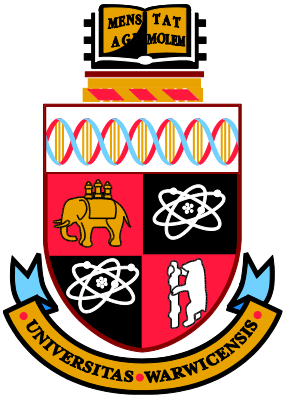


Multi-Variate Image Analysis for Detection of Biomedical Anomalies

Shan-e-Ahmed Raza

Thesis

submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

Department of Computer Science

The University of Warwick

January 2014

THE UNIVERSITY OF WARWICK

I dedicate this thesis to my parents, my teachers especially my supervisor Dr Nasir Rajpoot, my wife, my siblings and my sweet little niece.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Dr Nasir Rajpoot for offering me the opportunity to work with him at his lab at The University of Warwick under his kind supervision. I am grateful to his valuable ideas and suggestions and for his support during the writing process. His constant support, guidance and encouragement helped me to complete this thesis. I would also like to thank my co-supervisors Dr John Clarkson and Dr Andrew Thompson for setting up experiments and for their guidance to help me better understand the plant science.

I am also thankful to my PhD advisory panel Dr Sara Kalvala and Prof Chang-Tsun Li for their guidance and valuable suggestions on my PhD progress. I would especially like to thank Prof David Epstein and Dr Victor Sanchez for their kind suggestions and valuable remarks. I am also thankful to Prof Tim Nattkemper, Dr Metin Gurcan and Dr Khalid Niazi for their valuable feedback on my work.

I am very grateful to Dr Michael Khan, Prof Gail Taylor, Dr Graham Clarkson, Ms Sylvie Abouna, Mrs Gillian Prince and Ms Hazel Smith for setting up experiments for the collection of data for my experiments. I am thankful to Dr Roger Packwood, Mr Rod Moore and Mr Stuart Valentine for their technical help to set up equipment for capturing and analysing the images.

My special thanks to Horticultural Development Company (Project Number CP60a) and Department of Computer Science for funding my PhD and supporting this project.

I would like to thank all the members of my lab Adnan Khan, Ahmad Humayun,

Guannan Li, Korsuk Sirinukunwattana, Mike Song, Nicholas Trahearn, Samuel Jefferyes, Tian Ge and Violet Kovacheva for their help and support during my PhD.

Finally and most especially, I would like to thank my family for their support and encouragement during all these years.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I declare that, except where acknowledged, the material contained in the thesis is my own work, and has not been previously published for obtaining an academic degree.

Shan-e-Ahmed Raza

January 15, 2014

List of Publications

1. **S.-E.-A. Raza**, A. Humayun, S. Abouna, T. W. Nattkemper, D. B. Epstein, M. Khan, and N. M. Rajpoot, "RAMTaB: Robust alignment of multi-tag bioimages," *PLoS ONE*, vol. 7, no. 2, p. e30894, 2012.
2. **S.-E.-A. Raza**, V. Sanchez, G. Prince, J. Clarkson, and N. M. Rajpoot, "Registration of thermal and visible light images using silhouette extraction," *Pattern Recognition* (accepted), 2013.
3. **S.-E.-A. Raza**, H. K. Smith, G. J. J. Clarkson, G. Taylor, J. Clarkson, and N. M. Rajpoot, "Automatic detection of water stress in spinach canopies using combined visible and thermal imagery," *PLoS ONE* (under review), 2013.
4. A. Humayun, **S.-E.-A. Raza**, C. Waddington, S. Abouna, M. Khan, and N. M. Rajpoot, "A Framework for molecular co-expression pattern analysis in multi-channel Toponome fluorescence images," in *Microscopic Image Analysis with Applications in Biology (MIAAB 2011)*, 2011.
5. A. M. Khan, **S.-E.-A. Raza**, M. Khan, and N. M. Rajpoot, "Cell phenotyping in multi-tag fluorescent bioimages," *Neurocomputing*, 2014.
6. A. M. Khan, A. Humayun, **S.-E.-A. Raza**, M. Khan, and N. M. Rajpoot, "A novel paradigm for mining cell phenotypes in multi-tag Bioimages using a locality preserving nonlinear embedding," in *The International Conference on Neural Information Processing (ICONIP)*, 2012.

Abstract

Multi-modal images are commonly used in the field of medicine for anomaly detection, for example CT/MRI images for tumour detection. Recently, thermal imaging has demonstrated its potential for detection of anomalies (e.g., water stress, disease) in plants. In biology, multi-channel imaging systems are now becoming available which combine information about the level of expression of various molecules of interest (e.g., proteins) which can be employed to investigate molecular signatures of diseases such as cancer or their subtypes. Before combining information from multiple modalities/channels, however, we need to align (register) the images together in a way that the same point in the multiple images obtained from different sources/channels corresponds to the same point on the object (e.g., a particular point on a leaf in a plant or a particular cell in a tissue) under observation. In this thesis, we propose registration methods to align multi-modal/channel images of plants and human tissues. For registration of thermal and visible light images of plants we propose a registration method using silhouette extraction. For silhouette extraction, we propose a novel multi-scale method which can be used to extract highly accurate silhouettes of diseased plants in thermal and visible light images. The extracted silhouettes can be used to register plant regions in thermal and visible light images. After alignment of multi-modal images, we combine thermal and visible light information for classification of water deficient regions of spinach canopies. We add depth information as another dimension to our set of features for detection of diseased plants. For depth estimation, we use disparity between stereo image pair. We then compare different disparity estimation algorithms and propose

a method which can be used to obtain not only accurate and smooth disparity maps but also less sensitive to the acquisition noise. Our results show that by combining information from multiple modalities, classification accuracy of different classifiers can be increased.

In the second part of this thesis, we propose a block-based registration method using mutual information as a similarity measure for registration of multi-channel fluorescence microscopy images. The proposed block-based approach is fast, accurate and robust to local variations in the images. In addition, we propose a method for selection of a reference image with maximal overlap i.e., a method to choose a reference image, from a stack of dozens of multi-channel images, which when used as reference image causes minimum amount of information loss during the registration process. Images registered using this method have been used in other studies to investigate techniques for mining molecular patterns of cancer.

Both the registration algorithms proposed in this thesis produce highly accurate results where the block-based registration algorithm is shown to be capable of registering the images up to sub-pixel accuracy. The disparity estimation algorithm produces smooth and accurate disparity maps in the presence of noise where commonly used disparity estimation algorithms fail to perform. Our results show that by combining multi-modal image data, one can easily increase the accuracy of classifiers to detect anomalies in plants, which helps to avoid huge losses due to disease or lack of water at commercial level.

Chapter 1
Introduction

An image contains information about a particular *physical characteristic* of an object(s) being imaged, depending on the type of camera being used to capture the image. For example, a visible light imaging camera captures electromagnetic waves reflected by an object in the visible light spectrum, or we can say that one of the physical characteristics of the object to reflect particular wavelengths in the electromagnetic spectrum has been captured in the form of an image. There are many other imaging modalities which capture different information in the form of an image. Thermal imaging cameras capture heat signatures of an object using electromagnetic waves in the infrared region of the spectrum. Ultrasound and Sonar images contain information about the amount of sound waves reflected by the object of interest. Computed Tomography (CT) imaging is widely used in the field of medicine, CT images contain information about the amount of X-rays reflected and absorbed by a particular organ or part of body. Similarly, Magnetic Resonance Imaging (MRI) uses strong electromagnetic fields to visualize internal organs of a body. Fluorescence Microscopy uses fluorescent dyes to get information about protein structures present in a tissue under the microscope. There are many other imaging techniques which can be used to get information about an object in the form of images.

The sensors associated with almost all of the imaging modalities can individually capture only limited amount of information about certain physical characteristics. In some cases, it is useful and sometimes necessary to integrate this information,

captured using multiple modalities and at multiple angles, for in-depth analysis of an object under observation to overcome the limitations. For example it is often desired to combine information from CT and MRI for more accurate tumour definition [1–3]. With the development of new techniques such as fluorescence microscopy, the multi-channel image analysis can also be used for accurate tumour definition and understanding at the microscopic level. Similarly, multi-modal image analysis can be extended to other fields in biology for example, thermal imaging can be combined with visible light imagery for anomaly detection in plants.

The first step to combine information from multiple modalities is usually the spatial alignment of data from different modalities, commonly known as registration. However, automatic registration of multi-modal images is a challenging task compared to uni-modal registration. The main hurdle in the registration process is that the images are captured from different viewpoints and contain different lens distortions. This introduces complex transformations required to perform registration of multi-modal images. Another big challenge is that the images captured using multiple modalities contain different information. Sometimes information contained in the image captured by one modality is missing in the image captured by the other modality. Therefore, it becomes difficult to capture stable features which can be used as a cue for registration. Because of the difficulties involved in multi-modal image registration, many important biological questions remain unanswered as the amount of information captured by a single modality is usually limited.

In this thesis, we apply the idea of using multi-modal images for analysis to two different projects to accurately detect anomalies in plants and human tissues. We

present registration algorithms to align 1) multi-modal images of plants captured using conventional thermal and visible light cameras and 2) multi-channel images of tissues captured using an emerging new technology based on fluorescence microscopy known as Toponome Imaging System (TIS). After alignment of multi-modal data we add depth information to the plant images and propose method(s) to detect water stress and disease detection in plants using colour, thermal and depth information.

1.1. Thermal Imaging

Infrared radiation was first discovered by an astronomer, Sir William Herschel, in 1800, while he was trying to determine the colour among different sunlight colours which was responsible for heating [4]. Thermal imaging converts the thermal infrared radiation pattern of an object mainly in the electromagnetic spectrum (3-14 μm) to visible images. Thermal imaging techniques were mainly developed for military purposes but now have been used for a wide range of applications in agriculture, industry, civil engineering, aerospace, medicine and veterinary.

Similar to visible light imaging cameras, a thermal imaging camera comprises of a detector, signal processing unit and an image acquisition system. There are two types of detectors used in thermal imaging cameras to detect IR radiation, thermal and photon detectors [5]. In thermal detectors, the infrared radiation heats the detector element which is taken as a measure of radiation falling on the detector. In photon detectors, the radiation interacts at atomic or molecular level with the detector material, this interaction may involve a photon and an electron, resulting in the electron moving through quantum energy levels and producing charge

carriers that generate voltage across the detector. The photon detectors are more sensitive than the thermal detectors, but they need to be cooled down for the electrons to come back to the desired energy levels for interaction. Thermal detectors normally do not require cooling, but they are less sensitive and provide lower resolution. The signal processing unit and image acquisition system in thermal imagers are similar to a visible light imaging camera.

1.1.1. Thermal Imaging in the Horticulture Industry

Under water stress (deficit) conditions, plants generally tend to close their stomata, and the rate of transpiration is reduced. An important consequence of stomatal closure is that energy dissipation is decreased and hence the leaf temperature tends to rise. It also widens the range of temperature variation within the canopy which can be detected using infrared thermometry or by the use of thermal imagers [4,6,7]. The non-contact and non-destructive nature and repeatability of measurements makes thermal imaging quite useful in the agriculture and food industry [8]. However, there are some complications associated with the image data captured using thermal imagers. Various researchers have investigated and reported in their results that sunlight, shade, distance of plant regions from thermal imagers and leaf angles can affect the temperature information captured by the thermal imagers [9,10]. There has been a lot of research focus on stress analysis and disease detection of plants using thermal imaging; however, few researchers have exploited the information from the visible light images for analysis.

(a) (b)

Figure 1-1: (a) Cedip Titanium SC7000 (Image credit: SC7000 manual), (b) TH9100 WRI (Image credit: TH9100 Manual).

In the horticulture industry, thermal imaging has been used in predicting crop water stress, early disease detection in plants, determining genotype and phenotype, predicting fruit yield, bruise detection and detection of foreign bodies in food material. Thermal Cameras used by various researchers in the past include FLIR systems Thermovision 900LW, Thermacam P25, Thermacam PM250, Thermacam SC2000, Varioscan 3201 ST, Infrared solutions Snapshot 225, IR Snapshot 525. Thermal images used in this thesis were taken by Cedip Titanium SC7000 from the EPSRC instrument pool and a TH9100WR thermal camera, as shown in Figure 1-1.

1.1.2. Current Multi-modal Imaging Methods in the Horticulture Industry

Horticulture industry is adapting multi-modal imaging for fast and reliable

assessment of their crop. *LemnaTec*[1] are using infrared imaging combined with visible and fluorescence imaging for the purposes of plant phenotyping [11]. They have developed a hardware based system called Scanalyzer 3D which can take 3D visible images of the plants. It can scan the plants using different wavelengths which include infrared, visible and fluorescent light. Infrared light helps to quantify temperature differences within leaves and plants. Plant health, nutrients, senescence and phenotype can be studied using visible light images. Near infrared images are used to determine the water content of the soil. There are other systems such as Scanalyzer ESC and Scanalyzer HTS which have been developed to study phenotyping in environmental simulation chambers by taking images of the plants using a moving camera or a conveyer.

Phenovation[2] in Netherlands have developed systems based on multi-spectral and multi-fluorescence imaging to study the performance of photosynthesis in plants. The CropReporter system developed by *Phenovation* uses high power LED illumination as excitation source and high performance computers to process the images at several frames per second (800 frames/measurement in less than 3 seconds). These systems have limited use and can be used in a specific environment and not in greenhouse or open fields. *Green Vision Systems*[3] in Netherlands are using Near Infrared Spectroscopy to control the product quality and production process in the agro and food industry. Although different systems have been

[1] http://www.lemnatec.com/

[2] http://www.phenovation.com/

[3] http://www.greenvs.com/

developed in the horticulture industry which use multi-spectral and infrared imaging to assess the quality of product, there is still a lot of potential in joint multi-modal analysis of plants [12–14].

The National Plant Phenomics Centre (NPPC)[4] at Aberystwyth University features a state of the art greenhouse which can be used to study individual plants on a series of moving conveyer belts. The greenhouse contains computer controlled cameras which can provide 3D images of individual plants using multiple modalities including fluorescence, infra-red and near infra-red, laser and root imaging technology to monitor their growth.

In this thesis, we propose registration methods to align images of plants from multiple modalities and aim to investigate the potential of combined thermal, colour and depth information in order to detect anomalous conditions such as water stress or disease onset in plants. Our proposed registration algorithm produces highly accurate registration results (Chapter 3). After combining thermal and visible light information from normal and diseased/stressed plants we propose methods for anomaly detection in Chapter 5.

[4] http://www.plant-phenomics.ac.uk/en/

Figure 1-2: Scanalyzer HTS developed by LemnaTec to study plant phenotyping in simulated environments. Image Credit: [11].

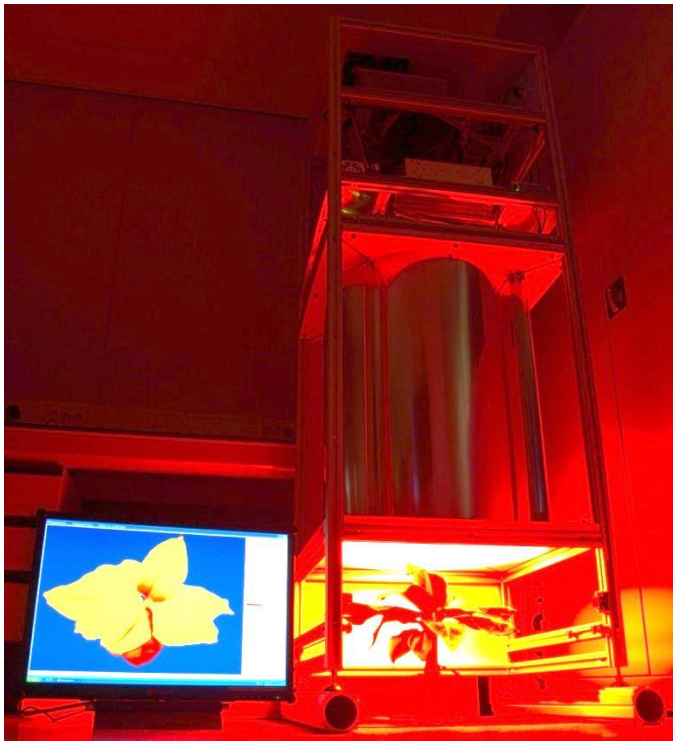


Figure 1-3: CropReporter High Resolution at Phenovations to study efficiency of photosynthesis in different light conditions. Image Credit: http://www.phenovation.com

1.2. Multi-Channel Fluorescence Microscopy Imaging

A cell in a human tissue can be defined as an assembly of hundreds of proteins which interact together to define cell functions [15,16]. These protein compositions can be decoded by using modern fluorescence imaging techniques. Most fluorescence microscopy techniques are limited to up to ten fluorescent tags which can point to simultaneous localisation of the corresponding biomolecules and protein structures inside the cells of a tissue specimen [17]. Toponome Imaging System is an automated robotic microscopy system based on fluorescence microscopy to locate dozens of different proteins or other biomolecules (in a cell or a tissue) by using fluorescently labelled antibodies, lectins or other specific ligands [18,19].

The TIS system provides us with a platform to decode and locate a large number (of the order of several thousands) of possible protein combinations at a given point in a cell. It uses a library of fluorescent tags to obtain phase-fluorescence pair images corresponding to each tag. One data set, a stack of grey value images, is recorded by performing N sequential cycles of fluorescent tagging, labelling and bleaching *in situ*. In each iterative step j, a fluorescence or tag image F_j and a corresponding phase contrast image I_j is recorded before and after incubation with a particular tag. So for each tag, e.g., an antibody against a specific protein or a dye such as DAPI that stains nuclei, we obtain fluorescence and phase contrast images.

There are other new technologies such as Matrix-Assisted Laser Desorption/Ionization (MALDI) imaging [20] or Raman microscopy [21] which record high dimensional images, organised as stacks of grey value images, encoding

the co-location or interaction of a large number of molecules. Another group of new bioimaging approaches achieve this by using different fluorophores, multispectral analysis, or bleaching with only one fluorophore [18,19,22–24]. The resultant image data consist of a stack of N grey value images I_j (j=1, …,N) where each image shows the spatial distribution of one molecule. Due to these techniques becoming ubiquitous, new computational approaches are needed to process and visualise multi-variate bio-images [25,26].

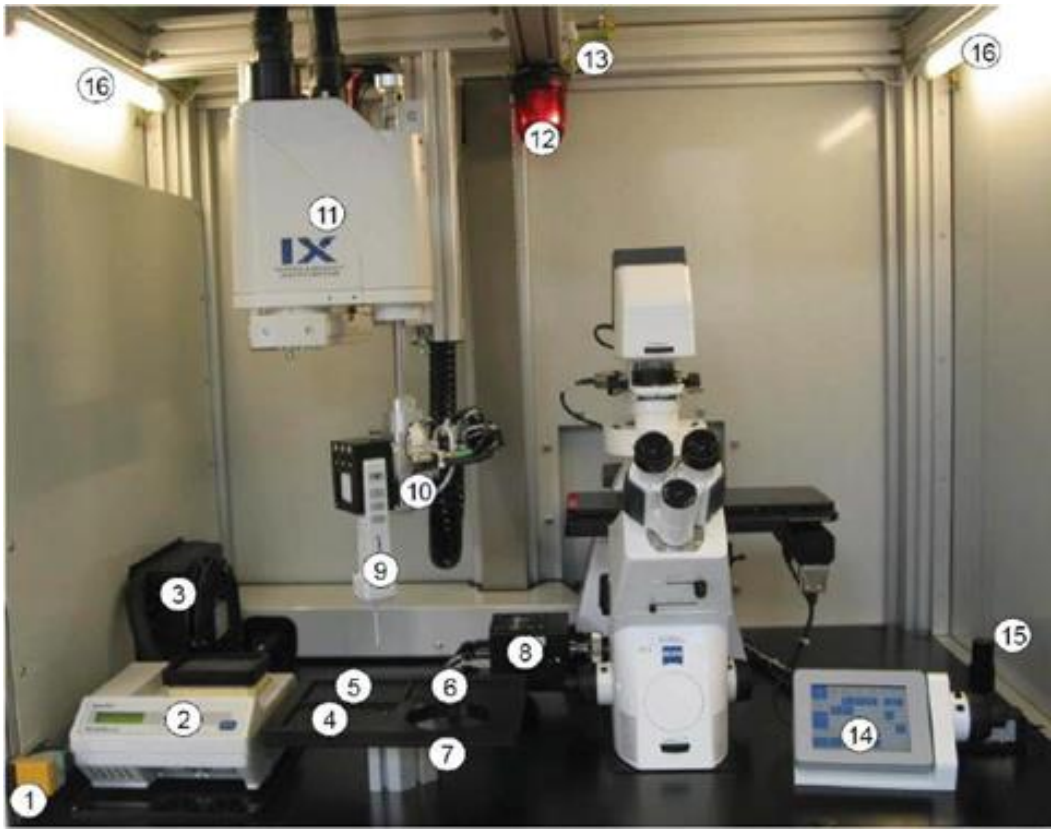


Figure 1-4 The Toponome Imaging System (TIS) installed at Warwick. The numbers marked in the figure represent (1) safety door sensor, (2) thermostat tag cooling, (3) heat exchanger and cooling fan, (4,5,6) holder(s) for deep well plates, (7) holder for beaker, (8) CCD camera, (9) pipette, (10) rotational axis, (11) SCARA robot (12) warning lamp, (13) temperature sensor, (14) touch screen control for microscope, (15) stage control.

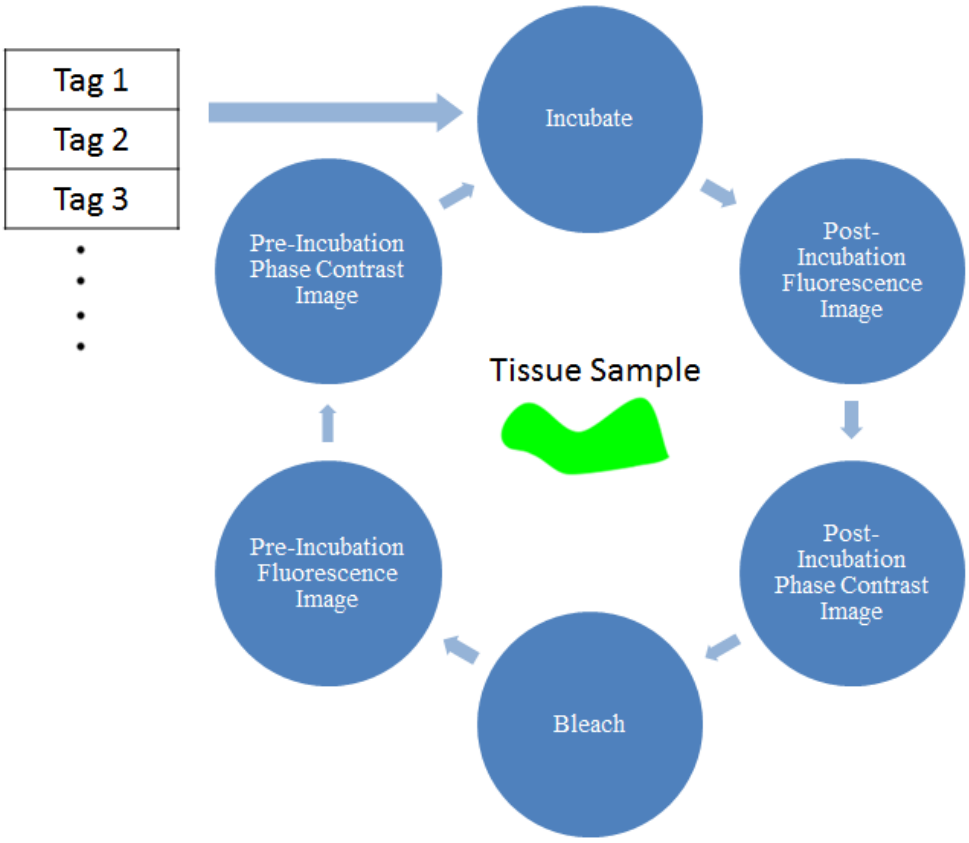


Figure 1-5: Toponome Imaging System data acquisition cycle.

Growth in molecular dimension is often accompanied with a growth in runtime of the imaging experiment. As a consequence, serious shifts can be observed between pairs of images in one stack, recorded for one field of view, making a direct analysis of the co-location signals meaningless. In experiments lasting for several hours, shifts can be caused by various external influences (mechanical perturbations, temperature changes, shift movements by the specimen due to repeated washes etc). Since most analytical approaches are based on processing N grey values $I_1, \dots, I_N(x, y)$ associated to a pixel (x, y) and searching the images for interesting patterns of co-location, for instance using clustering and dimension reduction [27], it is vital that all N images in a stack are accurately aligned. Unless this can be achieved, many important biological questions, such as cell classification and

discovery of functional protein networks within a cell at different points in time, cannot be addressed.

In this project, we address the problem of misalignment among different anti-body tags and present an approach to select a reference image with maximal overlap to minimise information loss during registration process. After registration, we discuss preliminary analysis performed in our lab using multi-channel fluorescence images of colon cancer tissue.

1.3. Aims of the Thesis

Many image analysis techniques have limited use because of limitation of the imaging modality being used to image the object under observation. The capacity of these analysis techniques can be greatly improved by using images from multiple modalities. In this thesis, we aim to improve the capability of image analysis techniques using multi-modal images both at the macro scale (using thermal and visible images of whole plant) and at the micro scale (analysis of human tissue using fluorescence microscopy 63× resolution) to analyse abnormal regions.

1.3.1. Thermal and Visible Light Imagery for Stress and Disease Detection in Plants

Thermal imaging has been used in the past for stress and disease detection in plants. One of the major problems associated with thermal imaging in plants is temperature variation due to canopy architecture, leaf angles, sunlit and shaded regions, environmental conditions and the depth (distance) of plant regions from the camera. We aim to combine information of stereo visible light images with thermal images

to overcome these problems and allow a precise 3-dimensional thermal profile of a crop to be quantified.

A moving stereo visible and thermal imaging setup (Figure 1-6) can then be installed on a rig over the crop (Figure 1-7) which can scan whole crops and automatically identify regions of crop under water stress or regions where a disease starts to develop. This can replace manual scanning which requires a lot of manual labour and usually involves scanning randomly selected small patches instead of whole crop scanning. The camera system can also be developed to scan individual plants on a moving conveyer belt.

Our aim is to investigate in a laboratory setting if combination of stereo visible and thermal can be used to enhance the capabilities of thermal images to identify water stressed or diseased regions in plants and canopies.

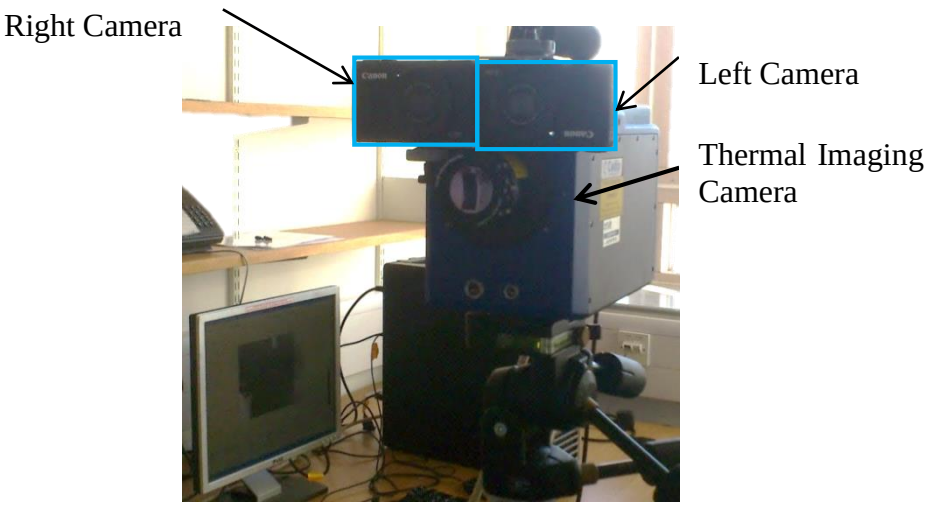


Figure 1-6: A thermal camera with stereo visible imaging setup can be installed above the crop for detecting anomalous regions.

Figure 1-7: Photo taken at DoubleH nurseries. The rigs above the crop provide a suitable place to mount camera setups for automatic detection of anomalous regions.

1.3.2. Alignment of Multi-Channel Fluorescence Microscopy Images

As outlined above in Section 1.2, the TIS image acquisition cycle produces images as a pair of fluorescence and phase contrast images for each anti-body tag. The fluorescence images provide information about the relative expression level of respective tags in sub-cellular compartments and the phase contrast images contain information about the structure of the tissue specimen. No significant misalignment is observed between fluorescence and the corresponding phase contrast pair as both images are taken within a short time interval of about 2-4 seconds. However, misalignment is observed between phase contrast images (and fluorescence images) for different anti-body tags. To analyse multiple molecular components in biological samples, fluorescence images from multiple tags must be aligned in a way that the same pixel location corresponds to the same physical location in the tissue being imaged.

The overall aim is to determine the transformations (in terms of translational shifts) necessary to align phase contrast images and then apply these transformations to register fluorescence images, such that a) the images are well aligned and b) the total number of non-overlapping pixels is minimised.

1.4. Main Contributions

The main contributions of this thesis are as listed below:

- We present a novel silhouette extraction method for extracting the object silhouette in highly noisy thermal and visible light image data. We have also presented a registration algorithm based on the silhouette extraction method to align objects in multi-modal image data (Chapter 3). This registration algorithm can be further extended to multiple object multi-modal image registration as far as the object boundary is distinguishable [28].
- We present a method to estimate smooth disparity maps compared to state of the art methods in noisy images of diseased plants (Chapter 4).
- In Chapter 5, we have introduced a novel set of features, by combining thermal, depth and visible light image data, which can be used to classify images of plants infected with disease or under various water stress conditions [29].
- We propose a block-based registration algorithm in Chapter 6 which can be used to align multi-channel fluorescence microscopy images up to sub-pixel accuracy. In addition, we have mathematically derived an objective function which can accurately and efficiently select a reference image in a stack of

hundreds of multi-channel images to minimise information loss during registration process [16].

1.5. Thesis Organisation

Chapter 2 contains a brief review of current literature on registration, multi-modal image registration, registration of multi-tag fluorescence microscopy images, and existing methods on thermal/visible image registration. It also reviews existing literature on identification of stressed/diseased plants using thermal images.

In Chapter 3, we propose a multi-scale method to extract silhouettes of diseased plants in thermal and visible light images. The extracted silhouettes in turn can be used to register images from different modalities.

Chapter 4 proposes a method of rectification using extrinsic marker points in a stereo image pair. Different disparity estimation algorithms are then compared and tested in this chapter on stereo images of plants with a noisy background. A novel method is also proposed to estimate smooth disparity maps in the presence of noise.

Automatic detection of water stressed plants using thermal and visible light images is proposed in Chapter 5. The selection of features is an important step and a set of good features are proposed in this chapter for more accurate classification of spinach canopies. The chapter then introduces a new dimension to the work by adding depth information for classification of diseased tomato plants using stereo visible and thermal images.

An algorithm for registration of multi-tag bio-images using mutual information as a similarity measure is proposed in Chapter 6. It includes a method for identification

of the best reference image plane which can be used to reduce the amount of information loss during the registration process.

Chapter 7 concludes the thesis, discusses limitations of the work, possible application at both experimental and industrial scale and future directions.

Chapter 2
Literature Review

In this chapter, we review image registration techniques and focus on multi-channel fluorescence and thermal/visible image registration methods. We cover commonly used disparity estimation techniques to include depth information to our analysis for disease detection in plants. This chapter also reviews previous work on water stress and disease detection in plants using thermal imaging.

2.1. Registration

Image registration is the process of spatially aligning two images of a scene so that corresponding points assume the same coordinates [30]. For image registration, one image is referred to as a reference image (I_r) and a second image is referred to as floating (target or sensed) image (I_j). Image registration is a process of finding the transformation $\tau_{rj}(.)$ such that each pixel location on (I_r) and $\tau_{rj}(I_j)$ correspond to same physical point in the scene being imaged.

2.1.1. Classification of Image Registration Methods

Registration methods can be divided into many different groups in many different ways. Maintz and Viergever [2] have quite comprehensively classified image registration methods on the basis of nine different criteria each of which is then subdivided into sub-groups. The nine different criteria are dimensionality, nature of registration basis, user interaction, modalities involved, subject, object, nature of

transformation, domain of transformation, and optimisation method.

On the basis of dimensionality, there are two major groups, spatial dimensions only and time series with spatial dimensions. Time series often involve e.g., monitoring of tumour growth, post-operative healing and evaluation of drug effects [31]. Each of the two groups is divided into 2D-2D, 2D-3D and 3D-3D registration depending on the number of spatial dimensions involved. On the basis of nature of registration, there are three major groups: extrinsic, intrinsic and non-image based. Extrinsic registration methods introduce foreign bodies introduced into the image space for registration; these foreign bodies are usually visible and accurately detectable in all the relevant modalities [32]. Intrinsic registration methods use features or landmarks from the image of the subject or the scene and do not include any foreign body. Non-image registration methods usually rely on camera calibration procedures [33]. The registration methods are divided into three groups on the basis of user interaction as interactive, semi-automatic and automatic. Automatic registration methods are usually desired; however semi-automated methods are also used to speed-up the process and to increase accuracy. Moreover the registration methods can be divided into uni-modal and multi-modal methods depending on the number of modalities involved and into inter- and intra-subject registration on the basis of same or different subjects involved in the registration process.

Maintz and Viergever [2] have divided registration methods into four groups based on the nature and domain of transformation i.e., Rigid, Affine, Projective and Curved. Rigid registration allows rotation and translations only, whereas affine transformation has a constraint that it maps parallel lines onto parallel lines and projective transformation has a constraint that it maps lines onto lines. All of these

transformation types mentioned above can be represented in terms of a constant matrix. Curved transformations are generally polynomial transformations and cannot be represented using a constant matrix as they generally map lines onto polynomial curves. Some researchers treat subsets of affine and projective transformations i.e., scaling and perspective transformation separate from affine and projective transformations [34]. Based on the domain of transformation the transformation functions are divided into two groups. If the transformation applies to the entire image the transformation is termed as global transformation whereas local transformations are composite of at least two transformations determined on sub-images (see Figure 2-1). The drawback of local transformation is that it can introduce discontinuities and is therefore seldom used. Finally registration methods can be divided on the basis of optimisation function used to minimise the cost function [35,36]. Optimisation functions are not necessary and only some registration methods use optimisation functions for registration.

2.1.2. Registration Methodology

At a higher level than the classification of registration methods in [2], there are two main approaches to solve the registration problem: area based and feature based. The feature based methods first extract salient features from the images before establishing correspondence between the images, whereas the area based methods skip the feature extraction step. In both kinds of approaches, the following major steps are performed while performing registration: matching, mapping function estimation and transformation.

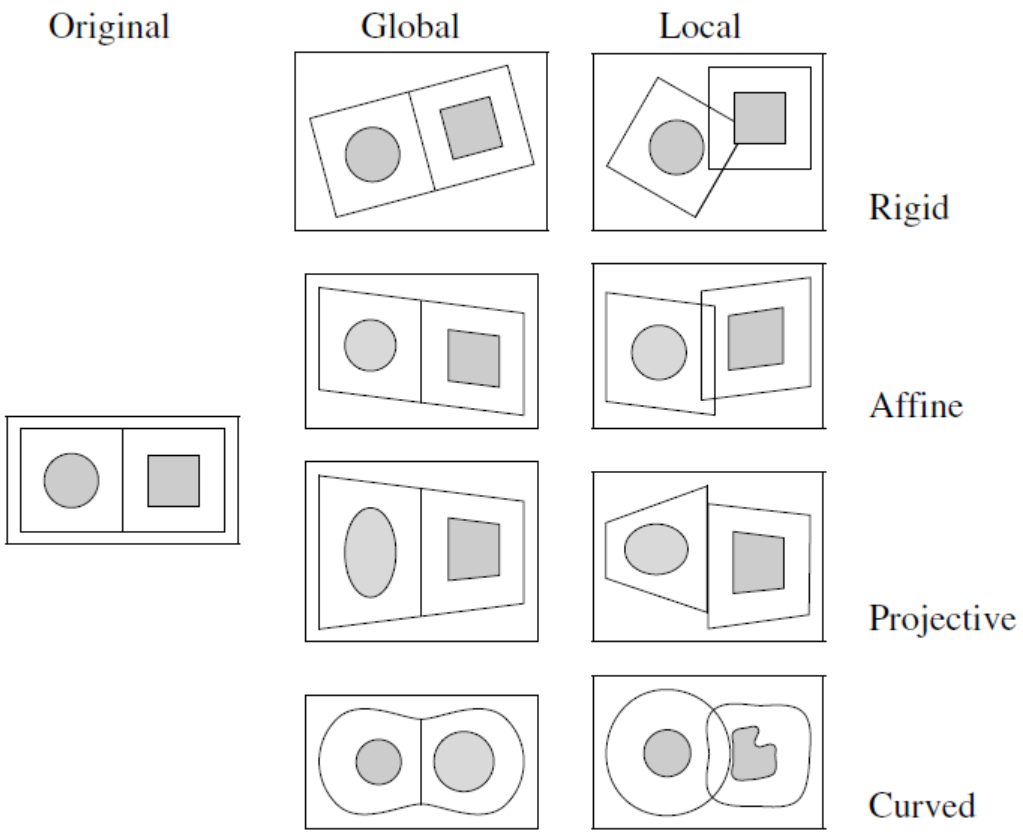


Figure 2-1: Examples of 2-D transformations. Global transformations apply to whole image whereas local are composite of at least two transformations on whole image. Image Credit: [2]

To establish correspondence between the images, the basic idea behind area based methods is to define a rectangular window in the reference image and search for the corresponding region in the floating image. The size of the rectangular window can be adjusted depending on speed and accuracy constraints, in some cases the size of the rectangular window can be of the size of the whole image. The next step is to choose a transformation function based on the complexity of the problem. The rectangular window in the floating image is transformed using a set of parameters in the transformation function and compared with a rectangular region in the reference image using a similarity or dissimilarity measure. The similarity measure acts as a cost function and the algorithm uses optimisation functions to search for

optimal transformation parameters. The simplest of similarity measures are l_1-norm and l_2-norm but normalized cross correlation, mutual information and phase correlation in the Fourier domain are regularly used for area based registration methods [16,37–40]. For a detailed list of similarity measures, we refer the reader to [30].

Feature based methods rely on the extraction of features before matching regions. These features can be extracted in the form of an edge using first order Prewitt, Sobel, Canny edge detectors, second order Laplacian, Marr-Hildreth operator, Laplacian of Gaussian, difference of Gaussian operators or phase congruency in the Fourier domain [41,42]. Another form of feature extraction method involves detection of lines, circles and shapes e.g., using the Hough transform [43] (see Chapter 4). However, the most common method for feature extraction is to extract localised features as it allows establishing point to point correspondences between images. Some of these popular methods to extract localised features are Harris corner detector, scale invariant feature transform and speeded up robust features [44–46].

Feature matching in feature based approaches is also different from area based methods. The feature based methods assume two sets of features in reference and floating image and find correspondence between them using spatial relations or feature descriptor values. Matching based on spatial relations uses geometric constraints to find correct correspondences. Spatial relations are more useful in extrinsic registration methods as foreign bodies are fixed in a certain pattern and their location can be accurately determined. But matching based on spatial relations can also be used in other registration techniques if the features can be accurately

determined and repeatedly detected in both reference and floating image. Chamfer matching and Iterative Closest Point Algorithm (ICP) are well known methods for feature matching using spatial relations [47–49]. It is not always easy to find a set of features in both the reference and the floating image in a certain pattern especially if the images have strong lens distortion. In this case, registration algorithms use features whose descriptors are invariant, unique, stable and independent [50]. Matching methods such as random sample and consensus (RANSAC) and Graph-based matching can be used for feature matching based on both spatial relations and invariant descriptors [30]. After feature matching, correspondences between the points in the reference and the floating image can easily be established. Based on these correspondences, mapping function can be estimated depending on the transformation applicable to the current problem. The last step of registration is the same in both area and feature based registration methods. After estimation of transformation parameters, the floating image is transformed according to the mapping function and matched with the reference image or ground truth to estimate accuracy and errors.

2.1.3. Multi-modal Image Registration

A major advantage of using multi-modal images for analysis of a subject is that they carry different types of information. However, this advantage also poses a major challenge when multi-modal image registration is performed. Multi-modal image registration is required when all the available imaging modalities are restricted to gather only limited information and there is a need for more information about the scene. The problem of multi-modal image registration has

also been associated in the literature with the inter-subject registration; see for example [51–53].

For multi-modal image registration, an algorithm which is invariant to radiometric changes, scale, occlusions and viewpoint is required for registration. In this type of registration, the similarity measure is of considerable importance. Skerl *et al*., [54] have proposed an evaluation protocol for a thorough, optimisation independent and systematic evaluation of similarity measures in multi-modal registration. Their results showed that the performance of similarity measure varies depending on the type of the multi-modal image pair used for registration. Area based mutual information is widely used and has become a standard measure of similarity between corresponding regions in multi-modal medical imaging applications. However, mutual information has its limitations and is often combined with feature based methods to increase accuracy [30,35,36,50,55]. Multi-modal image registration is a vast topic of study which can be applied to various fields. However, as our goal is to analyse thermal/visible images of abnormal plants and multi-channel fluorescence microscopy images of tissues, in this chapter we review in detail the approaches used for thermal/visible (section 2.1.3.1) and multi-channel fluorescence (section 2.1.3.2) image registration.

