Environmental Analysis through integration of Geographical Information and Machine Vision systems

by

Paul D. Kelly, M.Eng.

A thesis presented on application for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Engineering

The Queen's University of Belfast

School of Electrical & Electronic Engineering,

May 2004.

Chapter 1

Introduction

1.1 Types of Vision-based Analysis

Environmental analysis may be performed at various levels of detail. For example, satellite imaging such as that obtained from the current Landsat satellite provides an overhead view of the Earth's surface in strips corresponding to its orbit. But this provides relatively low-resolution images (several metres at best) and even with forthcoming high-resolution satellite images (resolution $< 1\,\mathrm{m}$) it is still only the surface canopy of vegetation, roofs of houses, etc. that are visible. Also, satellite images are only usable for remote sensing of ground if there is no cloud, which is much less likely for some parts of the world than others (Lunetta and Elvidge, 1999). Much closer to ground level, detailed analysis of e.g. individual plant or soil samples can be performed, to provide much information but highly localised to a small area and without much automation. In fact, researchers in these areas may wish to avoid excessive automation and may be more interested in analysing their data manually to 'make the most of it' (Anderson, 2000).

The gap between these two extremes, i.e. general surveying of the environmental landscape on a more 'human-sized' scale, is often currently filled by labour-intensive manual surveying or even manual subjective analysis of scenes, if the expense of a full survey is not justified. Some examples of the types of environmental situations that may be assessed are:

- erosion, e.g. of cliff faces or paths
- pollution, e.g. acid rain effects on trees and plants
- vandalism, e.g. damage to street furniture or rare plants
- planning violations, e.g. unauthorised new building
- fires, e.g. assessing extent of forest fires

The types of environmental resources being monitored may be, for example:

- forests
- special conservation areas, e.g. Northern Ireland ASSIs (Areas of Special Scientific Interest)
- urban landscapes
- coastal features

This thesis will explore methods to fill this gap in surveying techniques using economic camera-based systems and sensor fusion.

1.2 Data Structure for Objective Change Description

When considering methods of objectively describing change as detected by an automatic system, it is important to make the distinction between raster and vector data (Burrough, 1986; Russ, 1994). A 'rasterised' image space is divided up into a rectangular array of small cells or pixels. Raster data is a pixel-level attribute and describes the quantity measured by the transducer used to gather the data. A camera will measure the colour, or, more precisely, the spectral content of the light reflected from objects, as defined by three filter types corresponding to the red, green and blue components.

Vector data is an object-level attribute and describes the relative position and size of an object in the image frame (which can be transformed to a position in the world co-ordinate system). It typically consists of lines and points representing the outlines of objects in a scene.

One of the goals of a change detection system should be to manipulate and convert the data formats so that the different sources can be directly compared to each other. *Geo-referencing* the data, so that everything is referenced to the same co-ordinate framework, is an important pre-requisite for this.

1.3 Existing Methods of Change and Feature Analysis

'Change-detection systems' based on manual subjective analysis of scenes are numerous; the term may be applied to any situation where a person notices that a scene looks different and comprehends that something in particular has changed. Some examples of more sophisticated systems that still require manual measurements are

- Environmental Impact Assessment of a major construction project (Munn, 1979). Often conventional surveying techniques (Uren and Price, 1994) may be used and a map or plan created. This would be used as a base-line for follow-up monitoring of compliance with any requirements imposed.

- Road lighting Assessment. This is an example of a process that was originally carried out by subjective manual assessment, but more recent devlopments (Todd, 1990) have introduced partly automated systems, for example using a lux meter to measure street light output. This means the result of the assessment is more objective, however manual intervention to gather and process the measurements is still necessary.

There is much scope to use automated analysis based on image processing to fill the gap between these two extremes (subjective interpretation and detailed surveying). For example, it may give a better estimate of the amount of change in an area and indicate if detailed re-surveying is necessary. This is also likely to be a more economic approach, where the expense of manual surveying is only entailed if absolutely necessary.

More recently, complex and advanced systems have been designed for specific applications. Examples of these are

- GPS Van for roadway assessment and mapping (Novak, 1995). This uses precise positioning derived from sensor fusion between a dual-frequency kinematic GPS receiver, inertial sensor, gyroscopes and a wheel revolution counter. Cameras and other sensors are mounted on the van and the data gathered is post-processed using photogrammetric techniques. The process involves manual intervention and the end result is a vector dataset containing the locations and attributes of all the permanent roadside features in an area (roadsigns, streetlights etc.). This service is now a commercial product in the USA (Transmap Corp., 2003); as a result of the complex equipment used and manual effort involved in processing the data it is expensive. The data gathered using the system is integrated with existing GIS data for final presentation and visualisation.

- Shoreline erosion modelling system (Holland et al., 1997). This system makes use of a calibrated video camera and perspective transformations to compare successive images and detect change in the shoreline and wave effects over a period of time. It automatically compensates for changing light intensity, however manual final processing and quality control analysis of the data always takes place. The system is permanently installed in a fixed location and so position calculation and tracking is not an issue. The system is available as a commercial product and has been installed at several locations around the world (NorthWest Research Associates Inc., 2001).

While change detection is an integral part of the second system, the first system was designed as a mobile mapping system (Bossler et al., 1991) and does not have this explicit capability. Change detection is only possible at the object-based level, i.e. whether an object (e.g. a traffic sign or kerb) has been moved or removed. It is unclear whether a raster description (i.e. image) of the object is preserved in the database.

