Landmark Integration Using GIS and Image Processing for Environmental Analysis with Outdoor Mobile Robots


Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date: 06. Oct. 2020
Abstract—This paper describes the implementation of a system for integrating ground-level images with data from a Geographical Information System (GIS). The images are captured using a video camera on a moving platform. Amalgamation of data takes the form of both transforming GIS data into the image space, and transforming the image data into the 3-D world space for comparison with existing GIS data. In current implementations, e.g. inverse perspective transformation, a flat scene is initially assumed and image-based knowledge is used to compensate for errors. The method used here improves on this assumption by using a readily-available source of elevation data (the GIS) for the image scene. It is shown that typical functions available in a GIS are very suitable for this ground-based scene analysis; this rich source of data significantly reduces the image processing required in order to carry out a task.

Calibration of a camera and combination of the images with GIS and other sensor data is demonstrated. The effectiveness of the use of the GIS to ease planning and locate orientation features for an autonomous robot task is shown. The results of application of the technique to outdoor scenes are given and discussed.

I. INTRODUCTION

Use of aerial photographs and satellite images for change detection in a local environment has been well researched [1]. New techniques, including advanced perspective transformation and automated data acquisition are required to allow detailed ground level images to be used in task planning for robots. Change detection in a robot’s environment is an important factor in this observation application and in improving the robustness of the object-based RPY and position estimation.

The primary advantages of using ground level images are the high resolution and multiple views available, ease of acquisition and the potential low cost of the hardware. Consumer-grade cameras will be used instead of specialised photogrammetric cameras; thus experimental calibration and modelling of these cameras is necessary.

A. Background and Motivation for GIS Use

To analyse change in an area, based on changes in acquired images, it is first necessary to determine the correspondence between the pixels in the images and their 3-D world co-ordinates. Substantial work has been done in this area, but typically it is assumed that the ground is flat, i.e. all objects exist in the same plane [2]. Significant errors in the results can be caused by the ‘ground plane’ assumption not holding. An analysis of the deviations can be used to calculate more accurate information about the slope of the ground or heights of the objects [3].

The assumption of flat ground may be made because of the difficulty in obtaining and processing geographical data (i.e., ground elevation, building and perimeter outlines etc.) within existing image processing systems. These are the features that are required to locate the automated system and to carry out its change detection task.

For most practical applications it is appropriate to fuse all of the available sources of data [4]. Height data is readily available within GIS in the form of a digital elevation model (DEM). A GIS is the natural environment for this geographical data and it can process a wide variety of data formats. However, it is of vital importance that the GIS data available has been accurately surveyed and is of high quality (e.g. adequate resolution).

The spatial data processing capabilities of GIS are well-suited to the combination of observed images with GIS data. Much of the spatial analysis and processing may be done within the GIS, saving the time of re-implementing GIS-type functions in other programming environments. Thus, instead of extracting the elevation data from the GIS and using it externally, the GIS 3-D visualisation functions are extended for use in creating “camera images”. Feature points which need to be located in the real camera can easily be localised in these “camera images”; the robotic image analysis time is therefore reduced and accuracy of analysis is increased.

The current work builds on previous research related to three-dimensional image interpretation using low-cost digital video cameras [5]. This paper describes how the integration of these image processing and photogrammetric techniques with GIS increases the usefulness of the image data and enables fusion with many other spatial data sources.

Methods for visualisation of GIS vector data in the context of ground-level images and analysis of images in the real world co-ordinate system used by the GIS are also given. The use with ground-level images is novel and opens up the possibility of integration with many outdoor robot and planning systems.
B. Overview

The next three sections describe the main considerations in the combined image processing and GIS system:

1. Experimental determination of the interior orientation of the camera and implementation of the “camera image” in GIS using Computer Graphics techniques
2. Estimation of camera exterior orientation from GPS and inertial sensor measurements and accurate pose estimation making use of GIS vector data
3. Integration of ground-level images with existing GIS data, involving
   - Overlay of vector and sites data on images
   - Re-creation of perspective views from a series of images and orientations

Results (showing application of the techniques) are given with the descriptions of the methods. Following that a discussion is presented and conclusions are drawn.