2.1.3.1. Thermal/Visible Image Registration

A major focus in this thesis is on joint analysis of thermal and visible images of water stressed, diseased, and normal plants. As a pre-processing step before joint analysis, thermal and visible light images of plants must be aligned so that the same pixel locations in both images corresponds to the same physical locations in the

plant. Thermal and visible light images are usually captured using different type of sensors from different viewpoints and with different resolutions. Capturing information from different viewpoints may introduce misalignments in pixel locations of the same physical point of an object in images from different modalities.

To the best of our knowledge, there is no existing literature on automatic registration of thermal and visible light images of *diseased plants*. However, registration of infrared thermal and visible light images has been previously employed in video surveillance e.g., traffic, airport security, detection of concealed weapons, smoke detection and patient monitoring [10,56–58]. Registration of thermal and visible images of diseased plants is a challenging problem due to the fact that texture and edge information is often missing in the corresponding visible/thermal images. One approach for registration is to calibrate the stereo visual + thermal imaging camera setup and use transformations to align the resulting images [59–61]. A disadvantage of this approach is that the calibration parameters of the cameras may not be readily available. In such cases, a possible solution is to align the thermal and visible light images using exclusively image based information. Han *et al*., [62,63] proposed a method to register thermal/visible images of man-made environments by extracting geometrical structures using line extraction and aligning features extracted from these geometrical structures. Kim *et al*., [64] and Lee *et al*., [65] proposed registration methods that match magnitude and orientation of edge and gradient information between thermal and visible images of scenes with strong edge and gradient information. Jarc *et al*., [66] proposed a registration method based on texture features. However, the method is

not automatic and requires manual selection of features. A method based on dual-tree complex wavelet transform for registration of satellite images of thermal and visible images was proposed in [67]. The method uses a pyramidal, coarse to fine resolution, approach for registration and use dual-tree complex wavelet transform to extract line, edge information. The edge information is then used at each level to register images using mutual information as the similarity measure. De Vylder *et al.*, [68] have used the method proposed in [67] for registration of multispectral images of plants. In general, line, edge and corner based methods are reliable for images of man-made environments. However, they perform poorly on images of natural objects. Other methods based on mutual information and cross correlation of image patches rely on texture similarities between the two kinds of images [61,65,69]. Since there is a high probability that texture information may be missing in the corresponding visible/thermal image(s) of diseased plants, methods based on mutual information and cross-correlation may not be a good choice for registration.

Region-based methods, such as those based on silhouette extraction, usually provide more reliable correspondence between visible and thermal images than feature based methods [57,69,70]. Bilodeau *et al.*, [69] proposed registering thermal and visible images of people by extracting features from human silhouettes; however, the authors propose no specific method for silhouette extraction. Torabi *et al.*, [71] suggested a RANSAC trajectory-to-trajectory matching based registration method which maximises human silhouette overlap in video sequences. Han and Bhanu [58] proposed a hierarchical genetic algorithm (HGA) for silhouette extraction using an automatic registration method for human movement detection. The algorithm first extracts a preliminary human silhouette to match regions in

thermal and visible images. It then uses the HGA to register thermal and visible images. The authors improve the accuracy of the extracted human silhouette by combining silhouette and thermal/colour information from coarsely registered thermal and visible images. Human body temperature is generally higher than that of the background region and this characteristic has been used in [57,58,70] to extract human silhouettes. However, in the case of thermal images of diseased plants, the temperature profile does not exhibit this characteristic. It is possible that within the same plant, the temperature of different regions is higher or lower than that of the background. Another common method for silhouette extraction in video sequences is background subtraction. This method usually provides very good results because of the high frame rate of the sequences and due to the fact that the background between two consecutive frames is usually very similar. For the case of images of diseased plants, background subtraction is not a suitable strategy due to the limited number of consecutive still images and the fact that there may be a large interval between two consecutive still images. In Chapter 3 of this thesis, we propose an algorithm which features a novel multi-scale method for silhouette extraction of plants in thermal and visible light images, and uses the extracted silhouettes to register thermal and visible light image pairs of plants.

2.1.3.2. Multi-Tag Fluorescence Image Registration

Bioimage informatics is a recently established branch of bioinformatics research that has emerged in response to two major demands: increasing deployment of powerful new technologies for measuring molecular components (including genomics, transcriptomics, proteomics, metabolomics) and new biological knowledge (from the human genome project amongst others). Bioimage

informatics is concerned with the processing, analysis, and management of images recorded for biological specimens mostly using microscopy techniques [72–74]. The ultimate objective is to localise molecular components in biological samples (ranging from cell cultures to tissue sections) in order to overcome one of the most important limitations of most traditional destructive 'omics' technologies, in which molecular phenotype is acquired at the expense of anatomical and cellular spatial information [75–77]. A comprehensive molecular profiling of cellular and sub-cellular regions in a tissue is necessary to better understand a disease which is directly affected by the protein profile of a cell such as cancer. Analysis of protein co-localisation becomes viable because molecules can interact to perform certain functions only when they are located close to each other [24]. By using multiple optical filters and fluorescent dyes, fluorescence microscopy enables multiplexed quantitative analysis in cell and tissue specimen. However, its use is limited to a maximum of ten fluorescent dyes because of the limitations of the technology being used to capture the images [17]. Among other limitations, auto-fluorescence is a major problem in detection of molecules in tissues. There are emerging new technologies which can help to overcome these limitations and image several protein structures at subcellular level using multiple fluorescent dyes. One such high-tech microscope, known as Toponome Imaging System (TIS) [18,29], is installed at the University of Warwick.

As discussed above in Chapter 1, TIS uses N sequential cycles of fluorescent tagging and bleaching to produce a data set consisting of $2N$ pairs of fluorescent and phase contrast image pairs (Figure 1-5). There is a dearth of literature on registration algorithms for such image data, primarily because imaging systems for

such type of data have emerged only recently, although several researchers have proposed techniques for solving the somewhat related problem of automatic tracking of live cells by registering time consecutive frames (see for instance [78–80]). Wang [81] proposed the Multiplex fluorescence *in situ* hybridization (M-FISH) algorithm for registration of multi-channel images in the context of cancer diagnosis and research on genetic disorders. The algorithm searches for a transformation using mutual information to register the misaligned multi-channel FISH images. For a multi-channel image data, the selection of a suitable reference image (plane) is very important. An inappropriate selection of a reference image among the set of images may result in large transformations for the rest of images in the data set, resulting in large amount of information loss. The authors [81] selected DAPI as a reference image and did not address the problem of choosing the reference image to minimise information loss. In Chapter 3, we present a method to select a reference image to minimise this information loss. Kim *et al*., [82] have proposed registering multi-modal images using a three-step procedure: 1) Gaussian filtering, 2) rigid registration and 3) non-rigid registration. For rigid registration, the authors minimised the mean-squared intensity error and for non-rigid registration, a variant of the Demons' algorithm was used [83]. They also presented two approaches for selecting the reference image. The first approach uses the first image in time as the reference image and all the images are registered to this reference frame. The second approach uses information from previous time steps in an incremental scheme. Can *et al.*, [22] have used a mutual information based measure to register images from different histological imaging modalities. They mapped the multi-modal fluorescent images of the same tissue stained with

molecular biomarkers to the co-ordinate system of Hematoxylin-Eosin (H&E).

A new high-tech method has recently been developed by General Electric (GE) Global Research, known as the multiplexed fluorescence microscopy method (MxIF) that enables high-level multiplexing of protein and nucleic acid detection in a Formalin-fixed paraffin-embedded (FFPE) tissue similar to the TIS system [24]. MxIF data acquisition cycle stains the slide using fluorescent dyes. After staining the slide, the microscope uses highly efficient fluorochrome specific filter sets to capture multiplexed images of the tissue. The tissue sample is excited using fluorescent light and after a fixed exposure time the images are captured. The process continues in a cyclic fashion for a batch of fluorescent dyes. Among the images of different fluorescent dyes, the DAPI stain is used to capture the position of the nuclei in each cycle. The position of the nuclei in each image is then used to align fluorescence images from multiple cycles of the data acquisition run. Their alignment method consists of two steps: the first step performs a global translation estimation using normalised correlation in Fourier domain and the second step involves rotation using normalised mutual information as metric. To the best of our knowledge, the GE setup does not address the need for estimation of a reference image with maximal overlap to minimise information loss and use the first image in the imaging cycle as the reference image.

In this thesis, we specifically address the problem of multi-tag fluorescent microscopy image registration where multiple phase-and-fluorescence images of the same sample stained by different biological tags are obtained and registered. We also address the problem of defining a reference image with maximal overlap in multi-tag images.

2.2. Stress and Disease Detection in Plants using Thermal Imaging

Infrared thermometers have been used since the early 80's to determine the temperature differences in plants and different parts of the canopy by researchers for irrigation scheduling purposes. However, the development of thermal imaging cameras has extended the opportunities for more detailed and sensitive analysis of the thermal properties of plants and canopies. This has led to the development of different applications including early detection of water stress, plant disease and plant phenotyping. A recent review on advances in sensing plant diseases for crop protection suggests that "*A multi-disciplinary approach is essential to tap the full potential of these highly sophisticated, innovative technologies and high dimensional, complex data for precision crop protection. Using the expertise from phytopathology, geography, computer science and data mining, a new understanding of crop protection will be formed*" [14]. We aim to use our expertise in computer science for crop protection from disease and unwanted stress. Mahlein *et al.* [14] have reviewed the use of thermography, chlorophyll fluorescence, and hyperspectral sensors for disease detection in plants. For a more detailed review on the use of thermography for water stress, disease detection, genotyping and phenotyping, we refer the reader to [84]. This thesis focuses on the use of thermography techniques in the thermal infrared region for detection of stress and disease in plants.

2.2.1. Water Stress Detection in Plants

There has been a lot of work focused on water stress analysis of plants using thermal

imaging. However, few researchers have exploited information from the visible light and infrared thermal images simultaneously for analysis. Most of the work conducted uses stress indices formulated by Idso and Jackson [85,86] and researchers have conducted various experiments to investigate the relationship between different stress indices and temperature values determined by thermal imaging [87–89]. The use of thermal imaging as an indicator of plant stress has also been tested in a number of environmental conditions and the conditions best suited to its successful application have been explored. Jones [90] formulated a leaf energy balance equation to estimate stomatal conductance, which was dependent on a range of environmental factors and plant variables such as emissivity of the leaf surface, air density, and specific heat capacity. The complexity and associated difficulty in the process of accurately measuring these variables, made it difficult to obtain accurate estimates of stomatal conductance from leaf temperature. Consequently, Jones [7,91] used the 'Crop Water Stress Index' (CWSI) by Idso and Jackson and rearranged the leaf energy balance equation to derive thermal indices based on 'wet' and 'dry' reference surfaces, thus making stomatal conductance more straightforward to calculate from leaf temperatures. There is debate within the scientific community as to the ideal choice of reference surface and much work has been undertaken to find the best choice for reference surfaces and in what conditions they must be used [92]. Another such index "Water Deficit Index" (WDI) was introduced in [93] but it was found to be very close to CWSI [84,89].

The effect of sunlight and shade on thermal imagery has also been investigated and various options have been suggested to minimise the effect of these conditions. Grant *et al.*, [94] have analysed the robustness, sensitivity and limitations of thermal

imaging for detecting changes in stomatal conductance and leaf water status in plants. In a later study, Grant *et al.*, [95] suggested that the average temperature of the canopy was more useful to reduce the effect of leaf angles and other environmental factors when compared to individual leaf temperatures. Jones [96] compared techniques for image acquisition and performed experiments to investigate the potential of infrared thermography for irrigation scheduling and to evaluate the consistency and repeatability of measurements under a range of environmental conditions. They suggested excluding pixels which are outside the wet-dry threshold range to allow for semi-automated analysis of a large area of canopy. The authors suggested using thermal data from shaded leaves for improved data consistency, since there is less variability in temperature within an image (Figure 2-2), and smaller errors resulting from differences in radiation absorbed by reference and transpiring shaded leaves. Coefficient of variation of stress indices was found to be significant and discriminatory powers of the techniques for estimates of stomatal conductance were found to be limited. Stoll and Jones [9] found that sunlit leaves showed a wider range of temperatures, the authors believed that this is due to the fact that natural leaf orientation has little effect on the energy balance of shaded leaves compared to exposed leaves. Based on these observations, the information from temperature distribution can be combined with the leaf orientation for thermal analysis in high resolution images.

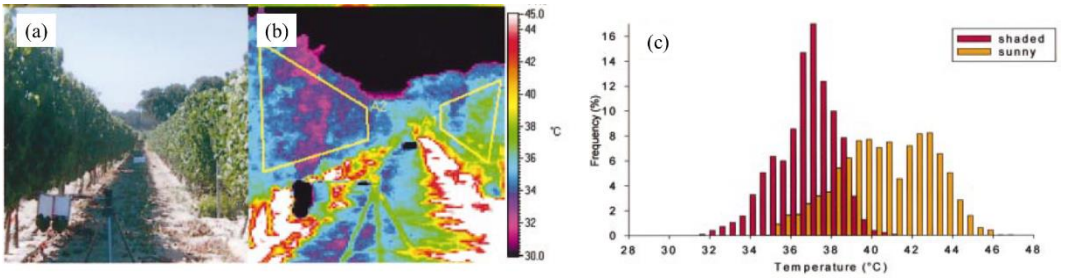


Figure 2-2: Corresponding visible light and thermal images (a) and (b) showing the shaded and sunlit sides, (c) temperature distributions for areas of the canopy outlined in yellow. Shaded region shows less variability ($\sigma = 0.93°C$) compared to sunlit region ($\sigma = 1.81°C$) as can be observed from the histogram. Image Credit: Jones *et al*. [96].

Combining information from thermal and visible light images has the potential for providing a better estimate of the stress indices and to identify stress regions in the canopy. Möller *et al.*, [97] studied the use of thermal and visible imaging to maintain mild to moderate water stress levels in grapevine. To estimate the canopy temperature, different sections of the canopy were used in this study, including: the whole canopy, the regions of canopy under the sunlight, the centre of the canopy and leaves from the centre of the canopy under the sunlight. The best correlation between CWSI and stomatal conductance was calculated from the centre of the canopy measurements (or its sunlit fraction). The authors observed that CWSI computed with wet and dry reference was the most robust index and suggested that the fusion of thermal and visible imaging can not only improve the accuracy of remote CWSI determination but also provide precise data on water status and stomatal conductance of grapevine.

Leinonen and Jones [98] introduced a partly automated method to study plant stress indices. They exploited colour information from visible light images to identify leaf area, as well as sunlit and shaded parts of the canopy. As an initial pre-processing step, images of constant temperature background were subtracted from the actual

image to correct for relative errors in calibration of the camera caused by internal warming of the camera. Ground Control Points (GCPs) were manually selected to overlay the thermal image on the visible light image. Different regions in the visible light images were classified using a supervised classification method into pixels which represent plant leaves, other parts of the plant and background. Statistical parameters and stress indices were calculated based on temperature values from the corresponding classified regions of the plant. The results showed that temperature distribution can be used as an indicator of stomatal conductance and plant stress. More recently, Cohen *et al.*, [99] have used automated methods to estimate water status using aerial thermal images of palm tree canopies. The authors used watershed segmentation of thermal images to detect the palm trees, and found the detected temperature to be a good indicator of the tree's water stress.

In this thesis, we aim to use combined information from thermal and visible light images of a spinach canopy to efficiently discriminate between plants undergoing water stress (deficiency) and well-watered plants. We present a new technique to enhance the 'discriminatory power' of thermal imaging to identify parts of the canopy which have lowered leaf water potential (i.e., they are under water stress). Instead of using stress indices to identify stress regions, we combine information from visible light and thermal images and use machine learning techniques to classify between water stressed and well-watered canopies. Furthermore, we have acquired information about the light intensity and green-ness of the plant, using pixel values in the RGB and Lab colour space from the visible light images. These data are subsequently used, along with statistical information from thermal images, to classify between stress and well-watered plants using Support Vector Machines

(SVM), Gaussian Processes Classifier (GPC) and a combination of both the classifiers.

2.2.2. Disease Detection in Plants

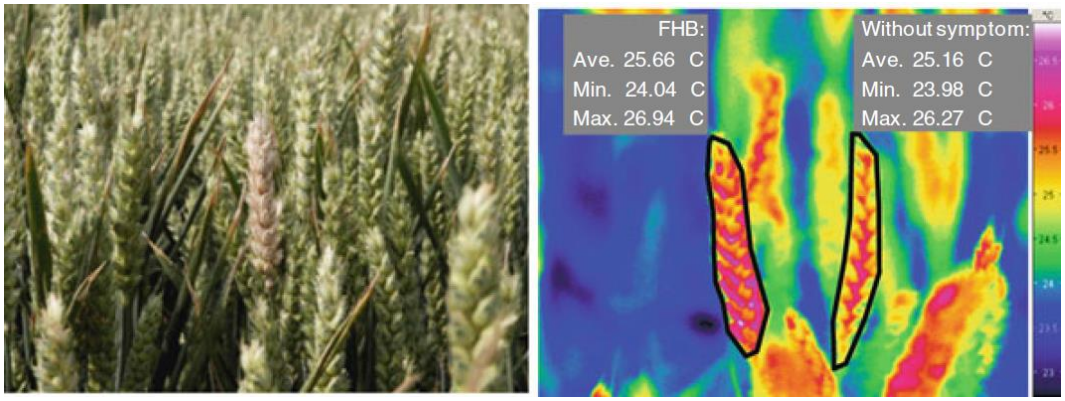
Thermal imaging has potential for early detection of disease, especially when the disease directly affects transpiration rate in plants. Early detection of disease is very important to control the spread of disease since late detection may result in reducing the quantity and quality of the crop yield [100].

Chaerle *et al.*, [101] studied the resistant tobacco plant infected with tobacco mosaic virus (TMV) and detected that the infected sites were 0.3-0.4°C warmer than the surrounding tissue hours before the initial appearance of the necrotic lesions. They used a localised-infection method and a high-resolution infrared camera to detect temperature increase at the site of inoculation. They observed a correlation between leaf temperature and transpiration by thermography and steady-state porometry. In a later study, Chaerle *et al.*, [102] studied the propagating cell death in bacterio-opsin transgenic tobacco plants. They found that the cell death was trailed by a coherent front of higher temperature. Lower temperature was observed at regions with visible cell death because of water evaporation from the damaged cells. It was observed that the stomatal closure preceded the tissue collapse. They obtained high resolution thermographic images by capturing several slightly overlapping images and then joining them. The sub-images were visualised within a temperature window of 1 °C, to maximise the temperature contrast. Chaerle *et al.*, [103] have studied the use of thermal and chlorophyll fluorescence imaging in pre-symptomatic responses for diagnosis of different diseases and to predict plant

growth. Both of these imaging techniques can be applied for detecting and diagnosing plant stresses. The authors concluded that conventional methods are time consuming and suitable for small number of plants, whereas imaging techniques can be used to screen large number of plants for biotic and abiotic stress and to predict the crop growth.

Oerke *et al.*, [104] studied the changes in metabolic processes and transpiration rate within cucumber leaves caused by pathogenesis of *Pseudoperonospora cubensis*. Under controlled conditions, a linear relation was found between transpiration rate and leaf temperature. They showed that healthy and infected leaves can be discriminated even before the visible symptoms of the downy mildew (caused by *Pseudoperonospora cubensis*) appear. The maximum temperature difference (MTD) was found to be related to the severity of infection and could be used for the discrimination of healthy leaves or canopies and those with downy mildew [105]. Conditions enhancing transpiration rate improved the detection of these changes at an early stage of infection. In another work, Oerke *et al.*, [106] studied the effect of *Venturia inaequalis* fungus on apple leaves and they found MTD to be strongly correlated to the size of infection sites.

Stoll *et al.*, [107] have investigated the use of infrared thermography to study the attack of *Plasmopara viticola* in grape vine leaves under varying water status conditions. They applied and studied different irrigation treatments for both inoculated and non-inoculated vines. A study on wheat canopies for detection of fungal diseases revealed that higher temperature was observed for ears (the part of wheat containing the fruit) infected with *Fusarium* (Figure 2-3) [108,109].



(a) (b)

Figure 2-3: Detection of *Fusarium Head Blight* (FHB) infected ears in wheat. (a) Colour image with an ear infected with FHB in pink shade; (b) effect of infection on absolute temperatures of ears. Image Credit: Oerke *et al*. [108].

In this thesis, we study the effect of a fungus *Oidium neolycopersici*, which causes powdery mildew, in tomato plants using thermal imaging and we investigate if combining stereo visible imaging with thermal imaging increases our ability to detect the disease before appearance of visible symptoms.

2.3. Stereo Vision for Depth Estimation

To add depth information for analysis of diseased plants, we use disparity estimation using stereo pair of images. In the literature, stereo vision is a vast subject and comprises of broad topics such as epipolar geometry, camera calibration, image rectification, disparity estimation, 3D reconstruction and occlusion detection [110]. A model of basic stereo vision setup consists usually of two identical cameras with overlapping field of view and parallel optical and vertical axis (Figure 2-4).

Let O_l and O_r be the centres of camera lens for the left and right cameras, respectively. The left and right camera project a point P to pixel location (u_l, v_l)

and (u_r, v_r) respectively. Since the cameras are identical and optical and vertical axes are parallel, the only difference between the location of projection of point *P* in the left and right images is along the horizontal axis. This difference in location of an object in left and right image plane is called disparity. If the stereo setup has vertical disparity among the image pair then image rectification is performed. One method to perform image rectification is by calibrating the stereo setup and then applying transformations to the image pair. In the absence of calibration parameters, rectification can be performed by using feature matching and estimating the fundamental matrix, which relates the point in a 3D-space to the point(s) in images captured from different angles. However, if the two cameras can be approximated by affine transformation, then the image pair can be rectified by affine transformation [110].

Let us assume that the point *P* is located at location (X, Y, Z) with O_l as origin of a Cartesian co-ordinate system, then the same point will be located at $(X - B, Y, Z)$ with O_r as origin in the camera co-ordinates. The pixel location (u_l, v_l) and (u_r, v_r) of point *P* can be estimated by using the pinhole camera model [111].

$$(u_l, v_l) = \left(f\frac{X}{Z}, f\frac{Y}{Z} \right), (u_r, v_r) = \left(f\frac{X-B}{Z}, f\frac{Y}{Z} \right)$$

where *f* is the focal length of camera(s) and *Z* is the distance(depth) of the point from the camera, disparity (*d*) can then be estimated as:

$$\begin{aligned} d &= u_l - u_r = \frac{fB}{Z} \\ Z &= \frac{fB}{d} \end{aligned} \tag{2.1}$$

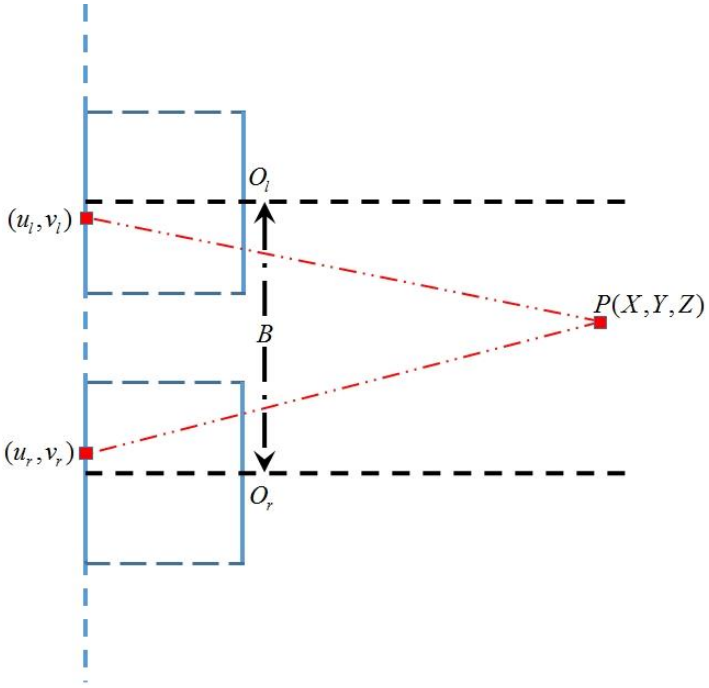


Figure 2-4 Model of a basic stereo vision setup.

Since focal length and baseline are constant for a stereo setup with known values of disparity, the depth of a point can be estimated using equation (2.1).

2.3.1. Disparity Estimation

The work done by Scharstein and Szeliski [112] is considered to be a landmark in the field of disparity estimation algorithms. Although their work is now more than 10 years old, the benchmarks presented by Scharstein and Szeliski have helped to improve the quality and quantity of work done on disparity estimation. They reviewed and compared most of the stereo correspondence algorithms at the time of publication. In addition, they produced multi-frame stereo data sets and made them publicly available along with the ground truth[5]. The authors presented a test

[5] http://vision.middlebury.edu/stereo/data/

bed to evaluate the performance of disparity estimation algorithms and the evaluation mechanism is still being used by researchers to evaluate the performance of their algorithms[6]. According to the authors, in general stereo algorithms perform four steps to estimate disparity i.e., matching cost computation, cost aggregation, disparity computation and disparity refinement. The sequence of these steps depends on disparity estimation algorithms which can be broadly divided into two types of methods: *local* and *global*. Local methods define a finite window at each point in the image and disparity is estimated depending on intensity values within this finite window. Matching cost is computed within the window and cost is usually aggregated over the whole window. Global methods, on the other hand, define a global energy function and minimise the energy function using optimisation algorithms. There is another class of algorithms which do not define a global function and use image pyramid in a hierarchical, coarse to fine, approach. The pyramid approach estimates a coarse resolution disparity image at the top of pyramid and then refines the disparity to original resolution at the pyramid base. Some authors have also introduced a feature based approach but feature based methods produce sparse disparity results if corresponding features are not reliably found in the image pair [113,114]. The amount of literature on disparity estimation algorithms is quite large and we will cover only commonly used disparity estimation schemes in this thesis.

A window based approach presented by Konolige [115] uses Laplacian of Gaussian (LOG) to enhance the images for correlation and then uses l_1-norm or sum of

[6] http://vision.middlebury.edu/stereo/eval/

absolute differences (SAD) as the cost function to search for correspondence along epipolar lines. As the disparity refinement step, the algorithm uses left/right check and confidence measure. The left/right check checks if disparity estimate is the same when the right image is fixed compared to when the left image is fixed. The confidence measure gives higher confidence to textured regions than to low textured regions. The window based algorithms suffer especially in the occluded regions, as it is not easy to find the matching patch in the corresponding image. The choice of window size and similarity measure also affects the result of window based algorithms.

Birchfield and Tomasi (BT) [116] presented a pixel dissimilarity measure using linearly interpolated intensity functions. The BT measure is easier to implement, is widely used in disparity estimation algorithms for cost computation, and has been shown to produce good results [112]. Birchfield and Tomasi [117] subsequently presented a pixel-to-pixel (p2p) approach based on the BT dissimilarity measure which can be used for disparity estimation and depth discontinuities. To formulate the cost function, the p2p approach adds a penalty for occlusion and a reward term for a match in addition to the dissimilarity measure. The algorithm adds hard constraints based on un-textured and occluded regions to facilitate possible matches and disallow certain unlikely mismatches. The algorithm uses a dynamic programming approach for cost function optimisation.

Kolmogorov *et al*. [118–120] presented an approach for energy minimisation via graph cuts. The authors defined a global energy function based on matching cost, smoothness and visibility and then reformulated the energy function to construct the graph and solve it using α-expansion. The α-expansion method assumes an

initial disparity α and then refines the disparity iteratively by searching for α which minimises the objective function. Later, the authors presented max-flow/min-cut algorithms as efficient global optimisation solutions. An iterative message passing approach commonly known as belief propagation approach was presented in [121–123]. This approach models the problem of stereo matching as a combination of three Markov random fields and proposes to find the optimal solution by using Bayesian belief propagation.

To overcome the limitation of window-based and global approaches, Hirschmüller [124,125] presented a semi-global approach for disparity estimation. The author approximated 2D energy function by solving 1D paths along eight/sixteen directions in the neighbourhood of each pixel. The method performed as well as global methods but the time complexity was reduced which was now linear to the number of pixels and disparities. They used a pixel-wise matching cost based on mutual information and proposed hierarchical calculation using image pyramid as hierarchical mutual information (HMI) [126]. In a subsequent study, Hirschmüller and Scharstein [127,128] compared cost functions with HMI in the presence of radiometric differences and found that HMI performed best for images with global radiometric differences but the results were not good for local radiometric differences. As a disparity refinement step, Hirschmüller [125] proposed “peaks removal” and “intensity consistent disparity checks”. Peak removal removes small patches of disparity which appear very different from their surrounding disparities. The next step of the algorithm in [125] then checks the consistency between intensity and disparity values, the underlying assumption being that discontinuity in disparity occurs when there is a change in intensity.

In recent years, many algorithms have been presented which use a combination of local and global approaches. We refer the reader to the Middlebury evaluation for a more detailed analysis of various different disparity estimation algorithms[7] [129–133].

Application of stereo vision in horticulture is not new and it has been used for plant quality assessment and phenotyping before. Ivanov *et al.*, [134] presented a feature based matching approach for disparity estimation in stereo images of plants but it was not fully automatic. Andersen *et al.*, [135] and Biskup *et al.*, [136] used area correlation combined with simulated annealing to estimate depth. Song *et al*., [137] presented a multi-resolution pyramid which uses Kalman filtering to update disparity results from one level to the next level. To increase the accuracy of 3D depth estimation, stereo vision has been combined by various researchers with Light Detection and Ranging (LIDAR) technology. Commonly used LIDAR technologies include time-of-flight and phase-shift LIDAR. For a detailed discussion on the use of stereo images with LIDAR technology, we refer the reader to [138–142].

Chapter Summary

In this chapter, we summarised existing techniques on registration of multi-modal images, in particular registration of multi-fluorescence and thermal/visible images. We also summarised existing literature on disease and stress detection in plants using thermal and visible images and depth estimation using disparity estimation

[7] http://vision.middlebury.edu/stereo/eval/

algorithms. The choice of registration method depends on the type of images and their application. Multi-fluorescence imaging technologies are relatively new and various techniques are being used for registration of multi-fluorescence images. However, in the limited multi-modal image registration literature, very few researchers have addressed the problem of selection of a suitable registration plane with maximal overlap in order to minimise information loss.

For stress and disease detection in plants, various researchers have used thermal imagery but few have simultaneously exploited information from visible images. An automatic method for disease and stress detection using combined visible/thermal imaging is non-existent to the best of our knowledge. Depth information has been used in the past for plant quality estimation and phenotyping. Some researchers have stated the effect of depth and leaf angles on thermal images of plants but the depth information has not been exploited for plant stress and disease detection. In Chapter 5, we will combine depth information with colour (visible light spectrum) and thermal images to analyse plant images undergoing stress and disease. In Chapter 6, we will focus on fast and accurate registration of multi-tag fluorescence images and address the problem of finding the reference plane with maximal overlap.

Chapter 3
Registration of Thermal and Visible Light Images

We propose here an algorithm for registration of thermal and visible light images of diseased plants. Let us consider the two main classes of registration algorithms discussed in Chapter 2 i.e., feature based methods and area based methods. Feature based registration methods require stable feature sets which can be used to match corresponding points in the reference and the floating image. However, in the case of thermal and visible light images of diseased plants it is difficult to find features which can be reliably detected in images from both the modalities. The major reason behind this problem is that the underlying temperature profile of a diseased leaf might be different from the visible light image. This can introduce unexpected features in one of the images that are missing in the other image. Therefore, we focus to area based registration methods in this Chapter.

Area based registration methods suffer from the same problem as the feature based methods if we directly use intensity values for registration, i.e., the intensity profile of the leaf in thermal and visible light image may be different. Therefore, it is difficult to find a cost function based on intensity values which can be minimised using optimisation methods for perfect registration. In section 3.4 we show that a cost function which directly uses intensity values perform poorly on these kind of images. An alternative method to overcome this problem is to use silhouettes of the objects of interest in thermal and visible light images for registration [57,70,143]. Various researchers have proposed silhouette based registration of thermal and

visible light in video sequences [58]. There are others who make use of the difference in temperature from the background for silhouette extraction [57,70]. In the case of diseased plants, the temperature profile might be very similar to the background in various portions of the plant and therefore silhouette extraction using background subtraction may not be a suitable choice.

The algorithm proposed in this chapter features a novel multi-scale method for silhouette extraction of diseased plants in thermal images. An overview of the proposed registration algorithm is shown in Figure 3-1. For the visible light image, the algorithm uses the strength of edges/gradient to detect and extract the silhouette whereas for the thermal image it uses a method based on the stationary wavelet transform (SWT). The latter follows a multi-scale approach that first estimates the silhouette at coarse scales by using the curvature strength as computed from the Hessian matrix of coefficients at each pixel location. It then uses these estimates to refine the silhouette at finer scales. After silhouette extraction, the algorithm employs a global + local registration method to register the thermal and visible light images.

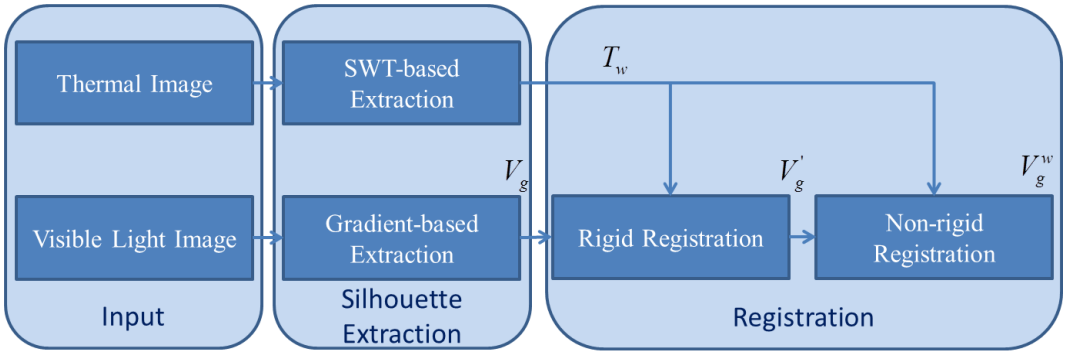


Figure 3-1: Overview of the proposed algorithm.

3.1. Image Acquisition

An experimental setup was designed and developed at the Department of Computer Science, University of Warwick, UK, to simultaneously acquire visual and thermal images of diseased/normal plants. The front view of the imaging setup is shown in Figure 1-6. The imaging setup consisted of two visible light imaging cameras **Canon Powershot S100**, and a thermal imaging camera **Cedip Titanium**. The setup was used to image tomato plants infected with the fungus *Oidium neolycopersici*, which causes powdery mildew disease. 10^6 conidia/ml and various control treatments were applied which resulted in different amounts of disease developments available for imaging. The disease symptoms consist of white powdery spots (first appearing after approx. 7 days) that expand over time and eventually cause chlorosis and leaf die-back. Thermal and visible light images of 71 plants under 10 different treatments were captured for 14 days (04 Oct 2012 to 17 Oct 2012) in a controlled environment at 20˚C.

3.2. Silhouette Extraction

3.2.1. Thermal Image

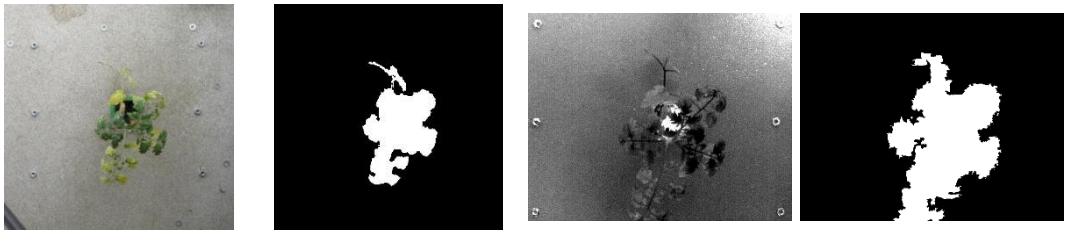
Extraction of plant silhouettes from thermal images obtained in our experiments is a difficult step because of high noise content. Since thermal images were obtained from diseased plants inoculated with powdery mildew, the intensity of the thermal profile changes within leaves. Figure 3-2 (c) shows a thermal image of a diseased plant enhanced by truncating the lower and upper 1% of pixel values and by contrast stretching where the thermal profile (i.e., intensity) of the background is very close to that of the leaves. Furthermore, the thermal profile of some of the leaves is

higher/lower than that of the background. Because of the presence of weak edges in the thermal image of the diseased plant, edge detection methods such as gradient, Canny edge detector, difference of Gaussian, Laplacian perform poorly on thermal images. Based on this observation, we propose an approach that is minimally affected by intensity changes within leaf.

It has been shown that the joint statistics of coefficients obtained after wavelet transformation (WT) show strong correlation among object boundaries in thermal and visible light images [144]. Thermal images, therefore, capture most of the object boundaries and thus WT can be used to extract silhouettes. Additionally, WT has shown to be very efficient in reducing noise, improving mutli-scale analysis and detecting edge direction information [145–147]. Multi-scale wavelet-based methods have also shown to be efficient in fusing thermal and visible light images [148,149]. In this work, we present a multi-scale wavelet-based method to extract plant silhouettes in thermal images. We use the stationary wavelet transform (SWT), which is similar to the discrete wavelet transform (DWT) except that it does not use down sampling. As a result, the resulting frequency sub-bands generated have the same size as the input image and contain coefficients that are redundant and correlated across different scales [150]. Our multi-scale SWT-based method uses the Haar filter to first decompose the thermal image into a number of sub-bands.

For an image of $m \times n$ pixels, it computes a matrix $\mathbf{H}_{i,j,s}$ equivalent to the matrix of second derivatives (Hessian) at each pixel location [145,151]:

$$\mathbf{H}_{i,j,s} = \begin{bmatrix} |V_{ijs}| & |D_{ijs}| \\ |D_{ijs}| & |H_{ijs}| \end{bmatrix} \tag{3.1}$$



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

Figure 3-2: (a) Example visible light image; (b) silhouette extracted from visible light image (V_g) using the gradient-based method; (c) corresponding thermal image (enhanced by truncating the upper and lower 1% pixel values and by contrast stretching); and (d) silhouette extracted from thermal image (T_w) using the SWT-based method.

where $H_{ijs}, V_{ijs}, D_{ijs}$ are the horizontal, vertical and diagonal coefficients, respectively, at scale s and pixel location (i, j); scale $s = 1$ corresponds to the first level of decomposition. Matrix $\mathbf{H}_{i,j,s}$ is then decomposed using Singular Value Decomposition (SVD). The largest singular value provides information about the direction of the highest curvature at pixel (i, j), also known as the curvature strength at pixel (i, j) [151]. We use the singular values of each Hessian matrix $\mathbf{H}_{i,j,s}$ to compute the $m \times n$ edge map, E_s, of the image at scale s as follows:

$$E_s(i, j) = \max(\lambda_{ijs}(1), \lambda_{ijs}(2)) \tag{3.2}$$

where $\lambda_{ijs}(1)$ and $\lambda_{ijs}(2)$ are the two singular values of $\mathbf{H}_{i,j,s}$ at pixel location (i, j) and scale s.

Due to the redundant properties of the SWT and the size of the wavelet filter used, edge information in E_s usually appears blurred. Furthermore, the position of edges tend to shift to the left and top from their actual position in the original image. This shift occurs within range Λ_s, as given by equation (3.3) and (3.4) [152,153]:

$$\Lambda_s = \sum_{\ell=1}^{s} \Delta_{\ell-1:\ell} \tag{3.3}$$

$$\Delta_{s-1:s} = 2^{(s-3)} \times (d_0^L + 3d_0^H - 4) \tag{3.4}$$

where d_0^H and d_0^L is the size of the high pass and low pass filters, respectively. For the Haar filter, $d_0^H = d_0^L = 2$. Figure 3-3 illustrates this *shifting* effect, where it can be observed that the round markers appear to have moved towards the left and top edge of the image as the scale increases. It is also important to note in Figure 3-3 that for low s values the boundary is thin but the noise content is too high. As s increases, weak edges become more prominent and the noise content decreases. Based on these observations, our method follows a multi-scale approach that first computes a binary image, $\mathbf{B}_s$, at scale s using thresholding and then refines it at the next lower scale $s-1$. Specifically, at scale $s-1$, $\mathbf{B}_s$ is multiplied point-wise with $\log(E_{s-1})$, i.e., $\mathbf{B}_s \circ \log(E_{s-1})$, is then thresholded to obtain $\mathbf{B}_{s-1}$, i.e., the binary image at scale $s-1$. This is done to discard any regions that may fall outside $\mathbf{B}_s$ before computing the binary image at scale $s-1$. Equation (3.4) establishes the correspondence between edges in two adjacent scales; namely, edges at scales *s* are $\Delta_{s-1:s}$ pixels thicker and appear shifted to the left and top by $\Delta_{s-1:s}/2$ pixels from the corresponding edge at scale $s-1$. Therefore, $\mathbf{B}_s$ is shifted by $\Delta_{s-1:s}/2$ pixels towards the right and bottom before computing $\mathbf{B}_{s-1}$ at scale $s-1$. Because a shift occurs at every scale, silhouette regions located at the top or left boundary of the image may be truncated. To avoid this, the method pads *r* rows and *c* columns to the top and left of the image using symmetric reflection of the first *r* rows and *c* columns of the original image before computing the wavelet coefficients, so that $r, c > \Delta_s/2$. Symmetric reflection also helps to avoid creating any extra edges at the boundary of the image.

