Development of an Integrated IMU, Image Recognition and Orientation Sensor for Pedestrian Navigation

C. Hide, *University of Nottingham*
T. Botterill, *Geospatial Research Centre (NZ) Ltd*

BIOGRAPHY

Chris Hide is a Research Fellow at the University of Nottingham, UK. He has a degree in Mathematics and Topographic Science from the University of Wales, Swansea and a PhD in engineering surveying from the IESSG, University of Nottingham. He worked on secondment for 3 years as a Senior Research Scientist at the Geospatial Research Centre in Christchurch, New Zealand before returning to Nottingham in 2010.

Tom Botterill obtained his MA degree in Maths and Computer Science at Cambridge University in 2005 and is currently studying towards a PhD with the Geospatial Research Centre and Department of Computer Science at the University of Canterbury, Christchurch, New Zealand. His research focuses on navigation using computer vision.

ABSTRACT

This paper describes a scheme for pedestrian navigation integrating measurements from a foot-mounted IMU with position and orientation updates from computer vision techniques. By mounting an IMU on a user's foot, the position drift can be substantially reduced since zero velocity updates can be applied every step. However, such a system will still suffer from position drift unless occasional measurements are available from other sensors. This paper describes a novel method for restricting such position drift using an image recognition algorithm. Firstly, a database of images and their locations is constructed over an area of interest. A user then navigates the area using foot-mounted inertial sensors and a video camera. As images are acquired, they are used to search the database of images using the Image Bag-of-Words algorithm. When new images are successfully matched with images in the database, the position from the database is used to update the inertial position using a Kalman filter. Furthermore, when images are successfully matched, orientation updates can be applied by estimating the relative orientation of the two cameras. These measurements can help overcome the limitations of the IMU, in particular the problem with heading drift. The integrated inertial and vision system is demonstrated to provide better than 10m accuracy (typically 1-5m) over a period of 21 minutes, and the paper demonstrates how orientation updates could be applied in the future.

INTRODUCTION

Pedestrian navigation represents one of the most challenging areas of navigation, with requirements to estimate position for long periods of time in areas where GNSS signals are difficult to receive. It is widely recognised that the solution to pedestrian navigation will comprise of a number of different positioning technologies combined to generate a single, seamless position solution that can work in many different environments. Dedicated local positioning systems based on technologies such as Ultra-wideband (UWB) may provide extra signals that can be used for positioning indoors, however their use is largely restricted to small areas due to the cost of installing the infrastructure. Instead the authors have looked towards using information and infrastructure that already exists such Wi-Fi signals or, as described in this paper, images, for providing extra navigation information primarily indoors.

One of the most promising areas of development for pedestrian navigation in the last decade is the use of foot mounted inertial sensors. Low cost IMUs (Inertial Measurement Units) consisting of 3 orthogonally mounted gyros and accelerometers can be used to compute a self-contained continuous position and orientation solution without the need for external measurements. However position accuracy using these sensors is typically very poor since the position error increases over time due to inaccuracies in the gyro and accelerometer measurements. One method to substantially improve the positioning performance from low cost IMUs is to mount the IMU on the pedestrian's foot and use Zero Velocity Updates (ZUPTs) every time the pedestrian's foot is detected as not moving. ZUPTs can not only be used to reset the velocity to zero, but can also be used to estimate other errors such as sensor biases, and position and attitude drifts using a Kalman filter. See, for example, Foxlin (2005); Godha et al. (2006).

Low cost inertial sensors can be used to provide a continuous position, but over time the position will still drift and therefore it is necessary to integrate IMUs with other position and orientation sensors. In particular, the position error using foot mounted IMUs is largely caused by heading error. This is due to two reasons. Firstly, heading error is difficult to initialise without using an aiding sensor such as a compass. However compasses are susceptible to large errors caused by local magnetic disturbances. Secondly, heading error is not observable using zero velocity updates alone since the gyros are not sensitive enough to measure the Earth's rotation. Position error will also be caused by other errors such as gyro and accelerometer noise, bias and scale factor errors.

As a result of all these factors, it is necessary to use a combination of technologies including inertial sensors (which are able to navigate continuously regardless of the environment), GNSS (when it is available) and other sensors which are able to provide measurements such as position, velocity and attitude updates. One such low-cost sensor is a camera, and recently the Computer Vision community have developed two approaches to make use of a single camera as a navigation sensor. Firstly, cameras may be used as a dead-reckoning sensor, for either simple odometry (Nistér et al., 2006) or in a Simultaneous Localisation and Mapping (SLAM) framework (Davison, 2003; Eade and Drummond, 2006, 2007), however these schemes are subject to gross errors and failure, for example due to featureless or dynamic environments. Systems integrating visual measurements with other sensors such as GPS receivers and IMUs have been more successful (Kim and Sukkarieh, 2007).