II. LOW-COST CAMERA CALIBRATION

In this section the calibration techniques necessary for the use of digital video cameras in photogrammetry are discussed. Then the implementation of a model of the camera using computer graphics is described. The resulting 2-D view of an area of GIS data that coincides exactly with the camera’s view of the same area in the real world is then derived.

A. Correction for Lens Distortion

The lenses used in consumer-grade cameras are not designed to produce geometrically-correct images (i.e. do not agree with the pinhole camera model, Fig. 1(a)). Before measurements can be made on an image it must be corrected to account for the lens distortion.

In previous research on camera calibration [5] measurements were made of the distortion on a grid of known points. Based on a model of the barrel distortion in the lens, a polynomial function was derived to model the transformation parameters for correction of the distortion.

A reverse mapping approach is used, i.e. for each pixel in the corrected image, the corresponding point in the distorted image is sampled and bilinear interpolation used to calculate the appropriate pixel value.

B. Interior Orientation Parameters

Detailed interior orientation parameters are not available for a consumer-grade camera; in this work it is being used beyond its designed purpose. The manufacturer specifies the minimum focal length as 2.9 mm and the sensor size as 1/4 inch, which corresponds to the industry standard charge-coupled device array (CCD) dimensions of 3.2 mm x 2.4 mm.

In an attempt to validate these measurements, the CCD dimensions were taken as correct and focal length calculated using a weak perspective model. This involves capturing images of targets of known size, positioned at known distances from the camera. It was found that either the calculated focal length or assumed CCD size differ from that provided by the manufacturer.

The next challenge in creating a computer-generated view that will be exactly aligned with the camera view is that the PAL Digital Video (DV) standard uses rectangular pixels whereas computer graphics uses square pixels. One of the images must be corrected to enable alignment.

For the camera the horizontal to vertical pixel aspect ratio is 59:54. In addition each horizontal video line in PAL DV is slightly longer than in the analogue PAL standard, so it must be cropped on each side to maintain the correct image aspect ratio of 4:3 [6]. In the vertical direction analogue PAL contains 576 discrete visible lines, which translates simply to 576 pixels in DV PAL.

The correction can be accomplished by back-projecting the 720 x 576 (rectangular pixels) image into a 768 x 576 (square pixels) grid; the pixels are selected as follows

$$\text{row}_{\text{rect}} = \text{row}_{\text{sq}}$$  (1)

$$\text{col}_{\text{rect}} = \frac{32}{59} + \left(\text{col}_{\text{sq}} \times \frac{54}{59}\right)$$  (2)

and linear interpolation is done between the two rectangular pixels either side of the calculated $\text{col}_{\text{rect}}$ value to give the correct pixel value for the square pixel.

C. Computer Graphics / GIS Implementation

In a camera that conforms to the pinhole model (i.e. lens distortion has been corrected) the centre of projection (COP) is located at the pinhole and is in front of the CCD array (or film), Fig. 1(a). The focal length is the distance the pinhole / COP is in front of the CCD.

When modelling a camera using computer graphics, the COP is located behind the near clipping plane (i.e. the image plane, Fig. 1(b)). The distance the near clipping plane is in front of the COP is the equivalent of focal length, i.e. the projection of an object at the same distance in front of the respective COP will appear the same size in both systems.

GIS 3-D view software, such as that in the GRASS GIS [7], has traditionally been designed with the aim of producing a ‘pretty picture’ visualisation of GIS data. This is almost always for human interpretation and does not take account of the exciting possibilities of image processing. Photogrammetric-grade correctness of the resultant image has not therefore been an absolute requirement. However, it is possible to modify the software to model the characteristics of a particular camera. This has been done with GRASS for the camera used. For field implementation this computing resource, in most cases, would be accessed remotely across a network. But in large autonomous vehicle systems this “camera image” creation system could be on-board.
Vertical field of view (FOV) is the only interior parameter directly specified by the user. The image aspect ratio is calculated from the size of the output view window on the computer screen, however this can also be sized by the user. In this case it is set to be 768 pixels wide by 576 high, i.e. corresponding with a square-pixel PAL image frame.