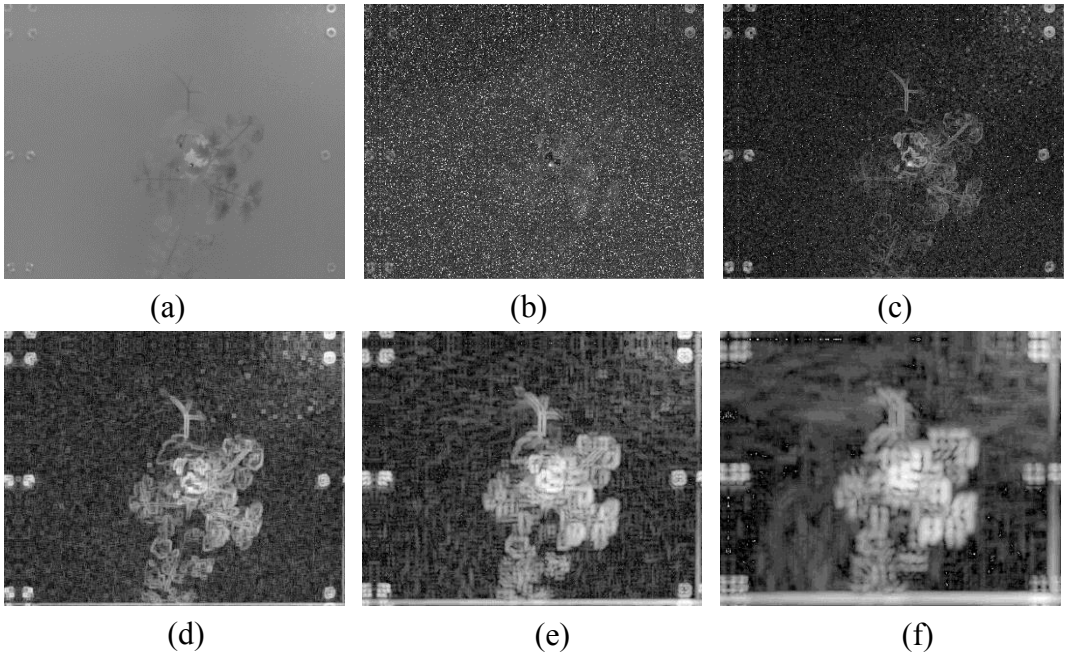


Figure 3-3: (a) Original thermal image, and (b)-(f) edge maps E_1-E_5. All images are enhanced by computing logarithm of the intensities. Artifacts due to symmetric reflection can be seen but do not affect the silhouette.

The flow chart of the SWT-based method for plant silhouette extraction in thermal images is shown in Figure 3-4. In this work, the method begins at scale $s = 5$, since it was found empirically that this scale provides the best trade-off between accuracy and computational complexity. Similarly, the method stops at scale $s = 3$ as this scale was empirically found to be a good trade-off between the amount of weak and blurred edges, which may affect the accuracy of the extracted silhouette (see Figure 3-3). We apply morphological operations on the binary image $\mathbf{B}_3$ to remove small objects. The last step is to discard the extra *r* rows from the top and the *c* columns from the left of the resulting binary image to obtain an *m*×*n* binary silhouette of the plant, denoted by T_w. Figure 3-2 (d) shows a sample binary image depicting the plant silhouette in the thermal image in Figure 3-2 (c) using the proposed SWT-based method.

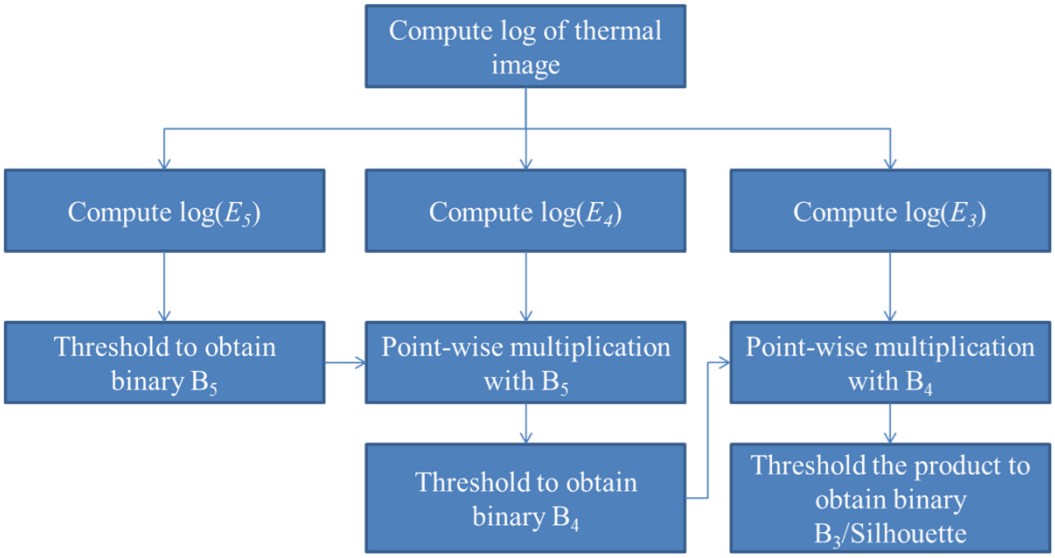


Figure 3-4: Flow chart of the SWT-based method for silhouette extraction in thermal images.

3.2.2. Visible Light Image

During the progress of the powdery mildew disease, some parts of the leaves change colour from green to yellow and then to white. It is therefore necessary to design a method that is robust to colour changes in leaves and that is capable of extracting plant silhouettes accurately from visible light images. To this end, we propose a gradient based-method that first converts the image from the RGB colour space to the Lab colour space in order to enhance the plant region by subtracting the 'a' from the 'b' channel. After colour space conversion, the method removes non-uniform illumination artifacts by subtracting the local mean [154]. It then removes noise by using anisotropic diffusion filtering, which helps to smooth the background noise while keeping the edges/boundary of the plant region sharp [155]. The method then detects edges by computing the gradient of the image using the Sobel operator. In order to enhance regions around high gradient values (plant boundary) and suppress low gradient values, the method performs a grayscale closing operation on the detected edges. The resulting image is thresholded to obtain a binary image

containing the silhouette. Morphological operations are then performed on the binary image to obtain the final plant silhouette denoted by V_g. Figure 3-2 (a) shows a sample visible light image of a diseased plant and Figure 3-2 (b) shows the corresponding plant silhouette as computed using our proposed gradient-based method. Note that the main motivation to use this method in visible light images, as opposed to the SWT-based method, is the low computationally complexity and good results. We further discuss this in Section 3.4.

3.3. Registration

The goal of registration is to align the thermal and visible light images in such a way that the same pixel locations in both the images correspond to the same physical location in the plant. Our particular registration method is a two-step process: global and local registration. For global transformation, a similarity transformation is parameterised by four degrees of freedom. A general similarity transformation matrix for a 2D image can be written as:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = S.\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{3.5}$$

where *S* is the scale factor, α is the angle of rotation along the z-axis, and t_x and t_y are the shifts in the *x* and *y* directions, respectively. The transformation in equation (3.5) maps a point (x_1, y_1) in a floating image to a corresponding point (x_2, y_2) in a static image. In our case, the binary image depicting the plant silhouette from the visible light image is the floating image and the binary image depicting the plant silhouette from the thermal image is the static image. The global registration step first finds the centroid of the plant silhouette in both the thermal and visible images.

It then calculates the difference between centroid locations and shifts the floating image by a number of pixels equal to this difference. It uses the sum of absolute differences as a cost function and an optimised pattern search algorithm [156] to search for the best approximation of similarity transformation between the two plant silhouettes. The search space range is chosen to be [0.9, 1.1] for scale factor *S*, $[-0.1^{\circ}, 0.1^{\circ}]$ for angle α, and $[-100, 100]$ pixels for translations (t_x, t_y). The resulting registered visible image silhouette obtained after applying similarity transformation is denoted by V_g'.

After global registration, the second step performs local registration using a free-form deformation (FFD) model based on multilevel cubic B-Spline approximation proposed by Rueckert *et al*. [35,36,157]. FFD models deform an object by manipulating an underlying mesh of control points Φ. For an image of *m*×*n* pixels, let $\Omega = \{(x, y) | 0 \le x < m, 0 \le y < n\}$ be the image domain on the *xy*-plane, and ϕ_{ij} be the value of the *ij*-th control point on lattice Φ represented by a $n_x \times n_y$ mesh with uniform spacing δ. The FFD approximation function can then be written as

$$\mathbf{T}(x, y) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(t) B_l(u) \phi_{i+k, j+l} \tag{3.6}$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $t = x/n_x - \lfloor x/n_x \rfloor$, $u = y/n_y - \lfloor y/n_y \rfloor$ and B_k and B_l represent cubic B-spline basis functions. This second step uses the hierarchical multi-level B-spline approximation proposed in [157] and an implementation of the limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) by Dirk-Jan Kroon as the optimisation function [36]. The similarity measure used here is the Sum of Squared Differences (SSD):

$$C_{similarity}(T_w, \tau(V_g)) = \sum_{x,y \in \Omega} (T_w(x,y) - \tau(V_g(x,y)))^2 \quad (3.7)$$

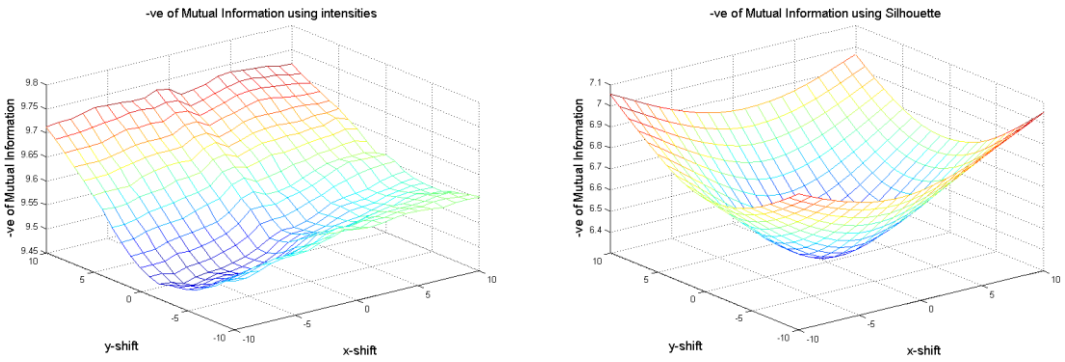
where $\tau(V_g)$ is the combined similarity transformation (equation (3.5)) plus local transformation $\mathbf{T}(x,y)$ (equation (3.6)) applied to V_g (the visible image silhouette) and T_w is the silhouette of the thermal image.

3.4. Results and Discussions

In this section, we first show that registration of thermal and visible images of diseased plants using silhouette extraction performs better than registration using exclusively intensity values (see Figure 3-5). To this end, we computed the mutual information of a pair of registered thermal and visible light images. Mutual information is a similarity metric commonly used for registration of multi-modal images [38]. We first converted the visible light image to grayscale image and then computed the mutual information of this grayscale image and the thermal image for various translation values in the *x* and *y* directions, as shown in Figure 3-5 (a). Ideally, local minima for the cost function (negative of mutual information) occur at a zero shift in both the *x* and *y* directions (since the images are already registered). However, it can be observed in the plot that minima do not occur at $(0,0)$ but at $(-8,-2)$. Mutual information is maximum when the joint entropy of both images is minimum, however this might not be the case in thermal and visible images of diseased plants as the intensity information from these images may have no direct correlation. We then performed the same experiment by using the plant silhouettes from the visible and thermal images, as extracted using our method. Figure 3-5 (b) shows the plot of mutual information using these plant silhouettes. It is obvious

from the plot that the global minima (negative of mutual information) occur at a zero shift in both the *x* and *y* directions.

To compare the accuracy of our algorithm, we obtained ground truth silhouettes, denoted by T_{GT}, by manually marking the plant region in 30 randomly picked thermal images from our dataset consisting of 984 pairs of images. For comparison purposes, plant silhouettes from the thermal images were also obtained by using the same (gradient-based) method as the one used for visible light images after log transformation. Since thermal images are grayscale, we skipped the first two steps, i.e., conversion to the Lab colour space and subtraction of the 'a' and 'b' channels. In this case, the thermal images were first enhanced by using log transformation followed by de-noising using anisotropic diffusion filtering. Table 3-1 details a list of notations used to denote plant silhouettes extracted and registered using different methods. V_w^g in Table 3-1 is not discussed as we will show later that silhouettes extracted by the gradient-based method (i.e., T_g) in thermal images are not a good approximation of plant silhouettes and therefore the registration of V_w with T_g is not relevant.



(a) (b)

Figure 3-5: (a) (Negative of) the mutual information using intensity values of images vs. shifts in the *x* and *y* directions. (b) (Negative of) mutual information using silhouettes of images vs. shifts in the *x* and *y* directions.

Table 3-1: Notations used to denote silhouettes extracted using different methods.

T_{GT}	Ground truth plant silhouette from thermal image by manual marking.
V_{GT}	Ground truth plant silhouette from visible light image by manual marking.
T_w	Plant silhouette from thermal image using the SWT-based method.
T_g	Plant silhouette from thermal image using the gradient-based method.
V_w	Plant silhouette from visible light image using the SWT-based method.
V_g	Plant silhouette from visible light image using the gradient-based method.
V_w^w	V_w after registration with T_w using the registration method.
V_w^g	V_w after registration with T_g using the registration method.
V_g^w	V_g after registration with T_w using the registration method.
V_g^g	V_g after registration with T_g using the registration method.

Figure 3-6 shows four pairs of thermal and visible images of plants at different stage of powdery mildew disease. The boundary of V_{GT} and T_{GT} obtained by manually marking four pair of images is shown in pink. Figure 3-7 shows the overlap between the ground truth T_{GT} and [T_g,T_w] for the four pairs of images in Figure 3-6. The amount of overlap is represented in yellow, T_{GT} is represented in green and [T_g,T_w] are represented in red. The results show that the percentage of overlap (yellow) is higher in the pair (T_{GT}, T_w) than it is in the pair (T_{GT}, T_g) (Table 3-2). It is important to note that although T_w is very similar to T_{GT}, there are still some non-overlapping regions (red) in (T_{GT}, T_w) (marked by 'η' in Figure 3-7). These non-overlapping regions occur when the leaf surfaces are clumped together and thus, the fine details are not captured by our SWT-based method. Let us recall that in our method there is a trade-off between the accuracy of the binary image and the amount of noise at scale $s-1$, compared to scale s. At high scales, binary images with more blurred boundaries (i.e, B_s) are usually generated whereas at lower scales, binary images

with less blurred boundaries are usually generated at the expense of high noise content.

Some of the leaves in thermal Image P1 in Figure 3-6 show both higher and lower temperatures than that of the background. In this case, the gradient-based method failed to extract the plant silhouette from P1 as it missed most of the leaves. Our SWT-based method was unable to extract fine stem details near the plant boundary in P1 where the leaf temperature is very close to that of the background; however the overall result (overlap) of our method was better than that achieved by the gradient-based method.

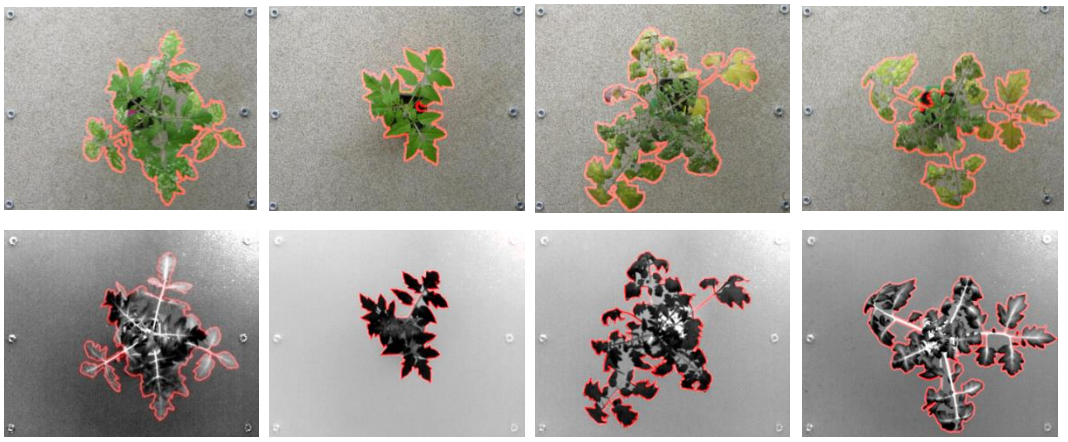


Figure 3-6: Top row: sample visible light images; bottom row: corresponding thermal images. Pink colour in top row represents hand-marked ground truth silhouettes.

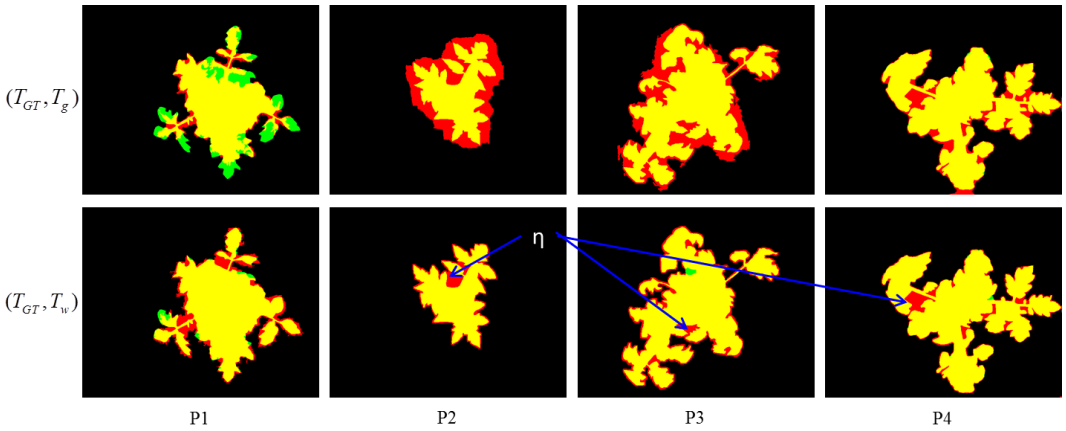


Figure 3-7 Amount of silhouette overlap for the four pairs of sample images in Figure 3-6. First row: overlap (yellow) between the manually extracted ground truth plant silhouette T_{GT} from thermal image (in green) and T_g (in red). Second row: overlap (yellow) between T_{GT} (in green) and T_w (in red). Region marked as 'η' shows non-overlapping pixels where the SWT- based method was unable to capture fine details due to clumped leaves at the plant boundary.

We quantified the accuracy of our silhouette extraction method by using the coverage metric Γ [57]:

$$\Gamma(I_1, I_2) = \frac{I_1 \cap I_2}{I_1 \cup I_2} \tag{3.8}$$

where I_1 and I_2 are any two binary images depicting silhouettes. In addition, we also computed the Sørensen–Dice index (Ψ) as another metric to quantify the results:

$$\Psi(I_1, I_2) = \frac{2\times|I_1 \cap I_2|}{|I_1|+|I_2|} \tag{3.9}$$

Table 3-2 reports the values of Γ and Ψ for the four pairs of images shown in Figure 3-6. These results show a higher percentage of overlap between the ground truth silhouette (T_{GT}) and the silhouette extracted by the SWT-based based method (T_w), than that between T_{GT} and the silhouette extracted by the gradient-based method (T_g).

Table 3-2: Values for the coverage metric Γ and the Dice index Ψ for four sample pairs of images in Figure 3-6.

	Silhouette Pairs	P1	P2	P3	P4
Γ	(T_{GT}, T_g)	0.7221	0.5328	0.6840	0.8302
	(T_{GT}, T_w)	0.7884	0.8048	0.8141	0.8718
	(T_{GT}, V_g^w)	0.7733	0.7535	0.7867	0.8162
	(T_{GT}, V_w^w)	0.8176	0.8054	0.8067	0.8632
	(T_w, V_g^w)	0.8459	0.89	0.8911	0.8372
Ψ	(T_{GT}, T_g)	0.8386	0.6952	0.8123	0.9072
	(T_{GT}, T_w)	0.8817	0.8919	0.8975	0.9315
	(T_{GT}, V_g^w)	0.8722	0.8594	0.8806	0.8988
	(T_{GT}, V_w^w)	0.8997	0.8922	0.8930	0.9266
	(T_w, V_g^w)	0.9165	0.9418	0.9424	0.9114

Figure 3-8 shows registration results obtained after registration. The top row shows the amount of overlap between the silhouettes T_w and V_g^w while the bottom row shows the thermal image overlaid on top of the registered visible light image. For the case of P1 and P4, the stem of the plant features a higher temperature than that of the background and as a consequence, the stem appears red in colour in the thermal image, which exactly coincides with the stem region in the visible light image. Leaves that are located far from the centre of the plant in P1 also feature a higher temperature than that of the background and appear in yellow colour. These leaves, when overlaid, also coincide with the leaves in the visible light image. Blue-

purple regions in P2, P3 and P4 show leaf regions in thermal images that coincide with the corresponding leaf regions in the visible light images when overlaid.

Figure 3-9 shows a cropped and enhanced section of P3 delimited by a white box in the silhouette overlap and by a black box in the thermal/visible overlay in Figure 3-8. The visible light image and the corresponding thermal image are shown in Figure 3-9 (a) & (b) respectively. The visible light image shows the leaf in red with a shade of pink. Because of this peculiar colour of the leaf, the extracted silhouette from the visible light image is not accurate and results in a disconnected region depicted in Figure 3-9 (c) in yellow and green colour. This inaccuracy causes a local mis-registration of the leaf region as depicted in Figure 3-9 (d), where the purple shade represents the leaf region and the yellow shade represents the non-leaf region in the thermal image. Note that the leaf region in the thermal image does not overlap the leaf region in the visible light image accurately. These local mis-registrations can occur if there are prominent inaccuracies in the extracted silhouettes.

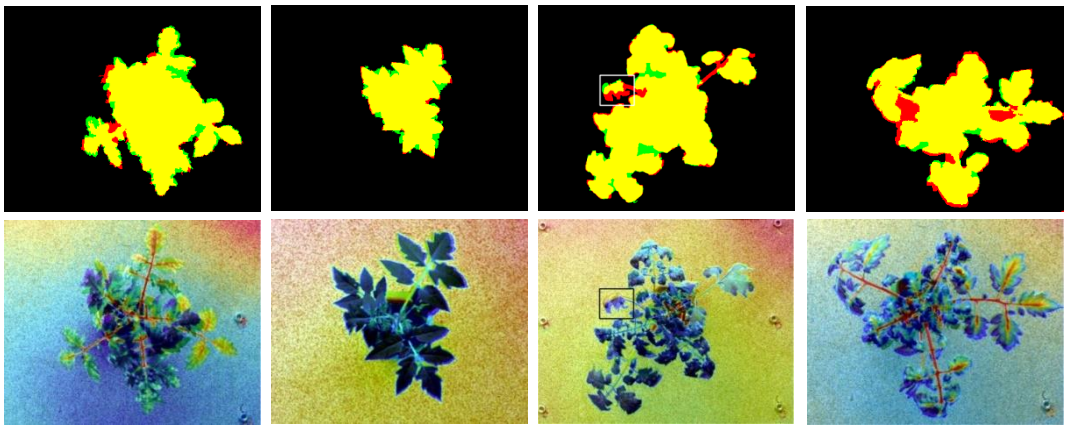
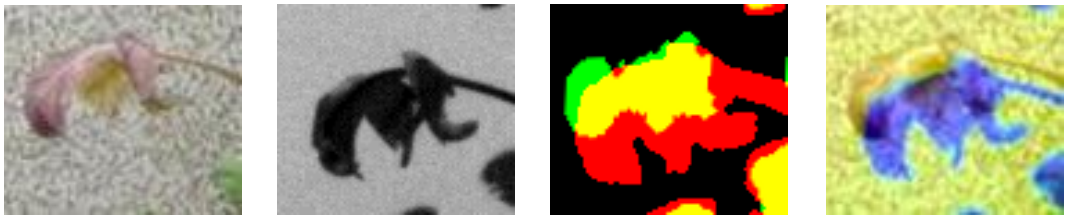


Figure 3-8: Registration results based on silhouettes. Top row: overlap (yellow) between the silhouettes of thermal (red) and registered visible (green) light images. Bottom row: thermal images overlaid on top of registered visible light images. Stem and leaf regions in thermal image coincide with stem and leaf regions in the visible light image. The ‘mis-registration’ represented by the white box in the P3 silhouette overlap and by the black box in the thermal/visible image overlap is explained in Figure 3-9.



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

Figure 3-9: Unregistered part of P3 marked with a box in Figure 3-8. Visible light image (a) with corresponding thermal image and (b) with a leaf in red-pink colour; (c) Local mis-registration of silhouettes shown in green and red colour; (d) Local mis-registration of visible and thermal images where purple shade represents the leaf region and yellow colour represents the non-leaf region in the thermal image.

The mean values of $\Gamma(T_w, V_g^w)$ and $\Psi(T_w, V_g^w)$ were calculated to be **0.8815** and **0.9364** with standard deviation of **0.0446** and **0.0260**, respectively. Table 3-3 tabulates Γ and Ψ values for 30 pairs of images randomly picked to mark the ground truth. The tabulated Γ and Ψ values are higher for the pair (T_{GT},T_w) than those for the pair (T_{GT}, T_g). Similarly, Γ and Ψ values are higher for the pair (T_{GT}, V_g^w) than those for the pair (T_{GT}, V_g^g), which shows that when the visible image silhouettes are registered with T_w, the overlap region is larger than that obtained when registered with T_g,. Figure 3-10 shows the histogram of $\Gamma(T_w, V_g^w)$ and $\Psi(T_w, V_g^w)$ for all pairs of images.

All results were obtained using MATLAB 2012a on an Intel 3.20GHz core-i5 PC with 16 GB of RAM running a Linux system. The proposed algorithm takes approximately 12.77 sec on average to register a pair of thermal and visible images. We also tested our SWT-based method to extract silhouette from visible light images. For this case, the subtraction of the 'a' channel from the 'b' channel was performed after converting the visible light image into the Lab colour space. We also used a multi-scale approach by first finding a coarse boundary at scale *s*=3 and then refining it at scale *s*=2. The resulted silhouette V_w was then registered with T_w

to obtain V_w^w. Note that in this case, the SWT-based method is generally computationally expensive as it takes over 15 sec, on average, to register a single pair of images.

Table 3-3 tabulates $\Gamma(T_w, V_w^w)$ and $\Psi(T_w, V_w^w)$ values for 30 pairs of images for which the ground truth was obtained. Table 3-4 tabulates *p*-values using the Wilcoxon signed-rank test for the null hypothesis H_0, median(D1-D2)=0, and the alternate hypothesis H_1, median(D1-D2)>0. D1 and D2 represent the coverage metric and Sørensen–Dice index, respectively, for the ground truth silhouette T_{GT} with silhouettes T_g, T_w, V_g^g,$V_g^w$and V_w^w. The significance level of the test is 0.01; *p*-values less than 0.01 indicate that the SWT-based method performs significantly better than the gradient-based method to extract a silhouette.

Figure 3-11 shows the overlap between the ground truth silhouette manually extracted from the visible light image V_{GT} and V_g, V_w. Note that the silhouette extracted using the SWT-based method shows the fine details of the boundary, which are relatively more visible in the P2 overlap. However, no major improvements in the registration results were achieved by extracting the silhouette using the SWT-based method. Given the additional computational cost of the SWT-based method, the gradient-based method is recommended to extract silhouette from visible light images for registration purpose.

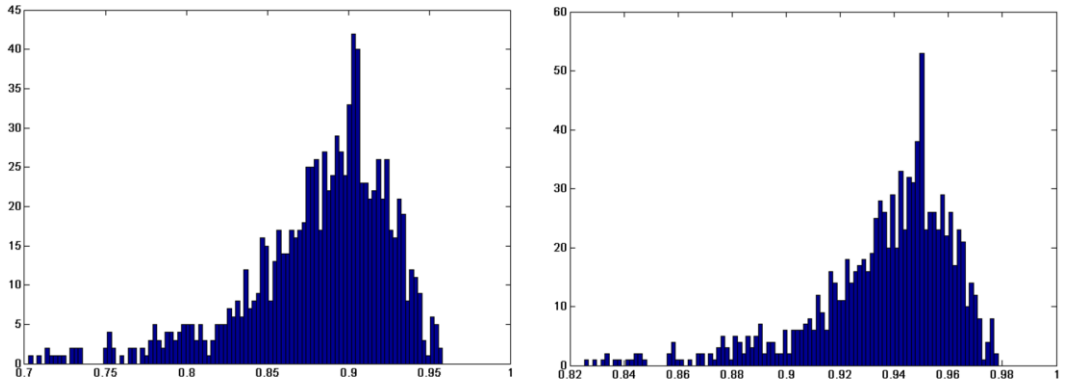


Figure 3-10: Histograms of $\Gamma(T_w, V_g^w)$ (*left*) and $\Psi(T_w, V_g^w)$ (*right*) for all 984 pairs of images.

Table 3-3: Mean and standard deviation of $\Gamma(T_w, V_w^w)$ and $\Psi(T_w, V_w^w)$ for 30 randomly picked pairs of images and mean and standard deviation of $\Gamma(T_w, V_w^w)$ and $\Psi(T_w, V_w^w)$ for all 984 pairs of images in our dataset are shown in bold.

Image Pair	No. of Sample Pairs	Γ		Ψ	
		μ	σ	μ	σ
(T_{GT}, T_g)	30	0.7273	0.0943	0.8385	0.0684
(T_{GT}, T_w)	30	0.8531	0.0385	0.9203	0.0226
(T_{GT}, V_g^g)	30	0.7464	0.0823	0.8522	0.0569
(T_{GT}, V_g^w)	30	0.8101	0.0460	0.8944	0.0285
(T_{GT}, V_w^w)	30	**0.8468**	**0.0420**	**0.9165**	**0.0250**
(T_w, V_g^w)	30	0.8754	0.0362	0.9332	0.0209
(T_w, V_g^w)	984	0.8815	0.0446	0.9364	0.0260
(T_w, V_w^w)	30	0.8952	0.0456	0.9441	0.0259
(T_w, V_w^w)	984	**0.9089**	**0.0539**	**0.9514**	**0.0312**

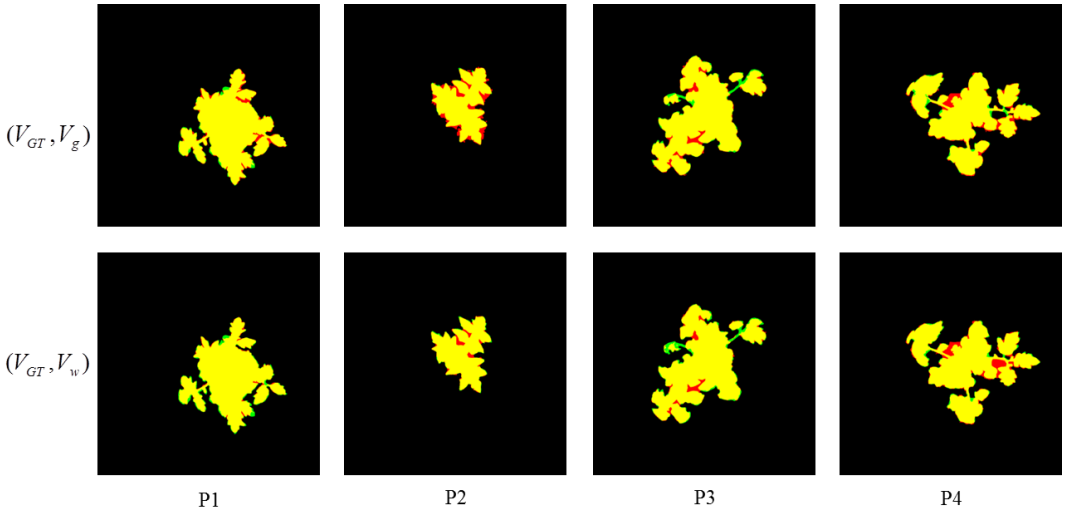


Figure 3-11: Top row: overlap (yellow) between the ground truth silhouette extracted from visible light image V_{GT} (green) and the silhouette extracted from visible light image using the gradient-based method V_g (red). Bottom row: overlap (yellow) between V_{GT} (green) and the silhouette extracted from the visible light image using the SWT-based method V_w (red).

Table 3-4: p-values using Wilcoxon signed-rank test to test the null hypothesis H0:median(D1-D2)=0, and the alternate hypothesis is H1:median(D1-D2)>0. The significance level of the test is 0.01.

D1	D2	p-value
$\Gamma(T_{GT}, T_w)$	$\Gamma(T_{GT}, T_g)$	1.0063×10^{-6}
$\Psi(T_{GT}, T_w)$	$\Psi(T_{GT}, T_g)$	1.0063×10^{-6}
$\Gamma(T_{GT}, V_g^w)$	$\Gamma(T_{GT}, V_g^g)$	9.4802×10^{-4}
$\Psi(T_{GT}, V_g^w)$	$\Psi(T_{GT}, V_g^g)$	9.4802×10^{-4}
$\Gamma(T_{GT}, V_w^w)$	$\Gamma(T_{GT}, V_g^w)$	6.1686×10^{-7}
$\Psi(T_{GT}, V_w^w)$	$\Psi(T_{GT}, V_g^w)$	6.1686×10^{-7}
$\Gamma(T_{GT}, V_w^w)$	$\Gamma(T_{GT}, V_g^g)$	6.1686×10^{-7}
$\Psi(T_{GT}, V_w^w)$	$\Psi(T_{GT}, V_g^g)$	6.1686×10^{-7}

Chapter Summary

In this chapter we proposed an algorithm for registration of thermal and visible light images of diseased plants using silhouette extraction. The main novelty of the algorithm is a multi-scale method based on the stationary wavelet transform (SWT) capable of extracting the silhouettes of diseased plants with high accuracy. Our proposed algorithm employs a gradient-based method to extract the plant silhouette in visible light images and the multi-scale SWT-based method to extract the plant silhouette in thermal images by iteratively refining a boundary across a number of scales. Using the extracted silhouettes, the algorithm then applies a global plus local registration approach to register salient features in pairs of thermal and visible light images.

Our test data set consisted of 984 pairs of images of 71 plants undergoing 10 different types of treatments acquired for 14 consecutive days. We first showed that silhouette based techniques give better results for registration of thermal/visible images than the methods that rely directly on pixel values such as mutual information. We then demonstrated the overall accuracy of our algorithms (average DICE coefficient value of **0.9514** and average coverage metric value of **0.9089**) to register the images in our dataset. We also evaluated the accuracy of our SWT-based method to extract silhouettes. To this end, we quantified the amount of overlap between extracted silhouettes and hand-marked ground truth silhouettes and showed that our method is capable of detecting plant silhouettes with high accuracy. Our SWT-based method can be extended to single/multiple object silhouette extraction and be used for registration of any kind of images provided that the objects of interest do not occlude each other.

Chapter 4
Depth Estimation using Stereo Images

To add depth information to the set of features which can be collected from registered thermal and visible light images we use disparity between the stereo image pair. For disparity estimation the images must be rectified and the disparity between the image pair must be only along the horizontal scan lines. If there is a vertical disparity we need to perform image rectification. In this Chapter, we first propose a marker based image rectification and then a multi-resolution disparity estimation method. We compare different disparity estimation algorithms on test data sets and noisy stereo images of diseased plants. The proposed method is fast, produces accurate and smooth disparity maps, and is less sensitive to background noise.

4.1. Image Rectification

Image rectification is a transformation process which maps a given stereo pair of images to a common image plane in a way that the disparity between the image pair is only in the horizontal direction. This simplifies the disparity estimation problem to one-dimensional scanning along the horizontal line parallel to the baseline. One method to perform rectification is to calibrate the stereo setup. However, if the calibration parameters are not available, then rectification can be performed by feature matching followed by estimation of the fundamental matrix. If the two cameras can be approximated by affine transformation, then rectification can be

performed by using just affine transformation [110,158]. In this chapter, we use extrinsic markers to match points in a stereo pair of images and then use affine transformation to perform rectification.

A sample pair of stereo images is shown in Figure 4-1. The colour images from left and right cameras are shown in Figure 4-1 (a) & (b) respectively. The colour images in (a) & (b) were converted to grayscale and then a RGB composite image was formed using grayscale left (a) image as red channel and grayscale right (b) image as green and blue channels. The composite image is shown in Figure 4-1 (c). The composite image shows that the same features of the plant in the left and the right image do not lie on the same horizontal line parallel to the baseline. For disparity estimation, we rectify the image pair so that the disparity between the left and the right image is only along the horizontal axis. The correspondence between extrinsic marker points placed on the ground is shown by arrows in Figure 4-1 (c) which indicates that the transformation function required for rectification is essentially a translation. However, to make our method more robust to minor affine transformations (if present), we align these markers using affine transformation. The goal is to align the points on the ground level, resulting in zero disparity at ground level and objects above the ground and closer to cameras showing positive disparity. The image rectification method proposed here consists of three steps:

1. Extraction of extrinsic markers
2. Establishing correspondence between marker points
3. Estimate and apply the transformation required for rectification to the right image.

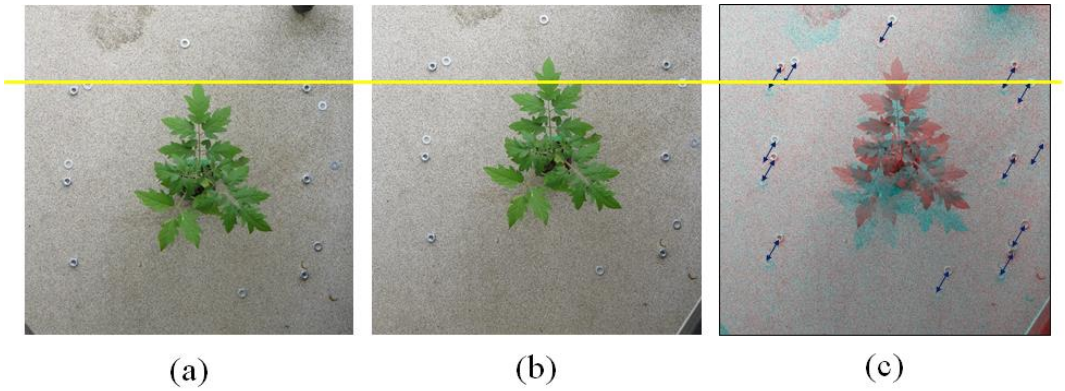


Figure 4-1: (a) & (b) show image from left and right camera; (c) RGB composite image formed by converting visible light images in (a) & (b) to grayscale and using grayscale image from left camera as red channel and grayscale image from right camera as green and blue channel. The yellow line shows that the same features of the plant do not lie on the same horizontal line and hence the stereo image pair requires rectification.

Figure 4-2 (a) & (b) show 'a' and 'b' channels of the Lab colour space for the image in Figure 4-2 (a). The marker points cannot be visually identified in the 'a' channel whereas they can be easily identified in the 'b' channel. Therefore, we perform PCA on the 'a' and 'b' channel, the projection of data on the 1st principal component contains mostly the plant region whereas marker points become more prominent in the projection of data on the 2nd principal component. The projection of data on the 2nd principal component was then smoothed with Gaussian filter and the size of the filter was chosen to enhance the regions belonging to the marker points (Figure 4-2 (c)). Figure 4-2 (d) shows the image after thresholding the image in Figure 4-2 (c) using the 1st percentile value. The 1st percentile value was chosen because the marker points show very low values as can be observed in Figure 4-2 (c).

After the binary image has been obtained the Hough transform was used to extract the location of the marker points. The Hough transform initially extracts all the edges in the image and then using the equation of a circle $(x - c_1)^2 + (y - c_2)^2 = r^2$, maps all the points on the edges to parameter space consisting of c_1, c_2 & r. If $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the set of points on a circle then all the points in the set $\boldsymbol{x}$ and $\boldsymbol{y}$

belong to the same point in 3-D parameter space. The high pixel concentration in the parameter space points to the presence of a circle corresponding to those particular parameters. The steps performed to extract the location of extrinsic marker points can be listed as below:

1. Convert RGB image to Lab colour space.
2. Perform PCA on ‘a’ and ‘b’ channels.
3. Smooth the projection on the 2nd principal component with a Gaussian filter.
4. Threshold the resulted image from step 3 to extract potential marker points.
5. Estimate the position of marker points using the Hough transform [43,159].

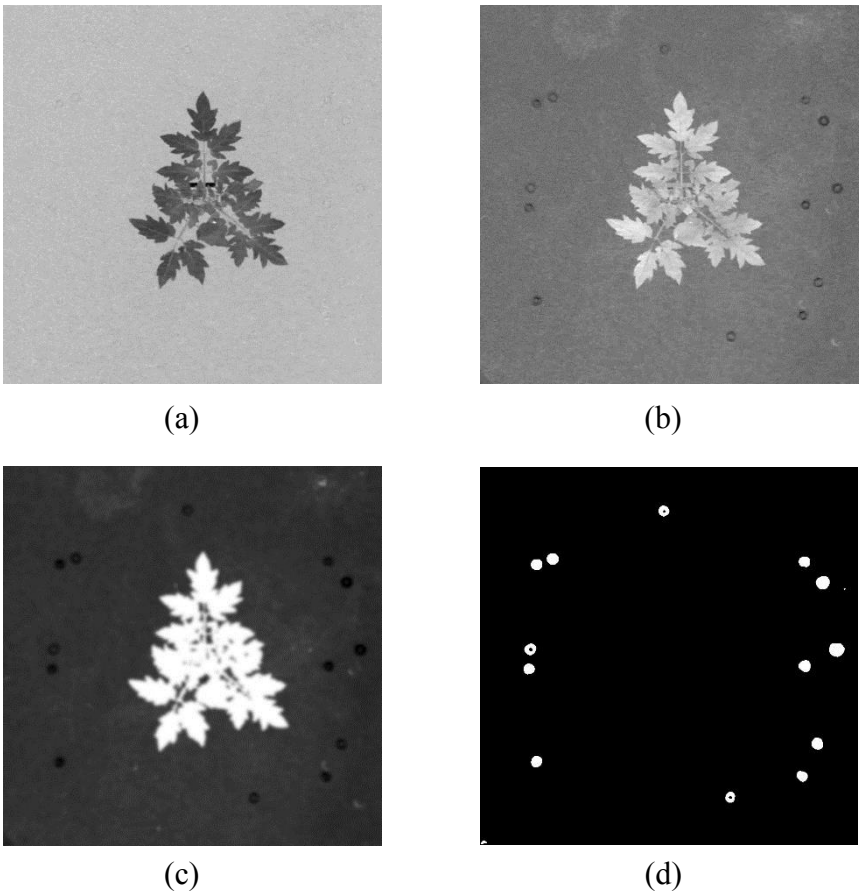


Figure 4-2: (a) & (b) show ‘a’ and ‘b’ channels of ‘Lab’ colour space for plant in Figure 4-1; (c) shows the result of projection on 2nd principal component after performing PCA on ‘a’ and ‘b’ channels and smoothing with Gaussian; (d) shows the result of threshold and the candidate marker points.