Alternatively, images may be used to recognise when the camera is viewing a scene that has been seen previously. The Image Bag-of-Words algorithm is the method most commonly used for this purpose, for example by Sivic and Zisserman (2004); Nistér and Stewénius (2006); Cummins and Newman (2008). These schemes index images into a database as they are captured. This database can then be queried to search for images that appear to be of the same place, hence providing occasional position updates when a location with an earlier position estimate is observed. These schemes are robust to changes in scale, illumination, camera position, and small changes in the scene, and databases of hundreds of thousands of images can be queried in real-time with error rates of just a few percent, depending on the frequency of update required and the nature of the environment (Cummins and Newman, 2008).

For this paper we use both the Image Bag-of-Words algorithm to provide position updates and a scheme to calculate the relative orientation of camera positions. During an initial 'mapping' survey thousands of images are captured at known positions throughout the area to be navigated. These images are added to a Bag-of-Words database. A pedestrian then travels through the area with a camera and IMU. When a new image is observed, the database can be searched to generate an estimate of the most likely location it was observed from, and the reliability of this match. This match is validated by checking that the scene geometry is compatible, and the relative orientation of the two cameras is computed from this image geometry. These position and orientation measurements, together with an estimate of the uncertainty in the difference from the mapped camera location, is used to correct the IMU navigation solution using a Kalman filter. The strengths of this innovative use of computer vision to provide position and orientation updates for pedestrian navigation are that it is low cost and passive, does not requiring any dedicated infrastructure or markers, and is feasible for use over wide areas.

This paper describes the development of a pedestrian navigation system based on an IMU, GNSS and computer vision. The paper first describes the algorithm used for generating position from the IMU measurements. Secondly the Image Bag-of-Words algorithm is described along with details of how orientation information can also be extracted from the images.

FOOT MOUNTED INERTIAL SENSORS

The position and orientation of the pedestrian is computed using inertial sensors attached to the user's foot. A three axis IMU is used that consists of three gyros and accelerometers and the measurements. The measurements are integrated using navigation frame mechanisation equations as described in Titterton and Weston (1997) and are numerically integrated at the high rate of the IMU (typically 100Hz). In order to initialise the IMU, the IMU measurements are averaged for a short period of 2 seconds. During this period, the gyro output is averaged and is used to estimate the initial gyro bias. Secondly, the accelerometer outputs are averaged and are used to resolve the initial roll and pitch of the IMU by comparison of the outputs with the local gravity vector. The initial heading of the IMU is computed using a one-off heading measurement derived from a magnetometer (the IMU used for this paper also contains a 3-axis magnetometer).

KALMAN FILTER

A Kalman filter is used to estimate the navigation and IMU errors. The state vector is defined as:

$$x = \begin{pmatrix} \delta p & \delta v^n & \delta\omega & \delta g^b & \delta a^b \end{pmatrix}^T \tag{1}$$

where δp is vector of latitude, longitude and height errors; v^n is the vector of navigation frame velocity errors; $\delta\omega$ is the vector of misalignments about the navigation frame axes; δg^b is the vector of gyro bias errors; δa^b is the vector of accelerometer bias errors. The Kalman filter is used to estimate the errors using a linearised inertial navigation model such as that described in (Titterton and Weston, 1997; Farrell and Barth, 1999). The model used in this paper is known as the ϕ-angle model (Scherzinger and Reid, 1994). The model describes the interaction between different error states and can be used to estimate the full state vector using position or velocity measurements and sufficient dynamics. The filter is used in feedback form so that when a measurement is available from a sensor, the error is computed using the Kalman filter which is then used to correct the inertial sensor measurements and navigation parameters. This is to ensure the navigation errors remain small and hence keep the linearised model valid. More information on Kalman filters and Kalman filters for inertial navigation can be found in (Hide, 2003; Foxlin, 2005; Farrell and Barth, 1999).

ZUPT measurements are applied every time the user's foot is detected as being stationary (i.e. at least every step). A simple algorithm is used to detect that the user's foot is stationary using a moving average filter with a window of 7 measurements. The magnitude of the acceleration is compared to the magnitude of gravity (i.e. $\left||f^b| - |g^n|\right|$) and a threshold is used to determine whether the foot is stationary. This is shown in Figure 1. The filtered IMU measurements are then decimated to a rate of 20Hz. A further check is applied so that a ZUPT is only detected if two consecutive filtered measurements both fall within the threshold, this is to remove a point during a typical step where a false zero velocity update is sometimes detected. Although the total acceleration also includes the accelerometer bias and scale factor error, in our tests, this algorithm has proved to be robust. When a ZUPT is detected, the current velocity is differenced with zero and is used as the measurement to update the Kalman filter. More details of the IMU mechanisation and integration are given in Hide et al. (2009).

