matching that involves cluttered images.

In clutter image, a lot of features from the background may not form a match with the features in the database, giving rise to false matches mixing with the correct matches. Hence, the correct features need to be removed from the set of matches. The filtering of correct features from a set of both false and correct features is done with the use of hash table in the implementation of Hough transform.

Each of the correct features (minimum of 3) that agree on an object and its pose is then subject to more thorough confirmation by a least squares estimate for the approximation of the object pose. Any other image keypoint vectors that are stable with this pose are identify. Image keypoint vectors that pass all these tests are identified as correct features.

3.2.3 Speed-Up Robust Feature (SURF)

The Speed-Up Robust Feature (SURF) uses the same principle and processing steps as the SIFT algorithm but it utilizes a different scheme. The SURF algorithm was introduced by Bay, H et al [70], in their paper "SURF: Speeded up Robust Features" in 2006. As the

name suggests, it is a speeded up version of the SIFT algorithm which consists of the following processing stages:

3.2.3.1 **Interest Point Detector**

In order to detect interest points on the image, the SURF algorithm uses a Hessian matrix because of its fast computational time and accuracy. Rather than selecting the location and scale using different measures, the algorithm relies on the determinant of the Hessian matrix for selecting location and scale. For a point $X = (x, y)$, in an image I, the Hessian matrix $H(x, \delta)$ in X at scale δ is given as:

$$H(Xx, \delta) = \begin{bmatrix} L_{xx}(x, \delta) & L_{xy}(x, \delta) \\ L_{xy}(x, \delta) & L_{yy}(x, \delta) \end{bmatrix} \qquad (3.21)$$

where $L_{xx}(x, \delta)$ is the convolution of the Gaussian second order derivative $\dfrac{\delta^2}{\delta x^2} g(\delta)$ with the image I at point x.

Similarly, $L_{xy}(x, \delta)$ and $L_{yy}(x, \delta)$ are given as:

$$L_{xy}(x, \delta) = I(x) * \frac{\delta^2}{\delta x \ y} g(\delta) \qquad (3.22a)$$

$$L_{yy}(x, \delta) = I(x) * \frac{\delta^2}{\delta y^2} g(\delta) \qquad (3.22b)$$

[70] and [71], show in their findings that the Gaussians are optimal for scale-space analysis. The convolutions is approximated and speed-up by using approximated kernels and integral images.

Integral images are two-dimensional lookup tables in the form of a matrix having the same size as the original image [71]. The sum of all pixels at the upper left area of the original image is contained in each element of an integral image with respect to the element's position. This enables the computation of the sum of rectangular areas in the image, at any scale or position with the aid of only four lookups:

$$I(x) = \sum (A + C - B - D) \qquad (3.23)$$

Where points A, B, C, and D, are elements of the integral image I, as shown in Fig.3.8.
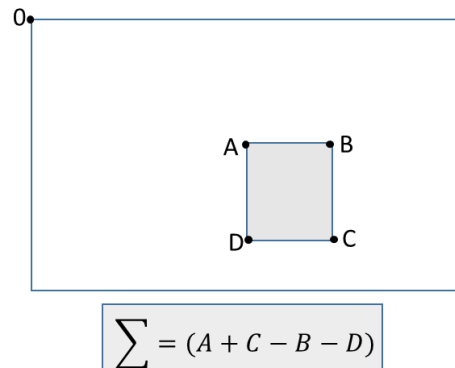
Fig.3.8. Finding the Sum of Shaded Area on an Integral Image.

In practise, the second order derivative Gaussian Kernels $\frac{\delta^2}{\delta y^2} g(\delta)$ that was used for the

Hessian matrix must be discretized and cropped before applying them to a kernel.

Afterwards, the kernel is approximated with a rectangular box filters. This is illustrated in

Fig.3.9, where grey areas are denoted as 0 in the kernel, white areas are denoted as positive

and black areas as negative.

(a)   (b)
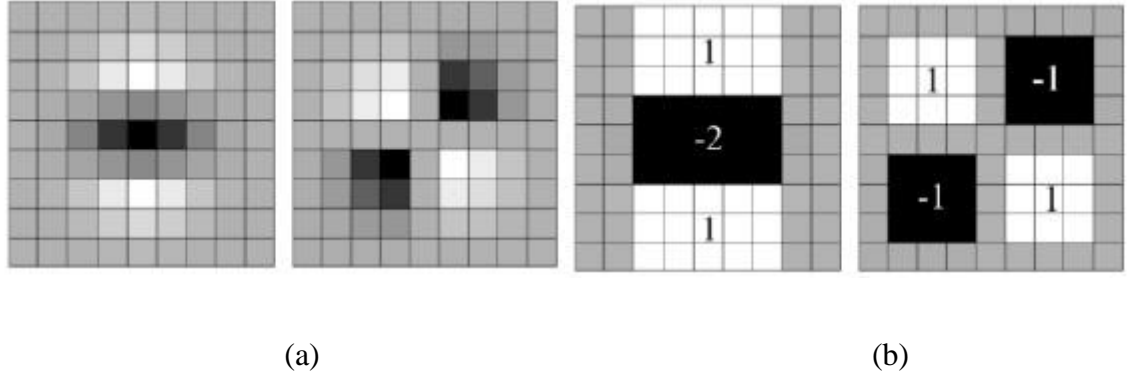
Fig.3.9 (a) Discretized Gaussians (b) D_{yy} and D_{xy} Approximations [70]

In this way, the approximated convolution is effectively calculated for a randomly sized

kernel making use of the integral image.

The approximated convolution for a randomly sized kernel, making use of the integral

image can be calculated by:

$$Det(H_{approx}) = D_{xx}D_{xy} - (wD_{xy})^2 \qquad (3.24),$$

where the approximations for discretized Gaussians $L_{xy}(x,\delta)$ and $L_{yy}(x,\delta)$ are denoted by

D_{xy} and D_{yy} respectively.

[69], shows in his paper that the lowest scale that can be processed by a SURF algorithm correspond to a δ value of 1.2. When calculating the determinant of the Hessian matrix, using the approximated kernels, it needs to be weighted (wD_{xy}), as shown in (3.24). The term **w** is known to be theoretically sensitive to scale but can be kept constant by assigning a value of 0.9 to it. Hence, the equation (3.25) becomes:

$$Det(H_{approx}) = D_{xx}D_{xy} - (0.9D_{xy})^2 \qquad (3.25),$$

In SURF, scale space are normally implemented as image pyramids. The images are repeatedly blurred with Gaussian kernels and subsequently sub-sampled in order to achieve a higher level of the pyramid. Features are detected across scales by examine many octaves and blur levels. Unlike the SIFT algorithm, where the image scales are progressively reduced and larger Gaussians Kernels are used. The SURF algorithm, is analysed by up-scaling the Gaussian kernels rather than reducing the image scale. This comparison is illustrated in Fig.3.10.
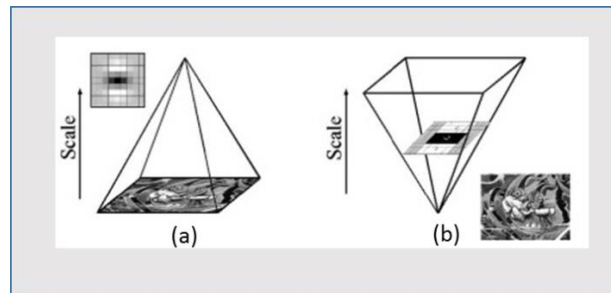
Fig 3.10 (a) SIFT: Image Size Progressively Reduced (b) SURF: Up-Scaling Gaussian Kernels Instead of Reducing Image [70]

A non-maximum suppression in a 3x3x3 neighbour is used to localise interest points in the image and over the scales. Then with the method proposed by Brown et al [57], the maxima of the Hessian matrix's determinant can be interpolated in the scale and image space.

3.2.3.2 **SURF Descriptor**

The purpose of a descriptor is to offer a unique and strong description of an image feature based on the pixels that surround the interest point. A descriptor must be computed for each interest point found in the image.

(a) **Orientation Assignment to Interest Point**

The SURF algorithm is very robust to rotations, an upright version of the SURF referred to as U-SURF is very robust to rotations within the range of $\pm 15^{0}$, without having to perform orientation assignment [70], [71].

In order to make the interest point rotation invariant, the Haar wavelet response in the x and y direction, which lies within the pixels in a circular neighbourhood of a user's defined radius around the interest point is computed. The Gaussian function centred at the sample point within a circular neighbourhood of the interest point is used to weigh the obtained Haar wavelet response (Fig.3.11), and the dominant orientation is detected by a sliding window of size $\Pi/3$ (Fig.3.12).
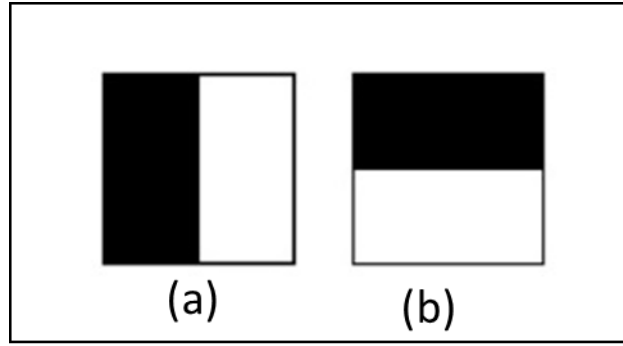
Fig.3.11 (a) Haar Wavelet Filters to Compute Responses in x Direction (b) and y Direction. Black

Side have Weight -1 and the White Side +1 [70].
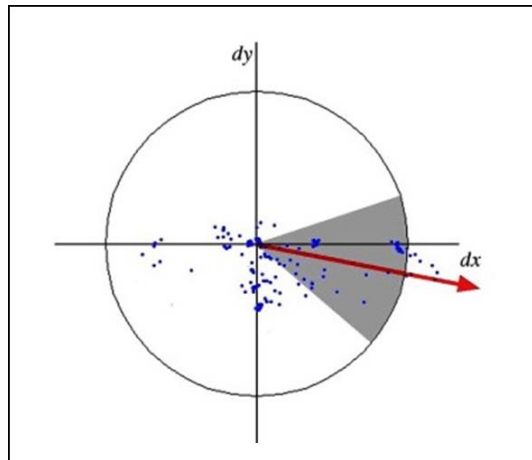
Fig.3.12 A Sliding Orientation Window Size $\frac{\Pi}{3}$ Detects the Dominant Orientation [70]

Both the vertical (dy) and horizontal (dx) responses within a window are summed up. Then,

the orientation of the interest point is defined by the longest vector over all windows.

(b) **Descriptor Based on the Summation of Haar Wavelet Responses**

The region around the interest point is described by a square region within an interest area

of window size of 20S (where S is the scale of the interest point at the point of detection).

The interest region is divided into 4 x 4 smaller sub-regions (Fig.3.13), as described by dx and dy wavelet response.
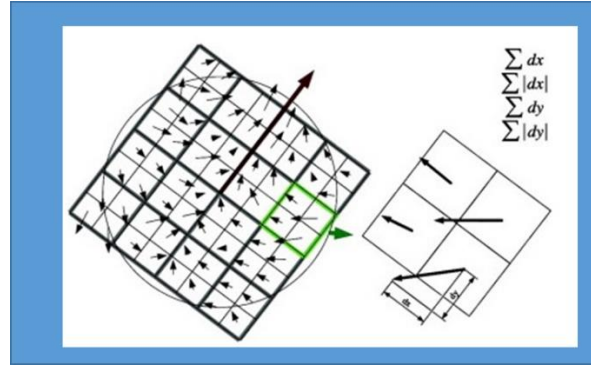
Fig.3.13 4 x 4 Square Sub-region is laid on the Interest Point on the Left [70].

For each sub-region a feature vector v is calculated based on 5 x 5 regular space sample points, using (3.26):

$$v = [\sum dx, \sum |dx|, \sum dy, \sum |dy|] \qquad (3.26)$$

To offer more robustness for noise, translation and deformation, the wavelet responses are weighted with a Gaussian.

3.2.3.3 SURF Feature Matching

Matching of features for object recognition and image matching with SURF algorithm uses the same principle as the SIFT algorithm. However, the use of integral image makes the computation of SURF faster than that of SIFT algorithm.

Also the vector size of SURF being smaller than that of SIFT, makes the SURF image matching operation faster. However, the SIFT has higher matching accuracy than SURF because the SIFT produces more feature points than SURF.

3.3 **Area Based Image Matching**

Area based matching techniques; unlike the feature based techniques make use of grey values as its matching entities. Grey values of neighbouring pixels are used rather than one pixel because of the uncertainty in matching one pixel. An image patch cut from one image known as the template, is used to search for a corresponding image patch in a sampled image. The template consists of m x n pixels and in most cases the m and n are equal, and are odd numbers.

In order to identify the matching area, the window template is compared with the sample image by sliding it from left to right, then, from up to down of the sample image. At each location of the template window in the sample image, a grey value is calculated and the location with the highest grey value is chosen as the highest matches.

The bigger the template, the more the requirement of matching entity is achieved. However, the geometric distortion caused by imaging instruments and change in orientation of images can affect image matching using big templates [72], [73]. For example, in sloppy areas the matching image patches are not geometrically identical. To solve this type of problem, the size of the template has to be small or its shape modified by geometric distortion (such as a trapezoid window). This approach is mostly applied in photogrammetry in the generation of a Digital Elevation model (DEM) [73].

One of the major tasks in image matching is avoidance of mismatches. In order to avoid outliers in area based image matching techniques, a threshold is set for similarity measures. Apart from the setting of a threshold for similarity check, geometric constraint and robust adjustment method are used for further computation to eliminate wrong matches. In area based matching, Correlation or Least Square methods are used for image matching.

3.3.1 Correlation Image Matching Techniques

Correlation matching techniques tends to find the similarity between two images by matching a patch of pixels in the first image, known as the template, and sliding it through the second image and by calculating and comparing the grey values at each area in the search image, the location with the highest value is selected as the best match. Thus, the normalised cross-correlation between a template image $(g_T(x_T, y_T))$ and a search image patch $(g_S(x_S, y_S))$, centred at the points (x_{T_1}, y_{T_1}) and (x_{S_1}, y_{S_1}) respectively, is given as:

$$\rho = \frac{\sum\limits_{m=-j}^{j} \sum\limits_{n=-k}^{k} [g_T(x_{T_1}+m, y_{T_1}+n) - \bar{g}_T] - [g_S(x_{S_1}+m, y_{S_1}+n) - \bar{g}_S]}{\sqrt{\sum\limits_{m=-j}^{j} \sum\limits_{n=-k}^{k} [g_T(x_{T_1}+m, y_{T_1}+n) - \bar{g}_T]^2 \sum\limits_{m=-j}^{j} \sum\limits_{n=-k}^{k} [g_S(x_{S_1}+m, y_{S_1}+n) - \bar{g}_S]^2}}$$

(3.26)

where:

$n_1 = 2j+1$ and $n_2 = 2k+1$ are the sizes of the template image, while

$$\bar{g}_T = \frac{\sum\limits_{m=-j}^{j} \sum\limits_{n=-k}^{k} [g_T (x_{T_1} + m, y_{T_1} + n)]}{n_1 \cdot n_2}$$ (3.27a)

and,

$$\bar{g}_S = \frac{\sum\limits_{m=-j}^{j} \sum\limits_{n=-k}^{k} [g_S (x_{S_1} + m, y_{S_1} + n)]}{n_1 \cdot n_2}$$ (3.27b)

are the average grey values in the template image and the search window of the right (sampled) image respectively. This technique is realised by comparing the average grey value in the template image to all $n_1 \times n_2$ pixels of a search window in the sample image. This comparison assigns a correlation coefficient value $\rho(x_{S_1}, y_{S_1})$ to each location (x_{S_1}, y_{S_1}) within the sample image search window. Using interpolation, the location of the template image, centred at a point in the search window of a sampled image where maximum value of the correlation coefficient is found, is selected as a match, provided that this maximum value of correlation coefficient does not exceed a certain predefied threshold.

To compute the coordinate of the pixel, where the maximum correlation function $\rho(x_S, y_S)$ is located, an estimate using second order polynomials in the x_S and y_S directions is used:

$$\rho(x_S) = a_0 + a_1 x_S + a_2 x_S^2$$ (3.28a)

71

$$\rho(y_S) = b_0 + b_1 y_S + b_2 y_S^2 \qquad (3.28b)$$

Then, the coordinates are determined as:

$$x_{S_m} = \frac{-a_1}{2a_2} \qquad (3.29a)$$

$$y_{S_m} = \frac{-b_1}{2b_2} \qquad (3.29b)$$

Other correlation matching approaches as proposed in [74] and [75] try to match the template window with the search window of a sampled image, by minimizing their differences. This is expressed by the normalized spatial root mean square deviation, given as:

$$m_d = \frac{1}{n_1 n_2} \sum_{m=-j}^{j} \sum_{n=-k}^{k} [[(g_T(x_{T_1} + m, y_{T_1} + n) - \bar{g}_T)] - (g_S(x_{S_1} + m, y_{S_1} + n) - \bar{g}_S)]]^2$$

$$(3.30)$$

Or by normalizing the absolute difference of the two windows, given as:

$$A_d = \frac{1}{n_1 n_2} \sum_{m=-j}^{j} \sum_{n=-k}^{k} \left| [g_T(x_{T_1} + m, y_{T_1} + n) - \bar{g}_T] - [g_S(x_{S_1} + m, y_{S_1} + n) - \bar{g}_S] \right|$$

$$(3.31)$$

72

The methods used in (3.30) and (3.31) are similar to the correlation coefficient method, the only difference here is that the window search aims in finding a local minima instead of local maxima.

3.3.2 **Least Squares Method**

Correlation coefficient method is very fast and efficient, if the images involved in the matching have the same geometric and radiometric properties. However, images that exhibit such conditions are very rare, especially when aerial images are involved, due to a lot of geometric distortions found in aerial images. This geometric distortion in aerial images may include, imaging sensor, change in height of the aerial platform, terrain slope, position and attitude differences. Illumination and reflectance conditions can also cause an image to be radio-metrically distorted.

In order to solve the matching problem caused by images that are geometrically and radio-metrically distorted, a Least Squares matching technique is employed. The least squares technique attempt to solve the problem of geometric and radio-metric distortion in a search image by minimizing the differences in the grey values between the template and search image window in an adjustment where corrections were made to the geometric and radio-metric distortion of the search image [75], [76].

Using a stereo pair of images as an example, if the left image $L_T(x, y)$ is made the template and the image on the right $R_S(x, y)$ as the sampled image. Then, ideally the correlation is established if:

$$L_T(x, y) = R_S(x, y) \qquad (3.32)$$

However, the presence of noise in both images or noise in only the search image, makes (3.32) unreliable. Hence, a noise vector is added, then (3.32) becomes:

$$L_T(x, y) - \ell(x, y) = R_S(x, y) \qquad (3.33)$$

ℓ is a noise vector caused by a geometric and radio-metric distortion to the images, in an ideal situation as in (3.32), the $\ell = 0$.

In order to find the match point, the position of the function values $R_S(x, y)$ must be determined. This is done by minimizing a function which measures the differences in grey values between the template image and the search image. The goal is to determine the geometric and radio-metric transformation parameters of the search window, such that vector ℓ is minimized. The least squares equation in (3.33) is a non-linear adjustment problem, this needs to be linearized and the location of the template is estimated.

The template location in the search image is described by a shift parameters $(\Delta x, \Delta y)$ with respect to an initial position of $R(x, y)$ and the estimation of the corresponding window $\ell(x, y)$ in the search image area $R^0(x, y)$. Apart from the shift parameters, the image shaping parameters and radiometric corrections must be introduced, in order to account for distortions caused by systematic image deformations [76].

If for example, $L_T(x_T, y_T)$ is the template of n_1 x n_2 pixels in the left image and $R_S^0(x_S^0, y_S^0)$ is the equal size estimation of the template location in the right image, the goal is to estimate a new location for the search image window in the right image $R_S(x_S, y_S)$, such that, the difference in the grey values will be minimized in a least squares approach. The estimation is achieved through a perspective transformation of the coordinates $R^0(x_S^0, y_S^0)$ and resampling of the corresponding grey values. The geometrical transformation relates the two image patches by a bivariate polynomial given by:

$$x_S = t_y^T A_a t_x \qquad (3.34a)$$

$$y_S = t_y^T B_b t_x \qquad (3.34b)$$

where:

$$t_x^T = [1, x_0, x_0^2, ... x_0^{m-1}]$$

$$t_y^T = [1, y_0, y_0^2, ... y_0^{m-1}]$$

(A_a, B_b) = the unknown m x m parameter matrix

(x_0, y_0) = the grid location of the sample image points of $R^0(x, y)$

With respect to very small size of the templates to be matched, the bivariate polynomial is usually substituted with a 6-parameters affine transformation:

$$x_S = \begin{bmatrix} 1 & y_0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \end{bmatrix} = a_{11} + a_{12}x_0 + a_{21}y_0 \qquad (3.35a)$$

and,

$$y_S = \begin{bmatrix} 1 & y_0 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \end{bmatrix} = b_{11} + b_{12}x_0 + b_{21}y_0 \qquad (3.35b)$$

The function $R_S(x, y)$ in (3.33) need to be linearized, in order to operate with the traditional least squares approach, while the transformation parameters in (3.35) need to be estimated. Hence,

$$L_T(x, y) - \ell(x, y) = R_S^0(x, y) + \frac{\delta R^0(x, y)}{\delta x} dx + \frac{\delta R^0(x, y)}{\delta x} dy \qquad (3.36)$$

where:

$$dx = \frac{\delta x}{\delta p_i} dp_i$$

$$dy = \frac{\delta x}{\delta p_i} dp_i$$

$$p_i = \text{ith transformation parameter in (3.34)}$$

Differentiating (3.35) gives,

$$dx = da_{11} + x_0 da_{12} + y_0 da_{21} \qquad (3.37a)$$

$$dy = db_{11} + x_0 db_{12} + y_0 db_{21} \qquad (3.37b)$$

76

From (3.36) and (3.37), with the addition of the shift parameter (r), results into the following observation equations:

$$L_T(x, y) - \ell(x, y) =$$
$$R^0(x, y) + R_x da_{11} + R_x x_0 da_{12} + R_x y_0 da_{21} + R_y db_{11} + R_y x_0 db_{12} + R_y y_0 db_{21} + r_2$$

$$(3.38)$$

where:

$$R_x = \frac{\delta R^0(x, y)}{\delta x}, \text{ and } R_y = \frac{\delta R^0(x, y)}{\delta y}$$

The unknown vector x becomes:

$$x^T = [da_{11}, da_{12}, da_{21}, db_{11}, db_{12}, db_{21}, r_S] \qquad (3.39)$$

For every pair of pixels from the left template image $L_T(x, y)$ and the right search image window $R_S(x, y)$, one observation equation is formed.

In matrix form, we have,

$$-\ell(x, y) = AX - L \qquad (3.40)$$

where each element of the vector L is in the form:

$$L = L_T(x_T, y_T) - R_S^0(x_S^0, y_S^0) \qquad (3.41)$$

And each row of matrix A, is:

$$A = [R_{S_x}, R_{S_x} xL_T, R_{S_x} yL_T, R_{S_y}, R_{S_y} xL_T, R_{S_y} yL_T] \quad (3.42)$$

Thus, the least squares solution vector is:

$$\hat{X} = (A^T PA)^{-1} A^T PL \quad (3.43)$$

where P is the weight matrix, which is usually a diagonal or identity matrix.

The parameters obtained through the least squares solution are used in updating the coordinates of the template image and to resample grey values, as a new right image window $R_S'(x_S', y_S')$ is selected as a match of a fixed left image template $L_T(x_T, y_T)$.

This means, each pixel (x_S', y_S') of this new window corresponds to a pixel (x_{L_T}, y_{L_T}) of the template according to the following transformation:

$$x_{R_S}' = (a_{11}^0 + da_{11}) + (a_{12}^0 + da_{12})xL_T (a_{13}^0 + da_{13}) yL_T \quad (3.44a)$$

$$y_{R_S}' = (b_{11}^0 + db_{11}) + (b_{12}^0 + db_{12})xL_T (b_{13}^0 + db_{13}) yL_T \quad (3.44b)$$

The final solution obtained through iterations does not completely remove the difference between the template window and its conjugate in the search image. This is because minor differences can be introduced by terrain slope, and change in camera orientation. The least squares approach tries to minimized these differences, which cannot be completely eliminated. The highest possible pixel coordinate difference between the initial estimation and the absolute solution that can be used in a least squares technique is referred to as pull-

range [77]. According to [64], the pull range values depend on the size of the window and are usually such that the matching windows conjugate has at least 50% overlap.

By introducing constraints, for intersecting conjugate rays, in addition to geometric and radiometric constraints, multiple image grid points can be simultaneously matched and hence, performing both image matching and object space coordinate determination [76], [77].

3.4 **Working in the Spatial and Frequency Domain**

Filtering in the frequency domain is much more efficient in convolving images with large kernels. It is also useful in the reverse process of deconvolution, which is when the output image and the Point Spread Function (PSF) (also known as the impulse response) are known, the input image can be reconstructed. In another words, the forward convolution may be processed by a multiplication of the frequency components and the PSF as in (2.4). While the deconvolution can be processed by dividing the frequency components of the output image with the corresponding PSF in the frequency domain.

During the reverse process of deconvolution, the frequency components of the PSF must be handled with care, so as not to blow-up noise from the input image into large artifacts in the reconstructed image.

Frequency domain filtering ignore the present of any object regions in the image and process the whole image signal at once. This makes filtering in the frequency domain robust to applications such as, smoothing, object recognition and image matching. In these cases, the frequency filtering is better than filtering in the spatial domain.

In contrast, applications that involves segmenting objects in the image from its background in order to recognise object or region of interest are better done in spatial domain than in the frequency domain.

3.5 **Chapter Summary**

This chapter presented the state of art of the most common image matching and image object detection techniques, which includes feature based and area based matching and detection techniques.

The feature based matching techniques are frequently used, when the information needed is mostly the local structure of the image rather than the image intensities. To test for image similarity, features in a referenced image are tested with the features of the second image. Image feature can be in the form of points, lines or regions in the image. With respect to feature points, the search for unique image points and the matching of images consists of three major processes: interest points selection, descriptor vector formation and descriptor vector matching. The interest point selection is the process of selecting locations in the image where there are repeatability of points. Afterwards, a descriptor vector is formed by the pixel around the interest point. Finally the descriptor vector, which must be unique and invariant to image deformations are matched with a referenced descriptor vector to determine similarity. The dimension of the descriptor of the two matching images determines the time it takes for matching the images. A descriptor vector with a smaller dimension takes less time for matching when compared with a descriptor vector with a larger dimension.

The earliest algorithm that uses feature based techniques is Harris corner detection. The Harris corner uses a mathematical formula to determine the window that produces large variations when moved in all direction on the image. Each of the windows is assigned a score. One can figure out which of the windows consists of a corner based on the score. The Harris corner detection algorithm not being invariant to scale, gives rise to the development of SIFT and SURF algorithms.

The SIFT algorithm, generates keypoints from images and these keypoints are used to compute descriptor vectors through octaves, which are used to generate difference of Gaussian.

The SURF is a speed up version of the SIFT algorithm and it makes use of integral image for the computation of a feature vector. SURF is faster in operation than the SIFT algorithm because of the smaller dimension of its descriptor vector and the use of integral image for the computation of its feature vector.

The second most commonly used matching and object recognition technique is the area based method. Area based matching makes use of grey level values for matching entities to measure the degree of image matching. To identify matching area, a window template of an image is slid in all direction through the sampled image. At each location of the template window on the sampled image, a grey value is computed and the location with the highest grey value is selected as the highest matches. Algorithms that make use of area based matching, includes correlation coefficient and least squares methods. The former is usually employed when images involved in the matching have the same geometric and radiometric properties. While the least squares method is employed where matching images are geometrically and radiometrically distorted.

CHAPTER FOUR

Detection of Objects in Video Streams from a Low-cost UAV

4.1 **Introduction**

A composite correlation filter is used for detection of objects of interest in the video streams acquired from a low-cost UAV in this thesis. A correlation filter was chosen for object detection because of its ability to handle general types of distortion. Also, since it is a Correlation Pattern Recognition (CPR) filter, it has the robust property of evaluating the whole input signal at once, unlike the feature-based techniques, which minutely extract information from piecewise examination of the signal and compare the relationships between the features. Matching the whole image against the template makes the CPR less sensitive to small mismatches. Composite correlation filters are designed from many training images, which represent the different views of the objects to be detected. The filter can be trained to detect any object with any kind of distortion as long as the expected variations can be captured by the training images. The major aim of all composite filters in this thesis is to be able to detect the object on which they were trained. To obtain a cross-correlation as a function of the relative shift between a query image and a set of training image templates, the query image is compared to the template. The whole operation is computed in the spatial-frequency domain for computational efficiency by performing the complex multiply:

$$Y(a,b) = X(a,b)*H(a,b) \qquad (4.1)$$

where X(a,b) is the 2D-Discrete Fourier Transform (DFT) of the query image, H(a,b) is the spectrum of the reference template, * denotes the complex conjugate of the filter spectrum and Y(a,b) is the DFT of the correlation output . A Fast Fourier Transform (FFT) [78] algorithm is used to implement the DFT efficiently. Whenever there is a match between the

query image and the template image, the correlation filters are designed to give a sharp peak at the centre of the correlation output plane and no peak if there is no match between the query image and the template image, as illustrated in Fig. 4.1.
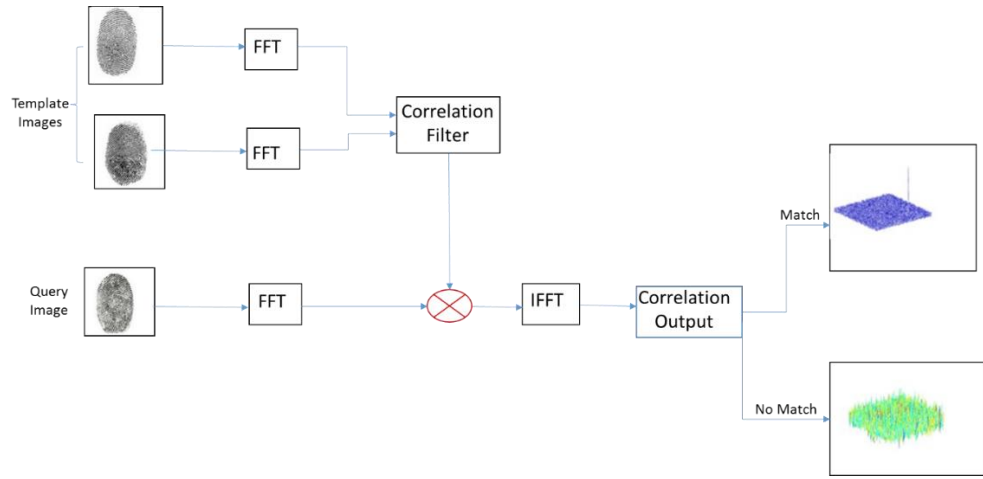
Fig.4.1 Matching (correlating) a query image with the template (correlation filter)

4.2 **Composite Correlation Filter Design**

The first composite correlation filter developed, was the Projection Synthetic Discriminant Function filter (PSDF) [79]. The PSDF filter is based on the assumption that the filter is a weighted sum of training images and the weights are determined so that the correlation output takes on pre-defined values in response to training images at the origin. However, PSDF filters suffer from not having any in-built robustness to noise and display large side lobes, making location of the correlation peak difficult. Hence synthetic discriminant function (SDF) filters are based on training images that contain examples of expected distortions. The Minimum Variance SDF (MVSDF) [79], [80] was then developed to

minimise the Output Noise Variance (ONV) in the projection SDF. The MVSDF is able to improve correlation peak height variations but cannot suppress white noise present in the correlation output plane. In order to solve the problem of white noise and high correlation peak side lobes generated by the earlier composite correlation filters, the Minimum Average Correlation Energy (MACE) [80] filter was developed. The. MACE filter is able to suppress the side lobes by minimizing the energy in the correlation plane. With the reduction of energy in the correlation plane, sharp correlation peaks were produce but it is still not as robust to noise as the MVSDF. It was discovered by [80] and [81] that both attributes of MACE and MVSDF can be integrated into a single filter by providing an optimal trade-off between the MACE and the MVSDF. To illustrate the optimal trade-off between these filters, let us consider the design of a filter that uses two performance criteria, the average correlation energy (ACE) and output noise variance (ONV), in order to satisfy a set of linear constraints. Since it is not possible to minimise both performance criteria ONV is minimized for every possible choice of ACE. This is illustrated in the following Lagrangian equation:

$$\phi(\rho) = ONV - \rho(ACE) - h^{+}X\delta \quad (4.2)$$

where ρ is a single Lagrange multiplier that forces ACE to a fixed value, δ is the vector of M Lagrange multipliers, which corresponds to the linear constraints on the correlation peaks in response to M training images. It can be seen that when ACE is fixed to any value, minimising ONV will also minimize $\phi(\rho)$.

