Towards Vision Based Navigation in Large Indoor Environments

Jaime Valls Miro, Weizhen Zhou and Gamini Dissanayake
Mechatronics and Intelligent Systems Group
University of Technology Sydney (UTS)
NSW2007, Australia
Email: {j.vallsmiro, w.zhou, g.dissanayake}@cas.edu.au

***Abstract*— The main contribution of this paper is a novel stereo-based algorithm which serves as a tool to examine the viability of stereo vision solutions to the simultaneous localisation and mapping (SLAM) for large indoor environments. Using features extracted from the scale invariant feature transform (SIFT) and depth maps from a small vision system (SVS) stereo head, an extended Kalman filter (EKF) based SLAM algorithm, that allows the independent use of information relating to depth and bearing, is developed. By means of a map pruning strategy for managing the computational cost, it is demonstrated that statistically consistent location estimates can be generated for a small (6 m x 6 m) structured office environment, and in a robotics search and rescue arena of similar size. It is shown that in a larger office environment, the proposed algorithm generates location estimates which are topologically correct, but statistically inconsistent. A discussion on the possible reasons for the inconsistency is presented. The paper highlights that, despite recent advances, building accurate geometric maps of large environments with vision only sensing is still a challenging task.**

I. Background and motivation

In mobile robotics, the incremental construction of a map of an unknown environment while concurrently generating an estimate for the location of the vehicle is known as the simultaneous localization and mapping (SLAM) problem. Significant progress has been made, in the past few years, in addressing a range of issues associated with the SLAM problem (see for example, [1], [2], [3], [4], [5], [6], [7] and the references therein).

The advances made over the years towards solving the SLAM problem are indeed related to performance improvements in both sensors and computer hardware. For indoors robots in particular, the SLAM problem was initially addressed mostly using sonars, and then came the ubiquitous laser range finder, which has proved a breakthrough in autonomous mobile navigation. Laser sensors provide accurate 2D depth measurements (extendable to 3D with additional tilting units), and many SLAM-related algorithms have been devised based on data obtained specifically from laser range finders.

Wide availability of low cost, low power light-weight cameras as well as maturity of computer vision algorithms have made real-time vision processing much more practical in recent times, and consequently there has been an increasing interest in visually based navigation systems in the robotics community. Vision SLAM in particular has seen many advances in recent years [8], [9], [11]. Cameras are interesting as they provide a wealth of geometric information from an unmodified scene, as well as perceptual information such as textures and colours, which can be matched by few other sensors. A low-cost and lightweight vision based solution to the SLAM problem in an indoor setting is of great value, particularly for robotic search and rescue scenarios.

Both monocular and stereo pairs have been used for mobile robot's vision-based mapping and navigation. The former can not directly retrieve depth information from the scene, therefore traditional Bayesian techniques to solve the SLAM problem such as the EKF can not be readily used with information from single cameras [13]. Special landmark initialization techniques have been proposed in the literature to overcome this, thus enabling a full Gaussian estimate of its estate and the application of EKF [14], [15], [16]. An interesting solution with mono-vision SLAM is provided by vSLAM [17], where SIFT features [18] are combined into landmarks to populate a visual database. These are in turn employed by a Kalman filter to build a map, onto which the robot localises itself by means of a particle filter. vSLAM has been demonstrated to provide navigation capabilities to indoor mobile robots operating in relatively small environments, for example, in a two bedroom apartment. Reliance of odometry for landmark initialisation (in effect providing the scale which is not observable from a bearing only sensor) is certain to make it unsuitable where wheel slip may be significant, for example in case of indoor search and rescue scenarios. Svedman [19] reports on using information from two cameras with vSLAM in an attempt to remove the need to drive the robot and take a number of sequential frames. Davison [9] uses a template to introduce features with known geometric relationships to determine the scale and completely avoid the use of odometry. The demonstrations so far has only been on relatively small environments and with hand-waved sensing where the camera is manoeuvred to maximise the information gain.