After the location of marker points has been identified, the correspondence between the points was established using the following steps:

1. A matrix of pairwise distances between location of marker points in the left and right images was computed.
2. The potential corresponding marker points were identified using the minimum distance along each row/column of the matrix.
3. As a verification step, all the lines joining the corresponding points in the left and right images form (almost) the same angle. The angle of the lines with the *x*-axis was used to verify the match in Figure 4-1 (c).

After the correspondence between the marker points was established, affine transformation parameters were estimated using inverse transformation and the right image was transformed using the estimated affine transformation. The composite image after rectification is shown in Figure 4-3, where red channel in the RGB composite image corresponds to left image and blue and green channels correspond to rectified right image. After performing rectification, features of the plant in both the left and the right image lie on the same horizontal row as indicated by the yellow line in the zoomed version of the composite image.

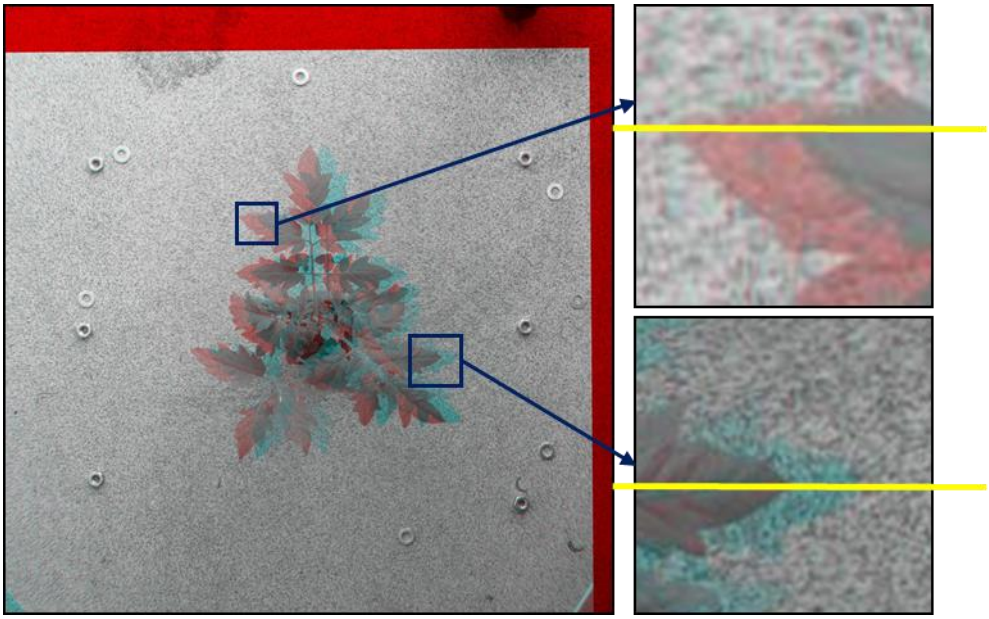


Figure 4-3: RGB composite image formed by converting images from the left and the right camera to grayscale and then using the grayscale left image as red channel and rectified grayscale right image as blue and green channels. The yellow line on the right shows that plant features in zoomed in parts lie on the same horizontal line.

4.2. Disparity Estimation

After the images have been rectified disparity estimation is simplified to one dimensional horizontal scanning. In this section, we compare results of six different disparity estimation algorithms on some of the images from the Middlebury Stereo data set[8] and then we test these algorithms on rectified stereo images of diseased plants (Figure 4-3). We selected these six particular algorithms for our study based on three criteria 1) they represent major disparity estimation schemes, 2) these methods have been used in the past for comparison purposes [112,128,137], and 3) they produce acceptable results on the plant images. The goal is to find an algorithm which produces an accurate and smooth disparity map, and is less sensitive to the

[8] http://vision.middlebury.edu/stereo/data/

background noise and colour variation in diseased plants.

4.2.1. Block-based Stereo Matching (KO'98) [114]

A basic block-based stereo matching algorithm was presented in [115]. The algorithm takes LoG (Laplacian of Gaussian) transform of the stereo images where standard deviation of Gaussian was chosen to be 1-2 pixels. It was shown in a later study that using the LoG transform instead of directly using intensity values reduces the disparity estimation error especially in block based methods [128]. The algorithm uses absolute differences (the so-called l_1-norm) to find the matching blocks and finally filters the disparity image to obtain dense stereo results. The post filtering gives high confidence to textured areas and low confidence to flat areas and then performs left/right consistency check. The left/right consistency check looks for consistency in matching from a fixed left image region to a set of right image regions and matched right region to a set of left image regions.

4.2.2. Multi-Resolution Stereo Matching (SO'07) [136]

A multi-resolution stereo matching algorithm was proposed for surface modelling of plants in [137,160]. The algorithm first divides the image into overlapping blocks at each level of the multi-resolution pyramid and then uses a variation of the Birchfield and Tomasi (BT) cost function (difference of linearly interpolated intensity functions) to match the corresponding blocks [116]. These matches are first computed at the highest level of the pyramid and propagated to the next (lower) level using bilinear interpolation to avoid the search window to be trapped in local minima. To avoid the propagation of poor results from the parent, the algorithm weighs the information at each point from direct parent based on the disparity

information and error propagated from neighbouring parents and the direct parent. Finally, the disparities at each level are refined using the Kalman filter [161] based on the disparities and errors associated with the coarser resolution.

4.2.3. Graph-cut based Stereo Matching (GC'01) [119]

This algorithm defines a global energy function and minimises the energy function using graph cuts [118–120,162]. For a pair of stereo images, the goal of this algorithm was to find a label f_p which denotes disparity value for each pixel p in the left (reference) image such that the following energy function for a label f can be minimised,

$$E(f) = E_{data}(f) + E_{smooth}(f) + E_{visibility}(f) \tag{4.1}$$

where $E_{data}(.)$ is the data term and was chosen to be squared differences between the matching pixels, $E_{smooth}(.)$ is a constraint which helps to create a smooth disparity map, the smoothness term involves the notion of the neighbourhood and adds a penalty if the neighbourhood has different disparity, and finally $E_{visibility}(.)$ is a penalty term for occluded pixels. The algorithm constructs a graph with the help of the above energy function defined above and then minimises the energy function. The algorithm initially defines a unique disparity α and then iteratively searches for an α which minimises the energy function. If the energy is minimised, it selects the new α; otherwise, it selects the current α as the disparity.

4.2.4. Non-local Cost Aggregation Method (YA'12) [129]

Cost aggregation methods usually perform summing/averaging matching cost over windows and, therefore, provide locally optimal solutions and are vulnerable to

local nature of the scene, for example the lack of texture. The method proposed in [130] uses the concept of bilateral filter by weighting the pixel intensity differences with intensity edges and provides a non-local solution by aggregating the cost on a tree structure derived from the stereo image pair. This method constructs a graph $G = (V, E)$, with weight function *w* using the guidance image (reference image for disparity) *I*. The vertices (nodes) of this graph are all the image pixels in the guidance image and the edges are established using weights between the nearest neighbouring pixels. All the edges carry weights defined by a weighing function *w* which is image gradient in this case. Edges with large weights (intensity edges) are removed as unwanted edges and a minimum spanning tree (MST) is constructed. If $D(p, q)$ denotes the shortest distance (sum of weights) between two nodes in MST, then the aggregated cost $C_d^A(p)$ at pixel *p* and disparity level *d* can be defined as

$$C_d^A(p) = \sum_{q \in N} \exp\left(\frac{-\mathrm{D}(p,q)}{\sigma}\right) C_d(q) \tag{4.2}$$

where $C_d(q)$ is the matching cost at pixel q at disparity level *d*, *N* is the user-specified support region and σ can be used as a scaling factor to adjust the cost function for spatial similarity. An initial cost volume is computed based on the initial disparity estimates. For disparity refinement, left/right consistency check is used to divide the disparity estimates into stable and unstable disparities. The non-local cost aggregation is then applied followed by winner-take-all operation to propagate disparity from stable pixels to unstable pixels.

4.2.5. Semi-Global Matching (SGM'08) [124]

A semi-global matching method was presented in [124,125] to overcome the

limitation of local and global methods e.g., computational cost. This method defines a global energy function as

$$E(D)=\sum_{p}\left(C(p,D_p)+\sum_{q\in N}P_1T[|D_{\mathrm{p}}-D_{\mathrm{q}}|=1]+\sum_{q\in N}P_2T[|D_{\mathrm{p}}-D_{\mathrm{q}}|>1]\right) \tag{4.3}$$

The first term in the energy function on the right hand side in the above equation is the pixel-wise matching cost, which was chosen to be BT in this case, for disparity D_p at pixel p. The second and the third terms add smoothness constraint to the function. The second term adds a small penalty for all pixels q in the neighbourhood N of pixel p for which disparity changes are small. The third term adds large penalty for large changes in disparity in the neighbourhood of pixel p. The energy minimisation problem was simplified by aggregating 1D *minimum* costs from all directions. The cost $L_r(p,d)$ of pixel p at disparity d along the direction r can be defined as

$$\begin{aligned}L_r(p,d)=\mathrm{C}(p,d)+\min(&L_r(p-r,d),L_r(p-r,d-1)+\mathrm{P}_1,\\&L_r(p-r,d+1)+\mathrm{P}_1,\min_i L_r(p-r,i)+\mathrm{P}_2)\\&-\min_k L_r(p-r,k)\end{aligned} \tag{4.4}$$

The disparity d at pixel p can then be determined by minimising the cost $S(p,d)$

$$S(p,d)=\sum_r L_r(p,d) \tag{4.5}$$

Similar to the previous method, disparity estimation is followed by disparity refinement steps. The first step removes peaks which appear as small patches that are very different to surrounding disparities. For further refinement of disparity, disparity in valid regions was extrapolated to occluded regions determined by left/right consistency check.

4.2.6. Multi-Resolution Semi-Global Matching (MRSGM'13)

We propose a multi-resolution approach followed by a post processing step to create a smooth and accurate disparity map which is robust to the background noise and variation in our data set. An overview of the proposed approach is shown in Figure 4-4. We first calculate the disparity at three different resolutions and then take the median of the disparity estimated at these three resolutions. For disparity estimation, we use the semi-global matching method as described in the previous section. However, we use block based BT as the matching cost instead of pixel-wise matching. After disparity estimation, we post-process the result using the colour information as proposed in [129].

The post processing method uses the concept of bilateral filtering to improve the disparity map, the underlying assumption is that colour discontinuity is a strong indicator of depth discontinuity. If D denotes the disparity map and I denotes the reference image, then for a pixel $p = \{x, y\}$, let us assume $\overrightarrow{d_p} = \{D(x-1, y), D(x, y-1), D(x+1, y), D(x, y+1)\}$, $\overrightarrow{u_p} = \{x-r, \dots, x+r\}$, $\overrightarrow{v_p} = \{y-r, \dots, y+r\}$, where r is the radius of the bilateral filter, we can update the disparity map D using the following equation

$$D(x,y) = \arg\min_{d \in \vec{d}_p} \frac{\sum_{u \in \vec{u}_p} \sum_{v \in \vec{v}_p} W(u,v) C(u,v,d)}{\sum_{u \in \vec{u}_p} W(u,v)} \tag{4.6}$$

where $$W(u,v) = \exp\left(-\frac{\| I(x,y), I(u,v) \|_2}{2\sigma_R^2}\right).\exp\left(-\frac{(x-u)^2 + (y-v)^2}{2r^2}\right) \tag{4.7}$$

and $$C(u,v,d) = \min(\upsilon\mathcal{L}, | D(u,v) - d |) \tag{4.8}$$

υ is a constant and was chosen to be 0.2 [129], $\mathcal{L}$ is the total number of disparities. The remaining parameters σ_r and r can be used to control the smoothness of the updated disparity map.

In addition to the disparity estimation algorithms described above, we also tested some other algorithms for disparity estimation, such as those presented in [117,118,122,129,131,162–164] but we found from our experiments that these algorithms did not produce a good disparity map on our data set and some of these algorithms were found to be very sensitive especially to the background noise. Therefore, we do not include these algorithms in our analysis.

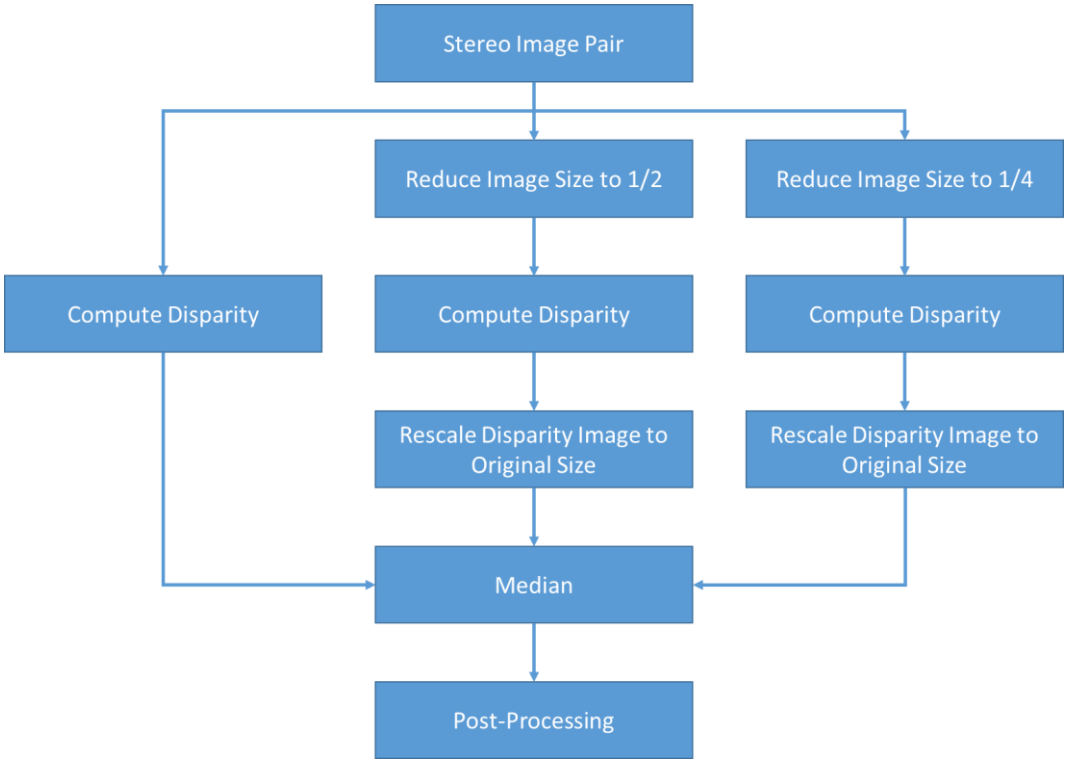


Figure 4-4: Overview of the proposed multi-resolution semi-global matching approach.

4.3. Experimental Results and Discussion

We first compare the results of the algorithms presented in the previous section on

Middlebury data set and then we will compare the results on stereo images of diseased plants. All the algorithms and results presented in this section were generated using a machine running Windows 7 on an Intel Core i3-2120 (3.3 GHz) CPU with 3GB RAM (665 MHz). The code for SO'07 (provided by the author) was implemented in MATLAB 2013a, whereas the C/C++ implementation of GC'01 and YA'12 were downloaded from the author's websites[9,10]. We used OpenCV[11] library to implement SGM'08 in C++ for our experiments. The KO'98[12] and MRSGM'13 were partially implemented in C++ and partially in MATLAB 2013a, where the post processing algorithm in MRSGM'13 uses C++ implementation by [129][13].

For KO'98, we chose 11×11 block size and for SO'07, we used 16×16 with 2 pyramid levels for our experiments. For GC'01, YA'12 and SGM'08, we chose default parameters provided by the authors. Finally, we chose 5×5 block-based BT as cost function and r=3, $\sigma_r = 15$ for MRSGM'13. All the parameters specified above other than the default parameters were chosen on the basis of their good results on stereo images of diseased plants.

Figure 4-5 shows the six images used to compare all the disparity estimation algorithms used in our experiments. Images labelled 'Aloe', 'Baby 1', 'Bull',

[9] http://www.cs.cornell.edu/People/vnk/recon.html

[10] http://www.cs.cityu.edu.hk/~qiyang/publications/cvpr-12/code/index.htm

[11] http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

[12] http://www.mathworks.co.uk/help/vision/ref/disparity.html

[13] http://www.cs.cityu.edu.hk/~qiyang/publications/code/cvpr-10-csbp/csbp.htm

'Flower Pots', 'Rocks 1' were taken from the Middlebury data set whereas the 'Plant' image shows a sample plant image from our data set. The ground truth disparity map for the first five images was also provided with the Middlebury data set. To measure the quality of our results, we compute two quality measures as suggested by [112]. If D_c represents the disparity map estimated by the algorithm being tested and D_{GT} represents the ground truth disparity then we define RMS and *B* (percentage of bad matching pixels) as follows,

$$RMS = \left(\frac{1}{N_d} \sum_{(x,y)} \Upsilon(x,y)^2 \right)^{1/2} \tag{4.9}$$

$$B = \left(\frac{1}{N_d} \sum_{(x,y)} \mathrm{T}(\Upsilon(x,y), \delta_d) \right) \tag{4.10}$$

where $\Upsilon(x,y) = |D_c(x,y) - D_{GT}(x,y)|$, N_d is the total number of pixels and δ_d is the disparity error tolerance. $\mathrm{T}(\Upsilon, \delta_d) = 1$ if $\Upsilon > \delta_d$ else $\mathrm{T}(\Upsilon, \delta_d) = 0$. We chose the tolerance value δ_d to be 1 pixel for the results presented here. We calculated RMS and *B* values for the whole image region and non-occluded region to test the algorithms.

Let us consider the plots shown in Figure 4-6 to Figure 4-10. Let us first consider SO'07 algorithm [137], which aims to produce smooth disparity maps but inadvertently increases the error in an attempt to produce smooth disparity maps, resulting in large errors in RMS and *B* plots as compared to all the other algorithms. In Figure 4-10 we do not include SO'07 as it was implemented in MATLAB and is expected to be slow compared to the algorithms implemented in C/C++. However, efficiency of the algorithm in terms of time is irrelevant if the RMS error and *B* are

very high. All the other algorithms produce comparable results, with GC'01 and YA'12 producing almost negligible RMS error whereas RMS error is less than 1% for MRSGM'13, SGM'08 and KO'98. Similarly, *B* is negligible in the non-occluded regions for GC'01 and YA'12, whereas it remains less than 4.2%, 3.3% and 2.9% for MRSGM'13, SGM'08 and KO'98 respectively.

Figure 4-10 compares computational efficiency of the algorithms. GC'01 was found to be roughly more than 100 times slower than MRSGM'13, and YA'12 was calculated to be at least 3.5 times slower than MRSGM'13 on the plant images. SGM'08 and KO'98 performed faster computation compared to MRSGM'13. We reiterate that GC'01 and YA'12 were implemented in C/C++ whereas MRSGM'13, SGM'08 and KO'98 were partially implemented in MATLAB and partially in C/C++. These results lead to the conclusion that although GC'01 and YA'12 produce more accurate results, they are slow compared to MRSGM'13, SGM'08 and KO'98.

Let us consider the performance of all the six algorithms on the plant images. Figure 4-11 shows that SO'07 performed poorly on the plant images and was found to be very sensitive to the background noisy pattern in the image. From the results on test images from Middlebury data set, we know that GC'01 and YA'12 produce accurate disparity maps but in the case of plant images these two algorithms were found to be highly sensitive to the noise content in the image. GC'01 is slow and produces artifacts along the scan lines on the plant images. YA'12 produces false disparity maps in the region which belong to the background. We know from the rectification step (section 4.1) that disparity is zero on the pixels which belong to the ground but

YA’12 produced disparity higher than zero. This is because the YA’12 algorithm divides the image into regions and assumes a constant disparity throughout this region. This introduces artifacts which can be observed in YA’12 result. KO’98 and SGM’08 results were found to be less sensitive to background noise but the disparity map produced by the algorithms were not smooth and showed small peaks/patches around some pixels which were inconsistent with the neighbouring disparity. When compared to all the other algorithms, MRSGM’13 not only produced smooth disparity maps but was also found to be less sensitive to the noise content.

Aloe

Baby1

Bull

Flower Pots

Rocks 1

Plant

Figure 4-5: ‘Aloe’, ‘Baby 1’, ‘Bull’, ‘Flower Pots’, ‘Rocks 1’ were taken from Middleburry dataset whereas ‘Plant’ image shows a sample plant image from our data set.

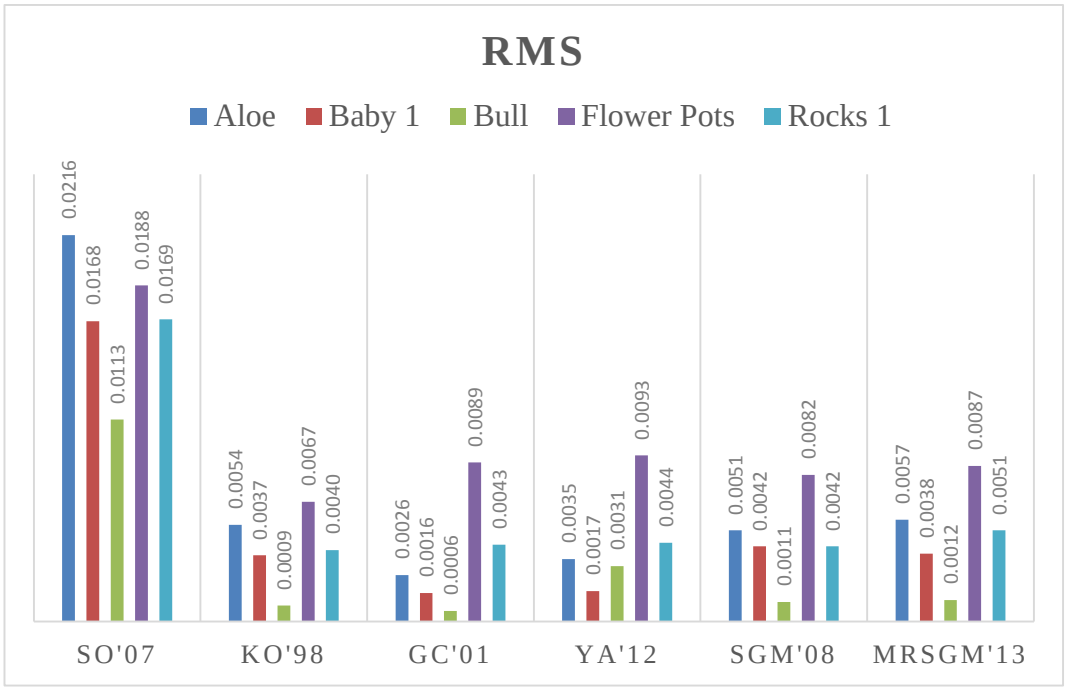


Figure 4-6: RMS plots for five different images from Middlebury dataset shown in Figure 4-5 using disparity estimation algorithms in section 4.2.

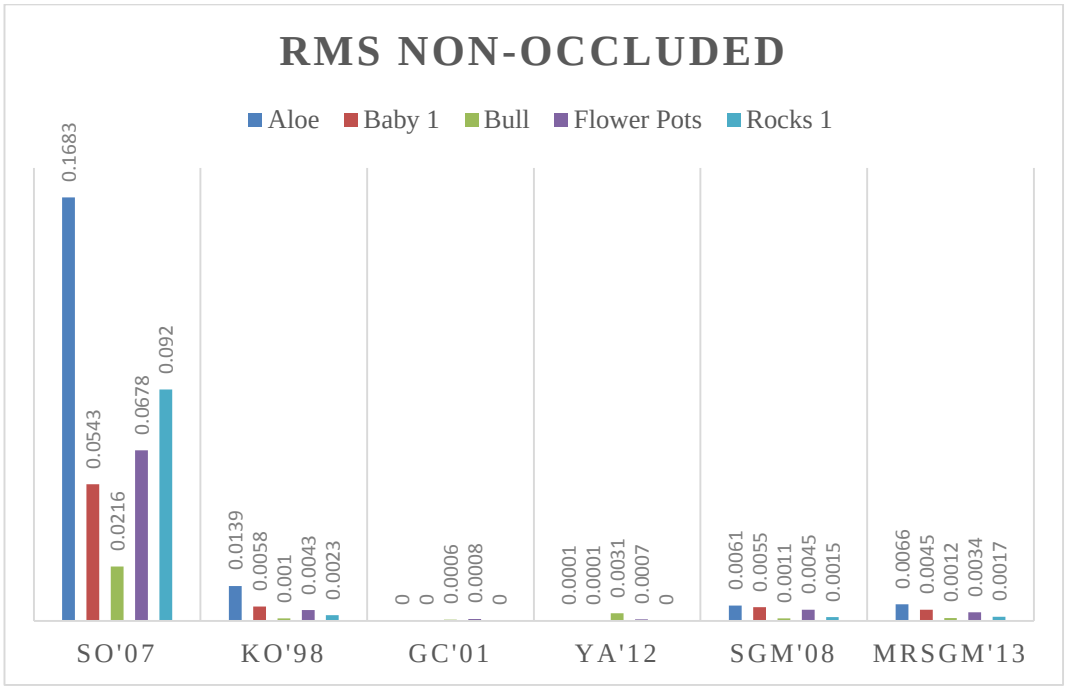


Figure 4-7: RMS plots of non-occluded regions for five different images from Middlebury dataset shown in Figure 4-5 using disparity estimation algorithms in section 4.2.

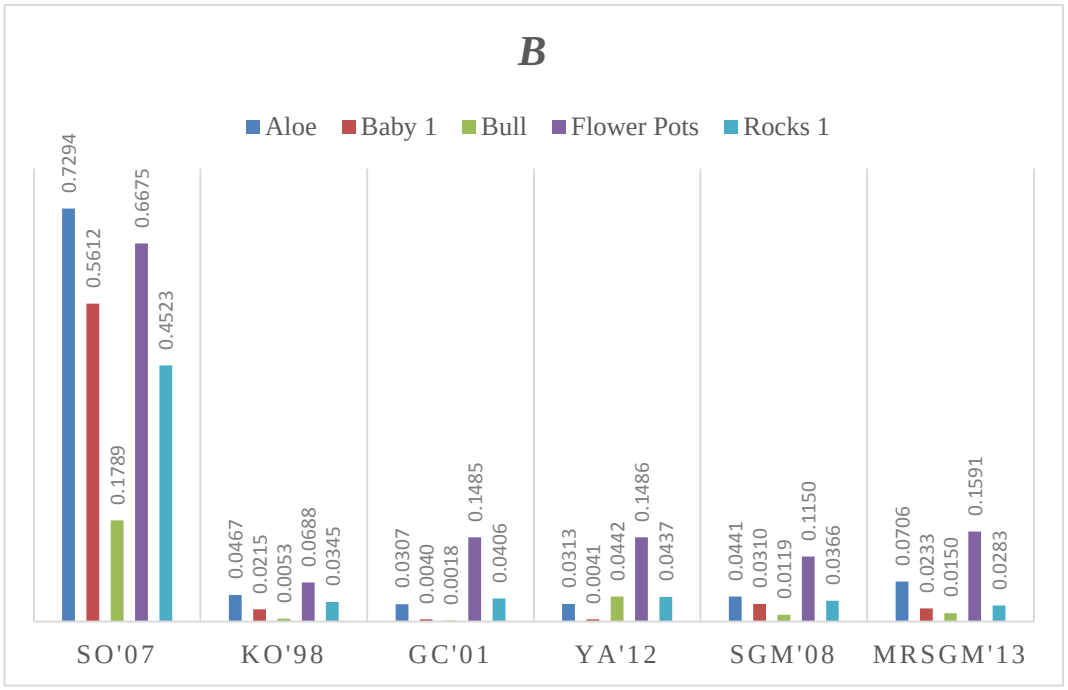


Figure 4-8: *B* value plots for five different images from Middlebury dataset shown in Figure 4-5 using disparity estimation algorithms in section 4.2.

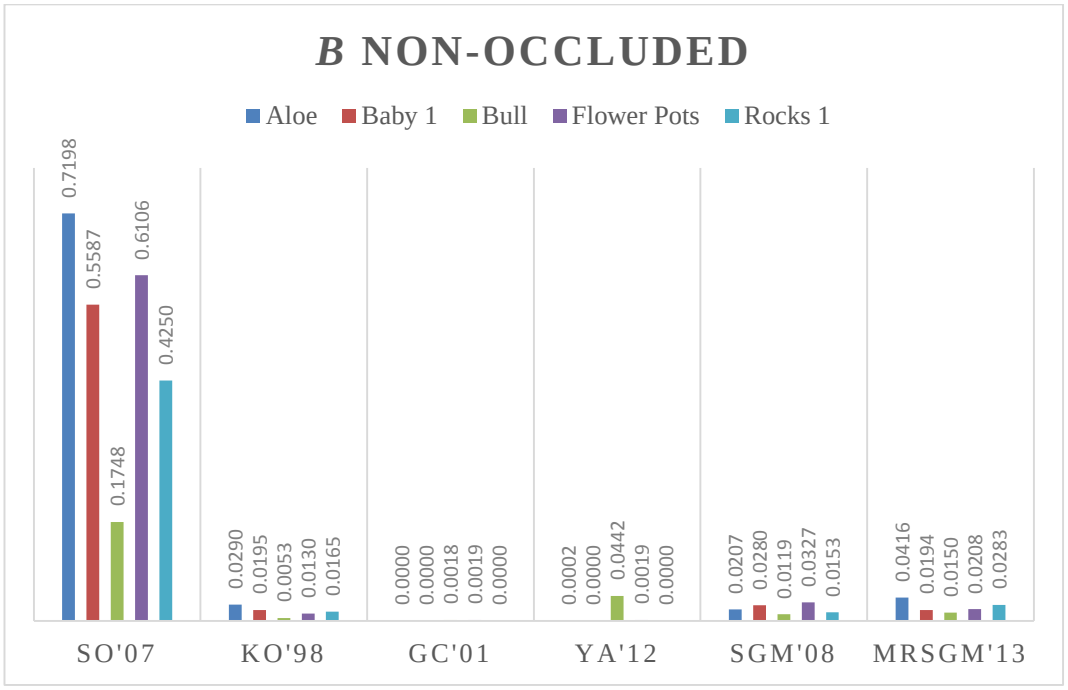


Figure 4-9: *B* value plots of non-occluded regions for five different images from Middlebury dataset shown in Figure 4-5 using disparity estimation algorithms in section 4.2.

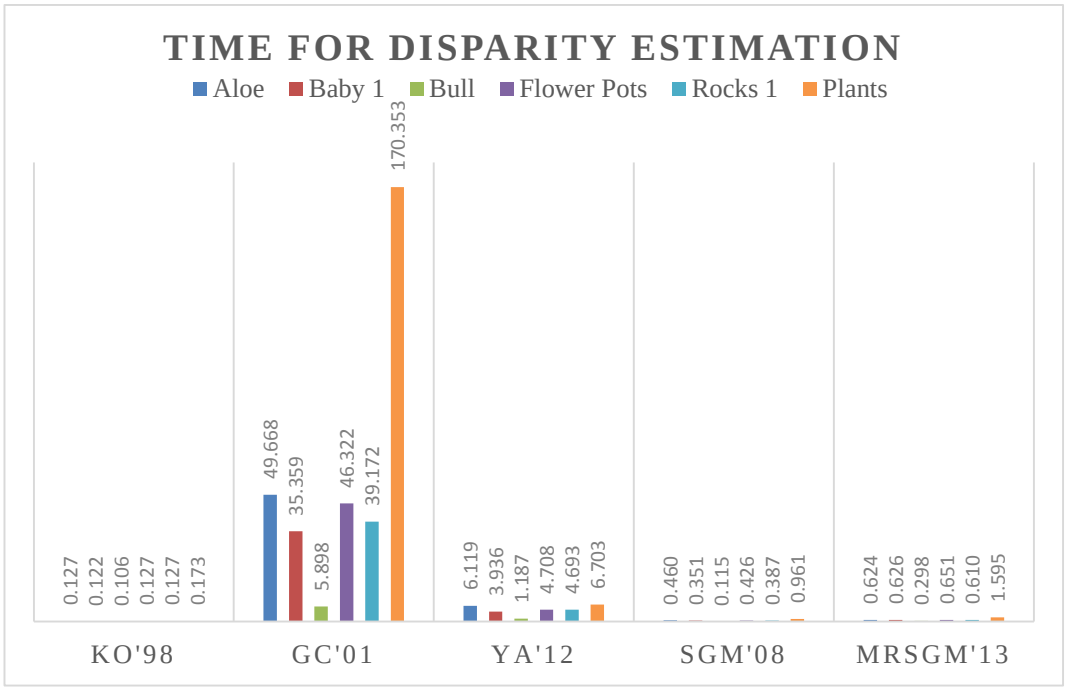


Figure 4-10: Time taken for disparity estimation of images in Figure 4-5 using disparity estimation algorithms in section 4.2.

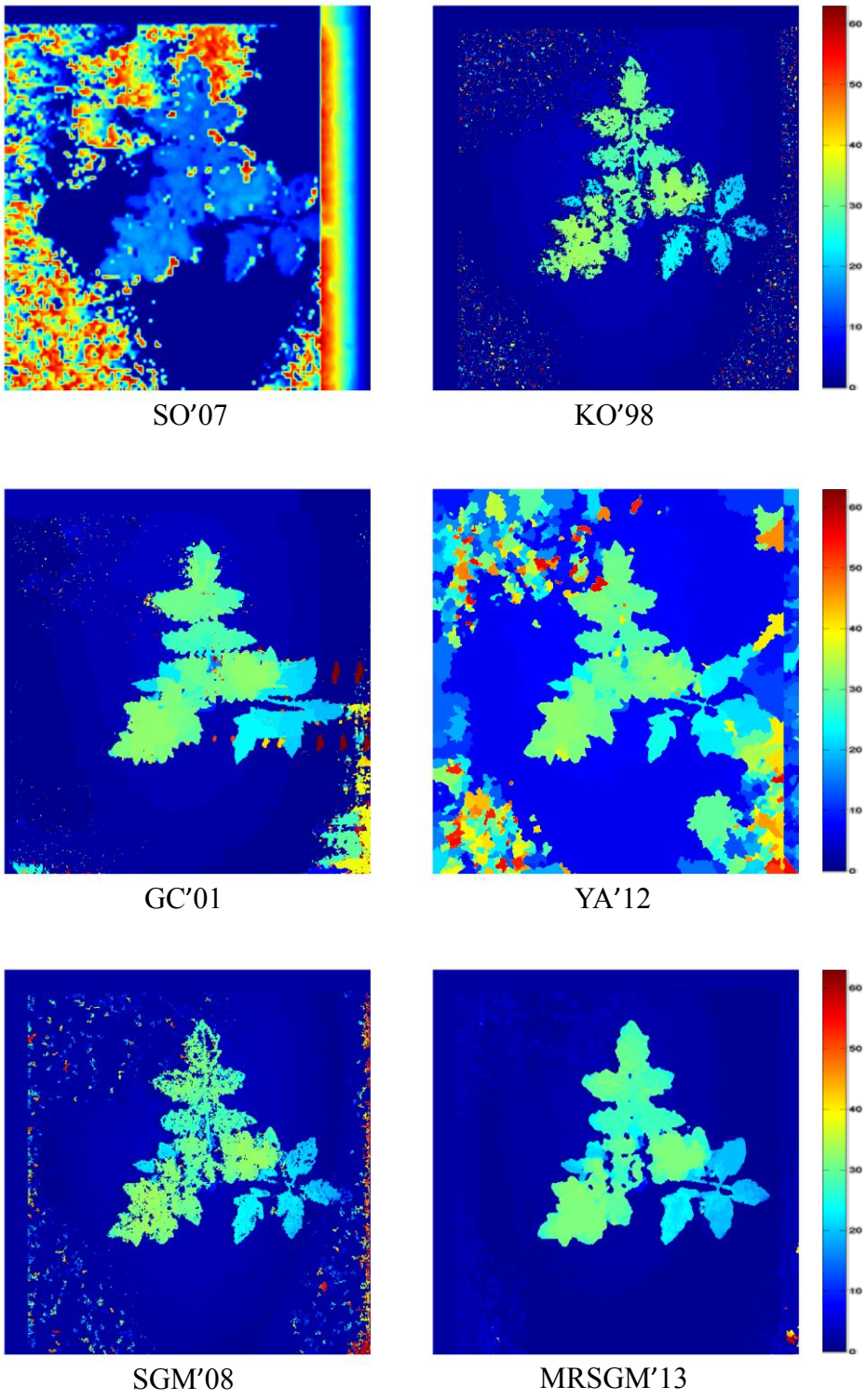


Figure 4-11: Disparity estimation results of algorithms in section 4.2 on the stereo plant image.

Figure 4-12 shows model of plant image after depth estimation, please follow the attached link[14] for full view of this plant model. To verify the depth estimated using MRSGM'13 we calculated the distance of an object as shown in Figure 4-13 using equation (2.1). The difference between the actual distance calculated manually and estimated distance was calculated to be no more than 2 cm in Figure 4-13. However, as disparity is directly related to depth, in our experiments we directly used disparity information instead of calculating the distance of each plant from the camera.

Chapter Summary

In this chapter, we proposed a method for estimation of depth in images of plants with high background noise. As a first step, we performed image rectification and then we proposed a disparity estimation algorithm. We compared six different algorithms for disparity estimation and showed that the proposed MRSGM'13 algorithm can be used to produce smooth and accurate disparity maps not only on the test data set but also on the diseased plant images. Although GC'01 and YA'12 performed best on the test data sets, our plant images with relatively more background noise than the Middlebury images proved to be quite challenging for these algorithms. In addition, GC'01 and YA'12 were calculated to be very slow compared to the proposed MRSGM'13 which was found to be not only less sensitive to the noisy pattern but also produced smooth and accurate disparity maps.

[14] http://www.youtube.com/watch?v=l3wqd0IdHl0&hd=1

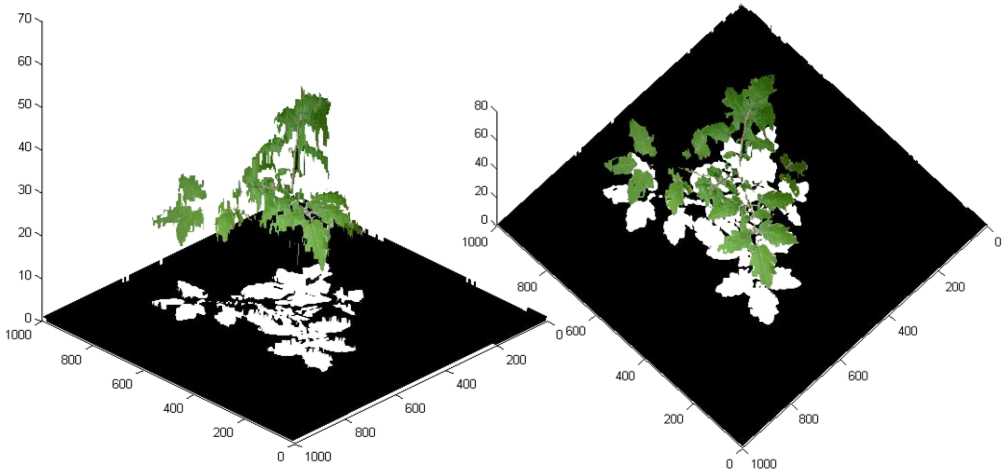


Figure 4-12: A 3D model of plant after depth estimation.

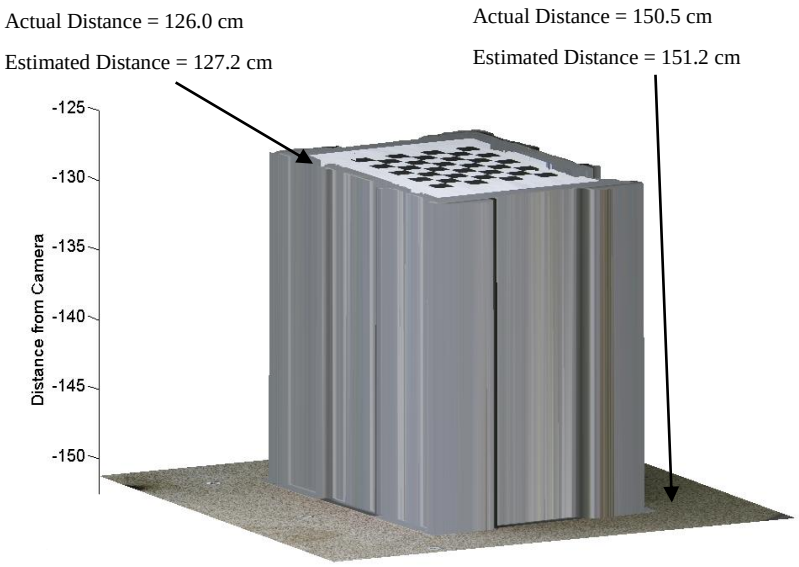


Figure 4-13: Estimated distance of an object using the proposed algorithm compared to actual distance.