1.4 Scope for Application of New Methods

The automatic systems described so far have either been stationary or required complex positioning hardware (e.g. dual frequency kinematic GPS systems) to determine the camera location and attitude (i.e. the *exterior orientation parameters*). Software-based image processing systems have been developed, which make use of landmarks visible in the image and their correspondence with a database of known artefact positions to estimate the camera exterior orientation. Examples of these systems (the individual algorithms used are described in more detail in Section 1.5.4) are

- Aircraft Position Determination based on oblique view of runway lighting configuration (Chatterji et al., 1998). The aim of this is to create a system that can be used to aid aeroplane pilots to land at night. Interpretation of the runway lighting configuration as seen through a camera on-board the plane, together with fusion of data from an inexpensive GPS system, enables the plane's position *relative to the runway* to be determined accurately. The algorithm used in the system requires prior knowledge of the correspondence between each light in the image and its ground co-ordinates.

- Aerial Photograph Orientation using GIS data and prominent topographic features in the image (Habib and Kelley, 2001). An algorithm and not a complete system is proposed here, but good matching results are shown although it is very computationally intensive. When initial approximations of the camera orientation are known, e.g. from a GPS / inertial measurement system, the algorithm has very good potential to provide an accurate automatic position determination system. It has the advantage of not requiring prior knowledge of the correspondence between the features in the image and the GIS data.

If these types of system were applied to a general environmental scene, it would enable *direct geo-referencing* of all the objects visible in the scene, although this assumes the camera has already been calibrated for its internal orientation parameters (Section 2.6). Together with the external parameters and GIS ground elevation data, a 3-D world position can be calculated directly for every pixel or object in an image of the scene (Section 4.3). Data from multiple images (representing different viewing angles, times and lighting conditions; Section 4.4.4) may be combined in a database.

Improvements continue to be made on the mobile mapping systems that were first substantially realised by Bossler et al. (1991). Those by Tao et al. (2001) (see more detailed discussion in Section 1.5.4) incorporate many useful and relevant techniques and ideas with regard to geo-referencing and indeed GIS integration. However because of the application, the emphasis is still on using high-end precision GPS / inertial sensor positioning systems and thus all measurements are based on this constraint of known camera motion.

The next section comprises a review of literature relevant to each component of a geo-referenced change-detecting system. Following that the system will be defined in detail.

1.5 State of Current Research (Literature Review)

This research is at the boundaries of many traditionally-defined disciplines. For example, the fields of Geographical Information Science, Photogrammetry and Remote Sensing, Optical Engineering, Intelligent Transportation Systems, Oceanic Engineering and Agricultural Engineering all contain relevant results.

The layout of this section is as follows: for each component of the system the relevant existing work by various researchers will be introduced and put into context with regard to other existing work. Indication will be given as to how the relevant existing work can assist in the goals of this work or how it is deficient and will be improved by this work.

1.5.1 Camera / video image acquisition and initial processing

Whilst the majority of the analysis techniques described in this thesis are independent of the means of image acquisition, it is nonetheless important to consider a suitable camera to use in testing the system and to realise its practical application.

In some of the non-cost-restricted applications described in Section 1.1, a very high resolution digital camera is used in order to obtain the maximum possible amount of information from the scene. For high-end aerial photography conventional film cameras are still used, but in most cases digital charge-coupled device array (CCD) cameras are the normal capture device. A common requirement especially in the systems already described is continuous grabbing of images from the camera. A camera and separate frame-grabber may be used for some applications, however for many applications a commerical video recording camera will be suitable and available at lower cost (albeit with potentially limited image resolution).

Many manual change detection systems augment a subjective description of the change with photographs taken with a conventional camera. Using a video camera is consistent with this principle of using commercially available equipment.

Correction for lens distortion is essential for any photogrammetric operation. Traditionally it has been done within a photogrammetric plotter by mechano-optical methods (Thompson, 1966). When a particular point in an image was being projected, the focal length for the projection would be automatically adjusted slightly depending on the distance the point was from the principal point of the image.

Computer-based image processing techniques, which involve the rasterisation of a photograph into pixels, enable more advanced 'pixel-shifting' distortion correction to be carried out (Tsai, 1987). This approach has developed naturally from the traditional ray-tracing calculations carried out by optical engineers (Conrady, 1960).

Distortion correction techniques have also been developed as part of work on various applied imaging problems, e.g. coastal erosion monitoring (Holland et al., 1997) and medical imaging (endoscopy) (Helferty et al., 2001). These typically involve using knowledge of the camera lens characteristics to develop a linear mathematical model of the distortion, as a function of the pixel location within the image. Each pixel can then, for example, be shifted to its correct undistorted position, or the model can be used to correct the image-space co-ordinates of objects after their identification by image analysis.

Modern cameras are almost all capable of capturing colour image data, rather than just a panchromatic brightness level over the whole visible spectrum (i.e. 'black and white'). However colour is a somewhat abstract concept; at a lower level, the relationship between wavelengths of reflected light and the various 3-dimensional colour spaces has been well defined (Wyszecki and Stiles, 1982). The RGB colour space used in most cameras is based on the standards originally developed for colour television (Sproson, 1983), and loosely based on the way the human eye works. Reduction of the full visible spectrum to three colour components represents a large reduction in the information content.