Despite the increased cost, using a stereo camera is advantageous as it makes the system fully observable in that the sensor provides enough information (range and bearing) to compute the full three-dimensional state of the observed landmarks. Some approaches rely on 2D projections of the features, such as the vertical edges corresponding to corners and door frames as proposed in [10], which are subsequently

tracked using an EKF. This type of visual feature representation is not sufficiently distinctive, and require ellaborate data association solutions to reject spurious matches. More recently, a stereo pair has been used in [11] with scale-invariant image features to solve the full 3D SLAM problem based on a Kalman filter framework. This work has demonstrated good results for a robot moving in a small room approximately $10x10\ m^2$. However, as cross-correlations are not fully maintained, and it only relies on local estimates, the algorithm is not globally consistent and will diverge when the area to be explored is large [1]. Precisaly to address the global localisation problem, the authors have recently extended the algorithm [12] by proposing a submapping strategy which relies on highly specific SIFT features to locally correct the odometry, and the global alignment of the submaps. Yet the backward correction step is constrained to the actual closure of the loop to correct for the effects of drifts and slippage, and results are still restricted to the same small room than in the previous paper. Three-dimensional metric maps are also obtained using stereo in [20] by implementing a Rao-Blackwellised particle filter to counteract for the sensitivity of EKF to outliers in landmark detection. A motion model based purely on visual odometry is also used, effectively generalising the problem to unconstrained 3D motion. Feature management and computational complexity, which grows exponentially with the number of particles, is likely to make this strategy infeasible in a large environment. Other approaches rely on iterative minimization algorithms from vision techniques such as ICP to perform local 3D alignments to solve the SLAM problem [21], which produce good results when restricted to fairly structured environments. Similar limitation are reported in [22], where 3D line segments become landmarks suitable to be used in a particle filter implementation where each particle also carries with it an EKF.

In this paper a number of innovations to overcome some of the inherent issues associated with stereo vision based SLAM algorithms are presented. For all its obvious advantages, even after meticulous calibration, stereo depth information of measurements beyond a few meters is generally not accurate to be fully relied upon. Also, range densities are entirely dependent on the textures of the surfaces being observed, so that positive stereo correlation of arbitrary visual features in a scene can not be guaranteed. Bearing to features, on the other hand, is a relatively reliable visual measurement: a known feature at infinity in fact provides accurate information on the orientation of a robot even when the estimate of the robot position is relatively inaccurate. In the approach presented in this paper, bearing and disparity to features on the camera image extracted through the SIFT algorithm are used. Unlike the range and bearing from a stereo head, bearing and disparity to a given feature can be treated as two independent *measurements*. It is thus possible to update the robot and feature states, even when only the bearing information is reliable. Performing an update using the bearing information first also enables reliable rejection of outliers due to errors in the disparity measurement, without sacrificing the information contained in the bearing measurement. Data association, particularly when the feature density is high and the quality of the range information is poor is also a significant issue in SLAM. Use of SIFT descriptors together with a bearing innovation gate is used to overcome this problem. A map management strategy to eliminate features that are not frequently re-observed is used to address the computational cost issues.

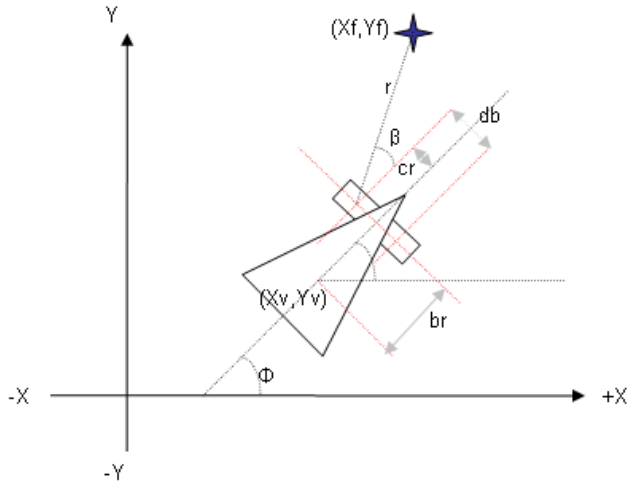

Fig. 1. Stereo-equipped vehicle and feature projection on 2D world coordinate.

