GEO-REFERENCING AND MOSAICING AGRICULTURAL FIELD IMAGES FROM A CLOSE-RANGE SENSING PLATFORM

BY

YANSHUI JIANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Advisor:

Associate Professor Lei Tian

ABSTRACT

Field images are becoming more frequently used for sensing the crucial properties of the crops in Precision Agriculture. Some of them contain distortions that need to be removed before further analysis. Data from images, such as the coordinates of the crops, also need specific algorithm to be extracted. For these purposes, two computer vision algorithms were developed for pre-processing the field images from two monocular vision systems. One algorithm was used for a tower remote sensing system data pre-processing for image distortion removal and the mosaic to generate geo-referenced images. The other was for the image data interpretation used for the vision system of a field robot.

Satellite and aerial remote sensing systems are the two major platforms for collecting remote sensing images for agriculture. However, due to the critical drawbacks of these systems, such as low spatial and temporal resolution, a tower remote sensing system with a 360-degree rotatable camera on the top has been established in the experiment field to obtain the multi-spectra images for monitoring the status of the plants. In this research, the geo-reference and image mosaic algorithms were developed for data acquisition.

While taking remote sensing images of the field, the camera will turn 360^o horizontally and 90^o vertically. This creates the difficulty of geo-reference because different images have different distortions. Therefore, traditional ways of geo-reference, such as using Ground Control Points (GCP), are no longer appropriate. A three-axis digital compass was used to provide the absolute orientation of the camera, which can be used to geo-reference a single image. The calibration of both camera and compass was introduced, and necessary parameters for geo-reference were estimated. Based on the angles, positions and optical parameters of the camera, a transformation from the image coordinate system to the ground coordinate system was introduced. After the transformation, the performance of the geo-reference method was evaluated with data from a Real-Time Kinematic Global Positioning System (RTK-GPS) to assure the accuracy.

Since there are not enough features in the field, the algorithm of the image mosaic for the tower system is based on geographical information rather than features. Moreover, comparing to aerial and satellite systems, the images from a tower remote sensing system have usually 10 to 15 times larger pitch angles that result in large geometric distortion. Thus further processing is needed to remove distortion. The algorithm firstly used coordinate transformation to compute top-view coordinates of all pixels in the image. The new coordinates were used for reorganizing the pixels. Due to the large geo-metric distortion, the Pixel Combination was applied. After computing the top-view image, global alignment was applied to generate an initial image mosaic. This global alignment method was based on the geographical information of each image, which allows a pixel-level mosaic, without limitation of detection of feature points. The mosaic image was then improved by local alignment. In local alignment, the movement of each pixel was computed by an optical flow method. For covering the entire field, seventy-one images were taken for the Energy Farm near the campus of the University of Illinois at Urbana-Champaign. The accuracy of the image mosaic was tested using several markers in the field.

The other algorithm of computer vision is the image data interpretation of a vision system. A monocular vision system for a field robot was developed to replace a former binocular stereo vision system. With this system, the 3D coordinates of plants can be geometrically estimated. Many approaches of monocular stereovision use mirrors and prisms to convert monocular images to binocular ones. However, the high frequency encoder on the robot makes it possible that without any optical accessories, one camera can still recover the 3D coordinates of the object, due to the availability of real time velocity measurement. The camera installed on the robot functions similarly to the tower camera, which can transform the coordinates in the image frame to the ground frame. Therefore, the monocular vision system is able to estimate the relative position of a plant in front of the robot. Furthermore, with the velocity of the robot detected, the vision system can even recover the height of the plant. The robot has an encoder installed that can measure the real-time speed of the robot. With this information, the displacement between two images can be obtained, thus the height of the plant can be estimated.

Therefore, with this method, the 3-D geometric information of a plant can be obtained.

Two types of experiments--laboratory test and field test--were conducted for evaluating accuracy. After the experiments, the results from both tests were compared to address the possible sources of errors. Compared to the ideal environment in a lab, the outdoor conditions could decrease the system's performance. However, the final results show that the method has the ability to provide relatively accurate measurement.

ACKNOWLEDGEMENTS

I would like to gratefully thank some people who have helped me a lot with this research and my life.

Dr Lei Tian, my advisor, has providing me resources and guidance during my entire study for my master's degree. His wisdom has been helpful for my research.

Dr. Hansen and Dr Grift are my committee members. Their expertise and knowledge have inspired me and encouraged me to work toward the success.

I want to thank Dr. Ahamed and Yuliang. Their work for the platform of the tower system was definitely remarkable. Yuliang also taught me a lot about the camera coordinate transformation.

My colleagues, Yonghua, Yan, Ben and Hongxin, have helped me with this project. They are like my brothers who are always glad to give a helping hand.

I also want to thank my wife, for her generousness and love. You are the one who supported me most for my life.

My parents, your support of my study in United States is really appreciated.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

Coordinate transformation is an important component in computer vision technology. Through this transformation, the real-world positions of objects in the images can be determined. In this research, two algorithms of this technology were developed. In agricultural remote sensing, this method is used for pre-processing of the image data from a tower remote sensing system. The algorithm includes geo-referencing and image mosaicing. The other algorithm is the image data interpretation for a vision system.

1.1 Agricultural Remote Sensing

Remote sensing for agriculture can be simply defined as the method that can observe a field or crop without having the sensor in physical contact (Xiang, 2008). Remote sensing image data have progressed to offer a quick method for estimating spatial and temporal crop characteristics and field variability. During the past few decades, the potential of remote sensing has drawn much attention in crop management, especially for determining nutrient status and for weed density mapping. Temporal resolution and a synoptic view are the main advantages of remote sensing. Also, the digital format of a remote sensing image allows fast processing of large quantities of data. Remote sensing was also used in optimization of inputs, such as fertilizer, herbicide and pesticide and outputs (yield) and its quality is improving. It is becoming more and more important for farmers. Other applications of remote sensing in agriculture are guidance for weed control and fertilizer management by monitoring crops and soil.

Examples of traditional remote sensing images for agriculture are satellite and aerial images. Both of them have been applied for many years and have turned out to be stable and reliable ways to obtain remote sensing images. Besides these two methods, some researchers have developed an Unmanned Aerial Vehicle (UAV) platform to provide more specific remote sensing applications (Xiang, 2008). Spatial, temporal and spectral resolutions are the major criteria for selecting a remote sensing method for site-specific management. The advantages of satellite imagery are that large areas can be captured in one image, in which information can be updated regularly for monitoring changes and the method is more cost-effective where high resolution is not essential. However, satellite-based remote sensing is facing problems including high temporal resolution of imagery due to longer satellite revisit times, total cost, cloud cover and limited spatial resolution (Cristain et al., 2007). Because of these limitations, the space-borne remote sensing platforms are not suitable for site-specific management. However, aerial images have become the focus of remote sensing researchers in order to overcome space-borne limitations. Aerial photographs are most useful when fine spatial detail is more critical than spectral information, as their spectral resolution is generally coarse when compared to data captured with electronic sensing devices. Temporal resolution is the major drawback to implementing space-borne and airborne remote sensing methods for site-specific management of energy crops. Low spatial and high temporal resolution is not suitable to adopt for site specific management. On the other hand, the ground-based sensing has an immense potential and could be popular for site-specific management.

1.2 The Vision System for a Field Robot

Agricultural mechanization is one of the most important engineering achievements of the last century. Mechanization actually is the major reason for the shift of the U.S. labor force employed on the farms. For the past one hundred years, from 1900 to 2000, the farm labor force changed from 38% to 3% (Constable and Somerville, 2003). The field robot is one of the applications of agricultural mechanization with the ability to perform field specific tasks. The tasks of guidance and sensing for field robots are usually conducted by a machine vision system. Machine vision has revealed its potential as a weed sensing method. For precision agriculture purposes, machine vision systems can identify individual weeds against crop plants and soil (Jeon, 2008). Advantages of using imaging technology for sensing are that it can be fairly accurate and nondestructive, and it yields consistent results. Applications of machine vision technology will not only improve agricultural industry's productivity, but also will reduce costs and make agricultural operations and processing safer for farmers.

1.3 Research Motivation and Outline

A tower remote sensing system was established at the Energy Farm near the campus of the University of Illinois at Urbana-Champaign, designed for providing near real-time remote sensing images. The major device of the system is a 3-CCD camera. The camera was installed on a platform that is capable of rotating vertically and horizontally, allowing the camera to scan over the entire field to acquire remote sensing images. Because each image can only carry a small portion of information on the field, an algorithm was developed to locate the current position in the image. The algorithm, named geo-reference, was based on readings from a digital compass installed on the camera.

After scanning over and taking images, multiple images of the entire field were acquired. There were many images with different distortions, and each image contained only part of the field. An image mosaic was needed for obtaining a more easily readable map of the entire field, for future analysis. Because the distance from the camera to the ground was small, and plants within one plot were similar, feature-based image mosaic algorithms may be not suitable. Therefore, the algorithm for mosaicing the images was designed for combining images based on their geographical information provided by geo-reference.

Many field robots have been equipped with a binocular stereo-vision system to identify and measure weed in the surroundings. Jeon (2008) has developed a stereovision system and an image processing algorithm to identify weeds and estimate their locations from a field image with respect to a pre-defined origin. The ability to estimate the 3D coordinates of plants is the major reason for using a binocular stereo vision system. However, the high frequency encoder make using monocular vision system to estimate 3D coordinates possible by the coordinate transformation. The reason of using monocular vision system is that it has lower cost and wider range of applications. The algorithm for the monocular vision system can be combined or integrated in other camera sensing systems, such as spectral sensing systems. The approach of measuring 2D coordinates of plants with a monocular vision system is explained in this thesis, as an application of a camera coordinate transformation. While measuring the movement of the robot, the heights of plants can be measured. This measurement can be the reference for other site specific applications of the robot system, such as cutting weeds and applying chemicals.

CHAPTER 2 LITERATURE REVIEW

This chapter reviews some related researches completed by other researchers. Specific topics include the geo-referencing and image mosaic. Some approaches of monocular stereovision systems are also presented.

2.1 Remote Sensing for Agriculture

The advantage of using remote sensing technology is that it can obtain spatially and temporally variable information for precision farming. This has been proved by many researchers. Remote sensing imagery can be obtained either through satellite-based sensors or video digital cameras on board of small aircraft. Remote sensing, a milestone in sensing technology, has became the most scientific and modern approach to sustainable agriculture (Ray et al., 2001). Images from remote sensing can be used as base maps in variable rate applications of fertilizers, pesticides or other chemicals. Moreover, the remote sensing enables the identification of visually undetectable properties of plants. Remote Sensing has the unique capability of recording data in visible as well as invisible (i.e. ultraviolet, reflected infrared, thermal infrared and microwave etc.) parts of the electromagnetic spectrum. Therefore, problems within a field can be identified and located remotely before they can be visually identified. Certain phenomenon, which cannot be seen by human eyes, can be observed through remote sensing techniques. For example, trees, or crops, that are affected by disease, or insect attack can be detected by remote sensing techniques before human eyes see them (Balaselvakumar and Saravanan, 2002) (Nowatzki et al., 2004). Remote sensing has also been applied in water management. Ines et al. (2006) explored water management options for irrigation in agriculture. They used both remote sensing simulation modeling and genetic algorithm optimization. Remote sensing, in their research, was used to recognize the regional system through a random data assimilation approach. Then their water management optimization model utilized the derived data as inputs. Bastiaanssen et al. (2000) have also done similar work. Their studies have demonstrated a process named “soil vegetation atmosphere transfer process”. The relationship between crop and water management was interpreted by this process. Moreover, they demonstrated examples in water management of relationship between researchers and practitioners. First, they illustrated where research tools and techniques have practical applications. Secondly, by identifying problems that remote sensing could solve, they introduced with additional research and development. Remotely sensed images can also be used to identify nutrient deficiencies, diseases, water deficiency or surplus, weed infestations, insect damage, hail damage, wind damage, herbicide damage, and plant populations without having the sensor in physical contact (Nowatzki et al., 2004). Fitzgerald et al. (2006) developed a system named EC Sampling, Assessment, and Prediction-Response Surface Sampling Design (ESAP-RSSD). They applied this system for direct ground sampling. The cotton crop attributes, including height and width could be predicted specifically using aerial imagery. All the predicting processes were from suitably calibrated regression equations. A three-year remote sensing experiment was conducted for evaluation. They examined both the reliability and the validity of method.

2.2 Geo-referencing

Traditionally, the concept of image geo-reference has relied on aerial triangulation, which is based on the presence of Ground Control Points (GCP) in a certain area (Skaloud et al., 1999). However, collecting GCP can consume a lot of time and labor, and, sometimes, it is difficult to define such points in the field. Researchers have contributed various approaches to the study of image geo-reference. Rocchini and Rita (2005) tested different polynomial functions to rectify the aerial images and applied the method to various terrain types. Their experiments showed that first-order terms were most important to fitting the curve to the terrain. Their research showed an increase of the error with more rugged terrains. Thomson et al. (2002) demonstrated a continuous geo-reference method for a video based remote sensing system on an agricultural spray plane. They tested several configurations of a Differential Global Positioning System (DGPS) to evaluate the accuracy of the geo-reference result when flying at altitudes of 21 to 420m. A spatial accuracy of only –11m to +83m was reported for this system. Zhou et al.

(2005) discussed developing a real-time video geometric geo-reference scheme to be used for forest-fire mapping. Their method relied on identification of tie points and ground control points. However, features required in their method rarely exist in agricultural fields. Price and Alli (2005) tried to automatically geo-rectify images using GPS. Spatial errors were 56m on average using standard GPS and 9m on average using a GPS with a Pulse Per Second (PPS) output. Mark and Hardin (2005) developed a remotely piloted aerial vehicle for monitoring natural and agricultural resources. In order to geo-reference aerial images, they assumed the GPS would give the coordinate of the center of the aerial photos, and pitch and roll variations were ignored by their process. Toth (2002) demonstrated an Airborne Integrated Mapping System (AIMS) integrating an Inertial Navigation System (INS) and GPS. The AIMS could achieve 30cm geo-reference accuracy. However, this system used a traditional Inertial Measurement Unit (IMU) that was heavy and expensive to maintain. Superior performance specification was attributed to the high cost of such a unit. Its use on the UAV for agricultural applications is prohibitive. Another drawback for such a system is the difficulty of lever-arm calibration. In the maps of farms, the most common way of geo-reference is using the location of a farmhouse. However, some researchers have tried to develop a new method to geo-reference large farms. They conducted a study using nearly five hundred actual farm boundaries to investigate the best representation of total farms by a single point. The conclusion is that the main farm building is the best geo-reference method for practical purpose (Durr and Froggatt, 2002). Not only feature points, but 3D building models can be used for geo-reference. The information of street vector data and computer-aided design models of buildings can be used as the first step for geo-reference. This method was used in simulated Synthetic Aperture Radar (SAR) images created by using a SAR simulator. This kind of images was similar to remote sensing images, with a view angle from top (Balz, 2006).

The angle of the camera is important for determining the transformation matrix and, therefore, it impacts geo-reference results. Xiang (2008) used a gyro and Real Time Kinematic Global Positioning System (RTK-GPS) to record the attitude of a UAV system with a camera.

After capturing the images, the transfer matrix corresponding to them can be computed with the data of the gyro and GPS. The error of this system is around 50cm with a 14.2m height. The disadvantages of this application are that the GPS and gyro are expensive and the gyro may have an accumulative error. Rather than measure the angle directly, it can be calculated or at least corrected by using the observable features in the images (Campion et al., 2002). This method has been demonstrated to reduce attitude errors by more than an order of magnitude to <1 mill radian. This method may be useful for more accurate geo-reference, but it is based on airborne image data.

2.3 The Image Mosaic

Zhu et al. (2005) have done some relevant experiments. They used a video captured by an airborne camera to mosaic a large image along the path the airplane has passed. The point of their work is to find a computationally efficient method for generating a seamless geo-referenced mosaic. Wu et al. (2007) have done a similar work. They used resample navigation data to estimate the attitude of the camera, and then mosaiced the images with video sequence. Because the video was captured by an airborne camera, there will be no large geometric distortion effect caused by the angle between camera frame and ground. Lin and Hsiung (2008) have developed a system of virtual reality GIS with stereovision. Their system can create merged panoramic 3D models of the surroundings around the camera, based on the images captured. The mosaic image in this case is composed of images with geometric distortion. However, the tower remote sensing system needs a large image of the whole field, not a 3D model. Because of the perspective effect, the images cannot be connected directly. Bertozzi and Broggi (1998) have developed an algorithm to remove the perspective effect. Given the parameters computed in camera calibration, the pixels of the image can be related to the ground. Through this relationship, a top view image of the field can be constructed. They have verified that this method did work for lane detection for vehicle, by constructing an image of the road and environment in front.

Shum and Szeliski (2000) have done some research on panoramic image mosaics. They use global and local alignment for accurate image mosaics. Rather than measuring the angles of the camera directly, their method can estimate the necessary parameters, including camera attitude and focal length. With these parameters determined, the translation can put the image together in a same plane. After global alignment, most of the accumulated registration errors were reduced. Then, to compensate for small amounts of motion parallax introduced by translations of the camera and other un-modeled distortions, local alignment, which was based on the results of pair wise local image registration, was applied.

Rather than find feature points for an image mosaic, Scheidt (2007) described a method for generating a seamless and radio metrically accurate Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) Thermal Infrared (TIR) mosaic of atmospherically corrected radiance and from that, extract surface emissivity for arid lands, specifically, sand seas. Canty et al. (2004) demonstrated a successful example of mosaic by automatically selecting Pseudo-invariant features (PIFs) between bi-temporal images using the Multivariate Alteration Detection (MAD) technique (Nielsen et al., 1998), and they emphasize a number of unique characteristics that are important to their mosaic technique: The selection of PIFs was not manual or subjective except for one decision threshold, based on scale-invariant criteria, and corresponded to physical characteristics of the land surface. Their results compared favorably with other manual methods, and their technique was fast and automatic. After testing, orthogonal linear regression of PIFs was preferred to Ordinary Least Squares (OLS) regression.

2.4 Monocular Vision System

There have been many attempts to use a monocular camera to replace normal stereovision systems with two cameras. The first attempt may be used monocular information as an additional for improving stereo algorithm (Mordohai and Medioni, 2006). They used monocular information, such as color, in the last post-processing stage to increase the performance in texture less areas and near depth discontinuities. To recover depth from images, the human visual system uses many monocular depth cues (Martin, 2005). Martin found out the “most powerful constraints” for image interpretations are “domain specific” in a monocular vision system and act as “visual sonar”. Saxena et al. (2009) tried to create 3D models from a single image. In their research, an algorithm to estimate the depth from a single still image is introduced. There are many monocular cues that were used for recover the depth in their research. However, in the plant detection process, only the height of the plant is necessary. Moreover, the height measurement for the plant need to be precise for future application, thus information merely form one frame of image is not accurate. Criminisi et al. (2000) provided an interactive method for computing 3D features. In this system, users can specify the object segmentations. Moreover, both the position and the height of the observed points and objects can me measured by this system. This method is based on perspective effect, thus most of the processed images were take from sides of objects. Moreover, the algorithm is effective to measure the height of regular cubes, since this algorithm needs the computation of a vanishing line. Therefore, this method may be difficult to apply in the measurement of irregular objects, such as plants.

Matsumoto et al. (1997) has developed a system that can convert a monocular image sequence to stereo. The depth of objects was sensed based on motion parallax. The grid points of two frames were matched and compared to extract depth information. The main feature of this system is that it can be applied to films already taken in the past. However, this algorithm needs iterative calculation, which is difficult to apply in future real-time system. Wang and Ishii (2009) used a monocular system to recovering depth information. A camera was installed on a helmet and then the researcher droved a motorcycle with it. The distance of trees from camera along the street was computed using captured images. The movements of trees in images were estimated by optical flow. However, limited by equipments, they did not have an accurate speed sensor for the motorcycle thus they estimated the acceleration. This limited the accuracy of the system. Pachidis and Lygouras (2007) had developed a “pseudo-stereovision” system for real-time robot application. They used three mirrors, a beam splitter and a single camera to act as a stereo-vision system. The mirrors were fixed in front of the camera with 45 degree angle, which generate two virtual cameras. The beam splitter permits the reflection of 50% of the incident light and propagation of the other 50%. As the incoming light came from two directions, 50% of the light in each direction is lost, whereas the other 50% is driven to the camera lens. Then a complex image is acquired by the system in one shot. The expected information, including the disparities and the depth of objects is obtained using the proper algorithm. This method can estimate the parameters of object with only one shot. However, the mirrors and beam splitter increase the complexity of the system. Also, since the image from two visual cameras overlapped with each other, it is only applied for simple environment. For the plants on filed, the disparity will be very difficult to obtain. Similarly, Teoh and Zhang (1984) developed a monocular stereovision system using a Sanyo video camera with a microcomputer. In their system, two mirrors are fixed at a 45^{o} angle with respect to the optical axis of the camera, along with a third mirror mounted on the shaft that can rotate through a 90^{o} angle. Consequently, the system will act the same as two cameras with parallel optical axes. Note that since two shots are required, the camera should only be used in static scenes. Nishimoto and Shirai (1987) also proposed a monocular stereo-vision system. They placed a glass plate instead of a mirror in front of the camera. If the glass plate rotates at a small angle, the optical axis of the camera will shift slightly, simulating two cameras with parallel optical axes. The pair of stereo images obtained can only provide coarse-depth values. This camera system requires two shots from a scene, and therefore, it can not be applied in plant identification process on a moving robot.

The heights of objects on the ground can be computed based on several sequence frames of input video. This method acts conversely as visual odometry, which can estimate the motion of a single moving camera (Nistér et al., 2005). In their research, the estimation of camera movement is for navigational purposes. The point features were matched between pairs of frames for estimating the camera motion. The “geometric hypothesize-and-test architecture” was implemented for feature tracking. Further, the camera pose can be measured using the 5-point algorithm between three of the frames (Nistér, 2003). The camera pose is important for coordinate transformation from camera frame to ground frame. His algorithm can provide accurate navigational information.

Through the literature review, several relevant researches were presented. The designs and methods have inspired the geo-referencing and image mosaic for the tower remote sensing system and the monocular vision system.

CHAPTER 3 GEO-REFERENCING FOR A TOWER REMOTE SENSING SYSTEM

This chapter introduces both concepts and experiments of the geo-reference for a tower remote sensing system. The details of the system's composition of the tower remote sensing system are introduced, including the camera, controllers and compass. The concepts of camera and compass calibration are also presented to support the algorithm of the geo-reference. The definitions of intrinsic and extrinsic parameters were presented. Finally, the results of geo-reference were compared with data from RTK-GPS to evaluate the system's performance.

3.1 Instrumentation and Methods

The section introduces the devises and major methods. The platform for this research is a tower remote sensing system. The coordinate transformation method was applied for geo-referencing.

3.1.1 System composition

A tower remote sensing system has been developed to provide continuous, near real-time remote sensing images for crops in the experimental fields. The tower has height of around 38 meters, standing in the middle of a 36 acres field. Figure 3.1 shows the basic layout of this system. This system can work automatically to monitor the field without an operator.

The tower remote sensing system was developed with a 4 band MS4100, a multispectral charged couple device (CCD) camera (Geospatial), a pan/tilt device (PT570P medium duty) and receiver (LRD41C21/22 Legacy®), a lens controller and digital compass. The entire system is shown in figure 3.2 (Ahamed, 2009). The multispectral camera is a digital camera with a high resolution of 1920x1080 pixels (figure 3.3). Differed from the normal CCD camera, the camera is available in two spectral configurations: RGB for high quality color imaging and CIR for multispectral applications. The camera has 3 CCD channels. These three channels have the

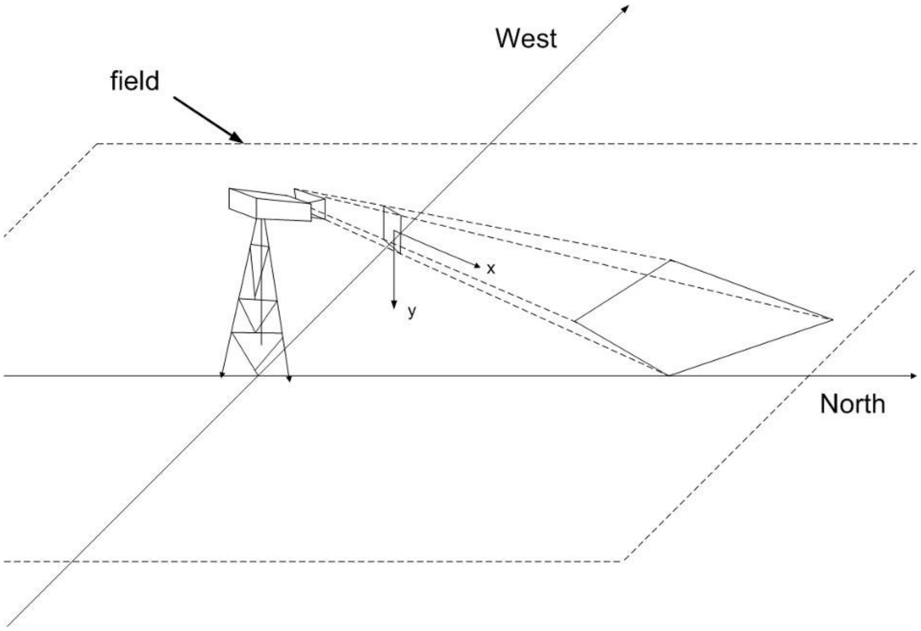


Figure 3.1. Sketch of concept of tower remote sensing system.

center wavelengths of 650nm, 800nm and 500nm, respectively and bandwidth of approximately 100nm for each (Redlake, 2000). The camera has the maximum frame rate of 10 frames per second. The camera can output 8-bit and 10-bit digital image for each channel. Through a serial interface, the external control of gain and exposure time for each independent channel via a standard RS 232 port was provided. Exposure time means the amount of time that each channel in the camera accumulates to charge before the electronic shutter is closed and resulting value is read out. The exposure time of the camera varies from 0.1ms to 108 ms that corresponds to 16-bit digital number from 1 to 1080, 1079 steps in total. The gain settings control the amount of the output signal amplification for each individual channel in the camera. The gain of the camera ranges from 0dB to 36dB corresponding to 95 to 1023 in 16-bit digital number representation and 928 steps in total.