Chapter 5
Combining Multi-Modal Image Data for Anomaly Detection in Plants

In this chapter, we show that by combining thermal information with the colour and depth information we can increase the accuracy of classification algorithms to detect anomalous plant regions compared to directly using thermal or colour information. This increase in accuracy of classification can be translated to huge savings in commercial crops. In section 6.1, we present a classification algorithm and results for detection of water deficient regions in spinach canopies and in section 6.2, we present classification algorithm and results for detection of diseased tomato plants infected with powdery mildew.

5.1. Detection of Water Deficient Regions

Thermal imaging has been used in the past for remote detection of water deficient regions of plant canopies. In this chapter, we present a new technique to enhance the 'discriminatory power' of thermal imaging to identify parts of the canopy which have reduced water availability (under water stress). For our experiments, thermal and visible light images of a spinach canopy with different levels of soil moisture were captured. We then combine information from thermal and visible light imagery along with machine learning techniques to identify water deficient regions of spinach canopy. Our approach is to extract information about the light intensity and green-ness of the plant from the visible light images and use this information

along with statistical measurements from thermal images to differentiate between water deficient and well-watered plants using Support Vector Machines (SVM), Gaussian Processes Classifier (GPC) and a combination of both. All three classifiers show promising results with the set of features extracted using combined information from thermal and visible light images.

5.1.1. Image Acquisition

The images used in this section were captured by our collaborators Miss Hazel Smith and Prof Gail Taylor at the University of Southampton in collaboration with Dr Graham Clarkson of the Vitacress Limited. Spinach (*cv. Racoon*) was drilled on 11 March 2010 at Mullens Farm, Wiltshire and was maintained with commercial practice. Measurements were taken on 27 April 2010 of two kinds of treatment areas in bright and clear conditions: well-watered and water-deficient. The former treatment had been irrigated during the preceding week while the latter had not, and were both harvested the following week for market. Both treatment areas were crops of spinach of the same age and variety and both had reached full canopy cover. Sampling consisted of taking a single image and soil moisture measurement at 20*m* intervals for the length of each row. Three rows were sampled per treatment, with five rows separating the sampled rows (Figure 5-1). Soil moisture measurements were made using a Delta-T ML2x Thetaprobe connected to a HH2 moisture meter (Delta-T Devices, Cambridge, UK), with the probe position being in the centre of the bed at a depth of approximately 7*cm*. Infrared thermal images were taken using a TH9100WR thermal camera (Figure 1-1 (b)) from a fixed distance of approximately 1*m* above the crop. All measurements were taken between 1100 hrs and 1300 hrs on a single day.

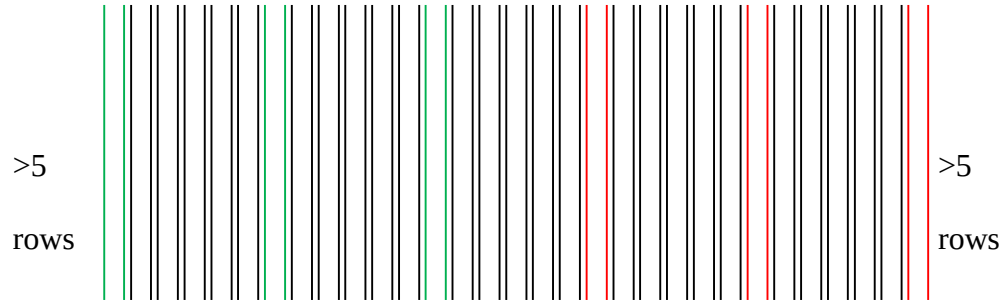


Figure 5-1: Sampling layout for the collection of thermal images and soil moisture measurements, 2010. The rows represent beds of Spinach (cv. *Racoon*), with those marked in green showing irrigated sample rows and the red indicating non-irrigated sample rows, the rows separating the sample rows are shown in black. Point measurements were made every 20*m* for the full length of each bed (n=54 for each treatment). Figure Credit: Hazel Smith, University of Southampton.

5.1.2. Pre-processing and Feature Computation

Information from both thermal and visible light images (Figure 5-2) was used for classification purposes. Thermal images were obtained as images with pixel intensity values ranging from 0 to 255. Initially, the image values were transformed to temperature values. A character recognition algorithm based on cross correlation was used, which automatically recognised the characters in the temperature bar (Figure 5-2 (c)) and identified the temperature range for the thermal image [165]. This made it possible to replace the image values, which ranged from 0 to 255, with temperature values. In order to extract useful information from thermal and visible light images, both must be aligned so that the pixel location in both images corresponds to the same physical location with respect to the plant. Since both thermal and visible light images are acquired using a single device, there is a fixed transformation between thermal and visible light images. In order to compute this transformation, the transformation between a single pair of thermal and visible light images was calculated by manually selecting control points. To reduce the amount of noise present in the visible light image, anisotropic diffusion filtering was applied

[155]. These pre-processing steps resulted in the images shown in Figure 5-3 and further calculations were conducted on these images.

In order to get good classification results, we extracted information from the data in the form of features which carry discriminating information from different treatments and similar information from the same treatment type. Features were selected on the basis of observations made by various researchers [9,94–98]. Average values and variation in the thermal profile of the canopy were selected and combined with information from the visible light image. As a first step, the colour space of the visible light image was transformed from RGB to Lab colour space (Figure 5-4). In Lab colour space, instead of Red, Green and Blue channels, an L-channel exists for luminance, as well as 'a' and 'b' channels for the colour components. Features selected for experiments along with their description and justification are given in Table 5-1.

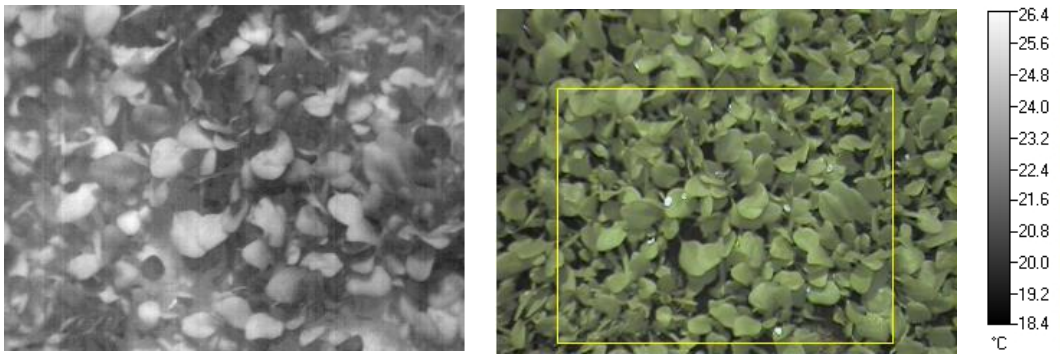
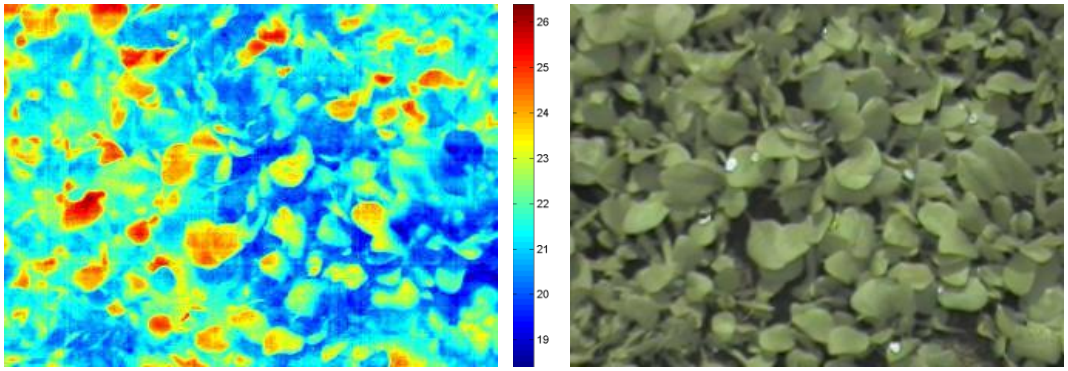


Figure 5-2: Image(s) obtained using a thermal imaging camera (NEC Thermo TracerTH9100 Pro). (a) thermal image with pixel values ranging from 0-255; (b) Visible light image of the scene, with yellow rectangle showing the region corresponding to the thermal image in (a); (c) corresponding temperature range for the thermal image in (a).



(a) (b)

Figure 5-3: Visible light thermal images of Figure 5-2 obtained after pre-processing: (a) the thermal image in Figure 5-2 (a) has been replaced by temperature values; (b) visible light image in Figure 5-2 (b) has been registered to match thermal image in a way that same pixel locations correspond to same point located on the plant.

Table 5-1: Features selected for our experiments. Feature type shows that the corresponding feature contains information about colour (C) or thermal (T) data or both (C/T). The rightmost column shows *p*-values of the features calculated using analysis of variance (ANOVA).

	Symbol	Description	Type	*p*-value
1.	μ_{LT}	Luminance has been found to be a major factor which affects the thermal profile of an image [9]. We linearly scaled the temperature values with the corresponding L-channel of the colour image so that the effect of light intensity was incorporated into the model. After scaling temperature data with the L-channel, mean scaled temperature value of an image was used as a feature.	C/T	0.154

2.	μ_a	The colour information indicates the amount of area covered by the plants or by other types of region. It can be seen in Figure 5-4 (b) that lower intensities correspond to green parts of the plant whereas the background shows a higher intensity value. For this reason, the mean of the a-channel in our set of features was used.	C	1.92×10^{-07}
3.	μ_b	Similar to Feature 2, it can be seen in Figure 5-4 (c) that darker regions correspond to background and hence the mean of b-channel was included in the set of features.	C	1.67×10^{-04}
4.	σ_{nT}	The amount of variation present in an image is also important [96]. Each row of the temperature data was, therefore, normalised by its median and then the standard deviation of the temperature values employed as a feature, to determine the amount of variation in the canopy region covered by the image.	T	2.89×10^{-19}

5.	μ_{aT}	In Lab colour space, lower values in a-channel correspond to green regions. The a-channel was thresholded using Otsu's method [42] to find the background regions as represented by white pixels in Figure 5-4 (d). Temperature values corresponding to the background were discarded and the mean of the temperature values corresponding to the rest of pixels calculated as a measure of the mean temperature of green parts of the plant.	C/T	1.88×10^{-21}
6.	σ_{aT}	Similar to Feature 5, the temperature values corresponding to background were discarded and the standard deviation of temperature values corresponding to the rest of the pixels calculated as a measure of variation in thermal intensities of green parts of the plant.	C/T	1.024×10^{-4}
7.	μ_T	Mean of temperature values	T	1.46×10^{-21}
8.	σ_T	Standard deviation of temperature values	T	1.12×10^{-04}

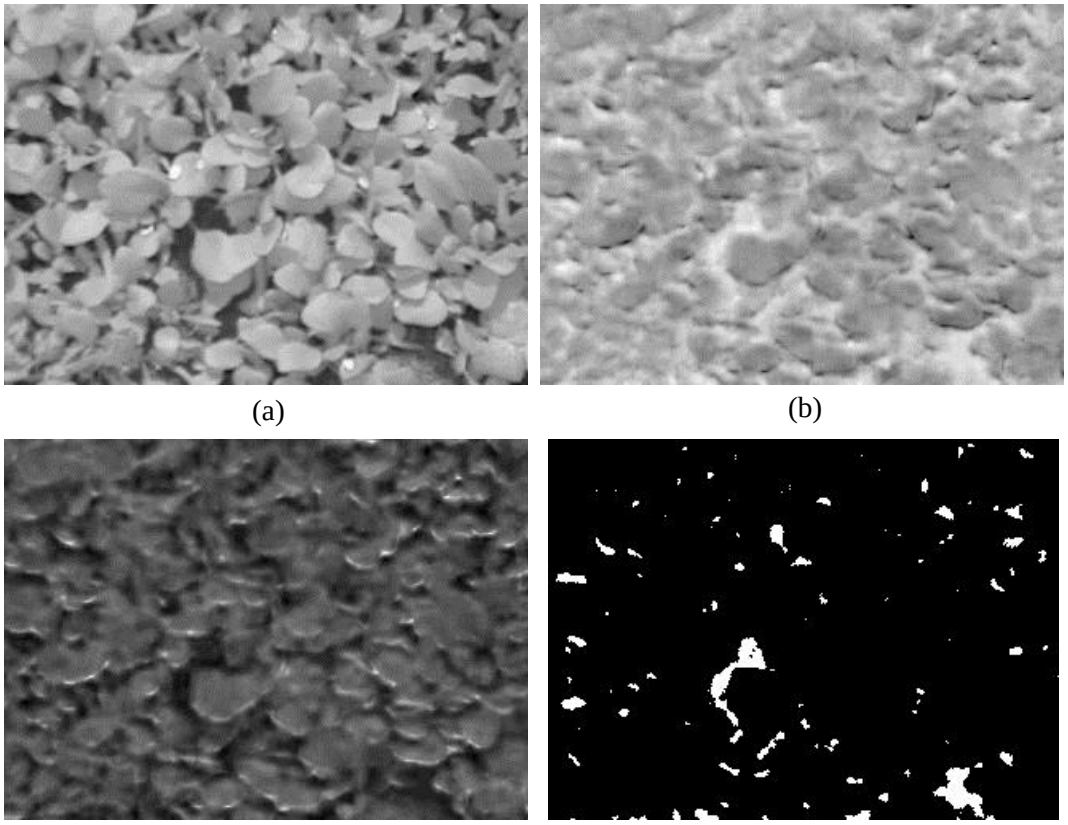


Figure 5-4: (a), (b) and (c) L, a and b channels of the visible light image, respectively; (d) thresholded a-channel.

5.1.3. Classification

5.1.3.1. Support Vector Machines (SVM)

SVM is a supervised learning method used for classification and regression analysis [166]. SVM constructs a hyperplane in high dimensional space and tries to find the hyperplane which maximises the separation between two classes of training data points. In this section, we used linear SVM which uses the model,

$$y = \mathbf{w}^T \mathbf{x} + b \tag{5.1}$$

where $\mathbf{x} = [\mu_{LT}, \mu_a, \mu_b, \sigma_{nT}, \mu_{aT}, \sigma_{aT}, \mu_T, \sigma_T]$ denotes the input feature vector and y denotes the classification output (+1 for plants undergoing water stress, and -1 for

well-watered plants). SVM models the parameters b and $\mathbf{w}$ to find the maximum margin hyperplane between data points from two classes.

5.1.3.2. Gaussian Processes for Classification

Gaussian Processes (GP) can be defined as a class of probabilistic models comprised of distributions over functions instead of vectors [167–169]. A Gaussian distribution can be expressed by a mean vector and a covariance matrix [170]. A GP is fully characterised by its mean and covariance functions. In machine learning, GPs have been used for regression analysis and classification. Similar to SVM, GPC also belongs to the class of supervised classification methods. However, instead of giving discriminant function values, it produces output with probabilistic interpretation, i.e., a prediction for $p(y = +1|\mathbf{x})$ which denotes the probability of assigning a label (y) value +1 to the input feature vector $\mathbf{x}$ [171]. GPC does not calculate this probability directly on the input variables and assumes that the probability of belonging to a class is linked to an underlying of a latent function. Given a training set $D = \{(x_i, y_i)|i = 1,2, \dots, n\}$ consisting of training images of both classes (water deficit and well-watered), with manually assigned labels y_i to the corresponding feature vectors $\mathbf{x}_i$ extracted from those images, GPC makes prediction about the label of the feature vector computed from an unseen image $\mathbf{x}_*$, using posterior probability,

$$p(y_* = +1 | D, \mathbf{x}_*) = \int p(y_* = +1 | f_*)\, p(f_* | D, \mathbf{x}_*) df_* \tag{5.2}$$

The probability of belonging to a class $y_i = +1$ for an input $\mathbf{x}_i$ (known data point) is related to the value f_i of a latent function f [172]. This relationship is defined with the help of a squashing function. In this case, a Gaussian cumulative

distribution function was used as the squashing function.

$$p(y=+1|\ f_i)=\frac{1}{2}\left[1+\frac{erf(y_i f_i)}{\sqrt{2}}\right] \tag{5.3}$$

where $erf(z)$ is the error function defined as $erf(z)=2/\sqrt{\pi}\int_0^z e^{-t^2}dt$. The second term in the integral in equation (5.2) is given by,

$$p(f_*|D,\mathbf{x}_*)=\int p(f_*|\mathbf{X},\mathbf{x}_*,\boldsymbol{f})\,p(\boldsymbol{f}|D)d\boldsymbol{f} \tag{5.4}$$

where $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]$ and $\boldsymbol{f}=[f_1,f_2,\ldots,f_n]$, n is the number of samples. $p(\boldsymbol{f}|D)$ can be formulated by the Bayes' rule as follows,

$$p(\boldsymbol{f}|D)=\frac{p(\boldsymbol{f}|\mathbf{X})}{p(\boldsymbol{y}|\mathbf{X})}\prod_{i=1}^{n}p(y_i|f_i) \tag{5.5}$$

and $p(y_i|f_i)$ can be calculated by equation (5.3) and $p(\boldsymbol{f}|\mathbf{X})$ is the GP prior. Since a GP is characterised by a mean function and a covariance function, a zero mean was used for symmetry reasons, and a linear covariance function selected which has been found to be effective in classification problems [168]. The normalisation term in the denominator is the marginal likelihood given by,

$$p(\mathbf{y}|\mathbf{X})=\int p(\boldsymbol{f}|\mathbf{X})\prod_{i=1}^{n}p(y_i|f_i) \tag{5.6}$$

where $\boldsymbol{y}=[y_1,y_2,\ldots,y_n]$. The second term in the above equation is not Gaussian and this makes the posterior in equation (5.5) analytically intractable. However, analytical approximations or Monte Carlo methods can be used. Two commonly used approximation methods are Laplace approximation and Expectation Propagation (EP). EP minimises the local Kullback-Leibler (KL) divergence

between the posterior and its approximation and has been found to be more accurate in predicting than Laplace approximation and hence EP was used for approximation in these experiments [167,168].

5.1.4. Results and Discussion

A total of 108 images of spinach canopies and corresponding soil moisture point measurements were acquired, with 54 images of well-watered beds designated as treatment A, while the remaining 54 images of the droughted canopy were designated as treatment B. The identity of the two treatments was not known during the development of image analysis. After pre-processing, six different features (1-6, Table 5-1) were obtained from each image. SVM and GPCs were used to classify the test images into one of the two classes: drought and well-watered. For the SVM, linear kernel was used and for the GPC, a zero mean and a linear covariance function were chosen. As discussed before, SVM gives discrete classification results and classifies each image as treatment A or treatment B, whereas GPC gives the probability (likelihood) of each image belonging to a particular treatment. Figure 5-5 shows the probability of an image belonging to treatment B (P_s) versus the values of soil moisture for one set of training and testing data. It was clear that classification results obtained by our classifier were roughly proportional to the manually calculated soil moisture values. Based on the probabilities given by GPC, each image was classified as an image from either treatment A or treatment B.

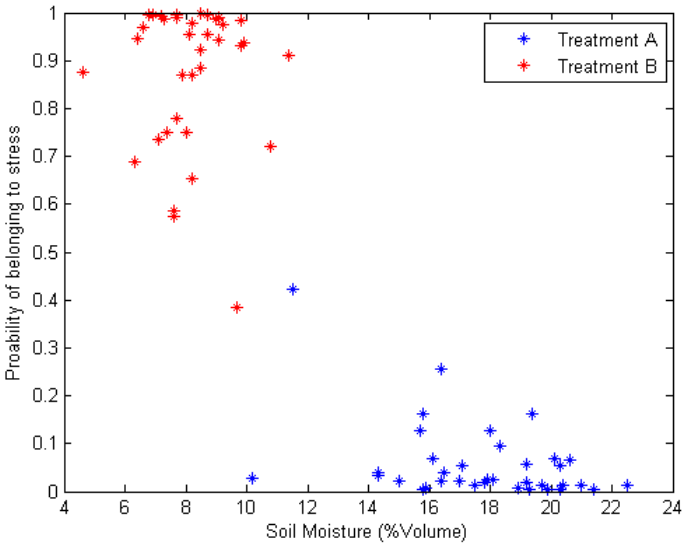


Figure 5-5: Probability of belonging to treatment B (P_s) versus Soil moisture values (correlation value = - 0.89, High moisture means less probability of stress). Classification accuracy for this particular set of training and testing data was 98.62% as given by GPC.

Since two different types of classifiers were used, disagreement between the results of both the different classifiers could be assessed, which occurred in some cases. This disparity was utilised to further refine the classification results, and, although this refinement is not very significant, it produces better results. Information from both classification methods was combined to reduce the error from classification. If an image was classified by SVM as treatment A and its probability of belonging to treatment B according to GPC was higher than 80%, then that image was labelled as treatment B. On the other hand, if an image was classified as treatment B and its probability according to GPC was less than 20%, the image was labelled as treatment A. Note that it was found experimentally that the 80-20% threshold gave the best classification results.

We conducted a total of 200 cross-validation trials to test the accuracy of the classifiers for different pairs of training and testing sets. In each trial, 36 images were chosen at random (18 from each treatment) for training purposes and the

proposed algorithm was tested on the other 72 images. Results showed that GPC demonstrated a higher level of accuracy than the SVM classifier (Table 5-2). However, if information from the results of both the classifiers was combined, results were improved in terms of sensitivity, specificity, positive predictive value (PPV), and accuracy. An average accuracy of 96.27% was obtained for SVM, 96.68% by using GPC and a slightly higher 97.12% when information from both classifiers was combined. When the results of colour-only and temperature-only features were compared, it was found that combining information from both temperature and colour data increased the accuracy of classification. Furthermore, including mean and standard deviation of temperature values without combining them with colour information diminished the accuracy of results; thus the mean (μ_T) and standard deviation (σ_T) were removed from the set of features. In these experiments, it was found that scaling with luminance intensity (μ_{LT}) plays an important role in classification. When the luminance intensity scaling feature was removed from our set of features, we found that the accuracy of the classifiers decreased (Table 5-3). In the case of GPC classification, accuracy fell by up to 7%.

Table 5-2: Comparison of average classification results of different classifiers using our algorithm.

Feature(s) selected	Classifier	Sensitivity (%)	Specificity (%)	PPV (%)	Accuracy (%)	σ_{accuracy}
Colour only (μ_a, μ_b)	SVM	67.28	70.29	70.98	67.74	3.36
	GPC	80.68	52.96	21.68	56.87	3.55
	Both Classifiers	67.32	70.42	71.11	67.80	3.40
Thermal only (μ_T, σ_T)	SVM	93.35	91.28	90.89	92.14	1.92
	GPC	93.06	80.30	76.67	85.42	2.29
	Both Classifiers	93.35	91.28	90.88	92.14	1.92
Features (1-8) Table 1.	SVM	95.52	96.39	96.30	95.85	1.97
	GPC	96.38	97.39	97.30	96.79	1.56
	Both Classifiers	96.62	96.93	96.84	96.70	1.60
Features (1-6) Table 1.	SVM	95.86	96.86	96.80	96.27	1.58
	GPC	96.53	96.99	96.90	96.68	2.00
	Both Classifiers	**96.97**	**97.38**	**97.31**	**97.12**	**1.52**

Table 5-3: Comparison of average classification results of different classifiers without using light intensity scaling feature (μ_{LT}).

	Sensitivity (%)	Specificity (%)	PPV (%)	Accuracy (%)	$\sigma_{accuracy}$
SVM	94.98	95.01	94.83	94.84	2.01
GPC	88.21	91.84	92.05	89.70	2.61
Both Classifiers	95.28	95.27	95.08	95.12	1.89

To further investigate the strength of the classifier with the proposed set of features, we created an artificial image with mixed conditions by combining randomly picked thermal and visible light images from Treatment A and Treatment B to form a mosaic. The ground truth pattern for the mosaicked image is shown in Figure 5-6 (a). Black colour represents image region corresponding to treatment A and white colour represents the image region which corresponds to treatment B. A 50×50 block was defined at each pixel location in the mosaicked image and the classifier was tested using the features extracted from each of these small blocks (307,200 blocks in total). The classifier for this experiment was trained in a similar way as for the real data (i.e., on 36 randomly selected images). It is important to mention here that by using 50×50 blocks to simulate mixed conditions, we reduced the amount of information, so the accuracy of classification is expected to deteriorate. However, the results show robustness of our proposed feature set when compared to thermal only features. The classification results using the combined classifier with thermal only and the proposed feature set are shown in Figure 5-6 (b) & (c) respectively. The classification accuracy using SVM, GPC and the combined classifier was calculated to be **89.10%**, **94.07%** and **92.46%** using the proposed

feature set compared to **78.32%**, **54.10%** and **76.25%** when using thermal only features. The classification accuracy for the combined classifier is less than GPC in the proposed feature set and less than SVM in the thermal only feature set in mixed conditions. However, we still consider this classifier to be important as it gives the best results on real data. Figure 5-7 shows GPC classification results using the proposed set of features in terms of the confidence score (C_s). For treatment A, $C_s = 1 - P_s$ and for treatment B, $C_s = P_s$, where P_s is the probability of belonging to treatment B as given by the GPC. The bright shade represents high confidence in classification results and dark shade represents low confidence in the classification. It can be observed that the classifier has higher confidence in the region where the image is from treatment A or treatment B. However, the confidence value is low, as depicted by low grey values around the boundary of two merging images from different treatments. The mean and standard deviation of C_s was calculated to be **90.48%** and **17.79%** using the proposed feature set and **51.12%** and **32.33%** using thermal only features respectively.

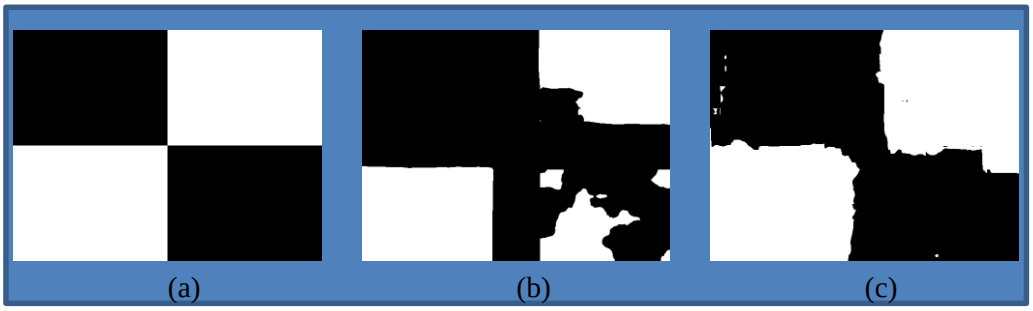


Figure 5-6: (a) The ground truth pattern for mixed condition mosaicked image. Black colour represents image region corresponding to treatment A and white colour represents the image region which corresponds to treatment B; (b) & (c) show classification results obtained using combined classifier with thermal only and the proposed feature set respectively.

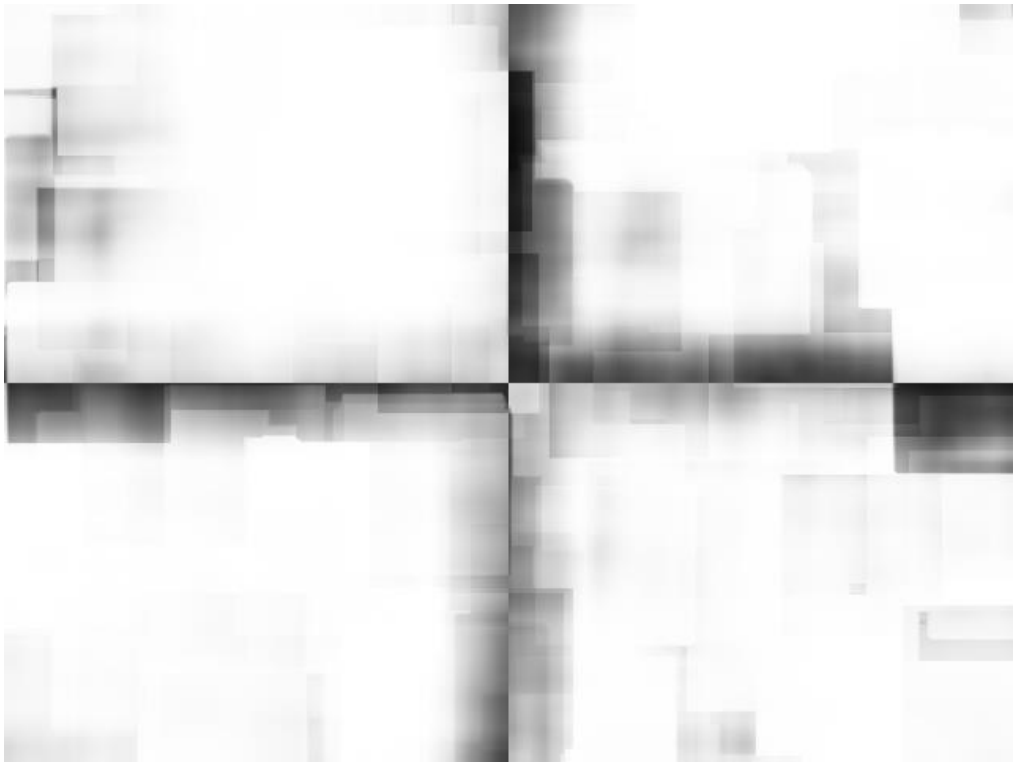


Figure 5-7: GPC classification result in terms of confidence score (C_s). Bright shade represents high confidence in classification results and dark shade represents low confidence in the classification. The classifier has higher confidence in the region with image from treatment A or treatment B, however the confidence value is low, as depicted by darker shade, around the boundary of two merging images from different treatments.

5.2. Disease Detection in Plants

In the previous section we combined the information from thermal and visible light images for stress detection in a commercial spinach canopy. In this section, we add depth information to temperature and colour information and present results of our study on thermal and stereo visible light images of tomato plants infected with powdery mildew. The images for this experiment were collected as described in section 3.1. The experiment was performed on 71 plants and the images were collected for 14 consecutive days (day 0 to day13). Out of these 71 plants, 17 plants were normal and were not inoculated with the disease whereas the remaining 54 plants were inoculated with *Oidium neolycopersici*, which causes the powdery mildew

disease in tomato plants. In this thesis, we present two different approaches for classification of diseased plants after adding depth information in the form of disparity from Chapter 4.

5.2.1. Pre-Processing

For our experiments, we first register colour (visible light) images with the corresponding thermal images as described in Chapter 3, we apply the same transformation to the disparity image, consequently the same pixel location in all the three images (thermal, colour and disparity) approximately correspond to the same physical point in the plant. As a first pre-processing step, we remove the background to obtain an image which contains only plant regions. This step was not required in case of water stress experiments, where the image consisted of mostly the canopy and very little of the background. To remove the background, we train an SVM classifier with a linear kernel using the RGB pixel values and classify each pixel into background/plant pixel. The result of extracting the plant region using our method on an image is shown in Figure 5-8 (b).

(a) (b)

Figure 5-8: (a) Colour image registered with thermal image; (b) Colour image obtained after background removal.

5.2.2. The Classification Approaches

We present two different approaches to detect diseased plants:

1. Pixel level classification approach
2. Region based classification approach

5.2.2.1. Pixel Level Classification Approach

This is a two-step classification approach in which we directly use pixel values to first roughly classify the plant regions into normal and diseased regions and then we classify whole plant into normal and diseased plant using the information extracted from the potential diseased regions. For the first step, we first convert the colour space of the image in Figure 5-8 (b) to Lab and directly use the pixel values corresponding to 'a' and 'b' channels. Similarly, we change the RGB colour space of the colour image to CMYK and use pixel values corresponding to 'C' & 'Y' channels. For depth and temperature (T) information, we directly use pixel values in disparity (D) and thermal images, respectively. Therefore, our classification algorithm uses a six dimensional feature vector **V** consisting of a, b, C, Y, D & T values at each pixel location to classify a pixel into normal or diseased pixel. For this purpose, we train the SVM classifier with radial basis function (RBF) kernel with $\sigma = 1$ to classify plant regions in an image into normal and diseased regions. The result of classifying plant regions into normal and diseased for the image in Figure 5-8 (b) is shown in Figure 5-9 (a) and Figure 5-9 (b) respectively.

For classification of a plant into normal or diseased plant, we assume that the normal plants carry less variation in the aforementioned feature measurements whereas the diseased plants carry large variation in the same measurements. Based

on this assumption, we can classify a plant into normal or diseased plant. It is possible that some regions in normal plants can be erroneously classified as diseased regions in the first step of our classification approach. According to our assumption, if a region in a normal plant is incorrectly classified as diseased, it will have less variation whereas a correctly classified diseased region will have high variation. Let us place all the feature vectors corresponding to the diseased regions in Figure 5-9 (b) in a matrix V and perform the principal component analysis (PCA) on V to test our hypothesis. We can compute the standard deviation of projection of features vectors along the first and second principal components as σ_{p1} & σ_{p2} respectively. The smaller values of σ_{p1} & σ_{p2} in Figure 5-10 for normal plants validates our assumption that there is low variation in data for normal plants compared to diseased plants, therefore we can classify the images on the basis of this information.

(a) (b)

Figure 5-9: The result of classifying plant in Figure 5-8 (b) into (a) normal and (b) diseased regions.

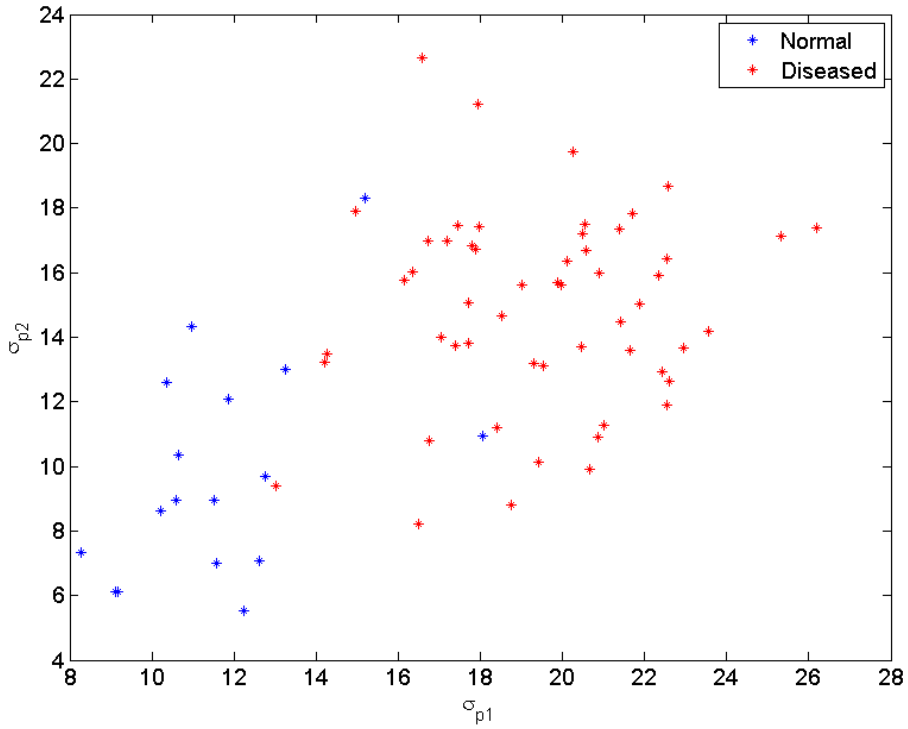


Figure 5-10: A scatterplot of the standard deviation (σ_{p1} & σ_{p2}) of projection of feature vectors, corresponding to diseased regions for normal and diseased plants (Day 13), plotted along first and second principal components respectively.

5.2.2.2. Results and Discussion

From the total of 71 plants, 54 plants were diseased and 17 plants were normal (not inoculated with the fungus). To test the strength of our features, we used SVM classifier with RBF kernel where $\sigma = 3$. We ran 200 cross-validation trials and tested the classifier using random pairs of training and testing data. In each trial, we randomly picked 17 out of 54 diseased plants for classification purpose. Once the number of diseased and normal plants was equal, we randomly picked 7 out of 17 normal and diseased plants each for training purpose and the remaining 10 for testing the classifier. The disease starts to appear 7 days after inoculation and, therefore, we concentrate on classification results for day 5 to day 13 after inoculation in this thesis (see for example, Figure 5-11). The classification results of the proposed classifier for 200 trials in terms of average

accuracy, sensitivity, specificity and positive predictive value (PPV) are shown in Figure 5-12. Figure 5-12 indicates that we can achieve an average accuracy of more than 75%, 9 days after inoculation. The highest average accuracy achieved in this case is on day 13 i.e., 89.93%, which is very significant. However, as the disease starts to appear 7 days after inoculation detecting the disease after day 9 is not very beneficial at the commercial level as it might spread across the crop, therefore we propose another approach for classification of diseased plants in the next section.

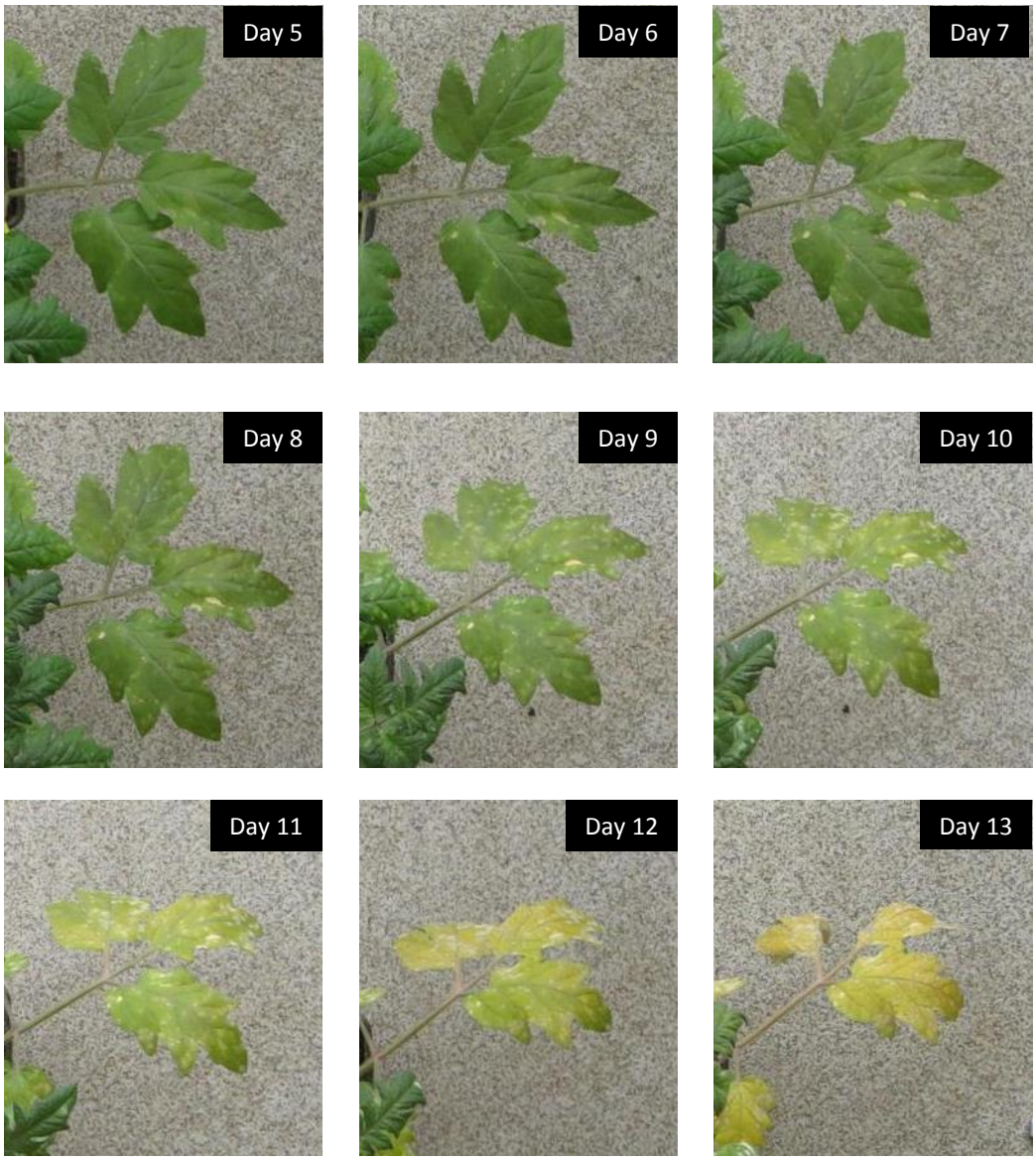


Figure 5-11: The appearance of disease symptoms with time on leaves of a diseased plant.