Lastly we consider the measurements that are available from using image updates. Image updates can be used to provide two types of measurement for the IMU using the Kalman filter. Firstly position measurements (which are defined here for image updates, but are also the same for GNSS position updates) are defined as:

$$z_k = \begin{bmatrix} \lambda \\ \phi \\ h \end{bmatrix}_{IMAGE} - \begin{bmatrix} \lambda \\ \phi \\ h \end{bmatrix}_{INS} \tag{2}$$

where λ, ϕ and h are the latitude, longitude and height respectively. The design matrix is defined as:

$$H = \begin{pmatrix} I_3 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3}$$

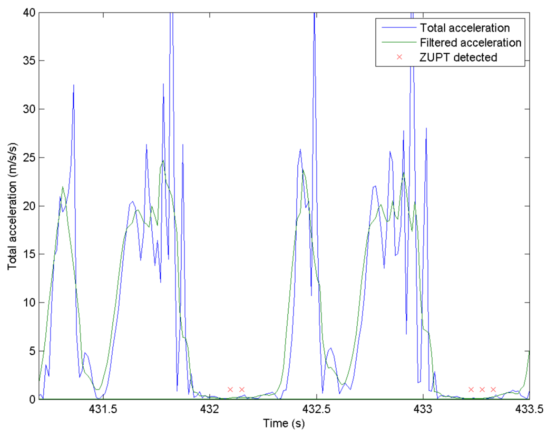

Figure 1. Auto detection of zero velocity from the accelerometer measurements

For attitude updates we consider the camera rotations to be defined as:

$$z_k = \begin{bmatrix} \varphi \\ \theta \\ \psi \end{bmatrix}_{IMAGE} - \begin{bmatrix} \varphi \\ \theta \\ \psi \end{bmatrix}_{INS} \tag{4}$$

where φ, θ and ψ are the roll, pitch and yaw respectively. The design matrix is defined as:

$$H = \begin{pmatrix} 0 & 0 & \Xi & 0 & 0 \end{pmatrix} \tag{5}$$

where Ξ is the matrix that relates body frame rotations to rotations about the north, east and down axes and is given as:

$$\Xi = \begin{pmatrix} \frac{\cos\psi}{\cos\theta} & \frac{\cos\psi}{\cos\theta} & 0 \\ -\sin\psi & \cos\psi & 0 \\ \cos\psi\tan\theta & \sin\psi\tan\theta & 1 \end{pmatrix} \tag{6}$$

BAG-OF-WORDS IMAGE MATCHING

An innovative method for generating position updates to aid the INS comes from the Computer Vision community and is known as the Image Bag-of-Words algorithm (Sivic and Zisserman, 2004). This allows a database of photos of different locations to be searched to find the most likely place a query image was captured.

In the scheme described in this paper we first 'survey' the area to be navigated. This survey involves capturing photographs throughout the area (including environments where GNSS reception is difficult) while using accurate navigation sensors to record where these photos are taken. Following the survey accurate position estimates are obtained using post-processing algorithms such as Kalman filter smoothing to give more accurate position estimates than would be possible for real-time applications. The survey results in a set of images along with their position that can be used to create an Image Bag-of-Words database that can later be searched. This provides a method that provides occasional position updates that can be used to aid the previously described foot mounted inertial sensors using the Kalman filter.

The Image Bag-of-Words scheme used in this paper is the scheme described by Botterill et al. (2008), which is based on the scheme of Nistér and Stewénius (2006). Each image observed is added to a Bag-of-Words database as follows: firstly images are converted to greyscale, then processed with the FAST corner detector (Rosten and Drummond, 2006). This identifies typically 400 points (corner features) in each image that are repeatable (i.e. can be detected in subsequent images) and are likely to be distinctive (i.e. have locally high changes in image intensity). Image patches sized 11×11 around these corners are used as 'descriptors'; vectors encoding the local appearance of each feature. The similarity of two of these descriptors is measured with the L_2 norm. Each of these descriptors is then mapped to the closest descriptor ('image word') from a quantised 'dictionary' (a representative subset of these features) so that the image is represented as the set of the image words it contains, i.e. as a Bag-of-Words . FAST Corner detection takes around 10ms per image, and adding an image to the Bag-of-Words database takes 2ms.

The idea behind the Bag-of-Words algorithm is that two images of the same place will contain many of the same features, and hence will have many image words in common. Therefore to find images appearing similar to a new image, the new image is again represented as a Bag-of-Words . This list of words is compared to the list of words representing each other image, and those with most words in common are selected, hence providing a list of surveyed locations that appear similar to the current location.