By defining $\rho = \dfrac{(\mu - 1)}{\mu} \in [0,1]$, where μ is the mean peak value of the constrained filter.

The parameter μ offers the way to optimally trade the properties of the correlation filter to obtain a suitable compromise between the two performance criteria.

Equation (4.2) becomes:

$$\phi(\mu) = \mu\phi(\rho) = \mu ONV + (1 - \mu)ACE - h^+ X_q \qquad (4.3)$$

where $q = \mu\delta$.

Hence, the weighted linear combination of ACE and ONV indicates the performance criterion to be minimised. Thus, as μ varies from 0 to 1.0, the emphasis sifts from minimizing ACE to minimizing ONV.

This approach can be applied to more than two performance criteria and to other unconstrained correlation filters like the MACH filter [82]. When the optimal trade-off performance criterion is applied to the MACH filter, it takes the form:

$$h = (\alpha ACE + \beta ASM + \gamma ONV)^{-1} m \qquad (4.4)$$

where α, β and γ are the optimal trade-off parameters, which are associated with the performance criteria ACE, Average Similarity Measures (ASM) and Output Noise Variance (ONV), respectively. Each parameter is varied while all others are held constant until a satisfactory value is found. The composite correlation filter, used in this research, is the Optimal Trade-off Maximum Average Correlation Height (OT-MACH) filter [83]. The OT-MACH filter is able to obtain the best balance between the filter's variance to noise,

sharpness of the correlation peak, and distortion variance. A detailed literature on correlation filter design is give in [79], [80], [81], [82] and [83].

4.3 Experimental Procedure

For detection of objects of interest in a video stream from a low-cost UAV, an experiment was carried out both in United Kingdom and Nigeria. Where video data were collected for approximately 45 minutes from a low-cost UAV and transmitted through a radio link to a portable PC on the ground control station in real time.

The UAV used in the experimental work reported in this thesis is the Spreading Wings S800, which is a product of Djibouti Dow Jones (DJI) Industry Ltd, as shown in Fig.4.1. It is a lightweight, multi-functional hexa-rotor integrated aircraft. Table 4.1, shows the main specification of the DJI S800. The UAV consists of integrated GPS/INS for position, altitude control and a 7.5R/C flight simulator, which is used for flying the UAV from the ground control station. The UAV supports up to 50 way-points in the flight plan.

Fig. 4.2 Spreading Wings S800 [83]

The GPS/INS has been integrated on board into the center frame. This allows raw images captured to be linked to the exact time of acquisition of images. The DJI S800 gives real-time flight data and video feeds with a 5.8G video downlink via a radio signal.

Diagonal Wheelbase	800mm
Frame Arm Length	350mm
Center Frame Diameter	240mm
Bi-Pod Size	500mm(Length)×415mm(Width)×320mm(Height) width: 145mm
Max Power	360W
Signal Frequency	30Hz – 450Hz
Total weight	2.6Kg

Table 4.1 DJI Spreading Wings S800 Specification [83]

Tables 4.2 and 4.3 show the specification of the sensors and camera employed.

Size	17.3 x 13 mm
Diagonal	21.64 mm
Surface Area	224.9 mm^2
Pixel Pitch	3.74 μm
Pixel Area	13.99 μm^2
Pixel Density	MP/cm^2

Table 4.2 Sensors Specification [83]

Resolution	17.20 Megapixels
Sensor Size	Four Thirds (17.3 x 13 mm)
Sensor Resolution	4620 x 3474
Crop Factor	2
Video Recording	1920 x 1080p

Table 4.3 Camera Specifications [83]

The procedural steps for object detection in a video stream are as follows:

- **Video Streams Resample**: - The first step in object detection in the video stream captured by the UAV, is the resampling of the video streams into frames at the rate of one frame per second of sizes 256 x 256. This is done with the use of the BPS video converter 1.4 software at the ground station. Then each image in the video frame is

processed individually with the goal of detecting objects of interest in the video frames with a very small number of false detections.

- **Training of Correlation Filter**: **-** The next step in object detection from the video frames, is the training of the correlation (OT-MACH) filter. This is done by carefully cropping images of objects of interest to a size of 64 x 64 and training them on multiple sample targets to form template images. These template (filter) images are trained with the aid of a computer aided design (CAD) model, under different viewing angles (0^0 – 360^0), different lighting conditions, and different image deformation that is expected from the target image in the video frame. This makes the correlation filter to be robust to different kinds of distortion that is expected from the target image on the video frame. The trained images used as examples are classified into two sets, the first set consists of a person CAD model, while the second set is a car CAD model. The two sets of training images are trained under different UAV Camera viewing angles. Afterwards, all the sets of trained images are used to synthesis the filter and then stored in a database.

- **Cross Correlation**: - The final step in object detection is the cross correlation between the test (query) image and the training image. To achieve cross correlation, a test image is matched by individually comparing it with each of the trained images in the database and finding candidate match images. Whenever the cross correlation output plane gives a sharp peak, it implies there is a match between the input image and one of the trained images in a class. Hence an object has been detected.

4.4 **Test Results and Discussion**

Fig. 4.3 and Fig. 4.4 shows the example results of the two sets of trained images, which have been derived from CAD models of a selected person and car respectively.

These trained sets are derived from different viewing angles (0^0 – 360^0) of the UAV's camera.

The correlation output peak, which is the result of the cross correlation is quantified by determining the peak-to-side lobe ratio (PSR). This is done by comparing the correlation peak height value with the peak value of the side lobes. This is given as:

$$PSR = \frac{Peak - \mu}{\delta}$$

(4.5)

where the peak is the largest value in the correlation output plane and, μ and δ are the average values and standard deviation respectively of the correlation values of some side lobes surrounding the peak.
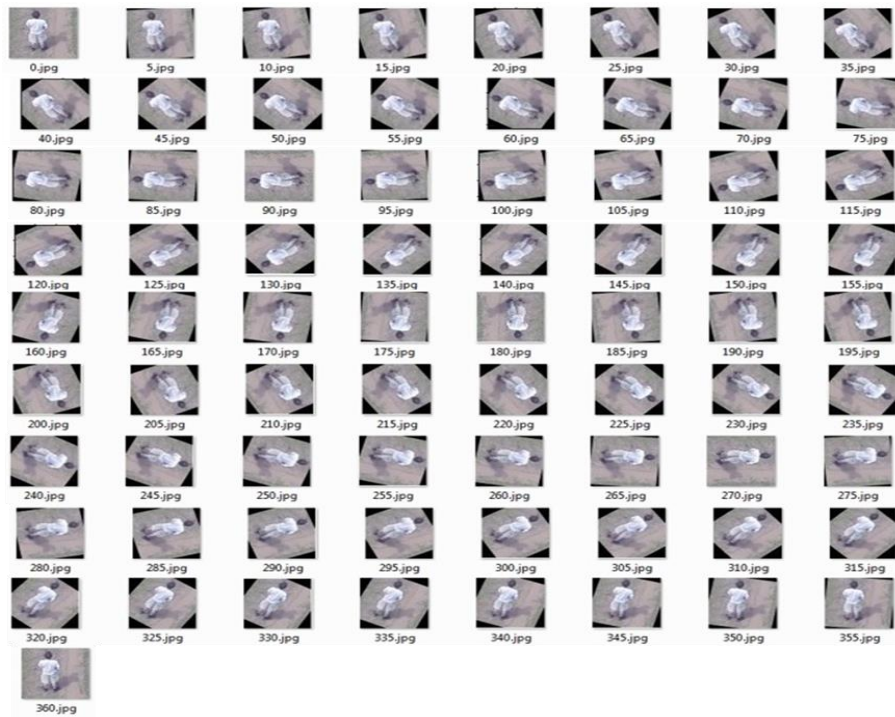
Fig. 4.3 Trained Images from Human CAD Model
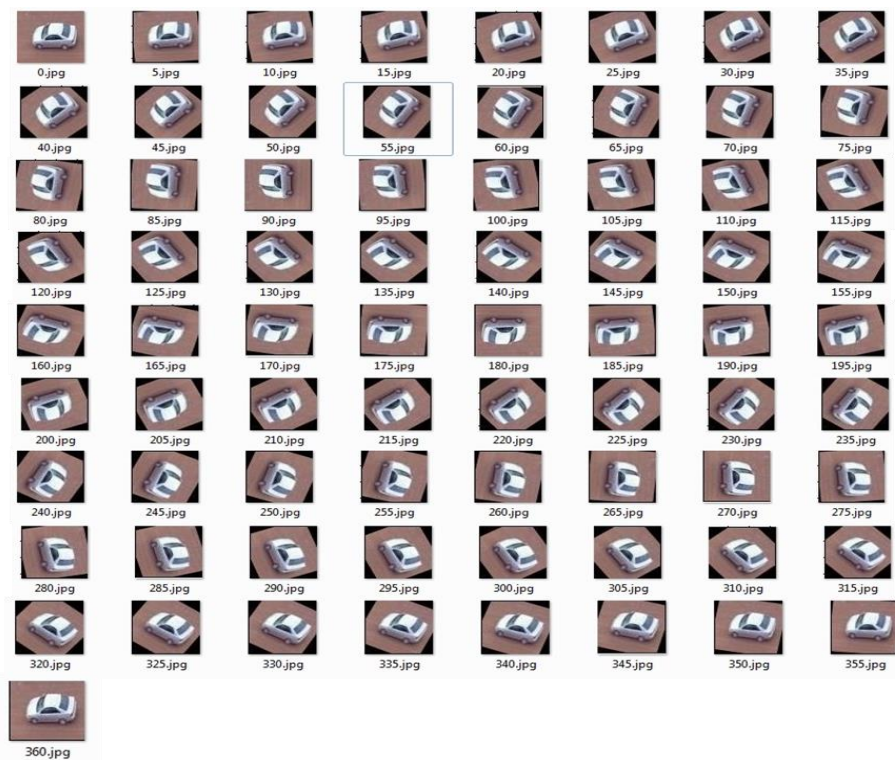
Fig. 4.4 Trained Images from Car CAD Model

PSR is designed to measure the relative height of the correlation peak with respect to the background. One advantage of PSR is that, it is invariant to constant change in illumination in the test image. PSRs are found to be larger for authentic test images and smaller for impostor images. A Fig. 4.5 and Fig. 4.6 show the resulting correlation peaks when the test image used is an impostor and is authentic respectively.
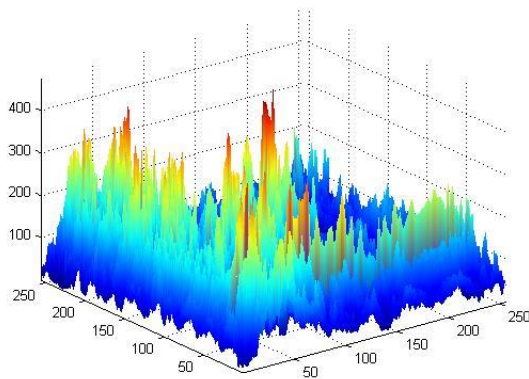
Fig. 4.5 correlation output for an imposter test image, the PSR value is found to be 8.12
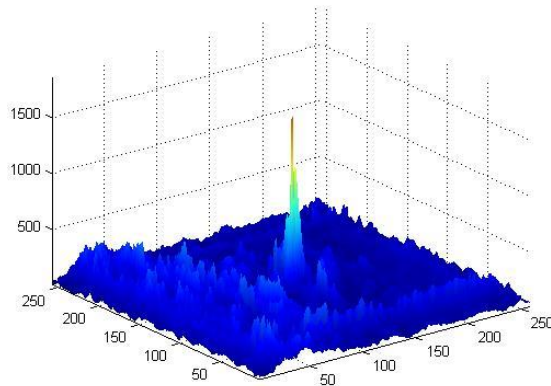
Fig. 4.6 correlation output for an authentic test image, the PSR value is found to be 28.51

In this experiment the OT-MACH filter parameters α, β and γ have been fixed to 0.0000009, 0.42 and 0.1 respectively. These values have been found to be near optimal for all the test images. However better results can be obtained by adjusting the parameters between correlation test operations.

Multiple example tests were conducted using different viewing angles of the UAV's camera of the target scenes. The results obtained are as follows:

(a)
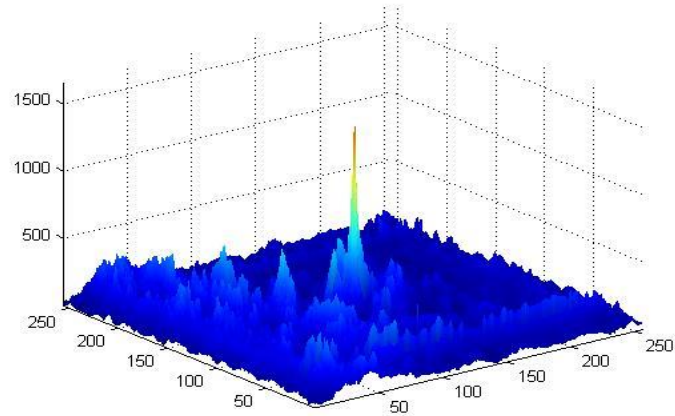
(b)
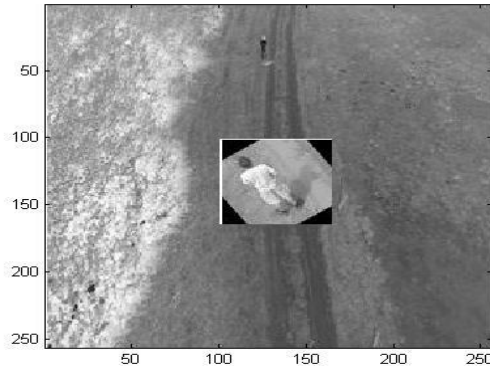
(c)

Fig.4.7 (a) Video frame (256x256) containing the target image (b) Correlation output when target (person) is view by the camera at an angle of 35^0 (b) Result image showing the corresponding (64 x 64) trained image.

The PSR value in the correlation output plane in Fig. 4.7 was found to be 28.12.
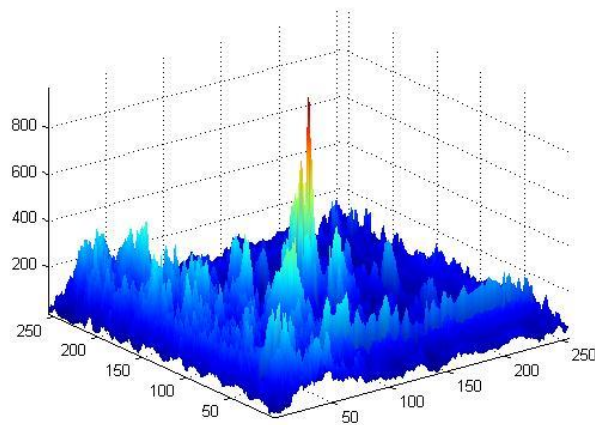
(a)
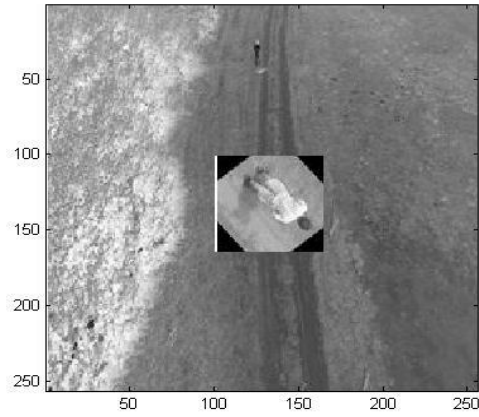
(b)

Fig. 4.8 (a) Correlation output when target (person) is view by the camera at an angle of 320^0 (b)

Result image showing the corresponding (64 x 64) trained image.

The PSR value in the correlation output plane in Fig. 4.8 was found to be 17.28 while the

correlation peak is 625.40. It can be observed that the PSR in Fig. 4.7 is larger than that of

Fig.4.8, this implies that the degree of similarity of object detected and the trained image is

larger in the test image in Fig. 4.7.

The test was also carried out using a car CAD model under different illumination condition,

using the same parameters and the results obtained are:

(a)
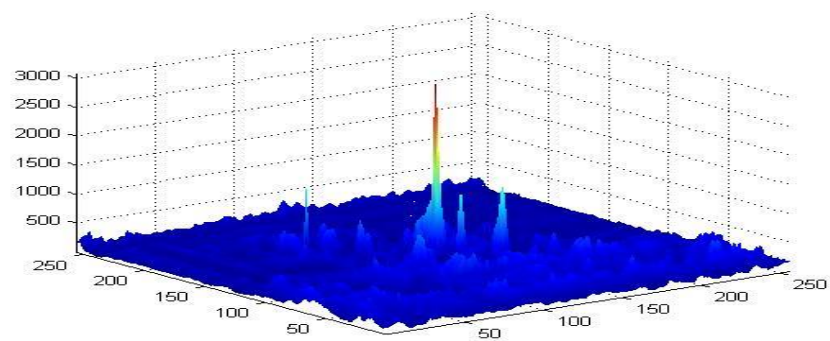
(b)
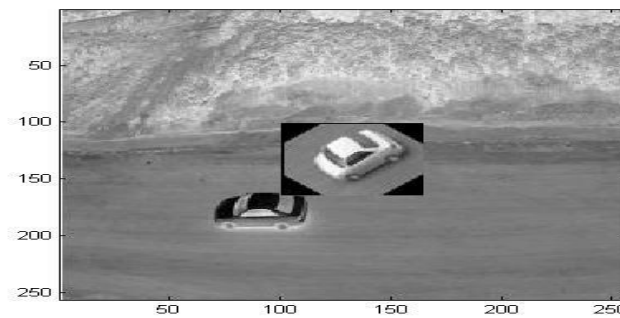
(c)

Fig. 4.9 (a) Video frame (256x256) containing the target image (b) Correlation output when target

(car) is viewed by the camera at an angle of 35^0 (b) Result image showing the corresponding (64 x
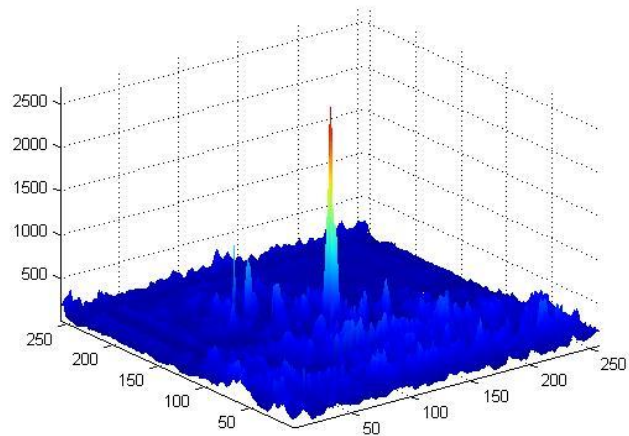
64) trained image.

(a)
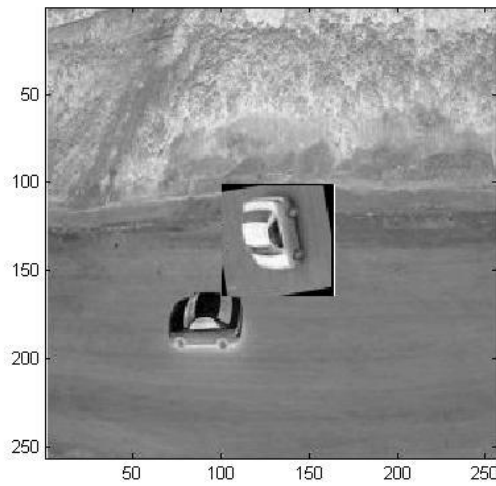
(b)

Fig. 4.10 (a) Correlation output when target (car) is view by the camera at an angle of 100^0 (b)

Result image showing the corresponding (64 x 64) trained image.

4.5 **Chapter Summary**

This chapter presents a brief overview of a composite correlation filter. The major aim of the correlation filters is to detect any object with any kind of distortion as long as these distortions can be captured by a set of training images. Objects are detected by matching a query image with a set of trained image templates. The correlation process is computed in the spatial frequency domain for fast and effective computation. A sharp peak at the centre of the correlation output plane, indicates a match between the query and a template image.

In order to get the best balance between the SDF filters variance to noise and the MVSDF filter's sharp correlation peak constraints, an OT-MACH filter is used for object detection.

In order to evaluate our method, experiments were conducted on a test field both in Nigeria and United Kingdom. Where video stream data were captured by the use of UAV. The data obtained were evaluated using MatLab/Simulink software and gave promising results in terms of object detection and recognition.

CHAPTER FIVE

PHOTOGRAMMETRY MODELLING

5.1 Photogrammetry

Photogrammetry is the technology that enables us to obtain reliable information about objects and their locations on the Earth's surface through the process of capturing, measuring and interpreting photographic images [84]. Photogrammetry was formally used for architectural survey [85], [86], but now it is widely used for remote sensing technology, it was the first methodology to make use of the geometric properties of objects from photographic images.

The development of photogrammetry consisted of three main phases. The mode of operation and type of equipment used to carry out measurements distinguished one phase from another and it took a many years for transition from one phase to the next phase.

In the first phase of photogrammetry, recording of images is carried-out with the aid of photographic film. Orientation reconstruction and other measurements are carried-out in an optical-mechanical way. Hence this phase is called Analogue Photogrammetry [86], [87].

In the second phase of photogrammetry, with the advent of computers, the reconstruction of the orientation process change from analogue to algorithmic, where formulas and their parameters are evaluated and stored in the computer. This phase makes use of an analytical plotter for photographic support, hence the name Analytical Photogrammetry [87].

The latest and current phase is the digital photogrammetry phase [88], this phase make use of digital images that are captured by digital cameras, which are then stored and processed on a computer system. With digital photogrammetry, many tasks are now automated (e.g., digital ortho-image formation and digital elevation model extraction). The output product

of digital photogrammetry are in digital form and hence they can easily be stored, managed and used for different applications.

In reality, photogrammetry gives correct and exact environmental information from a wide range of images. Measurements taken using the photogrammetric technique on images or photographs reflect measurements taken on the ground. Instead of having to make constant visits to the field to make measurements of distance and object's coordinate position on the Earth's surface. The photogrammetry technique permits the accurate gathering of information from images. This approach of collecting environmental information from images saves time and cost, and gives a high level of accuracy.

From digital photogrammetry that is based on digital recording instruments, the following processes can be distinguished:

- Frame Acquisition – This is the process of image acquisition in which the central perspective defines the geometrical link between the two dimensional image plane and the object in the 3D world.

- Whiskbroom and Push-broom Scanning Systems – This process is more complex when compared with the traditional central perspective, it consists of multiple central perspectives [88]. This process is time consuming and expensive.

The frame acquisition imaging process involves generating images through single central perspective geometry. As shown in Fig.5.1, the object points (P, Q R and S) generate the image points (p, q, r, and s) through a single perspective centre (O). The single central perspective method is used in this research and unlike the scanning systems, it is not expensive and uses less time.
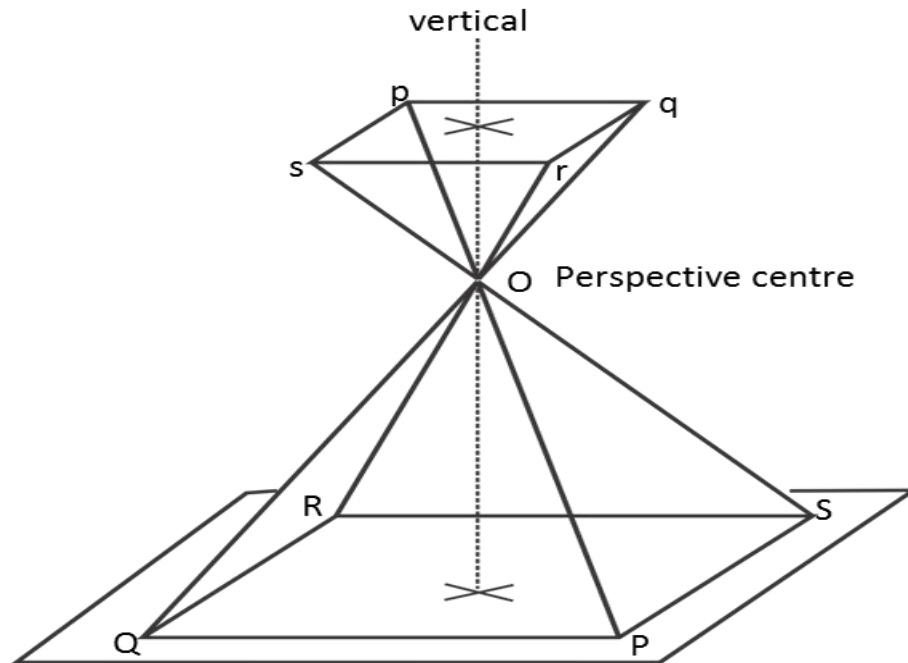
Fig. 5.1 Central perspective linking points on the image plane and points on the ground [88].

5.1.1 Types of Photogrammetric Processes

There are two types of photogrammetric process: the terrestrial photogrammetry and the aerial or aero-photogrammetry. These two processes are defined as follows:

- Terrestrial photogrammetry – This refers to acquisition of images by cameras positioned at ground level, in which objects are also located on the Earth's surface (such as landslide monitoring, building surveys, etc.).

- Aerial or aero-photogrammetry – This is where acquisition of images is carried out from above sea level, in which the camera is attached to an aircraft and the object is on the ground. Present day maps are derived from aero-photogrammetry survey.

5.1.2 Stereoscopic Viewing

From a single image frame, which consists of a two dimensional plane, we can only get two dimensional co-ordinates. In order to get three dimension views, photogrammetry uses a method similar to the principle of human vision [89].

We are able to see objects in the 3D-world in which we are in, because our two eyes are able to receive optical information as a central perspective between images captured by both eyes. The image captured by the left eye is slightly different from the image captured by the right eye. These two images are then combined in our brain to produce a spatial impression. This is the process that enables us to estimate the distance between us and an object. When this same principle is applied in photogrammetry to get the three dimensional information it is called stereoscopic viewing [90]. To illustrate stereoscopic viewing, consider the case of aerial photogrammetry in Fig.5.2:
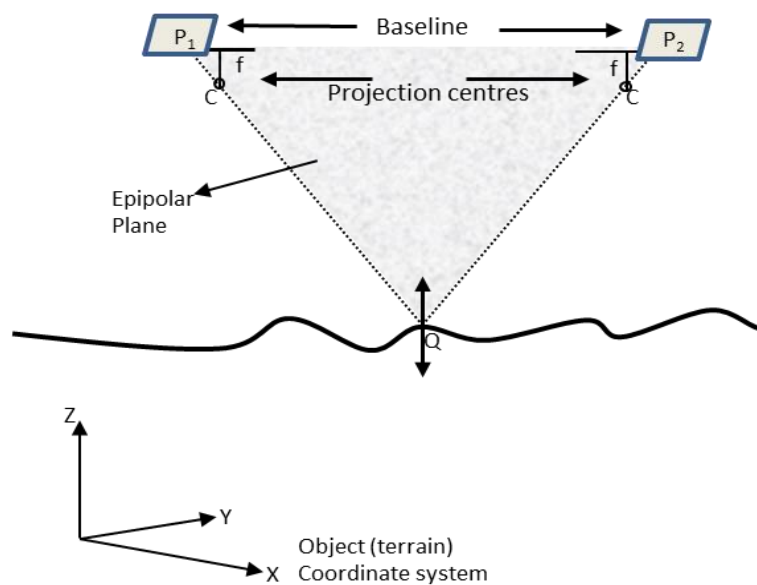
Fig.5.2 Example of Stereoscopic Viewing with two Different Camera Positions [88].

where P_1 and P_2 are the camera positions on the left and right respectively, C is the corresponding projection center for each camera position, the ray from the center of projection of the left camera position to the center of projection of the right camera position is called the base or baseline and Q is the image point in the 3D world scene.

If we are able to reconstruct the geometric parameters of the condition of the camera during exposure, we can then determine the height of an object point in the 3D world scene. This can be done by setting up the two equations of the rays, and then calculating their intersection.

To implement stereoscopy, the following rules must be obeyed in aerial imaging:

- Each image used for stereoscopy must have at least 50% view of the same terrain features. But a 50% view means, no room for error from the aerial images. Hence a 60% forward overlap and a 20-40% side overlap is the minimum requirement.

- The distance between exposure stations must be constant, so that all images will have approximately the same scale.

- For aerial imaging consisting of more than one flight strip, the flight strips must be close and parallel to each other as much as possible.