Distance to the near clipping plane is normally set automatically by the software to a value of 1 cm. For this application it was changed to 3.45 mm (focal length for our camera) and the software was recompiled. This highlights an important advantage of using Open Source Free Software (which GRASS is) over closed proprietary software where similar functionality may require a prioritised change request.

The fixed setting for distance to the near clipping plane, together with the user-specified vertical FOV are used to calculate the width $w$ and height $h$ of the image plane using the following equations

$$ h = 2 \tan \left( \frac{\text{FOV}}{2} \right) \times \text{dist}_{\text{near}} \tag{3} $$

$$ w = h \times \text{aspect} \tag{4} $$

where $\text{dist}_{\text{near}}$ is the distance to the near clipping plane. Re-arranging (3) gives

$$ \text{FOV} = 2 \tan^{-1} \frac{h}{2 \text{dist}_{\text{near}}} $$

For $h = 2.4\text{mm}$ and $\text{dist}_{\text{near}} = f.d. = 3.45\text{mm}$ this gives $\text{FOV} = 38.4^\circ$, which can be specified through the normal user interface of the software.

Fig. 2. Experimental set-up showing DV Camera, inertial sensor and GPS antenna on an equipment rack.
A dual-frequency Leica GPS receiver is operated in kinematic mode at its maximum sampling frequency of 10Hz. Post-processing results in centimetre-level accuracy. The primary aim of using the GPS receiver was to validate the output from the 3-D position sensor and the pose estimation algorithm. It was the most expensive component of the system and it is anticipated that in a production system its use would not be required.

The camera and GPS antenna have fixed co-ordinates in the M-space. Methods exist to determine the rotation of the M-space relative to the W-space using data from more than one GPS receiver [8]. With only one mobile receiver the rotation can be calculated by considering a vector representing the instantaneous direction of movement of the vehicle.

Owing to the high GPS sampling frequency and relatively smooth and slow (30–50 km/h) vehicle movement, we can approximate this vector as that between the previous A(x0i + y0j + z0k) and next B(x2i + y2j + z2k) GPS positions. i, j, and k are unit vectors in the W-space in the directions of the x, y and z-axes, respectively. The instantaneous direction vector is therefore given by

\[ \overrightarrow{AB} = (x_2 - x_0)i + (y_2 - y_0)j + (z_2 - z_0)k \]  

(5)

The angles this vector (along the y'-axis) makes with the axes can be calculated using the \( \text{atan2}(\cdot) \) function, and accounting for axis offsets the rotation of the mobile co-ordinate frame relative to the world is described as

\[
\begin{align*}
\theta_x &= \text{atan2}(y_2 - y_0, z_2 - z_0) - \pi \\
\theta_y &= \text{atan2}(z_2 - z_0, x_2 - x_0) + (\pi/2) \\
\theta_z &= \text{atan2}(x_2 - x_0, y_2 - y_0) - (\pi/2)
\end{align*}
\]

The co-ordinates of the camera in the W-space \((xyz)^T\) can now be calculated as

\[
\begin{pmatrix}
x \\
y \\
z
\end{pmatrix} = \begin{pmatrix}
o_x & 0 \\
o_y & 0 \\
o_z & 0
\end{pmatrix} + 
\begin{pmatrix}
C_2C_3 & S_1S_2C_3 - C_1S_3 & S_1C_2S_3 + S_1S_3 \\
C_2S_3 & S_1S_2S_3 + C_1C_3 & C_1S_2S_3 - S_1C_3 \\
-S_2 & S_1C_2 & C_1C_2
\end{pmatrix} \begin{pmatrix}
x' \\
y' \\
z'
\end{pmatrix}
\]

where \(S_1 = \sin \theta_x\), \(S_2 = \sin \theta_z\), \(C_1 = \cos \theta_x\), etc.; \((x' y' z')^T\) are the fixed co-ordinates of the camera in the M-space and \((o_x o_y o_z)^T\) are the current co-ordinates of the origin of the M-space (i.e. GPS Antenna) in W-space co-ordinates [9].