The image acquisition device in this project was the PCI 1428. It is an IMAQ board for PCI chassis that supported a diverse range of camera. The board was installed on a small computer with PCI expansion slot (SC241S) that could operate under extreme outdoor conditions. This computer had Windows® server version installed, which allowed access from any client

(a) (b)

Figure 3.2. (a) Picture of the tower standing the field, (b) the outside appearance of the tower remote sensing system.

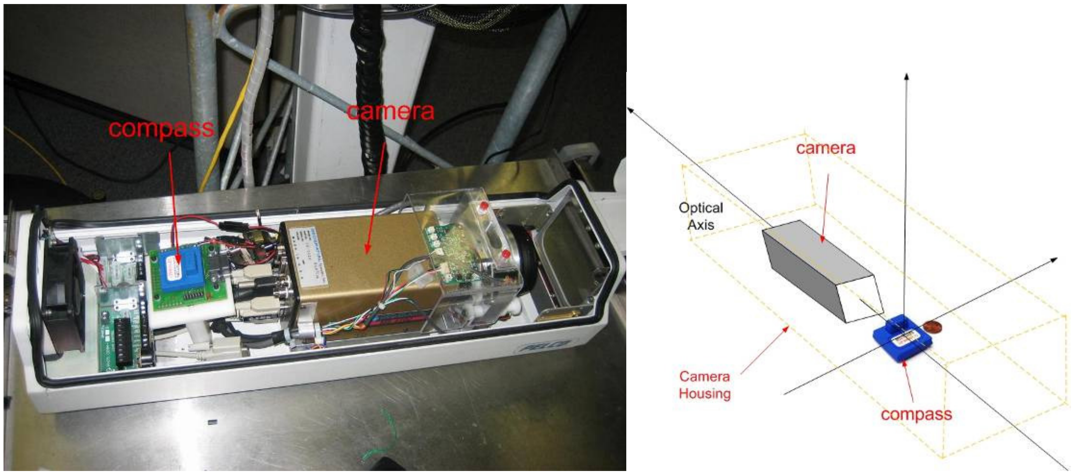
Figure 3.3. Camera, compass and lens controller inside the camera housing.

through internet. Therefore, the tower system can be controlled remotely, and all data can be transferred and stored every day. A serial port of the computer was connected to the external control port of the camera via a nine-pin serial cable. The pan/tilt device is being rotated in transverse and longitudinal directions to obtain the images according to the plot distributions. The Pelco D protocol was used to communicate with the pan tilt device and receiver using RS232 serial communication. The presets according to the field distribution are established using the caller identifications and automatic rotations of the pan/tilt device as developed. The presets provide the memory of all positions of camera that allows the automatic image acquisition. A Nikon F mount variable zoom lens (18 mm- 200 mm) which is controlled by motors and the lens controller is used to the precise movement of lens for zoom out and in. The lens motorization is developed externally and used two motors to control the zooming and focusing functions.

The novelty of this system is its image-acquisition system that can control the camera attitude to cover any specific point in the experimental field. The variation of the angle of the camera is ranged from 0 to 80 degree vertically, and 360 degree horizontally. However, this introduces the difficulty of geo-reference, since the Ground Control Point (GCP) method is no longer appropriate. Since it is impossible that to place GCPs all over the ground to make sure the existence of the makers in images with various angles of the camera. A solution might be the measurement of camera attitude, in order to determine the location in the image. To measure the attitude, a Sparton SP3004D digital compass was attached to the camera directly so that the yaw, pitch and roll of the camera could be measured based on magnetic field and gravity. The Sparton SP3004D digital compass incorporates software enabling optimized performance. A 2-D calibration algorithm has been developed to allow accurate in-field calibration for platform-based applications. Enhancements to the adaptive 3D calibration have been implemented to improve calibration speed and stability. The accuracy of the compass is 0.3 degree with 0.1 degree resolution (Sparton, 2008). The attitude information measured by the digital compass can be used for geo-reference without any information from the ground. Figure 3.4a shows the layout in the camera housing. The compass was installed behind the camera, fixed on a plastic support. For accurate measurement, the compass must be installed along the optical axis of the camera (figure 3.4b).

3.1.2 Camera coordinate transformation

The coordinates of ground and image are related by intrinsic parameters and extrinsic parameters (Forthy and Ponce, 2005) which are listed in Table 3.1. There are three frames of three coordinate systems defined in this system for the overall coordinate transformation.



(a) (b)

Figure 3.4. (a) Compass and the camera inside the camera housing, (b) sketch of compass installation.

They are: image frame, camera reference frame and ground frame. There are three frames in this overall coordinate transfer: ground frame, camera frame and image frame. Ground frame is the idealized plain of ground, which is absolutely horizontal. Camera frame is transformed from ground frame by extrinsic parameters, which is intercepted with ground frame and parallel to CCD sensors of camera. The final one is image frame, which is overlapped with CCD sensors. We distinguish the intrinsic parameters, which relate the image coordinate system to camera reference frame, from extrinsic parameters, which relate the coordinate system in camera reference frame to a fixed ground coordinate system.

The intrinsic parameters include the focal distance (*fc*) and the coordinate of the principle point (*cc*). The extrinsic parameters, including three rotation parameters and three translation parameter, cannot be determined in camera calibration. These parameters will be estimated by other sensors, such as compass and GPS.

There are two frames, image frame and camera reference frame, which associate with the intrinsic parameters (figure 3.5). The two frames were related by the pinhole, which in reality is the centroid of the camera lens. The object in front of the lens will be projected inversely on

Table 3.1. List of intrinsic and extrinsic parameters.

Intrinsic parameters		**Extrinsic parameters**		
fc	Focal distance	R	Φ	pitch
			θ	yaw
			ψ	roll
cc	Coordinate of principle point	T	X	
			Y	
			Z	

the camera sensor through the pinhole. This phenomenon is called pinhole perspective projection. The basic idea of perspective effect, which is essential in camera calibration, can be defined as:

(a) Farther objects appear smaller than closer ones;

(b) The projections of two parallel lines lying in some plane appear to converge on a horizon line formed by the intersection of the image frame with the plane parallel to lines lying plane and passing through the pinhole.

Intrinsic parameters explain the relationship of coordinate of camera reference frame and pixel frame, while extrinsic parameters are keys to transformation between ground frame and camera reference frame. The coordinate transferring from camera reference frame to image can be written as:

$$\begin{bmatrix} x_{image} \\ y_{image} \\ 1 \end{bmatrix} = \begin{bmatrix} fc(1) & 0 & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{cam\_frame} \\ y_{cam\_frame} \\ 1 \end{bmatrix} \qquad (1)$$

Focal distance (*fc*) is determined based on the focal length and the size of one pixel on physical CCD sensor. The image coordinates of one specified image point are expressed in pixel units. Therefore, *fc* has the unit of inverse physical units, for example, mm^{-1}. After *fc* applied, the coordinates in camera reference frame can be transformed into pixel units, represented in the image. In addition, since pixels are normally rectangular instead of square, so there are two elements in *fc* that are along transverse and longitudinal directions, respectively.

The principle point is the projection of the intersection of optical axis of lens and the image frame. In general the origin of the camera coordinate system is at a corner of the image instead of the center. In this paper, the origin was defined as the upper-left corner, while the image coordinates are the row and column indexes of a pixel. Also, the center of the CCD matrix usually does not coincide with the principle point. This adds the parameter *cc* that defines the position (in pixel units) of principle point in image coordinate system. For example, in this experiment, since the resolution is [1912 1076], the principle point should be close to but not exactly [956 538].

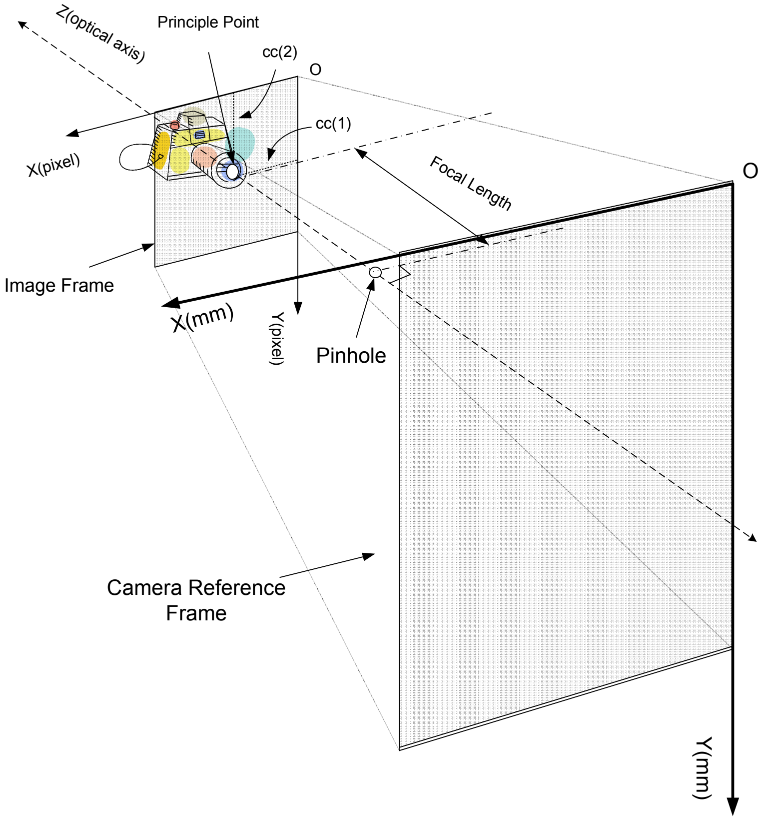


Figure 3.5. Definitions of pinhole, image frame and camera reference frame.

The distortion effect can affect the geo-reference results significantly. Thus, another transformation procedure needs to be applied. There are two types of distortion, round distortion and tangential distortion. The model for these two types of distortion can be expressed as follows (Heikkilä and Silvén, 1997):

$$\begin{bmatrix} x_{out} \\ y_{out} \end{bmatrix} = (1 + kc_1 \cdot r^2 + kc_2 \cdot r^4 + kc_5 \cdot r^6) \cdot \begin{bmatrix} x_{in} \\ y_{in} \end{bmatrix} + dx \tag{2}$$

where

$$r^2 = {x_{in}}^2 + {y_{in}}^2 \tag{3}$$

$$dx = \begin{bmatrix} 2kc_3 xy + kc_4 (r^2 + 2x^2) \\ kc_3 (r^2 + 2y^2) + 2kc_4 xy \end{bmatrix} \tag{4}$$

The parameters kc_1, kc_2 and kc_5 are the coefficients of the round distortion model, and dx is the tangential distortion vector. Therefore, the 5-vector 'kc' contains both radial and tangential distortion coefficients

The transformation matrix relates the ground frame to the camera frame. It is composed of the extrinsic parameters, including three angles of the camera and the distance from camera to ground. The three angles are pitch, roll, and yaw, which can be determined using a digital compass. The distance from the camera to the ground was measured during system installation. Figure 3.6 shows the definition of the three angles in extrinsic parameters. The mechanism of the compass determines the definition of the three angles. Pitch is the angle of the camera under the horizon plane, while yaw is the angle between camera's optical axis and due north. These two angles are most important in geo-reference for this project. Roll is the angle of camera turns along the optical axis. In most of the cases, the roll angle will be zero.

The relationship between the coordinates in camera reference frame and ground frame

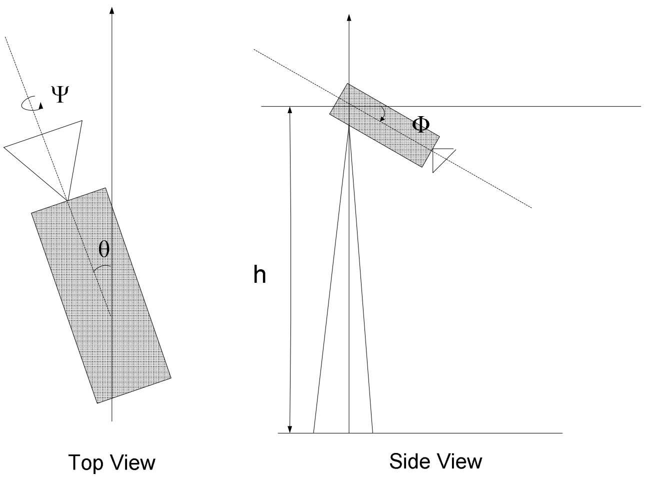


Figure 3.6. Definitions of three angles, pitch, roll and yaw.

was demonstrated in the following equation:

$$x_{cam\_frame} = T_{transfer} \cdot X_{ground} \tag{5}$$

where x_{cam_frame} is the coordinate in the camera reference frame, and X_{ground} is the corresponding coordinate of the ground. There are generally two parts in the transformation process: rotation (*R*) and translation (*T*). Then the relationship between two coordinates can be expressed by:

$$x_{cam} = R \cdot (X_{ground} + T)$$

$$R = \begin{bmatrix} \cos\theta\cos\psi & -\cos\varphi\sin\psi + \sin\varphi\sin\theta\cos\psi & \sin\varphi\sin\psi + \cos\varphi\sin\theta\cos\psi \\ \cos\theta\sin\psi & \cos\varphi\cos\psi + \sin\varphi\sin\theta\sin\psi & -\sin\varphi\cos\psi + \cos\varphi\sin\theta\sin\psi \\ -\sin\theta & \sin\varphi\cos\theta & \cos\varphi\cos\theta \end{bmatrix} \tag{6}$$

$$T = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}$$

R is the rotation matrix that represents the angles between the camera frame and the ground frame. *θ, Φ,* and ψ are the yaw, pitch, and roll of the camera attitude. R matrix here is a combination of three rotations due to pitch, yaw and roll angles, respectively. Because the rotation is according to the axis of camera coordinate system, the rotational axis will change after each rotation. Thus, the sequence of the three rotations is fixed. First is the rotation caused by pitch, and then followed by rotation of yaw angle. Finally, the roll angle rotation will be applied.

T is the translation matrix, which represents the position difference of camera and the origin of the ground coordinate. Considering the fact that, in the field, the origin of the ground coordinate system is the position of the camera's projection on the ground, then T has the form [0 0 *h*]' where *h* is the distance from the camera to the ground, which is the height of the tower.

The overall coordinate transfer from ground frame to image frame is a four-step process, which can be written as:

$$x_{cam} = R \cdot (X_{ground} + T) \tag{7}$$

$$\begin{bmatrix} x_{cam\_norm} \\ x_{cam\_norm} \end{bmatrix} = \begin{bmatrix} {}^{x_{cam}}\!\!\diagup_{z_{cam}} \\ {}^{y_{cam}}\!\!\diagup_{z_{cam}} \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} x_{image} \\ y_{image} \\ 1 \end{bmatrix} = \begin{bmatrix} fc(1) & 0 & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{cam\_norm} \\ y_{cam\_norm} \\ 1 \end{bmatrix} \tag{9}$$

$$\begin{bmatrix} x_{im\_undistorted} \\ y_{im\_undistorted} \end{bmatrix} = (1 + kc(1) \cdot r^2 + kc(2) \cdot r^4 + kc(5) \cdot r^6) \cdot \begin{bmatrix} x_{image} \\ y_{image} \end{bmatrix} + dx \tag{10}$$

The ground coordinates are initially transformed by extrinsic parameters, including translation matrix and rotation matrix. This transfer relates ground coordinates and camera frame coordinates. Following is normalization process. Because the camera coordinate system that transformed from ground coordinate system is a 3D system, so it has an extra Z-axis, comparing to 2-D coordinate system of image frame. Thus the normalization is necessary. Both x and y components were divided by z component for a normalized camera frame coordinates. Then this normalized coordinate was transformed by intrinsic parameters, creating coordinates in the image frame. Finally, distortion model was applied for adjusting computed coordinates. This is the whole process of coordinate transformation.

3.1.3 Camera and compass calibration

Camera calibration is essential in image geo-reference. The coordinates in ground and image are related by intrinsic parameters and extrinsic parameters. Both intrinsic and extrinsic parameters can be theoretically computed through the camera calibration process. The maximum-likelihood estimation was used to estimate the parameters (Zhang, 2000). In the calibration process, several images of the calibration panel from different angles were captured. With the detection of corners of grids on the calibration panel, several pairs of ground and image coordinates could be obtained. The least square method was applied to get the relationship between these pairs. After the calibration, the intrinsic parameters are invariant under different situations as long as the focal length does not change.

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\| m_{ij} - \hat{m}(A,R,T) \right\|^2 \tag{11}$$

In this equation, m_{ij} is a point in the image, and $\hat{m}$ is the projection of this point, computed by rotation and translation matrix. By minimizing the differences between these pairs of points, the parameters can be determined. Figure 3.7 shows ten images of a calibration panel taken for parameter estimation. The ten images were taken from various angles. While taking the images, the camera should be properly close to the panel, making the panel occupying the entire image.

To determine the intrinsic parameters, including focal length, principle point and distortion parameters, the camera was calibrated in the lab under different focal lengths. In this case, five different focal lengths, namely, 18mm, 50mm, 100mm, 150mm and 200mm, were selected. A calibration panel with 7×9 black-and-white squares was used to calibrate the camera. The size of each square is 107mm by 107mm. Twenty images for five different focal lengths were captured for the calibration process. These images were taken with different attitudes of the

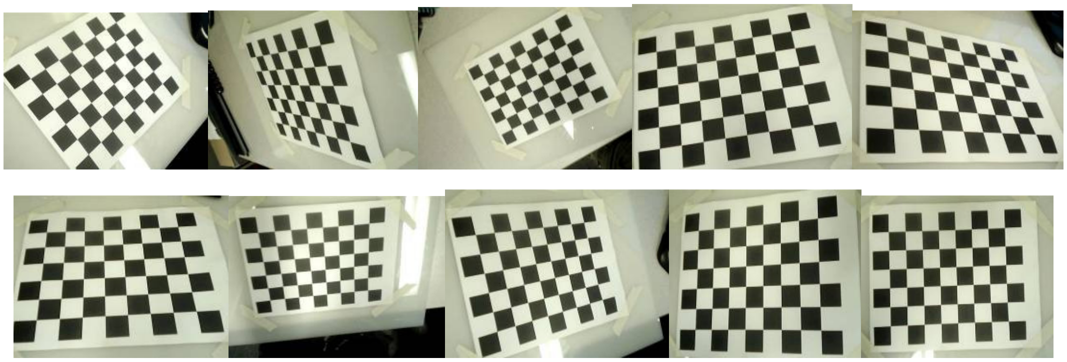


Figure 3.7. Image set for camera calibration.

camera. A calibration toolbox was used to compute the parameters. The calibration results are shown in Table 3.2.

After the calibration, the intrinsic parameters are invariant under various situations as long as the focal length does not change. However, although the calculation for intrinsic parameters can be easily performed in the lab, the extrinsic parameters were difficult to determine using camera calibration in the current circumstances because it is difficult to calibrate the camera every time if the attitude of the camera changes. In this situation digital compass is used to estimate the extrinsic parameters.

The compass also needed to be calibrated. There were generally two errors that needed to be minimized in digital compass measurement. One was the error caused by the magnetic disturbance around the compass. The other is the bias error caused by the fixed angle between the compass and the camera as a result of imprecise installation. The former error was minimized using the calibration algorithm provided by the compass producer (Sparton, 2008). The mechanism of this calibration is the determination of four disturbance factors (Xiang, 2008).

There are two types of magnetic distortions, one is the hard iron effect, and the other is the soft iron effect. The following figure 3.8 shows magnetic reading affected by both distortions and compensated compass readings. The effect of hard iron is adding a constant bias error along the two axes of the sensor’s coordinate system, which is demonstrated as a shift in the origin of

Table 3.2. Intrinsic parameters for different focal lengths.

Focal length: 18mm	fc = [2375.41096 2379.36332]
	cc = [960.24787 557.45995]
	kc = [-0.04386 -1.18455 -0.01754 0.00015 0]
Focal length: 50mm	fc = [5145.12694 5195.64427]
	cc = [1174.71208 416.64754]
	kc = [0.44058 -9.37502 -0.04806 0.03185 0]
Focal length: 100mm	fc = [10531.36215 10468.86581]
	cc = [857.69685 238.21515]
	kc = [1.24642 -38.61037 -0.06890 -0.0050 0]
Focal length: 150mm	fc = [19102.77152 19109.44275]
	cc = [753.47672 -251.45700]
	kc = [1.12519 12.27481 -0.09435 0.00140 0]
Focal length: 200mm	fc = [21905.27891 21683.32304]
	cc = [822.53759 165.15477]
	kc = [0.69227 845.9929 -0.08394 -0.00632 0]

the two circles. For compensation, all shift factors need to be determined. The soft iron distortion is due to the disturbance of the earth magnetic field and any magnetically soft material around the compass. To compensate for soft iron distortion, the two scale factors for each axes need to be determined to change the ellipsoid response to a circle.

For the determination of these four factors, the compass needs to be turn 360 degree horizontally while measuring values along X and Y direction, recorded as X_{max} ,X_{min} Y_{max} , and Y_{min}. Then the four factors are:

$$X_{sf} = 1 \ or \ (Y_{\max} - Y_{\min}) / (X_{\max} - X_{\min}) \text{whichever is greater} \tag{12}$$

$$Y_{sf} = 1 \ or \ (X_{\max} - X_{\min}) / (Y_{\max} - Y_{\min}) \text{whichever is greater} \tag{13}$$

$$X_{off} = [(X_{\max} - X_{\min}) / 2 - X_{\max}] \times X_{sf} \tag{14}$$

$$Y_{off} = [(Y_{\max} - Y_{\min}) / 2 - Y_{\max}] \times Y_{sf} \tag{15}$$

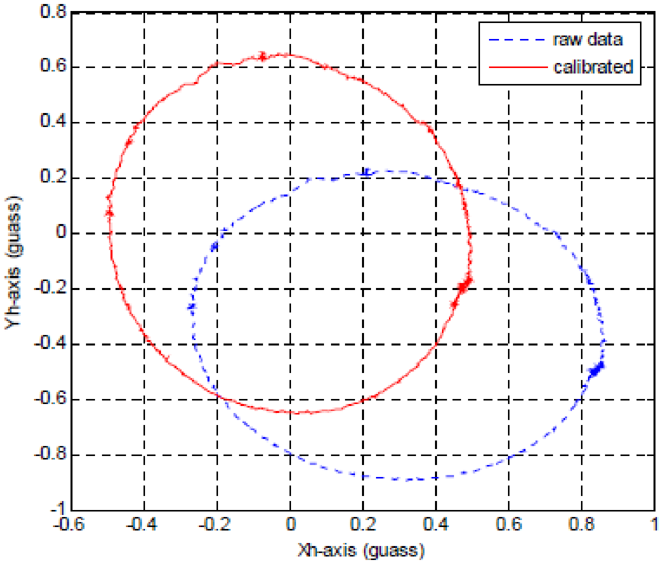


Figure 3.8. Compass calibrations (Xiang, 2008).

For the installation error, another calibration was conducted. The transfer matrix computed in the camera calibration process is relatively accurate. Thus if some image points and corresponding ground coordinates were measured, the rotation matrix can be estimated by Least Square method, just like in the camera calibration process. Then the orientation of the camera can be calculated. Comparing this result with the measurement of the compass, the bias angle between the two can be estimated.

Such calibration process was conducted before the geo-reference experiment. In this process, eight markers on the ground as objects were captured by the camera. After recording the

GPS reading and image coordinates of each marker, eight pairs of coordinates were stored (figure 3.9). Then least square method was used to estimate rotation matrix, from which the three angles were extracted. The bias angle of installation is the difference between this extracted angle and current compass reading (table 3.3).