- The distance between the two exposure stations must not be too great in order not to lose stereoscopic coverage.

In practical terms, stereoscopy between two images can only be provided by controlling the horizontal and vertical parallaxes in tie points on two images. The difference in the location of points in the x-direction or in the direction of the flight is called x-parallax. The x-parallax adjustment played a major role in the determination of the elevation of a point.

The y-parallax is the major factor affecting the formation of stereoscopic viewing, this means, with the present of y-parallax in imaging, stereoscopic viewing is not possible.

The y-parallax is caused by the difference in the location of tie points in the direction perpendicular to the flight direction. The major cause of y-parallax is a tilt in the photograph. An essential task that needs to be done by a photogrammetrist is the elimination of y-parallax within a model.

5.2. Aerial Image and Data Acquisition

During aerial image acquisition, overlapping images are captured along the UAV flight direction. Each point along the flight direction where the camera captures an image is known as the exposure station (Fig.5.3).
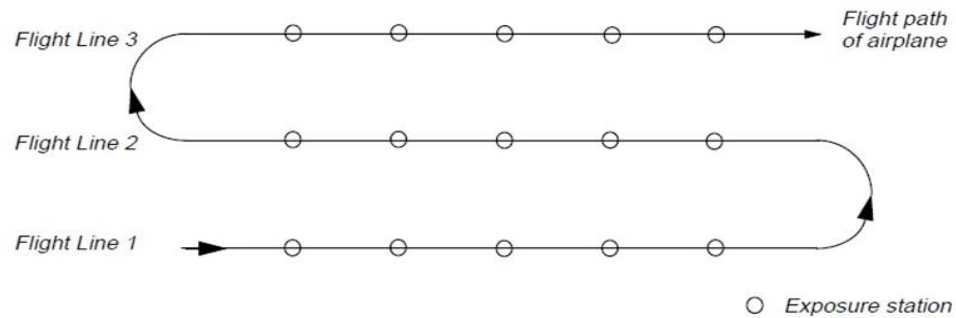
Fig.5.3 Exposure stations along a flight strip [84].

Images captured along the same flight line constitute a strip of images or photographs. In Fig.5.3, all the exposure stations in the flight line 1 for example, constitute a flight strip while the images captured by all the expose stations in flight line 2 constitute another

flight strip. All images in a flight strip are taken at approximately the same altitude and the distance between exposure stations is kept constant.

Images from several flight strips of the same UAV as a collective form a block of images, usually with a forward overlap of 60% and side overlap of between 20–30% (Fig.5.4). The use of more than one image in photogrammetry, gives the geometry related to the exposure stations, the image points on the image plane and the corresponding object points on the ground, high accuracies and precision.
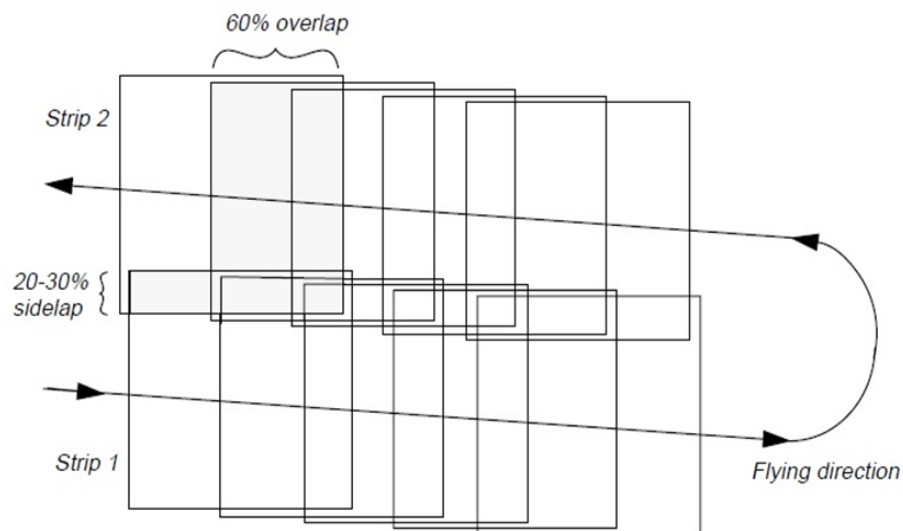
Fig.5.4 Strips of Images Forming a Block of Aerial Images [84].

5.3   Co-ordinate Systems

In order to understand the connection between the camera/sensor used to capture the object point on the ground, the object point itself and its corresponding image point, (which is the main task of photogrammetry), the coordinate system associated with these three variables must be defined.

5.3.1 Pixel Co-ordinate System

Pixel co-ordinate system is used to define file coordinates of the digital image. This pixel coordinate system usually has the origin of its coordinate system at the left upper corner of the image, its x-axis points to the right while its y-axis points downward, as shown by axis c and r in Fig.5.5. These file coordinates (c, r) represent the pixel column and row number respectively and is referenced as the pixel coordinates in this thesis.
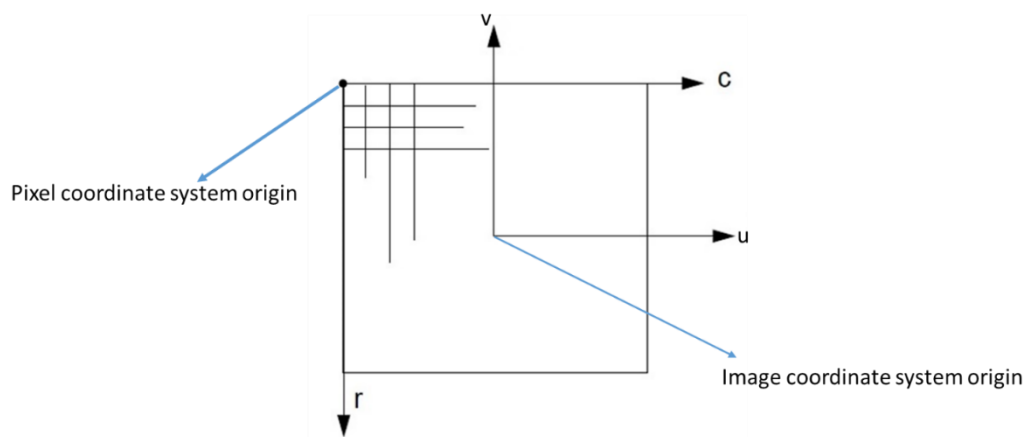
Fig.5.5 Pixel and Image Coordinates Systems

5.3.2 Image Plane Coordinate System

An image plane coordinate system is the 2D coordinate system which has its origin in the image center or at the intersection of the fiducial marks as shown by the u and v axis in Fig.5.5. Under perfect conditions, the origin is normally at the principal point. Image plane coordinates are used to describe image point positions on the film plane. The image coordinates are usually measured in millimetres or microns. The image plane coordinates are referenced as (u, v) in this thesis.

5.3.3 Image Space Coordinate System

An image space coordinate system unlike the image coordinate system adds a third (w) axis, making it a three-dimensional axis. The perspective center (C) as shown in Fig.5.6, defines the origin of the image space coordinate system.
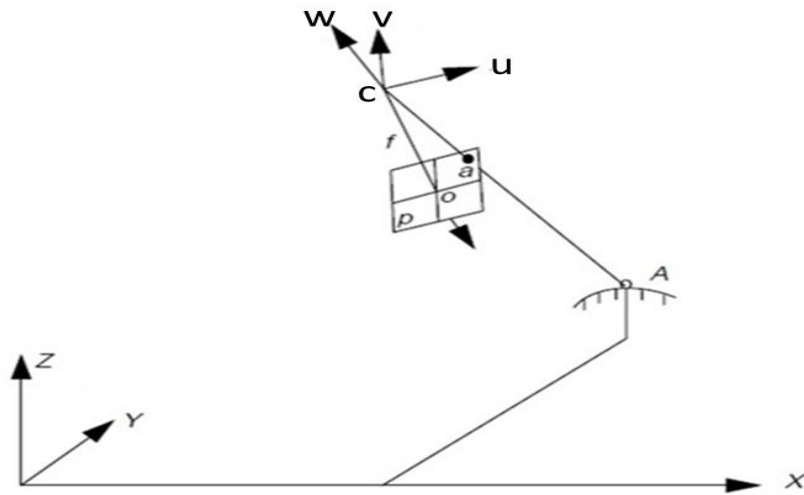
Fig.5.6 Image Space and Ground Coordinate Systems [84].

The w axis represents the optical axis and its value is usually equal to $-f$ (where f is the focal length of the camera). The image space coordinates described image point positions inside the camera and are usually express in the same units as the image plane coordinates. The image space coordinates are referenced as (u, v, w) in this thesis.

5.3.4 Ground Coordinate System

This is a 3D coordinate system made use of a known map projection. The ground coordinate system is referenced as (X, Y, Z) as shown in Fig.5.6, the Z value represents

the altitude above sea level for a given object point. The ground coordinates are usually measured in feet or meters.

5.4 Basic Photogrammetry Geometry

The major task in photogrammetry geometry is to determine the (X, Y, Z) object co-ordinates in the 3D-world given the corresponding 2D image plane co-ordinates (u, v) on two or more digital images, as illustrated in Fig.5.7.
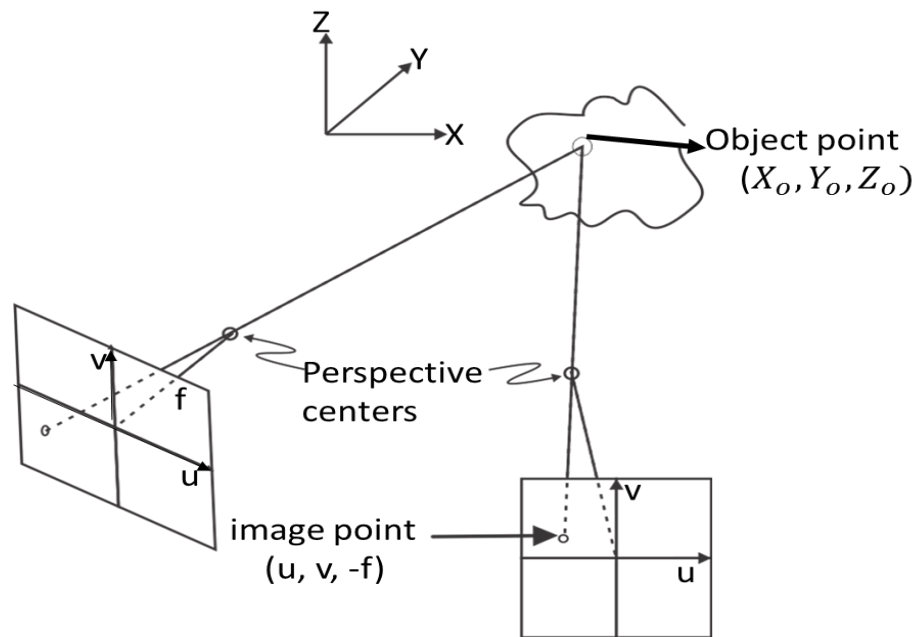
Fig.5.7 Photogrammetry geometry and triangulation [91]

To solve this task, the affine transformation of each of the intersecting rays and the location of the central perception of each exposure must be established with respect to the 3D-ground co-ordinate (X, Y, Z) system. The six parameters that are involved in this process are: three orientation angles and three camera station coordinates. These parameters describe what is termed the exterior orientation of the image (also known as

extrinsic parameters). The camera's interior orientation parameters (intrinsic parameters) are described by the relationship between the perspective centre and the image co-ordinate system. The interior orientation consists of three parameters: camera focal length, principal distance and the co-ordinates of the principal points.

With reference to Fig.5.8, the principal distance is the perpendicular distance from the perspective center to the image plane, and the point at which the optical axis intersects the image plane is called the principal point. The origin of the image plane co-ordinate system (u,v) should ideally be coincident with the principal point but this rarely happens.
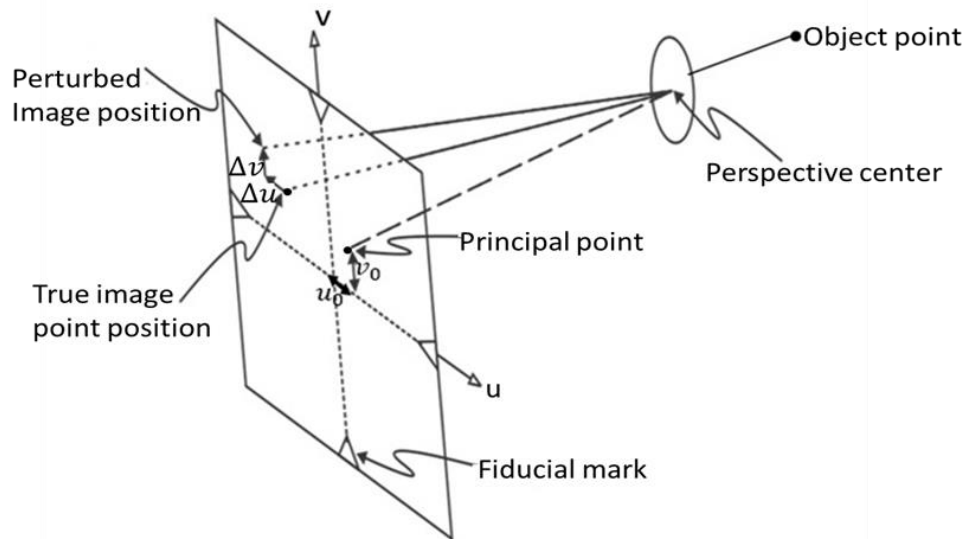
Fig.5.8 Perspective transformation and the effect of offsets to collinearity [91].

Hence the principal point offsets u_0 and v_0 define the shift from the origin (u,v) of the image plane, defined by fiducial (image reference) marks, and the principal point.

For digital cameras, which make use of a CCD matrix array in the image plane, instead of a fiducial system, a central row and column are used to define the origin of the (u, v) image plane coordinate system. CCD cameras usually have values up to 0.5mm for the origin co-ordinates [91], [92].

5.4.1 Principle of Collinearity

When the 2D coordinate system of the image plane (u, v) align very well with the object 3D-space coordinate system. Then the object point on the ground coordinate system, the perspective center and the image point in the 2D image plane, will all lie along the same straight line (as shown in Fig.5.6, points "A", "c" and "a", lie on a straight line). This is the principle of collinearity [93]. Although, this is not always the case, collinearity can only be achieved by capturing images with a perfect camera system. In a general situation, the image point, the object point and the perspective center does not always satisfy the principle of collinearity, giving rise to departures $\Delta u \; and \; \Delta v$ of the image point from its true position on the image plane, as shown in Fig.5.8. These offsets, which are caused by various geometric distortions, can be corrected by the calibration of the photogrammetric imaging system. Calibration of the imaging system in its simplest form is the shifting the actual image point coordinates ($\Delta u \; and \; \Delta v$) so as to satisfy the collinearity principle. Using the principle of collinearity, the perspective transformation between image and object space is given by:

$$
\begin{bmatrix} u - u_0 + \Delta u \\ v - v_0 + \Delta v \\ -f \end{bmatrix} = cR \begin{bmatrix} X_i - X_0 \\ Y_i - Y_0 \\ Z_i - Z_0 \end{bmatrix}.
\qquad (5.1),
$$

where,

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} =$$

$$\begin{bmatrix} \cos\phi\cos\kappa & \cos\omega\cos\kappa + \sin\omega\sin\iota\cos\kappa & \sin\omega\sin\kappa - \cos\omega\sin\phi\cos\kappa \\ -\cos\phi\sin\kappa & \cos\omega\cos\kappa - \sin\omega\sin\phi\sin\kappa & \sin\omega\cos\kappa + \cos\omega\sin\phi\sin\kappa \\ \sin\phi & -\sin\omega\cos\phi & \cos\omega\cos\phi \end{bmatrix} \quad (5.2)$$

This is the transformation matrix that describes the relative orientation between the image space and object space coordinate systems. After some mathematical manipulation (5.1) becomes:

$$u - u_0 + \Delta u = -f \frac{r_{11}(X - X_0) + r_{12}(Y - Y_0) + r_{13}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)} \quad (5.3a)$$

$$v - v_0 + \Delta v = -f \frac{r_{21}(X - X_0) + r_{22}(Y - Y_0) + r_{23}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)} \quad (5.3b)$$

where:

r_{ij} is an element of R,

(u, v) are the image coordinates,

(u_0, v_0) are the principal point coordinates,

(X_i, Y_i, Z_i) are the object point coordinates,

(X_0, Y_0, Z_0) are the exposure station coordinates,

f is the camera focal length,

$(\Delta u, \Delta v)$ are distortion in the image point coordinates.

It was shown by Gruen and Beyer [94] that it is possible to determine both the interior orientation elements and the exterior elements, which are the parameters that described the

departures from collinearity through analytical approach based on the solutions of both equation (5.2a) and (5.2b). Analytical self-calibration and test-range calibration represents this approach. Finding the solution of all parameters forming the collinearity equations: interior and exterior orientation, calibration coefficients and object space target point co-ordinates simultaneously without prior knowledge of the 3D co-ordinate information of the object space target area is known as Self-Calibration [94], [95]. Test-range calibration [95] has the same procedural calibration as the self-calibration but required object space control points in its calibration procedure, such a requirement is not necessary in self-calibration.

5.4.2 Direct Linear Transform (DLT)

The direct linear transform (DLT), which was originally presented in [96], is based on the principle of co-linearity (i.e. all points must be on a straight line). It is a projective geometry transformation between 3-D object space and 2-D image plane, it is expressed as:

$$u + \Delta u = \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_9 X + L_{10} Y + L_{11} Z + 1}, \qquad (5.4a)$$

$$v + \Delta v = \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_9 X + L_{10} Y + L_{11} Z + 1} \qquad (5.4b)$$

where u and v are the image or pixel co-ordinates not necessarily referenced to the principal point, $L_1 - L_{11}$ can be physically interpreted as the interior and exterior orientation parameters of the image. The DLT is used only for non-metric camera which does not require *a priori* knowledge of the interior orientation parameters (u_0, v_0, d). But

in the digital camera context, the DLT offers two major advantages when compared with the collinearity approach. The first, been the ability of the DLT method to provide a non-iterative and direct solution, which is independent of the initial parameter estimates. Hence resulting in a faster computation. The second advantage is the presence of the affine/shear image co-ordinate correction in the DLT model, which is very appropriate for CCD sensors. However DLT requires foreknowledge of the object space and image coordinates of a set of Ground Control Points (GCP) that is well distributed in three dimensions [95], [96].

5.4.3 Sources of perturbation

In finding suitable parameters for the functions Δu and Δv, it is important to study the three major causes of departures from collinearity (perturbation). These distortions which are physical in nature are: the symmetric radial distortion, the de-centering distortion, and focal plane unflattening [96], [97]. The total image shift at a point will be the accumulative effect of each of these perturbations (offsets). Thus,

$$\Delta u = \Delta u_r + \Delta u_d + \Delta u_u \qquad (5.5a)$$

$$\Delta v = \Delta v_r + \Delta v_d + \Delta v_u \qquad (5.5b)$$

where the subscripts r, d, and u are the radial distortion, decentring effects, and out of plane unflattening respectively. The relative magnitudes of these perturbations depend on the nature of the camera used for image acquisition.

5.4.3.1 Radial Distortion

The most common distortion is the radial lens distortion, which causes the real image point to be displaced radially in the image plane. The radial distortion is usually approximated using an odd-ordered polynomial series expression:

$$\Delta r = K_1 r^3 + K_2 r^5 + K_3 r^7 + ... \tag{5.6}$$

where $K_1, K_2...$ are coefficients of radial distortion and r is the radial distance from the principal point, i.e.

$$r = \sqrt{u^2 + v^2} \tag{5.7}$$

For most of the medium-angle, non-photogrammetric lenses used in close-range cameras, the third-order terms are sufficient to model the lens distortion. Wide-angle lenses employ higher order terms (very rarely above seventh-order term) for adequate lens distortion. For CCD cameras, one or two coefficients are enough to compensate for the distortion [97]. From (5.6) the necessary radial distortion correction for the u and v image plane co-ordinates is given by:

$$\Delta u_r = \frac{\bar{u}}{r} \Delta r \tag{5.8a}$$

$$\Delta v_r = \frac{\bar{v}}{r} \Delta r \tag{5.8b}$$

where $\bar{u} = u - u_0$ and $\bar{v} = v - v_0$.

5.4.3.2   Decentering Distortion

The centers of curvature of lens surfaces not always being collinear gives rise to another category of lens distortion known as decentering distortion. This type of distortion originated from both radial and tangential image displacements, which were demonstrated by Brown [98] in the following corrected equations:

$$\Delta u_d = P_1(r^2 + 2\bar{u}^2) + 2P_2\bar{u}\bar{v} \qquad\qquad (5.9a)$$

$$\Delta v_d = P_2(r^2 + 2\bar{v}^2) + 2P_1\bar{u}\bar{v} \qquad\qquad (5.9b)$$

where P_1 and P_2 are decentering distortion parameters.

Like the radial distortion, decentering distortion also varies with focussing, but the resultant image co-ordinate offsets are usually small and this offset is generally ignored in analytical photogrammetry.

5.4.3.3   Focal Plane Unflatness

In focal plane or film unflatness distortion, image displacement is caused by image focal plane unflatness. Which is a major factor affecting the accuracy of the photogrammetric process, this is common especially where nonmetric cameras are involved. Fig.5.9, illustrates the effect of the unflatness image plane and it can been seen that the radial image displacement Δr_u is a function of the incidence angle of the imaging ray.

Hence, the short focal length with wide-angle lenses is more greatly affected by out-of-plane image distortion than the long focal length with narrow-angle lenses. [99], shows

that a CCD array can show a degree of planarity that does not permit any unflatness correction.
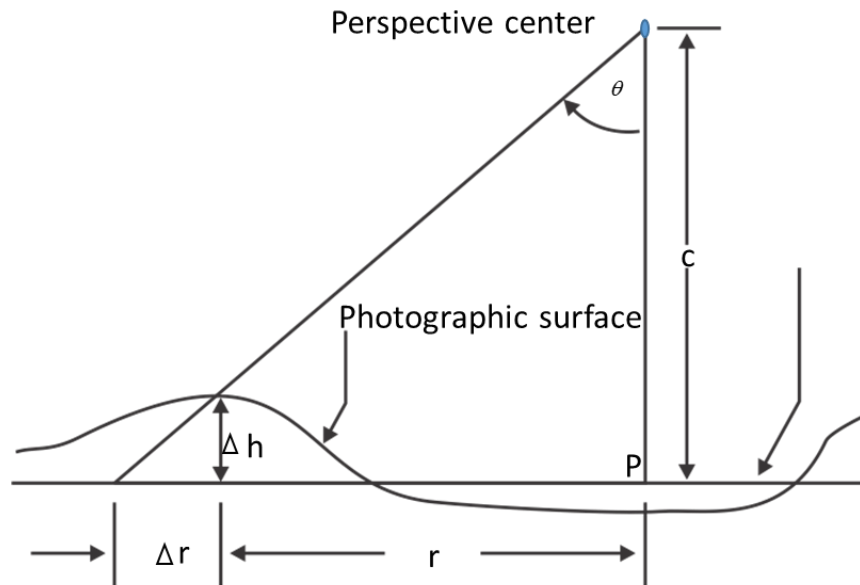
Fig.5.9 Image Displacement Due to Focal Plane Unflatness [100]

There are other types of distortion that have been proposed in the literatures [100], [101] and [102]. But in most cases the error is small and the distortion component is insignificant.

5.5   Photogrammetric Process and Results

The output products of photogrammetric processes range from ortho-image formation, digital elevation model extraction, feature gathering, stereo pair formation, control point extension and very accurate point determination [103]. In order to produce these aforementioned products of photogrammetric process, the relationship between the camera/sensor used to capture an object point on the ground, the object point itself and its

corresponding image point on the image plane must be defined. In order to define this relationship the following parameters must be determined for each image:

- Interior orientation parameters

- Exterior orientation parameters

- Precise representation of the ground

The most difficult task in photogrammetry is trying to use a very small number of ground control points (GCPs) for each image in a project. Most photogrammetric applications use a limited number of GCPs because of the intensive procedures associated with collecting GCPs.

5.5.1 Photogrammetric Techniques

The input data provided for a photogrammetric process determine the photogrammetric techniques that can be used for production of ortho-rectification, digital elevation extraction and other photogrammetry output products.

5.5.1.1 Space Resection

The method that is usually used to determine the exterior orientation parameters of one image or many images based on known GCPs is known as space resection [104]. Space resection is based on the principle of collinearity, which stipulates that, the object point on the ground, the perspective center of the camera and the image point of the corresponding object point on the image plane must lie along the same straight line.

With at least three GCPs (X, Y, Z) coordinate known and with the camera information available, the space resection can be used to calculate the exterior orientation parameters of an image.

Space resection is useful when performing single frame ortho-rectification, this mean processing one image at a time. But if it involves multiple images, for each image process, a minimum of three GCPs must be found on them.

To compute, the positions of the exterior orientation parameters, space resection uses the collinearity condition, in which light rays from at least three object points pass through its corresponding image points on the image plane and resect at the perspective center of the camera. Afterwards, the least squares adjustment method is used to compute the positions of the exterior orientation.

5.5.1.2 Space Forward Intersection

The techniques used to determine the X, Y, Z ground coordinates of points that appear on the overlap part of two or more images, with both the interior and exterior orientation parameters given, is known as space forward intersection [104].

To determine the X, Y, Z ground coordinates of points that appear on the overlap part of two images for example. The concept of space forward intersection is illustrated in Fig.5.10. Here, the principle of collinearity is applied, in which the corresponding light rays from the two exposure stations O1 and O2, pass through their corresponding image point's p1 and p2 on the two images respectively and intersect at the same ground point P.
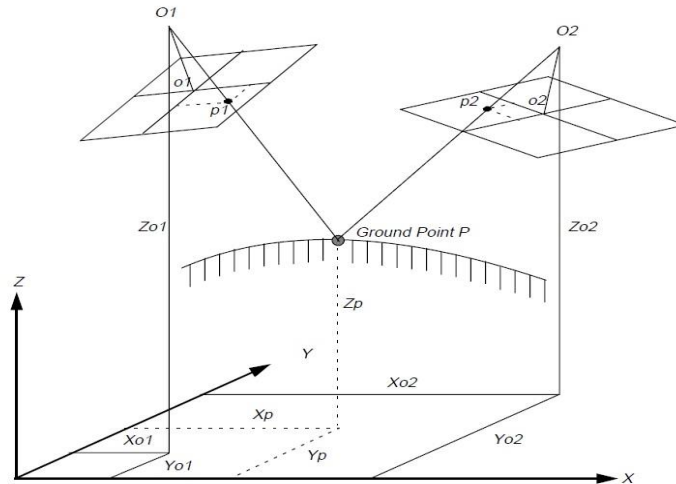
Fig.5.10 Space Forward Intersection [103].

Using the known exterior orientation parameters and the image coordinate measurements of point P on image 1 and image 2 as input in the collinear equations, then, the X_P, Y_P and Z_P ground coordinates of point P can be computed.

Space forward intersection techniques are used for applications, such as, cadastral mapping [103], [104], using airborne surveying method, GCPs collection and highly accurate point determination.

5.5.1.3 Bundle Block Adjustment

For applications involving many images, space forward intersection and space resection methods cannot give a highly accurate result. This can be attributed to the lack of accuracy in the values of the input parameters. For example, it is very rare for values of exterior parameters computed by the airborne GPS and INS techniques in aerial photogrammetry to be very accurate for each image or photograph in a project due to both systematic and non-systematic factors. Thus, on board GPS and INS techniques provide initial estimate

values of the exterior orientation parameters, however, the final value for these parameters must be adjusted to obtain higher accuracies.

In space resection, the more the images in a project, the more GCPs are required for the computation of exterior orientation parameters. For example in a project that consists of 70 or more images, to implement space resection, a minimum of 210 GCPS is needed. Hence, the time and the cost required to identify, match and measure all of the points will be high.

The high cost associated with block triangulation and ortho-rectification is dependent upon the number of GCPs used for the photogrammetric process. This cost can be reduced by collecting fewer GCPs. To ensure high accuracies with the use of reduced GCPs, a technique known as bundle block adjustment [104], [105], is employed.

Bundle block adjustment can be defined by examining each of the words that makes up the term. A bundled solution is a solution that is computed by determining each image position and its exterior orientation parameters, the tie points X, Y, Z coordinate system, and the GCPs. In one solution, all the images in a block in a project are processed at the same time. The least squares adjustment method [106] is used in the entire block model for the estimation of the final bundle adjustment, while error is been minimized and distributed along the entire block.

The process of defining the geometrical relationship between the images in a block, the camera or sensor used to capture these images, and the ground is known as block triangulation [107]. Once this geometrical relationship has been defined, an accurate representation of the imagery on the Earth's surface is established. When processing imagery captured from digital camera and videography camera, block triangulation is denoted as Aerial Triangulation (AT). While in processing imagery captured by a push broom sensor, block triangulation is denoted as Triangulation.

In photogrammetry, many models of block triangulation are used, among these are: the strip method, the independent model method, and the bundled method. The bundle method is the most difficult of the above methods, in view of its ability to minimize and distribute errors [108], [109].