Precise knowledge of the colour space used in the camera is therefore important to be able to make maximum use of this limited information for natural feature recognition. If full spectral reflectance information is desired, it could be obtained by means of filters or using a field spectrometer or similar instrument (Anderson, 2000). This would be important if natural object recognition was a high priority; certain natural features exhibit very specific spectral reflectance characteristics (Krinov, 1953). The spectrometer used in these types of measurement is limited however in its resolution and capability for covering a large area.

Calibration of the camera's response to varying luminance is important when using images of outdoor scenes. It is necessary to take account of changes in ambient solar illumination to enable accurate change detection. McMenemy et al. (2003) summarise the methods and techniques for calibration of a commercial digital video camera in order to use it for light source measurement; some components of this work are useful for determining the ambient illumination in an area.

The importance of focal length, film dimensions and other camera parameters has been well defined since the earliest studies of photogrammetry (Thompson, 1966). With the coming of digital cameras and the widespread availability of cheap photogrammetry, these parameters have been derived again and summarised as appropriate for various application areas, e.g. Holland et al. (1997), McMenemy et al. (2001) or Agouris et al. (2000).

Old or historic photographs may be used as a base-line for change detection. Known objects in the scene will be very important for calculating the camera position, and for famous photographs or those taken by well-known photographers it is possible in many cases to find details of the camera used and base the calibration on this information.

'Repeat Photography' is the technique whereby a landscape is re-photographed from the same location as an existing old photograph in order to easily identify the changes that have taken place. Many papers have been published on the subjectively-interpreted results of repeat photography (e.g. a good bibliography was compiled by Hart and Laycock (1996)). Not much work has been published on automation of the process or the techniques themselves, which often involve trial and error and manual interpretation of parallax error; Butler (1994) provides an introduction to the methods currently in use by geographers. The application of photogrammetric techniques to the problem has recently been investigated with some success (Strausz, 2001). However this work has not yet been fully developed nor successfully applied in practice.

1.5.2 Geographical data sources and GIS

Mobile systems that have previously only been used in laboratory conditions are being extended to operate in the 'real world' outdoors. There are many examples of this in the robotics literature, where taking account of the unknown or hard-to-model outdoor terrain forms a central part of the work. For example, Bertozzi et al. (1998a) guided an autonomous vehicle along roads successfully initially by assuming the terrain was flat, but developed methods of compensating for the slope of non-flat roads by analysing the deviations from the flat ground assumption. Anderson et al. (2003) used geographical information, i.e. a digital terrain model, to predict line-of-sight visiblity for mobile robots (in order to facilitate radio communication).

While the second example above is a useful application and very relevant to the field of mobile robotics, it nonetheless amounts to a standard line-of-sight analysis with additional manipulation of the results. This could be carried out within a Geographical Information System. Acquiring and interpreting / importing the geographical data, and converting it to a form suitable for the main processing task inevitably forms a substantial part of systems such as this. However a GIS will already include import and export filters for all the common formats geographical data is normally supplied in. In applications where the need to interface with existing geographical data is more obvious, e.g. change detection, many researchers are already integrating GIS with their work (Agouris et al., 2000).

Geographical Information Systems have become increasingly widely used since the 1970s for manipulating spatial data. The advantages of an electronic database (with paper maps used only for presentation of data from the database) over the traditional system where the paper map *was* the ‘database’ are enormously wide-ranging and well documented (Burrough, 1986). The ease and convenience with which the geographical data can be re-used and used as leverage for diverse purposes is greatly enhanced when it is stored in a GIS.

A characteristic of most GIS is the distinct separation between two-dimensional data as would be shown on a conventional plan or map, and information on the third dimension, i.e. ground elevation data. Nowadays GIS is used widely for planning and resource management, but most of this widespread usage amounts only to automated cartography, e.g. in many applications the possiblities of three-dimensional data are not even considered and GIS is just used as an automated way of working with conventional two-dimensional maps (Burrough, 1986; O’Sullivan and Unwin, 2003). Of course the way data are provided and marketed by the mapping agencies also has a bearing on how it is used, and 2-D vector topological map data and either vector contours or raster elevation models are usually quite separate data products (Perkins and Parry, 1996).

The more advanced analysis capabilities of GIS are used by researchers in environmental monitoring and GIS applications, etc. The field of improving and developing the capabilities of GIS (rather than studying its application) is known as geographical information science. Traditional GIS such as Arc/Info (late 1970s) and GRASS (early 1980s) have extensive data analysis capabilities and are very suitable for the type of development carried out as part of geographical information science.

More recent developments in GIS software are primarily data viewer applications that take advantage of developments in graphical user interfaces (e.g. Arc/View, early 1990s). These types of software have recently gained more data analysis capabilities but a traditional analysis-oriented GIS is still required for serious work.

As mentioned above, the elevation data provided by most mapping agencies (including the Ordnance Survey of Northern Ireland) is a separate data product to the main two-dimensional database used in production of paper maps. The latter is a vector product (i.e. lines, points and areas with associated attributes). The former is typically available in a more diverse range of formats, e.g.

- Individual spot heights, e.g. collected by GPS surveying
- Contours (i.e. lines of constant elevation)
- Gridded elevation data (raster format)

Raster grid format, where each cell is allocated an elevation value is one of the most useful to have the data in as it facilitates many GIS analyses, such as 3-D visualisation (Burrough, 1986). This format is known as 2.5-dimensional data and is adequate for modelling many rural areas, since it is very appropriate for representing a 'bare-earth' digital elevation model but not so good at modelling structures above ground. For urban areas a combination of raster bare-earth DEM and vector data for modelling above-ground structures is likely to be more appropriate (see Section 7.2.5).