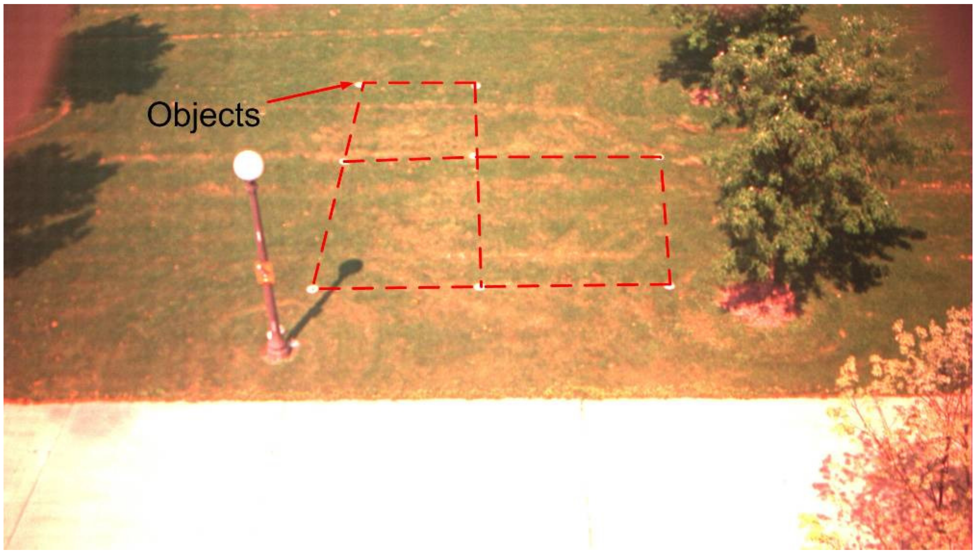


Figure 3.9. Image of calibration for bias angle.

Table 3.3. Bias angles due to installation.

	pitch(°)	yaw(°)	roll(°)
Offset	+5.1	-6.3	-4.0

3.2 Experimental Results

The experiments for the geo-referencing algorithm are introduced in this section. The experiments are conducted for evaluating the system performance. The necessary analysis of the sources of errors was also presented.

3.2.1 Geo-reference result

The geo-reference experiment was conducted at the campus of the University of Illinois at Urbana-Champaign. The top part of the tower was installed on the roof of Agriculture Engineering Science Building. Then the whole quad besides the building could be observed by the camera. The height of the tower was 15.193m, measured by RTK-GPS, which is the sum of the height of the building and tower. The image size of the camera was 1912×1076 pixels. The corresponding ground resolution ranged from 6mm to 21mm with the minimum 18mm focal length. For testing the accuracy of the geo-reference, several white boards were placed on the quad as objects. The positions of these objects were measured by a GPS. These measurements were recorded and stored for future analysis. Three images of the quad were captured with five objects in each, which means fifteen pairs of ground and image coordinates were collected to determine the accuracy (Table 3.4). The measured points in the ground coordinate with north as the x-axis, west as the y-axis, and the position of the tower as the origin, were recorded, along with compass reading for the three images (Table 3.5).

The ground coordinates were transformed into image coordinates through a transformation matrix determined by recorded compass readings. Then the computed image coordinates were compared with the objects' real image coordinates that were determined manually. The result is demonstrated in Figure 3.10. The red circle in the image represents the observed objects in the field, and the blue stars represent the result of the transformation of coordinate from ground to image. If these two points are compared, the error can be determined.

The error of this geo-reference method was 57.15 pixels on average. The corresponding ground error was between 56.2cm and 152.3cm according to the distance from the camera to the objects. The distribution of the pixel error is shown in figure 3.11, while the ground error was illustrated in Figure 3.12. There are mainly two reasons for the errors. One is the inaccuracy of the compass. Although the compass had been calibrated, the various disturbances around it could not be completely eliminated. The algorithm of geo-reference is sensitive to the change of compass, and 0.1 degree of error can cause more than 5 pixel errors in the image. The other reason could be related to the ground undulation. Further experiments and contours would be needed to understand the sloping profile of the ground. The altitude of objects could vary and was found within 20cm. The altitude error contribution also affected the geo-reference accuracy.

Table 3.4. Recorded coordinates of objects and corresponding error.

	North(m)	West(m)	Error(pixel)
Test A	-18.10	23.09	55.05
	-27.44	28.37	94.90
	-22.86	46.42	114.33
	-14.68	42.63	95.26
	-12.411	32.67	63.52
Test B	-1.42	32.73	33.52
	-10.73	37.68	33.71
	-6.63	65.34	13.37
	1.55	57.48	13.21
	4.55	45.87	7.22
Test C	-4.83	24.67	24.67
	-0.99	24.70	24.70
	3.35	24.65	24.65
	-4.95	30.80	30.80
	-1.36	30.86	30.86

Table 3.5. Recorded compass reading.

	yaw(°)	pitch(°)	roll(°)
Image1	33.3	110	3.4
Image2	6.3	114.5	4
Image3	8.7	128.2	4.4

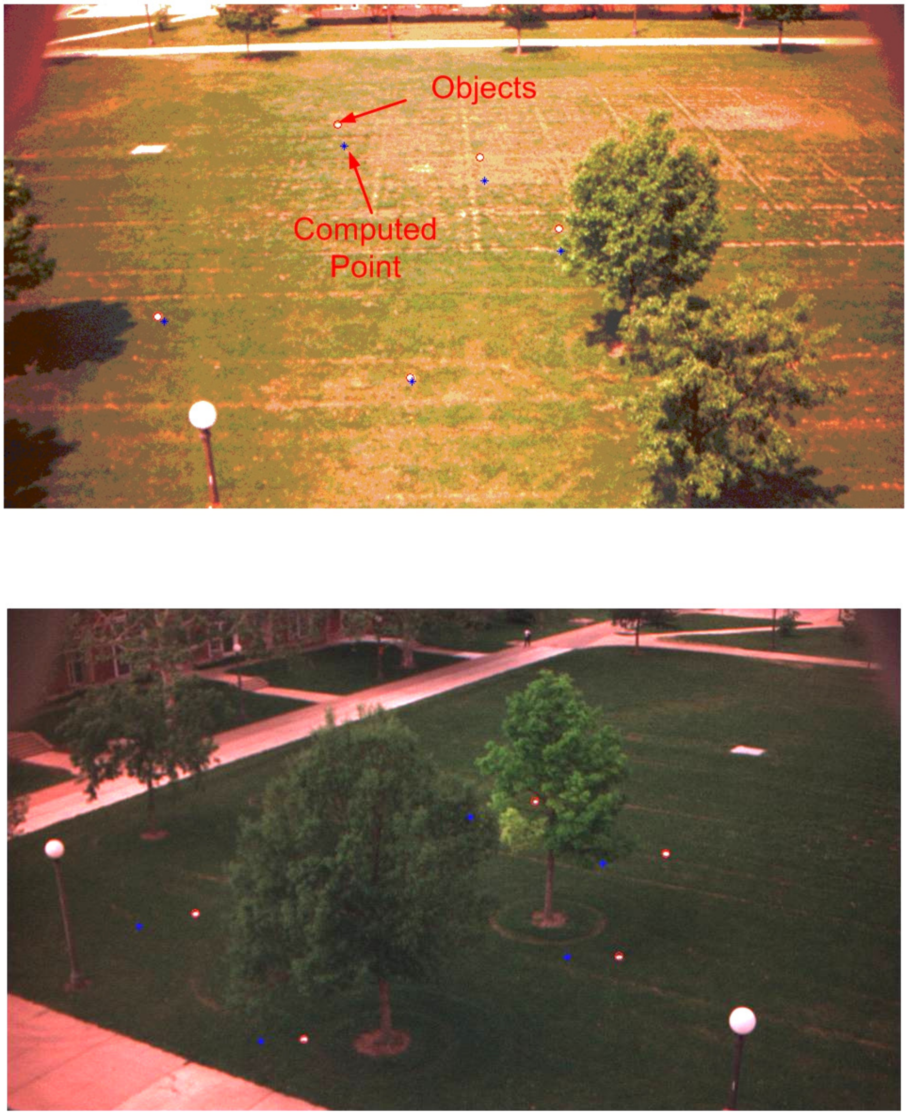


Figure 3.10. Two examples of result of geo-referencing.

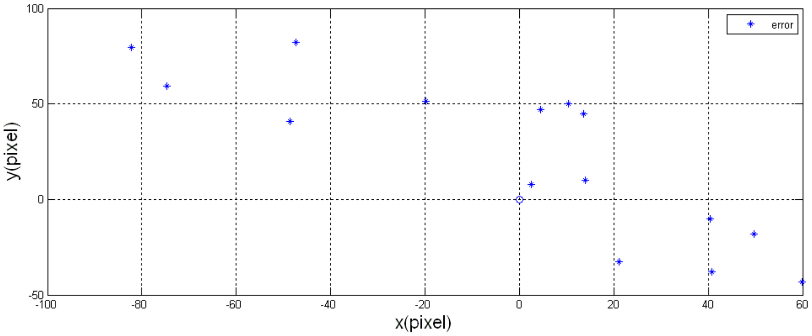


Figure 3.11. Error distribution in image coordinates.

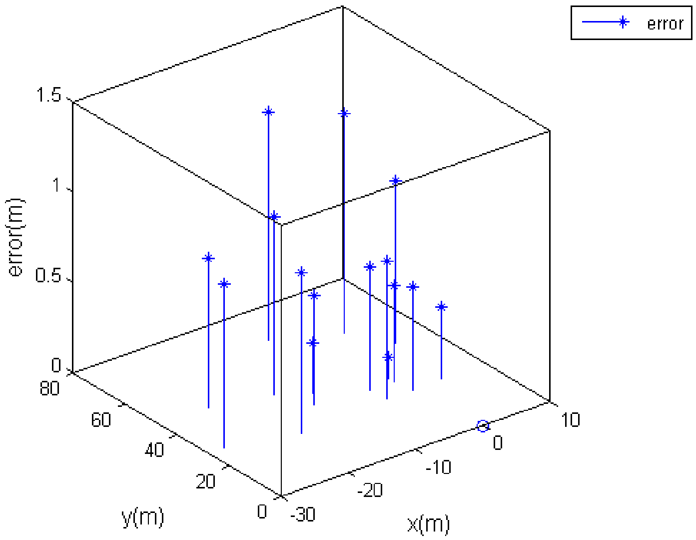


Figure 3.12. Error distribution in ground coordinates.

3.2.2 Spatial resolution analysis

Spatial resolution is the size of the area that one pixel covers in the ground. The pixel error of the system is 57.15 pixels. However, the ground error is not a number, but a range. The reason is that the spatial resolution is not consistent for different images. Further, even in the same image, the spatial resolution is not consistent. For the same pixel, the spatial resolutions along two different directions are not same. Due to the complexity of the spatial resolution, the ground error was provided as a range, since for different images, same pixel error can mean different ground error.

The difference between longitudinal (vertical) and transverse (horizontal) spatial resolution is caused by the pitch angle. Due to the angles between camera reference frame and the ground frame, the corresponding ground spatial resolution is not consistent. That means different areas of the same image have different spatial resolutions, and the area on the ground that one pixel covers is no longer square. The spatial resolution decreases along y axis within image frame. The effect is so called perspective effect. This introduces the problem that objects measured further from camera are less accurate.

Figure 3.13 demonstrates this inconsistency of spatial resolution. This image was taken in the lab. The reason of using this image as an example is that the grids can demonstrate the difference in spatial resolution. The camera was previously calibrated for the estimation of basic parameters. The panel in the image has 63 square grids with the size of 29mm X 29mm each. The pitch angle of the camera is 30 degree, causing large geometric distortion. Considering that all squares on the panel have the same size, the closer ones look much larger, with a size of 31 by 16 pixels. The farther squares have a size of 21 by 9 pixels. The transverse spatial resolution varies from 0.94 mm to 1.38 mm within the panel, while longitudinal spatial resolution varies from 1.81 mm to 3.22 mm. This picture shows how spatial resolution varies within the same image.

For further analysis, figure 3.14a shows the trends of variation of the spatial resolution of the tower system when pitch is 30 degrees. The spatial resolutions, both transverse and longitudinal, increase from top to the bottom of the image, because of the decreasing distance from the camera. When the pitch is 30 degrees, the lowest transverse spatial resolution is 22 mm while the longitudinal spatial resolution is 74 mm. The trends of this decrease will be different if the pitch angle changes. When the pitch angle becomes 60 degree, the image will be much more precise, since the distance from the camera to the scenes on the ground has decreased (figure 3.14b. The lowest transverse spatial resolution becomes 8.8mm while longitudinal spatial resolution becomes 11.8 mm. Figure 3.15a shows how the spatial resolution changes when pitch angle varies from 20 to 80 degree. The last figure is about the distribution of spatial resolution in an image of pitch angle as 30 degree. The maximum spatial resolution is nearly ten times larger than the minimum one.

For a same pixel in the image, the transverse and longitudinal resolutions are different. This is also because of the pitch angle. If the pitch angle is not 90 degrees, the ground frame and the image frame will not be parallel. For the instance which camera frame is absolutely horizontal (aerial images), the spatial resolution is only related to the distance from the camera to the ground. In this case, the transverse and longitudinal spatial resolutions are equal. The following equation shows this relationship.

$$R = \frac{S_{pixel} \cdot D}{fc} \tag{16}$$

R is the spatial resolution, including both transverse and longitudinal. S_{pixel} is the pixel size of the image sensor. D is the distance from the camera to the ground. From this equation, a conclusion can be drawn: the farther from camera to ground the lower spatial resolution.

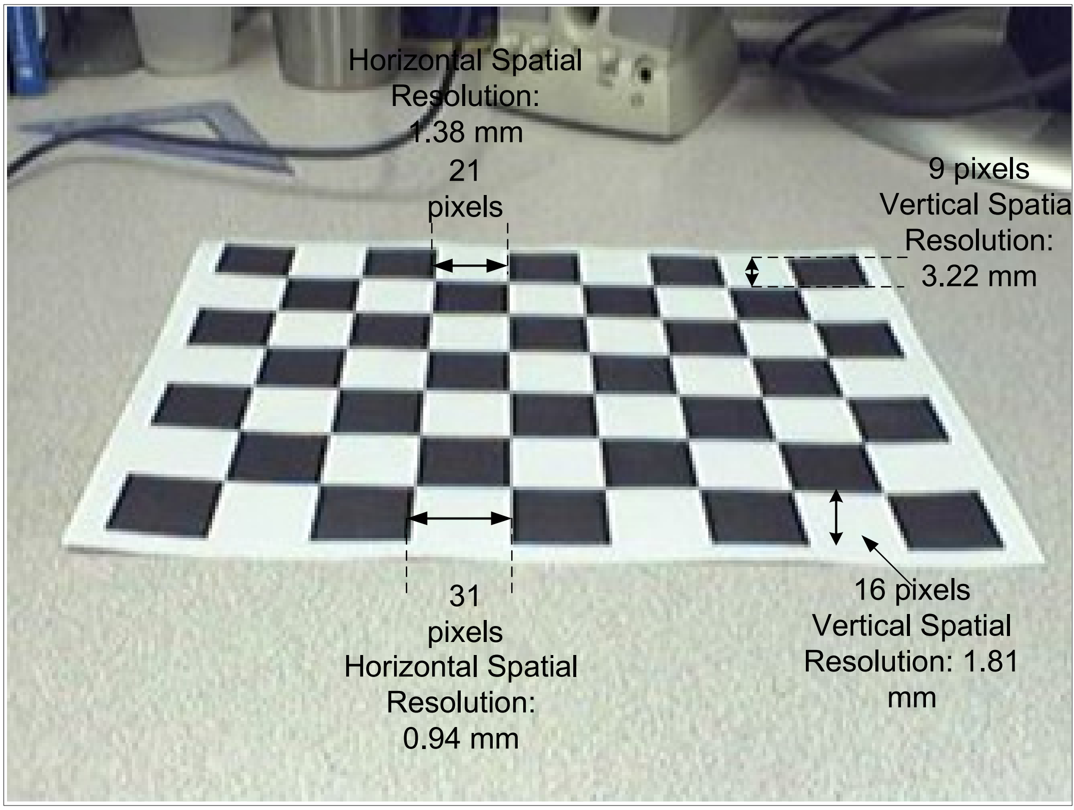
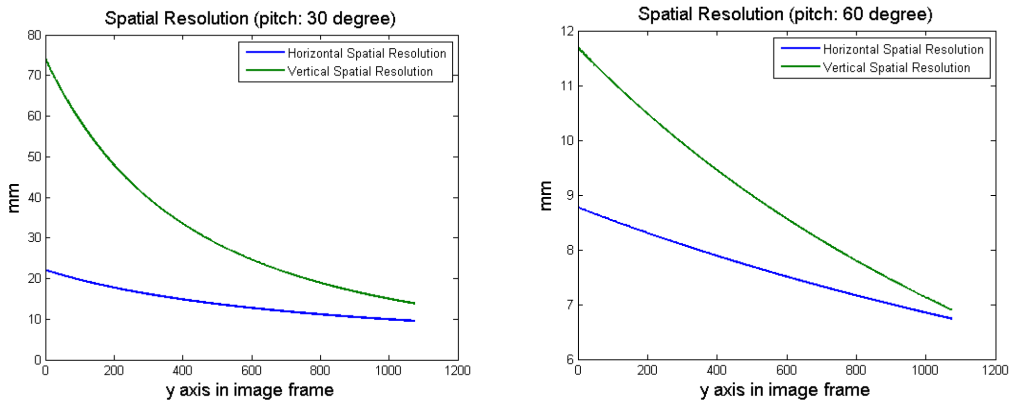


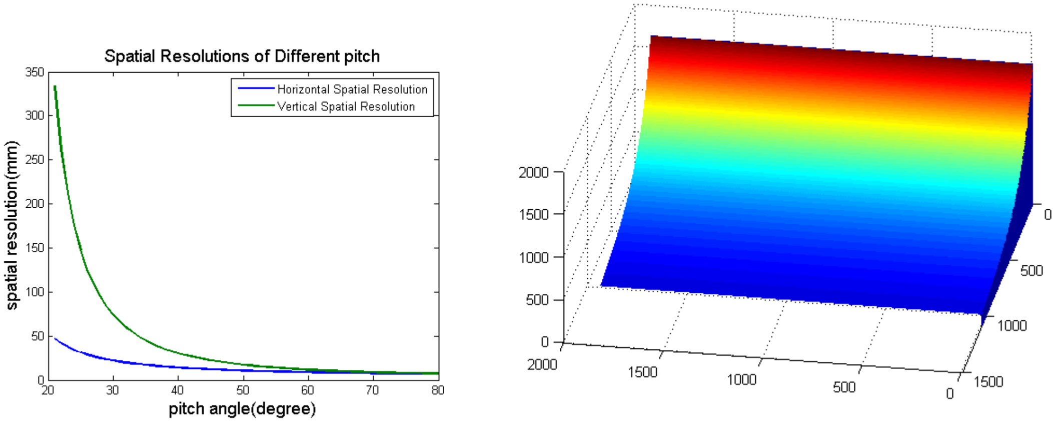
Figure 3.13. Difference between spatial resolutions.

However, in this project, the equation of spatial resolution has changed. The transverse spatial resolution will remain the same while longitudinal spatial resolution will be much larger than the transverse one. Figure 3.16 explains this difference. The left of the figure shows a camera pointing at the ground with a pitch angle. The border lines define a pixel in the camera frame. The projection of the pixel is the area that this pixel covers. In aerial images, when the



(a) (b)

Figure 3.14. (a) Variation of spatial resolution along y axis while pitch is 30°, (b) variation of spatial resolution along y axis while pitch is 60°.



(a) (b)

Figure 3.15. (a) Variation of spatial resolution when pitch changes, (b) the 3D version of variation of spatial resolution when pitch is 30°.

pitch is 90 degree, the pixel in camera frame and the projection will overlap. However, in this system, the various pitch angles make it much more complicated. The right of the figure shows more detail. *Rt* is the transverse spatial resolution, which remains same as the pixel size in the camera frame. *Rl* is the longitudinal spatial resolution. Because one pixel is very small comparing to the entire image, angle *a* is very close to 90 degree. Therefore, *Rl* is larger than *Rt*, because of the pitch angle.

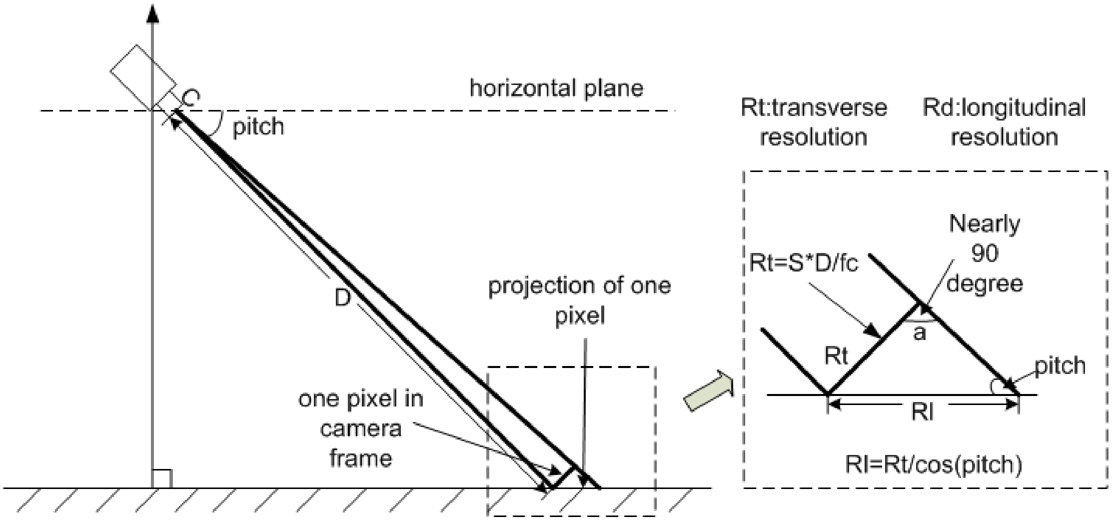


Figure 3.16. Sources of difference between transverse and longitudinal spatial resolution.

The final equations of both transverse and longitudinal spatial resolution are:

$$R_{transverse} = \frac{S_{pixel} \cdot D}{fc}, R_{longitudinal} = \frac{S_{pixel} \cdot D}{fc \cdot \cos(pitch)} \tag{17}$$

Because the errors in this experiment were all measured by pixels, a need further calculation for final ground error estimation is needed. In the image frame, the sizes of all pixels are consistent, so the errors of all experimental points can be averaged. However, because of the inconsistency of ground spatial resolution of the image, simply averaging the errors may not be an appropriate choice. Instead, the range will be provided for error analysis.

3.3 Conclusions

In this chapter, the algorithm for the geo-reference of a stand-alone tower remote sensing system was developed. Due to the rotation property of this system, traditional GCP method was no longer proper. A digital compass was used to provide necessary parameters for the geo-reference algorithm, including pitch, yaw and roll of the camera. These parameters, along with height of the tower, composed a transformation matrix for geo-referencing. Beside these extrinsic parameters, the camera was also calibrated in the laboratory, for the estimation of intrinsic parameter. Finally, a relationship between image coordinate system and ground coordinate system was established.

After calibration, a series of experiments were conducted to evaluate the accuracy. Several markers were placed in experimental fields for detection after image of this field taken. The positions of these makers were measured by RTK-GPS with a precision level of centimeters. Then these measured coordinates were transformed into the image frame. The geo-reference results, shown in the image were compared with detected markers. Then the accuracy of this method was determined in pixels. The result shows that this method of geo-reference can provide coordinate transformation with an error of 57.15 pixels, 2.9% of the measuring rage. Then this error was transformed into spatial error, according to the distance of field and camera. The corresponding spatial error ranged from 56.2cm and 152.3cm. The developed algorithm shows acceptable geo-reference performance. Therefore, the compass based geo-reference method is approved a acceptable replacement in the situation when GCPs are no longer possible.

CHAPTER 4 IMAGE MOSAIC FOR TOWER REMOTE SENSING IMAGES

After the image acquisition, all images of different view angles were stored for future image mosaic. The properties of this system resulted in some particular difficulties. Firstly, the images were taken closely to the ground, comparing to aerial and satellite remote sensing images, resulting in reduced features. Thus the feature based mosaic method was likely to be no longer feasible. Secondly, the pitch angle of the camera varies widely causing the geometric distortion very severe in this situation and the image mosaic became more difficult.

In this chapter, the image mosaic algorithm was demonstrated and divided into three steps. Firstly, the process of transforming all images into ground plane was introduced. This process, after which the images all seemed to be taken from top, was named top-view process. Then global alignment was illustrated, which combined all images together based on geographical information. Finally local alignment was introduced to reduce error of image mosaic. Some experimental results were demonstrated to evaluate the accuracy.

4.1 Methods and Algorithms

The devices used for image mosaic are same as the geo-referencing experiment. The algorithm of image mosaic is introduced step by step.

4.1.1 Top-view form

Before combining the images, each image needs to be transformed into top-view form. In this process, all positions of all pixels will be reorganized by a particular algorithm. After the rearrangement, the geometric distortion will be removed. In this tower remote sensing system, every image was taken from a different angle, which results in difficulty in image mosaic: images will not be globally consistent without transforming them into a same plane. The key of the process is the coordinate rotation, part of coordinate transformation, from image frame to ground frame. This process transformed all images into ground plane.