The yaw and pitch of the 3-D position RPY sensor may be calculated from the GPS data as

\[
\begin{align*}
yaw &= \theta_z \\
pitch &= \text{atan2}(\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}, z_2 - z_0)
\end{align*}
\]

(6) (7)

The pitch angle from the Tokin sensor is always small for travel on roads without steep hills and it is never more than 2–3° out from the GPS-derived value and usually closer. In general the sensor is qualitatively accurate for yaw and pitch and appears to be operating within its specification. It would therefore be adequate for providing an initial attitude estimate for use with pose estimation algorithms.

At present the RPY sensor is used for the roll measurement as this cannot be derived from GPS measurements with only one receiver. In the future, image features will be matched with expected GIS artefacts to improve the RPY estimation accuracy and compensate for drift.

B. Elevation Data Sources and Quality

The standard Ordnance Survey of Northern Ireland (OSNI) vector data used is only two-dimensional, and an accurate source of elevation data must be fused with this in the GIS before it can be of use for 3-D position estimation in a non-flat landscape.

30m resolution ASTER DEM data [10] was interpolated with GPS measurements gathered on the mobile system to give a raster digital elevation layer. The regularised spline with tension (RST) algorithm [7] was used for the interpolation, with smoothing factors of 7m for ASTER and typically 1–2 cm for the GPS data, depending on the accuracy output from the kinematic processing.

C. Accurate Pose Estimation using GIS Vector Data

Single-photo resection techniques [11] that avoid complete reliance on GPS and RPY data are a useful alternative method of pose estimation [12]. This is being investigated using easily identifiable features in the scene whose positions are described within the GIS data.

The technique is described in detail in [5] and has been successfully used with features visible from an aircraft. In brief this involves definition of a set of pairs of W-space co-ordinates and image pixel co-ordinates followed by least squares minimisation to find the camera orientation.

In comparison with GPS data the technique matches to within 10m horizontally and 2–3m vertically. For the detailed scenes described here, the GPS and map resolution is not accurate enough when the objects are close; objects sufficiently far away to minimise the RPY and location errors must be chosen. This method has good potential for automated processing of a video sequence as features only need to be identified once and can be tracked between images [13].

IV. INTEGRATION OF RESULTS WITH EXISTING GIS DATA

In this section two possible methods of integrating ground-level images and GIS data are discussed. The basis of these is to project the GIS data into a 2-D view that
aligns exactly (pixel-for-pixel) with the camera image. This is done by modelling the camera orientation in the GIS 3-D visualisation software (see Section II-C).

A. Overlay of Vector and Sites data on Images

A uniform colour is used for the terrain surface and background. This can be made transparent and the vectors extracted easily and overlaid on the image for visualisation. An example of this is shown in Fig. 4. A vector layer representing the centreline of boundaries and fences (lines) is overlaid on an image, together with a sites layer representing the ground height of the road (spheres), derived from GPS measurements. Different GPS sampling intervals of 0.2s and 0.5s were used when travelling in alternate directions along the road.

There is an accurate matching of visual data and GPS points on the road surface. The seemingly large visual error in the LHS road boundary is effectively only 1 m and appears to be caused by surveying and resolution errors (see Section III-B) in the original OSNI vector data.

B. Database of Multiple Images from a Video Sequence

Each pixel in the projected view represents a \( W \)-space point whose easting, northing and elevation may be obtained from the GIS and associated with a surface colour value from the camera image. The automated calculation of these values for every pixel is not a standard feature of the GIS and source code had to be adapted and scripts created.

To survey large areas of ground at a reasonable scale many images must be used and stored in a database with metadata describing viewing conditions. Different aspects of ground-level features will be visible when viewed from different angles and a strength of this method is that many features not visible from the air can also be seen.