However, data will typically be obtained in different formats from various different sources, often in different co-ordinate systems / projections, e.g. from national mapping agencies and international data sources such as ASTER (Welch et al., 1998), and combining it into a usable gridded model is a substantial problem. Cebecauer et al. (2002) summarised some interpolation techniques for combining point and vector data into a gridded elevation model but selection of the most appropriate technique depends on the data.

Any other data that may be conveniently represented as a collection of attributes, each one associated with a two-dimensional position on the ground, is suitable for processing by GIS. Vertical aerial photographs are a good example of this, particularly in their ortho-rectified state (see Section 1.5.3).

1.5.3 3-D visualisation and image perspective transformation

Historically, the requirement to represent the 3-D world in 2-D images (on paper or film etc.) has given rise to many innovations and diverse data manipulation techniques in this area. This section discusses the historical development and modern relevance of the techniques used to transform image data between these various 2-D representations of the 3-D world.

Aerial Photograph to 2-D Map

The effective transfer of information from an aerial photograph to a two-dimensional map has been one of the central aims of the field of photogrammetry for well over a century (Laussedat, 1898). Conventionally, complex opto-mechanical precision instruments were used to enable manual tracing of features from two adjacent aerial photographs onto a map. Such instruments, e.g. stereoplotters, can optically re-construct the path of the light rays through the camera to enable precise measurements to be made. Light is projected through photographic diapositives and measurements of the horizontal and vertical offsets of objects in the photgraphs may be made (Thompson, 1966).

The traditional photogrammetric technique that is of most interest and relevance to the current work is that of orthophoto generation. In contrast to the first technique described where only selected features from the photograph are transferred to the map, creation of an orthophoto involves re-projecting *all* the photographic content into the map co-ordinate system. This involves additional correction for the varying ground elevation over the area covered by the photograph as well as transformation from perspective to orthographic projection.

The first mechanical solution to orthophoto generation was patented by Ferber in 1936 and by the 1960s there were dozens of machines of a similar design in use throughout the world (Thompson, 1966). Throughout the 1960s and 1970s the availability of computers and scanners resulted in electronic methods of orthophoto generation being developed. The possibility of pixel re-sampling and interpolation algorithms made it very easy to implement much better quality solutions than was possible with a mechanical system. Another possibility with an electronic system was that oblique aerial photographs could be ortho-rectified (Thompson, 1966; Neteler and Mitasova, 2002). Mechanical systems were unable to correct the high levels of distortion inherent in these images to the degree required for orthophoto generation.

Orthophoto generation software became widely available in geographical information systems in the 1980s and is now a widely available standard technique (Neteler and Mitasova, 2002).

Recent developments in photogrammetry involve the use of a high resolution digital surface model to model individual objects such as buildings on the ground, and the orthophoto can then accurately represent ‘no data’ areas, e.g. that were hidden behind buildings or trees (Sheng et al., 2003). However it is still only a 2.5-D model (albeit a very high resolution one) and cannot model features with overhangs.

Ground-level Photograph to World Co-ordinate Space

The use of ground-level images for surveying and measurement goes back even further than aerial photography, to the very beginning of the science of photogrammetry (Laussedat, 1854) and the use of the camera lucida (a portable light projection device). When aerial photography became available, ground-level photogrammetry became less relevant. It is really only in the 1990s and in the field of image processing that it has come into widespread use for this purpose, driven by applications in environmental monitoring (e.g. Holland et al. (1997)) and autonomous vehicle control (e.g. Bertozzi et al. (1998b), Lutzeler et al. (2001)).

Some traditional users of aerial photogrammetry are now recognising the low-cost and high resolution advantages that ground-level images offer. The mobile mapping systems already mentioned (Section 1.4) replace (and improve on) the use of aerial photography for many urban mapping tasks. The nature of the ground-level images means that some objects in the scene will be very close to the camera, and small errors in the calculated camera position and orientation could lead to large errors in the correspondence with objects in the scene. Typically this is overcome by using extremely accurate positioning hardware (Tao et al., 2001)—see Section 1.5.4—to allow direct geo-referencing of the image data to 3-D world co-ordinates.

The widespread use of 3-D visualisation techniques, especially in GIS software (see next sub-section) in many cases offers a convenient 'reverse mapping' route from ground truth perspective view images back to the plan-view format that image data is often stored in. However until now this has remained largely untapped and in most cases specific solutions have been implemented to solve specific problems, examples of which have been given above.

2-D Image / Orthophoto & Ground Elevation Data to Perspective View

A long-standing visualisation technique in geography involves making a perspective drawing of a scene using elevation information from a contour map. This can be extended by transferring detail from a vertical aerial photograph, e.g. tree line, snow line, vegetation type and other prominent features. This type of drawing is tedious to create manually but can result in an increased understanding of the scene over conventional plan-view representations.

In the late 1970s computing power increased and computer graphics emerged as a research field. Researchers from various different areas turned their attention to use of the new technology for producing perspective views of geographic data, in particular, to combine satellite image data (pre-registered to a 2-D grid—data from the early Landsat satellites was widely available in digital format as it was transmitted from the satellites back to Earth in this format) with gridded elevation data. Examples are Tanaka and Suga (1979) (photogrammetry), Moellering (1980) (geography) and Dungan (1979) (computer graphics). In this early work the aims were primarily to show the capabilities of computer graphics and for simple visualisation purposes (i.e. a replacement for the perspective drawing technique).