The remainder of this paper is organised as follows: Section II summarises the mathematical framework employed in the study of the SLAM problem. Section III reviews the relevant aspects of the SIFT algorithm, with a discussion of data association issues in SLAM. Then, the proposed methodology for the solution to the visual-based SLAM problem is presented in Section IV. Detailed experimental setup, results obtained and a discussion are provided in Sections V and VI respectively. Finally, Section VII summarises the contribution of this paper.

II. The SLAM problem formulation

The setting for the SLAM problem is that of a vehicle with a known kinematic model, starting at an unknown location, moving through an environment containing a population of features or landmarks. The vehicle is equipped with a sensor that can take measurements of the relative location between any individual landmark and the vehicle itself.

The state of the system $\mathbf{x}_v(k)$ consists on the position and orientation of the vehicle together with the position of all landmarks. The motion of the vehicle through the environment is modelled by a conventional discrete-time state transition equation or process model:

$$\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{y}(k+1) \\ \phi(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{x}(k) + \Delta T \times V_v \times \cos\phi \\ \mathbf{y}(k) + \Delta T \times V_v \times \sin\phi \\ \phi(k) + \Delta T \times \omega_v \end{bmatrix} \quad (1)$$

where V_v corresponds to the vehicle linear velocity, and ω_v is the angular velocity. The vehicle is equipped with a sensor

that can obtain observations of the bearing $\mathbf{z}_\beta$ and disparity $\mathbf{z}_d$ of landmarks with respect to the vehicle according to:

$$\begin{bmatrix} \mathbf{z}_\beta \\ \mathbf{z}_d \end{bmatrix} = \begin{bmatrix} \arctan\left(\frac{\mathbf{y}_f - \mathbf{y}_v - br \times \sin\phi - cr \times \cos\phi}{\mathbf{x}_f - \mathbf{x}_v - br \times \cos\phi + cr \times \sin\phi}\right) - \phi \\ \frac{-\mathrm{db} \times f}{\mathrm{br} - \sin\phi \times (\mathbf{y}_f - \mathbf{y}_v) - \cos\phi \times (\mathbf{x}_f - \mathbf{x}_v)} \end{bmatrix} \quad (2)$$

The Kalman filter is the sensor fusion technique used in the approach to SLAM presented in this paper. The reader is referred to [1] and the references therein for further details about EKF SLAM. In essence, the filter recursively computes estimates for a state $\mathbf{x}(k)$ which is evolving according to the process model and which is being observed according to the observation model. The Kalman filter computes an estimate which is equivalent to the conditional mean $\hat{\mathbf{x}}(p|q) = E\left[\mathbf{x}(p)|\mathbf{Z}^q\right]$ $(p \geq q)$, where $\mathbf{Z}^q$ is the sequence of observations taken up until time q. The error in the estimate is denoted $\tilde{\mathbf{x}}(p|q) = \hat{\mathbf{x}}(p|q) - \mathbf{x}(p)$. The Kalman filter also provides a recursive estimate of the covariance $\mathbf{P}(p|q) = E\left[\tilde{\mathbf{x}}(p|q)\tilde{\mathbf{x}}(p|q)^T|\mathbf{Z}^q\right]$ in the estimate $\hat{\mathbf{x}}(p|q)$.

Note that unlike in traditional formulations where either the range and bearing from the robot to a feature or the Cartesian coordinates of a feature relative to the robot [11] is used as observations, the disparity and the bearing to a feature can be treated as two independent measurements.

III. Visual features processing

In the work proposed here, an efficient mechanism to detect and represent stable local features was required. An immensely popular choice drawn from computer vision as a fundamental component of many image registration and object recognition algorithms is SIFT [11], [18]. Whilst not the only one, a recent comparative study [23] of several local descriptors showed that the best matching results were obtained using the SIFT mechanism, which was identified as the most resistant to common image deformations. This made it the sensible choice for our research, and the work of other researchers working on SLAM also seem to agree with this judgement (see [8], [11] and [12] for instance).