OBTAINING RELATIVE CAMERA ORIENTATION

When surveying the area to be navigated, the orientation of the camera is recorded each time an image is captured. When navigating the area, if the area is sufficiently visually distinctive, then the Bag-of-Words algorithm will recognise that the pedestrian is seeing a scene that was photographed earlier. From these two images we can calculate the relative orientation of the cameras, and as we know the absolute orientation of the camera from the survey this will provide an orientation update to the Kalman filter. This section describes the procedure for computing these relative orientations from pairs of images. This procedure is summarised in Figure 2.

The first step in recovering relative camera orientation is to find a set of possible correspondences between the two images. A correspondence is a feature in one image matched to the same feature in the other image. Possible correspondences are easily obtained from the Bag-of-Words representations of the two images by finding image words common to both images. Matches between the locations of these features provide the correspondences. Often multiple features in each image are described by the same image word; in this case a pairwise comparison between their descriptors identifies those that are significantly more similar in appearance than other possible pairs as correspondences. When multiple (N) features in one image appear similar to multiple (M) features in the other image, all possible pairs are considered as possible correspondences (referred to as '$N-M$ correspondences').

Many of these correspondences found will be gross outliers so the next step is to remove these while fitting a model of the relative pose of the two cameras. This relative pose is described by an essential matrix (Hartley and Zisserman, 2003), which can be computed from a minimum of five correspondences (Stewénius et al., 2006), but is most accurately estimated by a least-squares fit to a large number of inlier correspondences (Hartley and Zisserman, 2003). This matrix is computed from the BoW correspondences, while simultaneously identifying inliers, using BaySAC (Botterill et al., 2009b). BaySAC, based on the popular RANSAC outlier removal scheme (Fischler and Bolles, 1981), requires that each correspondence is assigned a prior inlier probablity. These probabilities are determined from the number of possible matches in the other image that each feature could have (Botterill et al., 2009b). BaySAC now finds an essential matrix as follows: the five correspondences most likely to be inliers are selected and a model is fitted to these correspondences. The total number of correspondences compatible with this model being correct are now selected. When few compatible correspondences are found, the inlier probabilities of these five points are updated (at least one is likely to be an outlier), the most likely set of five correspondences is selected again, and the procedure is repeated. When a model compatible with many correspondences is found this model is very likely to be correct, in which case the compatible points are selected as inliers.

The matrix and inlier correspondences are further refined using top-down outlier-removal (Rousseeuw and Leroy, 1987): the essential matrix fitting all inlier correspondences is computed using the linear least-squares normalised eight-point algorithm (Hartley and Zisserman, 2003), then the correspondences least compatible with this matrix are removed. The essential matrix is decomposed to give four possibilities for the relative camera pose (relative orientation, and translation up to an unknown scale factor). The correct possibility is taken to be the one leading to most reconstructed points falling in front of the cameras (Hartley and Zisserman, 2003).

The error function minimised by the normalised eight-point algorithm is not a particularly good measure as it is biased towards points far from the points' centroid. A better measure of error is Sampson's error (Hartley and Zisserman, 2003). The essential matrix maps each point to a line (its epiline) in the other image. Sampson's error is given by the squared distance each matching point falls from its corresponding epiline. This error is minimised by Levenberg's method for nonlinear optimisation (Levenberg, 1944), giving our final refined relative camera orientation.

The uncertainty in the relative camera orientation (σ_{angle}) is quantified by fitting a distribution to errors in models fitted to simulated correspondences with added noise. While many factors contribute to this uncertainty, the number of inliers found D, and the uncertainty in point localisation (σ, about 0.69 pixels for the FAST corner detector) as a fraction of the mean length of feature tracks (m pixels) account for most of the error, giving $\sigma_{angle} = K \frac{\sigma}{m\sqrt{D-5}}$ with $K = 9.8$.

To estimate and refine the relative camera pose from 400 possible correspondences of which 50% are outliers takes around 70ms. 75% of this time is taken computing essential matrices from sets of five points. All C++ source code used in this section is available online (Botterill, 2010).

DISCUSSION

The use of image based position updates provides many potential benefits. Firstly, there is no requirement for a dedicated infrastructure such as that required for RFID or Ultrawideband positioning. Therefore large areas can be covered inexpensively without the need for any installation of

For each pair of frames:

1. **Find Matches**
   - Find a set of possible correspondences from the Bag-of-Words representations of the two images (typically 25%-75% of about 200 matches will be inliers)
2. **Find inliers and approximate solution**
   - Use BaySAC to identify which of these possible correspondences are inliers, simultaneously estimate first approximation to the essential matrix (inlier set now contains over 90% inliers)
3. **Refine inliers and solution**
   - Topdown refinement finds more inliers and rejects more outliers while simultaneously fitting an increasingly accurate essential matrix (inlier set now contains about 98% inliers)
   - Choose camera matrix from E and reconstruct 3d point positions (up to a scale factor); reject points falling behind camera (inlier set now contains about 99% inliers)
   - Refine camera matrix by Levenberg-Marquardt least-squares optimisation to reduce Sampson's error function (errors in relative orientation reduced by around 30%)
   - Relative orientation now given from refined camera matrix
4. **Estimate uncertainty in relative orientation**
   - Variance estimated from number of inliers and variance in image feature measurements.