As computer-based 3-D visualisation became more popular during the 1980s it was increasingly used out of the traditional context of academic geography. Military work on flight simulation was one important application. Architecture is another example of an important application area; when visualising proposed developments architects often wish to put 3-D CAD models of buildings into context with surrounding geographical data (K2VI, 2004).

Nowadays a lot of GIS software (e.g. GRASS, ERDAS Imagine) includes the capability to create 3-D visualisations of the content of the GIS database. However the main emphasis (and requirement for the majority of users of these features (Mitasova, 2002)) is better visualisation of data for improved human interpretation. Photogrammetric-grade correctness of the resultant image is not usually an absolute requirement and this is reflected in the user interfaces of the visualisation software (see Section 4.2).

In computer graphics these types of techniques in general are known as *texture mapping*. Weinhaus and Devarajan (1997) describe in detail the application of various texture mapping techniques to real-world scenes.

1.5.4 Accurate position and orientation measurement

Useful techniques for accurate position and orientation measurement have developed from applications in several areas, primarily navigation, surveying and photogrammetry. The levels of accuracy required vary between applications; however the accuracy of some of the underlying techniques has improved with time: hence the various techniques have been applied to different application areas at different stages in their development.

In surveying, normally the position of only one point at a time is required to be found, and this is achieved by measuring its relative location to other points and landmarks using levelling, trigonometric heighting and tacheometry techniques. This involves use of theodolites and other precision instruments; the mathematics of the techniques are described in detail by Uren and Price (1994).

Trigonometric heighting and triangulation (Uren and Price, 1994; Cheetham, 1965) were historically also the only techniques used for preparing conventional maps but throughout the first half of the twentieth century aerial photogrammetry came into much more widespread use for this purpose (as was discussed in Section 1.5.3). Normally the film is moved slowly through the camera to compensate for the forward motion of the aeroplane (Thompson, 1966) and the position determination task amounts to calculating the camera location using knowledge of known ground control points (i.e. landmarks visible in the image) and the camera interior orientation parameters. This is done using the perspective projection equations (Thompson (1966) and Section 4.2.3). The ground control points have to be identified manually and it is a very labour-intensive task. Effectively the location of the camera is found relative to the landmarks marking the ground control points. Their location is known in another coordinate system (e.g. the Irish Grid) and thence the camera's location relative to this system may be calculated.

The use of the compass (a freely rotating metal needle that aligns itself with the Earth's magnetic field) for navigation has been documented for many hundreds of years since its first recorded use in ancient China. Inertial sensors were first used for navigation by C S Draper in 1949 (Mostafa, 2001). These instruments contain one or more gyroscopes, which operate on the same principle as a compass, and accelerometers. They were primarily used for military navigation applications in the following decades, where the self-contained nature of the instrument, unaffected by poor weather or distance from navigation beacons, was a significant advantage.

Also initially used in military navigation were the satellite-based systems TRANSIT and NAVSTAR (more commonly known as GPS). TRANSIT was developed in the 1960s and allowed calculation of position only intermittently (whilst a satellite was passing overhead). The GPS system became fully operational in the early 1980s and includes enough satellites to enable calculation of a three-dimensional position at any time from any point on or near the Earth (Uren and Price, 1994).

Over the last two decades GPS has become very popular as an aid to navigation. The practical accuracy available to non-military users (10 m) was discovered to be much better than the theoretical designed accuracy (150 m). *Selective availability* was introduced to encode errors in the signal and reduce the available accuracy to 100 m for non-military users (Uren and Price, 1994).

However differential GPS (DGPS), where use is made of an additional static receiver as well as the roving one, can increase the accuracy to around 1–2 m. This is adequate for most navigation applications and DGPS has been used successfully in many autonomous vehicle navigation systems, e.g. Rogers et al. (1996). In 2000 selective availability was removed, greatly improving the accuracy that could be achieved with a single GPS receiver. However most accurate navigation systems already used DGPS which could compensate for the errors due to selective availability, and were thus mostly unaffected by this change.

Even the 1–2 m accuracy of DGPS can be improved on, however, by basing measurements on the phase of the carrier waves used for the GPS signals. This allows sub-centimetre levels of accuracy. As the necessary equipment and processing software became more affordable in the 1990s, carrier phase positioning became popular in surveying and eventually also in photogrammetry. Nowadays many aircraft used for air survey contain a carrier phase GPS system together with an inertial measurement system and these are used for *direct geo-referencing* of the images captured without recourse to ground control point analysis.

The distinction between position identification methods used for navigation and for surveying/photogrammetry has also become more blurred in that scene landmarks are now often used for navigation, particularly in the field of autonomous robotics: although in this field the landmarks may often be identified by laser scanning rather than vision, e.g. Madhavan and Durrant-Whyte (2004). Landmark-based methods have been developed as stand-alone algorithms (Lu et al., 2000) and have been used for aircraft landing assistance (Chatterji et al., 1998) and position determination for automated assessment of lighting installations (McMenemy et al., 2001; Zatari et al., 2003). The applications use an existing database containing the location of the ground-based landmarks to calculate the camera position; only an initial approximation to the current position such as can be found from a low resolution GPS receiver is required for an accurate result (providing the locations of the ground control points are known accurately). These represent only a small sample of the applications in use today.

Much interesting work is being carried out in the robotics-related field of research known as SLAM (simultaneous localisation and mapping) with regard to combination of laser scanning, vision-based etc. techniques for position determination of mobile robots, e.g. Williams et al. (2002); Morgenthaler et al. (1990). Many of the methods in this field have only been tested in indoor environments however and it is notable how many of the developments are analogous to GIS algorithms and data storage models.