5.5.1.3.1 The Concept of Bundled Block Adjustment

The concept of bundled block adjustment is based on the collinearity condition. To illustrate this concept, an example, consisting of two frames CCD camera data with three GCPs of known X, Y, and Z coordinates. Also provided are six tie points as illustrated in Fig.5.11. For every single ground point measured, the corresponding image point has coordinates (u, v). Hence, to establish a relationship between an object point on the ground and its corresponding image point on the image plane, two collinearity equations are created. These equations in the context of bundled block adjustment are known as observation equations [105], [106], [107].
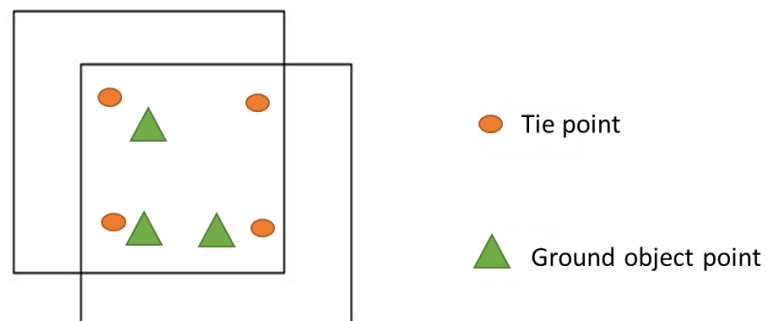
Fig.5.11. Two Frames CCD with 3 GCPs and 4 Tie Points

For an object point on the ground measured on the overlap areas of two images, four collinearity equations can be created. This includes two for the top image comprising a pair and two for the bottom image comprising another pair. This implies that an object point on the ground measured on the overlap part of two images consists of four collinearity equations that can be formulated as:

$$u_{a1} - u_0 = -f \frac{r_{11}(X_A - X_{01}) + r_{12}(Y_A - Y_{01}) + r_{13}(Z_A - Z_{01})}{r_{31}(X_A - X_{01}) + r_{32}(Y_A - Y_{01}) + r_{33}(Z_A - Z_{01})} \qquad (5.10a)$$

$$v_{a1} - v_0 = -f \frac{r_{21}(X_A - X_{01}) + r_{22}(Y_A - Y_{01}) + r_{23}(Z_A - Z_{01})}{r_{31}(X_A - X_{01}) + r_{32}(Y_A - Y_{01}) + r_{33}(Z_A - Z_{01})} \qquad (5.10b)$$

$$u_{a2} - u_0 = -f \frac{r'_{11}(X_A - X_{02}) + r'_{12}(Y_A - Y_{02}) + r'_{13}(Z_A - Z_{02})}{r'_{31}(X_A - X_{02}) + r'_{32}(Y_A - Y_{02}) + r'_{33}(Z_A - Z_{02})} \qquad (5.11a)$$

$$v_{a2} - v_0 = -f \frac{r'_{21}(X_A - X_{02}) + r'_{22}(Y_A - Y_{02}) + r'_{23}(Z_A - Z_{02})}{r'_{31}(X_A - X_{02}) + r'_{32}(Y_A - Y_{02}) + r'_{33}(Z_A - Z_{02})} \qquad (5.11b)$$

An image point measurement of ground object point A on image 1:

u_{a1}, v_{a1}

An image point measurement of ground object point A on image 2:

u_{a2}, v_{a2}

Camera exposure station of exterior orientation on image 1:

X_{01}, Y_{01}, Z_{01}

Camera exposure station of exterior orientation on image 2:

X_{02}, Y_{02}, Z_{02}

With reference to the example in Fig.5.11, since three ground object points have been measured on the overlap areas of the two images, this implies that twelve collinearity equations can be created. This includes, two equations for each ground object point (3 object points) on image 1, and also, two equations for each ground object point (3 object points) on image 2.

Also in the example above, since four tie points have been measured on the area of overlap of the two images, sixteen equations can be created (from each tie point). This gives a total of 28 observation equations. In this example (Fig.5.11), the known elements are:

- The six exterior orientation parameters of the image on top (i.e., X, Y, Z, ω, ϕ and κ).

- The six exterior orientation parameters of the image at the bottom (i.e., X, Y, Z, ω, ϕ and κ).

- The tie points X, Y, and Z coordinates, create twelve unknowns (i.e., four tie points multiply by three X, Y, Z coordinates).

From the above, the overall total of the unknowns equals 24.

The value and redundancy in the input data determines the overall quality of a bundle block adjustment. With reference to the example above, the redundancy in the project can be determined by subtracting the number of unknowns (24) from the number of knowns (28). Then the resulting redundancy is 4. Once all the observation equations have been created, the collinearity condition can then be solved using a least squares adjustment approach.

5.5.1.3.2   Least Squares Adjustment in Aerial Triangulation

Least squares adjustment technique is an algebraic method used for the estimation of the unknown parameters linked with a solution while also reducing error within the solution [106], [109]. The purpose of least square adjustment in block triangulation is:

- The adjustment of values linked with exterior and interior orientation parameters.

- The estimation of values of X, Y and Z coordinates linked with the tie points

- Minimizing and spreading of input data error through the entire bundle model.

The least squares adjustment technique uses repeated processing until a desired result is reached. A desired result is reached when the input data residuals are minimized. This includes, finding the corrections to the unknown parameters by reducing the input residuals. The residuals are obtained from the variance between the measured user's input and the computed value in a project.

The residuals to be minimized include: the image point coordinates of the object points on the ground, the tie points and the known object point ground coordinates. The least squares conditions can be formulated as follows:

$$V = AX - L \qquad\qquad (5.12)$$

where:

       V = a vector of image coordinate residuals

       A = the matrix of partial derivative associated with the unknown parameters, which includes, exterior orientation, interior orientation, tie points X, Y, Z and object point ground coordinates

       X = the matrix of the unknown parameters.

L = the matrix of the input observations, which includes image point coordinates and object point ground coordinates.

The X matrix, which holds the adjustments of the unknown parameters, can be calculated by the following formula:

$$X = (A^t P A)^{-1} A^t P L \qquad (5.13)$$

where, X, A and L are as in (5.12), t is the matrix transposed and P is the matrix holding the observations.

After the least squares iteration process is completed, the adjusted values of the unknown parameters are then added to the initial estimates. In most practical cases, the initial values are the exterior orientation parameter readings obtained from the on board GPS/INS sensor and to compute the final EOPs, the adjusted value of the EOPs that is computed by the least square adjustment process is added to the initial value of EOPs provided by the on board GPS/INS. The least squares iteration technique continues to adjust the values of the unknown parameters by an iteration process until a pre-defined threshold is reached. The results that can be obtained from the least squares adjustment process in an aerial triangulation includes:

- The absolute values of both the exterior and the interior orientation parameters of each image in a block.
- The image and GCP coordinate residuals.
- The coordinates of the adjusted X, Y and Z tie points.

The block triangulation result above can then be used as primary inputs for processing the following tasks:

- Ortho-rectification

- Formation of Digital Elevation Model (DEM).

- Creation of stereo pairs

- Accurate point determination

5.6 Ground Control Points

The photogrammetry component that is instrumental to the establishment of a highly accurate relationship between the camera/sensor used to capture the images in a project, the imagery itself and the Earth's surface is the ground control points (GCPs). These GCPs have features that are located on the Earth's surface, which have X, Y, and Z coordinates that are known. The X and Y coordinates of a full GCP specifies the horizontal control while the Z coordinate specifies the vertical control (elevation).

Ground control surveys for ortho-rectification and geo-referencing, consist of two procedural steps. The first procedure involves creating a setup of basic control on the project area. This basic control consists of the scaling of the project area by determining the X and Y coordinates of permanent features on the project, and the benchmark of vertical (Z) control elevation that serves as a reference model for subsequence surveys. Basic ground control surveys can be done by using either the conventional ground techniques or by using GPS techniques. [108], [106] and [109] found GPS techniques to be cheaper and very accurate when compared with the conventional ground technique.

The second step, involves the measurement of established image points on the image plane corresponding to object points on the scaled (control) ground.

The topographies on the Earth's surface used as GCPs for ground surveys must be permanent features, so that they can be reused for a subsequent visit. Commonly used features are:

- Side walk corners

- Intersection of roads

- Survey benchmarks

5.6.1 GCP for Mapping Project Area

The minimum number of GCP required for a photogrammetric mapping project depends on the size of the area. Theoretically, to establish a relationship between the image space and the object's ground space, at least two GCPs having X, Y, and Z coordinates and one GCP having a Z coordinate is needed. This gives a total of seven observations.

In processing one image, for the purpose of ortho-rectification (known as single frame ortho-rectification), a minimum of 3 GCPs is required. Each GCP must have X, Y and Z coordinates associated with it and must be evenly distributed to ensure high accuracy in camera/sensor modelling.

In processing a strip of adjacent images, a minimum of 2 GCPs for every third image will be adequate. In a strip or block of images, an accurate geometry can be applied in regions where there is less redundancy, such as the curve edges of the strip or block. Thus, choosing GCPs located at the curve ends of a strip or block increase the accuracy of ortho-rectification. The GCPs arrangement for a block of images consisting of four strips of

images with each strip having eight overlapping images is illustrated in Fig.5.12. The image locations of the GCPs are measured on the overlap parts of the images.
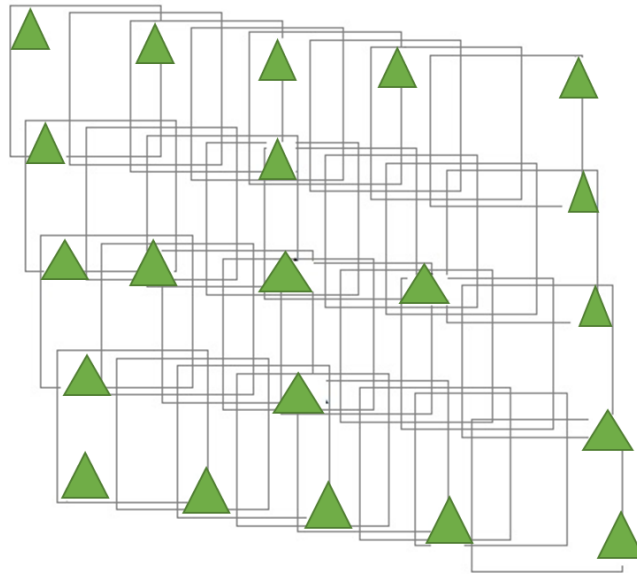
Fig. 5.12 GCP arrangement in a block of images

5.7   Tie Points

A point that is identifiable in the overlap areas between two or more images is known as a tie point. The tie point has ground coordinates that are not known, tie point ground coordinates are determined during aerial triangulation. Tie points in two overlapping images appearing in their overlap areas are identified and measured during block triangulation. Tie point measurement can be done either manually or automatically. Tie points must be well distributed over the area of the block to ensure high accuracy during block triangulation.

As illustrated in Fig.5.13, for a block of images with 60% forward overlap and 30% side overlap, nine points will be enough to tie together the block along with separate strips.
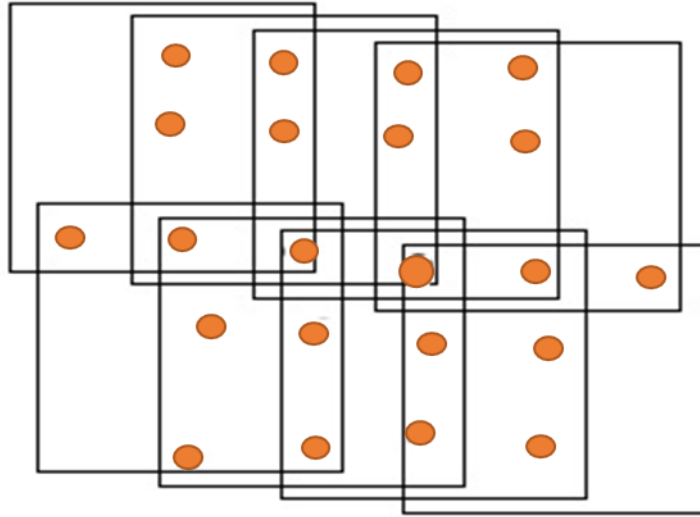
Fig.5.13 Tie points on a block of images

Image matching techniques which include, area based matching, and feature based, are used individually or in a combined form for the automatic identification of tie points and to tie together block images and individual strips.

5.8   Chapter Summary

The main purpose of photogrammetric technique is to thoroughly state the sizes of an object and hence define its coordinates. This was achieved based on the intersection of rays from a pair of photographic images. The origin of photogrammetry started with analog photogrammetry in which, images were recorded on photographic film. During this period, orientation reconstruction and other measurements are carried out in an optical mechanical way. The orientation reconstruction was later changed to algorithms and is called analytical photogrammetry. With advent digital cameras and high speed computer systems, computers are employed for all photogrammetric processes. There are two types of imaging method in photogrammetry, this includes terrestrial and aerial imaging. With

respect to aerial imaging, a photogrammetric camera is fixed on an aircraft and runs parallel flight strips as defined in the flight plan. The two dimensional images captured by the aircraft, are observed pair by pair in order to create a stereoscopic view. Tie points in a pair of images with a number of known ground control points are adjusted based on geometric equations, in which specific rules (collinearity) are followed. Thus, this process is called aerial triangulation. The results obtained from these equations are used as inputs for processing photogrammetric tasks such as: ortho-rectification, formation of DEM, etc.

CHAPTER SIX

Ortho-Rectification and Geo-referencing of Video Streams from a Low-Cost UAV

6.1 **Introduction**

The flight of the low-cost UAV is generally less-stable when compared with larger fixed-wing aircraft and because it flies at low altitude (approximately 500 m), the camera frame tends to jerk continuously along the flight path. This, coupled with the disorientation caused by long viewing of the video camera, makes the analysis of data acquired by the low-cost UAV difficult. Thus, the ortho-rectification and geo-referencing of the video data could serve as an important post-processing step preceding the analysis of the UAV data. Ortho-rectification is the process of geometrically correcting an aerial image such that the scale is uniform. The ortho-image formed from the process is corrected for lens distortion, topographic relief, and camera tilt. This can be used to measure true distances, because it is represents the exact measurement of the earth's surface. Ortho-rectification and geo-referencing are essential to pin point the exact location of targets in video imagery acquired at the Unmanned Aerial Vehicle (UAV) platform. This can only be achieved by comparing such video imagery with an existing digital map. However, it is only when the image is ortho-rectified with the same co-ordinate system as an existing map that such a comparison is possible. Many approaches have been presented for ortho-rectification and geo-referencing of low-cost UAVs by researchers in the past. These range from using different types of operational platforms such as satellite [110], fixed-wing aircraft [111], helicopters and UAVs; and different sensors like radar [112], visible and multi-spectral images [113]. Also, many mathematical models have been used for ortho-rectification, ranging from the simple affine transformation [114] to projective transformation [115]. Geo-referencing in general is based on direct geo-referencing using a GPS/INS system, but because of the low quality of the GPS/INS usually used for low-cost UAVs during data acquisition, the GPS/INS accuracy is very low.

The method used in this research for ortho-rectification and geo-referencing is based on the photogrammetry model [116], [117], this is a form of geometry imaging system that makes use of the aerial image central perspective and the principle of collinearity. A close alternative to this photogrammetry method, are the Laser scanner and the Push-broom [118]. These two approaches were observed by [119] and it was found to be time consuming and very expensive.

6.2 Experimental Procedures

The experimental procedure for ortho-rectification and geo-referencing of object detected from video streams captured by a low-cost UAV is as follows:

6.2.1 Data Collection and Test Field Calibration

The UAV used in the experimental work reported in this thesis is the Spreading Wings S800, the detailed configuration and specification of the Spreading Wings S800 is given in chapter three of this thesis. The Spreading Wings S800 UAV is equipped with GPS/INS and has the capability to geotag (time and position tagged) each video frame at the exact time of exposure.

The experiments were conducted on a calibrated control field at the federal capital territory Abuja, Nigeria. The control field coverage area is 0.11357 square kilometers (sq. km). Six ground control points (GCPs) were collected using a hand held GPS receiver, and these ground control points, which are located at the corners of sidewalk and crossroad (Fig.6.1), were observed till four satellites are locked simultaneously before readings were taken. The six GCPs collected are converted from a geodetic coordinate system to X, Y

Cartesian coordinate system in meters for use in the computation of the final bundle block adjustment.

Fig.6.1 Digital map of the test field used in the experiment showing the location of the 6 GCPs

The aerial platform includes the UAV and the sensor systems (camera and GPS/INS system). The UAV, which is remotely controlled by a pilot at the control station, flies along the test field and collects video streams of the test area, while the on board GPS/INS system simultaneously collects telemetry data of the test area. The telemetry data consists of the UAV's position and angular attitude. All data collected and their source are also summarized in Table 6.1.

Data	Purpose
Mpeg-based video stream	Captured by UAV video camera to pin point exact locations of targets.
Telemetry Data	Captured by the GPS/INS system on the UAV platform. It consist of the UAV's position and angular attitude
GCPs	Captured by hand-held GPS receiver. It is used to measure the X and Y coordinates of the GCPs.

Table 6.1 Data collection source and its purpose

The video and the telemetry data were collected for approximately 25 minutes and were transmitted through a radio link of 2.4GHz to the portable PC at the ground station in real time. In order to perform a near real-time image geo-referencing of the video streams acquired by the UAV platform, it is essential to generate an ortho-image from the image sequence. Due to the limited payload weight on the UAV platform we were forced to off-load this process to the ground station (Fig. 6.2).
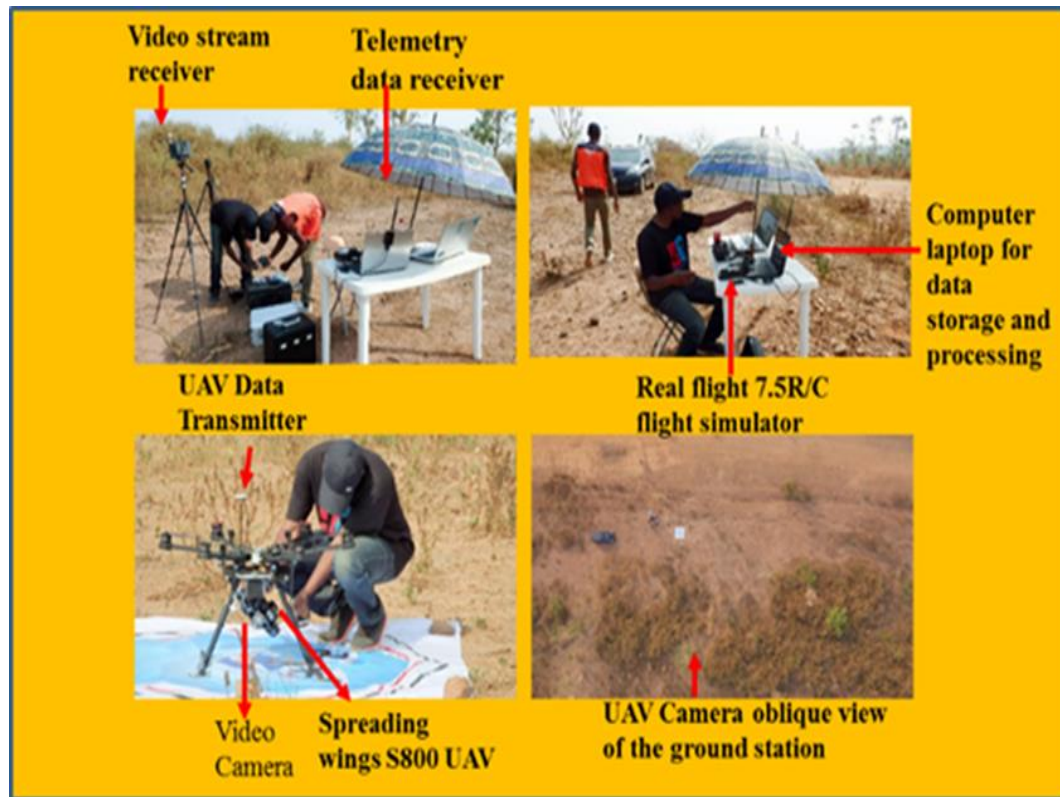
Fig. 6.2 Ground control Station

Since the GPS and INS are integrated into the camera on board, the received video sequence frames are time and position tagged (see appendix I). Fig.6.3 shows the GPS flight elevation against distance covered by the UAV. The unstable lines in the graph are an indication of continuous jerking of the aircraft along the flight line. Hence, the cause of distortion in the captured images and the need for ortho-rectification of these images.

Fig.6.3 GPS/INS Flight Track Log (Elevation Vs Distance)

6.2.2 **Multi-sensor System Data Fusion**

The multi-sensor digital systems used in this research consists of a digital camera system used for image acquisition and a GPS/INS integrated sensor system used for recording camera's positions and attitudes. When using multi-sensor digital systems, all these systems need to be calibrated in order to minimise the systematic errors that arises from these systems. The main calibration requirement in a multi-sensor systems is the camera and boresight calibration.

Boresight calibration involves computing the misalignment angles between the INS body frame and the camera frame. Since the INS sensor system and the camera cannot occupy the same center on the UAV there are is a gap between the two sensors coordinate system. The angle formed by the gaps between the coordinate system of the INS body frame and

the camera coordinate system is known as boresight angle. The concept of boresight

calibration is shown in Fig. 6.4.
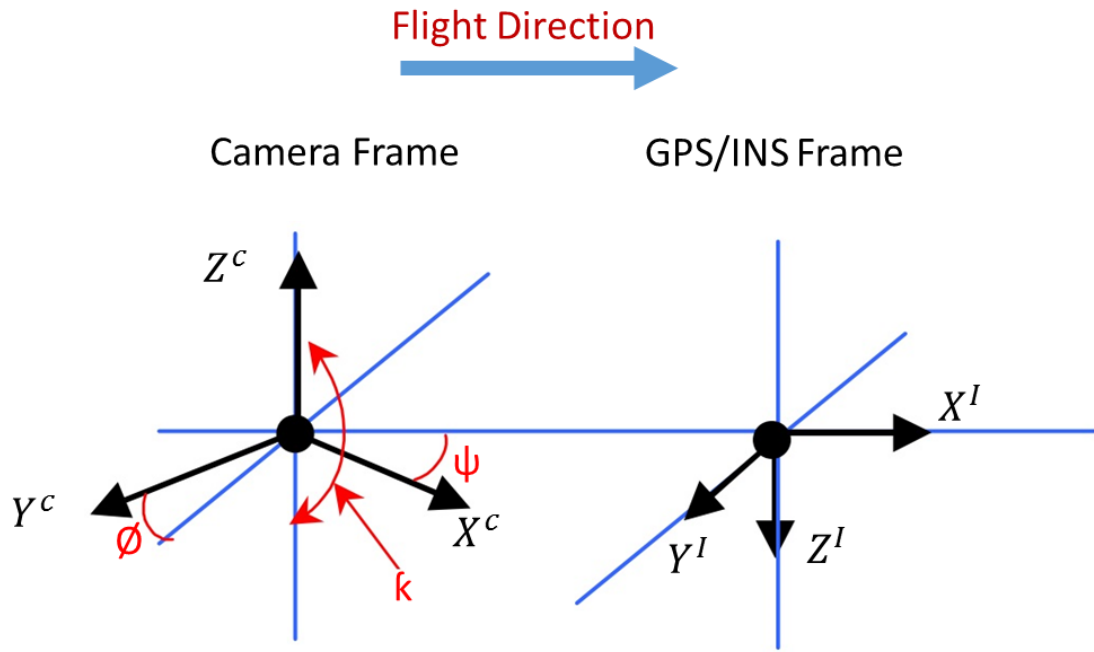
Fig. 6.4 INS/Camera Boresight Calibration

The relative orientation of the camera frame coordinate system with respect to the INS

body frame coordinate system is defined by the direction cosine matrix as:

$$
R_C^I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

$$= \begin{bmatrix} \cos\phi\cos\kappa & \cos\phi\sin\kappa & -\sin\phi \\ \sin\varphi\sin\phi\cos\kappa - \cos\varphi\sin\kappa & \sin\varphi\sin\phi\sin\kappa + \cos\varphi\cos\kappa & \sin\varphi\cos\phi \\ \cos\varphi\sin\phi\cos\kappa + \sin\varphi\sin\kappa & \cos\varphi\sin\phi\sin\kappa - \sin\varphi\cos\kappa & \cos\varphi\cos\phi \end{bmatrix}$$

(6.1)

Hence (6.1) is used as additional observations in the final bundle adjustment and estimating the boresight as one of the parameter.

6.2.3 Ortho-Rectification and Geo-Referencing of Video Streams

The basic steps for video imagery ortho-rectification and geo-referencing are described in the following sub-sections:

6.2.3.1 Feature Points Extraction and Tie Points Matching of Video Frames

After the de-compilation of the video streams into individual frames as discussed in chapter four of this thesis. The next step is to extract feature points from each video frame, track tie points between images and video frame matching. A feature based algorithm is used for feature points extraction, tie points tracking and image matching of the video frames. We chose feature based matching because, it is invariant to radiometric changes and it runs faster when compared with other matching techniques. Also the features used are points rather than lines. This enables us to avoid problems caused by broken lines and more so, points are very easy to describe and are invariant to central projection. Feature points

extracted and matched are based on a number of control parameters such as window sizes and threshold values. This matching technique is detailed in chapter three of this thesis.

Based on the SURF algorithm operations, a chain of overlapped video frames and corresponding tie points are generated for further (relative and absolute) bundle adjustment model.

6.2.3.2 **Finding of Interior Camera Orientation Parameter**

The next step is the finding of the interior orientation of each video frame, known as camera calibration. This involves finding the focal length of the camera, principal point co-ordinates and lens distortion of each video camera frame. For the calibration of the video camera, we used a mathematical model called the Direct Linear Transform (DLT). As mentioned in chapter five, the DLT model is based on the principle of co-linearity (i.e. all points must be on a straight line), and it requires foreknowledge of the image plane coordinates and Ground Control Points (GCPs). The DLT model based on a point q on the first image frame and for a ground control point G can be expressed as:

$$u_q + \Delta u = \frac{L_1 X_G + L_2 Y_G + L_3 Z_G + L_4}{L_9 X_G + L_{10} Y_G + L_{11} Z_G + 1} \qquad (6.2a)$$

$$v_q + \Delta v = \frac{L_1 X_G + L_2 Y_G + L_3 Z_G + L_4}{L_9 X_G + L_{10} Y_G + L_{11} Z_G + 1} \qquad (6.2b)$$

where, (u_q, v_q) are the coordinates of the image point q on the first image frame (X_G, Y_G, Z_G) are the coordinates of the ground control point G; and $(\Delta u, \Delta v)$ are optical errors, which can

be expressed as:

$$\Delta u = \mathrm{E}(L_{12}r^2 + L_{13}r^4 + L_{14}r^6) + L_{15}(r^2 + 2\xi^2) + L_{16}\xi\eta \quad (6.3a)$$

$$\Delta v = \Lambda(L_{12}r^2 + L_{13}r^4 + L_{14}r^6) + L_{15}\eta\xi + L_{16}(r^2 + 2\eta^2) \quad (6.3b)$$

where, $[\mathrm{E}, \Lambda] = [u - u_0, v - v_0]$ and $r^2 = \mathrm{E}^2 + \Lambda^2$.

From (6.2),

$$\frac{1}{D}u_q = \frac{1}{D}(L_1X_G + L_2Y_G + L_3Z_G + L_4 - L_9uX_G - L_{10}uY_G - L_{11}uZ_G) + \Delta u \qquad (6.4a)$$

$$\frac{1}{D}v_q = \frac{1}{D}(L_1X_G + L_2Y_G + L_3Z_G + L_4 - L_9vX_G - L_{10}vY_G - L_{11}vZ_G) + \Delta v \qquad (6.4b)$$

where, $\qquad D = L_9X_G + L_{10}Y_G + L_{11}Z_G + 1 \qquad\qquad (6.5)$

(6.4) in matrix form becomes:

$$\frac{1}{D}\begin{bmatrix} u_q \\ v_q \end{bmatrix} =$$

$$\frac{1}{D}\begin{bmatrix} X_G & Y_G & Z_G & 1 & 0 & 0 & 0 & 0 & -u_qX_G & -u_qY_G & -u_qZ_G & \mathrm{E}r^2D & \mathrm{E}r^4D & \mathrm{E}r^6D & (r^2+2\mathrm{E}^2)D & \mathrm{E}\Lambda D \\ 0 & 0 & 0 & 0 & X_G & Y_G & Z_G & 1 & -v_qX_G & -v_qY_G & -v_qZ_G & \Lambda r^2D & \Lambda r^4D & \Lambda r^6D & \Lambda \mathrm{E}D & (r^2+2\Lambda^2)D \end{bmatrix}\begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \end{bmatrix} \quad (6.6)$$

$$\begin{bmatrix} u_q \\ v_q \end{bmatrix} =$$

$$\frac{1}{D}\begin{bmatrix} \frac{X_G}{D} & \frac{Y_G}{D} & \frac{Z_G}{D} & \frac{1}{D} & 0 & 0 & 0 & 0 & \frac{-u_qX_G}{D} & \frac{-u_qY_G}{D} & \frac{-u_qZ_G}{D} & \frac{\mathrm{E}r^2D}{D} & \frac{\mathrm{E}r^4D}{D} & \frac{\mathrm{E}r^6D}{D} & \frac{(r^2+2\mathrm{E}^2)D}{D} & \frac{\mathrm{E}\Lambda D}{D} \\ 0 & 0 & 0 & 0 & \frac{X_G}{D} & \frac{Y_G}{D} & \frac{Z_G}{D} & \frac{1}{D} & \frac{-v_qX_G}{D} & \frac{-v_qY_G}{D} & \frac{-v_qZ_G}{D} & \frac{\Lambda r^2D}{D} & \frac{\Lambda r^4D}{D} & \frac{\Lambda r^6D}{D} & \frac{\Lambda \mathrm{E}D}{D} & \frac{(r^2+2\Lambda^2)D}{D} \end{bmatrix}\begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \end{bmatrix} \quad (6.7)$$

In (6.7), the values of $L_1\ldots\ldots L_{11}$ are the only unknowns. Also in the equation, it was assumed that the image frame consists of a point and one ground control point was used for the calibration of the camera. Thus, less than 16 parameters were used while the unused rows and columns were discarded.

From here, we will denote the right hand-side matrix of (6.7) by $\mathbf{F}$, the $L_1......L_{11}$ matrix

as $\mathbf{L}$, and the left hand- side matrix as, $\mathbf{h}$, so that (6.7) can be expressed as:

$$\mathbf{h} = \mathbf{F}.\mathbf{L} \qquad (6.8)$$

Calibration is achieved by solving for $\mathbf{L}$ in (6.8) but the matrix $\mathbf{F}$ is not a square matrix and

hence cannot be inverted. Thus, a least squares method is employed to solve for $\mathbf{L}$. A

simple way to do this is to use the 'Moore-penrose pseudo-inverse' method [120]. The step

is to multiple both sides of (6.8) by $\mathbf{F^T}$:

$$\mathbf{F^TFL} = \mathbf{F^Th} \qquad (6.9)$$

Since $\mathbf{F^TF}$ is a square, it can now be inverted. We then multiply both sides of (6.9) by the

inverted matrix square $(\mathbf{F^TF})^{-1}$:

$$(\mathbf{F^TF})^{-1}(\mathbf{F^TF})\mathbf{L} = (\mathbf{F^TF})^{-1}\mathbf{F^Th} \qquad (6.10)$$

An identity matrix I is formed on the right-hand side of (6.10), then the solution of $\mathbf{L}$ is

calculated:

$$\mathbf{L} = (\mathbf{F^TF})^{-1}\mathbf{F^Th} \qquad (6.11)$$

Through computational iteration, the $L_1......L_{11}$ parameters can be solved. Afterwards, the

interior and exterior parameters can be calculated.