The process can be explained by following sequence:

1. Compute new coordinates for each pixel in the image.
2. Reorganize each pixel by computed ground coordinate.
3. Apply pixel combination.

As introduced in Chapter 3, the major difference between top-view images and original images is that in top-view images, the spatial resolution is consistent, while in the images originally captured by the tower system, the spatial resolution is not. The spatial resolution decreases along y axis of the image, making farther objects appear smaller. This is the so called geometric distortion.

For computation of top-view images, this distortion must be eliminated by the rearrangement of pixels for uniform spatial resolution. In this procedure, each pixel in the image will have a new coordinate, a new location. Functioning similarly like geo-reference, each pixel coordinate was transformed by a rotation matrix, determined by corresponding compass reading.

$$x_g = R^{-1} \cdot x_p \tag{18}$$

$$R = \begin{bmatrix} \cos(roll) & \sin(roll) & 0 \\ -\sin(roll) & \cos(roll) & 0 \\ 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(pitch) & \sin(pitch) \\ 0 & -\sin(pitch) & \cos(pitch) \end{bmatrix} \bullet \begin{bmatrix} \cos(yaw) & 0 & -\sin(yaw) \\ 0 & 1 & 0 \\ \sin(yaw) & 0 & \cos(yaw) \end{bmatrix} \tag{19}$$

R is the rotation matrix determined by compass reading. The sequence of these four matrices can be explained as following. First matrix from right rotates the yaw angle, and then followed by pitch and roll angle. Each rotation is respect to axes of image frame.

After this transformation, each pixel in the image now has a new coordinate. These coordinates are all in ground frame. If all the pixels were placed based on these computed coordinates, the distances between every two adjacent pixels will be the same. In other words, the rearranged image will have consistent spatial resolution. Then the image will become top-view.

Figure 4.1. Top-view image of figure 3.13.

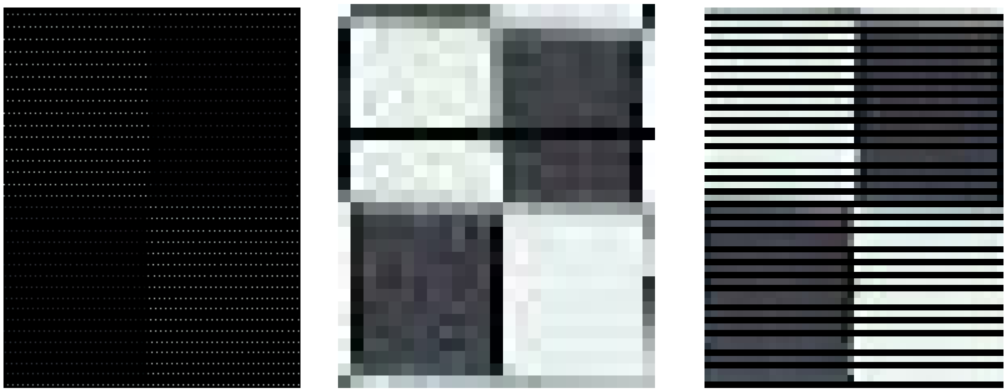
Figure 4.1 shows the top-view image. The original image is in figure 3.13. The spatial resolution now is consistent within the panel: no matter the distance to the camera, all squares now have the same sizes and same shapes. Because of the variation of the spatial resolution, the shape of the image will no longer be rectangular. The upper part of the image will be enlarged, since scenes in the upper part are farther than lower part. After this arrangement, the shape of the image was transformed from rectangular into trapezoid.

The originally computed top-view image may contain serious gaps among pixels, due to improper scales of images. To avoid this, we need another parameter to adjust the scale of transformed coordinates of the pixels. Because we are trying to use new coordinates as reference for rearranging pixels, a scalar is needed for shrinking the distance between two pixels to fit the appropriate spatial resolution. After choosing an appropriate scalar, the scale of new coordinates of pixels will fit the image size.

$$x_{p\_new} = s \cdot x_g \tag{20}$$

x_g is the computed coordinates in ground frame, while x_{p_new} is the pixel coordinates of the top-view image.

The following figure 4.2 shows different results under different scalars. If the scalar is too small, the distance between two pixels will fall below the appropriate spatial resolution, which will result in unnecessary overlapping that may lower the final spatial resolution. If the scalar is too large, the scale of computed coordinates will be oversized, resulting in gaps between every two adjacent pixels. The proper scalar is the optimized point of trade-off between spatial resolution and gaps among pixels.



(a) (b) (c)

Figure 4.2. Top-view image using different scalars, (a)image of a large scalar, (b) image of a small scalar, (c) image of a proper scalar.

4.1.2 Pixel combination

Shrinking the scale of computed coordinates to fit the appropriate spatial resolution will not solve all the problems. When pixels were tightly arranged in the transverse direction, there are still some gaps in longitudinal direction (figure 4.2c). The reason is: for the same pixel, the longitudinal and transverse spatial resolutions are different. This difference was explained in Chapter 3. This difference caused a problem that we need to make a choice between transverse spatial resolution and longitudinal spatial resolution. The shrinking of coordinate scales improves image quality but lower the final spatial resolution. If it is based on transverse spatial resolution, the gaps between pixels will lower the image quality. In other words, in this case the image is not shrunk of an enough amount. But if the scale of coordinates was chosen based on the longitudinal resolution, a lot of information will be lost. Thus, an algorithm, named pixel combination, was applied for reducing gaps between pixels. This algorithm can avoid the lost of information.

There are some concepts that need to be explained before developing a pixel combination algorithm. Considering a square object on a horizontal surface (for example, the squares on the panel in figure 3.14), if the image was taken from top, the shape of the object will be square in the image. However, if the image was taken with a pitch angle of 30 degree, the object will no longer be square. The shape will become trapezoid, and the size will shrink, due to lower longitudinal spatial resolution (figure 3.14). Thus, in top-view image, we are trying to use a shrunk, trapezoid area to cover a larger, square area, which will result in mismatches. Because the trapezoid is smaller than the square it should cover and the shape is different, some blank space will remain.

This phenomenon can also be explained of the pixel level. In aerial and satellite images, because the physical shape of a pixel on CCD sensor is square, the area that one pixel covers is also square. So the spatial resolutions of this pixel of both transverse and longitudinal directions are equal. However, in this tower remote sensing system, the shape of the area which one pixel covers is no longer square, but trapezoid, because of the large pitch angle (figure 4.3). The longitudinal spatial resolution is always larger than transverse spatial resolution of a same pixel. This means the image loses more information along longitudinal direction. So if the images were transformed into top-view, the lost information along longitudinal direction will cause some gaps.

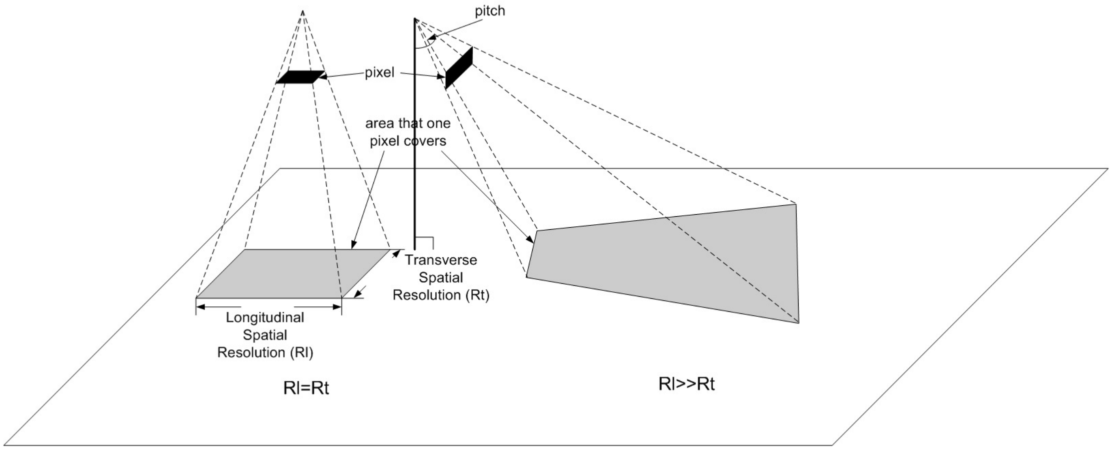


Figure 4.3. Causation of inconsistent spatial resolution along two directions.

One solution for these gaps is using one pixel to cover multiple areas that supposed to be covered (figure 4.4). That is, using one square pixel to represent the combination of information from several original pixels. The pixel value of the final combined pixel is the average of all pixel values of former pixels inside that area (Eq. 16). When the pitch angle is 90 degree, which means the captured image is top-view, there is no need to apply this algorithm. As the pitch angle decreasing, more and more gaps need to be filled. Figure 4.5a shows the top-view image without pixel combination transformed from a pitch angle of 30 degree. Figure 4.5b shows the image after pixel combination applied. Most of the gaps will be filled by this algorithm.

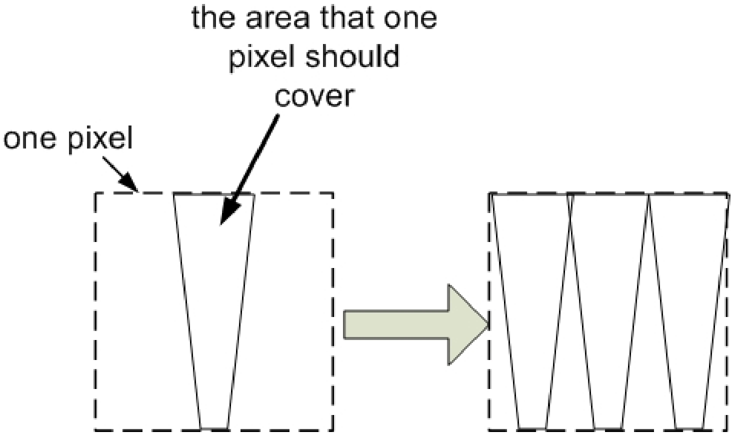


Figure 4.4. Pixel combination.

$$X = \frac{\sum_{i}^{n} x_i}{n} \tag{21}$$

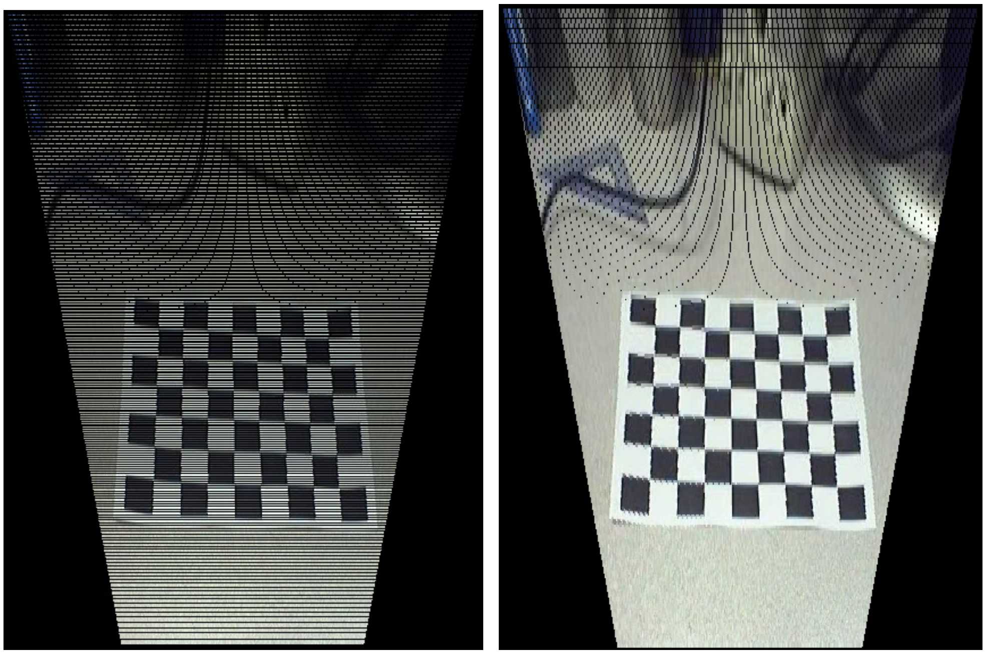
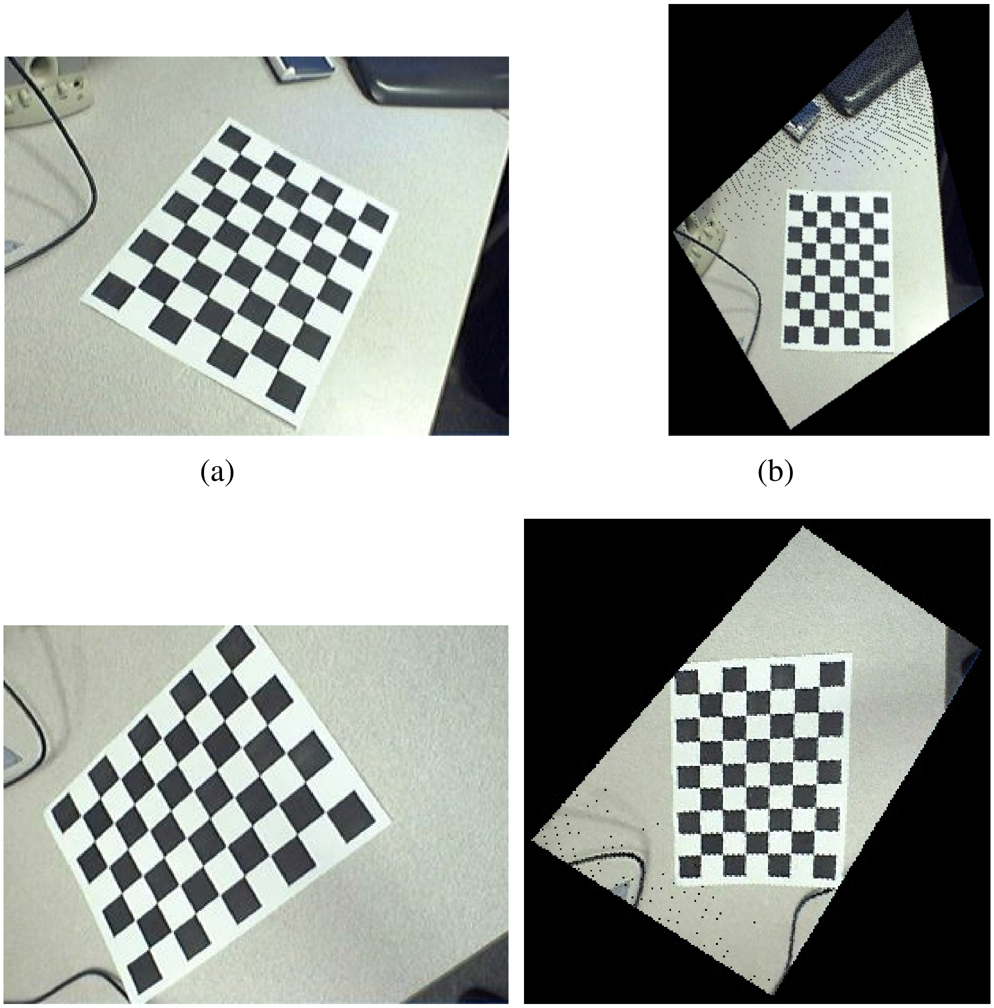

(a) (b)

Figure 4.5. Example shows the result of post processing, (a) image before post processing, (b) image after post processing.

In figure 4.6 there are some pairs of original and corresponding top-view image. These top-view images are all in a same plane, ready for image mosaic.

4.1.3 Global alignment

Traditionally, image mosaic was based on the detection of the features, such like corners or textures. After locating the same feature in two or more images, the way to combine the images was determined. However, in this project, the feature based method is no longer proper, since it is difficult to choose a unique feature in the field images. The textures and appearances of plants in the field are quite similar (figure 4.7).



(a) (b)

(c) (d)

Figure 4.6. Two pairs of original and corresponding top-view images, (a) and (c) original images, (b) and (d) the corresponding top-view images.

The compass can provide necessary attitude information of each image for image mosaic. The geo-reference algorithm developed in chapter 3 can provide the necessary geographical information of each pixel in ground frame. The transformation from image coordinate into ground coordinate unified various image coordinates into a same ground coordinate system. Then this relationship can be used for roughly matching the images, called global alignment.

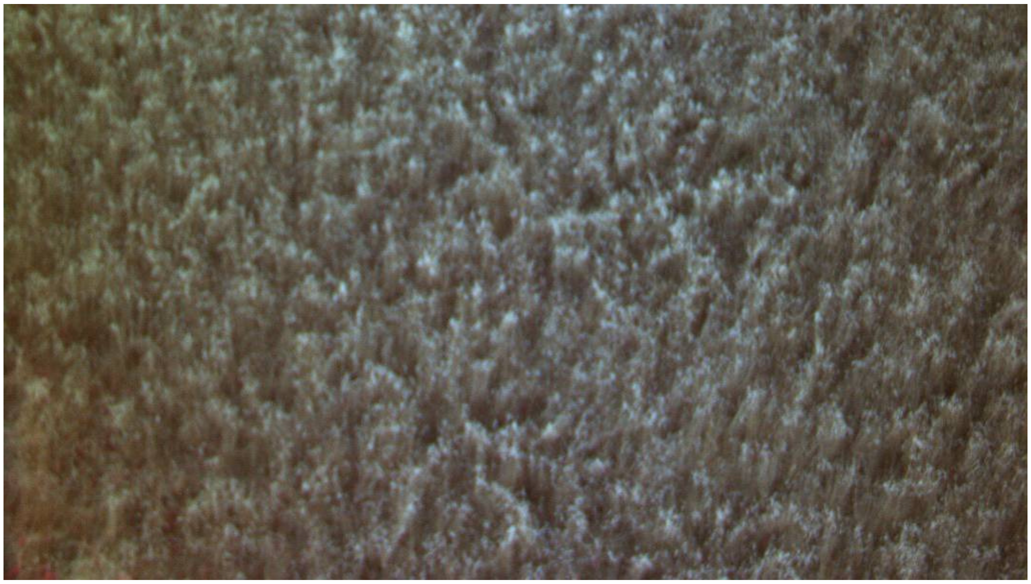


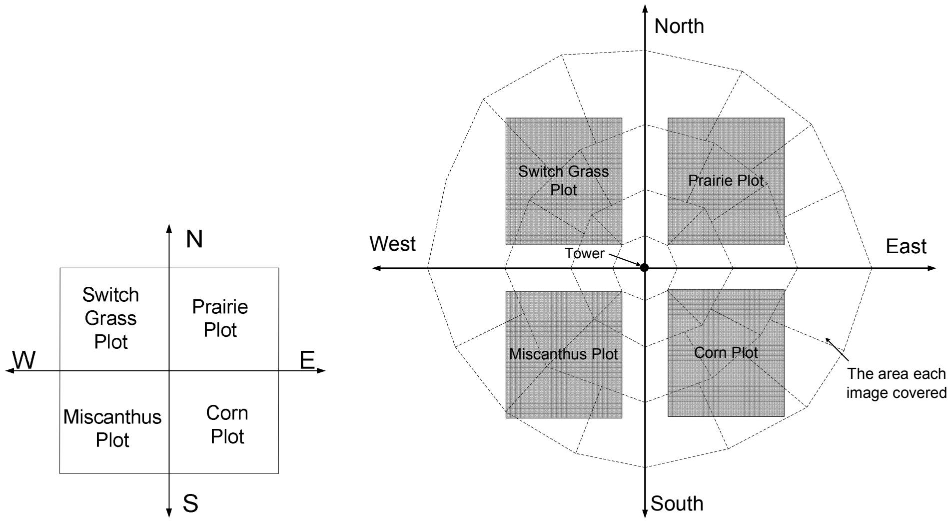
Figure 4.7. An example of lack of features in images from tower remote sensing system.

The two axes, x_i and y_i, of image coordinate system in image frame are transverse and longitudinal, respectively. The three axes of ground coordinate system in ground frame are: east direction as X, south direction as Y with Z axis pointing into the sky. After image coordinates transformed, the computed top-view images will be north oriented. In this project, there are 4 plots need to be covered. They are prairie, switch grass, miscanthus, corn fields, located at northeast, northwest, southwest, southeast, respectively (figure 4.8a).

For each single plot, 16 to 20 images were taken daily, which can cover the entire plot. All the camera attitude information for each image was stored in the camera housing controller as presets. The presets will be called one by one in fixed sequence during image acquisition process. While setting the presets for image acquisition, the overlaps between every two adjacent images were guaranteed. These overlaps were used for further local alignment. The arrangement for these images was calculated based on view angle of the camera and sizes of the plots. There are four to five rows of images for one plot. The images from same row will have same pitch angle. After determining the number of rows, four to five images per row were taken one by one. The entire arrangement will be similar to a spider web.

After the image acquisition process, all images will be transformed into top-view form.

Then, according to the geographical position information for each image, all the relative positions of top-view images in the image mosaic were determined. Since in the top-view transformation process, images were already north oriented, and all top-view images were arranged with only translation, the created image mosaic was also north oriented.



(a) (b)

Figure 4.8. (a) Four plots in the experimental field, (b) the image coverage.

Figure 4.8b shows four square plots around the tower. In the middle there are two roads across the entire field. The sizes of these plots are nearly identical. Figure 4.9 shows an example of image mosaic in corn field. 8 images were taken and jointed together.

4.1.4 Local alignment

After the global alignment, there may still be localized mismatches in the image mosaic. This is due to the error in compass reading. The error in readings may be because of the compass accuracy limitation, imperfect compass installation, and the swing of the camera. The accuracy of the compass is claimed 0.3 degree by manufacturer. However, since the compass need to be calibrated before measurement, due to the disturbance of the environment, the actual reading accuracy can not reach 0.3 degree level. When the compass was installed in the camera housing,

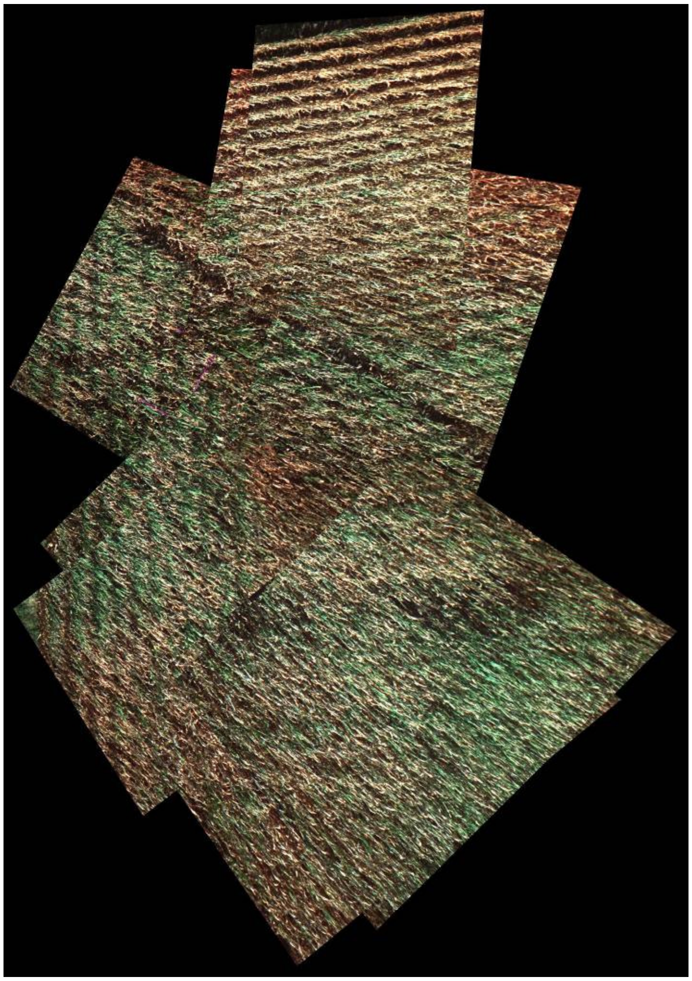


Figure 4.9. Result of an 8 image mosaic.

the axis of compass may not be absolutely parallel with camera axes. Moreover, the wind can make the tower swing severely, which brought error in measurement.

To compensate for this effect, the amount of these mismatches along both x and y axes in image coordinate system need to be quantified, and then the images can be locally adjusted. The local displacement between two images was estimated by optical flow. Optical flow is “the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene” (Brox et al., 2004). This method can estimate the “velocity” of each pixel between two similar images. This velocity, in fact, is the placement of pixels from one image to another. From the velocity distribution, the average placement between two images, which is the amount for adjustment, can be computed.