A. Data association with SIFT

The main strength of SIFT is to produce a compact (128th dimensional) landmark descriptor that allows quick comparisons with other regions, and is rich enough to allow these comparisons to be highly discriminatory. This is particularly so as the descriptor representation is designed to avoid problems due to boundary effects, i.e., smooth changes in location, orientation and scale do not cause radical changes in the feature vector. Furthermore, while the representation was not designed to be explicitly invariant to affine transformations , it is nevertheless surprisingly resilient to deformations such as those caused by perspective effects [23]. The location of each keypoint in the image is specified by 4 floating point numbers $[x, y, s, o]$ giving subpixel row and column location, scale, and orientation (in radians from -PI to PI) respectively.

The evident matching performance of the descriptors is what makes them an ideal candidate to the on-going problem in SLAM of robust data association. In particular when the pose estimate of the vehicle is in gross error, which means that despite the fact the vehicle might be in an area already mapped, loop closure solely based on the traditional geometry-based nearest neighbour innovation gating is not feasible, resulting in wrong re-mapping and erroneous global locations [8]. Figure 2 shows the relative insensitivity of SIFT to changes in viewpoint from the same scene by correctly matching corresponding keypoints. This is also applicable to image pairs obtained from the stereoscopic sensor, as seen in Figure 3. Features with spurious existence, and those which don't lie in the camera epipolar line can therefore be eliminated, and only surviving features that appear in both left and right images are then allowed to be initialized and integrated into the SLAM feature database. Furthermore, matches across a stereo pair can be used to generate an approximate range estimate to the features. This is particularly useful during the feature initialization step and when the stereo algorithm does not return disparity values due to lack of texture.

In an indoor environment, the potential is high for scenes or regions in an image that are very similar in appearance. Therefore it is conceivable that SIFT produces many incorrect

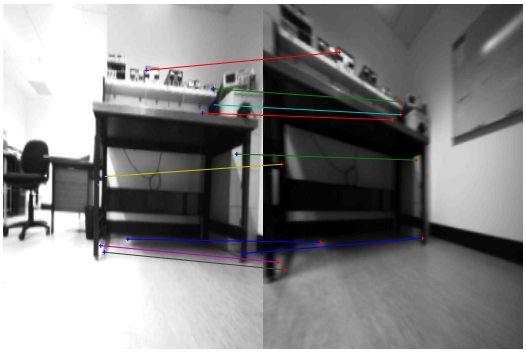

Fig. 2. Corresponding keypoints which show the robustness of SIFT to changes in view point.

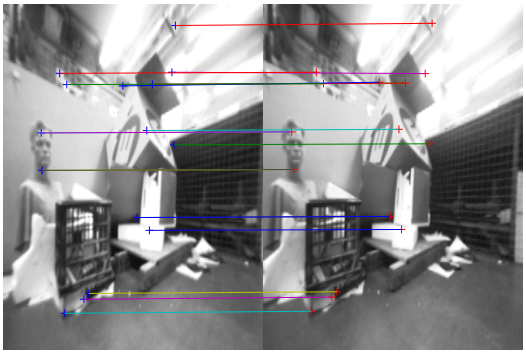

Fig. 3. Matched keypoints between a stereo pair of images in the rescue arena.

matches in an indoor scene. Use of the Bayesian innovation gate traditionally used in the Kalman filter estimator in conjunction with the SIFT descriptor was found to solve this problem.

IV. The Visual-SLAM algorithm

The algorithm for visual-SLAM as used in the experiments can be described as follows:

1) Initialization: set up a world coordinate frame at the initial robot location.
2) Initialization: obtain a stereo image pair of the scene from the camera and run SIFT on both left and right images.
3) Initialization: taking the left as the reference image, find matches by looking for the descriptor vector in the right image with closest Euclidean distance. Some further thresholding is carried out as suggested in [11] to keep only the most unique and distinctive features, discarding the feature if it is considered to be too similar to more than one keypoint. Potential mismatches are further filtered out by enforcing keypoints to remain on epipolar lines.
4) Initialization: taking disparity as the horizontal pixel difference between the left and right matches, together with the (pixel) position of the features in the image and the camera intrinsic parameters, triangulate to obtain and estimate of the 2D coordinates relative to the robot. Compute the feature coordinates in the world frame and incorporate this into the state vector.
5) Loop: predict robot motion using information from the encoders. Get stereo images and perform feature matching as in Step 2 and 3. Calculate a dense disparity map. A commercial stereo package is used for this purpose (see Section V-C below). Given the matched descriptors, search for positive associations with landmarks already in the current map, in a similar fashion to Step 3 above except epipolar restrictions are not exploited.
6) Loop: if an association is found, further validate this by first computing the bearing innovation and using a bearing innovation gate of 2σ. Range innovation of the feature is also calculated here to make sure the tentative match is indeed consistent: if range innovation is less than 20% of the predicted range, the feature becomes part of the matched list of bearing to features for later batch update.
7) Loop: extract the disparity from a $3x3$ pixel window around the feature location in the image plane. Select the median disparity of this window as the disparity for the feature. Predict an expected range for all the features for which a stereo disparity is available and use a gate to reject observations that are incorrect or regarded as out of the reliable range of the stereo sensor.
8) Loop: use a batch bearing update followed by a batch disparity update to generate new estimates for the feature and robot locations.

Fig. 4. The mobile platform with the mounted stereo and laser sensors.

9) Loop: initialise unmatched features using the procedure given in Step 4
10) Loop: map management. Some form of map maintenance needs to be implemented as the number of observed features can grow very large and tracking them all can become computationally very expensive. This is particularly true when many of the features might never be re-observed. A two-fold process has been devised to only retain the most significant point features: firstly, only after a predetermined number of frames (around 5 based on experimental results) are re-observed beacons regarded as a permanent features. Beacons that fail this test are deleted from the state vector and the state covariance matrix as described in [24]. How often map management needs to be carried out depends mostly on the processing power and memory available, as well as the run itself - longer runs need to prune more often. In our experience, performing map management every 20 to 30 frames was sufficient to produce a manageable map. These two values, as well as the number of successful hits to a feature before it is included in the map (3 to 5 in our experiments), are arbitrary and at the moment based on a trade-off between map density and accuracy, and computational complexity. We are currently investigating how to automatically adjust these parameters on the run. The median of the number of hits seems like a reasonable option, so that the least significant 50% of landmarks can be pruned at spaced intervals.
11) Loop again (back to Step 5).

V. Experimental setup

A. Robotic platform

To test the validity of the approach data was collected with an ActiveMedia Pioneer 2DX robot mounted with an stereoscopic camera, as depicted in Figure 4. The robot was also equipped with a SICK LMS200 laser rangefinder to evaluate the outcome of the vision based SLAM algorithm by superimposing them with range and bearing measurements of the environment.

The robot was driven through two distinctively different unmodified environments:

- a highly structured, low-texture open space office environment, with around $1.5\ m$ height partitions, narrow corridors and research students happily crammed in there.
- an arena being used to simulate search and rescue scenarios, with rubble and debris from a collapsed-like building, as pictured in Figure 3.

The robot was used to capture real-life stereo images, odometric poses and laser scan measurements at around 4Hz whilst being driven at speeds of $0.2 - 0.3\ m/sec$. Stereo and pose logged data was then processed by the algorithm described earlier in Section IV

B. Stereoscopic headset

The stereo head used is the STH-MDCS from Videre Design, a compact, low-power colour digital stereo head with an IEEE 1394 digital interface. It consists of two 1.3 megapixel, progressive scan CMOS imagers mounted in a rigid body, and a 1394 peripheral interface module, joined in an integral unit. Wide-angle lenses (FoV = 100 degrees) were fitted for this exercise (narrow angle lenses were also tested with poorer results as a lesser number of good quality distinctive features were picked up). The camera was mounted at the front and top of the vehicle at a constant orientation, looking forward. Images obtained were restricted to greyscale $320x240$ pixels.