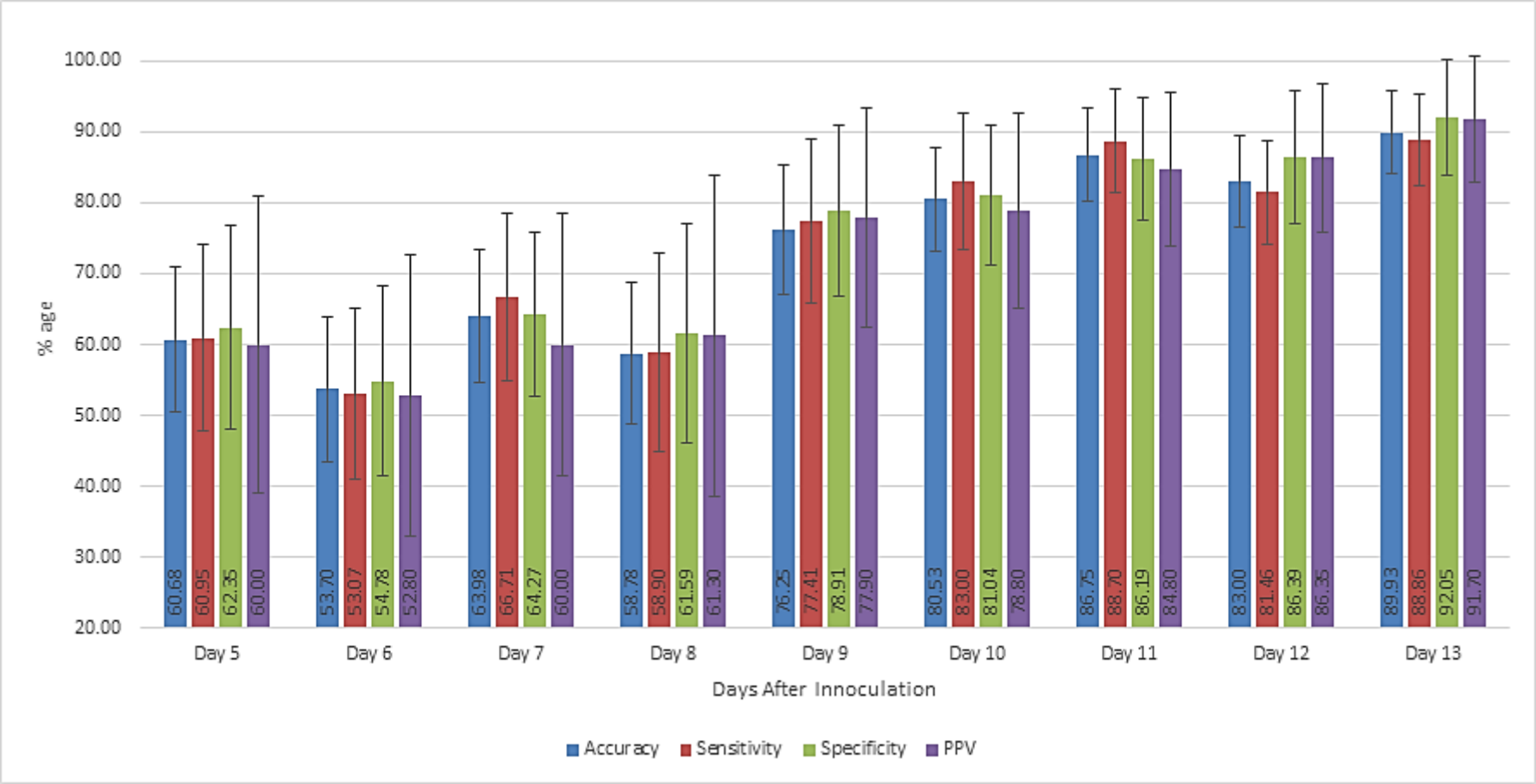


Figure 5-12: Average Accuracy, Sensitivity, Specificity and positive predictive value (PPV) results using the two-step pixel level classification approach.

5.2.2.3. Region based Classification Approach

This classification approach is similar to the approach in section 5.1 in that we directly compute features from the plant regions (Figure 5-8 (b)). The following six features were selected on the basis of *p*-values computed using analysis of variance (ANOVA) for different days after inoculation as shown in Table 5-4.

μ_C	Convert RGB image to CMYK and compute mean of 'C' channel.
μ_Y	Convert RGB image to CMYK and compute mean of 'Y' channel.
σ_{sca}	Convert RGB image to Lab and compute standard deviation of the corresponding thermal image after scaling with 'a' channel.
σ_{scL}	Convert RGB image to Lab and compute standard deviation of the corresponding thermal image after scaling with 'L' channel.
σ_D	Compute standard deviation of disparity corresponding to plant region.
μ_D	Compute mean of disparity corresponding to plant region.

The Cyan and Yellow in CMYK carry the green colour, yellow is very important because the leaf infected with powdery mildew turns yellow after showing white spots. Therefore, the presence of yellow colour can be directly translated to disease. The next two features carry temperature information where temperature information is scaled by Luminance and 'a' channel. As we have discussed already in Section 5.1, luminance is important and lower values of 'a' carry information about the green-ness of the pixel. The last two features carry depth information (in terms of disparity), the standard deviation of disparity must be higher in the diseased plants because of irregular leaves.

Table 5-4: *p*-values of the selected feature set for day 5 to day 13 after inoculation computed using ANOVA.

	μ_C	μ_Y	σ_{sca}	σ_{scL}	σ_D	μ_D
Day 5	7.59×10^{-01}	6.73×10^{-01}	4.17×10^{-06}	5.98×10^{-01}	2.06×10^{-02}	1.58×10^{-07}
Day 6	3.02×10^{-01}	4.61×10^{-01}	1.41×10^{-06}	8.15×10^{-01}	5.33×10^{-03}	8.71×10^{-08}
Day 7	1.14×10^{-01}	1.86×10^{-01}	9.07×10^{-06}	3.36×10^{-01}	6.85×10^{-04}	2.32×10^{-07}
Day 8	1.89×10^{-02}	8.53×10^{-02}	2.18×10^{-05}	1.08×10^{-01}	1.30×10^{-04}	5.36×10^{-06}
Day 9	4.39×10^{-04}	9.40×10^{-03}	3.78×10^{-05}	4.54×10^{-04}	1.14×10^{-07}	5.52×10^{-05}
Day 10	3.66×10^{-05}	8.54×10^{-05}	1.37×10^{-05}	4.07×10^{-05}	9.71×10^{-09}	4.51×10^{-03}
Day 11	3.77×10^{-05}	2.66×10^{-06}	1.94×10^{-07}	8.37×10^{-07}	7.49×10^{-11}	6.16×10^{-03}
Day 12	5.27×10^{-06}	4.09×10^{-09}	3.46×10^{-06}	3.12×10^{-10}	6.52×10^{-13}	1.98×10^{-01}
Day 13	1.35×10^{-06}	1.23×10^{-10}	5.47×10^{-05}	5.80×10^{-11}	1.48×10^{-12}	2.84×10^{-01}

5.2.2.4. Results and Discussion

We use the same classifier and the evaluation procedure as in section 5.2.2.2, i.e., we use SVM with RBF kernel (σ =3) and we use 7 images from each group for training and 10 images for testing for 200 cross-validation trials. Let us divide our analysis to colour only (μ_C, μ_Y), colour + thermal ($\mu_C, \mu_Y, \sigma_{sca}, \sigma_{scL}$), colour + depth ($\mu_C, \mu_Y, \sigma_D, \mu_D$), and colour + thermal + depth ($\mu_C, \mu_Y, \sigma_{sca}, \sigma_{scL}, \sigma_D, \mu_D$) to test how combining these different sets of features helps us to improve classification results. From Figure 5-13, we can see that if we use only colour information we achieve accuracy of over 70% only after day 10 of inoculation. We can increase this accuracy by combining colour information with thermal or depth, over 70% accuracy after day 9, which is an improvement but again is not very beneficial to use at commercial scale. Combining the features from colour, thermal and disparity images increase the accuracy of our classifier to more than 70% on day 5. Let us compare average accuracy

of colour + thermal + depth feature set in Figure 5-13 with average accuracy results in Figure 5-12. Average accuracy results in Figure 5-13 clearly outperform results in Figure 5-12. However, we can combine the features in this section with σ_{p1} & σ_{p2} to make a more robust classifier. The average accuracy results using the combined feature set are shown in Figure 5-14. Although average accuracy results of classifier are slightly less on day5, day6, day8 and day13, the combined feature set produces more stable results as shown in Table 5-5. The standard deviation of classifier accuracy using combined feature set is lower than the other two approaches. Figure 5-15 shows the results of average accuracy, sensitivity, specificity and PPV for 200 iterations using the combined feature set.

Table 5-5: Standard deviation of accuracy results for 200 cross-validation trials for feature sets in Figure 5-14. Low standard deviation for combined feature set shows more stable performance.

	Combined feature set	Region based Classification	Pixel Level Classification
Day 5	**8.54**	9.27	9.31
Day 6	**8.78**	9.12	10.22
Day 7	**7.79**	8.14	8.94
Day 8	**8.82**	8.86	10.43
Day 9	**7.61**	8.30	7.58
Day 10	**6.77**	7.70	7.80
Day 11	**5.68**	6.36	7.00
Day 12	**5.53**	6.22	7.43
Day 13	**5.22**	5.82	5.41

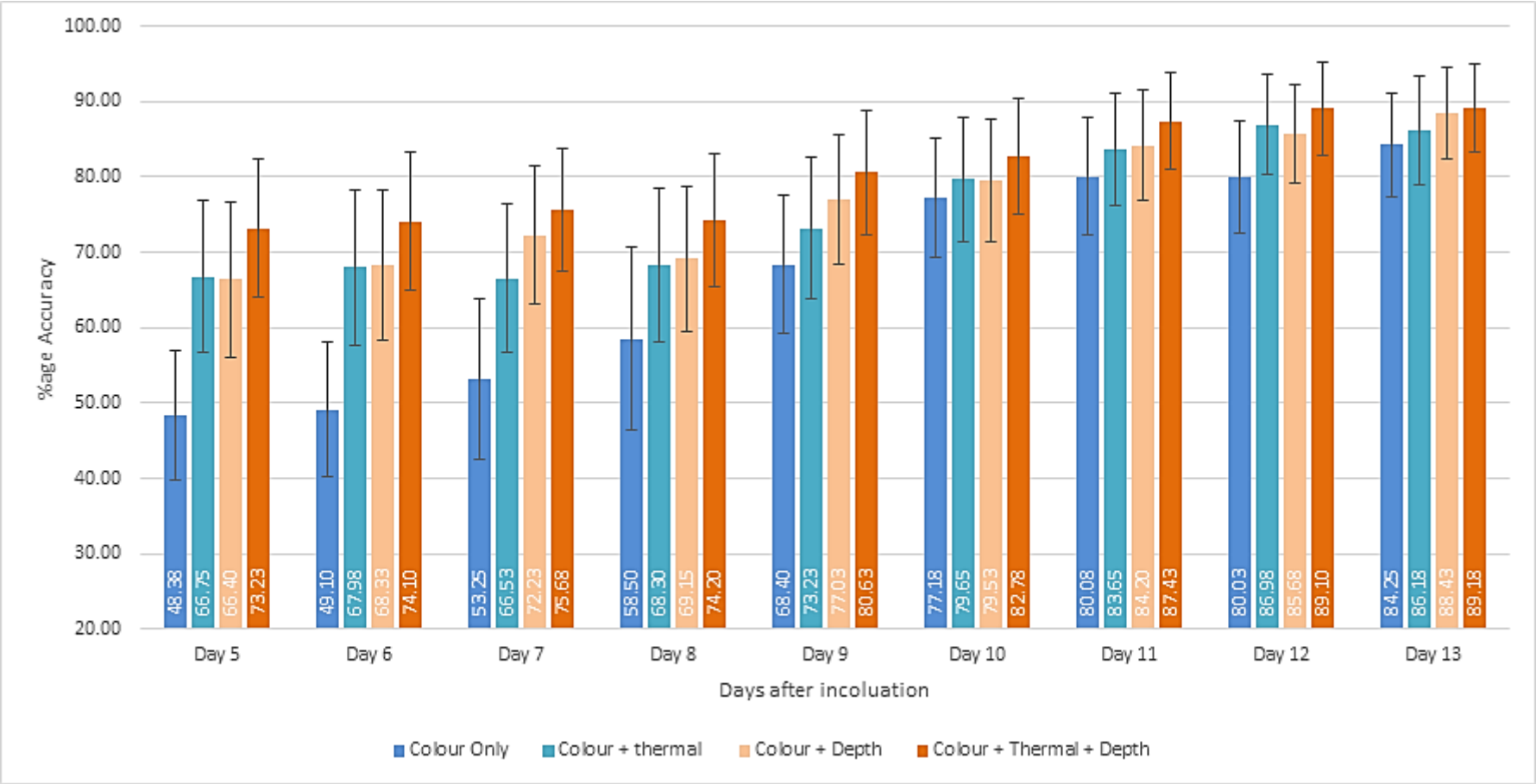


Figure 5-13: Accuracy of classifier using different set of features. Combining colour information with thermal or depth slightly increases the accuracy of the classifier, however combining colour information with thermal and depth improves the accuracy to more than 70% on day 5.

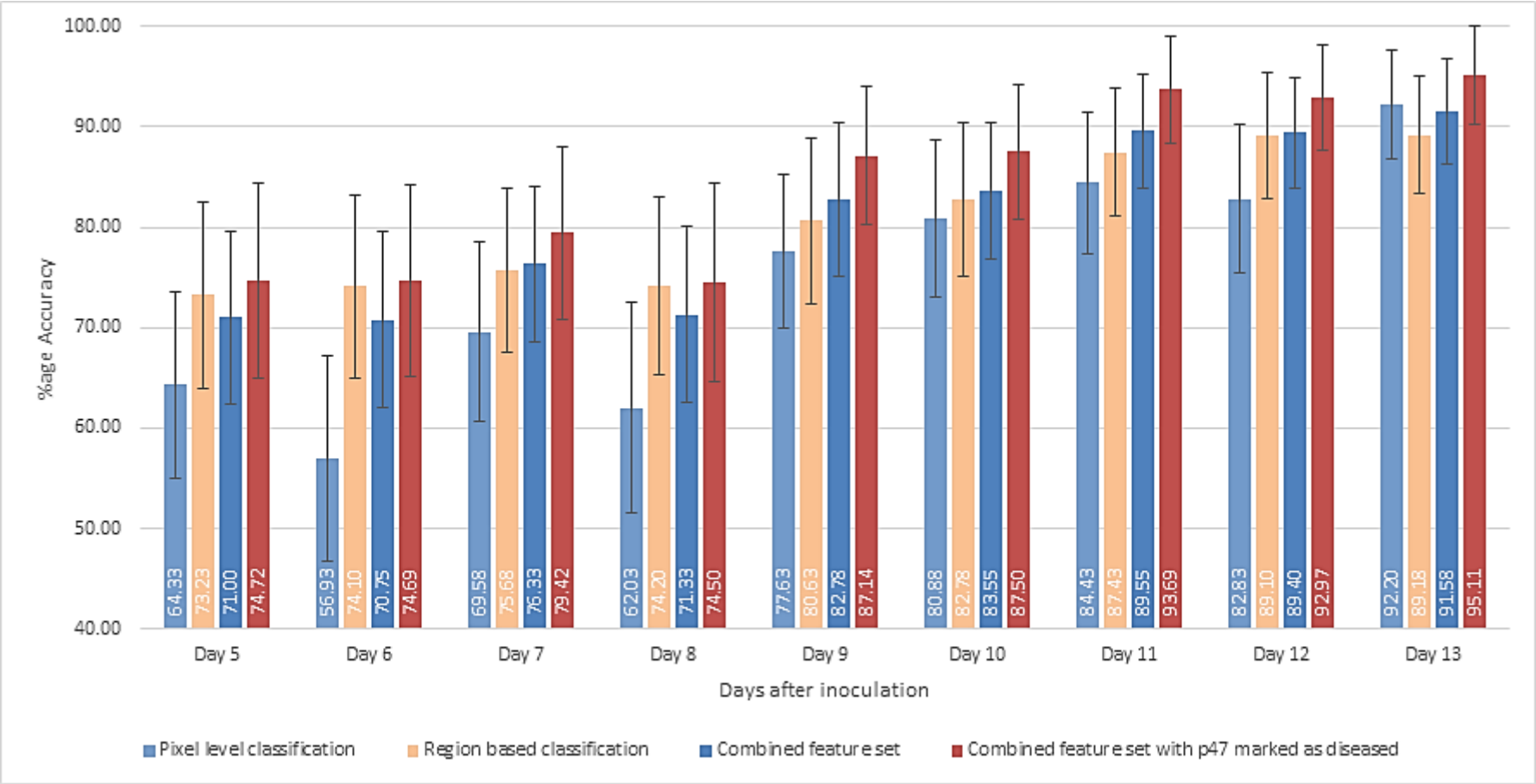


Figure 5-14: Average accuracy results using the combined feature set compared to average accuracy results of both the previous approaches.

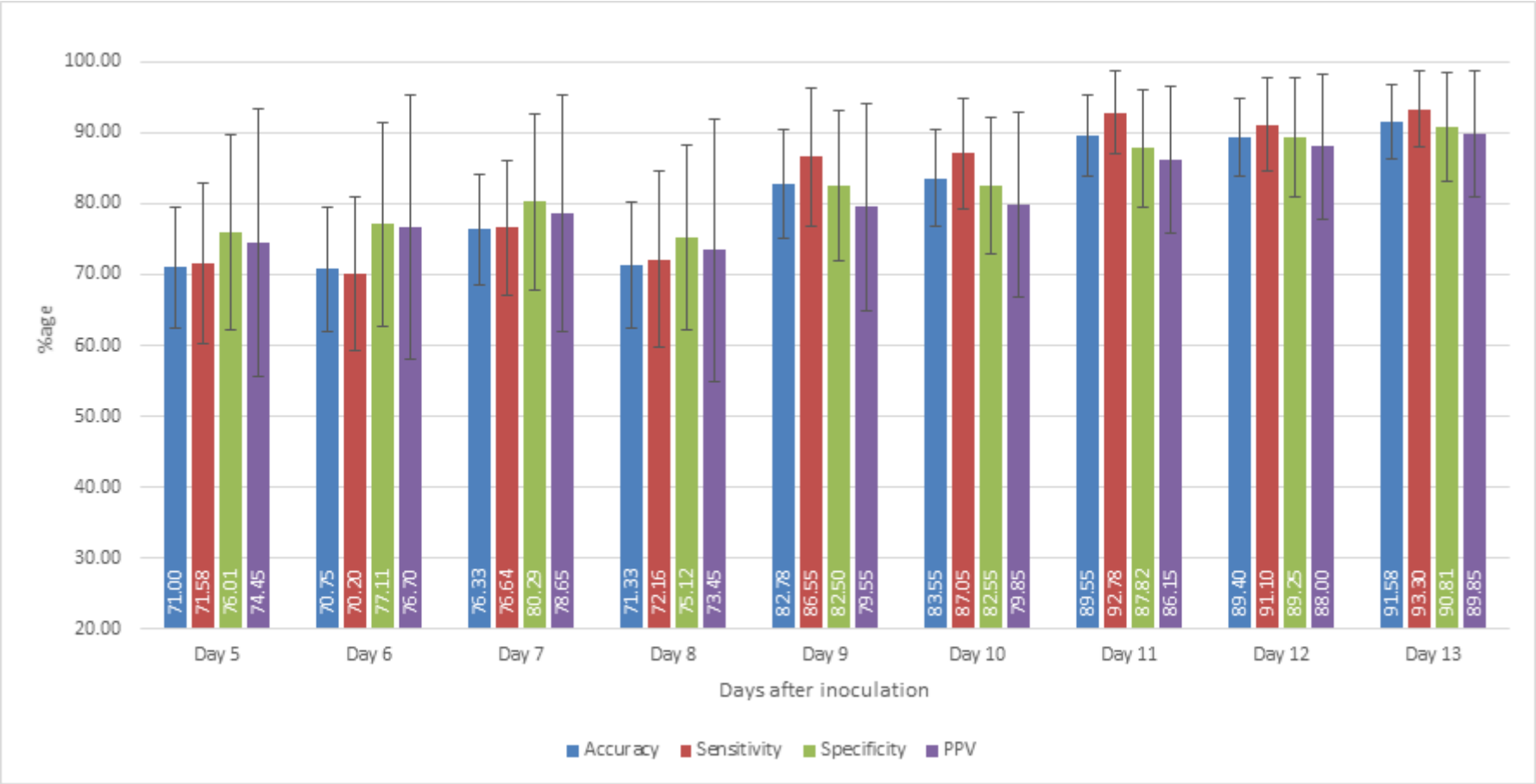


Figure 5-15: Average Accuracy, Sensitivity, Specificity and positive predictive value (PPV) results using the combined feature set.

Let us consider projection of combined feature sets using PCA on 1st and 3rd principal component, as shown in Figure 5-16. The projection shows feature values corresponding to a couple of normal plants in diseased regions. One such plant is marked as p47 and is shown in Figure 5-17. This plant was initially not inoculated with disease but showed symptoms of disease later during the experiment. If we mark this plant as diseased and then run the proposed classifier for 200 random cross-validation trials, we can achieve an average accuracy of more than 95% on day 13 as shown in Figure 5-14. The identification of a diseased plant among the non-inoculated plants shows the quality of our features set and reliability of the proposed classification method.

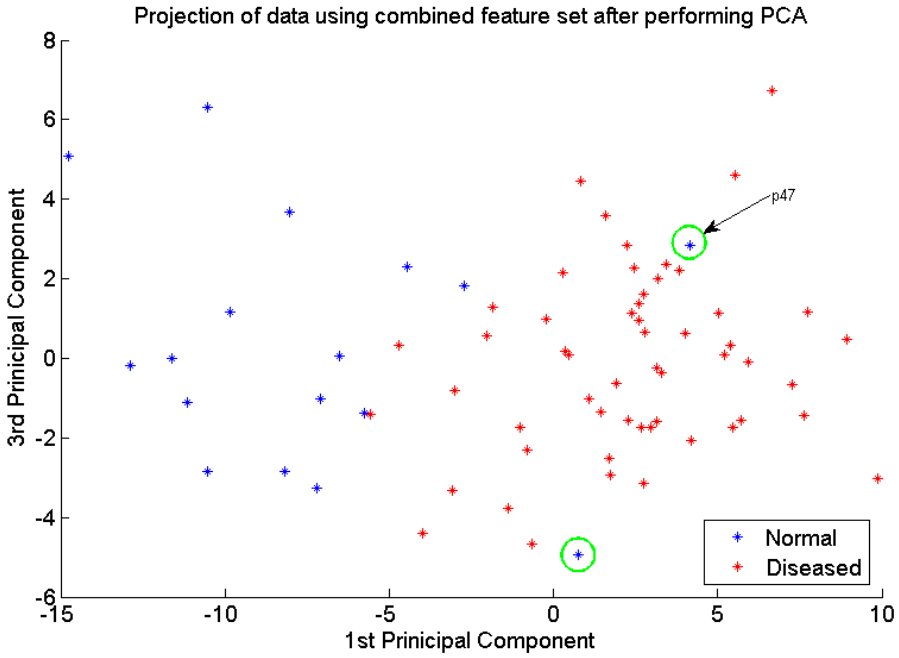


Figure 5-16: Projection of combined feature set on 1st and 3rd principal component after performing PCA. The projection shows feature values corresponding to some of the plants which were not inoculated with any disease, occur in disease regions. One of these plants is marked as p47 and is shown in Figure 5-17.

Figure 5-17: p47 shown for illustrative purpose the plant was not inoculated with any disease but later showed symptoms of the disease. These plants were successfully captured by our feature set.

Chapter Summary

Our results show that by combining information from thermal and stereo visible light images and using machine learning techniques, canopies which are experiencing water deficits or under threat of a disease can be identified with high accuracy – more than 95%. Thus we have considerably improved the use of remote images in the detection on canopy stress and disease onset using this combined approach. This improvement can be translated to huge savings and fewer losses in a commercial setting. Our approach consisted of three steps: removal of background regions, extraction of feature sets from plant regions, and classification using the extracted feature set. We showed that extraction of a good set of features can be useful for classification of water deficient and diseased plants. We can conclude from our results that based on information from stereo visible light and thermal images, a worthy set of features can be extracted which can be used at commercial scale for detecting anomalous regions in crops.

Chapter 6
Registration and Preliminary Analysis of Multi-tag Bio-images

We have shown in previous chapters that by combining multi-modal images we can increase the accuracy of classifiers for anomaly detection in plants. In this chapter, we perform multi-channel image analysis at microscopic level for detection of cancer cells in human tissue. The techniques discussed in this chapter not only allow us to differentiate between cancer and normal cells but can also be used for better understanding of cancer. Similar to the pre-processing steps required to analyse multi-modal images of plants, the multi-channel microscopic image data requires alignment of images prior to analysis. We present a novel framework, Robust Alignment of Multi-tag Bio-images (RAMTaB), to align images in a multi-tag fluorescence microscopy image stack. The proposed framework is applicable to multi-tag bio-imaging systems which (a) acquire fluorescence images by sequential staining and (b) simultaneously capture a phase contrast image corresponding to each of the fluorescence images. In addition, we derive a shift metric in order to select the Reference Image with Maximal Overlap (RIMO) in a stack of hundreds of images, in turn minimising the total amount of non-overlapping signals for a given number of tags.

The data used for study in this chapter was acquired using a multi-tag fluorescence imaging microscope known as Toponome Imaging System (TIS). TIS uses N cycles of fluorescent tagging and bleaching to produce $2N$ pairs of fluorescent and phase contrast images. As illustrated in Figure 1-5, N pair of images are captured before,

and the remaining *N* pair of images are captured, after tagging the tissue with the fluorescent tag. The latter *N* pairs contain the actual protein signal and the former contain the auto-fluorescent signal. For simplicity we use the notation *N* for number of pair of images in this text.

As discussed in Chapter 1 and Chapter 2, misalignment can be observed between images of even a single visual field. For direct analysis of co-located signals, alignment of images from different tags is necessary. We did not observe any significant misalignment between fluorescence and corresponding phase contrast images. However, we observed misalignment between phase contrast images for different antibody tags. The misalignment is manifested in terms of translational shifts. Other forms of misalignment, such as rotation, do not appear in our context, and we assume that all alignment transformations are translations. Figure 6-1(a) & (b) show misaligned composite RGB colour images made up of CD57, CD166 and DAPI (DAPI binds to nuclei, while CD57 and CD166 are protein markers) tags displayed in red, green, and blue channels respectively.

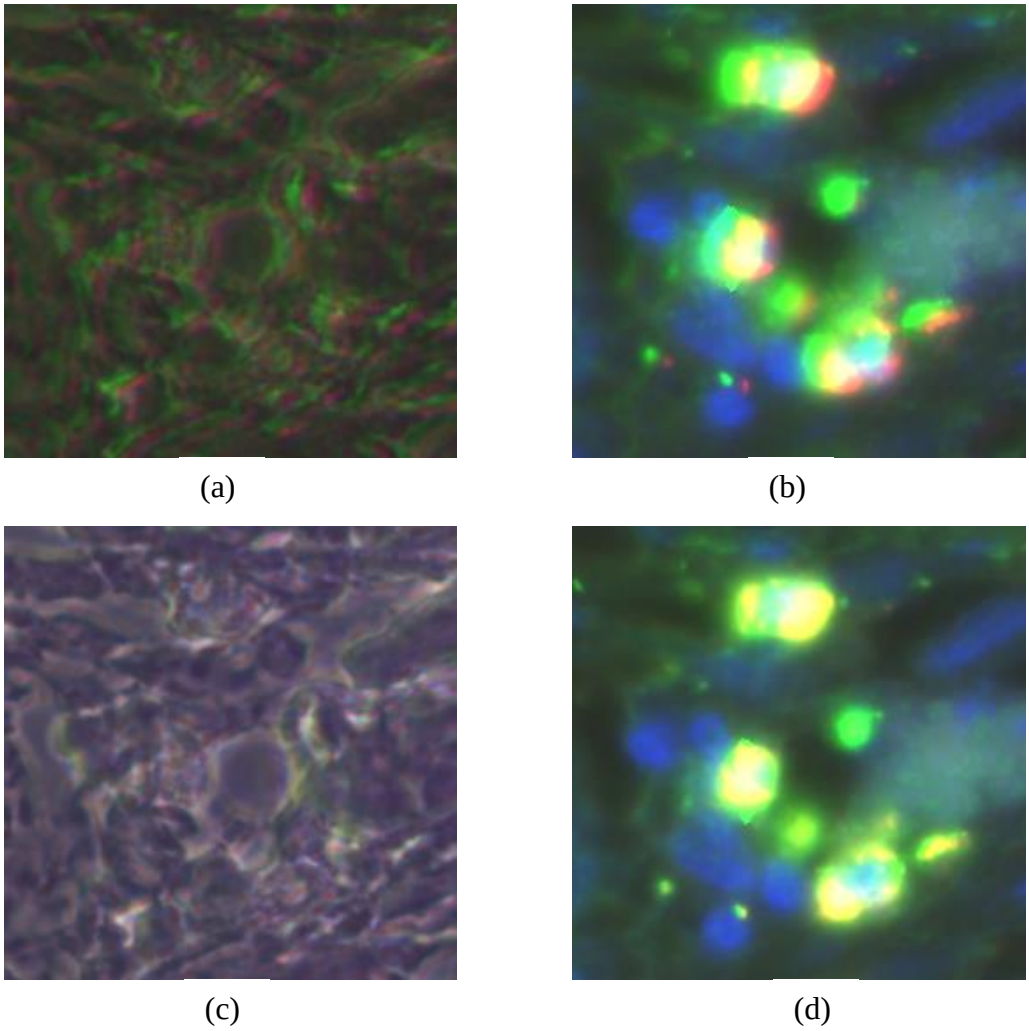


Figure 6-1: The RGB composite image before and after registration: R,G and B channels belong to CD57, CD166 and DAPI tags respectively. (a) and (b) respectively show composite image formed by using phase images I_j and fluorescence images F_j, before alignment, (c) and (d) show the RGB composite images after the images were aligned using the proposed framework. The red colour fringes in (a) show the degree of misalignment among the three tags. These colour fringes have been replaced by white and grey regions in (c) after alignment.

6.1. Image Acquisition

The human tissue was collected from operative samples at the George Eliot Hospital, Nuneaton, UK. For TIS imaging, tissues were incubated in sterile Phosphate Buffered Saline (PBS), in PBS containing normal goat serum, and then washed in PBS. See [15,173] for more details. The images were acquired by Sylvie Abouna under supervision of Dr Michael Khan using the TIS machine installed at

the University of Warwick. A library of 26 tags [174] consisting of a variety of cell specific markers, control tags, nuclei marker and tumour and stem cell markers was employed for imaging. Tags were applied sequentially to the tissue section. In TIS cycle, an image is acquired before and then again after incubation with each fluorophore-conjugated antibody or other fluorescent dye, and washed to remove unbound tags. Each image is captured at 63× and has a spatial resolution of 1056×1027 pixels, where each pixel has a resolution of approximately 200nm. Non-destructive photo bleaching clears the fluorescence after each tag incubation once the image has been acquired. The cycle of incubation, wash, image acquisition and photo bleach is repeated then for another tag (see Figure 1-5).

6.2. The Proposed Registration Framework (RAMTaB)

The proposed framework for multi-tag fluorescence image registration has three sequentially connected components in the following order: registration of all tag images in the stack to a tag image with the best focus [175] as a reference image, selection of RIMO, and re-alignment of all images in the stack to RIMO. Below we describe the core registration algorithm based on mutual information used by the first and the third components. A side benefit of RAMTaB is that even if the arbitrarily chosen reference image is different from the RIMO, the core registration algorithm does not have to be executed again.

6.2.1. Aim(s)

The overall aim of this work is to compute transformation (in terms of translational shift) parameters for each tag image in a stack, such that a) the images are well aligned and b) the total number of non-overlapping pixels φ is minimised. The total

number of non-overlapping pixels which can also be termed as loss of information φ may vary from one reference image to another. Suppose we have a stack of N images $I_1, \ldots, I_N$, all of the same scene, though possibly not perfectly aligned with each other. We choose a reference image I_r, and then for each floating image I_j we find transformation τ_{rj} so that each point on $T_{rj} = \tau_{rj}(I_j)$ for all j=1,2,…N corresponds to one and the same point in the tissue specimen being imaged. We will assume that the alignment transformations τ_{rj} are always translations, which is a reasonable assumption in the situation to which we will apply our theory. The aligned images can be mosaicked and arranged in a larger frame of reference as shown by the green dashed line in Figure 6-2.

There may be several ways in which the registration results can be used for a follow-up analysis. Here we consider four possible options. A first option is that we select only those co-ordinates where all the signals overlap as shown by $C_\cap$ in Figure 6-2. Note that we use the word *coordinates* instead of *pixels* because we wish to allow alignment transformations τ_{rj} with sub-pixel accuracy. We consider all the aligned images in a universal co-ordinate system and let C_j denote the set of co-ordinates of T_{rj}. We define the set $C_\cap$, shown as the yellow shaded area in Figure 6-2, as the set of co-ordinates of the overlapping region as follows,

$$C_\cap = \bigcap_{j=1}^{N} C_j \tag{6.1}$$

A major limitation of using $C_\cap$ is that this set may be very small if we have a few images that are far from the centre. We risk losing a large amount of signal in the non-overlapping region, when we can benefit from this signal at the cost of

eliminating from consideration only a few of our images.

A second option is that we select signals from all the aligned images and form a larger mosaicked image as shown by the green dashed line in Figure 6-2. Let C_0 be the set of co-ordinates of empty spaces in the larger mosaicked image, and let C_{U} be the set of co-ordinates of all the aligned images in the dataset as shown by shaded region with different colours in Figure 6-2. Let *C* be the set of co-ordinates of this larger rectangular image with green dashed boundary. Then

$$C = C_0 \bigcup C_{\mathrm{U}} \tag{6.2}$$

where

$$C_{\mathrm{U}} = \bigcup_{j=1}^{N} C_j \tag{6.3}$$

In this case, there may be a substantial region covered by only a few images, or even by no image at all, as in C_0 rendering the post processing of those elements potentially meaningless.

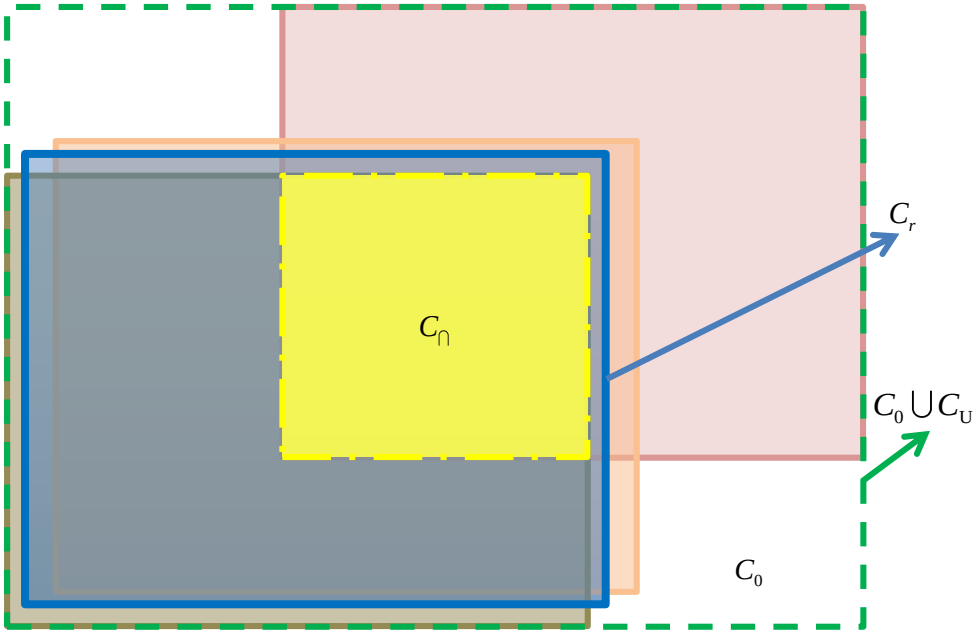


Figure 6-2: An illustration of mosaicked image containing co-ordinates of all the aligned images.

Let us consider another scenario and fix r with $1 \leq r \leq N$, and let us restrict our analysis to the region C_r. The signal from the ith image comes only from the region $C_r \cap C_i$. Then

$$\psi_r = \sum_i Area(C_r \bigcap C_i) \tag{6.4}$$

represents the sum of the areas overlapped by C_r that provide meaningful signal. Then to find RIMO, we find C_r corresponding to I_r with $1 \leq r \leq N$ such that ψ_r is maximal. Selection of RIMO has been further discussed in section 6.2.3.

6.2.2. The Core Registration Algorithm

Let us first consider different approaches used for registration of images for our data. As the fluorescent channels contain information captured by a particular tag, the information varies in fluorescent images across different tags, which translates to unstable features and therefore fluorescent images as floating or reference images

may not be a good choice for registration. We use instead phase contrast images corresponding to different tags for registration as they contain information about the structure of the tissue and provide more stable features across different tags. The intensity of phase contrast also varies across different tags and sometimes there are some image which are blur and may have local wear and tear. Feature based methods which rely on corner or edge detection fail to provide stable features in the presence of blur across different tags. Therefore, area based methods provide a more suitable choice for registration.

The choice of similarity measure is an important factor in the accuracy of registration results produced by area based methods. Various similarity measures such as $l1$-norm and cross correlation require intensities to be perfectly matched however this might not be the case with the phase contrast images as the intensity varies across different tags. We employ a mutual information-based framework [38] for registering one phase contrast image with another. Mutual information is based on the principle that for a registered pair of images, we have less uncertainty about the information contained in the reference image when we have information about the floating image and vice-versa. Several other researchers have shown mutual information to be a good similarity measure for microscopic images [81,176,177].

Mutual information based on Hartley's entropy measure is defined as follows. Let $H(I_A)$ and $H(I_B)$ denote the entropies of I_A and I_B, respectively, and let $H(I_A, I_B)$ be the joint entropy of I_A and I_B. Then $M(I_A, I_B)$ the mutual information between the two images I_A and I_B is defined by the formula

$$M(I_A, I_B) = H(I_A) + H(I_B) - H(I_A, I_B) \tag{6.5}$$

One approach to registering the two images is to maximise $M(I_A, \tau_{AB}(I_B))$, by varying τ_{AB} over some set of transformations. Maximising mutual information implies minimising the joint entropy. Marginal and joint entropy can be calculated from the joint histogram, which is formed using the intensity values of the two images. When mutual information is high, the joint histogram is sharp and closely resembles a diagonal matrix. In a mutual information based registration framework, we transform the floating image I_B to match the reference image I_A by searching for a transformation which maximises the mutual information between the reference image and the transformed floating image. Mathematically, this can be written as,

$$\tau_{AB} = \arg\max_{\tau} M(I_A, \tau(I_B)) \tag{6.6}$$

where τ denotes the transformation between reference and floating images required to align them. The optimisation is done using the *pattern search* method [156,178]. At each step, the search algorithm creates a set of points called a *mesh* around the optimal point of the previous step. The pattern search finds a point that improves the objective function. If the algorithm fails to find such a point, it decreases the size of the mesh, otherwise it chooses the new point which has improved the objective function as the new optimal point and increases the size of the mesh in the next step. This search continues until $|\Delta\tau|$ is less than a specified threshold $\theta_{\Delta\tau}$ or the number of iterations reaches the maximum allowed number of iterations. In general, the transformation τ could consist of affine and perspective transformations. In our case, however, rotations and non-rigid transformations are not required, and therefore we are only concerned with horizontal and vertical

movements between the target and reference images. Sub-pixel accuracy is achieved using bicubic interpolation [179] for sub-pixel shifts.

6.2.2.1. Measure of Confidence in the Registration Results

The method of registration described above is prone to get stuck in local maxima while optimising for mutual information. There are several other problems. In the formula for mutual information, we need to get round the problem of the changing size of the intersection as τ changes. It is also possible that no meaningful registration is possible. This would be the case if, for example, repeated washes during a TIS run were to tear the specimen, or if new extraneous material were to float into the visual field.

To obtain more reliable registration results capable of detecting such failures, we select *K* disjoint square sub-images from the reference, and *K* somewhat larger disjoint square sub-images from the floating image, as in Figure 6-3. The size of the floating sub-image is determined by the size of reference sub-image and the estimated search range required to register the two images. Each such square in the floating image corresponds to exactly one square in the reference image, and the corresponding squares have centres at the same positions in floating and reference images.

We register each of these square sub-images of the reference image within the corresponding larger square in the floating image. More precisely, we find *K* transformations $\tau_{rj}^1, \tau_{rj}^2, \ldots, \tau_{rj}^K$, where τ_{rj}^k is the optimal translation registering the *k*-th smaller square in the reference image within the *k*-th larger square in the floating image. Now τ_{rj}^k can be represented in the form of translation Δ_{rj}^k by a

certain 2-dimensional vector, which we denote by

$$\Delta_{rj}^{k} = (\Delta x_{rj}^{k}, \Delta y_{rj}^{k}) \tag{6.7}$$

where $\Delta x_{rj}^{k}, \Delta y_{rj}^{k}$ are shifts in *x* and *y* direction respectively for transformation τ_{rj}^{k}.

We then calculate the pairwise Euclidean distances d_{kl} between Δ_{rj}^{k} and Δ_{rj}^{l}, for *k*, $l = 1,2,\ldots,K$. If, for fixed *k* and for all *l*, the value of d_{kl} is greater than ω pixels (we chose $\omega = 1$ for sub-pixel accuracy), we mark *k* as an outlier and the user can be warned that this has occurred, making visual inspection possible. If for fixed *r* and *j*, fewer than *k*/2 translations Δ_{rj}^{k} are marked as outliers, their *k* indices are added to the outlier set η, to be excluded from any further calculation. This ensures that if a small number of registrations (fewer than *k*/2) disagree with the majority, they are safely removed from the computation.