Figure 2. Procedure for estimating the relative orientation of two camera positions

equipment or power supplies. For image updates to be successful (i.e. match correctly) it is necessary for distinctive features to be present within the image. Fortunately, many indoor locations provide feature rich environments with objects such as posters, pictures and signs making good features by which to identify locations. The Image Bag-of-Words algorithm provides an efficient method for searching large numbers of images. Imaging sensors such as digital stills and video cameras are inexpensive and are widely available. One obvious limitation of such a system is that it is necessary to first 'survey' an area by capturing images with known positions. This is achieved in this paper by optimising the data collection with short inertial surveys, and also using Kalman filter smoothing in post processing. However, a better method would be to use a better quality IMU that is capable of navigating accurately for longer periods of time such as the Honeywell HG1900 which could be used for a one-off survey. Multiple lower cost IMUs could then be used to locate multiple users within the surveyed area.

Some implementation issues arise from using position updates from image matching. Firstly, the Image Bag-of-Words algorithm does not always correctly match images such as when two distinct places appear similar (this is especially a problem when navigating away from mapped areas), or due to different places coincidentally having distinctive features in common. These erroneous position estimates are unlike errors from other navigation sensors, and if accidentally incorporated into the Kalman filter they can destroy previous good position estimates, so it is important to identify errors in advance. Previously we handled this by checking the Kalman filter innovations (the difference between the predicted INS position and the new measurement) and rejecting matches when measurements does not agree to within a certain number of standard deviations. An additional check is now provided by BaySAC: if no inlier set is found then it is likely that the features matched by Bag-of-Words do not actually lie on the same 3D objects, so the Bag-of-Words match is likely to be incorrect.

The procedure for estimating relative orientation described in the previous section also estimates the direction of one camera relative to the other. The magnitude of this direction is not known however without knowing the actual scale of the observed objects in the scene, so this direction is not currently used. Actual scales could obtained by using stereo cameras, by 3D reconstruction during the survey, or alternatively by using Object Recognition techniques to recognise objects of known size and from this information to estimate the scale of the world (Botterill et al., 2009a). Although not used in this work, it again highlights the vast potential of using images to provide positional information.

DATA COLLECTION AND PROCESSING

Data was collected at the University of Canterbury campus, New Zealand in August 2009. The equipment used is shown in Figure 3. A Microstrain 3DM-GX1 IMU was attached to the shoe of a pedestrian as shown in the Figure. The Microstrain IMU used was limited in that it is only able to measure rotations of up to 300 deg/s and accelerations of 5g. For pedestrian applications, it is typical to sense rotations of 600-900 deg/s. Therefore, in an attempt to minimise rotational velocity (which is maximum about the rotational axis of the ankle), the IMU was mounted on the foot at an angle of 45 degrees. The user hand carried a Sony Handycam DCR-SR47 digital video camera which records interlaced video frames at 25Hz at a resolution of 720x576 pixels. The video frames were decimated to 4Hz, and only the even horizontal lines were kept to remove the interlacing resulting in an image resolution of 720x288 pixels. The video frames were approximately synchronised to GPS seconds by starting video capture at the same time as the pedestrian's first step. The synchronisation requirements are not very strict for image updates since the accuracy of the position update is relatively low, and the pedestrian also travels at a low velocity (typically less than 1.5 metres per second). The IMU measurements were synchronised to GPS time using the Geospatial Research Centre's Precise Time Data Logger (PTDL) which is able to time stamp the serial data and record the data to SD card. The time stamp accuracy of the PTDL is better than 1 millisecond which ensures consistent timestamps for the IMU and has higher accuracy than is required for the application.

Figure 3. Sony Handycam DCR-SR47; GRC PTDL data logger; Microstrain 3DM-GX1 IMU; pedestrian using system

One problem that arises with the hardware shown in Figure 3 is that it is not possible to use orientation updates directly from the camera to update the IMU. This is because there is an unknown non-constant orientation difference between the camera (held by the user) and the IMU (on the user's foot). Therefore at this stage, the benefit of using orientation updates cannot be utilised directly. In the future it will be necessary to identify the possibility of moving the IMU from the user's foot to the camera (usually held in the user's hand). From this method it is clear that the benefit of frequent ZUPTs will be lost since the IMU will not necessarily remain stationary. Instead either higher grade IMU components will be required; another method will be required to transfer the alignment from the camera to the foot mounted sensor; or another aiding sensor will be needed. This is currently an active area of research by the authors.