The main energy function was minimized to compute the optical flow. The computation is a combination of both appearance and smoothness terms of the images. Appearance means the textures and the difference in brightness, while the smoothness means the variation of brightness changing. Optical flow computes a flow field by trying to minimize the difference between the two images. The main energy function can be expressed as:

$$\begin{aligned} &E(u,v) = E_{data}(u,v) + \alpha E_{smooth}(u,v) \\ &E_{data}(u,v) = \int_{i,j} (|I(x+w) - I(x)|^2 + \gamma |\nabla I(x+w) - \nabla I(x)|^2) dx \\ &w = (u,v) \\ &E_{smooth}(u,v) = \int_{i,j} (|\nabla_3 u|^2 + |\nabla_3 v|^2) dx \end{aligned} \tag{22}$$

α here represents the regularization term. Usually the value is high when smooth surfaces without textures are involved. The variable γ determines how important that the gradient values is. The spatio-temporal gradient can be expressed as:

$$\nabla_3 := (\partial x, \partial y, \partial t)^T \tag{23}$$

These equations demonstrate the basic ideas for the computation of optical flow.

Figure 4.10a shows two similar images from two adjacent frames of a video (Brox et al., 2004). The difference can be quantified by optical flow, by determining the velocity of the each pixel. The movements of the different objects in the image are not the same. For example, the tree has moved much more than the house since the tree is closer to the camera. Figure 4.10b shows the computed movement of several pixels of the trees.

For this project, theoretically, the displacements of all pixels should be the same, because all the objects are in the same plane. Therefore, to average the movement of all pixels would be a proper solution for simplification. The process of optical flow application in local

(a) (b)

Figure 4.10. (a) A pair of adjacent images, (b) computed velocities of every pixel.

alignment has five steps (figure 4.11). First one is to read two images that need to be adjusted. Usually they are the two images from two adjacent presets. The second step is to compute the corresponding top-view images. This will transform the two images in same plane. Then the overlapped parts of the two images will be both cropped for optical flow computation. Although the sizes of the remote-sensing images were very large, the overlapped part will be only a small portion. This guaranteed the acceptable time consumption of optical flow computation. After step four, all displacements of the pixels between the two overlapped images were computed. This displacement is therefore the quantified mismatches we need for adjusting. Final step is to adjust image locations based on this computation.

The following figures use some images from this tower remote sensing system to demonstrate the whole process of local alignment. The white marker in the image was used as a reference for determining the error. Figure 4.12 shows the original images for local alignment. After top-view computation, the overlapped parts were cropped, shown in figure 4.13. Figure 4.14 shows the computed optical flow of all pixels in cropped images. A comparison is displayed in figure 4.15. Before local alignment, the mismatches are obvious (figure 4.15a), while after local alignment, most of the mismatches disappear.

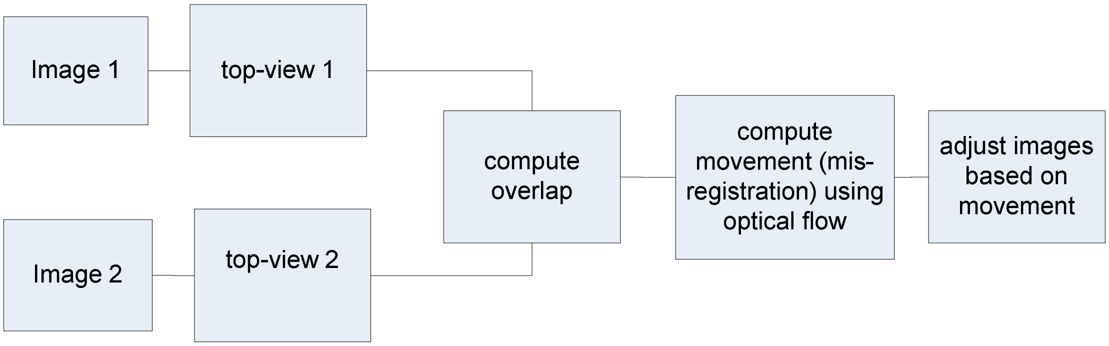


Figure 4.11. Flow chart of local alignment algorithm.

Figure 4.12. A pair of images from two adjacent positions.

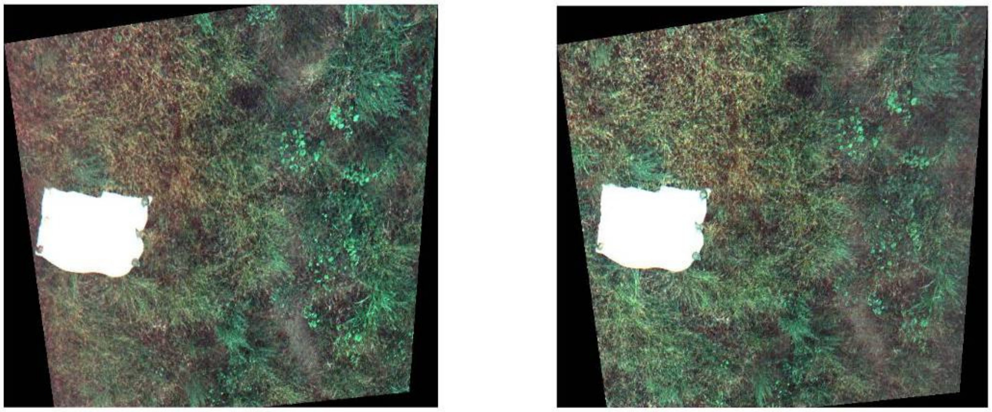


Figure 4.13. Overlap parts of two top-view images.

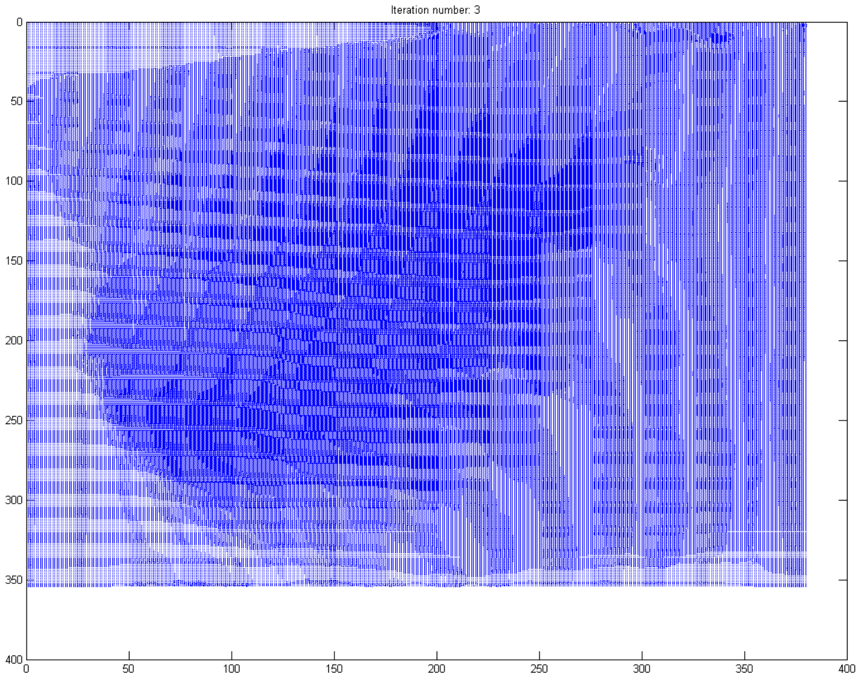


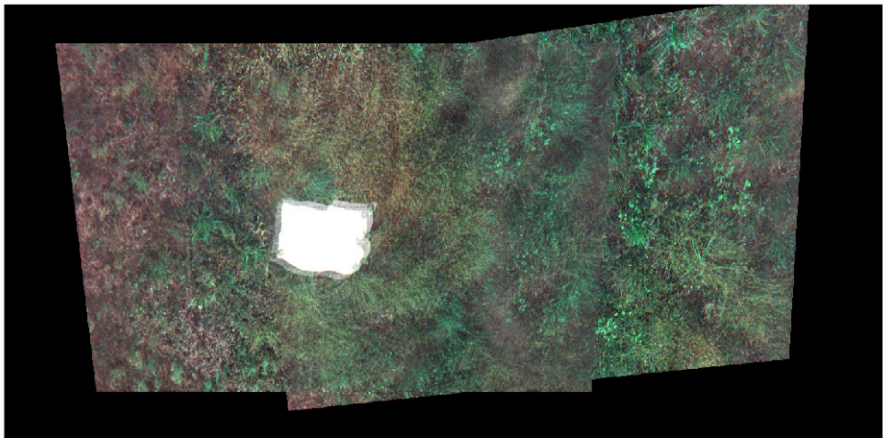
Figure 4.14. Computed optical flow of the two images in Figure 4.13.

4.2 Experiment of Image Mosaic

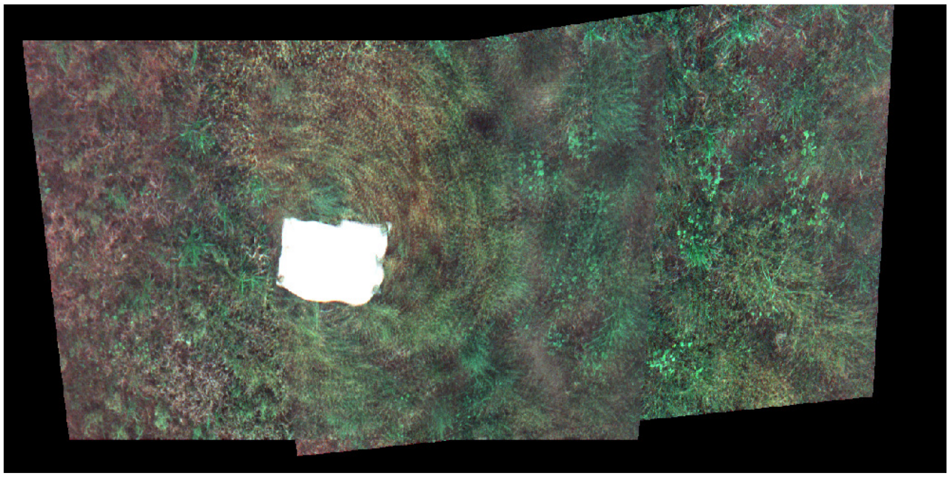
The experiment of the image mosaic was conducted in the Energy Farm near the University of Illinois at Urbana-Champaign. After the image acquisition, all images were processed in the laboratory. The performance of the image mosaic algorithm was tested and evaluated based on the experimental results.

4.2.1 Image mosaic and spatial resolution

The simplest way to estimate the spatial resolution is: after obtaining the mosaic of the entire field, use the dimension of the field and the total number of pixels to estimate the final spatial resolution. The switch grass field, for example, has the size of 190 X 190 meters. After images mosaic, 15 images were combined, obtaining a map of resolution of 1929X1613. Therefore, the final spatial resolution is about 0.12 m.



(a)



(b)

Figure 4.15. (a) Mosaiced image without local alignment, (b) mosaiced image after local alignment.

The following figure 4.16 shows the size of the entire field and resolution of the map. Due to the error in image mosaic, the spatial resolutions for each field may not be exact the same. Therefore, the final spatial resolution of the map was evaluated in the scale of the entire field. The resolution of the map after mosaic is 0.12 m.

The images were taken every day. There are 71 images in total. 19 are for miscanthus; 15 for switch grass; 18 for prairie and 19 for corn. The image acquisition process usually lasts about 40 minutes. While taking images, the date, time and compass reading will be automatically

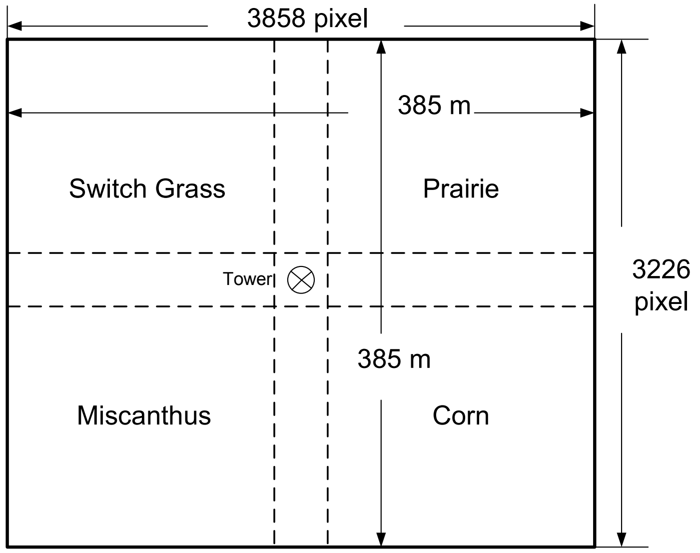


Figure 4.16. The size of the field and resolution of the map.

stored. After acquisition, the captured images were processed based on the recorded compass information. For decreasing the bias error, the compass reading was slightly adjusted for improving performance.

The figure 4.17 is the map of the entire field. In this map, 71 images were jointed together. The image quality was limited by the image acquisition algorithm since currently there is no white balance algorithm. Therefore the various illuminations cause the differences among images.

For evaluating the accuracy, some markers were placed in the switch grass. Figure 4.18 is a mosaic of NIR image of the switch grass field. 15 images were combined to provide an image of the entire field. Table 4.1 contains the angles of the images associated. These angles are the raw compass reading which were automatically stored while image acquisition process. The image of NO.1 is at the bottom left of the mosaic image. The images were taken row by row with subsequent numbers.

4.2.2 Evaluation of system performance

Because of the lack of obvious features in the field, it is usually difficult to determine mismatches in the mosaic images. Therefore, some markers were placed in the field to be used as reference for evaluation. These markers are some sticks with bright color which makes them easy to identify. Figure 4.19 shows an example of such marker. After mosaic, the mismatches can be determined after locating the markers in the image. Moreover, there are still some identical mismatches in the image, such as the edges and grooves in the field. These features were also used to determine the accuracy. Figure 4.20 are several examples showing how mismatches determined.

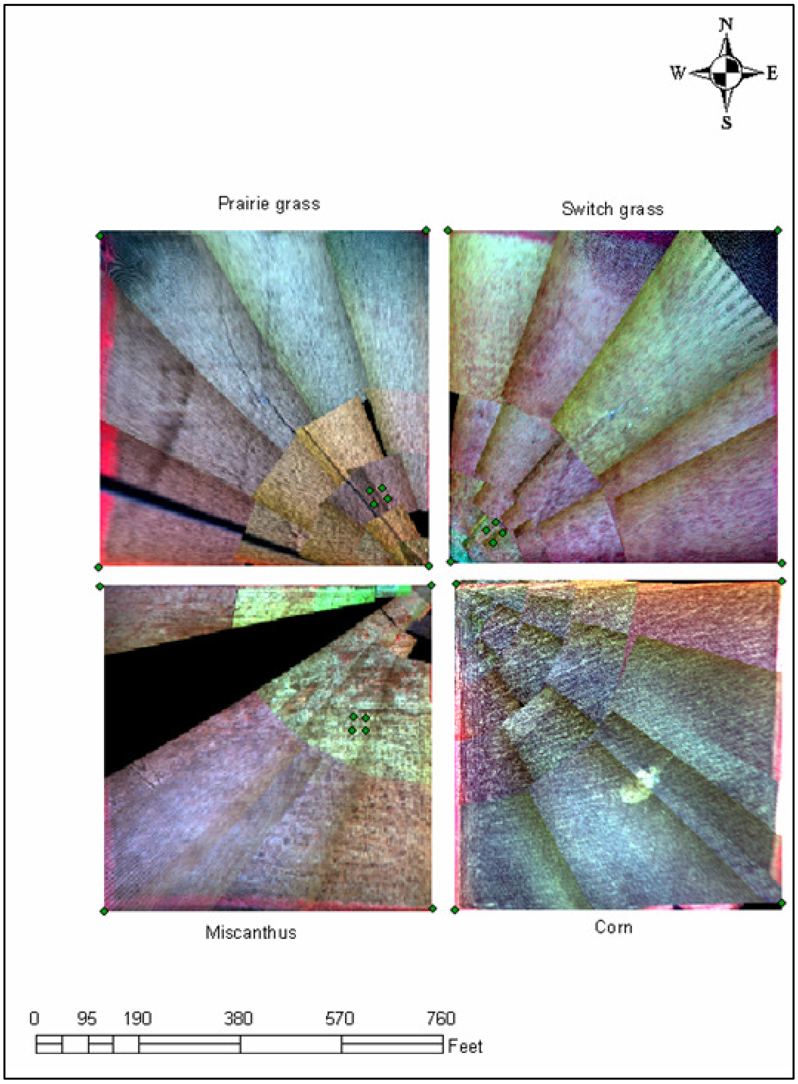


Figure 4.17. Image mosaic of the entire field.

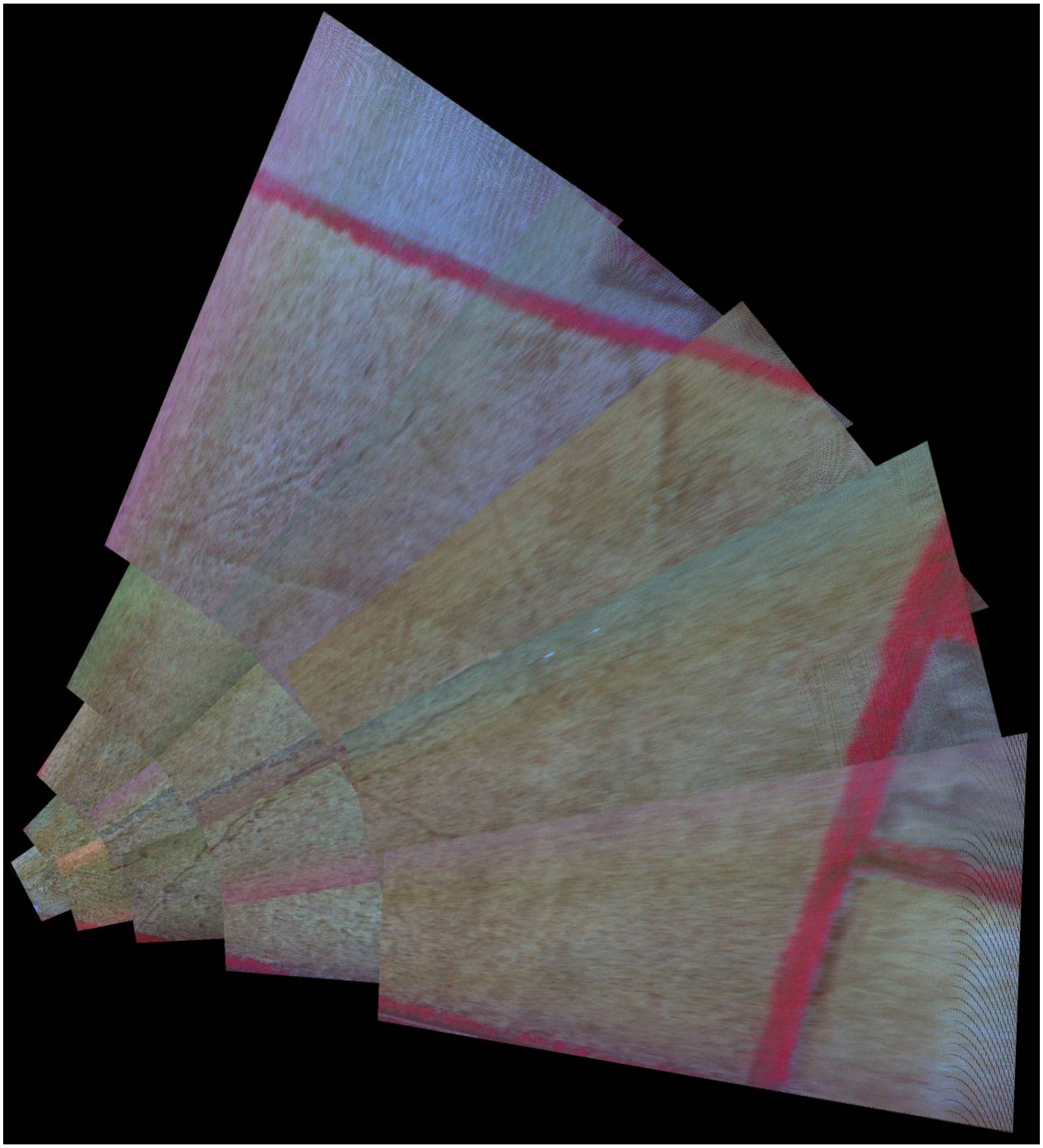


Figure 4.18. The Image mosaic of Switch Grass Field in which fifteen images were combined.

For evaluating the error in mosaic, five maps, in which 75 images were combined, were chosen. In these five maps, totally 39 featured were particularly picked for determining the error, including the edges of the field and the markers. Table 4.2 lists all the features and their errors. The errors were categorized into transverse error and longitudinal error along different directions. The errors were manually determined by reading the images in pixel level. Finally, the average error along transverse direction is 14.5 pixels and error long longitudinal direction error is 15.2 pixels.

Table 4.1. Angles of images that were combined.

	yaw	**pitch**	**roll**
1	63.54	57	0
2	54.3	47.89	0
3	73.57	47.19	0
4	44.44	37.31	0
5	66.79	36.45	0
6	81.14	36.78	0
7	87.21	25.56	0.13
8	76.12	25.98	0.35
9	57.73	25.96	0
10	40.39	26.22	0.18
11	35.82	16.65	0
12	53.86	16.46	0.26
13	63.98	16.35	0.09
14	75.94	16.13	0.35
15	91.7	15.95	0.44

4.2.3 Possible sources of errors

The error of this algorithm can be categorized as: the compass error and the focal length error. Beside these two main factors, there are also minor factors such as wind and variation in illumination which can result in small amount of error.

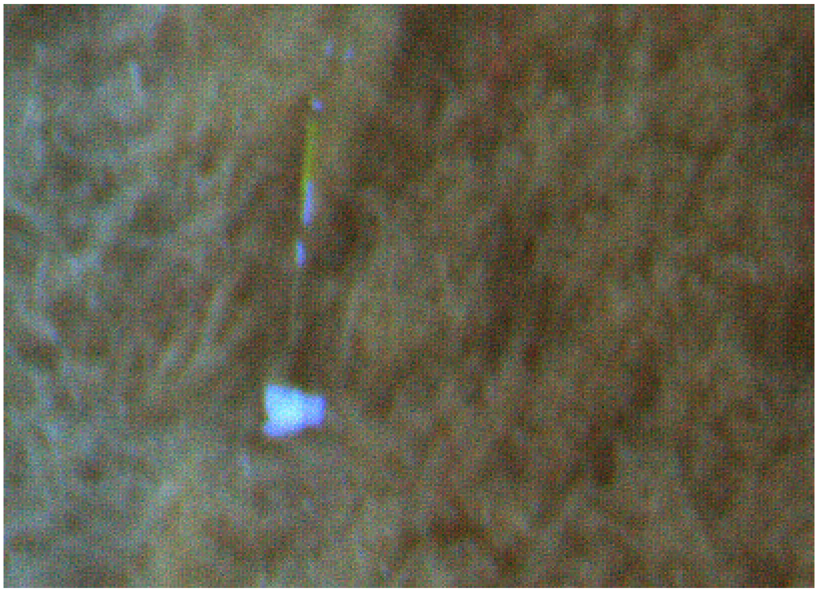
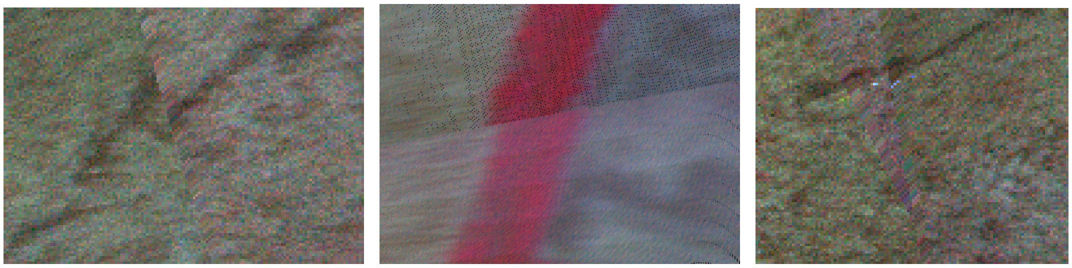


Figure 4.19. The marker used in the field. The bright color makes it easy to identify.



(a) (b) (c)

Figure 4.20. Some examples of mismatches in the mosaic images, (a) image shows the groove in the field used to determine the error, (b) image shows the mismatch at edges of field, (c) image shows the marker.

(1) Compass:

The compass error was discussed in Chapter 3. The reason for this error is due to imperfect installation and the disturbance of the environment. According to the distance and pitch angle, 1 degree error in compass reading can result in more than 100 pixel errors in mosaic. It depends on the distance form ground to camera. Therefore, the compass calibration is very important in this project to provide relatively accurate result.

Table 4.2. Pixel error of image mosaic.