 The interior orientation parameters can be calculated from:

$$u_0 = -(L_1 L_9 + L_2 L_{10} + L_3 L_{11})/(L_9^2 + L_{10}^2 + L_{11}^2) \qquad (6.12)$$

$$v_0 = -(L_5 L_9 + L_6 L_{10} + L_7 L_{11})/(L_9^2 + L_{10}^2 + L_{11}^2) \qquad (6.13)$$

$$f_u^2 = -u_0^2 + (L_1^2 + L_2^2 + L_3^2)/(L_9^2 + L_{1O}^2 + L_{11}^2) \qquad (6.14a)$$

$$f_v^2 = -v_0^2 + (L_5^2 + L_6^2 + L_7^2)/(L_9^2 + L_{1O}^2 + L_{11}^2) \qquad (6.14b)$$

$$f = \frac{f_u + f_v}{2} \qquad (6.15)$$

The exterior orientation parameters can be calculated from:

$$m_3 = L_9 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} \qquad (6.16a)$$

$$n_3 = L_{10} \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} \qquad (6.16b)$$

$$s_s = L_9 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} \qquad (6.16c)$$

$$m_1 = \frac{1}{f_u} \left[L_1 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + m_3 u_0 \right] \qquad (6.16d)$$

$$n_1 = \frac{1}{f_u} \left[L_2 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + n_3 u_0 \right] \qquad (6.16e)$$

$$s_1 = \frac{1}{f_u} \left[L_3 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + s_3 u_0 \right] \qquad (6.16f)$$

$$m_2 = \frac{1}{f_v} \left[L_5 \Big/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + m_3 v_0 \right] \qquad (6.16g)$$

$$n_2 = \frac{1}{f_v}\left[L_6 \bigg/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + n_3 v_0 \right] \qquad (6.16h)$$

$$s_2 = \frac{1}{f_v}\left[L_7 \bigg/ \sqrt{(L_9^2 + L_{10}^2 + L_{11}^2)} + s_3 v_0 \right] \qquad (6.16i)$$

The rotation matrix can be expressed as in (6.22) by:

$$Q = \begin{bmatrix} m & m_2 & m_3 \\ n_1 & n_2 & n_3 \\ s_1 & s_2 & s_3 \end{bmatrix} \qquad (6.17)$$

The perspective centre coordinates (X_s, Y_s, Z_s) can be calculated from:

$$m_3 X_S + n_3 Y_S + s_3 Z_S + (\sqrt{L_9^2 + L_{10}^2 + L_{11}^2}) = 0 \qquad (6.18)$$

$$u_0 + f_u(m_1 X_S + n_1 Y_S + s_1 Z_S)/W + L_4 = 0 \qquad (6.19)$$

$$v_0 + f_v(m_2 X_S + n_2 Y_S + s_2 Z_S)/W + L_8 = 0 \qquad (6.20)$$

where $W = \sqrt{L_9^2 + L_{10}^2 + L_{11}^2}$.

6.2.3.3 **Finding the Relative Orientation**

The process of orienting images in relation to one another, by recreating the relative position and angular attitude of images with respect to one another at the instants of exposure is called relative orientation. This is achieved in a stereoscopic pair image, by

making the first image fixed and setting the Z coordinate of the first (fixed) image to its focal length, while the X and Y coordinates and the angular attitude (ω, ϕ, κ) of the fixed

image are all set to zero. The X coordinate of the last image is then equal to the photo base. After the determination of the interior orientation parameters of each video frame separately, using the DLT model as mentioned in section 6.2.2.2, the two image coordinates system are clearly known.

Hence, the relative orientation unknown parameters can be easily solved by using tie point coordinates as observations in a least squares solution.

The results obtained can then be used to attach relatively oriented image sets to each other for an entire flight strip. The entire strip is then adjusted to absolute coordinates using ground control points in a final transformation, which is discussed in the next section.

6.2.3.4 **Finding the Absolute Orientation**

The absolute orientation is the process of levelling and scaling the stereo model with respect to a reference datum using ground control points. With the relative orientation parameters computed, both the camera's interior and exterior parameters (IOPs and EOPs)

are solved simultaneously by a calibration model called bundle adjustment model.

The bundle adjustment model used the principle of collinearity to specify the relationship between the images coordinates of points, the image space coordinate at exposure, and the ground coordinates of points.

The bundle adjustment process like the relative orientation, start with the first and second frame (a stereo pair of images) and afterwards, extended to the entire flight strip. Hence, for any ground point Q, the first video frame is given by the following mathematical model:

$$u_{q1} - u_0 + \Delta u =$$

$$-f \frac{r_{11}^1(X_Q - X_{S1}) + r_{12}^1(Y_Q - Y_{S1}) + r_{13}^1(Z_Q - Z_{S1})}{r_{31}^1(X_Q - X_{S1}) + r_{32}^1(Y_Q - Y_{S1}) + r_{33}^1(Z_Q - Z_{S1})} \qquad (6.21a)$$

$$v_{q1} - v_0 + \Delta v =$$

$$-f \frac{r_{21}^1(X_Q - X_{S1}) + r_{22}^1(Y_Q - Y_{S1}) + r_{23}^1(Z_Q - Z_{S1})}{r_{31}^1(X_Q - X_{S1}) + r_{32}^1(Y_Q - Y_{S1}) + r_{33}^1(Z_Q - Z_{S1})} \qquad (6.21b)$$

For the second video frame, the mathematical model is expressed as:

$$u_{q2} - u_0 + \Delta u =$$

$$-f \frac{r_{11}^2(X_Q - X_{S2}) + r_{12}^2(Y_Q - Y_{S2}) + r_{13}^2(Z_Q - Z_{S2})}{r_{31}^2(X_Q - X_{S2}^2) + r_{32}^2(Y_Q - Y_{S2}) + r_{33}^2(Z_Q - Z_{S2}^1)} \qquad (6.22a)$$

$$v_{q2} - v_0 + \Delta v =$$

$$-f \frac{r_{21}^2(X_Q - X_{S2}) + r_{22}^2(Y_Q - Y_{S2}) + r_{23}^2(Z_Q - Z_{S2})}{r_{31}^2(X_Q - X_{S2}) + r_{32}^2(Y_Q - Y_{S2}) + r_{33}^2(Z_Q - Z_{S2}^1)} \qquad (6.22b)$$

where (u_{q1}, v_{q1}) and (u_{q2}, v_{q2}) are the image coordinates of point q in the first and second video frame respectively; $(\Delta u, \Delta v)$ is as in (6.3); (X_Q, Y_Q, Z_Q) are object space coordinates of point Q; (X_{S1}, Y_{S1}, Z_{S1}) and (X_{S2}, Y_{S2}, Z_{S2}) are object space coordinates of the exposure station of the first and second video frames respectively; (u_0, v_0) are the principal point coordinates in the corresponding frame; f is the focal length of the camera; and r^1 and r^2 are the elements of the three rotational angles as in (6.1) in the corresponding frames.

In this mathematical model, the unknown elements, which consist of the camera's interior orientation parameters $(u_0, v_0, f, \Delta u, \Delta v)$, and the exterior orientation parameters of the first and second video frames, $(X_{S1}, Y_{S1}, Z_{S1}, \omega_1, \phi_1, \kappa_1)$ and $(X_{S2}, Y_{S2}, Z_{S2}, \omega_2, \phi_2, \kappa_2)$ respectively, can be solved by linearizing (6.21) and (6.22) using a Taylor series expansion. The linearized equation is expressed in matrix form by:

$$v_1 = A_1 X_1 + A_2 X_2 - l \qquad (6.23)$$

where v_1 is a vector of image coordinate residuals, X_1 is a vector of the exterior orientation parameters of the two video frames, X_2 represents the vector of the camera interior orientation parameters, A_1 and A_2 are the coefficients of X_1 and X_2 respectively.

On the basis of the 6 GCPs collected as described in section 6.2.1, the co-ordinates of the image plane in the first and second video frames were measured and all the unknown parameters in (6.23) are computed.

With the formation of the ortho-image from the above process, all ortho-images (video frames) are mosaicked together to create a mosaic image covering the test area. This mosaic image is then mapped onto the digital globe containing the map of the test area and the accuracy achieved is evaluated by the use of the six ground control points obtained from the test area.

The work flow of the ortho-rectification and geo-referencing procedure is shown in Fig.6.5:
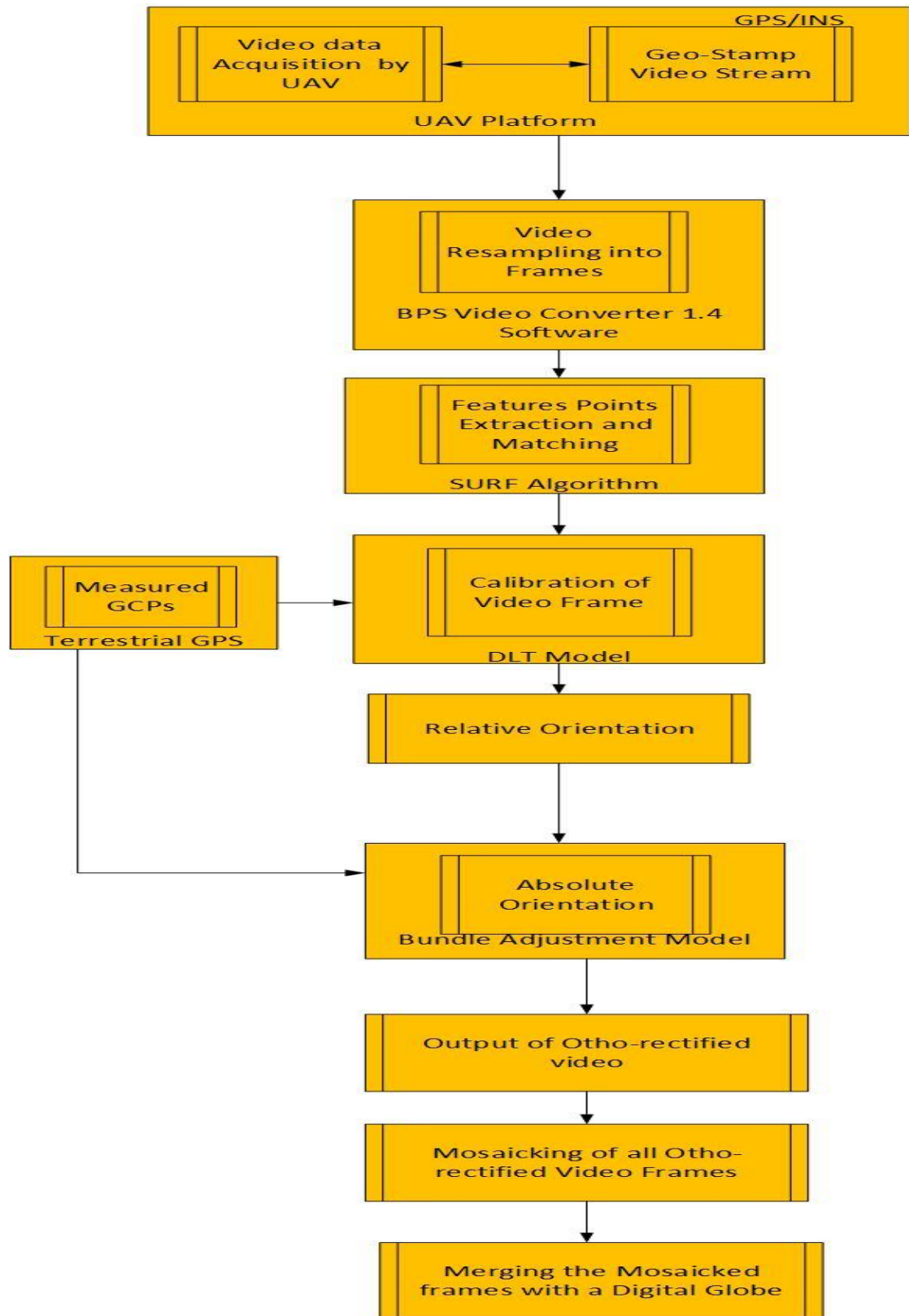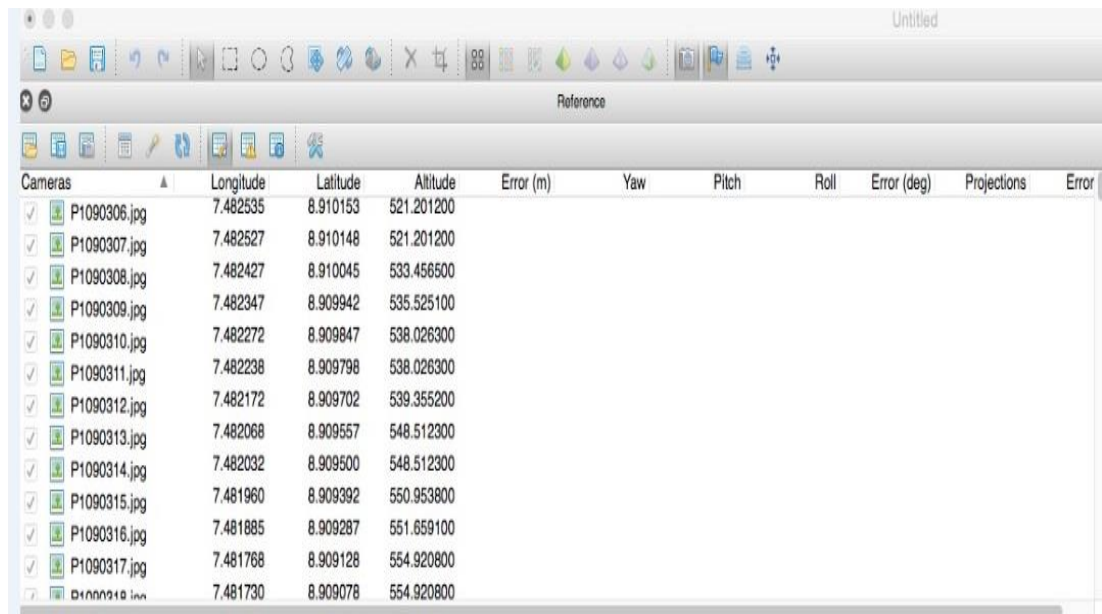
Fig.6.5. Work Flow of Geometric Ortho-rectification and Geo-referencing of Video Stream

6.3 **Experimental Results and Discussion**

As shown in Fig.6.6, each video frame is geo-tagged, that is each video frame is tagged with its corresponding geodetic coordinate (Latitude, Longitude, Altitude), apart from the value of the altitude, values of both Latitude and Longitude are converted to X, Y, Z Cartesian coordinate system in meters (appendix I) . Thus, these values can then be used during computation of in aerial triangulation. It can also be observed from Fig.6.6 that the values for attitude for each video frame are not the same, despite the fact that the UAV travels at constant altitude above the sea level. This is because as the UAV travels, it height above the ground height varies with its topography.

Cameras	Longitude	Latitude	Altitude	Error (m)	Yaw	Pitch	Roll	Error (deg)	Projections	Error
P1090306.jpg	7.482535	8.910153	521.201200							
P1090307.jpg	7.482527	8.910148	521.201200							
P1090308.jpg	7.482427	8.910045	533.456500							
P1090309.jpg	7.482347	8.909942	535.525100							
P1090310.jpg	7.482272	8.909847	538.026300							
P1090311.jpg	7.482238	8.909798	538.026300							
P1090312.jpg	7.482172	8.909702	539.355200							
P1090313.jpg	7.482068	8.909557	548.512300							
P1090314.jpg	7.482032	8.909500	548.512300							
P1090315.jpg	7.481960	8.909392	550.953800							
P1090316.jpg	7.481885	8.909287	551.659100							
P1090317.jpg	7.481768	8.909128	554.920800							
P1090318.jpg	7.481730	8.909078	554.920800							

Fig.6.6 Each Video Frame Geo-tagged with it Corresponding Geodetic Coordinate

Fig.6.7 shows, the 78 frames obtained as a result of video stream de-compilation, the video frames are decompiled at the rate of two frames per second by BPS video converter 1.4 software. Hence, leading to a forward overlap of 60%, no side overlap was recorded because the experiment was conducted for only one flight strip. The frames are shown as not overlapping for the purpose of clarification.
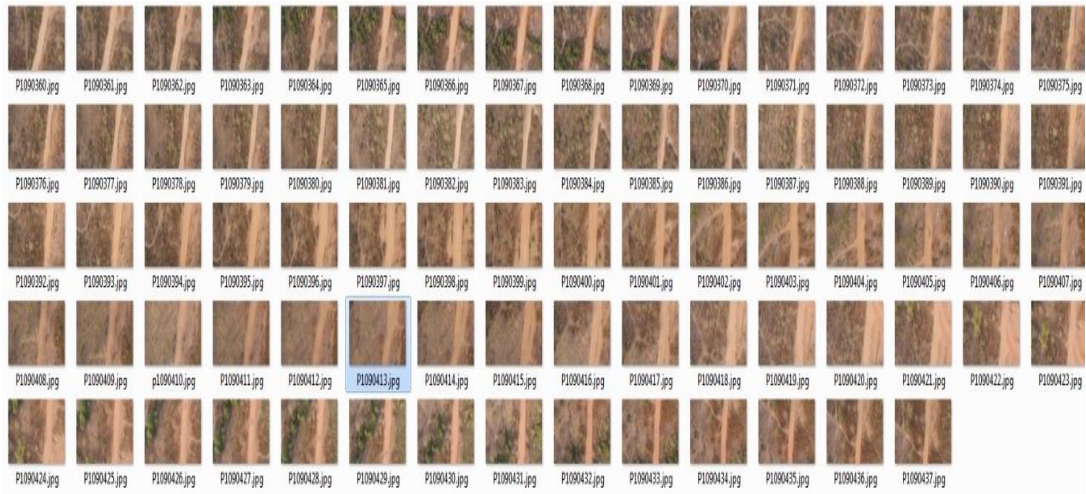
Fig.6.7 De-compilation of Video Streams into 78 Frames

Fig.6.8 show the result of interest points matching between two overlapping frames (first and the second frame) using the SURF algorithm. A total of 20516 tie points were successfully matched within 78 frames with 72983 projections. See appendix IIA for detailed survey data.
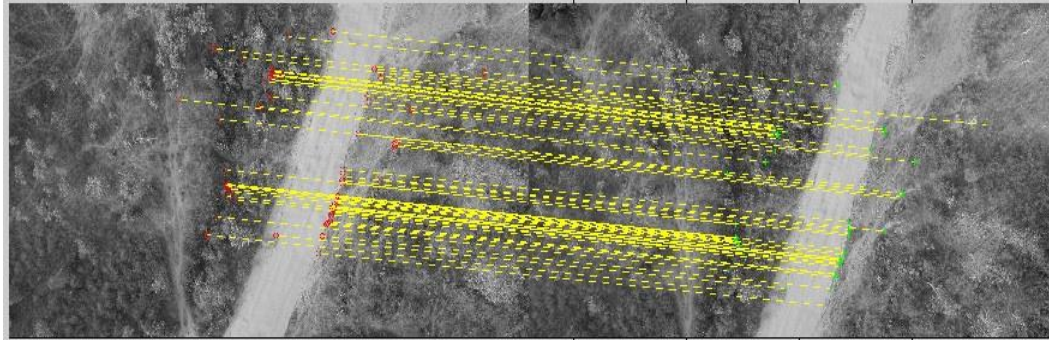
Fig.6.8 Conjugate points between two overlap frames (outliers removed)

Fig.6.9 shows the image residuals of the camera, which was the outcome result of the camera calibration.
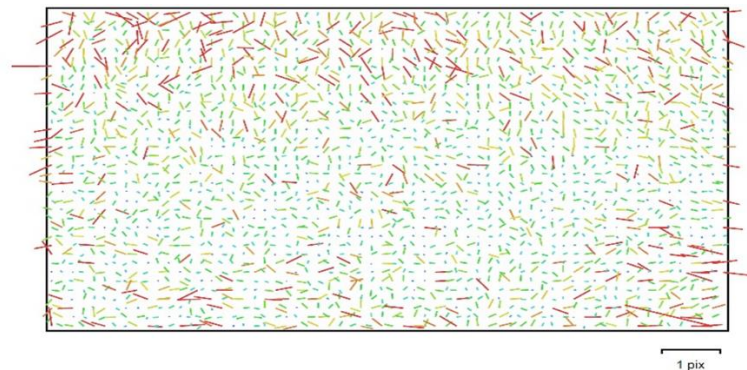
Fig.6.9 Image residuals for video frames

It can be observed from Fig.6.9, that the residuals are randomly distributed in magnitude and direction. This implies that there is no obvious systematic error in the residuals. The root mean square residual is 0.672782 pix.

Through the relative orientation stereo model the average camera location error was computed (Table 6.2). See also appendix IIB for detailed report on the camera locations and error estimates.

X error (m)	Y error (m)	Z error (m)	Total error (m)
8.820588	5.003659	3.798771	10.829129

Table 6.2.Average Camera Location Error

The on board GPS/INS provides the real time UAV's attitude information (Table 6.3), which is used in the calculation of the Boresight matrix by adding the GPS/INS readings to the result of the absolute bundle adjustment.

On-board GPS/INS Attitude Readings	Row (ω)	Pitch (ϕ)	Yaw (κ)
	0.0607932	0.00986701	1.0098690

Table 6.3 UAV Attitude Information

The result from the DLT model and the bundle adjustment method to calculate the interior orientation parameters and exterior orientation parameters is shown in Table 6.3 and Table 6.4 respectively.

In Table 6.3, u_0 and v_0 are the coordinates of the principal point. Also the lens distortion is not considered here, this is because the DLT method does not give very accurate results, hence, the solved interior orientation parameters and exterior orientation parameters will be used as initial values in (6.1).

The computation of absolute orientation is automatic but not autonomous because it involves human interaction, where we are prompted to enter the values of the selected ground control points.

Row (ω)	-0.0524789
Pitch (ϕ)	0.0042136
Yaw (κ)	-1.037061
u_0	2409.78
v_0	2380.12
f_u	4836.75
f_v	4836.75

Table.6.4 DLT Model Results

Row (ω)	-0.0545097
Pitch (ϕ)	0.0087668
Yaw (κ)	-1.0087921
u_0	2420.61
v_0	1335.46
K_1	-0.0310589
K_2	0.0280943
K_3	-0.0188507
P_1	0
P_2	0

Table 6.5 Results from Self-calibration

In Table 6.5, (K_1, K_2, K_3) represents radial distortion coefficients while (P_1, P_2) represents the tangential distortion coefficients. The results of ortho-rectification of each video frame is given in appendix III.

After the ortho-rectification process, the ortho-images (corrected frames) are mosaicked together to form a map covering the test area, the result of the mosaicking of video frames is shown in Fig.6.10. The reconstructed digital elevation model obtained as a result of the mosaicking of video frames is also shown in appendix IV. This mosaic digital elevation model is then integrated with a virtual globe (Google Earth). The result of integration of mosaicked images and the Google Earth is shown in Fig.6.11.

Fig.6.10 Mosaicked Images Covering Test Area

Fig.6.11 Mosaicked images registered on Google earth for geo-referencing

It can be seen from the result of the merging in Fig.6.11, that the image points on the mosaicked images aligns well with the 6 GCPs on the Google earth map with a misalignment error of between 3-5 meters.

6.4 **Chapter Summary**

This chapter presented the procedural steps for near real-time ortho-rectification and geo-referencing of video streams obtained from a low cost UAV equipped with a multi-sensor system. A mathematical model and DLT algorithm, which is based on a photogrammetry model, was used to calculate the internal geometric parameters of the camera and the exterior orientation parameters of each video frame. The values obtained from the solved DLT equations were used as inputs in the stereoscopic pair image of the relative orientation to eliminate the x-parallax and y-parallax in the developed model.

The DLT values are also employed as initial values for the calibration process for the final bundle block adjustment, in which the IOPs and EOPs of each video frame were solved simultaneously at the point of exposure. A test field located in Abuja, Nigeria was used to evaluate our method. Video and telemetry data were collected for about twenty-five minutes, and they were processed using our method. The results obtained demonstrated that the geometric measurement of the test field from ortho-images is more accurate, when compared with those from original perspective images and also, the accuracy of the two-dimensional DSM obtained, when compared with the 6 control points in the digital globe has misalignment errors between 2 and 3 metres.

CHAPTER SEVEN

Integration of Object Detection and Geo-referencing Processes

7.1 **Introduction**

After the two independent experiments conducted for object detection using the correlation filter and ortho-rectification and geo-referencing of the object detected using the photogrammetry model. Another experiment was conducted which integrated the two processes of object detection and geo-referencing.

In monitoring of a buried oil pipeline right-of-way (ROW) using a remotely controlled low-cost UAV equipped with sensors, objects of interest lying on the buried oil pipeline ROW and which may act as a potential threat to the buried oil pipeline ROW must be detected and the exact geographical location of the object detected must be determined. This will enable the pipeline operator to know the exact location where to deploy security operatives in order to apprehend the third party intruder to the pipeline ROW. The basic steps and results of experiments performed for the developed system for monitoring of buried oil pipeline ROW is illustrated in the following sections. The general work flow of the developed system is shown in Fig. 7.1.
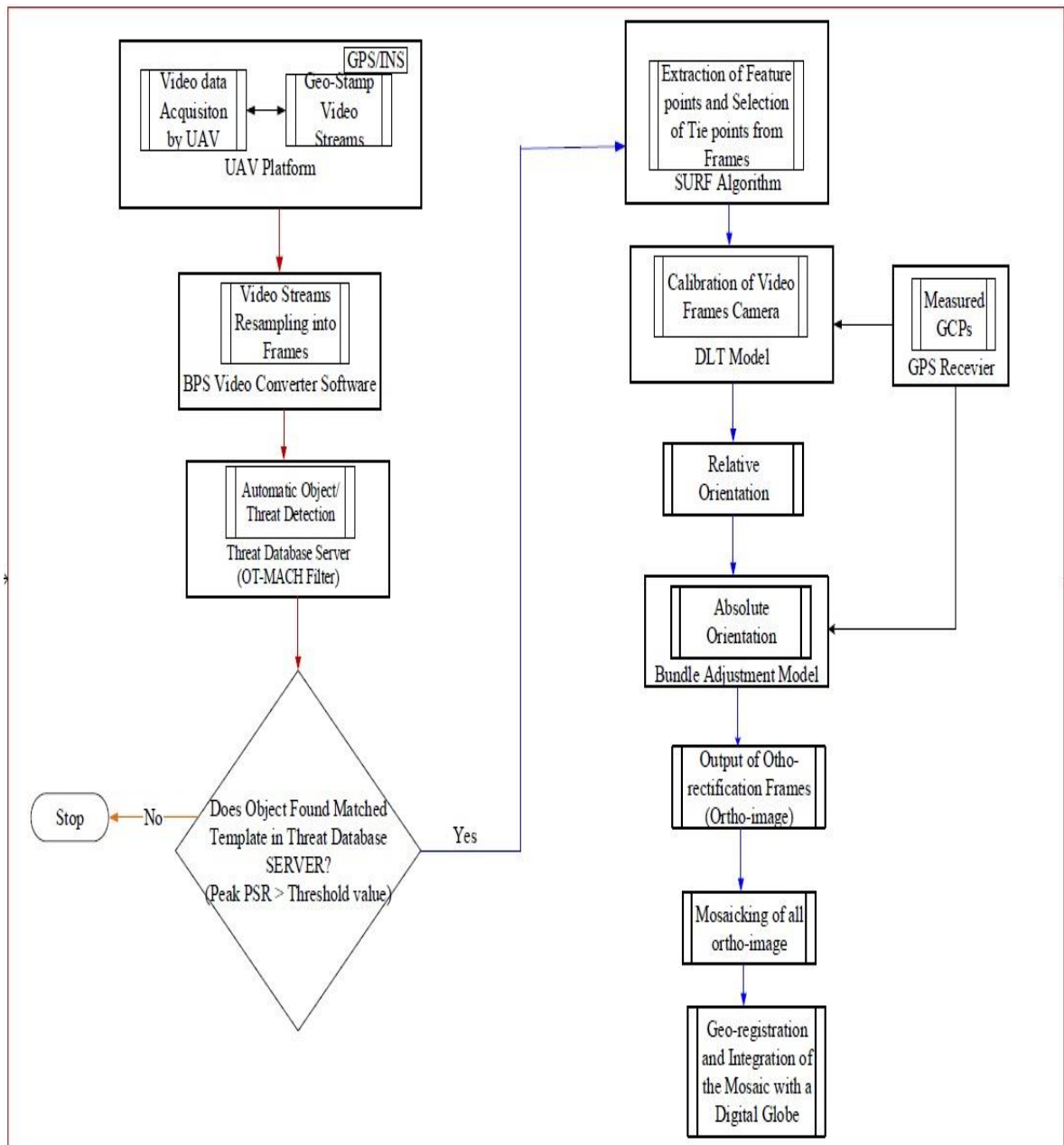
Fig. 7.1 Work Flow of the Developed Monitoring System

7.2 **Object of Interest Detection Process**

The first step in monitoring of the pipeline ROW is to detect objects of interest which may act as a threat to the pipeline ROW. The three basic steps for detecting objects of interest in a video streams are:

7.2.1 **Data Acquisition**

Experiments were conducted on a calibrated control field, located at the federal capital territory Abuja, Nigeria. Where video streams and telemetry data were collected with the aid of a Spreading Wings S800 UAV, which is equipped with GPS/INS and has the capability to geotag (time and position tagged) video frames at the exact time of exposure.

The DJI S800 UAV transmits real-time flight data and video feeds with a 5.8G video downlink via a radio signal to the ground station where all other processing takes place.

The ground station consists of a portable PC and three servers: the image processing server, the threat database server and the base map server. The configuration of each server is the same and is given in Table 7.1.

Operating System	Window 7 Enterprise, Service Pack 1 64-bit Operating System
Processor	Inter® Core (TM) i7-3770 CPU@ 3.40GHz
RAM	8.00GB

Table 7.1 Server Configuration.

7.2.2 **Resampling of Video Streams**

The video streams captured by the UAV are resampled into individual frames (JPG format) at the rate of one frame per second of sizes 256 x 256 using the BPS video converter 1.4 software in the image processing server. Conversion rate of one frame per second was chosen to allow for an image forward overlap of 65% between sequential frames to allow for stereoscopic reconstruction of image points during bundle adjustment. Breaking the video streams into individual frames allows for each image in the video frame to be processed individually with the aim of detecting objects of interest in the frames with a very small number of false detections. Fig. 7.2 shows the result of video streams broken down into 98 frames.
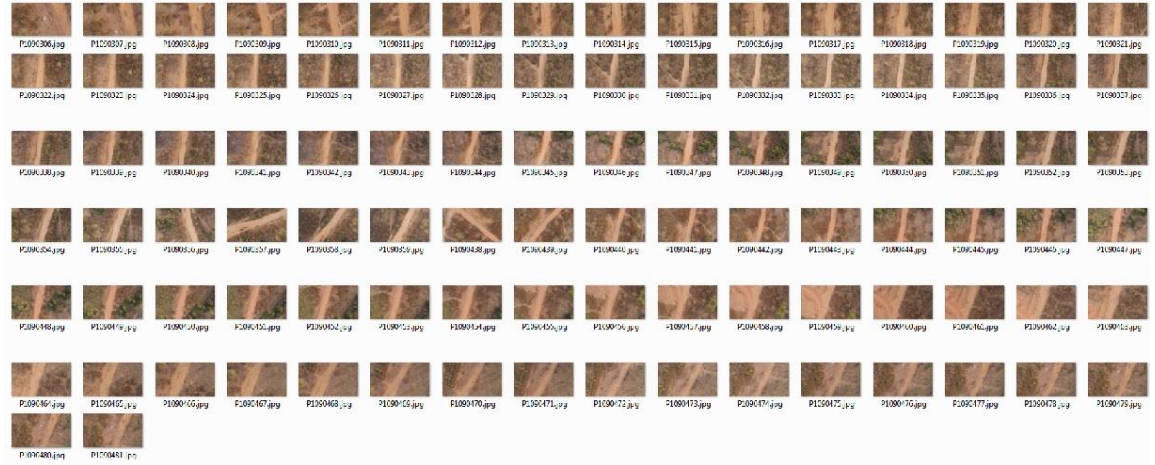
Fig. 7.2 Video streams converted into 98 frames.