RESULTS

Firstly it was necessary to generate a database of images. The same equipment was used for the navigation trial and the database generation, except for the database generation, the data was collected in short loops of less than 5 minutes duration. The navigation data from the IMU was computed using Kalman filter smoothing which combines a forward and reverse processing sweep on the data. This allows accurate reference positions and orientations to be generated without the need for expensive sensors (in the future the data collection step will need to be optimised). The image database that was constructed using the survey data consisted of 4322 images and their coordinates.

A second navigation dataset was collected on the following day. By collecting the navigation data and reference data on different days it gave a greater test of the Image Bag-of-Words algorithm since the lighting conditions were very different (a cloudy day compared to bright sunlight) as well as differences in the environment such as people, furniture, bicycles or vehicles.

The navigation trial spanned a duration of 21 minutes, over which 4Hz image capture resulted in 5066 images being collected. Out of the 5066 images, the Image Bag-of-Words algorithm identified 1374 images (27%) as successful matches. From the 1374 matched images, 51 images (4%) were incorrectly matched. This figure was identified by manually checking the results where an incorrect match is defined as having no common features between images.

Figure 4 shows the navigation solution using the IMU with ZUPTs and demonstrates that the position error accumulates with time as expected. The majority of the position drift is as a result of heading drift as the scale of the trajectory appears correct. This was anticipated since heading updates are not used, and heading error is not observable using a low cost IMU with only ZUPTs. It is likely that this trajectory could be substantially improved with measurements from a compass, although use of a compass has been avoided since the measurements can be unreliable in many environments. One other consideration for the performance of the integrated IMU/ZUPT solution is that the gyro sensors are operating very close to their saturation level of 300 deg/s which means that some large rotations may not be correctly measured. Therefore it is expected that the integrated IMU and ZUPT performance can be improved in the future.

Figure 5 shows the integrated IMU/ZUPT/image solution.

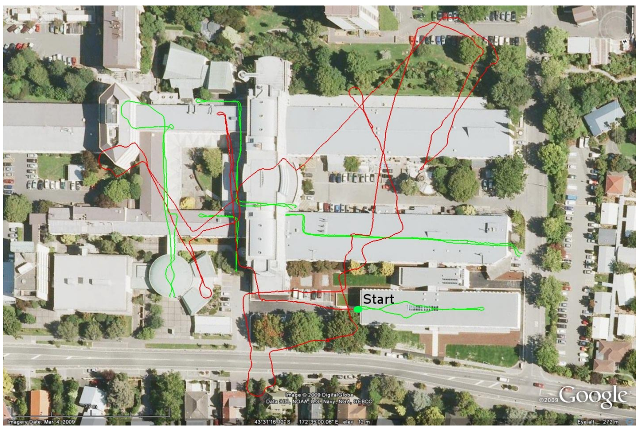

Figure 4. IMU/ZUPT solution (green line shows database; red line shows IMU/ZUPT solution)

The figure shows that the integrated system provides a continuous trajectory although it is clear that the image based position updates are able to substantially improve the position solution, even without orientation updates. Figure 6 can be used to identify the locations where image updates were available. The area that has most notable position drift is on the left hand side of the image and corresponds to an area with less frequent image updates. Through knowledge of the user's trajectory and measuring position error in Google Earth, the largest known position error is measured as 14m. It should be noted that by using the position updates in the Kalman filter, the filter is able to estimate heading error which improves the accuracy of the position solution during periods without image updates. Figure 6 also shows that there are some outliers in the position updates caused by incorrect image matches. These can be identified by the green line that connects successive images matches. These outliers are successfully removed by checking the innovations, although in the future, these incorrect images matches may also be identified using the relative camera orientation algorithm.

It is difficult to determine the accuracy of the integrated solution in the absence of a reference trajectory, which is very difficult to generate accurately in indoor environments.