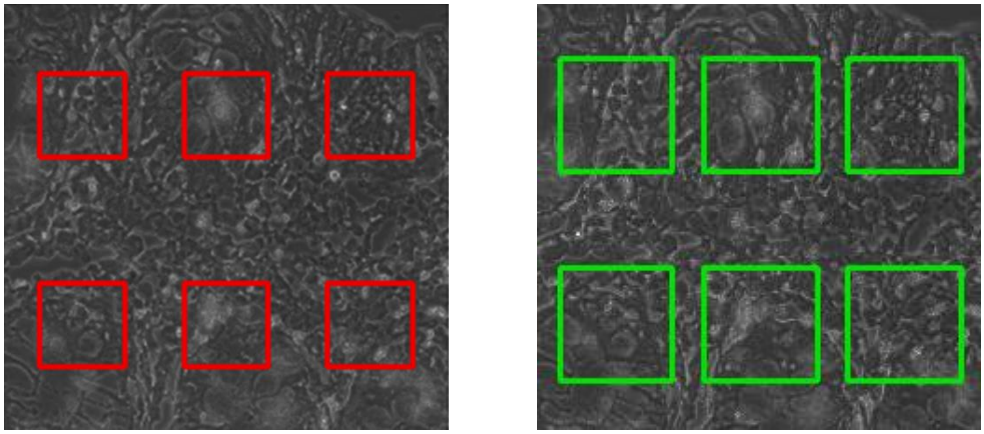


Figure 6-3: A total of *K* sub-images are extracted from different locations in the reference image I_r (left) and optimal transformations τ_{rj}^{k} for $k \in \{1,2,\ldots,K\}$ are calculated between the *k*-th sub-image in I_r and its corresponding sub image in the floating image I_j (right) using a search neighbourhood. Based on these individual transformations, the overall rigid transformation τ_{rj} required to register the images is calculated.

A major benefit of registering with sub-images is that one can easily compute a measure of confidence in the registration results in terms of the standard deviation of the shifts

$$\sigma_{rj} = \sqrt{\frac{1}{K}\sum_{\substack{k=1 \\ k\notin\eta}}^{K}(\Delta_{rj}^{k} - \bar{\Delta}_{rj})(\Delta_{rj}^{k} - \bar{\Delta}_{rj})^{T}} \tag{6.8}$$

where

$$\bar{\Delta}_{rj} = \frac{1}{K}\sum_{\substack{k=1 \\ k\notin\eta}}^{K}(\Delta x_{rj}^{k}, \Delta y_{rj}^{k}) \tag{6.9}$$

The standard deviation σ_{rj} can be used as a measure of confidence in the registration results. If this value is larger than a specified threshold, then the registration process is performed again using a slightly different set of square sub-images. If the confidence value is again larger than the specified threshold, we flag the floating image as a potentially bad quality image or an image that cannot be registered well. If the standard deviation is below the specified threshold for satisfactory registration, the transformation τ_{rj} is computed from the average of all non-outlier local translations $\bar{\Delta}_{rj}$ between sub-images as given in equation (6.9).

6.2.3. Selection of Reference Image with Maximal Overlap (RIMO)

In this section, we utilize the above calculated transformations between all images I_j, for j=1,2,…,N, and I_r in order to select the RIMO maximising the total overlap between the aligned images as shown in Figure 6-2.

6.2.3.1. Registration Graphs

First, we choose any arbitrary image I_r having good enough quality as a reference image and calculate the transformations τ_{rj}, which in this case are represented by translations $\Delta_{rj} = (\Delta x_{rj}, \Delta y_{rj})$ required to align all the images I_j, for j=1,2,…,N with I_r. Once these shifts have been calculated, we can compute the pairwise transformations $\Delta_{ij} = (\Delta x_{ij}, \Delta y_{ij})$ between any two images I_i and I_j in the dataset **I**, as shown in Figure 6-4. The pairwise transformations can then be arranged in the form of two *inter-tag shift matrices* as given below.

$$\Delta X = \begin{bmatrix} \Delta x_{11} & \Delta x_{12} & & \Delta x_{1N} \\ \Delta x_{21} & \Delta x_{22} & & \Delta x_{2N} \\ \vdots & & \ddots & \vdots \\ \Delta x_{N1} & \Delta x_{N2} & & \Delta x_{NN} \end{bmatrix} \tag{6.10}$$

$$\Delta Y = \begin{bmatrix} \Delta y_{11} & \Delta y_{12} & & \Delta y_{1N} \\ \Delta y_{21} & \Delta y_{22} & & \Delta y_{2N} \\ \vdots & & \ddots & \vdots \\ \Delta y_{N1} & \Delta y_{N2} & & \Delta y_{NN} \end{bmatrix} \tag{6.11}$$

Δx_{ij} and Δy_{ij} represent shifts along x-direction and y-direction of image I_j with I_i as the reference image. The above matrices can also be represented in the form of a registration graph, as shown in Figure 6-5. The registration graph can then be used to find shifts between any pair of images in the set **I**, as shown in Figure 6-4. We can now complete the matrices ΔX and ΔY with the help of the equations obtained from the registration graph,

$$\Delta x_{ij} = -\Delta x_{ri} + \Delta x_{rj} \tag{6.12}$$

$$\Delta y_{ij} = -\Delta y_{ri} + \Delta y_{rj} \tag{6.13}$$

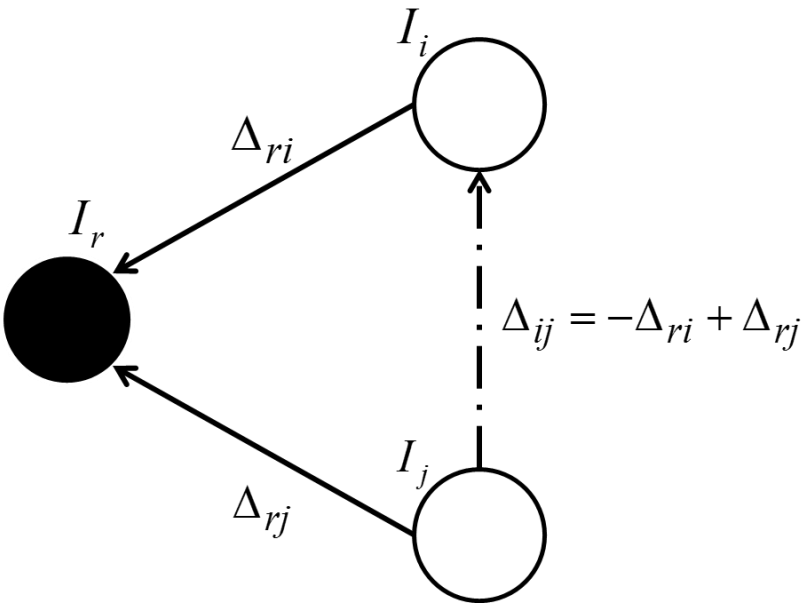


Figure 6-4: Finding the shift Δ_{ij} between the images I_i and I_j, using the previously calculated shifts Δ_{ri} and Δ_{rj} with image I_r. This is similar to vector diagrams where Δ_{ij} is the resultant vector.

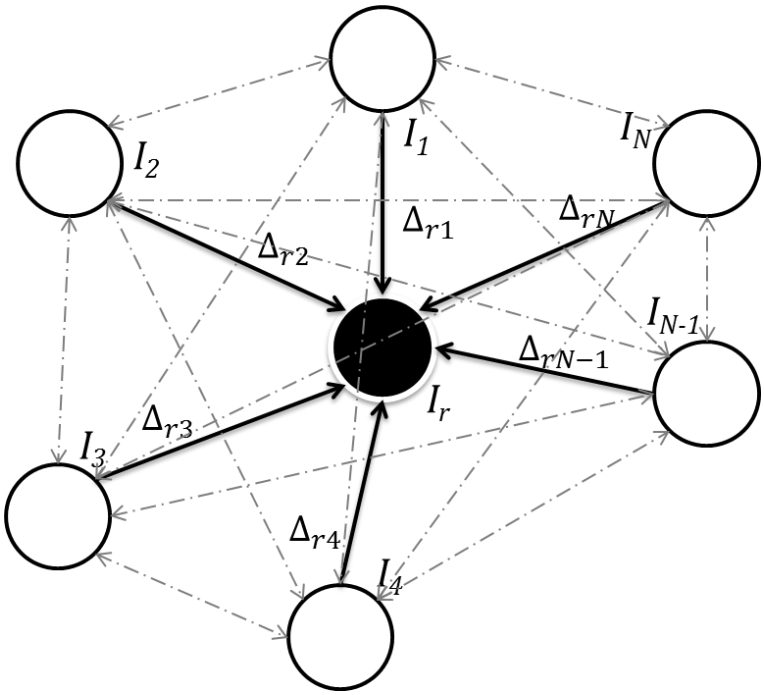


Figure 6-5: Registration graph showing shifts calculated between I_r and all the other tag images in the dataset $\mathbf{I} = \{I_j\}$, $j = 1,2,\ldots,N$. Nodes in the graph represent multi-tag images in an image stack **I**, solid edges represent transformations with respect to I_r, and dashed edges represent transformations that can be determined using this graph as shown and described in Figure 6-4.

The above equations give shifts required by any image I_j considering I_i as the reference image. Since the resultant matrix is skew-symmetric, we can first

compute the upper diagonal matrix and then compute the lower diagonal by just flipping the matrix about the diagonal with a negative sign, to reduce the amount of computation. The total number of registrations performed for N tag images is N-1, producing $\binom{N}{2}$ shift values using equations (6.12) and (6.13).

6.2.3.2. The Objective Function

We wish to compute the value of r that maximises ψ_r defined in equation (6.4). This could be done by direct computation. However, we will show that this expression is also given in terms of a certain metric that we will define. The metric will be a special case of a very general metric coming from a measure in the sense of mathematical Measure Theory.

6.2.3.3. The Shift Metric

We now discuss the metric associated with our objective function. For this discussion, we need a collection S of subsets of a fixed set X and a function $\mu: S \rightarrow [0, \infty)$ satisfying the conditions for S to be a semiring, and for μ to be a measure. The only examples that we will use in this section are:

1. S is the set of all finite subsets of the plane, and μ(S) is equal to the number of elements in S (counting measure). In fact, we will restrict our attention to the situation where the plane is divided into a fixed set of pixels, and each point of S is at the centre of some pixel. Then μ(S) is just a count of pixels.
2. S is the set of all rectangles in the plane, not necessarily with vertices at integer points, and μ(S) is the usual area of the rectangle. We will assume

that the plane is divided into square pixels of height and width one, so that $\mu(p) = 1$ for any pixel p.

The symmetric difference of two subsets $A, B \subset X$ is defined as,

$$A\Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A) \tag{6.14}$$

Lemma 1: For any $A, B, C \subset X$,

$A\Delta C \subset (A\Delta B) \cup (B\Delta C)$.

Proof:

Suppose $x \in A \setminus C$. Then $x \in A$ and $x \notin C$. If $x \in B$, then $x \in B \setminus C \subset B\Delta C$. If $x \notin B$, then $x \in A \setminus B \subset A\Delta B$. This shows that $A \setminus C \subset (A\Delta B) \cup (B\Delta C)$. Now suppose $x \in C \setminus A$. Then $x \in C$ and $x \notin A$. If $x \in B$, then $x \in B \setminus A \subset A\Delta B$. If $x \notin B$, then $x \in C \setminus B \subset B\Delta C$. This shows that $C \setminus A \subset (A\Delta B) \cup (B\Delta C)$.

Thus $A\Delta C = (A \setminus C) \cup (C \setminus A) \subset (A\Delta B) \cup (B\Delta C)$.

Recall that a *pseudometric d* satisfies the same axioms as a metric, except that $d(x,y)=0$ does not necessarily imply $x = y$.

Theorem 1. Let μ be a measure on X, and let F be the set of subsets of finite measure. Then we obtain a *pseudometric* d on F by defining $d(A, B) = \mu(A\Delta B)$. This is a metric if μ has the property that $\mu(A) = 0$ implies $A = \emptyset$, the empty set.

Proof:

For all $A, B \in F$, we have $d(A, B) = \mu(A\Delta B) \geq 0$. Since $A\Delta B = B\Delta A$, we see that $d(A, B) = d(B, A)$. For any $X_1, X_2 \in F$, we know that $\mu(X_1 \cup X_2) \leq \mu(X_1) + \mu(X_2)$, with equality when X_1 and X_2 are disjoint. (This is true for any measure, and

can be directly checked for our two examples of counting measure and area.) It follows from Lemma 1 and this inequality that

$$\begin{aligned} d(A,C) &= \mu(A\Delta C) \\ &\le \mu((A\Delta B)\cup(B\Delta C)) \\ &\le \mu(A\Delta B)+\mu(B\Delta C) \\ &= d(A,B)+d(B,C) \end{aligned} \tag{6.15}$$

This shows that d is a *pseudometric*. If, in addition, $\mu(F) = 0 \Longrightarrow F = \emptyset$, then

$$d(A,B) = 0 \Leftrightarrow A\Delta B = \varnothing \Rightarrow A = B \tag{6.16}$$

which is the final axiom needed in order to show that d is a metric.

Let us apply this result to the example of Figure 6-6. We fix a reference image I_r, and floating image I_j. Let $\Delta_{rj} = +(\Delta x_{rj}, \Delta y_{rj})$, for $1 \le j \le N$. Then $\Delta_{ij} = +(\Delta x_{ij}, \Delta y_{ij}) = (-\Delta x_{ri} + \Delta x_{rj}, -\Delta y_{ri} + \Delta y_{rj})$ as shown by the registration graphs in Figure 6-4 and Figure 6-5. Using the above metric d on subsets of the plane, we define $\left|\left|\Delta_{ij}\right|\right|_d = d(C_i, C_j)$, where C_i, C_j are the set of co-ordinates for I_i and I_j, though we caution that it is not a norm on the set of translations.

Lemma 2. Let d be the metric that arises from Theorem 1, applied to one of our two examples. Recall that each I_j has height h and width w for all $j = 1,2,\ldots,N$. Then, for $1 \le i, j \le N$, we have

$$\| \Delta_{ij} \|_d = 2(|\Delta x_{ij}|.h + |\Delta y_{ij}|.w - |\Delta x_{ij}|\,|\Delta y_{ij}|) \tag{6.17}$$

Provided that $|\Delta x_{ij}| \le w$ and $|\Delta y_{ij}| \le h$. If $|\Delta x_{ij}| \ge w$ or $|\Delta y_{ij}| \ge h$ then $\left|\left|\Delta_{ij}\right|\right|_d = 2hw$. This is the area lost by I_j with I_i.

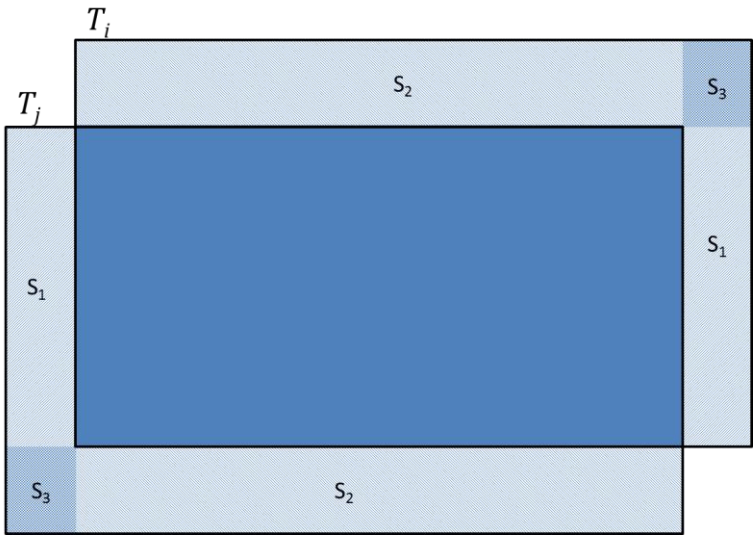


Figure 6-6: Calculating the area occupied by the non-overlapping region.

Proof:

If $|\Delta x_{ij}| \geq w$ or $|\Delta y_{ij}| \geq h$ then C_i and C_j do not intersect, and so $C_i \Delta C_j$ is the disjoint union of C_i and C_j, and this has area $2hw$. Otherwise, the situation will be similar to that shown in Figure 6-6. From this figure, we see that the area of $C_i \Delta C_j$, which is the area lost, is given by

$$\| \Delta_{ij} \|_d = d(C_i, C_j) = \mu(C_i \Delta C_j) \tag{6.18}$$

The region $C_i \Delta C_j$ consists of two congruent components, each comprising three sub-rectangles S_1, S_2 and S_3, meeting only along their edges, with S_3 occupying the corner position. We calculate $\mu(C_i \Delta C_j)$ as follows:

$$\mu(C_i \Delta C_j) = 2(\mu(S_1 \cup S_2 \cup S_3)) = 2(\mu(S_1) + \mu(S_2) + \mu(S_3)) \tag{6.19}$$

From Figure 6-6,

$$\begin{aligned} \mu(S_1) + \mu(S_3) = \mu(S_1 \cup S_3) &= |\Delta x_{ij}|.h \\ \mu(S_2) + \mu(S_3) = \mu(S_2 \cup S_3) &= |\Delta y_{ij}|.w \\ \mu(S_3) &= |\Delta x_{ij}|.|\Delta y_{ij}| \end{aligned} \tag{6.20}$$

It follows that

$$\mu(C_i \Delta C_j) = 2(|\Delta x_{ij}|.h + |\Delta y_{ij}|.w - |\Delta x_{ij}||\Delta y_{ij}|) \tag{6.21}$$

In practice though, $|\Delta x_{ij}|$ is typically much smaller than w and $|\Delta y_{ij}|$ is much smaller than h. Moreover, the value at which the objective function is optimised is unchanged if the objective function is multiplied by a constant. As a result, the third term in the sum can be ignored, and a good approximation to the exact answer is a scaled version of the l_1-metric, given by

$$d(C_i, C_j) = |\Delta x_{ij}|.h + |\Delta y_{ij}|.w \tag{6.22}$$

and this value can be used to specify the objective function. We can now revisit the objective function

$$\psi_r = \sum_i Area(C_r \cap C_i) = \sum_i \mu(C_r \cap C_i) \tag{6.23}$$

Note that $\mu(C_i) + \mu(C_j) = 2\mu(C_i \cap C_j) + \mu(C_i \Delta C_j)$. If C_i and C_j move, while keeping each of $\mu(C_i)$ and $\mu(C_j)$ constant, then $\mu(C_i \cap C_j)$ increases as $\mu(C_i \Delta C_j) = d(C_i, C_j)$ decreases. That is, the larger the area in common between two images, the smaller will be the distance between them. It follows that maximising ψ_r is equivalent to minimising,

$$\varphi_r = \sum_i \mu(C_r \Delta C_i) = \sum_i d(C_r, C_j) \tag{6.24}$$

as r varies. Minimising the above objective function gives r^* the index of the RIMO

image.

6.2.3.4. Using the RIMO

The objective function can easily be calculated and minimised by using the matrices obtained by the help of registration graphs in equation (6.10) and (6.11). Once we have computed all the shifts and found the RIMO and its distance to each of the after tag images, we realign all the tag images with reference to the RIMO. Furthermore, we can also identify which of the fluorescence protein images it might be best to ignore, if for some reason it is advisable to ignore one or more images. Of course, one will often want to ignore fluorescence images of poor quality. But it may also be advisable to eliminate, at least temporarily, images that are distant from the RIMO (using the distance function defined above).

6.3. Results on Registration

6.3.1. Experiments on Synthetic Data

Synthetic data was generated by selecting a phase contrast image I_{sel} from one of the TIS image stacks. Two random vectors x' and y' of length 500 were drawn from a uniform distribution of real-valued numbers in the range $[-x_{min}, +x_{max}]$ and $[-y_{min}, +y_{max}]$ with $x_{min} = x_{max} = y_{min} = y_{max} = 10$. Let (x_{centre}, y_{centre}) denote coordinates of the centre of the selected image I_{sel} and let I_{syn}^{0} denote a cropped section of I_{sel} with (x_{centre}, y_{centre}) as its centre. A new set of centre coordinates for the synthetic tag images is then calculated by adding x' and y' to (x_{centre}, y_{centre}) as follows,

$$x = x' + x_{center} \tag{6.25}$$

$$y = y' + y_{center} \tag{6.26}$$

A synthetic stack of TIS images $\mathbf{I}_{syn} = \{I_{syn}^{j}\}$, where $j = 1,2,\ldots,500$, was generated by taking cropped sections of I_{sel} with (x_j, y_j) as their centres and having the same pixel resolution as. The amount of actual shift for the synthetic tag image I_{syn}^{j}, for all *j* from the original *reference* image I_{syn}^{0} is given by the corresponding values (*i.e.*, the *j*-th elements) in *x* and *y*. Nearly a quarter of the synthetic tag images were randomly picked and a contrast change using gamma correction [42] with γ in the range 0.5 to 2 was applied to them. Another quarter of the synthetic images were randomly picked and Gaussian blurring with kernel bandwidth σ = 1 and a filter size of 6×6 pixels was applied to them. The remaining 50% of the images did not go through any intensity transformation, and were only translated by random shifts. The parameters for blur and translations were chosen based on observations from the real dataset. So a dataset consisting of randomly shifted images was generated, with contrast and smoothing artifacts added to half of them randomly. Figure 6-7 shows an illustration of how the synthetic data set is generated using a single phase contrast image from a TIS image stack as I_{sel}.

An image was randomly selected from our artificial data set and $\boldsymbol{x}_{cal}$ and $\boldsymbol{y}_{cal}$ shifts were calculated using our registration algorithm. The mean difference between the actual and estimated shifts was calculated to be (0.1128, 0.1165) in the *x* and *y*-directions respectively. We can achieve more accurate results by using different values of *K*, *S* and $\theta_{\Delta\tau}$ as shown in Table 6-1, but there is always a trade-off between time and accuracy. Using $\boldsymbol{x}_{cal}$ and $\boldsymbol{y}_{cal}$ the RIMO was calculated by using our algorithm. The RAMTaB successfully found the image which had minimum shift

with respect to all of the other images of the stack, therefore providing experimental verification that the algorithm is capable of finding RIMO.

Table 6-1 shows time consumed using a different number *K* of the sub-images and for different threshold specified for $\theta_{\Delta\tau}$ for a stack of 26 images with two visual fields on a 2.66 GHz Quad Core CPU. It was found empirically that $K = 9$, $S = 200$, and $\theta_{\Delta\tau} = 0.01$ gave us a good compromise between the algorithm's runtime and the accuracy of registration. Using these parameters, our approach takes about 22 minutes and 15 seconds to register a stack of 26 tag images with two visual fields on a 2.66 GHz Quad-core CPU using non-optimised MATLAB® code running on a Linux platform. The MATLAB source code, 32-bit Windows executable, and a sample TIS stack can be downloaded from the project website[15].

[15] http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/projects/bic/ramtab/

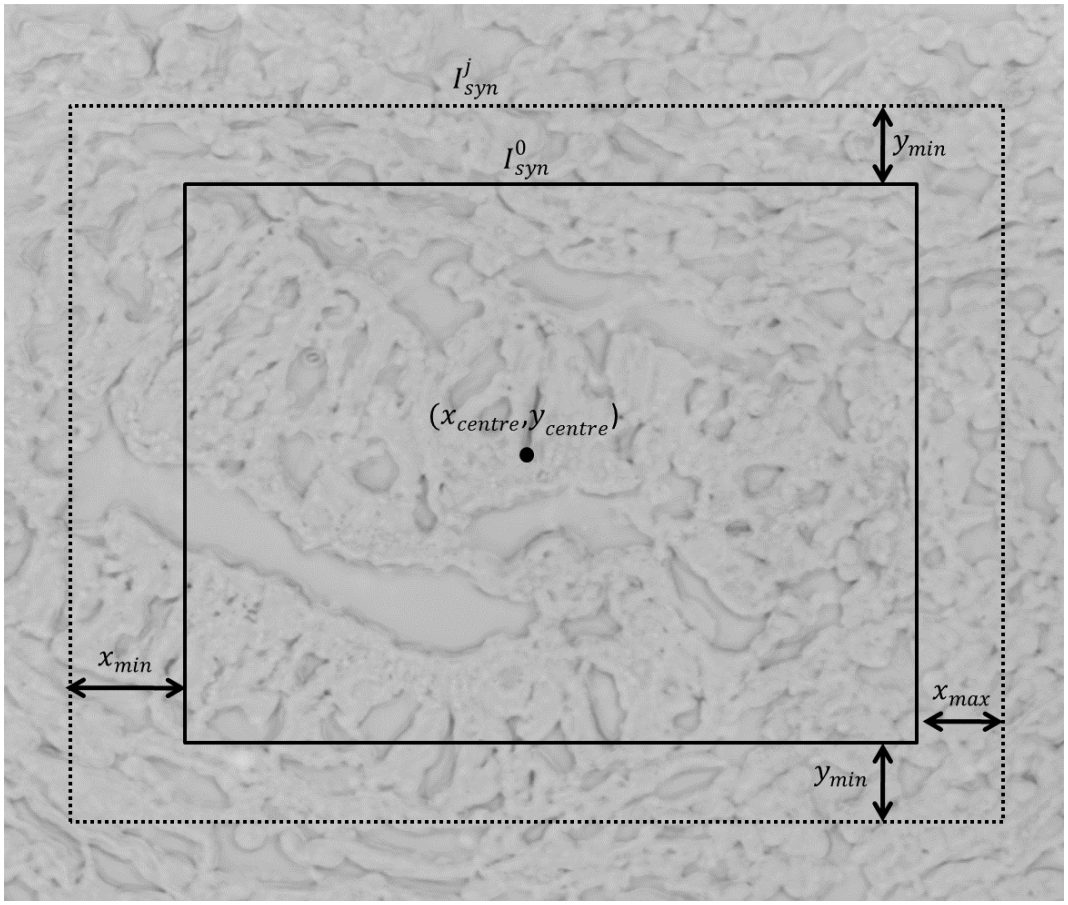


Figure 6-7: Construction of a synthetic data set $\mathbf{I}_{syn}$ using a single tag image I_{sel} with center coordinates (x_{centre}, y_{centre}) and uniformly distributed random shifts $x_n \in [-x_{min}, +x_{max}]$ and $y_n \in [-y_{min}, +y_{max}]$ in both directions. Intensity variations such as contrast stretching and Gaussian blurring are also introduced randomly in 50% of the images to mimic the random perturbations in pixel intensities during the image acquisition process (The image here has been inverted for visibility purpose).

Table 6-1: Time required to register a stack of 26 images with two visual fields for different values of *K*, *S*, and $\theta_{\Delta\tau}$ where *K* is the number of sub-images used while calculating the translations, *S* is size of the sub-images, and $\theta_{\Delta\tau}$ is the threshold for $|\Delta\tau|$ between two consecutive iterations of the pattern search algorithm.

K	S	$\theta_{\Delta\tau}$	Time taken to register real data containing 26 tag images with two visual fields	Approximate error to register 500 images of synthetic data			
				Normal	Corrupted	Mean	Standard Deviation
9	100x100	0.01	9 min 48 sec	0.4229, 0.3230	0.4228, 0.3556	0.4229, 0.3393	1.7801, 1.3568
		0.001	20 min 55 sec	0.3101, 0.2225	0.3294, 0.2638	0.3198, 0.2432	1.7731, 1.3647
		0.0001	30 min 25 sec	0.3010, 0.2133	0.3233, 0.2585	0.3122, 0.2359	1.7744, 1.3667
9	200x200	0.01	22 min 15 sec	0.1107, 0.1161	0.1149, 0.1169	**0.1128, 0.1165**	**0.0662, 0.0652**
		0.001	51 min 38 sec	0.0135, 0.0142	0.0166, 0.0153	0.0151, 0.0148	0.0121, 0.0120
		0.0001	81 min 27 sec	0.0020, 0.0023	0.0072, 0.0067	0.0046, 0.0045	0.0095, 0.0099
6	300x300	0.01	42 min 54 sec	0.0779, 0.0789	0.0760, 0.0698	0.0770, 0.0744	0.0580, 0.0560
		0.001	81 min 41 sec	0.0088, 0.0095	0.0127, 0.0113	0.0107, 0.0104	0.0118, 0.0115
		0.0001	113 min 41 sec	0.0021, 0.0025	0.0078, 0.0070	0.0050, 0.0047	0.0105, 0.0097

6.3.2. Experimental Results on Real Data

We ran the proposed registration framework and the algorithm for selection of RIMO on a large number of TIS stacks. Here we report results of a TIS run on a cancerous colon tissue captured by the biologists S. Abouna and M. Khan in October 2010. The antibody tag library for the experiment consisted of tumour

markers, stem cell markers, and proliferation markers. First, we choose any arbitrary tag image (e.g., DAPI) as a reference image I_r and calculate the transformations τ_{rj} required to align all the images I_j, for j=1,2,..., N with I_r. Using the results of registration, the RAMTaB gave Ki67 tag image as the RIMO. Registration results were also generated by arbitrarily choosing the Bax tag as reference.

The results of registration using 3 reference images (DAPI and Bax selected arbitrarily, and Ki67 as RIMO) are shown in Figure 6-8 in the form of a plot of magnitude of shift required to register a tag image to the corresponding reference image. The plots clearly show that by using Ki67 as reference tag, the total amount of shift required to register the images is much smaller than by using the other two reference images. When Ki67 was used as reference image, there was only one tag for which magnitude of shift was found to be greater than 10, whereas, when DAPI1 or Bax were used as reference images, there were more than 8 images for which the magnitude of shift calculated was greater than 10. Since our goal is to minimise φ_r it is clear from Figure 6-8 that the RAMTaB framework has been successful in minimising the magnitude of shifts. The amount of pixels lost when using Ki67, DAPI and Bax as I_r was calculated to be **129.10**, **208.42** and **168.89** respectively.

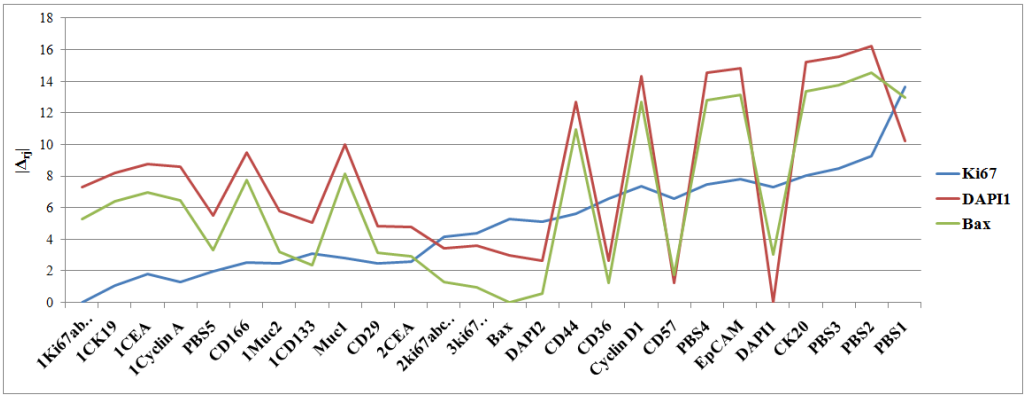
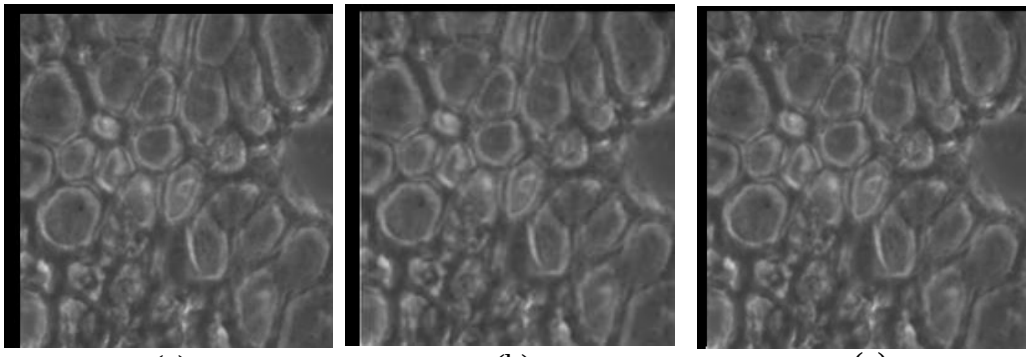


Figure 6-8: Magnitude of shift required to register different tag images to the corresponding reference image (tags along *x*-axis are arranged in a way that the first image is the RIMO, second image is the second in line as RIMO and the same trend continues towards right).

Figure 6-9 shows the upper-left part of the phase contrast image for CK20 tag from the same image stack, after it has been aligned to the phase contrast images corresponding to DAPI1, Bax, and Ki67. The blank rows and columns (having zero intensity values) near the top-left corner of the image are due to the amount of shift which was required to align the image to the respective reference image. The number of blank pixels near the top left corner in this image is equal to the number of pixels lost at the bottom left corner of the image. It can be observed from this figure that when the Ki67 image is used as a reference, φ_r is minimised, once again showing in empirical terms that the proposed RAMTaB framework selects RIMO as reference for registration. Figure 6-10 shows the amount of translational shift calculated using the proposed RAMTaB framework for images acquired during a single TIS run plotted against time. In this particular instance, the amount of shift decreased as the TIS run progressed but in other cases, the trend may be different. This indicates that the TIS machine settles down to a stable state as the run continues.



(a) (b) (c)

Figure 6-9: Phase contrast image for the CK20 tag registered to (a) DAPI1, (b) Bax and (c) Ki67 tag image; the black region on the top and the left of the image shows the amount of shift required to register the image to respective reference.

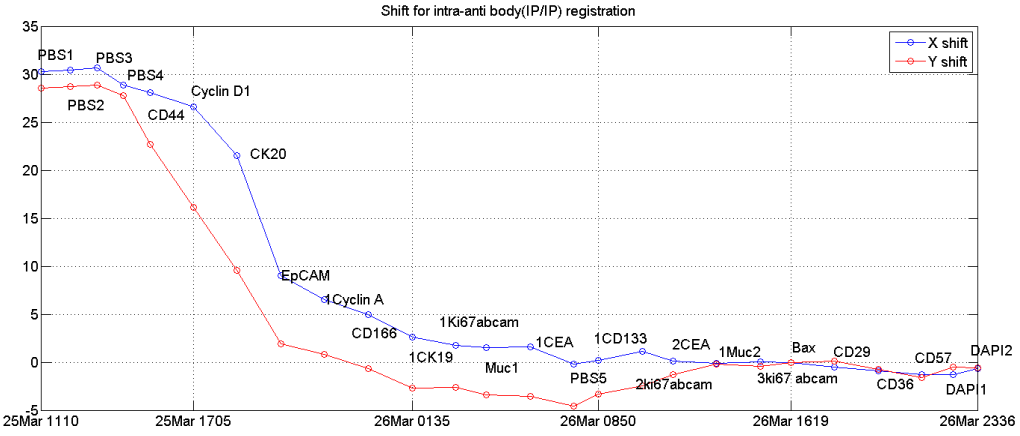


Figure 6-10: Shift, as estimated by RAMTaB, in both *x* and *y*-directions during one TIS run versus time. This indicates the TIS machine "settling down" to an equilibrium state as time passes, probably due to the temperature reaching a stable value.

Figure 6-1 shows composite RGB images obtained by using three different phase contrast and fluorescence images for CD57, CD166 and DAPI as red, green and blue channels, respectively. In Figure 6-1(a) & (b), phase contrast and fluorescence images from the original data set obtained after a TIS run are used as red, green and blue channels. The colour fringes in Figure 6-1(a) show the degree of misalignment present between these three phase contrast images which should ideally be aligned to each other. We have aligned the images using the proposed algorithm and formed

the composite as shown in Figure 6-1(c) & (d). If pixel intensities from all three phase contrast images are in agreement with each other, we should only see shades of grey in the composite RGB image. It can be seen in the alignment results of both the algorithms that the colour fringes have been removed in Figure 6-1(c). We have calculated the root mean squared (RMS) difference between the red, green and blue channels, for the phase contrast images shown in Figure 6-1 to numerically illustrate the misalignment, using the equation below,

$$RMS = \sqrt{\sum((I_R - I_G)^2 + (I_R - I_B)^2 + (I_G - I_B)^2)/3B} \tag{6.27}$$

where I_R, I_G, and I_B denote the red, green, and blue channel phase contrast images, respectively, and B denotes the number of pixels in each of the channel images. For the images shown in Figure 6-1(a) & (c), the RMS difference was found to be 7.14 for the misaligned images (Figure 6-1(a)) and 2.98 for the registered images (Figure 6-1(c)).

6.4. Preliminary Analysis

Preliminary analysis of the data was performed after registration in later studies in collaboration with other members of our group at the Bioimage Analysis Lab and with our collaborators at the Bielefeld University Germany [180–185]. We proposed a molecular co-expression analysis framework for analysis on a whole range of protein expression levels in close vicinity of cells, instead of previously used binarisation techniques [15,18,182]. The framework consists of three major components: 1) Registration using the proposed RAMTaB method (Section 6.2), 2) Segmentation of Nuclei and its neighbourhood region and 3) Analysis and visualization of molecular patterns using the clustering method. We first performed

segmentation of the nuclei to filter out noise from non-cellular pixel locations such as lumen and stroma. For this purpose, were first normalise intensity values of the aligned DAPI channel to the range [0,1] and then segmentation is performed using Gaussian Mixture Modelling (GMM) over the normalised intensity values. For data analysis we used the bottom-up hierarchical clustering method [186]. For *n* number of pixels, hierarchical clustering starts with *n* clusters and iteratively merges these clusters based on a specified criteria, the criteria chosen in this case was within class variance and the minimum number of clusters was chosen to be 20. We have shown that there is a clear difference in tissue morphology and molecular expression at sub-cellular level in normal and cancer specimens [182].

In a later study, instead of pixel level analysis, we performed cell level analysis [180]. We proposed a refined cell segmentation method and showed that nonlinear embedding of high dimensional protein co-expression vector performs better than linear dimensionality reduction methods. For cell segmentation we used the DAPI channel [187], the image was first binarised using the graph-cut based method to extract the foreground, the binarisation was followed by multi-scale Laplacian of Gaussian (LoG) filter for initial segmentation. The initial segmentation was further refined using a graph-cut based algorithm and followed by post-processing to secure very small nuclei. For each cell, an *N*-dimensional raw expression vector (REV) was formulated across *N* antibody tags using mean of intensity values under each cell. Non-linear t-Distributed Stochastic Neighbour Embedding (t-SNE) was compared with linear Principal Component Analysis (PCA) to embed *N*-dimensional data to lower dimensions and it was shown t-SNE preserved the pairwise relationships between REVs whereas PCA failed to preserve the

relationship which was present in high dimensional space. Affinity Propagation Clustering (APC) and Agglomerative Hierarchical Clustering (AHC) were performed and it was shown that nonlinear t-SNE embedding outperforms original high-dimensional REVs in their ability to discriminate between cancer and normal tissues on the basis of phenotypic distributions. This work was further extended in a later study and clustering was performed using Self Organizing Maps (SOM) in addition to AHC and APC, it was shown that the results were consistent and t-SNE embedding performed best and showed higher inter-class variability [181].

Chapter Summary

The TIS method uses a library of fluorescent tags to obtain phase-fluorescence pair images corresponding to each tag. Accurate alignment of tag images in a multi-tag fluorescence microscopy image stack is an essential pre-processing step prior to any analysis of protein co-expression. The proposed framework determines sub-pixel shifts between phase contrast images in a multi-tag fluorescence image stack. Subsequently, these shifts can be used to register the fluorescence images to co-localise signals from different protein molecules or find molecular co-expression patterns for different biomolecules. Importantly, our system is highly effective on real as well as on synthetic data. It has been shown to be robust to luminance and contrast variations, yields a confidence value in the quality of alignment results, and removes the need for a biologist to eyeball all phase contrast images in the stacks to select an appropriate reference image. Our block-based registration algorithm ensures that the alignment is robust to any damage caused during sequential bleaching or washing to a small part of the tissue. On the synthetic data, the proposed framework gives almost perfect alignment, up to two decimal places sub-

pixel accuracy for a selected set of parameters (see Table 6-1).

After registration, selection of the RIMO image is the next step. We have shown that we can collect maximum amount of data from the image stack after registration using the RIMO image. All the images were registered to this reference image using a novel shift metric. The RAMTaB framework was successfully used to register several stacks and has been shown to be robust to brightness and contrast variations. We have shown in later studies that normal and cancer tissues can be discriminated on the basis of high dimensional protein co-expressions from different anti-body tags.

Chapter 7
Conclusions and Future Directions

This thesis proposed approaches for registration, disparity estimation and classification for the purposes of anomaly detection from multi-variate bioimage data. This chapter summarises and concludes the work presented in this thesis and discusses some future directions.

In Chapter 1, we introduced the reader to the existing multi-modal techniques being used for analysing the quality of plants at research and commercial level. Most of the existing techniques at the research level do not use the full potential of multi-modal images as their analysis is limited to one modality at a time. The techniques being used at the commercial scale use multi-spectral images of plants but these techniques are again not using the full potential of an important part of the electromagnetic spectrum i.e., thermal imaging for plant quality assessment. This is partly because of the difficulties involved in aligning images from thermal cameras to other modalities. The motivation behind our work was to explore the opportunities offered by combining thermal and visible light images for anomaly detection in plants. We also presented an introduction to the existing multi-channel fluorescence microscopy techniques and discussed their potential for better understanding of diseases which are developed at sub-cellular level.

In Chapter 2, we reviewed the existing literature on registration, multi-modal registration, disparity estimation, and the use of multi-modal images for water stress

and disease detection. The review covered general approaches for registration of images and then we focussed our discussion on registration of multi-modal images, especially thermal and visible light images of diseased plants. We discussed limitations of existing approaches for registration of thermal and visible light images of diseased plants and emphasised the need for a new robust approach for registration of these kind of multi-modal images. In the same chapter, we also discussed the need for a robust registration method for registration of multi-channel fluorescence microscopy images. We reviewed some of the existing approaches for disparity estimation and discussed how disparity among the stereo image pair can be used for depth estimation. Finally, we reviewed the existing literature on water stress and disease detection in plants using thermal images. It was recognised that most of the literature on crop water stress was based on crop water stress indices, which are dependent on environmental conditions, which led us to a departure from these techniques and establish techniques which are less dependent on environmental conditions and can potentially be applied at commercial scale.