C. Software environment

The widely used Player open source robotics architecture, running under Linux, was the software of choice to interface with the robotic platform and the sensors to perform the synchronous data collection and actual control of the robot. The SRI Small Vision System (SVS) software was employed to calibrate the stereo head and perform stereo correlation within the Player framework.

VI. Results and Discussion

Stereo SLAM results are shown in Figures 5, 6 and 7. In the first example, the robot is driven in the office environment around two adjacent partitions in an area of around $6x6\ m^2$, closing an outer and inner loop. A similar run in an unstructured experimental search and rescue arena is shown in Figure 6, which covers an area of approximately $7x7\ m^2$. The longer run shown in Figure 7 stretches over approximately $6x18\ m^2$ in the office environment with many interweaving loops, covering a total travelled distance of approximately $150\ m$. The landmarks locations are shown as red stars. The landmarks that appear in the open spaces are those due to features detected on the ceiling. The filter was tuned based on a short run in the office environment by selecting the process and measurement noises using information obtained from the laser as the ground truth, and our previous experiences with laser-based SLAM on the same Pioneer platform. The same noise parameters were then used in all subsequent experiments. We currently assume $\sigma_v^2 = 0.0056$ and $\sigma_\omega^2 = 0.0056$ for process errors in velocity and turn rate respectively, a bearing observation error of $\sigma_b^2 = 0.0056$, and in accordance with [11] a disparity variance of $\sigma_d^2 = 1.0$.

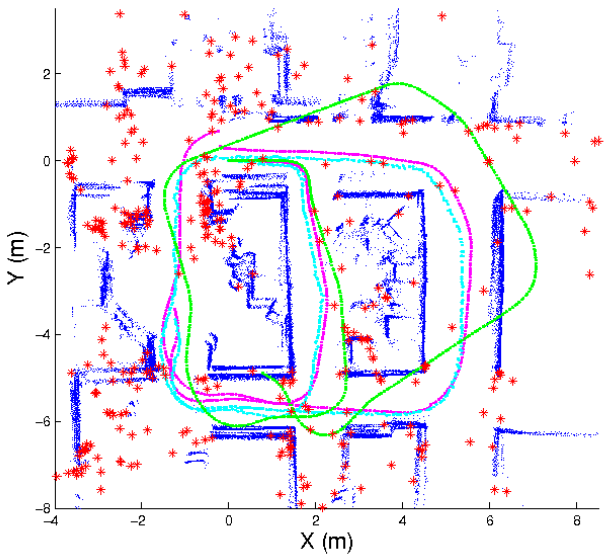

Fig. 5. Short office run: SLAM in pink. Odometry in green, ground truth from laser ICP in cyan. Walls from superimposed laser scans using SLAM poses. Stars in red are landmarks.

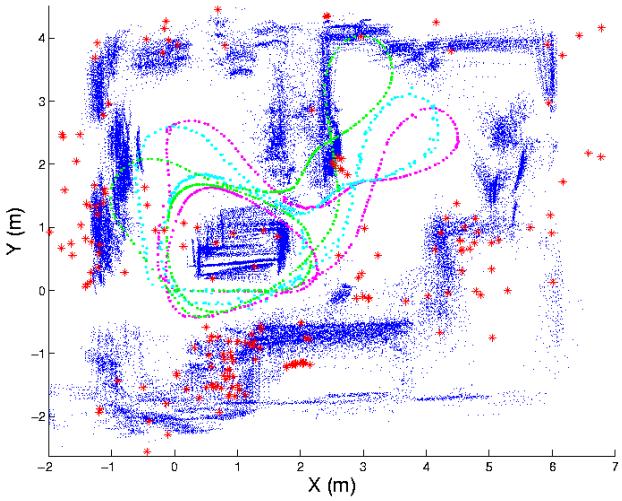

Fig. 6. Short search and rescue arena run: SLAM in pink. Odometry in green, ground truth from laser ICP in cyan. Walls from superimposed laser scans using SLAM poses. Stars in red are landmarks.