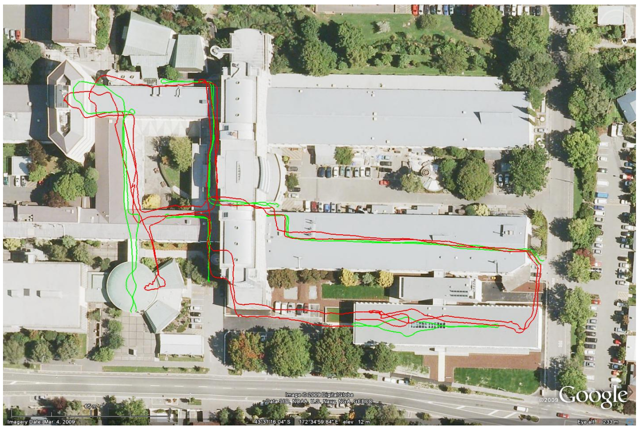

Figure 5. Integrated IMU/ZUPT/image update solution

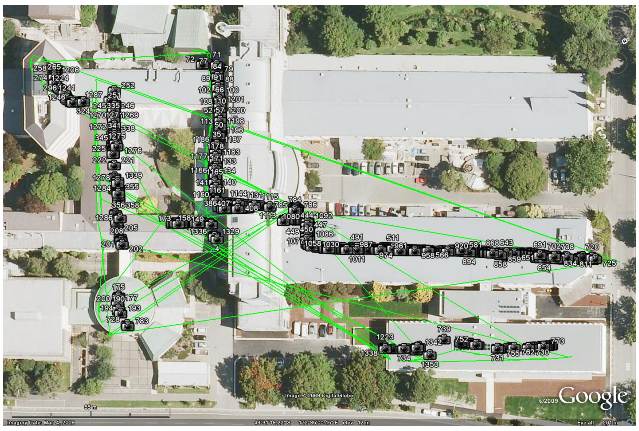

Figure 6. Location of image matches

Analysis of the trajectory using aerial imagery such as that provided by Google Earth also indicates that the position solution is usually better than 10m, and typically 1-5m.

ORIENTATION UPDATES

As previously described, it is not possible to directly use orientation updates in the Kalman filter as the IMU and camera are not rigidly mounted together, therefore the orientation between the two sensors can change. However, the algorithm has been shown to be able to successfully identify corresponding features between new images as shown in Figure 7. Here the features used in the Image Bag-of-Words algorithm have been marked with red and yellow dots in both the new image (shown on the left) and the database image (shown on the right). The yellow dots and green lines identify the correspondences that have been identified by the relative orientation algorithm. The algorithm uses the correspondences to identify the relative orientation and direction of one camera relative to the other. At this stage with the data that is available, it is not possible to compare the angles with the orientation of the IMU because of the unknown rotation between the camera and IMU. The next steps are to investigate the accuracy of the orientation updates and incorporate the measurements into the Kalman filter using the algorithm described in this paper. The initial results indicate that such an approach can be used which is of substantial interest to overcome the issues of heading observability in the IMU filter.

DISCUSSION AND CONCLUSIONS

This paper has demonstrated a novel pedestrian navigation system for positioning in areas of poor GNSS reception. The system consists of a low cost IMU, GNSS receiver and image capture sensor. The IMU is used to provide a continuous position and orientation which can be corrected by measurements from other sensors and ZUPTs when the user's

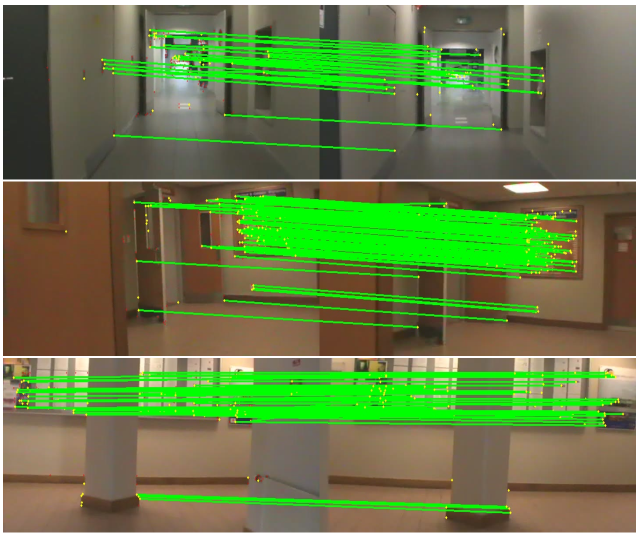

Figure 7. Feature correspondences using relative orientation algorithm

foot is stationary. Position updates from image matching have been demonstrated to be reliable, frequent, and effective in controlling the position drift from the foot mounted IMU.

The main advantages of such a system is that there is no need for a dedicated infrastructure; there is no need to receive signals; and the system can be deployed over very large areas. The paper has also demonstrated the possibility of obtaining orientation updates through the image matching algorithm which can help to reduce the problem of heading drift for the IMU/ZUPT solution. In the future it is also thought that it will be possible to refine the position obtained from the image update by making use of multiple images to resolve the unknown scale of the direction vector obtained from the algorithm. Such a system will work in any environment as long as there are distinctive features within the image. The system has also demonstrated that it is robust to issues such as illumination angle and different objects being contained in the image as the environment changes.