Notwithstanding the recent trend towards using high-end GPS and inertial systems for direct geo-referencing of photogrammetric images captured by air survey, further developments in the last few years (Habib and Kelley, 2001) involve use of ground control points again, but with automatic correlation with image features and hence automatic calculation of the camera position. Only an initial approximation to the camera location is required—this technique will be discussed in more detail in Section 5.1.3. This work has recently been extended by applying the techniques to automatic geo-referencing of multiple satellite images (Habib and Alruzouq, 2004), but it should also be possible to develop the technique for use in geo-referencing of ground-level images. This is a promising research area.

It is worth noting that even when the object of the task is change detection (and hence the objects visible in the image might not correspond with the last known state of the landscape as stored in a GIS database), many objects will not have moved since the last time the scene was surveyed and will have their correct location recorded in a GIS database. Thus in most scenes, even greatly changed ones, there will still likely be many suitable landmarks for these types of positioning techniques.

1.5.5 Illumination and Colour

The light or radiance reflected by an object is a function of both the surface reflectance of the object and the incident illumination. In order to accurately measure the spectral reflectance of an object (from which the colour can be calculated) it is necessary to also measure or characterise the illumination incident on the object. For outdoor images the incident illumination will almost all come from the sun.

In many applications however, e.g. aerial photography for map surveying (Perkins and Parry, 1996), this level of accuracy is not required and care is simply taken to capture images at the same time of day so that the sun will be in the same relative position. Even in geographical change detection and repeat photography (Butler, 1994) the methodology often employed requires the images to be compared only to be captured at the same time of year and in similar weather conditions.

If there is a lot of cloud cover the illumination from the sun may appear diffuse, while with no cloud cover the illumination coming from one particular direction (that of the sun) will be of a much higher intensity than from the other directions. The light observed reflected from an object will depend also on the position of the observer (camera), i.e. there are two angles (illumination and observation) to take into account.

The Bi-directional Reflectance Distribution Function (BRDF) can be calculated at the time of image capture if the sun and observer positions are known. It is a theoretical concept that describes directional reflectance phenomena (Sandmeier, 2000). A related instrument is the goniometer (Walter-Shea et al., 1993), which can measure light coming from multiple directions simultaneously.

Accurate methods of measuring and compensating for illumination have been developed in the biological sciences (such as the examples given above) and more recently in remote sensing. For example, the multi-angle imaging spectroradiometer (MISR), one of the instruments aboard the NASA Earth Observing System Terra satellite, acquires images of the same ground area at approximately the same time but from different observation angles (Diner et al., 2002).

When the sky is clear, shadows cast by the sun can provide much additional information about the scene if interpreted correctly. In the field of remote sensing shadows are often interpreted manually, using knowledge of the sun position to estimate the height of features (Campbell, 2002). Automated techniques based on the same principle have been used in the field of machine vision (Stauder et al., 1999).

Under certain conditions, the colour of the outdoor illumination may vary considerably. For example, at dusk and dawn light coming directly from the sun often has a red hue while that reflected from the sky has a blue hue (Hubel, 2000). A generalised approach to measuring the surface reflectance of objects under any illumination is needed. This approach is known as the 'Colour Constancy Problem' and has been an important subject of research for many years (Sangwine and Horne, 1998).

The aim (in attempting to solve the problem) is to split the received light from any point in an image into two components: the actual surface reflectance of the object at that point, and the illumination incident on the object. If the actual reflectance of the object can be recovered then it can always be identified as this will be constant under all lighting conditions.

In his research in the 1930s, Krinov (1953) used standard white tiles with known constant surface reflectances. These were then placed somewhere in the scene with the object whose spectral reflectance was being measured. As the surface reflectance of the tile was already known, the spectrum of the incident illumination could be easily found by dividing the received signal by the reflectance of the tile. The reasonable assumption was then made that the illumination of the adjacent natural object would be the same and thus its surface reflectance could be found by dividing by the illumination signal.

White tile standards are still used today but are inconvenient to use in field conditions; in any case it would only be feasible to use a standard for calibration purposes and certainly not for anything approaching real time operation with a moving camera.

In the 1970s an attempt was made at solving the colour constancy problem by Land and McCann (1971), using the 'Retinex' theory. Another important approach was made by Gershon et al. (1987). Both these worked on the assumption that data from only one image of the scene was available and they wanted to recover the true spectral reflectance of the objects in the image. They constrained the optimisation problem by assuming that the colours in the scene could be related to absolute values in the following ways: the brightest part of the scene was white (Land and McCann, 1971) or the average of all the colours in the scene was a constant known value (Gershon et al., 1987).

Constant illumination is of course quite an obvious assumption that must be made, but both these approaches required a wide area to be covered by this assumption to get the range of colours and objects they needed for their other assumptions to hold, and neither algorithm was particularly effective.

The difference in the method put forward by Tsukada and Ohta (1990) is that data from multiple images of the same objects with different illumination are used to solve the problem. This is an obvious approach to take in computer vision where the colour values of objects in previous scenes can be recalled exactly, unlike human vision. The multiple images do not have to be of exactly the same scene as long as there are at least two images with different but uniform illumination, with at least two objects common to both the scenes.

Many different approaches to illumination compensation have been proposed in the literature over the years, but it appears to be still a developing field, certainly when applied to large-scale outdoor images.