7.2.3 **Cross Correlation**

After the video streams had been resampled into individual frames, the video frames are then transferred into the threat database server, where each frame is cross correlated with each trained (image template) filter in the threat database server. Whenever the cross correlation output peak gives a PSR value that is greater than the chosen PSR threshold value of 20.01. Then there is a match between the video frame and one of the trained in-class images. Hence a target object has been detected. The PSR value is given by equation (4.5).

Fig. 7.3 shows the example results of trained images, which have been derived from CAD models of a selected car. These training sets are derived from different viewing angles (0^0 – 360^0) of the UAV's camera.

Fig. 7.4a shows the result obtained from cross correlation of the target image with the trained images using the OT-MACH correlation filter. The image at the upper left most corner is the 64 x 64 filter image (template) that correlates with the target image.

Fig. 7.3 Trained images from a car CAD Model under different viewing angles ($0^0 – 360^0$).
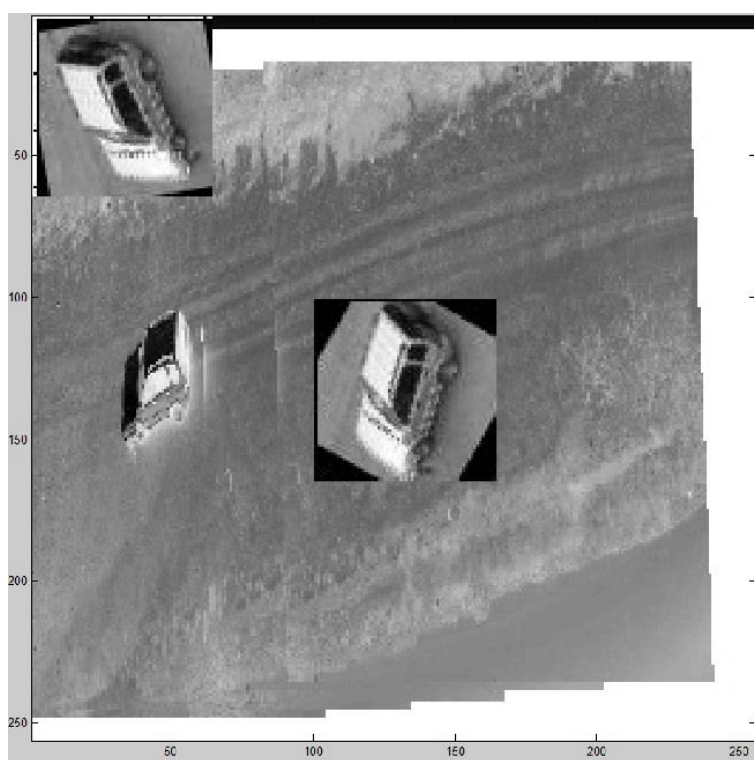
(a)
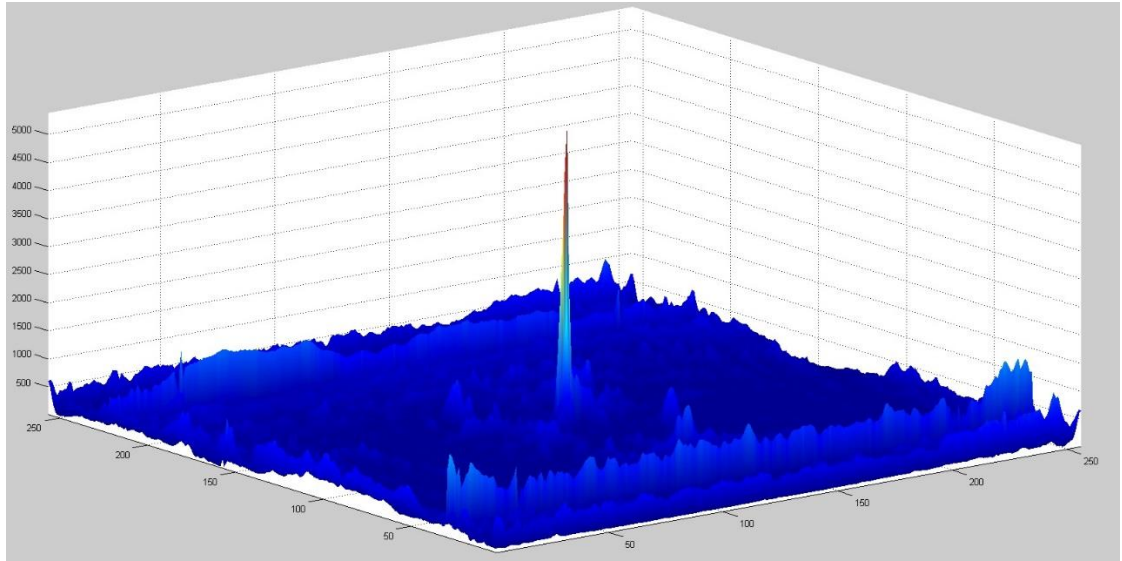
(b)

Fig. 7.4 (a) Result image showing the viewing angle of the UAV camera and the corresponding template image (b) Result of Cross Correlation output, the sharp peak indicate an object is detected.

In Fig. 7.4b the PSR threshold value was set to 17.01. This means if after cross correlation the output peak has a PSR value less than 17.01, then the object found is not a threat, it only has some features similar to the object of interest. However, all video frames which produce a correlation output peak with a PSR value greater than the chosen PSR threshold value (17.01) implies that objects of interest are detected on them. We chose a small PSR value to enable the filter to detect any little features of the object of interest that may be present in any frame, since it is possible to have the object features of interest in more than one frame. However, care must be taken so as not to choose a PSR value that is too low or too high as threshold, in order to prevent false detection or passing objects of interest without detecting them. Hence a lot of tests need to be conducted before a value is chosen as threshold for the PSR. The PSR value of the output peak in Fig. 7.4b was found to be 625. 55.

7.3 **Ortho-rectification of Video Frames**

The next step after objects of interest have been detected from the video frames is the ortho-rectification of video frames. This process is performed in the base map server, which consists of the SURF algorithm, the Agisoft Photoscan and Google Earth software. The X, Y, Z coordinates position data of each image is given in appendix V. The process of ortho-rectification and results obtained are explained in the following subsection.

7.3.1 **Extraction of Feature Points and Tie Points Search**

After the detection of objects in the video frames, the frames are transferred from the threat database to the base map database. The feature point based algorithm (SURF) is used for the extraction and matching of feature points from adjacent video frames.

Tie points are conjugate points on two images in the direction of the flight of the UAV. The SURF algorithm automatically searches for tie points between image frames. The algorithm exploits the fact video frames contain imagery that are very close to each other resulting into a small change of images between frames. This means any feature points found on an image in a frame will have image coordinates in close proximity to the coordinates of the tie points found on the image in the frame directly preceding and succeeding it. Hence the problem of occlusion is greatly reduced. Fig. 7.5 shows the result of putatively matched points including the outliers between two successive frames. Only the 20 feature points that best matched were selected while the outliers were discarded, the result of matching of the matched points inliers is shown in Fig. 7.6. The SURF algorithm m file used for the extraction and matching process is given appendix VI.
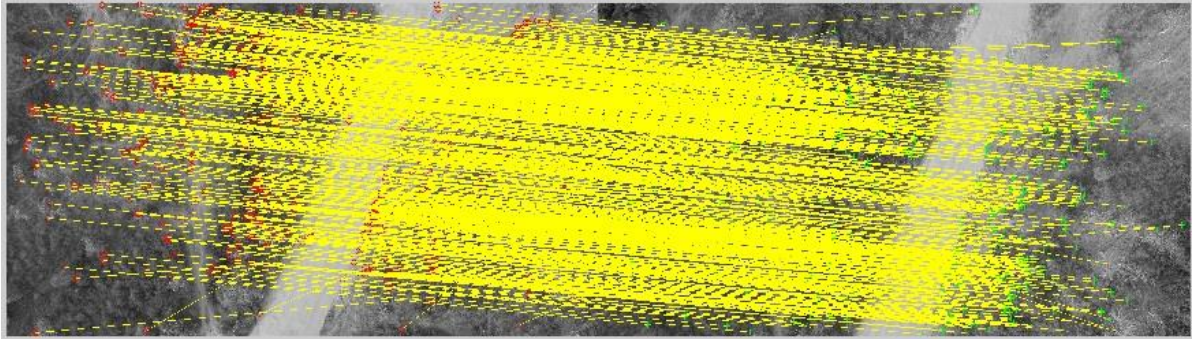
Fig. 7.5 Putatively Matched Points Including the Outliers between two successive Frames
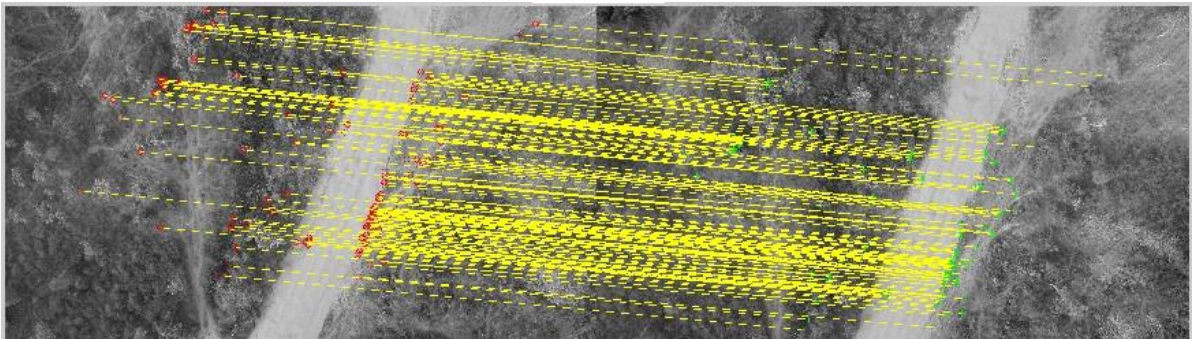
Fig. 7.6 Matched points (Inliers only) between two successive Frames

Tie points are used in the relative orientation step for estimating the image frame base. In an ideal situation where the flight line is a straight line in the x direction of the image planes, the sequential tie points will be found in positions with the same image coordinates with other image plane. But due to continuous jerking of the low-cost UAV along the flight line, the attitude and position of the camera is continuously changing. This produces some image point off-set on the image plane. This is corrected in the remaining steps of ortho-rectification in the following section. The survey data for this experiment is shown in appendix VIIA

7.3.2 **Camera Calibration**

The next step of ortho-rectification is the calibration of each camera video frame. Due to the off-set of the image point position on the video frame which is caused by lens distortion and other systematic and non-systematic distortion of the camera, there is the need to calibrate each camera video frame to correct these errors and obtain the true position of the image point on the frame. This involves finding the focal length, principal coordinates and lens distortion (interior orientation parameters) and the exterior orientation parameters of video camera frame using a DLT model. The results obtained for the calibration of the camera are given in Table 7.2 and the residual of the camera in Fig. 7.6. The DLT model results are used as initial values during iteration process during the final bundle adjustment.

Row (ω)	-0.0524789
Pitch (ϕ)	0.0042136
Yaw (κ)	-1.037061
u_0	2409.78
v_0	2380.12
f_u	4836.75
f_v	4836.75
K1	-0.025369
K2	0.0240245
K3	-0.014159

Table 7.2 Result of Video frame Calibration

where u_0, v_0 are the coordinates of the principal point, f_u and f_v are the focal lengths in the x and y direction respectively.
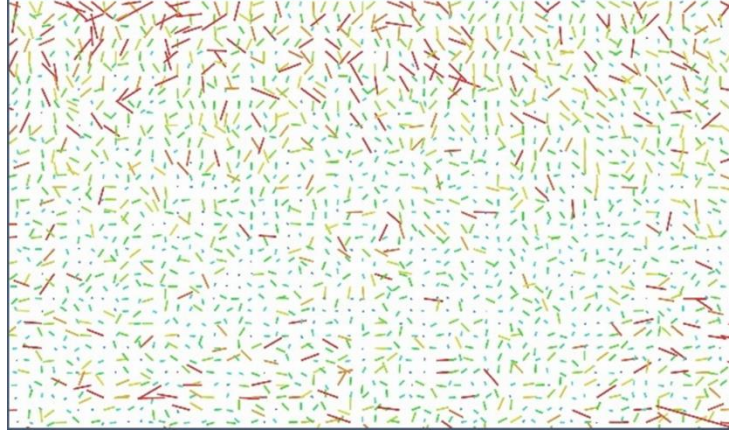
Fig. 7.6 Image residuals for DMC-GH3 Camera is approximately 1 pixel

7.3.3 **Recreation of Position and Angular Attitude of Video Frames**

After the calibration of video frames, the frames are further adjusted by forming a stereo pair image by placing video frames beside each other and recreating the relative position and angular attitudes of one image frame with respect to the one beside it, so that the baselines of all the image frames lies on the same horizontal plane using an affine coordinate transform. The output result of this process is a tie point in the first image frame and the tie point on the last image frame holding the entire sequence of images in the flight strip together. The results obtained from the experiment are shown in Table 7.3. A detailed camera locations and error estimates is also shown in appendix VIIB

X error (m)	Y error (m)	Z error (m)	Total error (m)
7.087607	4.113576	3.460632	8.895598

Table 7.3 Average Camera Position Error

7.3.4 **Bundle Adjustment for the Entire Flight**

The final step under ortho-rectification is the process of levelling and orienting the stereo model form in the relative orientation process in the previous section to the ground control system by adjusting the entire ray in the stereo model with reference to six ground control points through a self-calibration process. The results obtained from the self-calibration are given in Table 7.4. Where (K_1, K_2, K_3) represents radial distortion coefficients while (P_1, P_2) represents the tangential distortion coefficients. The results obtained from ortho-rectification of each video frame is given in appendix VIII.

After the ortho-rectification of video frame images, the ortho-images formed are mosaicked together to form a digital elevation model covering the test area (appendix IX). Fig. 7.7 shows the result of mosaicking of ortho-images from the experiment.

Row (ω)	-0.0547097
Pitch (ϕ)	0.0084768
Yaw (κ)	-1.0087921
u_0	2470.61
v_0	1435.46
K_1	-0.0314587
K_2	0.0281640
K_3	-0.0187615
P_1	0
P_2	0

Table 7.4 Results from Bundle Adjustment of the Entire Flight

Fig. 7.7 Result of Ortho-images Mosaicking

7.4 Geo-referencing

The final step in monitoring of the buried pipeline ROW is the geo-referencing of the object detected on the pipeline ROW. This is necessary in order to pin point the exact geographical location of the object detected for a quick response of pipeline security to the place intruded. The process of geo-referencing takes place in the base map server which contains the virtual globes.

7.4.1 Virtual Globe

There are different types of web based 3D geo-information services that are based on virtual globes that are available today. The two prominent ones are the Google Earth and the Microsoft Virtual Earth. These virtual globes have the capability to geo-register and integrate a very large amount of geo-spatial content like ortho-mosaic, terrain model and 3D objects.

7.4.2 Ortho-mosaic registration and Integration

The ortho-mosaic formed after ortho-rectification of each video frame is geo-registered and integrated with a virtual globe (Google Earth). This allows for real-time geo-referencing of the ortho-mosaic that contains the detected object. Fig. 7.8 shows the results of the integration and geo-registration of the ortho-mosaic with the Google Earth with a misalignment error of between 1.5 and 2 meters. While Fig. 7.9 shows the exact location of the object detected on the ortho-mosaic on the Google Earth.

The object detected is found to be at Latitude of 8.902066^0 and Longitude of 7.481951^0.

Fig. 7.8 Geo-registration of Ortho-mosaic into Google Earth
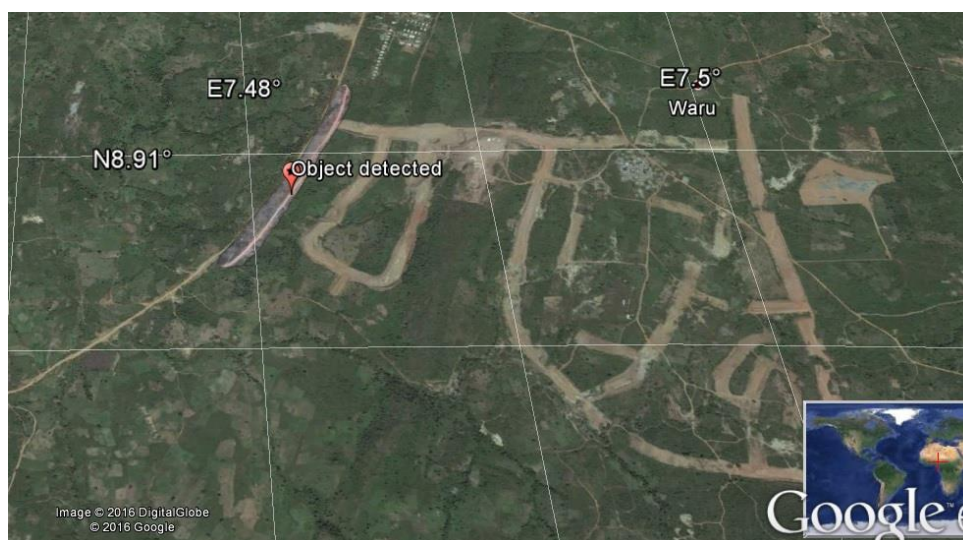
Fig 7.9 Object Detected at Lat. 8.902066^0, Long.7.481951 on Google Earth

Once the geographical location of the object detected has been found, an alert containing location of the object is then sent to the pipeline security personnel by the operator for necessary action.

7.5 **Chapter Summary**

This chapter illustrates the procedure and results of the experiment conducted for the developed system. Where both the object detection process and geo-referencing processes are integrated for testing the developed system for monitoring and protecting the buried oil pipeline ROW from oil theft or deliberate sabotage.

Objects which may act as a potential danger to the pipeline ROW were used to train the system for object detection, while a photogrammetry model was used for ortho-rectification and geo-referencing of video imagery captured by a low cost multi-sensors UAV. The resulting ortho-mosaic was then merged with a digital globe for geo-referencing of the detected objected contained on the ortho-mosaic.

CHAPTER EIGHT

Conclusion and Recommendation

8.1 Buried Oil Pipeline Monitoring

Oil carrying pipelines are buried in utility rights-of-way, which traverse remote environments. These rights-of-way, which are usually maintained by the pipeline operators, are sometimes damaged by construction equipment owned by a third party, with the intention of oil theft, vandalism and terrorist attacks. Damage to the pipeline rights-of-way are caused by digging or drilling along the pipeline rights-of-way can be devastating, sometimes causing explosions and death and, loss of millions of dollars by pipeline operators.

Many approaches have been used for protecting and monitoring of buried oil pipeline ROW, these approaches include the use of security force ground patrol of the pipeline ROW, the installation of sensors along the pipeline ROW and the use of helicopters for aerial patrol of the pipeline ROW. However, the frequent reports of pipeline vandalism and oil theft are pointers to the inadequacy of these methods used for protecting buried oil pipelines.

In this research, we were able to develop an automated image analysis system with the aid of a UAV equipped with sensors, to detect potential hazards (construction equipment, person, car, etc.) along the pipeline rights-of-way and also to pin-point the exact geographical location of the detected objects and then alert the pipeline operator.

Object detection is one of the major aspects of a remote monitoring system. Towards this purpose, we investigated different detection techniques, which include the feature based (Harris corner, SIFT and SURF) methods and area based (cross correlation, least squares)

detection methods. The cross correlation method was found to be the most suitable for this research.

A correlation filter was specially designed for the task of object detection and an optimal trade-off between the parameters of the filter were used to improve the performance of the pattern detection tasks.

After the detection of the objects of interest, the developed system must be able to pin point the precise geographic location of the object detected from a wide range of video frame images. This was achieved with the aid of a photogrammetry model.

Since data acquisition in this research was done by the use of a UAV equipped with camera and sensors; Video frames acquired at the UAV platform suffer from of a lot of geometric distortions. It was observed that these geometric distortions cannot be corrected to a reasonable accuracy without the use of ground control points (GCPs), especially when dealing with systematic and/or non-systematic errors and images covering rough terrain. Hence photogrammetric techniques with the aid of least squares bundle adjustment were used for correcting the aforementioned distortions to the video images acquired by the UAV platform.

The photogrammetric technique was able to produce the precise geographic location of the object detected from a wide range of images. This means, any measurement taken with the aid of a photogrammetry processed image reflects a measurement taken on the ground. Hence there is no need to constantly go to the test field to make measurements, since with the use of photogrammetry tools, information obtained from imagery is the same as that obtained from the field.

Our method was tested by performing three experiments on test fields located in Nigeria and United Kingdom. The first experiment was a test experiment for object detection conducted in a test field at Gatwick, United Kingdom. Where objects that may act as a potential threat to the pipeline ROW were used to train the developed system for recognition using a correlation filter.

Also another independent test experiment was conducted at Abuja, Nigeria to test our method for ortho-rectification and geo-referencing by calibrating the test field and using a photogrammetry model for ortho-rectification of video imagery. The ortho-image formed from the ortho-rectification of the video frame imagery are mosaicked and then integrated with a digital globe for geo-referencing.

The last experiment was also conducted in Abuja, Nigeria. This is the main experiment on the developed system for monitoring buried oil pipeline ROW for third party incursion. In this experiment the process of object detection is integrated with the process of ortho-rectification and geo-referencing. In other words, after the first process detects an object, which the system has been trained to recognise. The detection system then hands over to the ortho-rectification and geo-referencing system. Where the detected object geographical location is determined and an alarm is sent to the pipeline security personnel for necessary action.

The overall objectives of this thesis focus on developing an automated system that will protect the pipeline from damage caused by oil pipelines vandals and saboteurs. Such a technology would result in a safer and more reliable oil carrying pipeline systems and solve a long time standing problem of the natural oil and gas industry.

The prevention of third party incursion onto the buried pipeline right-of-way will go a long way in reducing pipeline service interruptions and cost of frequent repairs of the pipeline systems.

8.2 Future Work

While much work has been done, there is still room for improvement to help address the following outstanding limitations:

- To increase object detection capability of the correlation filter, future works need to be focused on designing correlation filters that are sensitive to the colour of objects. Colours are known to provide additional information about an object and can easily be used to add to the criteria for object detection in correlation filters.

- Also, future research may focus on how to reduce training of correlation filters, which is based on the number of constrained variants of the target image. A solution to this problem, will go a long way in reducing computational time and memory space needed for the detection process.

- Future works may also focus on designing UAVs specifically for the purpose of aerial monitoring where most of the image processing process will take place on-board the aircraft.

- Since the UAV flies at low altitude, future works should focus on equipping the UAV with sense and avoid sensors to prevent it from collision with flying objects (such as balloons, birds, etc.).

- Moreover, future research may focus on a UAV payload consisting of a combination of electro-optical sensors which can provide: day light surveillance, the infrared sensors for night time surveillance and synthetic aperture radar which is not affected by bad weather. The data obtain from these sensors can be processed and integrated to provide better information or information that could not be obtained when a single type of sensor is used.

APPENDIX I

Results for the test experiments for ortho-rectification and geo-referencing is as follows:

		X	Y	Z
Cameras		Longitude	Latitude	Altitude
☑	P109036...	7.478762	8.905067	577.014400
☑	P109036...	7.478832	8.905163	577.860200
☑	P109036...	7.478898	8.905252	578.186000
☑	P109036...	7.478992	8.905383	578.624900
☑	P109036...	7.479057	8.905477	578.847000
☑	P109036...	7.479122	8.905572	579.624700
☑	P109036...	7.479155	8.905617	580.007400
☑	P109036...	7.479248	8.905765	580.160400
☑	P109036...	7.479283	8.905815	580.347100
☑	P109036...	7.479353	8.905918	580.103700
☑	P109037...	7.479428	8.906032	578.675200
☑	P109037...	7.479492	8.906153	577.656600
☑	P109037...	7.479605	8.906307	578.253400
☑	P109037...	7.479690	8.906410	578.273200
☑	P109037...	7.479732	8.906463	577.914500
☑	P109037...	7.479853	8.906610	578.256800
☑	P109037...	7.479895	8.906658	578.773500
☑	P109037...	7.479968	8.906775	578.796200
☑	P109037...	7.480083	8.906957	578.113100
☑	P109037...	7.480162	8.907068	578.147400
☑	P109038...	7.480242	8.907173	579.027700
☑	P109038...	7.480328	8.907272	579.018600
☑	P109038...	7.480413	8.907372	578.777900
☑	P109038...	7.480487	8.907488	578.544800
☑	P109038...	7.480520	8.907550	577.053200
☑	P109038...	7.480588	8.907672	575.381600
☑	P109038...	7.480703	8.907843	575.260100
☑	P109038...	7.480738	8.907902	575.719000
☑	P109038...	7.480812	8.908010	575.820500
☑	P109038...	7.480887	8.908133	576.775700
☑	P109039...	7.481008	8.908310	578.418400
☑	P109039...	7.481093	8.908418	578.679000
☑	P109039...	7.481137	8.908467	578.632300
☑	P109039...	7.481222	8.908572	579.038200
☑	P109039...	7.481340	8.908743	578.591600
☑	P109039...	7.481378	8.908798	578.371400
☑	P109039...	7.481462	8.908908	579.455600
☑	P109039...	7.481577	8.909083	580.395100
☑	P109039...	7.481652	8.909192	579.351300
☑	P109039...	7.481728	8.909305	578.235300
☑	P109040...	7.481810	8.909427	577.670400
☑	P109040...	7.481848	8.909482	577.789500
☑	P109040...	7.481977	8.909660	577.885700
☑	P109040...	7.482057	8.909768	578.361700

Each Video Frame Geographical Location Data

☑	📥	P109040...	7.482097	8.909820	579.094600
☑	📥	P109040...	7.482217	8.909997	578.542600
☑	📥	P109040...	7.482252	8.910060	578.284600
☑	📥	P109040...	7.482353	8.910248	579.178200
☑	📥	P109040...	7.482417	8.910383	580.004800
☑	📥	P109040...	7.482447	8.910452	579.602400
☑	📥	P109041...	7.482578	8.910702	578.745300
☑	📥	P109041...	7.482668	8.910892	580.597500
☑	📥	P109041...	7.482725	8.911027	581.542200
☑	📥	P109041...	7.482788	8.911155	580.384500
☑	📥	P109041...	7.482857	8.911272	580.008500
☑	📥	P109041...	7.482887	8.911328	580.197100
☑	📥	P109041...	7.482968	8.911542	580.252100
☑	📥	P109041...	7.482993	8.911613	580.715500
☑	📥	P109041...	7.483037	8.911690	580.715500
☑	📥	P109042...	7.483147	8.911933	581.359600
☑	📥	P109042...	7.483173	8.912005	580.839200
☑	📥	P109042...	7.483237	8.912125	580.293800
☑	📥	P109042...	7.483305	8.912243	579.515600
☑	📥	P109042...	7.483368	8.912367	578.284600
☑	📥	P109042...	7.483460	8.912557	578.290000
☑	📥	P109042...	7.483492	8.912622	578.609600
☑	📥	P109042...	7.483555	8.912758	577.426000
☑	📥	P109042...	7.483635	8.912962	575.664000
☑	📥	P109042...	7.483693	8.913093	574.397600
☑	📥	P109043...	7.483725	8.913158	574.042000
☑	📥	P109043...	7.483812	8.913352	574.057400
☑	📥	P109043...	7.483872	8.913477	573.701200
☑	📥	P109043...	7.483935	8.913602	572.784000
☑	📥	P109043...	7.483967	8.913662	572.356300
☑	📥	P109043...	7.484033	8.913788	570.970600
☑	📥	P109043...	7.484095	8.913938	569.363300
☑	📥	P109043...	7.484117	8.913985	570.336700
☑	📥	p109041...	7.482518	8.910577	578.912700

Each Video Frame Geographical Location Data

Survey Data
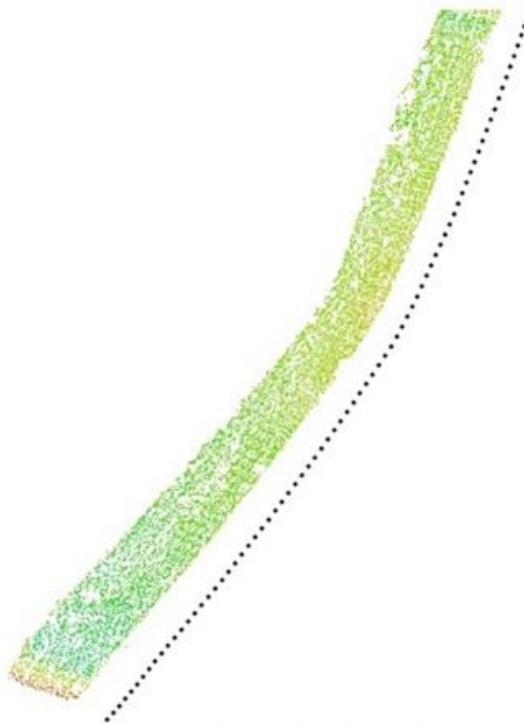
Camera locations and image overlap.