Registration of multi-modal (thermal & visible light) images of diseased plants is a challenging task due to the fact that different parts of a diseased plant may express different temperature or colour in thermal and visible light images. For example, there might be local variations within thermal image of a diseased plant which might not reflect in the corresponding visible light images if the disease symptoms have not yet become visible. In such cases, standard feature extraction methods are not able to produce stable features which can be used for registration purposes. Similarly, many robust similarity measures such as mutual information perform poorly because of these kinds of variations. Therefore, we proposed an alternative

approach employing silhouettes as global appearance instead of directly using intensity values for measuring similarity between two plant regions and using it for registration of plant regions in multi-modal images.

A novel multi-scale method for silhouette extraction of diseased plants in thermal and visible light images was proposed in Chapter 3. The method demonstrated high registration accuracy in terms of coverage metric and DICE coefficient of the extracted silhouettes when compared with the ground truth. We showed that the proposed multi-scale method is highly accurate as compared to gradient based methods especially on thermal images. The silhouettes extracted using the proposed method were used for registration of multi-modal images of diseased plants and we found the registration results to be very promising. The proposed silhouette extraction method is not limited to plants and can be extended to silhouette extraction of single/multiple objects in images provided that the objects of interest do not overlap.

In Chapter 4, we captured stereo images and calculated the disparity maps from stereo pairs of images for depth estimation purposes. We performed image rectification of the stereo images by identifying extrinsic marker points placed on the ground level. After image rectification, we compared six different algorithms for disparity estimation. In addition to the disparity estimation algorithms mentioned in Chapter 4, we also tested some other algorithms including [117,122,129] but these algorithms were found to be highly sensitive to the noise content in the diseased plants and, therefore, we did not include their results in our analysis. Our goal was to estimate smooth and accurate disparity maps which were insensitive to background noise and noisy patterns in diseased images. The

proposed multi-resolution method (section 4.2.6) was computationally fast and less sensitive to the noise content while simultaneously producing smooth disparity maps at the expense of slightly higher RMS and B (bad matching pixels) compared to Graph-cut based [120] and non-local cost aggregation method [130]. When compared to the algorithm [137], proposed in a previous HDC project (CP37), the proposed method performed best in all aspects, i.e., RMS error, B and computational efficiency.

Chapter 5 combines multi-modal features including depth for anomaly detection and consists of two parts: the first part proposed a classification approach to detect water deficient regions in canopies and the second part proposed classification approaches to detect diseased plants. For water deficiency detection, we were able to detect regions of the canopy, which were experiencing soil moisture deficit, by using a machine learning approach instead of stress indices, thus enabling us to identify water deficient regions without relying on environmental factors. Initially, the effect of reflected light and background information was reduced in order to extract features. In the second step, these features were classified using SVM, GPC and a combination of both classifiers. The colour information in visible light images provides information about the amount of reflected light intensity from the plant. Using this information, temperature values were scaled on the basis of reflected light. Plant regions can also be identified in the registered thermal image using colour information. This helped to discard temperature values belonging to the background and extract useful information from plant regions as identified in [96]. We also tested the proposed classifier on an artificially generated, mixed condition image. The classification results in this image showed a significant improvement

using the proposed feature set when compared to the thermal only feature set. We found the proposed set of features robust to the amount of input information and to mixed-condition images. In future, this work can be extended to identify canopies under multiple levels of stress.

For the disease detection part, we proposed two different approaches: a two-step pixel-level classification approach and a region-based classification. Before application of these two approaches, we extracted plant regions from the background region. In the first classification approach, we combined information to first detect potential disease regions in a plant and then we classified the plant into diseased or normal plant based on the information collected from the potential disease regions. The second approach is very similar to the water deficient region detection, whereby we directly collect information from plant regions. The results showed that by combining colour information with thermal and depth information, we can increase the accuracy of disease detection. As a refinement step, we combined feature sets from the two approaches and showed that the new feature set produced more stable results.

We also showed that our feature set was able to identify plants which were not inoculated with any disease but later captured the disease probably due to being in close proximity of other inoculated plants. Although we were able to detect diseased plants with an accuracy of about 95%, there is still need for improvement before this technique can be used at commercial scale. Although we can use the same technique across different crops for water deficiency detection purposes, our technique will need further development for disease detection on different plants and different diseases since different plants may respond differently in terms of

thermal signature to the same disease and therefore further testing is necessary before application. For our experiments, we collected the imaging data on a day-to-day basis. There may be some diseases which show symptoms of disease just hours before it visibly appears, and so the time interval between consecutive images may need to be reduced.

Chapter 6 proposed an approach for registration of multi-channel fluorescence microscopy images using the corresponding phase contrast images. The presented approach used mutual information as a similarity measure and proposed a block based method for fast registration with sub-pixel accuracy. In addition, we proposed a method for selection of reference image with maximal overlap (RIMO) amongst dozens of images in a stack which when used as a reference image causes minimum amount of information loss during registration process. Our results showed that the amount of information loss was minimal when we chose RIMO to register a multi-channel stack of images. The current method to find RIMO is limited to situations where only translations define the image transformation.

The images registered using the proposed method were used in later studies [180–182,184,185] which demonstrated the heterogeneity of cancer cells and showed different molecular co-expression patterns and cell phenotypes. Similar studies on a large data set will help us to better understand cell phenotypes which are involved in the development and progression of cancer. In future, we can introduce more complex transformations to the method to select RIMO. In addition, we can investigate different data normalisation techniques to study how intensity in fluorescent images from different channels relates to the actual signal, resulting in not only helping us to better understand protein-protein interactions but also

removal of artifacts such as auto-fluorescence. Another interesting aspect can be the study of behaviour of different cell segmentation methods on fluorescence images and the temporal analysis of the data from TIS machine.

Concluding Remarks

We have shown in this thesis that combining information from multiple modalities increases the accuracy of classification algorithms for biomedical anomaly detection. We presented algorithms which can be used to classify individual plants and canopy regions with anomalies at plant and canopy level instead of following the classical approach of studying disease development in individual leaves. There is a lot of further work to be done for validation and scaling-up before such detection systems can be installed in a commercial setting, for instance on rigs above the crop for anomaly detection. The approaches developed in this thesis can be rigorously tested on multiple types of disease with multiple control treatments. However, as the imaging technology advances and becomes available at cheaper price, we can expect to see these kinds of systems being installed in glass houses and commercial crops very soon.

The TIS machine uses multi-channel fluorescence microscopy technique to identify protein complexes in tissues at subcellular level and hence can be used to study cell behaviour. This can help us better understand and subtype certain diseases which are caused by malfunctioning of the cell, for example cancer. Similarly, temporal analysis of tissues can help us to identify origin of certain types of cells e.g., beta cells during pregnancy. The heterogeneous nature of certain cancer types e.g., breast cancer makes it difficult to decide a particular therapy to treat the disease. Large scale TIS studies may reveal certain biomarkers common to specific cancer

types, thus making the treatment decision process targeted to those biomarkers.

The techniques developed in this thesis can be applied to multi-modal image analysis problems in several other fields. The silhouette extraction method developed for estimating the silhouette of plants can be extended to segmentation of objects in noisy images where it is very difficult to segment the object of interest because of blurred boundaries for example segmentation of nuclei in microscopic images. Similarly, the multi-modal registration algorithm can be extended to registration of any kind of multi-modal image registration problem where the objects of interest do not occlude each other. The multi-channel image registration algorithm RAMTaB can not only be used to align images from several different channels/modalities, but also can be used to find the best reference image (RIMO) and to automatically test the quality of registration. We have found that many existing disparity estimation algorithms which produce excellent results on test data set fail to produce good quality results on real data in the presence of noise. The disparity estimation algorithm proposed in this thesis can be applied to estimate depth information in the presence of noise, especially to the cases with salt and pepper noise. Finally, we have shown that by combining multi-modal/channel image data the accuracy of detection of anomalies increases, therefore this idea can be extended to automatically scan whole plant canopies.

References

1. Wang X, Li L, Hu C, Qiu J, Xu Z, et al. (2009) A comparative study of three CT and MRI registration algorithms in nasopharyngeal carcinoma. Journal of applied clinical medical physics / American College of Medical Physics 10: 2906.

2. Maintz JB, Viergever MA (1998) A survey of medical image registration. Medical Image Analysis 2: 1–36.

3. Collignon A, Maes F (1995) Automated multi-modality image registration based on information theory. Information processing in Medical Imaging: 264–274.

4. Manickavasagan A, Jayas D (2005) Applications of thermal imaging in agriculture–A Review. Canadian Society for Engineering in Agricultultural, Food and Biological Systems: 1–11.

5. Williams T (2010) Thermal imaging cameras: characteristics and performance.

6. Fuchs M (1990) Infrared Measurement of Canopy Temperature and Detection of Plant Water Stress. Theoretical and Applied Climatology 261: 253–261.

7. Jones HG (1999) Use of infrared thermometry for estimation of stomatal conductance as a possible aid to irrigation scheduling. Agricultural and Forest Meteorology 95: 139–149.

8. Vadivambal R, Jayas DS (2010) Applications of thermal imaging in agriculture and food industry—A Review. Food and Bioprocess Technology 4: 186–199. doi:10.1007/s11947-010-0333-5.

9. Stoll M, Jones H (2007) Thermal imaging as a viable tool for monitoring plant stress. International Journal of Vine and Wine Sciences 41: 77–84.

10. Ju X, Nebel J, Siebert JP (2004) 3D thermography imaging standardization technique for inflammation diagnosis. In: Gong H, Cai Y, Chatard J-P, editors. Proc. SPIE, Photonics Asia. pp. 266–273. doi:10.1117/12.577055.

11. Eberius M (2011) Automated image based plant phenotyping - Challenges and chances. 2nd International Plant Phenotyping Symposium.

12. Zwiggelaar R (1998) A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. Crop Protection 17: 189–206. doi:10.1016/S0261-2194(98)00009-X.

13. Zwiggelaar R, Yang Q, Garcia-Pardo E, Bull CRR (1996) Use of Spectral Information and Machine Vision for Bruise Detection on Peaches and Apricots. Journal of Agricultural Engineering Research 63: 323–331. Available: http://linkinghub.elsevier.com/retrieve/pii/S0021863496900359.

14. Mahlein A-K, Oerke E-C, Steiner U, Dehne H-W (2012) Recent advances in sensing plant diseases for precision crop protection. European Journal of Plant Pathology 133: 197–209. doi:10.1007/s10658-011-9878-z.

15. Bhattacharya S, Mathew G, Ruban E, Epstein DBA, Krusche A, et al. (2010) Toponome Imaging System : in situ protein network mapping in normal and cancerous Colon from the same patient reveals more than five-thousand cancer specific protein clusters and their subcellular annotation by using a three symbol code research articles. Journal of Proteome Research: 6112–6125.

16. Raza S-E-A, Humayun A, Abouna S, Nattkemper TW, Epstein DB, et al. (2012) RAMTaB: Robust alignment of multi-tag bioimages. PLoS ONE 7: e30894.

17. Murphy RF (2006) Putting proteins on the map. Nature Biotechnology 24: 1223–1224.

18. Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, et al. (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nature Biotechnology 24: 1270–1278.

19. Bode M, Krusche A (2007) Toponome Imaging System (TIS): imaging the proteome with functional resolution. Nature Methods Application Notes 4: 1–2. doi:10.1038/nmeth998.

20. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM (2007) MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. Nature Methods 4: 828–833.

21. Van Manen H-J, Kraan YM, Roos D, Otto C (2005) Single-cell Raman and fluorescence microscopy reveal the association of lipid bodies with phagosomes in leukocytes. Proceedings of the National Academy of Sciences of the United States of America 102: 10159–10164.

22. Can A, Bello M, Cline H, Tao X, Ginty F, et al. (2008) Multi-modal imaging of histological tissue sections. International Symposium on Biomaedical Imaging. Paris: IEEE. pp. 288–291.

23. Barash E, Dinn S, Sevinsky C, Ginty F (2010) Multiplexed analysis of proteins in tissue using multispectral fluorescence imaging. IEEE Transactions on Medical Imaging 29: 1457–1462.

24. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, et al. (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1300136110.

25. Herold J, Zhou L, Abouna S, Pelengaris S, Epstein D, et al. (2010) Integrating semantic annotation and information visualization for the analysis of multichannel fluorescence micrographs from pancreatic tissue. Computerized medical imaging and graphics the official journal of the Computerized Medical Imaging Society 34: 446–452.

26. Herold J, Loyek C, Nattkemper TW (2011) Multivariate image mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1: 2–13. doi:10.1002/widm.4.

27. Langenkämper D, Jan K, Khan M (2011) Towards protein network analysis using tis imaging and exploratory data analysis. Workshop on Computational Systems Biology.

28. Raza S-E-A, Sanchez V, Prince G, Clarkson J, Rajpoot NM (2013) Registration of thermal and visible light images using silhouette extraction. Pattern Recognition (accepted).

29. Raza S-E-A, Smith HK, Clarkson GJJ, Taylor G, Clarkson J, et al. (2013) Automatic detection of water stress in spinach canopies using combined visible and thermal imagery. PLoS ONE (Under Review).

30. Goshtasby AA (2012) Images Registration Advances in Computer Vision and Pattern Recognition.

31. Prakosa A, Sermesant M, Delingette H (2013) Generation of synthetic but visually realistic time series of cardiac images combining a biophysical model and clinical images. Medical Imaging, IEEE transactions on 32: 99–109. doi:10.1109/TMI.2012.2220375.

32. Sun Y, Luebbers H-T, Agbaje JO, Schepers S, Vrielinck L, et al. (2013) Validation of anatomical landmarks-based registration for image-guided surgery: An in-vitro study. Journal of Cranio-Maxillofacial Surgery 41: 522–526. doi:http://dx.doi.org/10.1016/j.jcms.2012.11.017.

33. Liu J, Collins RT, Liu Y (2013) Robust autocalibration for a surveillance camera network. Applications of Computer Vision (WACV), 2013 IEEE Workshop on. pp. 433–440. doi:10.1109/WACV.2013.6475051.

34. Fitzpatrick JM, Hill DLG, Maurer CR (2000) Image Registration. Handbook of Medical Imaging: Medical image processing and analysis. pp. 447–514.

35. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, et al. (1999) Nonrigid

registration using free-form deformations: application to breast MR images. IEEE transactions on Medical Imaging 18: 712–721. doi:10.1109/42.796284.

36. Kroon D-J (2011) Segmentation of the mandibular canal in cone-beam CT data Enschede, The Netherlands: University of Twente. doi:10.3990/1.9789036532808.

37. Viola P, Wells WM (1997) Alignment by maximization of mutual information. International Journal of Computer Vision 24: 137–154.

38. Pluim JPW, Maintz JBA, Viergever MA (2003) Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 22: 986–1004. doi:10.1109/TMI.2003.815867.

39. Foroosh H, Zerubia JB, Berthod M (2002) Extension of phase correlation to subpixel registration. Image Processing, IEEE Transactions on 11: 188–200. doi:10.1109/83.988953.

40. Wolberg G, Zokai S (2000) Robust image registration using log-polar transform. Image Processing, 2000. Proceedings. 2000 International Conference on. Vol. 1. pp. 493–496 vol.1. doi:10.1109/ICIP.2000.901003.

41. Nixon MS, Aguado AS (2012) Feature extraction & image processing for Computer Vision. 3rd ed. Press A, editor Elsevier.

42. González RC, Woods RE (2008) Digital image processing. Pearson/Prentice Hall.

43. Hough P (1962) Method and means for recognising complex patterns.

44. Harris C, Stephens M (1988) A combined corner and edge detector. Proceedings of the Fourth Alvey Vision Conference. pp. 147–151.

45. Lowe D (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60: 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

46. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110: 346–359. doi:10.1016/j.cviu.2007.09.014.

47. Barrow HG, Tenenbaum JM, Bolles RC, Wolf HC (1977) Parametric correspondence and chamfer matching: Two new techniques for image matching.

48. Besl PJ, McKay ND (1992) Method for registration of 3-D shapes. Robotics-DL tentative. pp. 586–606.

49. Borgefors G (1988) Hierarchical chamfer matching: A parametric edge matching algorithm. Pattern Analysis and Machine Intelligence, IEEE Transactions on 10: 849–865.

50. Zitova B (2003) Image registration methods: a survey. Image and Vision Computing 21: 977–1000. doi:10.1016/S0262-8856(03)00137-9.

51. Rohde GK, Pajevic S, Pierpaoli C, Basser PJ (2003) A Comprehensive Approach for Multi-channel Image Registration. Biomedical Image Registration: 214–223. Available: http://www.springerlink.com/index/TCDPF15V3A0CF666.pdf.

52. Rohde GK, Pajevic S, Pierpaoli C (2004) Multi-channel registration of diffusion tensor images using directional information. ISBI: 712–715. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1398637.

53. Li Y, Verma R (2011) Multichannel image registration by feature-based information fusion. IEEE Transactions on Medical Imaging 30: 707–720.

54. Skerl D, Likar B, Pernus F (2006) A protocol for evaluation of similarity measures for rigid registration. Medical Imaging, IEEE Transactions on 25: 779–791. doi:10.1109/TMI.2006.874963.

55. Moigne J Le, Netanyahu N, Eastman R (2011) Image registration for remote sensing. Cambridge: Cambridge University Press.

56. Chen HM, Lee S, Rao RM, Slamani MA, Varshney PK (2005) Imaging for concealed weapon detection: a tutorial overview of development in imaging sensors and processing. Signal Processing Magazine, IEEE 22: 52–61.

57. Verstockt S, Poppe C, Van Hoecke S, Hollemeersch C, Merci B, et al. (2011) Silhouette-based multi-sensor smoke detection. Machine Vision and Applications 23: 1243–1262. doi:10.1007/s00138-011-0359-3.

58. Han J, Bhanu B (2007) Fusion of color and infrared video for moving human detection. Pattern Recognition 40: 1771–1784. doi:10.1016/j.patcog.2006.11.010.

59. Zhao J, Cheung SS (2012) Human segmentation by geometrically fusing visible-light and thermal imageries. Multimedia Tools and Applications. doi:10.1007/s11042-012-1299-2.

60. Krotosky SJ, Trivedi MM (2007) Mutual information based registration of multimodal stereo videos for person tracking. Computer Vision and Image Understanding 106: 270–287. doi:10.1016/j.cviu.2006.10.008.

61. Torabi A, Bilodeau G-A (2013) Local self-similarity-based registration of

human ROIs in pairs of stereo thermal-visible videos. Pattern Recognition 46: 578–589. doi:10.1016/j.patcog.2012.07.026.

62. Han J, Pauwels E, Zeeuw P De (2012) Visible and infrared image registration employing line-based geometric analysis. Computational Intelligence for Multimedia Understanding 7252: 114–125.

63. Han J, Pauwels EJ, de Zeeuw P (2013) Visible and infrared image registration in man-made environments employing hybrid visual features. Pattern Recognition Letters 34: 42–51. doi:10.1016/j.patrec.2012.03.022.

64. Kim YS, Lee JH, Ra JB (2008) Multi-sensor image registration based on intensity and edge orientation information. Pattern Recognition 41: 3356–3365. doi:10.1016/j.patcog.2008.04.017.

65. Lee JH, Kim YS, Lee D, Kang D, Ra JB (2010) Robust CCD and IR image registration using gradient-based statistical information. IEEE Signal Processing Letters 17: 347–350. doi:10.1109/LSP.2010.2040928.

66. Jarc A, Perš J, Rogelj P, Perše M, Kovacic S (2007) Texture features for affine registration of thermal (FLIR) and visible images. Computer Vision WinterWorkshop: 1–7.

67. Ghantous M, Ghosh S, Bayoumi M (2009) A multi-modal automatic image registration technique based on complex wavelets. 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE. pp. 173–176. doi:10.1109/ICIP.2009.5414196.

68. De Vylder J, Douterloigne K, Vandenbussche F, Van Der Straeten D, Philips W (2012) A non-rigid registration method for multispectral imaging of plants. Proc. SPIE. Vol. 8369. pp. 836907–836907–8. doi:10.1117/12.918752.

69. Bilodeau G, St-Onge P, Garnier R (2011) Silhouette-based features for visible-infrared registration. Computer Vision and Pattern Recognition Workshop.

70. Chen H, Varshney P (2001) Automatic two-stage IR and MMW image registration algorithm for concealed weapons detection. Vision, Image and Signal Processing, IEEE Proceedings 148: 209–216. doi:10.1049/ip-vis.

71. Torabi A, Massé G, Bilodeau G-A (2012) An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications. Computer Vision and Image Understanding 116: 210–221. doi:10.1016/j.cviu.2011.10.006.

72. Peng H (2008) Bioimage informatics: a new area of engineering biology. Bioinformatics 24: 1827–1836.

73. Swedlow JR, Eliceiri KW (2009) Open source bioimage informatics for cell biology. Trends in Cell Biology 19: 656–660. Available: http://www.ncbi.nlm.nih.gov/pubmed/19833518.

74. Swedlow JR, Goldberg IG, Eliceiri KW (2009) Bioimage informatics for experimental biology. Annual review of biophysics 38: 327–346.

75. Megason SG, Fraser SE (2007) Imaging in systems biology. Cell 130: 784–795. Available: http://www.ncbi.nlm.nih.gov/pubmed/17803903.

76. Starkuviene V, Pepperkok R (2007) The potential of high-content high-throughput microscopy in drug discovery. British Journal of Pharmacology 152: 62–71. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1978277&tool=pmcentrez&rendertype=abstract.

77. Verveer PJ, Bastiaens PIH (2008) Quantitative microscopy and systems biology: seeing the whole picture. Histochemistry and Cell Biology 130: 833–843. Available: http://www.ncbi.nlm.nih.gov/pubmed/18830616.

78. Rieger B, Molenaar C, Dirks RW, Van Vliet LJ (2004) Alignment of the cell nucleus from labeled proteins only for 4D in vivo imaging. Microscopy Research and Technique 64: 142–150. Available: http://www.ncbi.nlm.nih.gov/pubmed/15352085.

79. Wilson CA, Theriot JA (2006) A correlation-based approach to calculate rotation and translation of moving cells. Image Processing, IEEE Transactions on 15: 1939–1951. Available: http://www.ncbi.nlm.nih.gov/pubmed/16830914.

80. Matula P, Matula P, Kozubek M, Dvorák V (2006) Fast point-based 3-D alignment of live cells. IEEE Transactions on Image Processing 15: 2388–2396. Available: http://www.ncbi.nlm.nih.gov/pubmed/16900692.

81. Wang YP (2004) M-FISH image registration and classification. Biomedical Imaging Nano to Macro 2004 IEEE International Symposium on. IEEE. pp. 57–60. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1398473.

82. Kim I, Yang S, Baccon P, Heard E, Chen YC, et al. (2007) Non-rigid temporal registration of 2d and 3d multi-channel microscopy image sequences of human cells. Biomedical Imaging From Nano to Macro 2007 ISBI 2007 4th IEEE International Symposium on: 1328–1331. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4193539.

83. Thirion JP (1998) Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2: 243–260. Available: http://linkinghub.elsevier.com/retrieve/pii/S1361841598800224.

84. Costa JM, Grant OM, Chaves MM (2013) Thermography to explore plant-environment interactions. Journal of Experimental Botany. Available: http://www.ncbi.nlm.nih.gov/pubmed/23599272. Accessed 1 October 2013.

85. Idso SB, Jackson RD, Pinter Jr PJ, Reginato RJ, Hatfield JL (1981) Normalizing the stress-degree-day parameter for environmental variability. Agricultural Meteorology 24: 45–55. doi:10.1016/0002-1571(81)90032-7.

86. Jackson RD, Idso SB, Reginato RJ, Pinter PJ (1981) Canopy temperature as a crop water-stress indicator. Water Resources Research 17: 1133–1138.

87. Alchanatis V, Cohen Y, Cohen S, Moller M, Sprinstin M, et al. (2009) Evaluation of different approaches for estimating and mapping crop water status in cotton with thermal imaging. Precision Agriculture 11: 27–41.

88. Reinert S, Bögelein R, Thomas FM (2012) Use of thermal imaging to determine leaf conductance along a canopy gradient in European beech (Fagus sylvatica). Tree Physiology 32: 294–302.

89. Maes WH, Steppe K (2012) Estimating evapotranspiration and drought stress with ground-based thermal remote sensing in agriculture: a review. Journal of Experimental Botany 63: 4671–4712. doi:10.1093/jxb/ers165.

90. Jones HG (1992) Plants and Microclimate: a quantitative approach to environmental plant physiology. 2nd ed. Cambridge: Cambridge University Press.

91. Jones HG (1999) Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surfaces. Plant, Cell and Environment 22: 1043–1055. doi:10.1046/j.1365-3040.1999.00468.x.

92. Leinonen I, Grant OM, Tagliavia CPP, Chaves MM, Jones HG (2006) Estimating stomatal conductance with thermal imagery. Plant, Cell and Environment 29: 1508–1518.

93. Moran MS (1994) Irrigation management in Arizona using satellites and airplanes. Irrigation Science 15: 35–44.

94. Grant OM, Chaves MM, Jones HG (2006) Optimizing thermal imaging as a technique for detecting stomatal closure induced by drought stress under greenhouse conditions. Physiologia Plantarum 127: 507–518. doi:10.1111/j.1399-3054.2006.00686.x.

95. Grant OM, Tronina L, Jones HG, Chaves MM (2007) Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes. Journal of Experimental Botany 58: 815–825. doi:10.1093/jxb/erl153.

96. Jones HG (2002) Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine. Journal of Experimental Botany 53: 2249–2260. doi:10.1093/jxb/erf083.

97. Möller M, Alchanatis V, Cohen Y, Meron M, Tsipris J, et al. (2007) Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. Journal of Experimental Botany 58: 827–838. doi:10.1093/jxb/erl115.

98. Leinonen I, Jones HG (2004) Combining thermal and visible imagery for estimating canopy temperature and identifying plant stress. Journal of Experimental Botany 55: 1423–1431. doi:10.1093/jxb/erh146.

99. Cohen Y, Alchanatis V, Prigojin A, Levi A, Soroker V (2011) Use of aerial thermal imaging to estimate water status of palm trees. Precision Agriculture 13: 123–140.

100. Oerke E-C, Gerhards R, Menz G (2010) Precision Crop Protection - the Challenge and Use of Heterogeneity. Springer.

101. Chaerle L, Caeneghem W Van, Messens E (1999) Presymptomatic visualization of plant-virus interactions by thermography. Nature 17: 813–816.

102. Chaerle L, De Boever F, Montagu M Van, Straeten D Van Der (2001) Thermographic visualization of cell death in tobacco and Arabidopsis. Plant, Cell & Environment 24: 15–25. Available: http://doi.wiley.com/10.1046/j.1365-3040.2001.00654.x.

103. Chaerle L, Leinonen I, Jones HG, Van Der Straeten D (2007) Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. Journal of Experimental Botany 58: 773–784.

104. Oerke E-C, Steiner U, Dehne H-W, Lindenthal M (2006) Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions. Journal of Experimental Botany 57: 2121–2132. Available: http://www.ncbi.nlm.nih.gov/pubmed/16714311. Accessed 24 August 2011.

105. Lindenthal M, Steiner U, Dehne H-W, Oerke E-C (2005) Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography. Phytopathology 95: 233–240.

106. Oerke E-C, Fröhling P, Steiner U (2011) Thermographic assessment of scab disease on apple leaves. Precision Agriculture 12: 699–715. doi:10.1007/s11119-010-9212-3.

107. Stoll M, Schultz HR, Berkelmann-Loehnertz B (2008) Thermal sensitivity of grapevine leaves affected by Plasmopara viticola and water stress. Vitis-

Geilweilerhof 47: 133.

108. Oerke E-C, Steiner U (2010) Potential of digital thermography for disease control. Precision Crop Protection—the Challenge and use of Heterogeneity. Dordrecht, Netherlands: Springer Netherlands. pp. 167–182.

109. Lenthe J-H, Oerke E-C, Dehne H-W (2007) Digital infrared thermography for monitoring canopy health of wheat. Precision Agriculture 8: 15–26. doi:10.1007/s11119-006-9025-6.

110. Hartley R, Zisserman A (2000) Multiple view geometry in Computer Vision. Cambridge University Press.

111. Heikkila J, Silven O (1997) A four-step camera calibration procedure with implicit image correction. Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. pp. 1106–1112. doi:10.1109/CVPR.1997.609468.

112. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47: 7–42. doi:10.1023/A:1014573219977.

113. Fua P (1991) Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. Proceedings of the 12th international Joint Conference on Artificial Intelligence: 1292–1298.

114. Hoff W, Ahuja N (1989) Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11: 121–136.

115. Konolige K (1998) Small vision systems: hardware and implementation. Robotics Research International.

116. Birchfield S, Tomasi C (1998) A pixel dissimilarity measure that is insensitive to image sampling. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20: 401–406.

117. Birchfield S, Tomasi C (1999) Depth discontinuities by pixel-to-pixel stereo. International Journal of Computer Vision 35: 269–293.

118. Kolmogorov V, Zabih R (2002) Multi-camera scene reconstruction via graph cuts. Computer Vision—ECCV 2002.

119. Boykov Y, Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. Pattern Analysis and Machine Intelligence, IEEE Transactions on 26: 1124–1137. Available: http://www.ncbi.nlm.nih.gov/pubmed/15742889.

120. Kolmogorov V, Zabih R (2001) Computing visual correspondence with occlusions using graph cuts. Proceedings Ninth IEEE International Conference on Computer Vision 2: 508–515. doi:10.1109/ICCV.2001.937668.

121. Sun J, Zheng N, Shum H-Y (2003) Stereo matching using belief propagation. Pattern Analysis and Machine Intelligence, IEEE transactions on 25: 787–800. doi:10.1109/TPAMI.2003.1206509.

122. Felzenszwalb PF, Huttenlocher DP (2006) Efficient Belief Propagation for Early Vision. International Journal of Computer Vision 70: 41–54. doi:10.1007/s11263-006-7899-4.

123. Klaus A, Sormann M, Karner K (2006) Segment-Based stereo matching using Belief Propagation and a self-adapting dissimilarity measure. 18th International Conference on Pattern Recognition (ICPR'06): 15–18. doi:10.1109/ICPR.2006.1033.

124. Hirschmüller H (2005) Accurate and efficient stereo processing by semi-global matching and mutual information. Vision and Pattern Recognition, 2005 CVPR 2: 807–814. doi:10.1109/CVPR.2005.56.

125. Hirschmüller H (2008) Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE transactions on 30: 328–341. doi:10.1109/TPAMI.2007.1166.

126. Kim J, Kolmogorov V, Zabih R (2003) Visual correspondence using energy minimization and mutual information. Proceedings Ninth IEEE International Conference on Computer Vision. IEEE. pp. 1033–1040 vol.2. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1238463. Accessed 30 May 2012.

127. Hirschmüller H, Scharstein D (2007) Evaluation of cost functions for stereo matching. Computer Vision and Pattern Recognition, IEEE Conference on. IEEE. pp. 1–8. doi:10.1109/CVPR.2007.383248.

128. Hirschmüller H, Scharstein D (2009) Evaluation of stereo matching costs on images with radiometric differences. IEEE transactions on Pattern Analysis and Machine Intelligence 31: 1582–1599. doi:10.1109/TPAMI.2008.221.

129. Yang Q, Wang L, Ahuja N (2010) A constant-space belief propagation algorithm for stereo matching. Computer Vision and Pattern Recognition (CVPR), IEEE Conference on: 1458–1465. doi:10.1109/CVPR.2010.5539797.

130. Yang Q (2012) A non-local cost aggregation method for stereo matching. IEEE Conference on Computer Vision and Pattern Recognition: 1402–1409. doi:10.1109/CVPR.2012.6247827.

131. Grauer-Gray S, Kambhamettu C (2009) Hierarchical belief propagation to reduce search space using cuda for stereo and motion estimation. Applications of Computer Vision (WACV), 2009 Workshop on. pp. 1–8.

132. Mei X, Sun X, Dong W, Wang H, Zhang X (2013) Segment-Tree based Cost Aggregation for Stereo Matching. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 313–320.

133. Min D, Lu J, Do M (2013) Joint Histogram Based Cost Aggregation for Stereo Matching.

134. Ivanov N, Boissard P, Chapron M, Andrieu B (1995) Computer stereo plotting for 3-D reconstruction of a maize canopy. Agricultural and Forest Meteorology 1923. Available: http://www.sciencedirect.com/science/article/pii/016819239402204W. Accessed 29 October 2013.

135. Andersen HJ, Reng L, Kirk K (2005) Geometric plant properties by relaxed stereo vision using simulated annealing. Computers and Electronics in Agriculture 49: 219–232. doi:10.1016/j.compag.2005.02.015.

136. Biskup B, Scharr H, Schurr U, Rascher UWE (2007) A stereo imaging system for measuring structural parameters of plant canopies. Plant, Cell & Environment 30: 1299–1308. doi:10.1111/j.1365-3040.2007.01702.x.

137. Song Y, Wilson R, Edmondson R, Parsons N (2007) Surface modelling of plants from stereo images. Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007). IEEE. pp. 312–319.

138. Alenya G, Dellen B, Torras C (2011) 3d modelling of leaves from color and tof data for robotized plant measuring. Robotics and Automation (ICRA), 2011 IEEE International Conference on. pp. 3408–3414.

139. Song Y, Glasbey CA, van der Heijden GWAM, Polder G, Dieleman JA (2011) Combining stereo and Time-of-Flight images with application to automatic plant phenotyping. Image Analysis. Springer. pp. 467–478.

140. Omasa K, Hosoi F, Konishi A (2007) 3D lidar imaging for detecting and understanding plant responses and canopy structure. Journal of Experimental Botany 58: 881–898. doi:10.1093/jxb/erl142.

141. Rosell JR, Sanz R (2012) A review of methods and applications of the geometric characterization of tree crops in agricultural activities. Computers and Electronics in Agriculture 81: 124–141. doi:10.1016/j.compag.2011.09.007.

142. Van der Heijden G, Song Y, Horgan G, Polder G, Dieleman A, et al. (2012) SPICY: towards automated phenotyping of large pepper plants in the

greenhouse. Functional Plant Biology.

143. Torabi A, Bilodeau G (2011) Local self-similarity as a dense stereo correspondence measure for themal-visible video registration. Computer Vision and Pattern ….

144. Morris NJW, Avidan S, Matusik W, Pfister H (2007) Statistics of infrared images. IEEE Conference on Computer Vision and Pattern Recognition: 1–7. doi:10.1109/CVPR.2007.383003.

145. Nashat S, Abdullah A, Abdullah MZ (2011) A stationary wavelet edge detection algorithm for noisy images.

146. Kong SG, Heo J, Boughorbel F, Zheng Y, Abidi BR, et al. (2006) Multiscale fusion of visible and thermal IR images for illumination-invariant Face Recognition. International Journal of Computer Vision 71: 215–233. Available: http://link.springer.com/10.1007/s11263-006-6655-0. Accessed 7 August 2013.

147. Olivo-Marin J-C (2002) Extraction of spots in biological images using multiscale products. Pattern Recognition 35: 1989–1996. doi:10.1016/S0031-3203(01)00127-3.

148. Pajares G, Manuel de la Cruz J (2004) A wavelet-based image fusion tutorial. Pattern Recognition 37: 1855–1872. doi:10.1016/j.patcog.2004.03.010.

149. Chen S, Su H, Zhang R, Tian J (2008) Fusing remote sensing images using à trous wavelet transform and empirical mode decomposition. Pattern Recognition Letters 29: 330–342. doi:10.1016/j.patrec.2007.10.013.

150. Nason G, Silverman B (1995) The stationary wavelet transform and some statistical applications. Wavelets and Statistics: 1–19.

151. Morse BS (2000) Lecture 11 : Differential Geometry.

152. Dima A, Scholz M, Obermayer K (2002) Automatic segmentation and skeletonization of neurons from confocal microscopy images based on the 3-D wavelet transform. IEEE transactions on Image Processing : a publication of the IEEE Signal Processing Society 11: 790–801. doi:10.1109/TIP.2002.800888.

153. Fowler J (2005) The redundant discrete wavelet transform and additive noise. Signal Processing Letters, IEEE 853048: 629–632.

154. Kale K V (2008) Advances In Computer Vision And Information Technology. I.K. International Publishing House Pvt. Limited.

155. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic

diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12: 629–639.

156. Audet C, Dennis JE (2002) Analysis of Generalized Pattern Searches. SIAM Journal on Optimization 13: 889. Available: http://link.aip.org/link/SJOPE8/v13/i3/p889/s1&Agg=doi.

157. Lee S, Wolberg G, Shin SY (1997) Scattered data interpolation with multilevel B-splines. IEEE Transactions on Visualization and Computer Graphics 3: 228–244. doi:10.1109/2945.620490.

158. Trucco E, Verri A (1998) Introductory techniques for 3-D Computer Vision. Prentice Hall.

159. Atherton T, Kerbyson D (1999) Size invariant circle detection. Image and Vision computing 17: 795–803.

160. Song Y (2009) Modelling and analysis of plant image data for crop growth monitoring in Horticulture University of Warwick.

161. Welch G, Bishop G (1995) An introduction to the Kalman filter.

162. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23: 1222–1239.

163. Felzenszwalb PF, Huttenlocher DP (n.d.) Efficient belief propagation for early vision. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 CVPR 2004 1: 261–268. doi:10.1109/CVPR.2004.1315041.

164. Weiss Y, Freeman WT (2001) On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. Information Theory, IEEE Transactions on 47: 736–744.

165. Eikvil L (1993) Optical Character Recognition. Oslo.

166. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20: 273–297.

167. Rasmussen C (2004) Gaussian processes in machine learning. Advanced Lectures on Machine Learning: 63–71.

168. Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning. Cambridge: MIT Press.

169. Haranadh G, Sekhar CC (2008) Hyperparameters of Gaussian process as features for trajectory classification. Neural Networks (IEEE World

Congress on Computational Intelligence), IEEE International Joint Conference on: 2195–2199.

170. Sammut C, Webb GI (2010) Encyclopedia of machine learning. Springer.

171. Bazi Y, Melgani F (2010) Gaussian process approach to remote sensing image classification. Geoscience and Remote Sensing, IEEE Transactions on 48: 186–197.

172. Ebden M (2008) Gaussian Processes for Regression : A Quick Introduction.

173. Friedenberger M, Bode M, Krusche A, Schubert W (2007) Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures. Nature protocols 2: 2285–2294. Available: http://www.ncbi.nlm.nih.gov/pubmed/17853885. Accessed 26 June 2010.

174. Bhattacharya S (2011) Identification of colon cancer specific protein clusters by toponome imaging system and their cellular annotation by using a three symbol code University of Warwick.

175. Pertuz S, Puig D, Garcia MA (2013) Analysis of focus measure operators for shape-from-focus. Pattern Recognition 46: 1415–1432. doi:10.1016/j.patcog.2012.11.011.

176. Yang S, Kohler D, Teller K, Cremer T, Le Baccon P, et al. (2008) Nonrigid registration of 3-d multichannel microscopy images of cell nuclei. IEEE Transactions on Image Processing 17: 493–499. Available: http://www.ncbi.nlm.nih.gov/pubmed/18390358.

177. Likar B (2001) A hierarchical approach to elastic registration based on mutual information. Image and Vision Computing 19: 33–44. Available: http://linkinghub.elsevier.com/retrieve/pii/S0262885600000536.

178. Kolda TG, Lewis RM, Torczon V (2006) A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints. Energy.

179. Keys RG (1981) Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics Speech and Signal Processing 29: 1153–1160. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1163154.

180. Khan AM, Humayun A, Raza S-E-A, Khan M, Rajpoot NM (2012) A novel paradigm for mining cell phenotypes in multi-tag Bioimages using a locality preserving nonlinear embedding. The International Conference on Neural Information Processing (ICONIP).

181. Khan AM, Raza S-E-A, Khan M, Rajpoot NM (2013) Cell phenotyping in multi-tag fluorescent bioimages. Neurocomputing.

182. Humayun A, Raza S-E-A, Waddington C, Abouna S, Khan M, et al. (2011) A Framework for molecular co-expression pattern analysis in multi-channel Toponome fluorescence images. Microscopic Image Analysis with Applications in Biology (MIAAB 2011). Heidelberg.

183. Evans RG, Naidu B, Rajpoot NM, Epstein D, Khan M (2012) Toponome imaging system: multiplex biomarkers in oncology. Trends in molecular medicine 18: 723–731. Available: http://www.ncbi.nlm.nih.gov/pubmed/23122853. Accessed 25 November 2013.

184. Kölling J, Langenkämper D, Abouna S, Khan M, Nattkemper TW (2012) WHIDE--a web tool for visual data mining colocation patterns in multivariate bioimages. Bioinformatics (Oxford, England) 28: 1143–1150. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324520&tool=pmcentrez&rendertype=abstract. Accessed 25 November 2013.

185. Kovacheva VN, Khan AM, Khan M, Epstein D, Rajpoot NM (2013) DiSWOP: A Novel Measure for Cell-Level Protein Network Analysis in Localised Proteomics Image Data. Bioinformatics: 1–8. doi:10.1093/bioinformatics/btt676.

186. Ward Jr JH (1963) Hierarchical grouping to optimize an objective function. Journal of the American statistical association 58: 236–244.

187. Al-Kofahi Y, Lassoued W, Lee W, Roysam B (2010) Improved automatic detection and segmentation of cell nuclei in histopathology images. Biomedical Engineering, IEEE Transactions on 57: 841–852.