Spectral Reflectance and Colour Colour is a perceived phenomenon related to the physical properties of the light reflected from objects. This can be described by the intensity of electromagnetic radiation in specified wavelength intervals in the visible spectrum (i.e. the range from 380 nm to 780 nm). A plot of light intensity against wavelength contains more information about an object than its colour does and so is often used for accurate characterisation.

Creating such a plot involves recording the brightness of an object (i.e. intensity of radiation) repeatedly when viewed through multiple bandpass filters allowing only a small range of visible wavelengths to pass. The response over the entire visible spectrum can thus be built up with a suitable choice of filters. This method (involving twelve filters) was used by Krinov (1953). Of course radiation with wavelengths outside the visible spectrum (e.g. infra-red or ultra-violet) is usually also emitted along with visible light; this does not affect the colour of an object and can only be measured using special-purpose filters.

As indicated above, details of the light intensity reflected by an object at every visible wavelength contain more information than the colour of the object; the colour is *overspecifed* by the complete set of spectral reflectance characteristics. Research on the human visual system has shown (Wyszecki and Stiles, 1982) that a person with normal colour vision has three types of colour receptors (*cones*) in the eye, which are sensitive to light from three different parts of the visible spectrum. The sensitivities of the three types of cones correspond approximately to red light (long wavelengths), green light (medium wavelengths) and blue light (short wavelengths). A detailed explanation of the operation of the eye can be found in Wyszecki and Stiles (1982).

Most colour systems are based on the use of three components, as in the human eye, meaning that any colour can be represented as a position in three-dimensional space. The XYZ system was established by the Commission Internationale de l'Eclairage (CIE) in 1931 (Commission Internationale de l'Eclairage, 1931) and is based closely on the average response of the human eye. The 1931 *colour matching functions* (University of California, 2004) relate the three colour values to the spectral response.

Colour for Electronic Imaging Digital photography has been based on the standards used in television: the RGB system(s) developed in the 1950s based on the response of the light-sensitive phosphors used in vidicon cameras and was standardised in the 1970s (Sproson, 1978).

RGB and XYZ are absolute luminance values; it is the ratio of the three components to each other that determines the colour of the object, but if the light is of a different intensity or brightness the absolute values will be different. To overcome this, normalised *chromaticity co-ordinates* can be used:

$$r = \frac{R}{R+G+B}, \qquad \text{etc.}$$

Normalised chromaticity co-ordinates have been used by several researchers in the 1990s to characterise and identify natural objects, e.g. Woebbecke et al. (1995) identified different species of weeds this way.

At the point of interpretation by the brain, the colours seen by the eye have already been processed into the hue-saturation-intensity (HSI) system in a manner described by Wyszecki and Stiles (1982). This avoids the varying intensity problem mentioned above as the intensity information is already separated from chrominance (i.e. hue and saturation, where hue is the dominant pure colour and saturation is the brightness, ranging from black / white / grey (no colour) through pale to a bright hue). This system is summarised in the diagram in Figure 1.1. Lambert and Carron (1999) discuss further the advantages offered by the separation of components inherent in the HSI system.

Colour for Natural Object Recognition A lot of research effort has been concentrated in the area of agriculture, particularly large-scale intensive farming. The economics of scale meant this was an early area to invest in this technology.

Ali and Aggarwal (1977) used a flying spot film scanner to digitise aerial photographs of citrus orchards. False colour film sensitive to green, red and infra-red light was used; this meant that extra details that would not be captured by normal colour

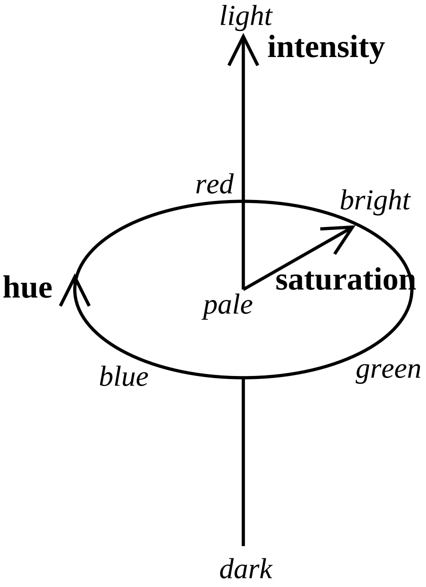

Figure 1.1: HSI Colour Space

film could be used with the conventional scanning equipment. A computer picked out a regularly spaced pattern of dots to scan and calculate xy normalised chromaticity co-ordinates and intensity. The trees in the orchard were all planted in a regular arrangement so it was easy to relate the position of the trees to the colour of the scanned dots. The data for each dot was used successfully to pick out diseased trees. However the system required a lot of user intervention for calibration and to set thresholds for picking out features etc.; these would change depending on the ambient sunlight.

Slaughter and Harrell (1989) devised a real-time system for detecting the centres of oranges to enable them to be automatically picked from trees. The image captured by the camera in the end-effector of the robot picker was digitised in RGB format but with only 32 levels instead of the usual 256. This meant there were only 32768 possible colours. A Bayesian classifier was trained using typical scenes from an orange grove, to classify specific RGB values as either part of an orange or background. The classifier was then implemented as a look-up table (only 4 kB in size owing to small number of possible colours) and could be used to perform binary segmentation of an image in real time.

Another algorithm then used the binary image to analyse the shapes of the orange regions and find centres of oranges. This could all be computed faster than the video frame rate and this real-time operation was very important as the oranges could be swaying in the wind and neeeded to be picked quickly before they moved.