Number of images:	78	Camera stations:	78
		Tie-points:	20516
Ground resolution:	0.0251572 m/pix	Projections:	72983
Coverage area:	0.0454991 sq km	Error:	0.672782 pix

Camera Model	Resolution	Focal Length	Pixel Size	Precalibrated
DMC-GH3 (12 mm)	4608 x 2592	12 mm	3.77706 x 3.77706 um	No

Camera Locations

		8.69579 m
		6.95663 m
		5.21747 m
		3.47831 m
		1.73916 m
		0 m
		-1.73916 m
		-3.47831 m
		-5.21747 m
		-6.95663 m
		-8.69579 m
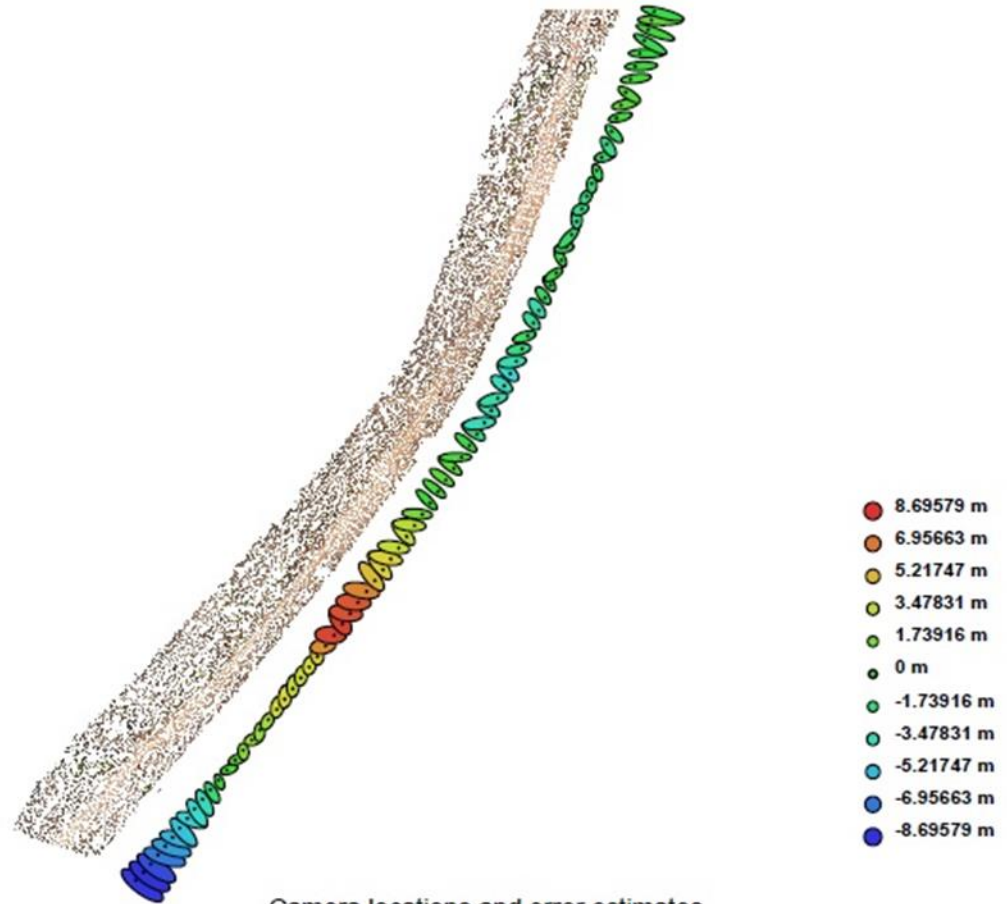
Camera locations and error estimates.
Z error is represented by ellipse color. X,Y errors are represented by ellipse shape.
Estimated camera locations are marked with a black dot.

X error (m)	Y error (m)	Z error (m)	Total error (m)
8.820588	5.003659	3.798771	10.829129

Average camera location error.

Results of Ortho-rectification of Each Video Frame

| Cameras	Long. est	Lat. est	Alt. est	Error (m)	Yaw est	Pitch est	Roll est	Error (deg)	Projections	Error (pix)
P109036...	7.478626	8.905166	568.372865	20.439107	27.855	10.339	44.510		817	0.696
P109036...	7.478705	8.905250	569.164392	19.047707	28.749	9.377	44.456		957	0.609
P109036...	7.478779	8.905334	570.104670	17.877910	29.091	9.190	44.397		900	0.685
P109036...	7.478854	8.905426	571.759994	17.282172	28.957	8.932	44.019		871	0.728
P109036...	7.478928	8.905516	572.828765	15.959574	29.074	8.850	44.050		863	0.655
P109036...	7.479004	8.905605	574.088080	14.515909	29.353	8.445	43.973		857	0.751
P109036...	7.479078	8.905700	574.977262	13.515002	29.367	8.166	43.979		936	0.627
P109036...	7.479154	8.905798	575.522792	11.932629	29.499	7.889	44.065		959	0.739
P109036...	7.479229	8.905895	576.248419	11.453067	29.535	7.658	43.902		949	0.749
P109036...	7.479311	8.905994	576.534072	10.171724	30.647	6.204	44.313		945	0.694
P109037...	7.479391	8.906091	576.993810	7.906115	30.472	6.094	44.526		946	0.746
P109037...	7.479476	8.906189	577.864122	4.323177	30.238	5.870	44.440		959	0.682
P109037...	7.479562	8.906291	578.771995	5.098639	30.179	5.622	44.396		948	0.732
P109037...	7.479647	8.906394	579.439655	5.150570	30.064	5.456	44.603		879	0.652
P109037...	7.479728	8.906494	580.156612	4.076356	29.987	5.327	44.672		836	0.588
P109037...	7.479816	8.906599	581.003730	5.118324	29.710	5.260	44.578		820	0.712
P109037...	7.479901	8.906702	581.076246	5.342200	29.827	4.805	44.837		813	0.682
P109037...	7.479987	8.906806	581.569167	4.902374	29.410	5.047	44.741		896	0.655
P109037...	7.480073	8.906910	582.189426	6.639936	29.239	4.951	44.474		919	0.645
P109037...	7.480155	8.907014	582.529448	7.444838	29.003	4.918	44.566		954	0.693
P109038...	7.480238	8.907121	582.715824	6.856267	28.798	4.797	44.692		966	0.723
P109038...	7.480320	8.907233	582.901487	5.875581	28.964	4.119	44.761		918	0.712
P109038...	7.480406	8.907342	582.995737	5.420577	28.791	3.883	44.877		875	0.715
P109038...	7.480490	8.907451	583.224456	6.274432	28.589	3.704	44.691		882	0.821
P109038...	7.480574	8.907560	583.395972	8.784907	28.789	3.104	44.778		864	0.787
P109038...	7.480662	8.907667	583.601811	11.543029	28.482	2.806	44.871		880	0.738
P109038...	7.480745	8.907781	583.609686	11.738989	28.163	2.701	44.946		889	0.729
P109038...	7.480825	8.907889	583.600714	12.409583	28.320	2.447	45.136		909	0.818
P109038...	7.480908	8.908002	583.552354	13.140016	28.140	2.297	45.152		916	0.702
P109038...	7.480995	8.908113	583.351417	13.782670	28.112	1.940	45.361		922	0.695
P109039...	7.481080	8.908225	583.555647	13.316100	27.723	1.929	45.261		893	0.648
P109039...	7.481162	8.908336	583.334400	12.720640	27.539	1.686	45.386		903	0.774
P109039...	7.481242	8.908445	583.260281	12.675307	27.645	1.199	45.248		930	0.718
P109039...	7.481324	8.908558	582.733171	11.956089	27.378	0.955	45.418		933	0.655
P109039...	7.481405	8.908674	582.173425	11.072798	27.231	0.801	45.436		929	0.599
P109039...	7.481487	8.908788	581.726998	12.431459	27.134	0.547	45.458		905	0.557
P109039...	7.481566	8.908900	581.001700	11.567906	26.746	0.589	45.713		892	0.604
P109039...	7.481646	8.909016	580.459386	10.613999	26.496	0.587	45.549		939	0.577
P109039...	7.481728	8.909133	579.772360	10.638318	25.853	0.422	45.679		971	0.543
P109039...	7.481808	8.909251	579.190793	10.591329	25.602	0.296	45.659		1039	0.530
P109040...	7.481889	8.909368	578.524569	10.864098	25.248	0.104	45.626		1102	0.545
P109040...	7.481968	8.909483	578.021715	13.128710	24.991	-0.132	45.572		1020	0.567
P109040...	7.482048	8.909602	577.368368	10.146964	24.631	-0.284	45.528		947	0.531

☑ 📥 P109040...	7.482127	8.909719	576.334231	9.675133	24.510	-0.594	45.770	937	0.530
☑ 📥 P109040...	7.482203	8.909834	575.604060	12.337839	24.248	-0.908	45.699	962	0.521
☑ 📥 P109040...	7.482275	8.909960	575.204842	8.281768	23.860	-0.983	45.511	1014	0.514
☑ 📥 P109040...	7.482340	8.910085	575.232629	10.550235	22.504	-0.135	45.469	948	0.494
☑ 📥 P109040...	7.482406	8.910211	575.547774	7.964124	21.137	0.732	45.466	929	0.526
☑ 📥 P109040...	7.482469	8.910336	575.854678	8.855500	20.578	0.871	45.336	994	0.550
☑ 📥 P109040...	7.482530	8.910462	576.327116	9.757090	20.031	0.954	45.278	1055	0.722
☑ 📥 P109041...	7.482648	8.910722	577.586647	8.082427	17.675	2.019	45.083	1005	0.654
☑ 📥 P109041...	7.482706	8.910852	578.091532	6.523684	16.807	2.361	45.076	972	0.579
☑ 📥 P109041...	7.482762	8.910981	578.575520	7.143013	16.381	2.405	45.017	925	0.746
☑ 📥 P109041...	7.482825	8.911109	579.224502	6.549705	15.926	2.255	44.829	936	0.723
☑ 📥 P109041...	7.482882	8.911238	579.318027	4.717531	15.458	2.575	44.974	942	0.812
☑ 📥 P109041...	7.482940	8.911364	579.916686	7.030826	15.190	2.357	44.876	961	0.668
☑ 📥 P109041...	7.482994	8.911494	579.798910	5.981487	14.831	2.309	44.986	983	0.737
☑ 📥 P109041...	7.483046	8.911623	579.594015	5.964345	14.196	2.368	44.945	1009	0.843
☑ 📥 P109041...	7.483098	8.911754	579.335183	9.897654	14.019	2.006	45.143	991	0.762
☑ 📥 P109042...	7.483149	8.911888	579.032370	5.581799	13.615	1.985	45.072	910	0.701
☑ 📥 P109042...	7.483203	8.912017	578.822816	4.055064	13.135	1.837	45.079	851	0.633
☑ 📥 P109042...	7.483252	8.912147	578.515357	3.453999	12.901	1.849	45.106	869	0.743
☑ 📥 P109042...	7.483304	8.912276	577.982803	3.913083	12.598	1.823	45.219	879	0.765
☑ 📥 P109042...	7.483353	8.912408	577.599292	4.949006	12.282	1.485	45.130	972	0.661
☑ 📥 P109042...	7.483407	8.912543	576.967296	6.200899	12.099	1.201	45.341	1060	0.623
☑ 📥 P109042...	7.483457	8.912676	576.505086	7.409192	11.802	0.947	45.270	1019	0.627
☑ 📥 P109042...	7.483508	8.912807	576.319523	7.533239	11.587	0.817	45.359	994	0.641
☑ 📥 P109042...	7.483558	8.912943	575.784175	8.756247	10.988	0.879	45.413	937	0.654
☑ 📥 P109042...	7.483604	8.913074	575.372507	10.123296	10.781	0.692	45.370	991	0.743
☑ 📥 P109043...	7.483653	8.913205	574.776723	9.521197	10.102	0.918	45.544	1031	0.706
☑ 📥 P109043...	7.483702	8.913339	574.368182	12.196572	9.399	1.235	45.288	1017	0.636
☑ 📥 P109043...	7.483750	8.913471	573.607526	13.429849	9.140	1.144	45.454	963	0.585
☑ 📥 P109043...	7.483797	8.913603	572.674505	15.205513	8.767	0.948	45.627	913	0.667
☑ 📥 P109043...	7.483842	8.913734	571.654961	15.907937	8.840	0.740	45.700	892	0.748
☑ 📥 P109043...	7.483881	8.913844	571.161233	17.815094	8.749	0.017	45.315	900	0.666
☑ 📥 P109043...	7.483916	8.913943	570.534911	19.699294	8.674	0.152	45.243	953	0.560
☑ 📥 P109043...	7.483945	8.914031	569.696375	19.574444	8.399	-0.257	45.316	799	0.629
☑ 📥 p109041...	7.482589	8.910590	576.769327	8.214680	18.885	1.476	45.282	1022	0.786
Total Error				10.829129					

APPENDIX IV

Digital Elevation Model

545.456 m

430.602 m
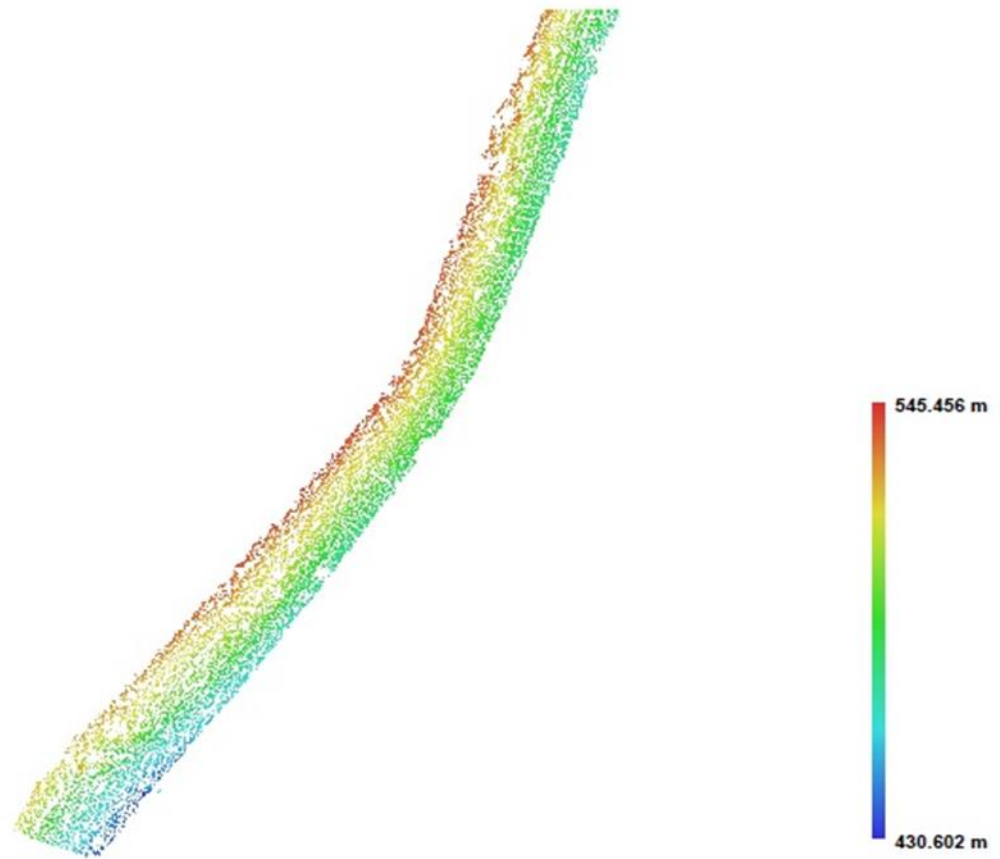
Reconstructed digital elevation model.

Resolution: 3.05284

Point density: 0.107285 per sq. m

APPENDIX V

Results for the experiments conducted for the integrated developed system is as follows:

Each Video Frame Geographical Location Data

Cameras	Longitude	Latitude	Altitude
P109036...	7.478762	8.905067	577.014400
P109036...	7.478832	8.905163	577.860200
P109036...	7.478898	8.905252	578.186000
P109036...	7.478992	8.905383	578.624900
P109036...	7.479057	8.905477	578.847000
P109036...	7.479122	8.905572	579.624700
P109036...	7.479155	8.905617	580.007400
P109036...	7.479248	8.905765	580.160400
P109036...	7.479283	8.905815	580.347100
P109036...	7.479353	8.905918	580.103700
P109037...	7.479428	8.906032	578.675200
P109037...	7.479492	8.906153	577.656600
P109037...	7.479605	8.906307	578.253400
P109037...	7.479690	8.906410	578.273200
P109037...	7.479732	8.906463	577.914500
P109037...	7.479853	8.906610	578.256800
P109037...	7.479895	8.906658	578.773500
P109037...	7.479968	8.906775	578.796200
P109037...	7.480083	8.906957	578.113100
P109037...	7.480162	8.907068	578.147400
P109038...	7.480242	8.907173	579.027700
P109038...	7.480328	8.907272	579.018600
P109038...	7.480413	8.907372	578.777900
P109038...	7.480487	8.907488	578.544800
P109038...	7.480520	8.907550	577.053200
P109038...	7.480588	8.907672	575.381600
P109038...	7.480703	8.907843	575.260100
P109038...	7.480738	8.907902	575.719000
P109038...	7.480812	8.908010	575.820500
P109038...	7.480887	8.908133	576.775700
P109039...	7.481008	8.908310	578.418400
P109039...	7.481093	8.908418	578.679000
P109039...	7.481137	8.908467	578.632300
P109039...	7.481222	8.908572	579.038200
P109039...	7.481340	8.908743	578.591600
P109039...	7.481378	8.908798	578.371400
P109039...	7.481462	8.908908	579.455600
P109039...	7.481577	8.909083	580.395100
P109039...	7.481652	8.909192	579.351300
P109039...	7.481728	8.909305	578.235300
P109040...	7.481810	8.909427	577.670400
P109040...	7.481848	8.909482	577.789500

☑	P109040...	7.481977	8.909660	577.885700
☑	P109040...	7.482057	8.909768	578.361700
☑	P109040...	7.482097	8.909820	579.094600
☑	P109040...	7.482217	8.909997	578.542600
☑	P109040...	7.482252	8.910060	578.284600
☑	P109040...	7.482353	8.910248	579.178200
☑	P109040...	7.482417	8.910383	580.004800
☑	P109040...	7.482447	8.910452	579.602400
☑	P109041...	7.482518	8.910577	578.912700
☑	P109041...	7.482578	8.910702	578.745300
☑	P109041...	7.482668	8.910892	580.597500
☑	P109041...	7.482725	8.911027	581.542200
☑	P109041...	7.482788	8.911155	580.384500
☑	P109041...	7.482857	8.911272	580.008500
☑	P109041...	7.482887	8.911328	580.197100
☑	P109041...	7.482968	8.911542	580.252100
☑	P109041...	7.482993	8.911613	580.715500
☑	P109041...	7.483037	8.911690	580.715500
☑	P109042...	7.483147	8.911933	581.359600
☑	P109042...	7.483173	8.912005	580.839200
☑	P109042...	7.483237	8.912125	580.293800
☑	P109042...	7.483305	8.912243	579.515600
☑	P109042...	7.483368	8.912367	578.284600
☑	P109042...	7.483460	8.912557	578.290000
☑	P109042...	7.483492	8.912622	578.609600
☑	P109042...	7.483555	8.912758	577.426000
☑	P109042...	7.483635	8.912962	575.664000
☑	P109042...	7.483693	8.913093	574.397600
☑	P109043...	7.483725	8.913158	574.042000
☑	P109043...	7.483812	8.913352	574.057400
☑	P109043...	7.483872	8.913477	573.701200
☑	P109043...	7.483935	8.913602	572.784000
☑	P109043...	7.483967	8.913662	572.356300
☑	P109043...	7.484033	8.913788	570.970600
☑	P109043...	7.484095	8.913938	569.363300
☑	P109043...	7.484117	8.913985	570.336700

APPENDIX VI

SURF M FILE

```matlab
%Read the reference image containing the object of interest
g = imread('1.jpg');
boxImage = rgb2gray(g);
figure; imshow(boxImage);
title('Image of a known object');
%Read the target image containing a cluttered scene
r = imread('2.jpg');
sceneImage = rgb2gray(r);
figure; imshow(sceneImage);
title('Raw image of ROW captured by UAV');
%Detect feature points in both images
boxPoints = detectSURFFeatures(boxImage);
scenePoints = detectSURFFeatures(sceneImage);
%Visualize the strongest feature points found in the reference image
figure; imshow(boxImage);
hold on;
plot(boxPoints.selectStrongest(20));
%Visualize the strongest feature points found in the target image
figure; imshow(sceneImage);
title('20 Strongest Feature Points from image capture by UAV');
hold on;
plot(scenePoints.selectStrongest(20));
%Extract feature descriptors at the interest points in both images
[boxFeatures, boxPoints] = extractFeatures(boxImage, boxPoints);
[sceneFeatures, scenePoints] = extractFeatures(sceneImage, scenePoints);
%Match the features using their descriptors
boxPairs = matchFeatures(boxFeatures, sceneFeatures)
%Display putatively matched features.
matchedBoxPoints = boxPoints(boxPairs(:, 1), :)
matchedScenePoints = scenePoints(boxPairs(:, 2), :)
figure;
showMatchedFeatures(boxImage, sceneImage, matchedBoxPoints, ...
    matchedScenePoints, 'montage')
title('Putatively Matched Points (Including Outliers)');
%estimateGeometricTransform calculates the transformation relating the
matched points,
%while eliminating outliers. This transformation allows us to localize
the object in the scene
[tform, inlierBoxPoints, inlierScenePoints] = ...
    estimateGeometricTransform(matchedBoxPoints, matchedScenePoints,
'affine');
%Display the matching point pairs with the outliers removed
figure;
showMatchedFeatures(boxImage, sceneImage, inlierBoxPoints, ...
    inlierScenePoints, 'montage');
title('Matched Points (Inliers Only)');
%Get the bounding polygon of the reference image
boxPolygon = [1, 1;...                              % top-left
        size(boxImage, 2), 1;...                    % top-right
        size(boxImage, 2), size(boxImage, 1);... % bottom-right
```

```matlab
        1, size(boxImage, 1);...                % bottom-left
        1, 1];                       % top-left again to close the polygon
%Transform the polygon into the coordinate system of the target image.
The transformed polygon
%indicates the location of the object in the scene.
newBoxPolygon = transformPointsForward(tform, boxPolygon);
%Display the detected object
figure; imshow(sceneImage);
hold on;
line(newBoxPolygon(:, 1), newBoxPolygon(:, 2), 'Color', 'y');
title('Detected object');
```

APPENDIX VIIA

Survey Data
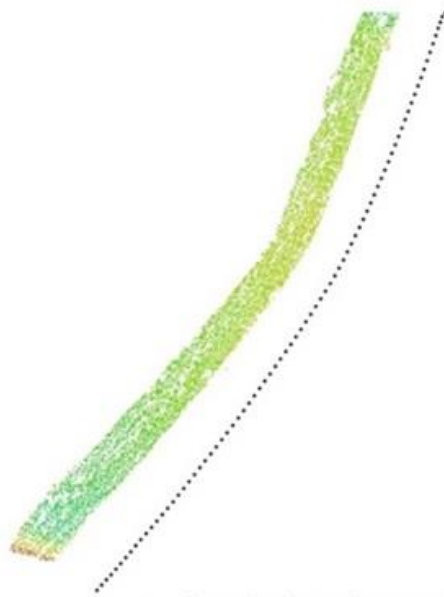
Camera locations and image overlap.

Number of images:	80	Camera stations:	80
		Tie-points:	20974
Ground resolution:	0.0248157 m/pix	Projections:	74005
Coverage area:	0.0420028 sq km	Error:	0.71731 pix

Camera Model	Resolution	Focal Length	Pixel Size	Precalibrated
DMC-GH3 (12 mm)	4608 x 2592	12 mm	3.77706 x 3.77706 um	No

Camera Locations

	7.81248 m
	6.24999 m
	4.68749 m
	3.12499 m
	1.5625 m
	0 m
	-1.5625 m
	-3.12499 m
	-4.68749 m
	-6.24999 m
	-7.81248 m
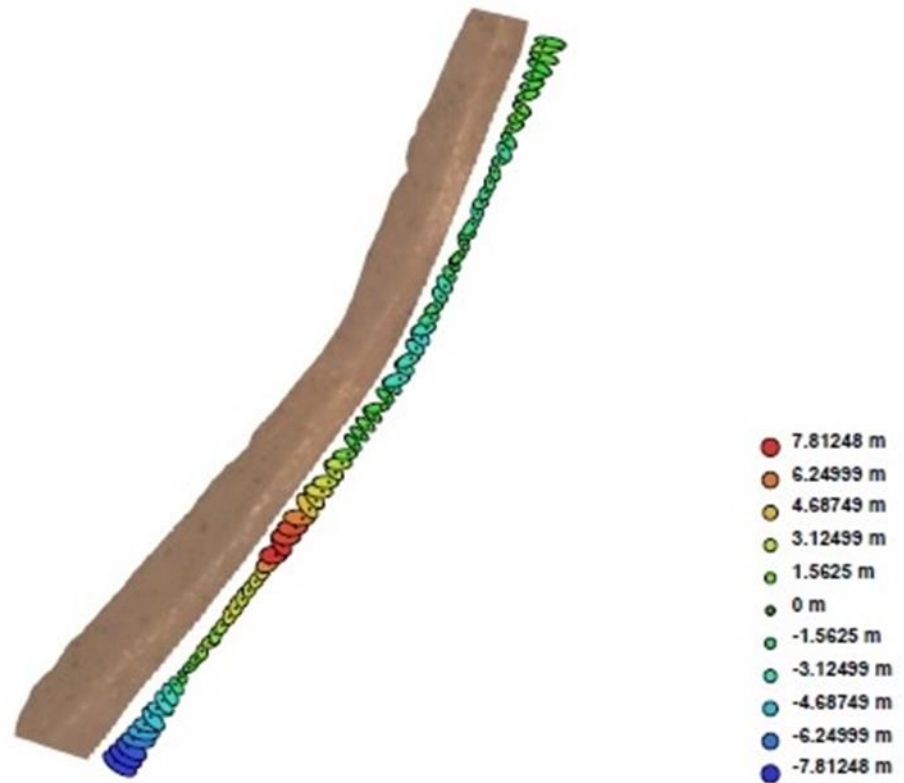
Camera locations and error estimates.
Z error is represented by ellipse color. X,Y errors are represented by ellipse shape.
Estimated camera locations are marked with a black dot.

X error (m)	Y error (m)	Z error (m)	Total error (m)
7.087607	4.113576	3.460632	8.895598

APPENDIX VIII

Results of Ortho-rectification of Each Video Frame

Cameras	Long. est	Lat. est	Alt. est	Error (m)	Yaw est	Pitch est	Roll est	Error (deg)	Projections	Error (pix)
P109036...	7.478659	8.905135	569.413518	15.538960	25.745	8.552	39.741		821	0.689
P109036...	7.478735	8.905221	570.199703	14.557072	26.732	7.691	39.641		966	0.598
P109036...	7.478807	8.905306	571.078168	13.697585	27.117	7.570	39.562		909	0.675
P109036...	7.478879	8.905402	572.662820	13.898971	27.035	7.337	39.164		878	0.735
P109036...	7.478951	8.905493	573.650557	12.903372	27.175	7.288	39.171		882	0.667
P109036...	7.479024	8.905585	574.836172	11.801820	27.507	6.933	39.084		883	0.736
P109036...	7.479096	8.905682	575.582050	10.693321	27.565	6.684	39.084		936	0.649
P109036...	7.479170	8.905782	575.951287	9.764132	27.749	6.455	39.173		969	0.737
P109036...	7.479243	8.905881	576.543044	9.332388	27.830	6.260	39.005		966	0.763
P109036...	7.479323	8.905981	576.708153	8.365035	29.082	4.935	39.405		956	0.694
P109037...	7.479402	8.906080	577.046587	6.275342	28.944	4.848	39.628		953	0.741
P109037...	7.479485	8.906180	577.883403	3.059178	28.751	4.631	39.540		955	0.672
P109037...	7.479568	8.906284	578.735649	4.805217	28.742	4.415	39.494		951	0.728
P109037...	7.479652	8.906388	579.326173	4.923138	28.668	4.276	39.709		876	0.661
P109037...	7.479732	8.906490	579.969400	3.591878	28.625	4.172	39.780		868	0.589
P109037...	7.479817	8.906596	580.787934	4.967127	28.387	4.117	39.687		837	0.699
P109037...	7.479901	8.906701	580.793656	5.134646	28.569	3.699	39.950		828	0.690
P109037...	7.479985	8.906807	581.287207	4.724291	28.159	3.936	39.841		903	0.677
P109037...	7.480070	8.906913	581.918310	6.351120	28.025	3.853	39.564		929	0.670
P109037...	7.480151	8.907018	582.186683	6.981266	27.816	3.824	39.667		959	0.698
P109038...	7.480232	8.907126	582.297202	6.261735	27.656	3.724	39.799		986	0.720
P109038...	7.480312	8.907239	582.394343	5.283981	27.911	3.102	39.861		941	0.735
P109038...	7.480397	8.907349	582.463777	4.820039	27.792	2.891	39.978		885	0.708
P109038...	7.480480	8.907459	582.693763	5.352442	27.638	2.726	39.780		900	0.802
P109038...	7.480564	8.907569	582.859667	7.813997	27.904	2.161	39.860		888	0.778
P109038...	7.480650	8.907677	583.097842	10.295585	27.659	1.877	39.955		886	0.743
P109038...	7.480732	8.907792	583.072578	10.160998	27.391	1.794	40.022		885	0.732
P109038...	7.480811	8.907901	583.040595	10.824507	27.593	1.585	40.205		919	0.812
P109038...	7.480894	8.908014	582.972767	11.511816	27.446	1.442	40.225		925	0.713
P109038...	7.480980	8.908126	582.802096	11.934308	27.476	1.119	40.432		954	0.684
P109039...	7.481064	8.908238	583.056631	11.061980	27.120	1.111	40.328		942	0.656
P109039...	7.481146	8.908349	582.838058	10.404066	26.984	0.885	40.458		945	0.794
P109039...	7.481225	8.908459	582.816611	10.581386	27.173	0.457	40.304		965	0.728
P109039...	7.481307	8.908572	582.301086	9.956274	26.946	0.209	40.480		955	0.662
P109039...	7.481388	8.908689	581.740652	8.619243	26.850	0.085	40.495		942	0.615
P109039...	7.481469	8.908802	581.329824	10.445264	26.804	-0.143	40.515		913	0.563
P109039...	7.481548	8.908914	580.619959	9.618197	26.451	-0.084	40.774		908	0.606
P109039...	7.481628	8.909030	580.131768	8.168022	26.238	-0.065	40.600		952	0.560
P109039...	7.481711	8.909147	579.487831	8.206060	25.639	-0.257	40.742		998	0.533
P109039...	7.481791	8.909265	578.938276	8.195794	25.432	-0.367	40.725		1064	0.529
P109040...	7.481872	8.909381	578.363224	8.497906	25.127	-0.556	40.681		1118	0.536
P109040...	7.481951	8.909497	577.972781	11.433377	24.920	-0.785	40.607		1032	0.541
P109040...	7.482032	8.909615	577.411793	7.855344	24.609	-0.930	40.552		932	0.528

APPENDIX VIII CONT.