No.	transverse error (pixel)	longitudinal error (pixel)	No.	transverse error (pixel)	longitudinal error (pixel)
1	1	1	21	25	31
2	4	2	22	1	0
3	25	12	23	16	2
4	42	6	24	6	32
5	3	11	25	27	42
6	11	17	26	1	2
7	9	30	27	14	9
8	7	30	28	20	4
9	0	0	29	2	12
10	23	14	30	11	1
11	20	3	31	30	45
12	3	7	32	13	40
13	14	2	33	1	0
14	29	38	34	15	16
15	22	40	35	30	7
16	40	8	36	3	10
17	2	6	37	18	1
18	13	12	38	27	58
19	27	3	39	7	29
20	4	10	Avg.	14.5	15.2

The figure 4.21 shows the error distribution when compass reading has a 0.3^o error. The contour appears circles around the tower, which shows that the points on the same circle will have same level of error of mosaic. The reason is these points all have the same distance to the camera and same pitch angles. Thus compass is a main source of mosaic error.

(2) Focal length error:

The focal length used in the top-view algorithm is based on the feedback from the lens controller. However, the feedback has been proven not exactly reliable. The mechanism of this feedback is based on the position of the gears that is not accurate. Moreover, the relationship between the focal length and focal distance is not a strong linearity. Therefore, there might be differences between the estimated focal distance and the true value.

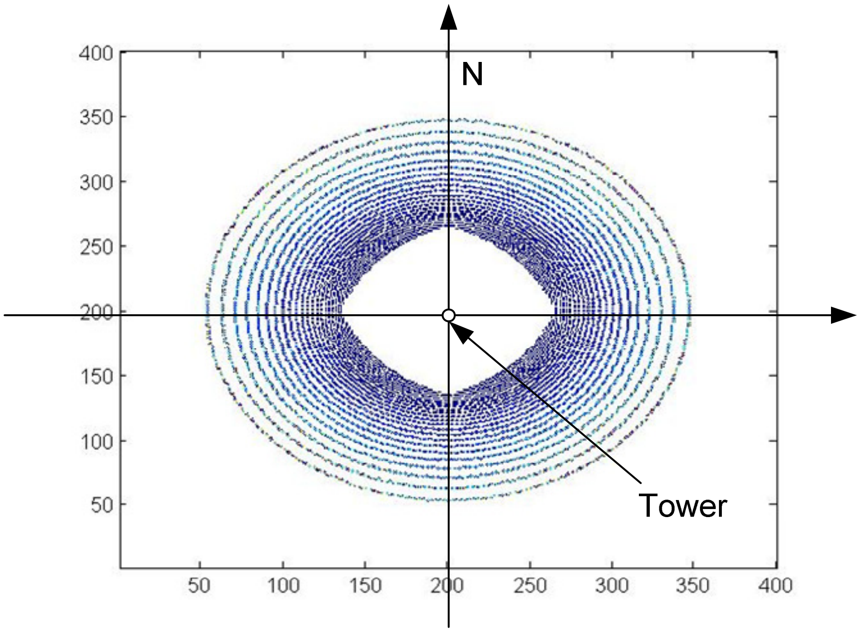


Figure 4.21. The mosaic error distribution while the compass error is 0.3°.

4.3 Conclusions and Future Work

In this chapter, the image mosaic for the tower remote sensing system was introduced in detail. There are mainly three steps of the mosaic algorithm. After the image acquisition process, the images were stored with corresponding orientation information. These information, including pitch, yaw and roll angles, were used for transforming the image into top-view format. Then the transformed images were combined based on their geographical location information. The final step is using optical flow to minimize the mismatches.

In total 5 maps, 75 images were chosen for evaluating the accuracy of the algorithm. The result shows that the algorithm can provide relatively accurate image mosaic. The pixel error was limited within 14.5 transversely and 15.2 longitudinally.

Future work may focus on minimizing the error in pixel value that is crucial for determining the spectral properties of the crops. The current image mosaic is lack of adjustment to various illuminations. Thus the brightness of adjacent images has slight difference. Moreover, the gain and exposure time of the three channels need to be adjusted to make histogram consistent. The difficulty is that due to the requirement of flexibility, the camera change the orientation, so developed method such as white balance panel was not possible. Therefore, a new method of white balance needs to be developed.

CHAPTER 5 IMAGE DATA INTERPRETATION FOR MONOCULAR VISION SYSTEM OF FIELD ROBOT

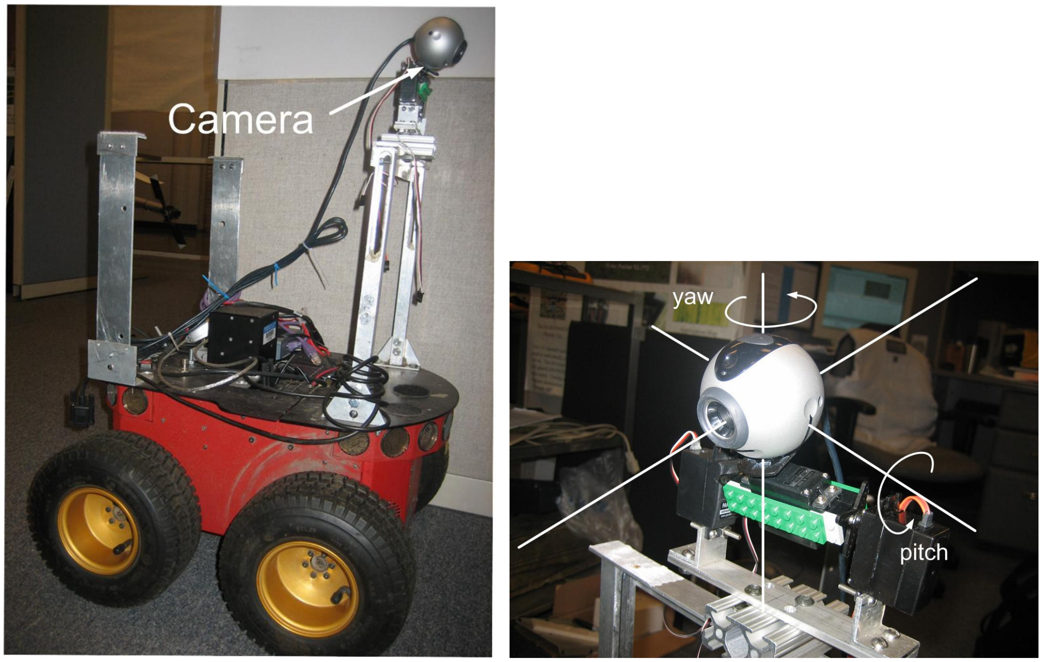
The method of coordinate transformation between camera and ground frames can be used in measurement for any scale. The same algorithm can also be applied in field robot systems. The motivation of this approach is to replace traditional binocular stereo-vision system with a single camera. This monocular vision system can be used for plants measurement, including both 2D position and height information. Similar to developed geo-reference system for tower remote-sensing system, a camera installed on the robot can provide 2D coordinates in ground frame of the plants that are in front of the robot. Further, with the movement of the robot detected, a single camera can function as a stereo vision system. Given the displacement between two sequent images, the height information of the plants can be obtained.

In this chapter, the development and performance of this system were presented in four sections. At first, the devices used in this project were introduced, including the camera, robot platform and necessary sensors. Further, the algorithm of measurement system was explained. First is the 2D position algorithm, and then followed by algorithm of estimation of plants' heights using robot speed measurement. Then two types of tests, the laboratory test and field test were introduced to validate the method with experimental data.

5.1 Devices Used in Vision System

ActiveMedia Pioneer 3 All Terrain (P3-AT) functions as the base robotic platform in this project. The platform is designed to be easily customized and equipped for outdoor and rough-terrain projects. The P3-AT has an aluminum body, with the dimension of 50 cm X 49 cm X 26 cm. The wheels are 21.5 cm in diameter. The base platform, including the wheels, minimal frame and motors, has weight of 13.03 kilograms.

The P3-AT utilizes four Pittman GM9236E349-R1 12 volt motors with 38.3 to 1 gear ratios. The platform implements differential steering, which allow rotation in the same place within a 40 cm radius circle. Two motors drive each side of the robotic platform, by means of a belt, to distribute power to wheels. A 100-tick encoder on each motor provides feedback information to the on-board microcontroller. The platform is capable of speeds up to 700 mm per second. The platform is also capable of carrying additional 30 kilograms of weight (ActiveMedia Robotics, 2003).



(a) (b)

Figure 5.1. Devices used in this research, (a) the robot platform, (b) the camera of the vision system.

The camera used in the vision system is a Logitech Webcam, a RGB camera with the resolution of 640X480. The RGB camera is only used to verify the basic concepts which mean it can be easily replaced with other professional models. The camera is installed on the brackets with 2 degrees of freedom of rotation. The camera platform, which allows the camera to pitch and yaw, is driven by two electrical motors. The camera on the robot functions similarly as the one in tower remote sensing system which can measure plants through coordinate transformation.

A Panasonic ToughBook® was attached on the robot as the on-board computer for data acquisition and processing. The control program of the robot was developed in C++ and the image acquisition and processing program was developed by Matlab. The robot can be controlled both by joystick and keyboard through serial port. The communication between the camera and computer is through USB port.

5.2 Methods and Algorithm

The algorithm of the monocular vision system is similar as the geo-referencing. The coordinate transformation can estimate the coordinates of the plant in the ground frame. Further, with the speed of the robot measured, the height of the plant can be obtained.

5.2.1 Definition of coordinate systems

Similar to the tower remote sensing system, the measuring method of this vision system is based on coordinate transformation. The coordinates of the plants in the image frame will be transformed to ground frame to determine the locations. There are three frames defined in this system. They are image frame, camera reference frame and ground frame. The definitions of image frame and camera reference frame are same as tower remote sensing system. For clear demonstration, figure 5.2 shows the relationship between ground coordinate system and the robot platform. The origin of the ground coordinate system is the projection of camera on the ground. The moving direction is along Y axis. Figure 5.3 shows the definition of two coordinate systems for the same object.

5.2.2 The 2D coordinates

The position of the plant, which means the coordinates of objects in ground frame, is directly determined by coordinate transformation. The process is described as following:

1 Obtain the coordinates of objects in image frame by image processing.

2 Transform this coordinates into camera reference frame by intrinsic parameters.

3 Further transform the coordinates into ground frame using extrinsic parameters.

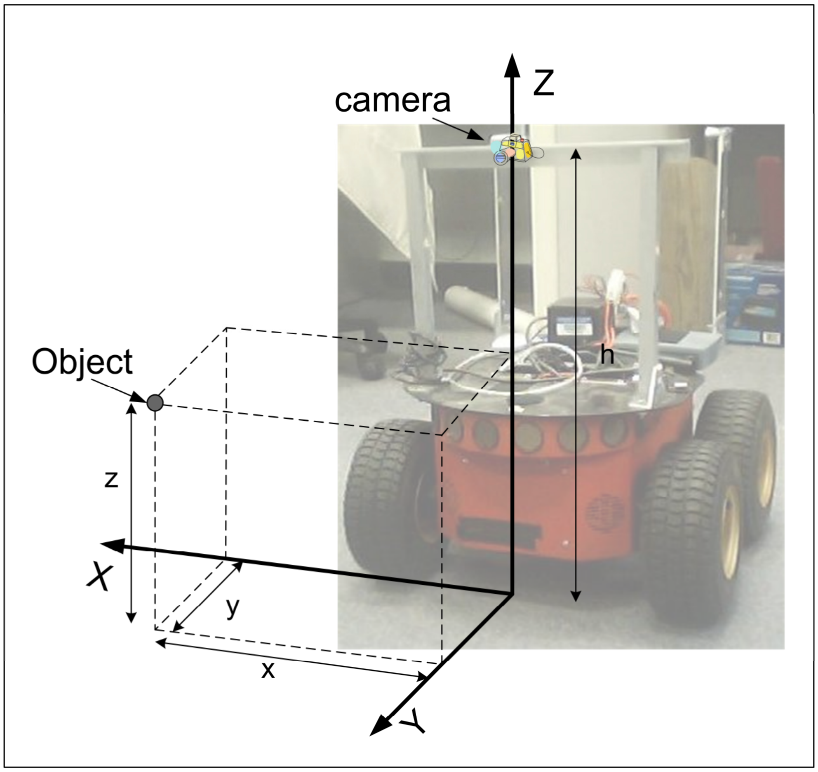


Figure 5.2. Definition of ground frame.

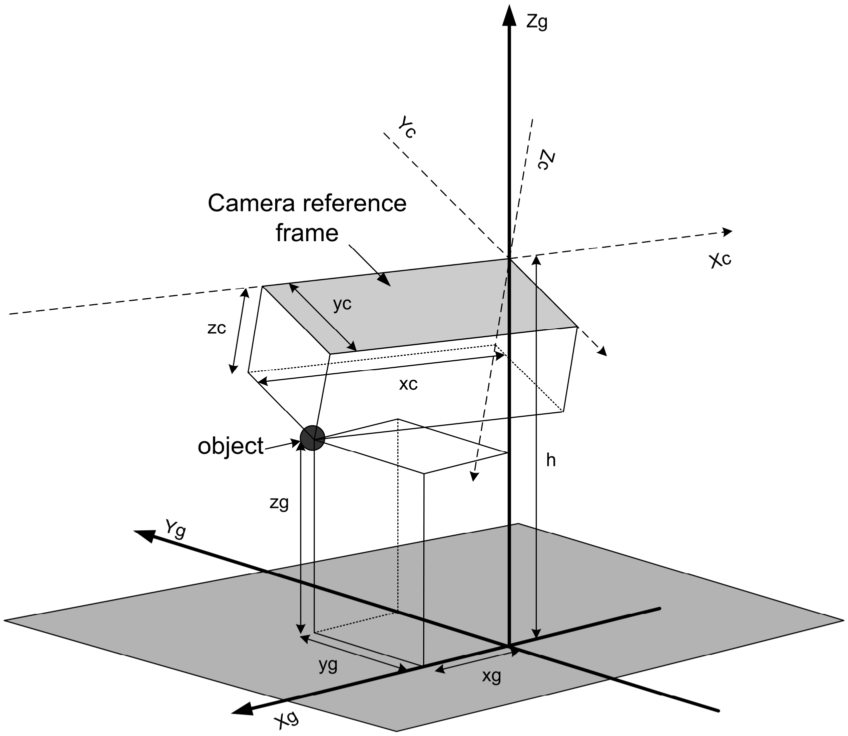


Figure 5.3. Definition of coordinate systems.

The x and y coordinates of the object are determined through former transformation after the camera has been calibrated. Figure 5.4 shows the area that the camera can cover. The plants inside this area can be measured.

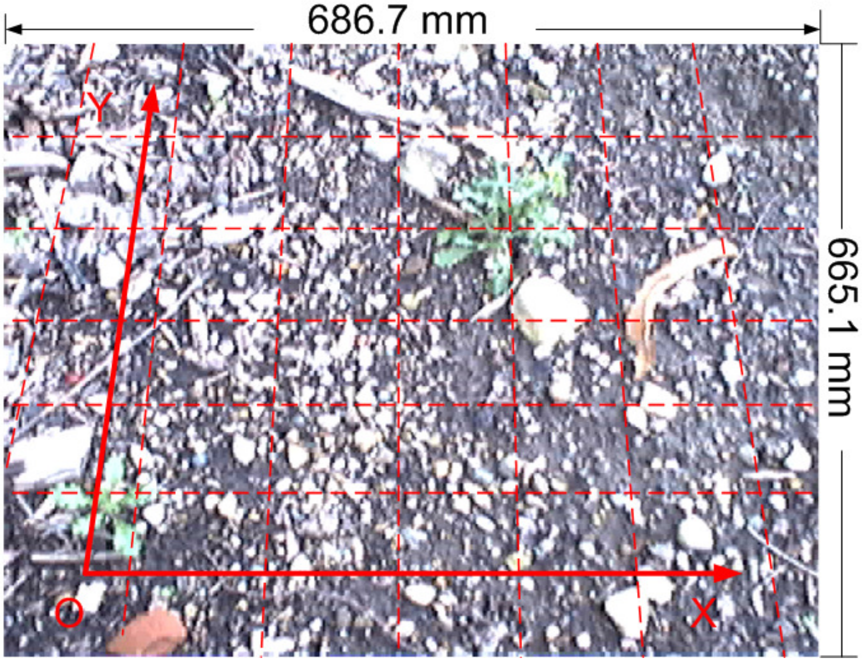


Figure 5.4. The image from the vision system which shows the area that the camera covers.

5.2.3 Encoder and compass

The method of height estimation in this project is to use two sequence frames to substitute the traditional stereo-vision system. Thus, all the physical relationships of the two frames, including difference in position and angle, need to be measured. Since the robot keeps moving while taking pictures, the displacement between two frames can be computed based on velocity measurement.

While the robot is moving in the field, not only the position, but the orientation changes, even executing a “straight-line-moving” program. The reason is the uneven field surface. This changing results in errors in coordinate transformation matrix. Therefore, the orientation variation was recorded as a reference to estimate the system error.

In this monocular stereo-vision system, the positions and orientations of camera in both frames are measured. The position difference, which is the displacement of robot between two images, is measured by encoder on the robot platform. The turning angle between two frames is detected by comparing speed difference of wheels from different sides.

To measure the height of the plant, the coordinates of the plant in both images are needed. However, this coordinate can not be obtained directly. But the difference of two coordinates can be deduced based on the measured displacement between two images. With this condition, a set of equations was established. After solving this equation, the height of the plant was calculated. The relationship is shown in details in figure 5.5

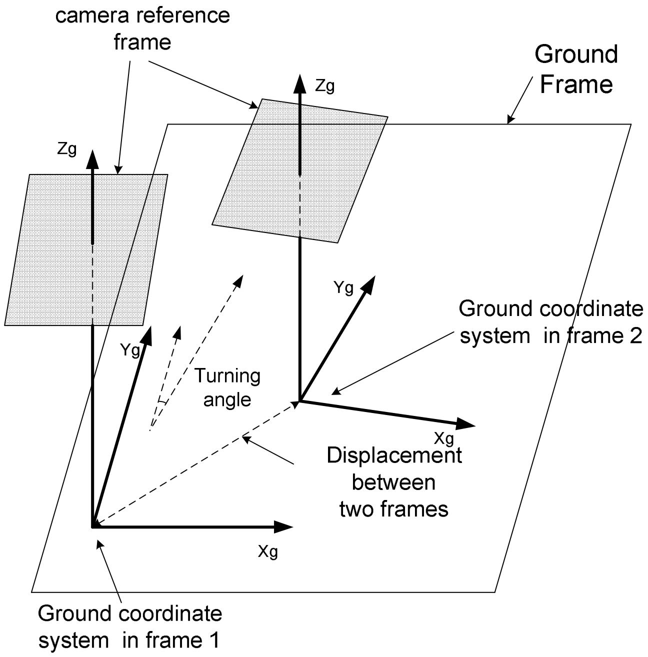


Figure 5.5. Definition of distance between two images.

5.2.4 Height estimation based on two-point coordinate transformation

The height of the object can be calculated by solving three coordinate transformation equations. There are two coordinate systems to define the position of the object. One is ground reference frame, and the other is pixel frame. The transformation matrix, H, related the ground coordinates with pixel coordinates of objects.

$$H \cdot X_g = X_p \tag{24}$$

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \cdot \begin{bmatrix} x_g \\ y_g \\ z_g \\ 1 \end{bmatrix} = s \cdot \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} \tag{25}$$

For a single image, the height information cannot be extracted since *Xg* is a vector with three degree of freedom while *Xp* has two degree of freedom. However, if one extra image is used, the height information can then be extracted. Given the movement of object between two images, the positions of object in ground reference frame are known. Then one extra equation is obtained, thus height along z axis in ground reference frame can be calculated by solving the three geo-reference equations.

$$\begin{matrix} H \cdot X_{g1} = X_{p1} \\ H \cdot X_{g2} = X_{p2} \end{matrix} \Rightarrow \begin{matrix} \dfrac{h_1 \cdot X_{g1}}{h_3 \cdot X_{g1}} = x_{p1}; \\ \dfrac{h_2 \cdot X_{g1}}{h_3 \cdot X_{g1}} = y_{p1}; \\ \dfrac{h_2 \cdot X_{g2}}{h_3 \cdot X_{g2}} = y_{p2}; \end{matrix} \tag{26}$$

Where h_i is the i th row of H matrix. X_{g1} and X_{g2} are the ground coordinates of objects in two images, x_{p1}, y_{p1}, and y_{p2} are the x and y value of objects in pixel coordinates of two images, respectively.

5.2.5 Plant identification

The plant identification algorithm is based on the connected component method. Because the objects that need to be identified in the project are plants, of which the most significant features are their green color. They can be identified based on the ratio of green channel pixel value to the total pixel value of three channels (figure 5.6). The pixels with ratio lager than the threshold are considered as plant, marked as 1 in binary image, while the others are background, marked as 0 (figure 5.7a). Threshold is defined based on the properties of the identifying plant and noise in background. From the binary image, the location of the plant can be easily obtained.

After the plant identified, the next step is the calculation of shift of plants in pixel level of two images. The definition of this shift is the distance between the two centroids of plants (figure 5.8). To determine the centroid of a plant, the image needs to be cropped first (figure 5.7b). After all connected component in the image identified, the chosen object was cropped and then displayed as a smaller image (figure 5.7b). Then the centroid can be determined.

5.2.6 Image crop and comparison

Matching the objects in two images is an essential procedure to ensure the accuracy in the determination of the shift. There are two factors, size and shape, were taken into account to match the objects. After all the connected components identified, all of them were labeled. The objects displayed along x axis in pixel coordinate were numbered as 1, 2, etc. respectively. The objects were cropped from the original images for further comparison. The cropped area was determined by the minimum and maximum pixel coordinates of selected object.

After the same objects in two images identified, both size and shape information were used to check the consistency. For the size comparison, the total numbers of pixels of the object in each image were compared. Only when the difference of total number of pixels is under a certain threshold, the objects in two images are considered the same plant. The threshold is determined by the lighting condition, since area of identified plant is effect by brightness of the image.

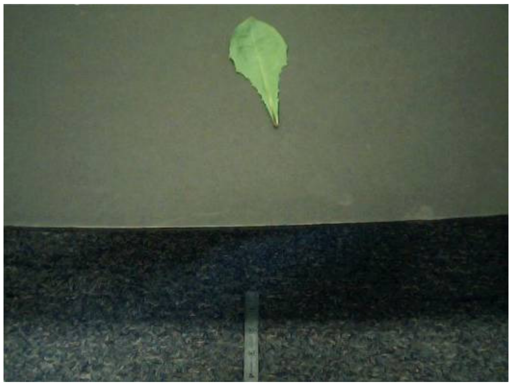


Figure 5.6. The original image, prepared for detection.

(a) (b)

Figure 5.7. (a) Binary image after plant identification, (b) cropped image.

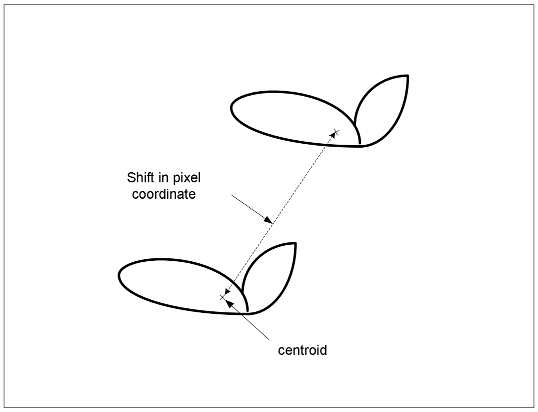
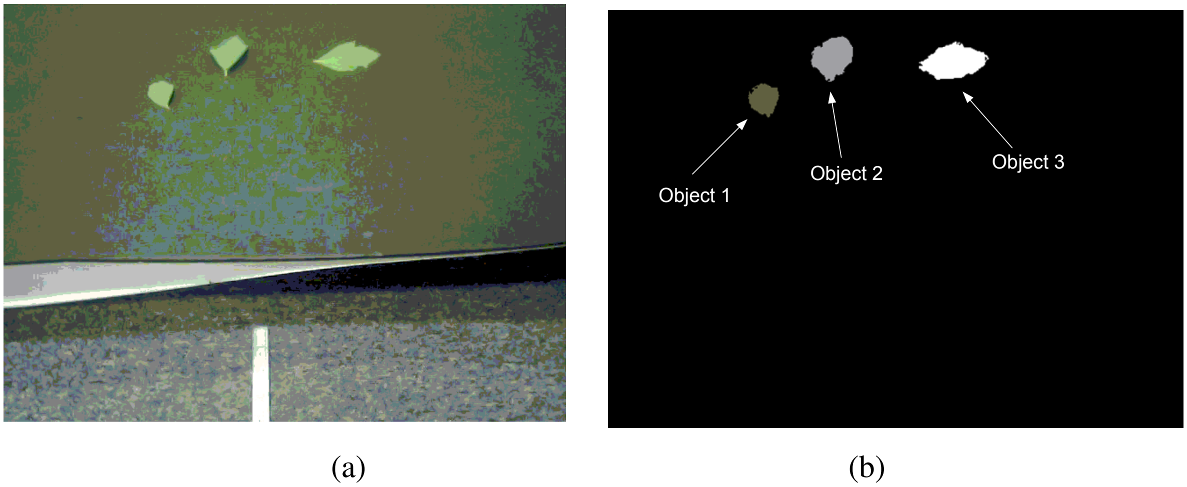


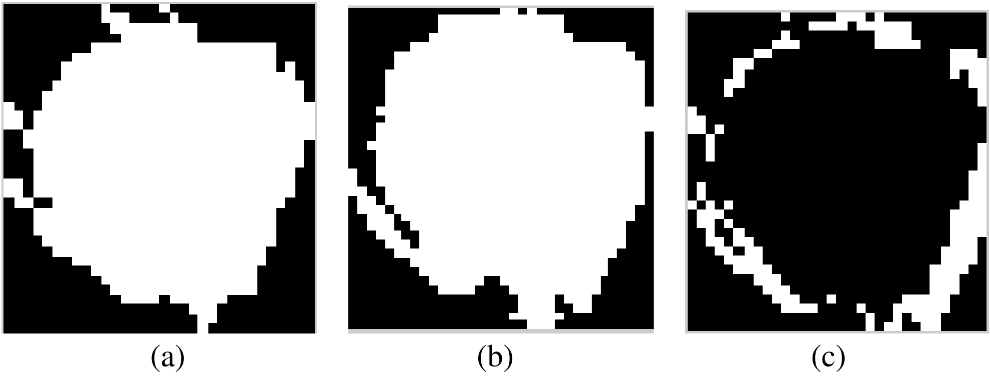
Figure 5.8. Displacement of objects in pixel coordinate.