The second example described above was designed to be inherently robust, i.e. it was trained to identify oranges under many levels of illumination intensity. The first example used normalised chromaticity co-ordinates, which removed the effect of changes in illumination intensity, however it also used intensity information directly. Without calibration by a human operator this might have been a problem. It is speculated also that the robotic fruit harvesting system might not have worked correctly if there was a change in the hue of ambient light or sky colour as might occur at sunrise or sunset. For example, it may be confused between the RGB value of the sky at sunset and that of an orange under direct bright sunlight. Section 6.3.2 will investigate methods of compensating for this type of problem.

1.5.6 Change detection

Automated processing of aerial images is a well-developed research area. Many books have been written on the subject, e.g. Campbell (2002) and established techniques are implemented in GIS (Neteler and Mitasova, 2002). Ortho-rectification involves warping an image so that it corresponds with the projected co-ordinate system in use on the ground; this is easy to accomplish in GIS when the camera parameters are known and a ground elevation model is available. It is then possible, by classifying the pixels in the image into groups, to determine the land use or type of building in particular areas.

An initial image classification stage is the starting point for many types of change detection, e.g. change vector analysis (CVA) (Bruzzone and Cossu, 2003). Most change detection techniques in current use involve analysing a raster image to identify specific objects or classes of pixels, encoding the object represented by these as a vector, and then comparing the vector data to other multi-temporal data sources (Neteler and Mitasova, 2002). Agouris et al. (2000) provide another example of a method of change detection using a GIS as a reference for existing features.

3-D Object Extraction If an overlapping stereo pair of images is available then features can be identified simultaneously in both images and measurement of height of the features made. This has been done manually for many years (Thompson, 1966) and can be easily automated with image processing. In the last decade, significant research has taken place on the automatic or semi-automatic extraction of 3-D object information from aerial photographs. In particular, guided building extraction from monocular images (Gülch and Müller, 2001) or making use of image stereo pairs (Tseng and Wang, 2003) is a popular research area. These methods make use of the fact that aerial photographs are rarely exactly vertical and part of the sides of buildings are usually visible.

1.6 Thesis and Scope

The main object of this thesis is to present techniques and algorithms that may be used in the creation of a system for geo-referencing objects in a general environmental scene, with specific application to change detection over time. This covers the following areas

- Image data acquisition and processing
- Geographical data acquisition and processing
- Position determination
- Illumination and colour measurement

1.6.1 Thesis focal theory

It is argued that it is possible to combine techniques and methodologies from these areas, in a manner not previously done, to produce a highly automated and relatively inexpensive system that performs as described.

In support of this *thesis*, the following are presented:

1. A review of the literature in the relevant areas and identification of promising methods and techniques for use in the system
2. A critical analysis of the signal flow and data requirements for the most relevant of these methods, resulting in a detailed and justified design for the proposed system
3. Experimental results indicating the successful operation or feasability of many of the components of the system

1.6.2 Scope

The work involves combination of ideas and techniques from the various relevant research areas into a low-cost, mobile, ground-based, temporal, environmental analysis system. The goal is to provide an economic and versatile method of assessing moderately-sized areas of land or materials.

One important aspect of the work is to maximise the use of image processing to reduce the requirement for calibration, i.e. to obtain information from images rather than external sensors. Fusion with data from other sensors is still important for rough localisation, e.g. GPS (global positioning system) and compass data for position and orientation information, but the information contained in the image of the scene should provide the fine level of detail.

Mobility is also an important requirement of the system being developed. Depending on the configuration, it may possibly be used as a handheld system, may be based in a car or perhaps a miniature helicopter or kite (Aber et al., 2002). The platform used will depend on the application and these usage configurations are all considered in the research.

As the system is a new concept (i.e. it makes change detection possible using low-cost additions to existing mobile vehicles, thus avoiding the requirement for special fieldwork expeditions), thorough attention to detail in the initial steps is very important. This will increase the efficiency with which others can continue to further develop the system and provide a solid foundation on which to base further work. For this reason, much attention will be given to the subject of data acquisition and correspondence of disparate data sources with each other. This applies to image data, existing geographical data and mobile positioning data and forms the bulk of the work presented in this thesis.

Substantial work on change detection and illumination and colour compensation is also presented. Further development of the change detection system is discussed and a complete system signal/data flow diagram is presented in Chapter 7.

1.7 Layout

The next chapter discusses camera selection, image acquisition and pre-processing. Distortion due to pixel aspect and lens effects is investigated and methods are developed to compensate. The characteristics of image pixel values and use of old photographs are also discussed. Chapter 3 discusses geographical data: its sources, accuracy and limitations. A suitable geographical information system is selected and used to process the data into forms suitable for fusion with image and position data. Chapter 4 treats one area of GIS functionality in detail: 3-D visualisation techniques are adapted for use with camera image data. Methods of combined visualisation, image perspective transformation and plan-view generation are all presented.

Chapter 5 proposes an inexpensive method of calculating camera position accurately by augmenting conventional GPS positioning with image-based methods. Some experimental results of this are presented. In Chapter 6 a proposed system for integrating ground-level images with existing GIS data and performing change detection is presented. Several different techniques regarding illumination and colour correction, and image segmentation for object identification are experimentally evaluated.

Finally Chapter 7 summarises the achievements and limitations of the work presented so far. A theoretical plan and signal flow diagram for the complete system are presented, and suggestions are made for further work.