☑ P109040...	7.482112	8.909731	576.445447	7.549628	24.555	-1.203	40.797	920	0.519
☑ P109040...	7.482188	8.909846	575.811111	10.985828	24.370	-1.482	40.720	953	0.540
☑ P109040...	7.482260	8.909971	575.440103	6.374661	24.022	-1.559	40.518	1013	0.530
☑ P109040...	7.482326	8.910095	575.429224	9.469083	22.641	-0.788	40.489	954	0.489
☑ P109040...	7.482392	8.910221	575.719125	6.222517	21.255	0.015	40.499	930	0.536
☑ P109040...	7.482455	8.910345	576.047350	7.171049	20.735	0.163	40.354	1013	0.577
☑ P109040...	7.482516	8.910471	576.483014	8.468465	20.214	0.233	40.298	1067	0.710
☑ P109041...	7.482576	8.910598	576.875570	7.036727	19.070	0.700	40.314	1035	0.783
☑ P109041...	7.482635	8.910730	577.666164	7.043895	17.854	1.176	40.114	1025	0.658
☑ P109041...	7.482693	8.910860	578.127225	5.094751	16.994	1.480	40.116	980	0.562
☑ P109041...	7.482750	8.910988	578.576202	5.873822	16.602	1.532	40.058	947	0.749
☑ P109041...	7.482813	8.911116	579.283561	5.197634	16.182	1.363	39.864	940	0.711
☑ P109041...	7.482871	8.911244	579.353740	3.534877	15.714	1.678	40.012	928	0.807
☑ P109041...	7.482930	8.911370	579.976925	6.629631	15.487	1.457	39.917	949	0.652
☑ P109041...	7.482986	8.911500	579.811651	5.079891	15.163	1.415	40.033	976	0.741
☑ P109041...	7.483039	8.911627	579.564251	5.346710	14.578	1.483	39.995	1027	0.841
☑ P109041...	7.483093	8.911757	579.255576	9.787763	14.463	1.136	40.200	1020	0.751
☑ P109042...	7.483146	8.911889	578.927115	5.444041	14.107	1.130	40.121	964	0.704
☑ P109042...	7.483201	8.912018	578.726633	4.000152	13.666	0.960	40.130	890	0.641
☑ P109042...	7.483252	8.912147	578.409538	3.524626	13.457	0.982	40.146	862	0.744
☑ P109042...	7.483306	8.912275	577.892498	3.831519	13.194	0.972	40.248	923	0.745
☑ P109042...	7.483357	8.912406	577.513212	4.603937	12.940	0.645	40.153	916	0.672
☑ P109042...	7.483413	8.912540	576.880970	5.664567	12.817	0.387	40.374	979	0.662
☑ P109042...	7.483466	8.912671	576.448427	6.512509	12.575	0.141	40.303	1025	0.637
☑ P109042...	7.483519	8.912801	576.308058	6.215839	12.406	0.026	40.394	922	0.646
☑ P109042...	7.483572	8.912935	575.804787	7.554589	11.830	0.057	40.445	935	0.658
☑ P109042...	7.483620	8.913065	575.426388	8.708277	11.684	-0.098	40.393	997	0.740
☑ P109043...	7.483672	8.913195	574.883089	7.172613	11.018	0.102	40.570	1024	0.725
☑ P109043...	7.483723	8.913327	574.560429	10.122763	10.336	0.409	40.309	1000	0.648
☑ P109043...	7.483774	8.913457	573.857231	10.910105	10.121	0.338	40.478	970	0.598
☑ P109043...	7.483825	8.913588	572.985385	12.216072	9.797	0.139	40.653	903	0.661
☑ P109043...	7.483873	8.913716	572.040627	11.973147	9.923	-0.023	40.709	868	0.772
☑ P109043...	7.483915	8.913826	571.673642	13.668899	9.925	-0.725	40.304	888	0.681
☑ P109043...	7.483953	8.913924	571.135794	15.856229	9.862	-0.576	40.222	944	0.557
☑ P109043...	7.483984	8.914010	570.355706	14.912623	9.641	-0.990	40.283	789	0.598
Total Error				8.895598					

Digital Elevation Model

534.071 m

424.478 m
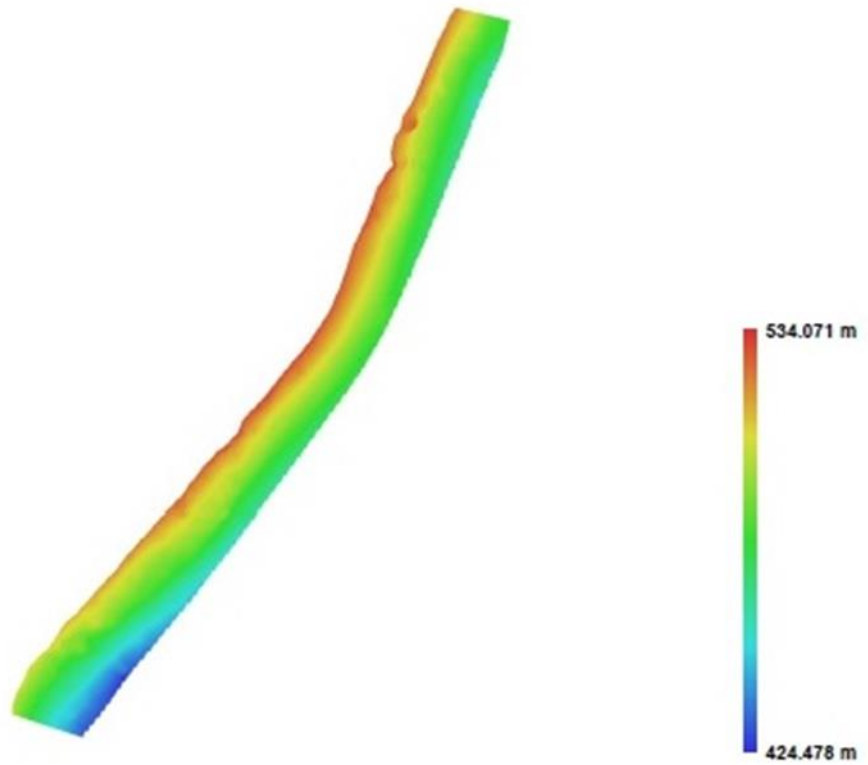
Reconstructed digital elevation model.

Resolution: 3.05161 m/pix
Point density: 0.107384 points per sq m

REFERENCES

[1] M. Kontitsis, M., Valavanis, K., N. Tsourveloudis, "A UAV Vision System for Airborne   Surveillance." *In Proc. of the IEEE International Conference on Robotics and Automation,* 2004, pp. 77–83.

[2] D.W. Casbeer, D.B. Kingston, R.W. Bear , T.W. McLain, and S.M. Li, "Cooperative Forest Fire Surveillance Using a Team of Small Unmanned Air Vehicles.", *Intl. Journal of System Science,* January 2005, pp 1-18.

[3] A. Ollero, J. Ferruz, F. Caballero, S. Hurtado, L. Merino, "Motion Compensation and Object Detection for Autonomous Helicopter Visual Navigation in the Comets System". *In: Proc. of the IEEE International Conference on Robotics and Automation, 2004, pp. 19–24.*

[4] B. Colfman, M. McCord, and K. Redmill, "Surface Transportation Surveillance from Unmanned Aerial Vehicles" *Proc. Of the 83rd Annual Meeting of the Transportation Research Board*, 2004.

[5] J. Allen, and B. Walsh, "Enhanced Oil Spill Surveillance, Detection and Monitoring through the Applied Technology of Unmanned Air Systems". *In: Proceedings of the 2008 international oil spill conference.*

[6]   R. W. Beard, T. W. McLain, D. B. Nelson, D. Kingston, and D. Johanson,"Decentralized cooperative aerial surveillance using fixed wing miniature UAVs," *Proc. IEEE, July 2006, vol. 94, no. 7, pp. 1306– 1323.*

[7]   D.Fransaer, F. Vanderhaeghen, and J. Everaerts, "PEGASUS: Business Plan for a Stratospheric Long Endurance UAV System for Remote Sensing". The International Archives of the Photogrammetry, Remot Sensing and Spatial Information Sciences, Istanbul, Turkey, GITC, Lemmer, Netherlands, 2004.

[8] A. Ollero, J. Ferruz, F. Caballero, S. Hurtado, L. Merino, "Motion Compensation and Object Detection for Autonomous Helicopter Visual Navigation in the Comets System". *In: Proc. of the IEEE International Conference on Robotics and Automation, 2004, pp. 19–24.*

[9]   http://pipelinesinternational.com/news (Accessed: 14 July 2013).

[10] B.O. Olawale, C.R. Chatwin, R.C.D. Young, P.M. Birch, and F.O. Faithpraise "RealTime Monitoring Of Buried Oil Pipeline Right-Of-Way for Third-Party Incursion", *International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 2,February 2015, pp. 163-173.*

[11] B.O. Olawale, C.R. Chatwin & R.C.D Young, P.M. Birch, F.O. Faithpraise, and A.O. Olukiran, "A Four-Step Ortho-Rectification Procedure for Geo-Referencing Video Streams from a Low-Cost UAV," *Int. J. Comput., Elect., Automation, Control Inform. Eng., vol. 9, no. 8, pp. 1445 – 1452, July, 2015.*

[12] P.Baronti, P. Pillai, V.Chook, S. Chessa, F. Gotta, and H. Fun, "Wireless sensor networks: a survey on the state of the art and the 802.15.4 and zigbee standards". Communication Research Centre, UK, May 2006.

[13] A. Carrillo, E. Gonzalez, A.Rosas, and A. Marquez, New distributed optical sensor for detection and localization of liquid leaks. Pat I. Experimental Studies, Sens, Actuators, A(99):229–235, 2002.

[14] N.Mohamed, and I. Jawhar, I. "A Fault-Tolerant Wired/Wireless Sensor Network Architecture for Monitoring Pipeline Infrastructures". *In Proceedings of the Second International Conference on Sensor Technologies and Applications (SENSORCOMM 2008), Cap Esterel, France,25–31 August 2008; pp. 179-184*.

[15] Y. Tu, H.and H. Chen, "Design of oil pipeline leak detection and communication systems based on optical fiber technology". *Proc. SPIE 1999, 3737, 584-592*.

[16] I. Jawhar, N. Mohamed, M. Mohamed, and M.A. Aziz, Routing Protocol and Addressing Scheme for Oil, Gas, and Water Pipeline Monitoring Using Wireless Sensor Networks. *In Proceedings of the Fifth IEEE/IFIP International Conference on Wireless and Optical Communications Networks (WOCN2008), Surabaya, East Java, Indonesia, 5–7 May 2008*.

[17] H. Van der Werff, M. Vander Meijde, F. Jansma, F. Van der Meer, and G.J. Groothuis. A Spatial-Spectral Approach for Visualization of Vegetation Stress Resulting from Pipeline Leakage. *Sensors 2008, 8, 3733-3743*.

[18] I. Stoianov, L. Nachman, S. Madden, T. Tokmouline. PIPENET: A Wireless Sensor Network for Pipeline Monitoring, In *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*, *Cambridge, MA, USA, 25–27 April 2007; pp. 264-273*.

[19] Y. Jin, and A. Eydgahi. Monitoring of Distributed Pipeline Systems by Wireless Sensor Networks. In *Proceedings of the International Conference on Engineering and Technology*, *Nashville, TN, USA, 17–19 November 2008*.

[20] Murphy, F., Laffey, D., O'Flynn, B., Bukley, J., Barton, J. Development of a wireless sensor network for collaborative agents to treat scale formation in oil pipes. LNCS 2007, 4373, 179-194.

[21] Anon, "*Functional Specification for a Satellite Surveillance System*", Andrew Palmer and Associates Report NR01003, March 2001.

[22] F.J Ahern, J. Murphy. Radiometric calibration and correction of Landsat 1, 2 and MSS data. Research Reports 78-4, November 1978. Canada Centre for Remote Sensing (Energy, Mines and Resources), Ottawa, Canada.

[23] Anon, "*A Cost Benefit Study of Satellite Surveillance*", Andrew Palmer and Associates report NR98029/R2, March 2001.

[24] Bento, M. D. F. Unmanned Aerial Vehicles: An Overview. *Inside GNSS* (January/February 2008), 54-61.

[25] http://www.ctie.monash.edu/hargrave/rpav_home.html. (Accessed on 12 August 2013).

[26] A. Reg. Unmanned Aircraft Systems: UAVs Design, Development and Deployment. A John Wiley and sons, Ltd., publication, 2010.

[27] R. Frampton, and F. Dstl.UK Ministry of Defence. 'UAV Autonomy'. (*Defence Codex – The Journal for Defence Engineering and Science*, Issue 1, summer 2008, UK Ministry of Defence.)

[28] R.A Frampton, and J.M Keirl. UK Ministry of Defence. 'Autonomy and its Application to Unmanned Systems'. (Proceedings of the 1st Moving Autonomy Forward Conference, *21–22 June 2006, De Vere Belton Woods Hotel*, Grantham, UK, Muretex Ltd.).

[29] D. Potts. 'The Observer Concept', *Proceedings of the 14th Bristol International UAV Systems Conference*, University of Bristol, UK, 1999.

[30] D. Hausamann, "Civil Applications of UAVs – User Approach", Shepherd's Civil UAV Symposium, London, UK, 17 – 19 July (2002).

[31] R.A. Sschowengerdt, *Remote sensing: Models and Methods for Image Processing,* 3rd ed., Academic Press, 2007, p. 1.

[32] J.G. Liu & P.J. Mason, *Essential Image Processing for GIS and Remote Sensing.* Wiley-Blackwell, 2009, p. 2.

[33] S.E. Umbaugh, *Computer Imaging: Digital Image Analysis and Processing.* CRC Press 2005, p. 15.

[34] R. Pouncey, K. Swanson & K. Hart, *ERDAS Field Guide.* 5th ed., Atlanta: ERDAS, 1999, p. 5.

[35] P.M. Mather, *Computer Processing of Remotely-Sensed Images: An Introduction.* 2nd ed., Wiley, 1999, p. 28.

[36] e-Cognition (2003); most recent information under www.definiens-imaging.com.

[37] Anon, "*Functional Specification for a Satellite Surveillance System*", Andrew Palmer and Associates Report NR01003, March 2001.

[38] http://www.dataflows.com/GPS_Overview.shtml (Accessed: 28th February 2016).

[39] http://www.gps.gov/systems/gps/control/ (Accessed: 28th February 2016).

[40] E. V. Hinueber, "Design of an Unaided Aircraft Attitude Reference System with Medium Accurate Gyroscopes for Higher Performance Attitude Requirements". *Inertial Sensors and Systems - Symposium Gyro Technology, Karlsruhe / Germany* (iMAR Navigation / DGON) 2011.

[41] J. Wendell, J. Metzger & A. Moenikes, "A Performance Comparison of Tightly Coupled GPS/INS Navigation Systems Based on Extended and Sigma-Point Kalman Filters," *J. Institute Navigation,* vol. 53, no. 1, 2006.

[42] A. C. Bovik. *Introduction to Digital Image and Video Processing*, Chapter 1.1, Morgan Kaufmann, 2000, pp. 3–17.

[43] R. C. Gonzalez and R. E.Woods. *Digital Image Processing*, 3rd edition, Prentice Hall, Upper Saddle River, NJ, 2008, pp. 50-54.

[44] P. W. Wong. *Image Quantization, Halftoning, and Printing*, Chapter 1.1, Morgan Kaufmann, 2000, pp. 657–667.

[45] O. Marques, *Practical Image and Video Processing using MATLAB.* New Jersey: John Wiley & Sons, Inc, 2011, pp. 21–26.

[46] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. In International Conference on Computer Graphics and Interactive Techniques, ACM Press, New York, NY, 2007.

[47] http://uk.mathworks.com/help/images/what-is-image-filtering-in-the-spatial domain.html (Accessed: 4th March 2016).

[48] D. Ballard and C. Brown. *Computer Vision*, Prentice-Hall, 1982, pp 24 – 30.

[49] A. Jain. *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989, pp 15 - 20.

[50] http://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm (Accessed: 4th March 2016).

[51] P.K. Sinha, *Image Acquisition and Pre-processing for Machine Vision Systems.* SPIE, pp. 495–520.

[52] T. Botterill, S. Mills, R. Green, Real-time aerial image mosaicing, in: International Conference of Image and Vision Computing New Zealand (IVCNZ), 2010, pp. 1–8.

[53] D. Ghosh & N. Kaabouch, "A survey on image mosaicing techniques," *J. Vis. Commun. Image R.*, vol. 34, pp. 1–11, 2016.

[54] M. Vivet, S. Peleg, X. Binefa, Real-time stereo mosaicing using feature tracking, in: IEEE International Symposium on Multimedia (ISM), 2011, pp. 577– 582.

[55] D. Ghosh, N. Kaabouch, W. Semke, Super-resolution mosaicing of unmanned aircraft system (UAS) surveillance video frames, Int. J. Sci. Eng. Res. 4 (2013).

[56] D.C. Brown. *The Bundle Adjustment: Progress and Prospects*, XIIIth international congress for photogrammetry, Helsinki, Finland, 1976.

[57] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11):1475–1490, 2004.

[58] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California, June 201O, pp 2544-2550.

[59] X. L. Dai & J. Lu, "An Object-based Approach to Automated Image Matching," *Geoscience Remote Sensing Symp.,igarss '99 Proc.ieee Int.* , vol. 2, 1999.

[60] M.F. Calitz and H. Rüther, H. L-norm methods in area based matching, International Archives for Photogrammetry and Remote Sensing, 30 (5), 1994, pp. 22.

[61] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. IJCV, 74(1), 200759–73.

[62] H. Moravec, Rover visual obstacle avoidance. In International Joint Conference on Artificial Intelligence, Vancouver, British Columbia, 1981, pp. 785-790.

[63] C. Harris and M.I. Stephens. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, 1988, pp. 147-151.

[64] http://opencv-python tutroals.readthedocs.org/en/latest/py_tutorials/py_feature2d/py_features_harris/py_features_harris.html (Accessed: 4 May 2015).

[65] C. Schmid, and R. Mohr. Local gray value invariants for image retrieval. IEEE Trans. On Pattern Analysis and Machine Intelligence, 19(5), 1997, pp. 530-534.

[66] P. Torr. Motion Segmentation and Outlier Detection, Ph.D. Thesis, Dept. of Engineering Science, University of Oxford, UK, 1995.

[67] D.G Lowe. Object recognition from local scale-invariant features. In International Conference on Computer Vision, Corfu, Greece, 1999, pp. 1150-1157.

[68] http://aishack.in/tutorials/sift-scale-invariant-feature-transform-features/ (Accessed: 4 April 2014).

[69] D.G. Lowe. Local feature view clustering for 3D object recognition. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, 2001, pp. 682-688.

[70] H. Bay, A. Ess and T. Tuytelaars, "Computer Vision and Image Understanding," *Elsevier*, no. 110, pp. 346–359, Sept. 2007. [Online]. Available: http://www.journals.elsevier.com/computer-vision-and-image-understanding. (Accessed 26 Feb. 2016).

[71] A. Baumberg. Reliable feature matching across widely separated views. In CVPR, 2000, pp. 774 – 781.

[72] F. Ackermann. Digital Image Correlation: Performance and Potential Application in Photogrammetry, Photogrammetric Record, Vol. 11, no. 64, 1984, pp. 429 – 439.

[73] E.P. Baltsavias. Hierarchical Multiphoto Matching and DTM Generation, Report No. 156, Institute for Geodesy and Photogrammetry, ETH-Zurichi, 1989, pp. $1 - 17$.

[74] C. Heipke, "Overview of image matching techniques", OEEPE Workshop on the Application of Digital Photogrammetric Work-Stations, OEEPE Official Publications, No.33, pp.173-189, 1996.

[75] B. Zitová and J. Flusser, "Image registration methods: a survey", Image and Vision Computing, vol.21, no.11, pp.977-1000, 2003.

[76] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Artificial Intelligence Journal, vol.78, no.1-2, pp. 87-119, 1995.

[77] F. Zhao, Q. Huang and W.Gao. Image Matching by Normalized Cross-Correlation, IEEE Conference on *Acoustics, Speech Signal Processing,* vol. 2, 2006.

[78] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. Numerical Recipes in C - 13.1 Convolution and Deconvolution Using the FFT. Cambridge University Press, Cambridge, 1988, 2nd ed., pp. 538–545.

[79] M. Savvides and B.V.K. Vijaya Kumar. Efficient design of advanced correlation filters for robust distortion-tolerant face recognition. Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, July, 2003. pp. 45–52.

[80] A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent. Minimum average Correlation Energy Filters. Appl. Opt., 26(17), 1987, P. 3633.

[81] K.H. Jeong, S. Han, and J.C. Principe. The fast correntropy MACE filter. In Proceedings of the International Conference on Acoustics, Speech, Signal Processing (ICASSP), 2007.

[82] M.D. Rodriguez, J. Ahmed, and M. Shah. Action MACH a spatio-temporal maximum average correlation height filter for action recognition. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008.

[83] B.V. Vijaya Kumar, A. Mahalanobis & R. D. Juday, *Correlation Pattern Recognition.*, New York: Cambridge University Press, 2005, pp. 205–224.

[84] R. Pouncey, K. Swanson & K. Hart, *ERDAS Field Guide.* 5th ed., Atlanta: ERDAS, 1999, p. 261.

[85] M.A. Gomarasca, *Basics of Geomatics.* 2nd ed., New York: Springer, 2009, pp. 79–85.

[86] W. Linder, *Digital Photogrammetry Theory and Applications.*, New York: Springer, 2003, p. 4.

[87] E.M. Mikhail, J.S. Bethel & J.C. Mcglone, *Introduction to Modern Photogrammetry.*, New York: Wiley & Sons Inc, 2001.

[88] W. Linder, *Digital Photogrammetry: A Practical Course.* 3rd ed., Heidelberg: Springer, 2009.

[89] H.A. Beyer. Line jitter and Geometric Calibration of CCD Cameras. ISPRS Journal of Photogrammetry & Remote Sensing 45, 1990, pp. 17-32.

[90] P.K. Sinha, *Image Acquisition and Preprocessing for Machine Vision Systems.*, Washington: SPIE, 2012, pp. 311–321.

[91] A. Gruen & T.S. Huang, "Calibration and Orientation of Cameras in Computer Vision," in *Springer series in information sciences,,* vol. 34, T.S. Haung, T. Kohonen, T. & M.R. Schroeder, Eds. New York: Springer, 2001, pp. 95–116.

[92] W. Boesemann, R. Godding & W. Riechmann. Photogrammetric Investigation of CCD Cameras. *Proc. Close-Range Photogrammetry Meets Machine Vision, ISPRS Comm. V,* Zurich, SPIE vol. 1395, 1990, pp. 119-126.

[89] H.A. Beyer. Line jitter and Geometric Calibration of CCD Cameras. ISPRS Journal of Photogrammetry & Remote Sensing 45, 1990, pp. 17-32.

[93] L. Chen, C.W. Armstrong & D.D. Raftopoulos, "An Investigation on the Accuracy of Three-dimensional Space Construction using the Direct Linear Transformation Technique," *J. Biomech*, vol. 27, 1994, pp. 493–500.

[94] A. Gruen & H.A. Beyer. System Calibration through Self-Calibration. *Presented to Workshop on Calibration and Orientation of Cameras in Computer Vision.* ISPRS XVII Congress, Washington, DC, 1992, pp. 2-14.

[95] M. Cramer and D. Stallmann, "System calibration for direct georeferencing," in *Proc. Int. Soc. Photogrammetry Remote Sens.*, Graz, Sep. 2002, vol. 34, 2002, pp. 79–84, Com. III, Part A.

[96] Y. I. Abdel-Aziz & H. M. Karara, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," in *Proc. Symp. Close-Range Photogramm,* 1971, pp. 1–18.

[97] J.Wang, F. Shi, J. Zhang & Y. Liu, A new calibration model and method of camera lens distortion. PR **41**(2), 2008, pp. 607–615.

[98] D.C. Brown. Decentering Distortion of Lenses. Photogrammetric Engineering **32(3),** 1996, pp. 444-462.

[99] W. Boesemann, R. Godding, W. Riechmann. Photogrammetric Investigation of CCD Cameras. *Proc. Close-Range Photogrammetry Meets Machine Vision,ISPRS Comm. V,* Zurich, SPIE vol. 1395, 1990, pp. 119-126.

[100] X. Chen, J. Yang, and A. Waibel. Calibration of a hybrid camera network. In *Proceedings of the 9th International Conference on Computer Vision*, IEEE Computer Society Press, Nice, France, 2003, pp. 150–155.

[101] M.R. Shortis, A.W. Burner, W.L. Snow, W.K. Goad. Calibration Tests of Industrial and Scientific CCD Cameras. *Proc. 1st Australian Photogrammetry Conference, Sydney,* Nov. 7-9, Paper 6, **11** p., 1991

[102] A.W. Burner, W.L. Snow, M.R. Shortis, W.K. Goad. Laboratory Calibration and Characterization of Video Cameras. *Proc. Close- Range Photogrammetry meets Machine Vision, ISPRS Comm. V,* Zurich, SPIE vol. 1395, 1990, pp. 664-671.

[103] W.J. Greening, B.A. Chaplin & D.G. Sutherland, "Commercial Applications of GPS-Assisted Photogrammetry," *Gps/gis Conf.*, vol. 4, 1994, pp. 391–402.

[104] F.S. Remondino & C.S. Fraser, "Digital camera calibration methods: considerations and comparisons," *Iaprs*, vol. XXXVI, Sept. 2006, pp. 25–27.

[105] K.P. Schwarz, M.A. Chapman, M.E. Cannon & P. Gong, "An Integrated INS/GPS Approach to the Georeferencing of Remotely Sensed Data," *Photogrammetric Eng. Remote Sensing*, vol. 55, no. 12, 1993, pp. 1667–1674.

[106] E.J. Krakiwsky, "The Method of Least Squares: A Synthesis of Advances," *UCGE Report*, no. 10003, 1990.

[107] P. Friep, "Experience with GPS Supported Aerial Triangulation," *Proc. Isprs*, vol. 29, 1992, pp. 299–305.

[108] M.A. Gomarasca, *Basics of Geomatics.* 2nd ed., New York: Springer, 2009, pp. 88–89.

[109] F. Remondino & C.S. Fraser, "Digital camera calibration methods: considerations and comparisons," *IAPRS,* vol. XXXVI, part 5, Dresden 25–27, Sept. 2006.

[110] Anon, "*Functional Specification for a Satellite Surveillance System*", Andrew Palmer and Associates Report NR01003, March 2001.

[111] M.D.F Bento, "Unmanned Aerial Vehicles: An Overview. *Inside GNSS* ", February, 2008, pp. 54-61.

[112] N. Mohamed, I. Jawhar, "A Fault-Tolerant Wired/Wireless Sensor Network Architecture or Monitoring Pipeline Infrastructures", in Proc. Int. Conference on Sensor Technologies and Applications (SENSORCOMM 2008), Cap Esterel, France,25–31 August 2008; pp. 179-184.

[113] Y. DU, P.M. Teillet, and J.Cihlar, "Radiometric Normalization of Multi-Temporal High-Resolution Satellite Images with Quality Control for Land Cover Change Detection: Remote Sensing of Environment", 2002, pp. 123–134,

[114] S. Chen, B. Mulgrew, and P. M. Grant, "A Clustering Technique for Digital Communications Channel Equalization Using Radial Basis Function Networks," *WASET Trans. Neural Networks*, vol. 4, 1993, pp. 570–578.

[115] B. Horn *Robot Vision*, MIT Press, 1986, pp 314 – 315.

[116] K. Kobayashi, C. Mori, "Relations between the Coefficients in the Projective Transformation Equations and the Orientation Element of the Photograph". Journal of Photogrammetric Engineering and Remote Sensing. Vol. 63, No. 9, September 1997, pp.1121-1127.

[117] Z. Guoqing, ''Near Real-Time Orthorectification and Mosaic of Small UAV Video Flow for Time-Critical Event Response' 'in Proc. O Geoscience and Remote Sensing, IEEE, 47(3), March 2009.

[118] Z. Li, J. Chen, and E. Baltsavias. Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences, 2008 ISPRS Congress Book. CRC Press, Taylor & Francis Group, Boca Raton, London, New York, Leiden, p. 527.

[119] E.M Mikhail, J.S. Bethel, J.C. McGlone. Introduction to Modern Photogrammetry. John Wiley & Sons Inc., New.

[120] B Zheng and R. B. Bapat. Generalized inverse A (2) T, S and a rank equation, Applied Mathematics and Computation 155 2004, pp. 407-415.

List of Published Papers

1. **B.O. Olawale**, C.R. Chatwin, R.C.D. Young, and P.M. Birch. "RealTime Monitoring Of Buried Oil Pipeline Right-Of-Way for Third-Party Incursion", *International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 2, February 2015, pp. 163-173.*

2. **B.O. Olawale**, C.R. Chatwin & R.C.D Young, P.M. Birch, F.O. Faithpraise, and A.O. Olukiran, "A Four-Step Ortho-Rectification Procedure for Geo-Referencing Video Streams from a Low-Cost UAV." In proc. of the 17th International Conference on Computer Vision Theory and Applications (WASET), Venice, Italy, August 13-14th, 2015.

3. **B.O. Olawale**, C.R. Chatwin & R.C.D Young, P.M. Birch, F.O. Faithpraise, and A.O. Olukiran, "A Four-Step Ortho-Rectification Procedure for Geo-Referencing Video Streams from a Low-Cost UAV," *Int. J. Comput., Elect., Automation, Control Inform. Eng., vol. 9, no. 8, pp. 1445 – 1452, July, 2015.*

4. Fina Faithpraise, Chris Chatwin, Joseph Obu, **Babatunde Olawale**, Rupert Young and Philip Birch. "Sustainable control of Anopheles mosquito population." *International Journal of Environment, Ecology and Management, vol. 3, ISSN 2051-2945, 2014.*

The following papers are about to be published:

1. Title: "**Control and Monitoring of borders with the aid of multi-sensor UAVs**."

This paper focus on the control and monitoring of national borders with the aid of UAV equipped with sensors using a combination of trained composite correlation filters for detecting illegal immigrants and vehicles along the border line and a photogrammetry model to pin point the exact geographical location of the person or vehicle found for a quick response to the scene by security operatives.

2. Title: **"Crop Monitoring using Low-Cost UAVs Equipped with Optimised Sensor Configurations."**

This paper focus on the monitoring of a farm in order to check the health status of crops. This was achieved by the use of k-means color segmentation algorithm and image change detection analysis for observing change in the status of the crops and photogrammetry model for determining the exact location of the affected crops.