Limitations of the system include the need to generate the survey dataset. This is still a problem since no navigation sensor is currently able to navigate inside buildings for long periods of time with high accuracy. A solution has been identified in this paper using Kalman filter smoothing, however it would be necessary to improve on this in the future. More expensive equipment could be used since the survey would need to be required infrequently. Issues such as the image database becoming out of date could also be addressed where new images and positions are obtained by users. Significant changes in illumination (such as day turning to night) may cause issues in some situations, however most of the areas where the system will be used will be areas such as offices and shopping centres which often have constant artificial lighting. In order for orientation updates to be used, it will also be necessary to colocate the IMU and camera which may mean that it will not be possible to mount the IMU on the foot.

This paper has identified a combination of IMU, GNSS and image sensor which have strong complimentary characteristics. The paper has demonstrated that a position error of less than 10m can be achieved over a period of 21 minutes using an IMU with ZUPTs and position updates from image matches. The paper has identified a wide range of areas in which the integration can be improved in the future.

REFERENCES

Botterill, T., 2010. Computer vision c++ source code. Online. URL `http://hilandtom.com/tombotterill/code`

Botterill, T., Green, R., Mills, S., November 2009a. A Bag-of-Words Speedometer for Single Camera SLAM. In: Image and Vision Computing New Zealand. Wellington, NZ, pp. 1–6.

Botterill, T., Mills, S., Green, R., Nov. 2008. Speeded-up Bag-of-Words algorithm for robot localisation through scene recognition. In: Proceedings of Image and Vision Computing New Zealand. pp. 1–6.

Botterill, T., Mills, S., Green, R., 2009b. New conditional sampling strategies for speeded-up RANSAC. In: To appear in Proceedings of the British Machine Vision Conference.

Cummins, M., Newman, P., May 2008. Accelerated appearance-only SLAM. In: Proceedings of the IEEE International Conference on Robotics and Automation. pp. 1828–1833.

Davison, A., Oct. 2003. Real-time simultaneous localisation and mapping with a single camera. In: Proceedings of the IEEE International Conference on Computer Vision.

Eade, E., Drummond, T., 2006. Scalable monocular SLAM. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 1. IEEE Computer Society, Los Alamitos, CA, USA, pp. 469–476.

Eade, E., Drummond, T., October 2007. Monocular SLAM as a graph of coalesced observations. In: Proceedings of the IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil.

Farrell, J. A., Barth, M., 1999. The Global Positioning System and Inertial Navigation. McGraw-Hill.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381–395.

Foxlin, E., November 2005. Pedestrian tracking with shoe-mounted inertial sensors. IEEE Computer Graphics and Applications 25, 38–46.

Godha, S., Lachapelle, G., Cannon, M., September 2006. Integrated GPS/INS system for pedestrian navigation in a signal degraded environment. In: Proceedings of ION GNSS 2006.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision, 2nd Edition. Cambridge University Press, Cambridge, UK.

Hide, C., September 2003. Integration of GPS and low cost INS measurements. Ph.D. thesis, University of Nottingham.

Hide, C., Botterill, T., Andreotti, M., 2009. An integrated imu, gnss and image recognition sensor for pedestrian navigation. In: Proceedings of the Institute of Navigation GNSS Conference. Fort Worth, Texas, pp. 1–10.

Kim, J.-H., Sukkarieh, S., 2007. Real-time implementation of airborne inertial-SLAM. Robotics and autonomous systems 55 (1), 62–71.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. The Quarterly of Applied Mathematics 2, 164–168.

Nistér, D., Naroditsky, O., Bergen, J., 2006. Visual odometry for ground vehicle applications. Journal of Field Robotics 23 (1), 3–20.

Nistér, D., Stewénius, H., 2006. A minimal solution to the generalized 3-point pose problem. Journal of Mathematical Imaging and Vision 27, 67–79.

Rosten, E., Drummond, T., 2006. Machine leaning for high-speed corner detection. In: Proceedings of the European Conference on Computer Vision. pp. 430–443.

Rousseeuw, P. J., Leroy, A. M., 1987. Robust regression and outlier detection. John Wiley & Sons, Inc., New York, NY, USA.

Scherzinger, B., Reid, D., April 1994. Modified strapdown inertial navigator error models. In: IEEE Position Location and Navigation Symposium. pp. 426–430.

Sivic, J., Zisserman, A., Jun. 2004. Video data mining using configurations of viewpoint invariant regions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Stewénius, H., Engels, C., Nistér, D., Jun. 2006. Recent developments on direct relative orientation. ISPRS Journal of Photogrammetry and Remote Sensing 60, 284–294.

Titterton, D. H., Weston, J. L., 1997. Strapdown Inertial Navigation Technology. Institution of Electrical Engineers.