(a) (b)

Figure 5.9. (a) Original image before image identification, (b) processed image.

The morphology of the plant is considered as the distribution of all ‘1’ pixel values in corresponding binary image. If the two objects have the similar sizes, the similarity can be determined by subtracting one image by another. The absolute value of the sum of the residues is used to evaluate the similarity. In this experiment, only when this value is within 40%, the objects are considered to have identical morphology.



(a) (b) (c)

Figure 5.10. The comparison of plants in two images, (a) the object in image 1, (b) the object in image 2, (c) the difference of the two objects.

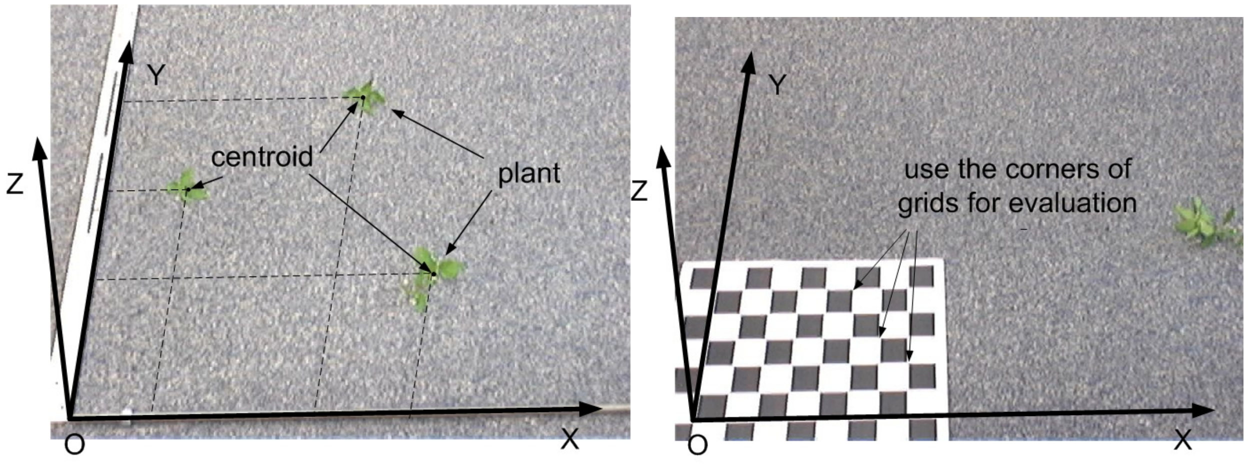
5.3 Experiments and System Performance Evaluation

Two types of tests were conducted in this research, a laboratory test and a field test. Laboratory test was for evaluating the basic methods and determining the accuracy in an ideal environment. Since the environment was the lab was usually simple (flat floor and consistent illumination), the results were expected to be more accurate than the field test. The experiments focused on testing the accuracy of algorithm for 2D positioning and height estimation. The field test was conducted after the laboratory test. The purpose of field test is to evaluate the performance of the system in real environments. The complexity of the outdoor environment may lower the accuracy. Finally, the results of both tests were compared for analyzing the sources of errors.

5.3.1 2-D positioning

Several fresh plants were placed on the lab floor for testing. The origin of ground coordinate system is randomly chosen near left bottom part of the image. This point can be used as the origin of an applicator (for example, mechanical arm) for future research (figure 5.11a). Two tubes were placed along x and y axes of the ground coordinate system as the reference for manually measuring the position of each individual plant. Every plant was measured by tape and the data were recorded as ‘measured’ values. Then, after the image was taken, the program estimated the positions of plants. Every centroid of each plant was identified individually, and then the coordinates were computed through coordinate transformation. These data were recorded as ‘estimated’ values.

The area that the camera covers has a size of 686.7 mm by 665.1 mm. Thus the coordinates of plants that can be measured must be within this range. The transverse RMS error along x axis is 2.8 mm while the longitude RMS error along y axis is 5.9 mm, without regression. The major reason for the error is that it is hard to determine the centroids in manual measurement. For testing the weight of this factor, a panel with black and white grids was placed instead of plants (figure 5.11b). The corners of the grids were chosen as objects for measurement. This could avoid the inaccuracy in manually choosing centroids. The accuracy of this test is within 0.5 mm.



(a) (b)

Figure 5.11. Example of 2D positioning, (a) the image captured image, (b) use the corners of grids for evaluating the causation of error.

To compensate for the errors of the vision system, a linear regression between the measured and estimated 2D coordinates was applied. Strong correlation between two coordinates was found (figure 5.13). The adjusted 3D coordinates using the regression decreased the transverse RMS error to 2.3 mm and longitude RMS error to 3.3 mm (table 5.1).

For comparing with the results of laboratory tests, the robot was moved outside for field test. The camera was accordingly calibrated in field environment. Several plants were chosen as the objects for testing (figure 5.12). Similarly, after data acquisition, the results were processed by linear regression (figure 5.13).

The large error in longitudinal direction is due to the height of the plant and the angle between camera and ground. The plants chosen for field test were relatively large and with many overlapping leaves, which brings the inaccuracy in centroids determination. Further, because of the angle of the camera, if the plant is higher than the ground, the estimated value will be larger than the true values. That is because the camera coordinate transformation matrix is no longer appropriate in this situation. Then we need to estimate the height of the plant to compensate for this error.

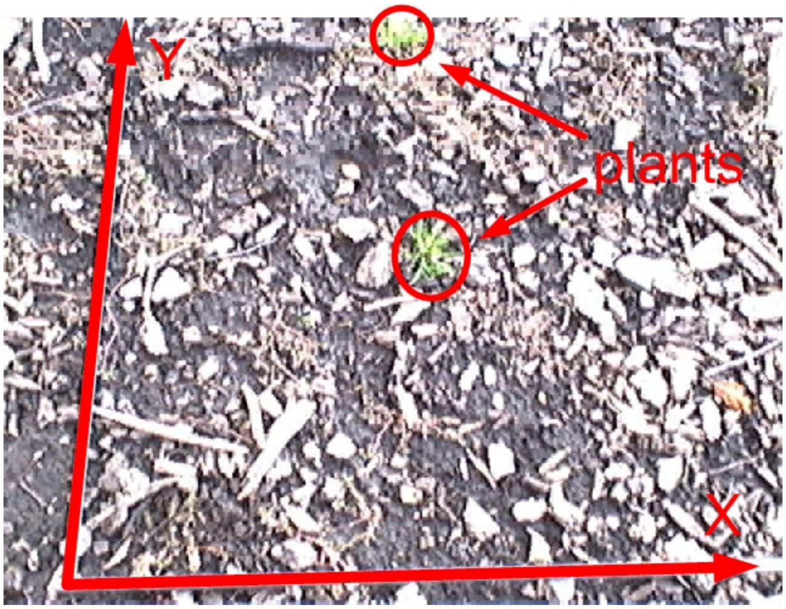
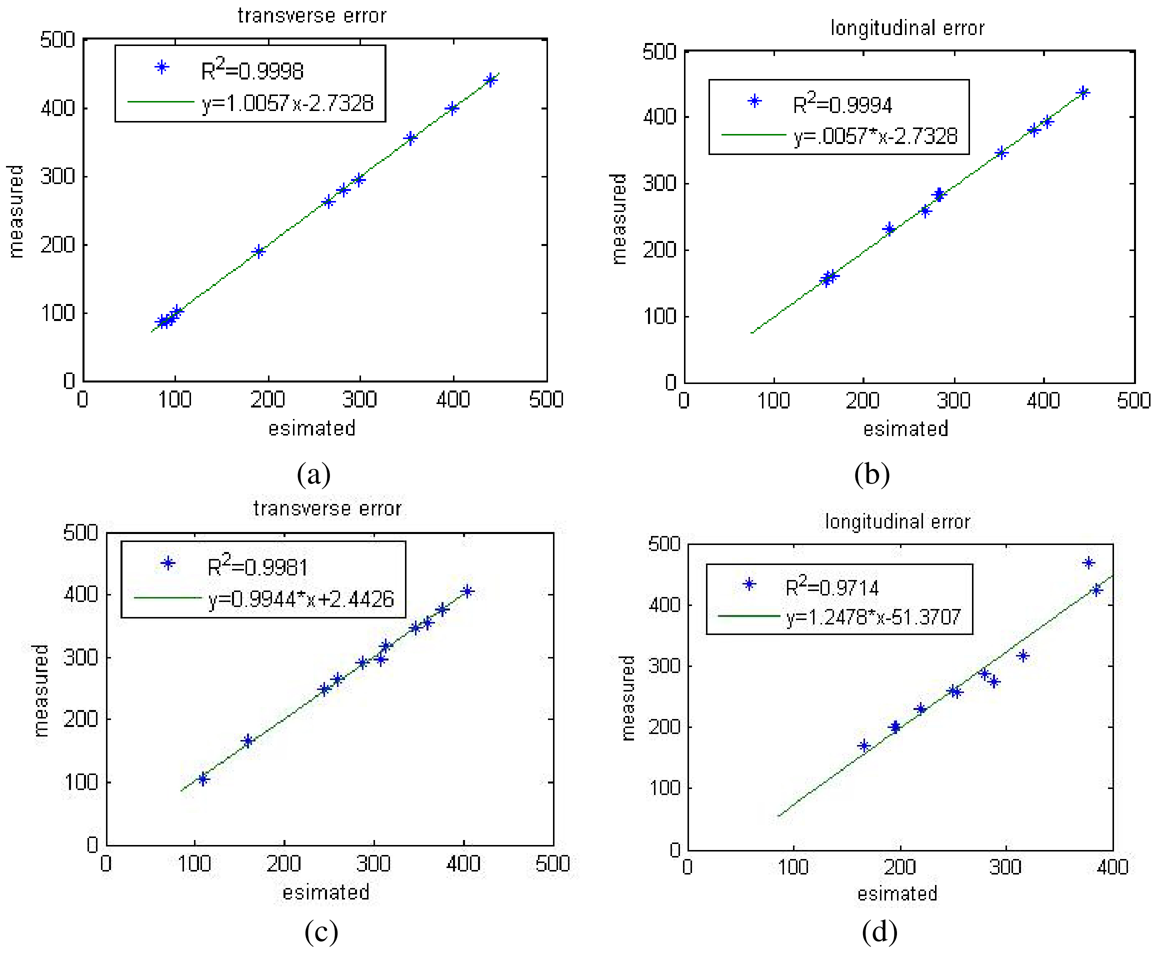


Figure 5.12. Image captured in field test.

Table 5.1 Error of 2D positioning.

Laboratory test				
	RMS error without regression (mm)	**Percentage without regression**	**RMS error with regression (mm)**	**Percentage with regression**
Transverse distance (x, mm)	2.8	0.4%	2.3	0.3%
Longitude distance (y, mm)	5.9	0.9%	3.3	0.5%
Field test				
	RMS error without regression (mm)	**Percentage without regression**	**RMS error with regression (mm)**	**Percentage with regression**
Transverse distance (x, mm)	5.5	0.9%	5.4	0.9%
Longitude distance (y, mm)	30.7	4.4%	21.0	3.0%



(a) (b)

(c) (d)

Figure 5.13. Linear regression of the results from both two tests, (a) the regression of transverse direction for lab test, (b) the regression of longitudinal direction for lab test, (c) the regression of transverse direction for field test, (d) the regression of longitudinal direction for field test.

5.3.2 Height estimation

There are two aspects of the height estimation. The robot data shows the displacement measurement, and the plant identification provides the shift.

(1) Robot data:

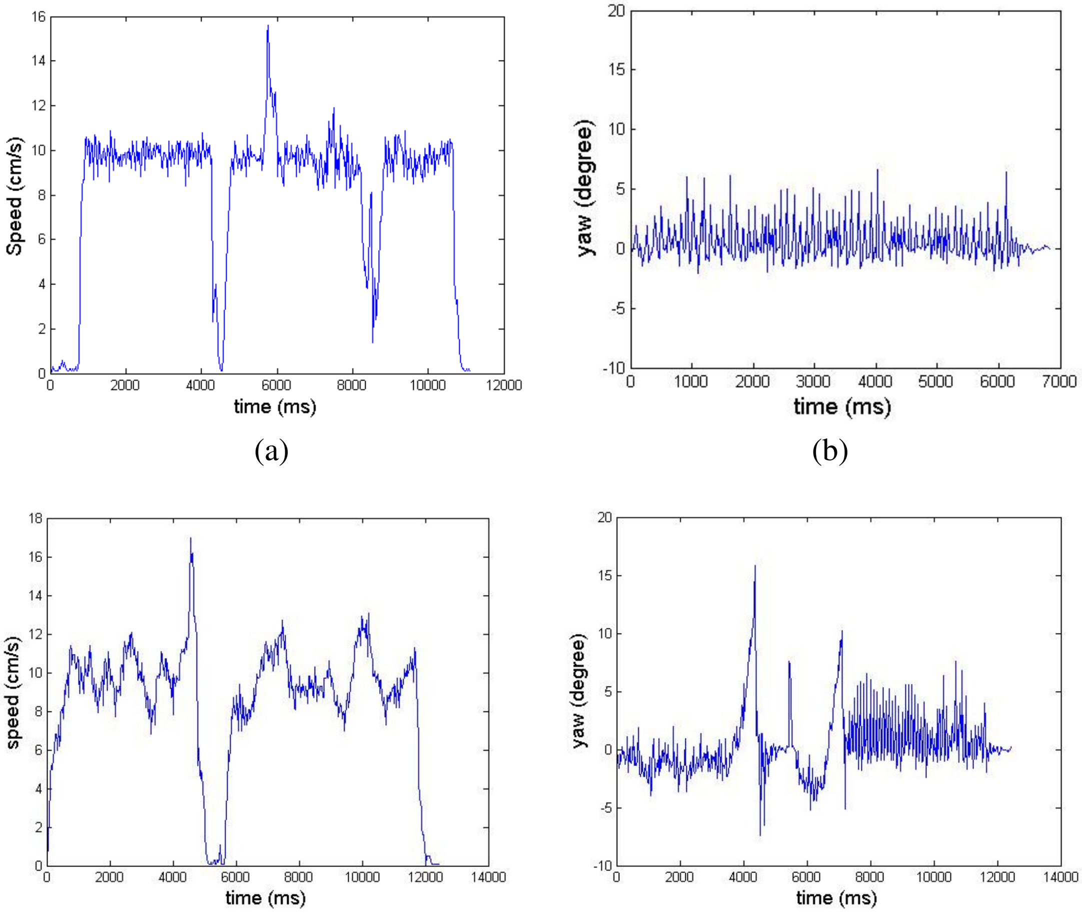
The speed of the robot should be within a certain level in order to make the displacement of robot between two images in a proper range to ensure the existence of plants in two images, considering the image acquisition rate. The maximum speed of the robot was set to 10cm per second. Then the average displacement between two images was about 90 mm. This guaranteed that the plants were taken inside both images. The speed was measured by the encoder very frequently, with the time interval of 25 ms. Figure 5.14 shows an example of recorded robot data. Figure 5.14a shows the speed of the robot measured by the robot. Except when beginning and stopping, during most of the time the speed was relatively even. There were two brief stops, which can be easily identified in the figure. Figure 5.14b shows the recorded yaw angle of the robot while moving. It is measured based on the differential speed of two wheels. The variation is small in the lab environment since the floor is flat with equal wheel-floor friction.

Figure 5.14c and 5.14d are the recorded encoder data in field tests. Comparing to the recorded data in laboratory tests, the variation of robot speed was much larger when the robot is moving on the soil. This was due to the tough surface of the ground. When the wheel left the surface, the reading of the encoder will be much larger than real speed. In contrast, when the wheel move to a groove, the resistance will be much larger, causing low speed reading. Figure 5.14c shows this variation during the period of the robot's movement. The jumps in recorded data are caused by a significant obstacle. Figure 5.14d is the recorded yaw angle of the robot. While the robot was moving, the tough surface of the ground caused several unexpected turning that becomes a major reason for errors.

(2) Plant identification:

For testing the plant identification algorithm in laboratory, several leaves were placed on a wood supporter in front of the robot. The function of the supporter is to provide some heights of the leaves for testing height measurement algorithm. While the robot is moving, the camera can capture the images of plants continuously. Several images were selected for testing. While reading the images, the program can recognize plants and extract them from the background. Then the same plant in two images was identified, thus the distance in pixels were obtained. Figure 5.15a and 5.15b show the two adjacent images chosen form the image sequence. In figure 5.15c and 5.15d are the results after plant identification. After leaves extracted, each individual leave was labeled for determining the same one in both images. The results of labeling were shown in figure 5.15e and 5.15f. The same leaves were marked in same grey level.

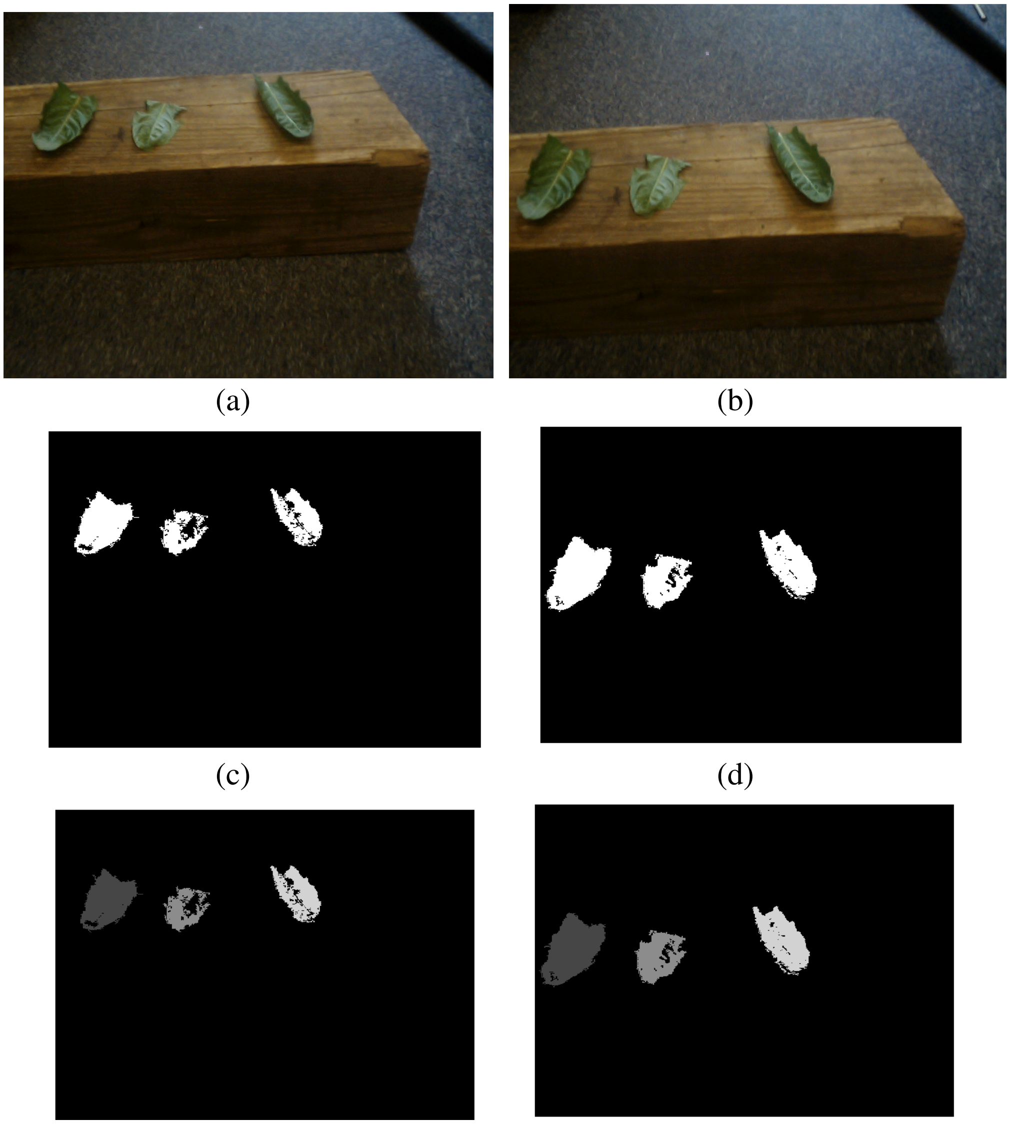
In the field tests, several plants with different heights were chosen (figure 5.16). There were much more noises in the field, therefore the parameters for plant identification were adjusted.



(a) (b)

(c) (d)

Figure 5.14. (a) The recorded speed data of the robot in lab test, (b) the recorded yaw data in lab test, (c) the recorded speed data in field test, (d) the recorded yaw data in field test.



(a) (b)

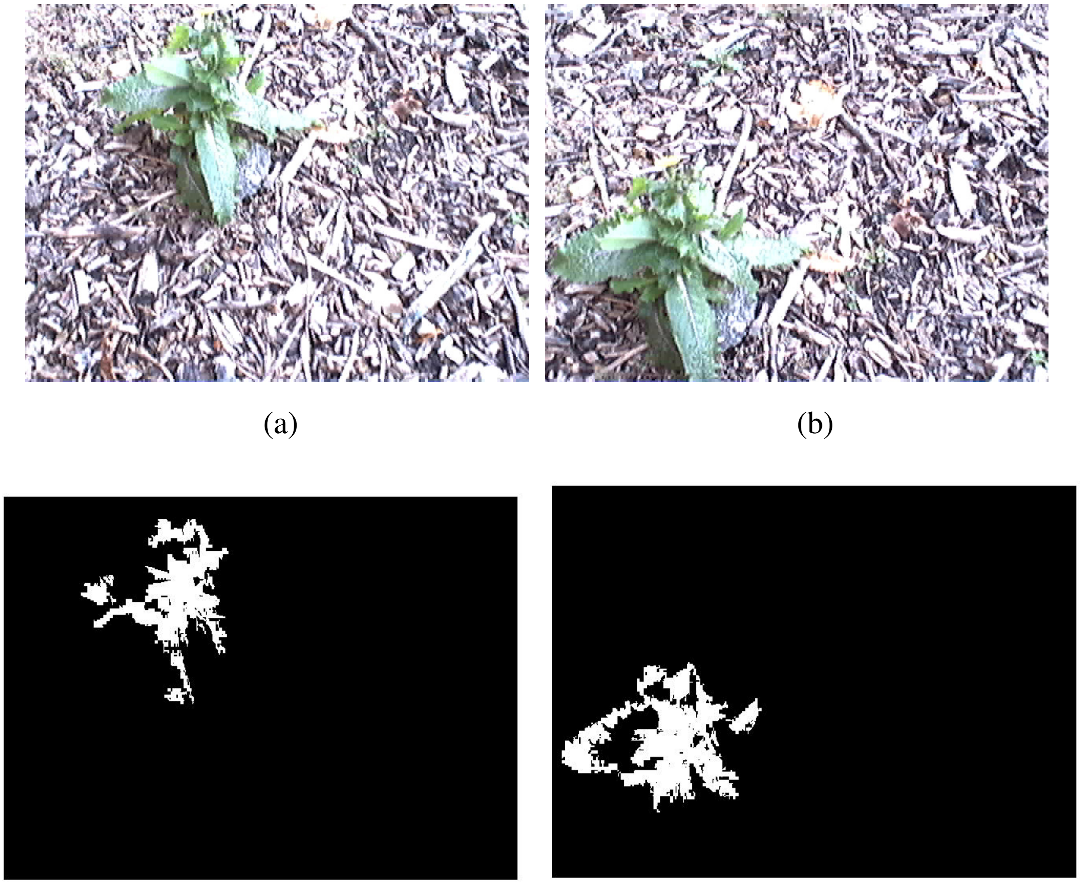
(c) (d)

(e) (f)

Figure 5.15. (a) and (b) A pair of sequent images used to estimate the height of the plants, (c) and (d) the images in which the plants were detected, (e) and (f) the same plant was marked with same grey level.