The information is somewhat similar to that stored in a mobile mapping system [14] although here it is used to construct a perspective transformed view of the scene by interpolating between the transformed points. Each image from the video sequence will contribute an extremely uneven distribution of surface colour and brightness values to the database. For a ground-level image there will be a large number of points close to the camera with very small distances separating them, while further away there may be large distances separating the points in the \( W \)-space that the pixels have been projected to.

Rules for pixel inclusion are derived, based on
- the closeness of the viewing angle from which the pixels were originally captured to the virtual camera angle they are now being viewed from
- the closeness of the original distance to the camera to the current viewing distance, etc.

As an illustration of the method, all the pixel values from three images taken at short intervals along the same stretch of road have been used in an interpolation (with 1m resolution) to create an aerial view. The plan-view image (Fig. 5) is formed from an interpolation between the points using inverse-distance squared weighting. This is much more than simple texture mapping as this image data is directly connected to its 3-D position. Interpolation must only be done for the area defined by the field of view and combined visual horizon of the three camera positions. A GIS line-of-sight analysis indicates a small region in the centre of Fig. 5—white area—that was not visible from any of the 3 images.

The image shows GIS vector boundaries data overlaid on the interpolated image and for areas with an unobstructed view from the camera the alignment is very good (e.g. the road where the grey line passes through the centre of the image coincides accurately with the two parallel curves that form the road boundary.) There are some small discrepancies of a similar magnitude to those in Fig. 4 and likely caused by the same sources. The surface colours of the crops in the different fields surrounding the road can be clearly identified.
The white patches in other areas have been caused by specular effects as a result of the sun being low in the sky; these have not yet been modelled. The main errors are caused by above-ground objects that are not modelled by the DEM being projected into the wrong location, e.g. some of the hedges at the side of the road.

Further implementation of the pixel-selection rules and improved image processing will reduce the detrimental effect of above-ground objects on image quality. An important aid to this will be better GIS vector data, e.g. much of the data used does not distinguish between hedges, walls and fences. Such information will be used in image processing to help identify features. Image flow methods can be used to locate discrepancies of the flow vectors in areas (e.g. the GIS indicates that the expected data point is far away, but the area flow vector deviates from this) and either remove them in mapping of GIS features or else model them as a new unmapped object.

V. DISCUSSION AND CONCLUSIONS

GIS applications originally designed purely for visualisation of data can be modified to take account of the characteristics of a particular camera. This opens many possibilities for using image processing in conjunction with a GIS and makes it an ideal processing tool for combining images with the underlying data.

GIS provides good potential for use in autonomous vehicle planning and control. It is most of use where the task involves geographical analysis, e.g. change detection. GIS is most useful for landmark extraction, speeding up the necessary image processing and improving forward planning of the motion task. The technique will also be used for rapid assessment of the amount of change in an area and indicate if it needs to be re-mapped in more detail.

Use of GIS elevation data for a non-flat area enables data from images of the area to be transformed into the 3-D world co-ordinate system used by the GIS and also GIS data to be transformed into the image space for visualisation. The transformation technique does not require stereo cameras or additional image-derived knowledge to cope with a non-flat area.

The elevation data is often readily available in GIS format but the quality must be carefully considered and improved if necessary. To a certain extent some of the limitations of GIS data can be overcome by using artificial intelligence rules and image processing.

Dual-frequency kinematic GPS surveying provides almost all the camera orientation information required to the necessary degree of accuracy. Pose estimation techniques that utilise the correspondence between GIS vector data and landmarks in an image are starting to approach the same level of accuracy but with a much lower cost than the GPS system. The landmark techniques allow the position and orientation of a calibrated camera to be determined even for an unstructured scene, as long as adequate mappable features are visible.

VI. ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of Northern Ireland Department for Employment and Learning and the Ordnance Survey of Northern Ireland, and the assistance of Ashraf Zatari and John Meneely with fieldwork.

VII. REFERENCES