(3) Height measurement result:

After the plants were labeled, the placement of each leaf was computed in pixel unit. The placement along both axes were computed, recorded as *dx* and d*y*. Combining *dx*, *dy* and the displacement between images that was computed from robot data, the height of the leaf can be computed.



(a) (b)

(c) (d)

Figure 5.16. (a) and (b) Two adjacent images captured in field test, (c) and (d) the plant was identified.

The height estimation in laboratory was relatively accurate since there were fewer disturbances. However, in the field test, the accuracy of height estimation decreased. The heights of ten plants were compared with computational results to evaluate the accuracy. The height of the plant ranges from 85 mm to 361 mm. The figure 5.17 shows the error of height estimation in field test. Because there is no strong linear regression between estimated values and measured values, the RMS error was calculated from the raw data. The reason for lack of linearity is the variety of causations for error. The summary of performance of the system in height estimation is demonstrated in table 5.2.

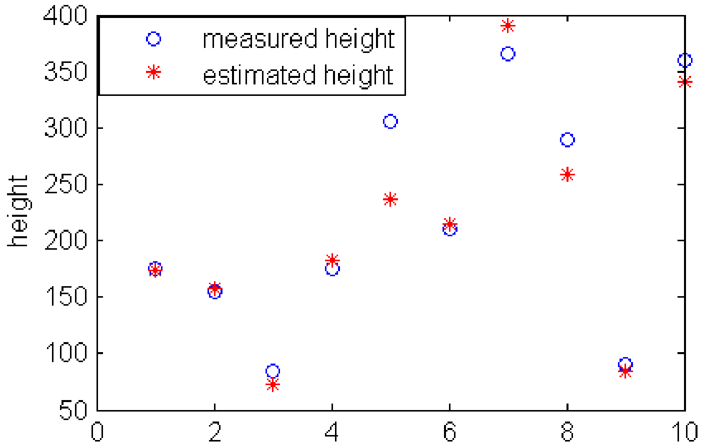


Figure 5.17. Error of the height estimation in field test.

Table 5.2. Accuracy of the height estimation.

	Laboratory Test	**Field Test**
The range of height variety (mm)	48 -141	85 – 361
RMS error without regression (mm)	6.9	23.7

After the height estimated, the coordinate transformation matrix can be adjusted accordingly. Then the accurate 2D positioning can be provided for plants with various heights.

5.3.3 Error analysis

Two major reasons of errors are: centroids determination in plant identification and speed information from encoder.

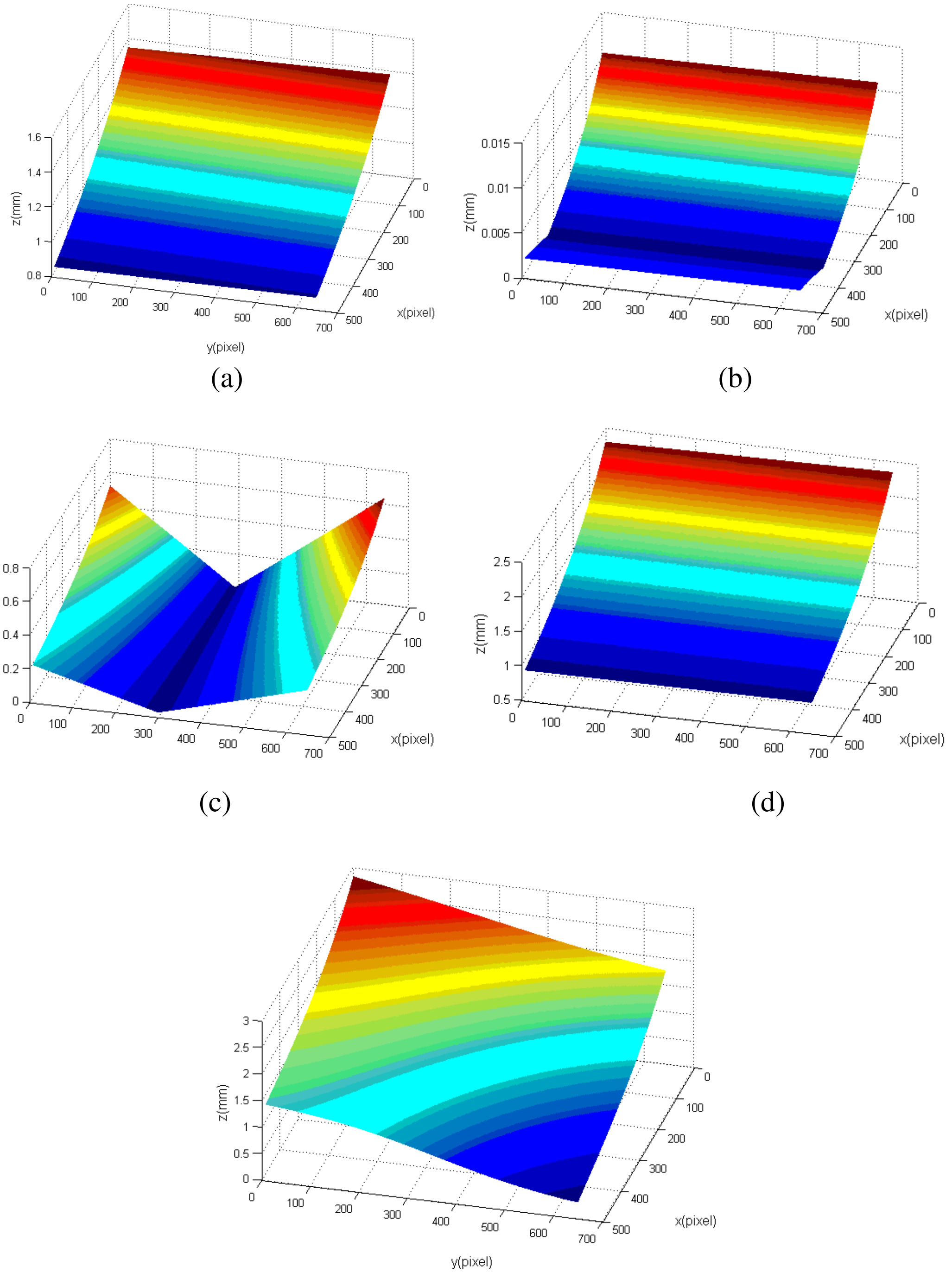
The plants in the field usually have many leaves overlapping each other. The heights of the leaves, however, are different. This makes the plant green from bottom to the top. The plant identification method in this project assumes the z coordinates of the centroids to be the heights of plants. In some situations, this assumption can provide reliable simplifications (figure 5.18a). However, sometimes this brings large errors (figure 5.18b).

(a) (b)

Figure 5.18. (a) Example of similar z coordinate of the centroid and the height of the plant, (b) example of different z coordinate and height.

As mentioned before, due to the angle between camera reference frame and the ground frame, the corresponding ground resolution in ground frame is not consistent. That means different areas of the image have different resolutions, and on the ground the area that one pixel covers is no longer a square. This in fact results in the problem that objects measured further from camera were less accurate. Figure 5.19 shows the trends of variation of the error. One pixel error along x axis in pixel coordinate may cause corresponding different real error in ground coordinate. Figure 5.19a is the error (mm) along x axis in ground frame while corresponding pixel error is 1 along x axis in pixel frame. Figure 5.19b is the error (mm) along y axis in ground frame while corresponding pixel error is 1 along x axis in pixel frame. Figure 5.19c is the error(mm) along x axis in ground frame while corresponding pixel error is 1 along y axis in pixel frame. Figure 5.19d is the error(mm) along y axis in ground frame while corresponding pixel error is 1 along y axis in pixel frame. Figure 5.19e is the error distribution with 1 pixel error along both two axes.

The other source of errors is the error in velocity and displacement measurement. The displacement between two images is the key for measuring the heights accurately. However, while such displacement is easy to measure in ordinary stereo vision system, which is merely the distance of two optical axes of two cameras, it is very difficult to determine in this monocular



(a) (b)

(c) (d)

(e)

Figure 5.19. Relationship between pixel error and spatial error.

stereo vision system. In this particular system, the subsequent two frames are chosen to mimic the two parallel images in stereo vision system, therefore, the distance of optical axis become the displacement that the robot has moved between two frames. This movement is difficult to measure, so the estimation based on velocity and time measurement is a good solution.

The velocity measurement is based on the encoder of the robot, which introduces new errors, distance error. From the figure 5.20 the relationship between distance error and the computed heights are obviously identified. This figure indicates the variation of computed height while the displacement measured varies within a small range. This relationship is almost linear. Therefore, the error in displacement will not worsen the condition of the algorithm.

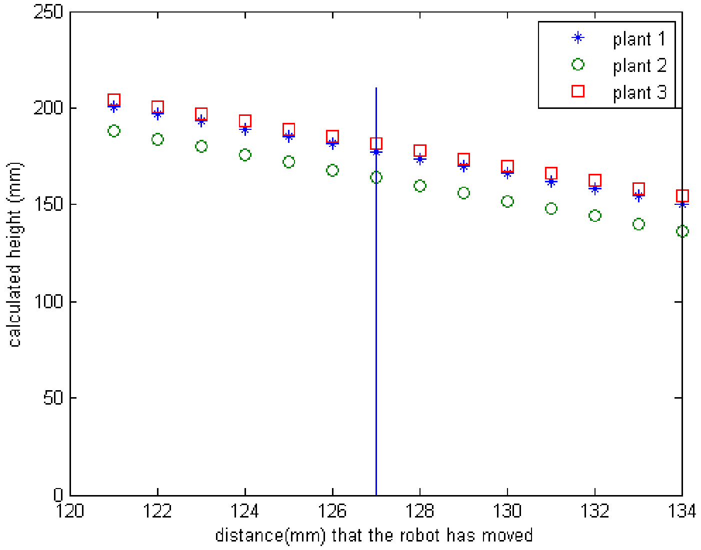


Figure 5.20. Sensitivity of the algorithm with the distance as input. The error in detected distance of robot moving will result in relatively large errors in height estimation.

The reason of error in displacement measurement can be explained graphically. Figure 5.21 is the curve of the real trajectory of the robot that is moving in the field, while the curve line is the measurement of robot's movement. The displacement between two frames is the displacement of two positions of camera where the images were captured, thus this displacement is a straight line in the ground frame. Our goal is to use the movement measurement as the estimation of such displacement. The function of the encoder is velocity measurement. Therefore the estimation of movement is the integration of speed. Considering the time resolution of the encoder, the final integration of velocity is therefore a curve line, which may be a little longer than a straight line. This results in errors in estimation of displacement between frames. This error occurs when the robot's movement is no longer a straight line, which happens frequently in off-road application.

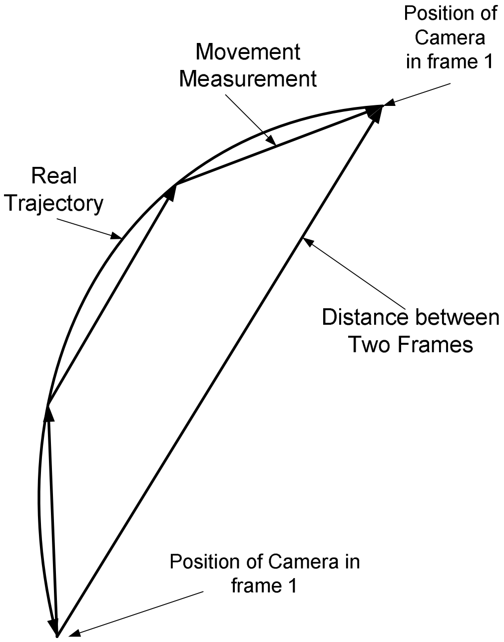


Figure 5.21. Distance detection from encoder.

There are some minor reasons for errors, such as wind. The traditional stereo-vision system can take two images at the same time that guarantee the maximum similarity of same plant in two images. However, in monocular system, the two images were taken one by one. The appearance of the plant may change during the time interval, due to the wind. This difference of appearance will result in errors in plant identification process.

5.4 Conclusions

This chapter discussed another application of camera coordinate transformation, the monocular vision system for a field robot. Similarly to the geo-reference of tower remote-sensing system, 2D plant positioning was introduced using a single camera. Another approach is the algorithm of measuring attitudes of plants by combining the coordinate transformation and the robot's speed information.

For evaluating the accuracy of algorithms, two types of experiments, laboratory test and field test, were conducted. The reason for developing two tests is that by comparing the results, the sources of errors can be revealed. In the laboratory test, the 2D coordinate measurement had shown relatively accurate results. The error ranged from 2.8 mm to 5.9 mm. After linear regression, the average errors decreased to smaller than 3.3 mm. For height measurement, the linear regression was not proper anymore since the linearity was weak. The RMS error of height measurement in the lab test was 6.9 mm. In the field test, the errors of both 2D coordinates and heights increased. The RMS error of 2D after regression were 5.4 mm and 23.0 mm along x and y axes, respectively. The RMS error in the plant height estimation is 23.7 mm. By comparing the two results, the two major sources were the error of encoder and the error in plant identification.

CHAPTER 6 CONCLUSION AND FUTURE WORK

6.1 Summary of Work and Conclusion

In this research, two applications of camera coordinate transformation were developed. For the tower remote sensing system, the geo-reference algorithm was developed based on this coordinate transformation with which the ground and image coordinates of objects were related. The application of coordinate transformation successfully solved the problem of the lack of GCPs. Necessary experiments were conducted for evaluating the accuracy. The error of the algorithm was in an acceptable range, with an average of 57.15 pixels. The main source of inaccuracy is the error from compass reading.

Image mosaicing is the process of combining several separate images together to provide integral information of a certain field. Unlike the other feature based method, the mosaic algorithm in this project was based on geographical information from the geo-referencing. In addition, since the pitch of the tower remote sensing images was usually 10 to 15 times larger than the aerial images, an algorithm named pixel combination was applied. The first step of this algorithm transformed all the images to a top-view format. In global alignment, the images were placed in the same plane based on their position in the ground frame. Then the mismatches were reduced with a local alignment procedure. A total of 71 images were captured for testing the algorithm. By comparing the positions of markers in different images, the performance of the image mosaic algorithm was evaluated. Final results revealed the error was 14.5 to 15.3 pixels.

The other application of coordinate transformation was for the monocular vision system of a field robot. Unlike other monocular stereovision methods, such as using mirrors and prisms to acquire two parallel images, camera coordinate transformation plus the speed information could extraction of the 3D coordinates of the object directly. With the transformation, the positions of plants that were in front of the robot could be measured. Furthermore, with the speed information measured by the encoder on the robot, the heights of the plant could be estimated by the camera. Two types of experiments were conducted to test the accuracy of this algorithm. In the lab tests, the accuracy of both 2D coordinates and height estimation was acceptable. In the field tests, the accuracy was lower because of complex environment. By comparing the results of the two tests, the sources of errors were addressed.

6.2 Recommendation for Future Work

Future work should focus on improving accuracy and extending applications. For geo-reference, since the main source of error was the compass reading, a better method of calibrating the compass is needed. Combining GCP method with coordinate transformation could be a solution for improving the geo-reference accuracy. With several permanent markers in the field as the reference, the compass reading can be corrected. Currently the only sensor for geo-reference is the compass, thus sensor fusion may be the future way of improving accuracy.

The value of the monocular vision system is that it can be easily incorporated with other visual sensing systems for robots. For example, a handheld NDVI sensing system was under development in the agricultural remote-sensing lab at the University of Illinois at Urbana-Champaign. The sensing system is a single camera that can measure the NDVI of plants by taking images. If this system is installed and combined with the monocular vision algorithm developed in this project, it can add spectral information for plant measurement. Also, the monocular vision algorithm can be easily implemented in other camera based sensing systems.

There is one possibility to combine these two applications. The field robot can be a site specific solution that can be a complement for the tower remote sensing system. The future plan is as follows: every time a point of interest is found in the field, the operator can remotely control the robot move to specified point for future application, such as data acquisition, sampling or applying herbicides.

REFERENCES

ActiveMedia Robotics. 2003. Poineer 3TM&Pioneer 2TM H8-Series Operations Manual, Version 3. ActiveMedia Robotics, LLC. Amherst, New Hampshire. Available at: http://www.mobilerobots.com/ResearchRobots/ResearchRobots/P3AT.aspx. Accessed 20 April 2010.

Ahamed, T., L. Tian, Y. Zhang, Q. Zhang, T. Grift, and K.C. Ting. 2009. Engineering solutions for biomass feedstock production - pre-harvest crop monitoring system. In *Proc. ASABE Annual International Meeting,* 1618-1631, Reno, Nevada.

Balaselvakumar, S., and S. Saravanan. 2002. Remote sensing techniques for agriculture survey. Available at: http://www.gisdevelopment.net/application/agriculture/overview/agrio014.htm. Accessed 20 April 2010.

Balz, T. 2006. Automated CAD model-based geo-reference for high-resolution SAR data in urban environments. *IEE Procedings: Radar Sonar Navig.* 153(3): 289-293.

Bastiaanssen, W. G. M., D. J. Molden, and I. W. Makin. 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. *Agricultural Water Management* 46(2): 137-155.

Bertozzi, M., and A. Broggi. 1998. GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection, *IEEE Transactions on Image Processing* 7(1): 62-81.

Brox, T., A. Bruhn, N. Papenberg, and J. Weickert. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conference on Computer Vision 2004,* 25-36, Prague, Czech.

Campion, D.C., J. P. Dugan, C. C. Piotrowski, and A. G. Evans. 2002. Direct geo-reference technique for rapid positioning of targets and environmental products using tactical grade airborne imaging data. In. *Proc. Oceans Conf Rec IEEE,* 1603-1608, Mississippi City, MS

Canty, M. J., A. A. Nielsen, and M. Schmidt. 2004. Automatic radiometric normalization of multi-temporal satellite imagery. *Remote Sensing of Environment* 91(3-4): 441-451.

Constable, G., and B. Somerville. 2003. *A Century of Innovation, Twenty Engineering Achievements that Transformed Our Lives.* Washington. D.C.: Joseph Henry Press, ISBN-13: 978-0309089081.

Criminisi, A., I. Reid, and A. Zisserman. 2000. Single view metrology *International Journal of Computer Vision* 40(2): 123–148.

Cristian F., S. Gheorghe, C. Vasile, and S. Elena. 2007. A validation of MODIS snow cover products in Romania: challenges and future directions. *Trans. GIS.* 11(6): 927–941.

Durr, P.A., and A. E. A. Froggatt. 2002. How best to geo-reference farms? a case study from

Cornwall, England. *Preventive Veterinary Medicine* 56: 51-62.

Forthy, D., and J. Ponce. 2002. *Computer Vision: A Modern Approach.* Upper Saddle River, NJ.: Prentice Hall, ISBN-13: 978-0130851987.

Fitzgerald, G. J., S. M. Lesch, E. M. Barnes, and W. E. Luckett. 2006. Directed sampling using remote sensing with a response surface sampling design for site-specific agriculture. *Computers and Electronics in Agriculture* 53(2): 98-112.

Heikkilä, J., and O. Silvén. 1997. A four-step camera calibration procedure with implicit image correction. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1106-1112, Los Alamitos, CA.

Ines, A. V. M., K. Honda, A. D. Gupta, P. Droogers, and R. S. Clemente. 2006. Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. *Agricultural Water Management* 83: 221-232.

Jeon, H. 2008. Plant specific direct chemical application field robot. PhD diss. Urbana, IL.: University of Illinois at Urbana-Champaign, Department of Agricultural and Biological Engineering.

Jun, W., Z. Dong, Z. Liu, and G. Zhou. 2007. Geo-registration and mosaic of UAV video for quick-response to forest fire disaster. In *Proc. SPIE Int. Soc. Opt. Eng.* 6788, No. 678810 Wuhan, China.

Lin, T. T., Y. K. Hsiung, G. L. Hong, H. K. Chang, and F. M. Lu. 2008. Development of a virtual reality GIS using stereo vision. *Computers and Electronics in Agriculture* 63(1): 38-48.

Mark, J., and P. J. Hardin. 2005. Applications of inexpensive Remotely Piloted Vehicles (RPVs) for collection of high-resolution digital imagery. In *Proc. 20th Biennial Workshop on Aerial Photography, Videography, and High-Resolution Digital Imagery for Resource Assessment,* Weslaco, TX.

Martin, M.C. 2000. Evolving visual sonar: depth from monocular images *Pattern Recognition Letters* 27: 1174-1180.

Matsumoto, Y., H. Terasaki, K. Sugimoto, and T. Arakawa. 1997. Conversion system of monocular image sequence to stereo using motion parallax. In *Proc. SPIE - The International Society for Optical Engineering* 3012: 108-115.

Mordohai, P., and G. Medioni. 2006. Stereo using monocular cues within the tensor voting framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(6): 968-982.

Nielsen, A. A., K. Conradsen, and J. J. Simpson. 1998. Multivariate Alteration Detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: new approaches to change detection studies. *Remote Sensing of Environment* 64: 1-19.

Nishimoto, Y., and Y. Shirai. 1987. Feature-based stereo model using small disparities. In *Proc. IEEE Comput. Soc. Conf. CVPR,* 192-196.

Nistér, D. 2003. An efficient solution to the five-point elative pose problem, In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* 2: 195-202.

Nistér, D., O. Naroditsky, and J. Bergen. 2005. Visual odometry for ground vehicle applications. Available at: http://www.vis.uky.edu/~dnister/Publications/2006/JFR/JFR1.pdf. Accessed 20 April 2010.

Nowatzki, J., R. Anders, and K. Kyllo. 2004. Agricultural remote sensing basics. Available at: http://www.ag.ndsu.edu/pubs/ageng/gis/ae1262.pdf. Accessed 20 April 2010.

Pachidis, T. P., and J. N. Lygouras. 2007. Pseudostereo-vision system: a monocular stereo-vision system as a sensor for real-time robot applications. *IEEE Transactions on Instrumentation and Measurement* 56: 2547-2560.

Price, R., and P. Alli. 2005. Development of a system to automatically geo-rectify images and allows quick transformation into a prescription map. ASAE Paper No. 051011. St. Joseph, MI.

Redlake, Inc. 2000. User's Manual. San Diego, CA. Available at: http://www.syncron.co.kr/product/data/MS4100.pdf. Accessed 20 April 2010.

Ray, S., S. Panigrahy, and J. S. Parihar. 2001. Precision farming in Indian context – role of remote sensing. Available at: http://www.gisdevelopment.net/application/agriculture/overview/agrio0003pf.htm. Accessed 20 April 2010.

Rocchini, D., and A. D. Rita, 2005. Relief effects on aerial photos geometric correction. *Applied Geography* 25: 159-168.

Saxena, A., M. Sun, and Y. A. Ng, 2009. Make3D: learning 3D scene structure from a single still image *IEEE Transactions on Pattern Analysis and Machine intelligence* 31: 824-840.

Scheidt, S., and M. Amsey. 2007. Radiometric normalization and image mosaic generation of ASTER thermal infrared data: an application to extensive sand sheets and dune fields. *Remote Sensing of Environment* 112(3): 920-933.

Shum, H. Y., and R. Szeliski, 2000. Systems and experiment paper: construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2): 101-130.

Skaloud, J. 1999. Optimizing geo-reference of airborne survey systems by INS/DGPS. PhD diss. Calgary, Canada.: University of Calgary, Department of Geomatics Engineering.

Sparton Electronics 2008. Specification Sheet. Brooksville, FL. Available at: http://www.thedigitalcompass.com/SP3004D.pdf. Accessed 20 April 2010.

Teoh, W., and X. Zhang. 1984. An inexpensive stereoscopic vision system for robots. In *Proc. Int. Conf. on Robotics and Automation*, 1: 186-189.

Thomson, S. J., J. E. Hanks, and G. F. Sassenrath-Cole. 2002. Continuous geo-reference for video-based remote sensing on agricultural aircraft. *Trans. of the ASAE*. 45: 1177-1189.

Toth, C. K. 2002. Sensor integration in airborne mapping. *IEEE Trans. on Instrumentation and Measurement* 51(6): 1367-1373.

Wang, S., and K. Ishii. 2009. Depth perception using a monocular vision system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5506 LNCS (PART 1): 779-786.

Warren, D. H., and E. R. Strelow. 1985. *Electronic Spatial Sensing for the Blind: Contributions from Perception*. Berlin, Germany.: Springer. ISBN: 9024726891.

Xiang, H. 2008. Development and investigation of autonomous agricultural remote sensing systems with high spatial and temporal resolutions. PhD diss. Urbana, IL.: University of Illinois at Urbana-Champaign, Department of Agricultural and Biological Engineering.

Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 22(11): 1330-1334.

Zhou, G., C. Li, and P. Cheng. 2005. Unmanned Aerial Vehicle (UAV) real-time video registration for forest fire monitoring. In *Proc. International Geoscience and Remote Sensing Symposium,* 3: 1803-1806.

Zhu, Z., E. M. Risemen, A. R. Hanson, and H. Schultz. 2005. An efficient method for geo-referenced video mosaicing for environmental monitoring. *Mach. Vision Appl.* 16(4): 203-216.