Although there are some errors present, the SLAM estimate is significantly superior to that obtained from dead-reckoning as expected. This is particularly apparent in the long run in Figure 7, where odometry falls mostly outside of the detail shown here (features have been removed to make the figure more readable). The approximate geometry of the environments recovered by superimposing laser range scans using the robot poses generated by the SLAM algorithm closely resembles the actual maps, thereby provide a qualitative indication of the validity of the robot location estimates. The apparent thickness of the boundaries of the workspace is small

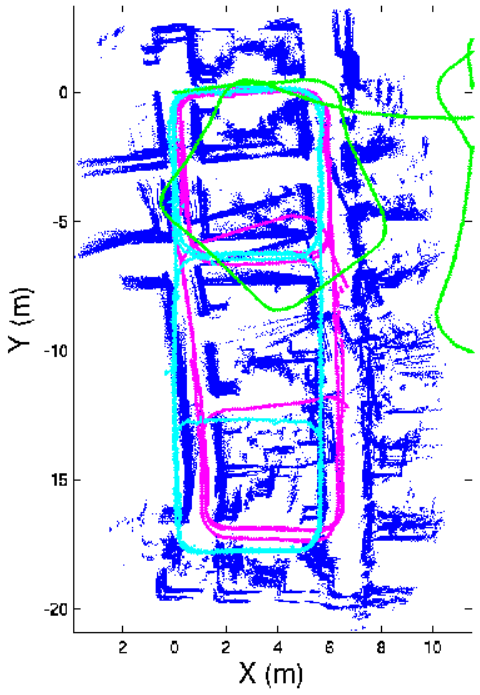

Fig. 7. Long office run: SLAM in pink. Odometry in green (mostly off figure), ground truth from laser ICP in cyan. Walls from superimposed laser scans using SLAM poses. Landmarks not shown.

indicating that the shape of the robot trajectory generated by SLAM is locally consistent. This is also true in the longer run office environment, although accumulated errors increase more significantly as the robot travels further away from its initial point, hence producing less consistent geometric maps. Discontinuities are due to errors present before loop closures which result in sudden jumps in the robot location estimates. Loop closures were facilitated by the ability of SIFT to reliably associate new features to their previously seen counterparts. Thus the algorithm developed appears to be adequate to enable a robot to navigate in unknown indoor environments.

Figure 8 shows a histogram of matches between observations and the landmarks in the final map. This forms the basis of the map management strategy described in Section IV. Clearly, a large number of features have not been repeatedly observed and may be deleted from the map without incurring a significant loss of information. Figure 9 shows errors in the robot location estimates (x, y, ϕ) together with the associated 95% confidence limits for the three experimental runs. At loop closures, indicated by the sudden reductions in the confidence limits, the position errors are in the range of $0.2\ m$. However, it is clear that the filter estimates for the long office run (subplot c) are statistically inconsistent as the error in the robot location estimate is predominantly outside the 95% confidence limits. Even using larger than expected values for measurement noises, it was not possible to tune the estimator to achieve a statistically consistent result. As the process model used in the EKF has been proven using laser-based localisation as well

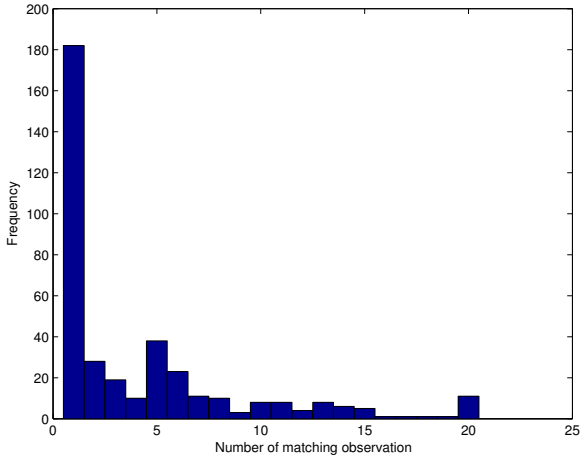

Fig. 8. Histogram of matched landmarks.

as SLAM, this points to an inadequate measurement model, although the model used in this paper, as depicted by Figure 1, is geometrically straightforward and is based on the current literature on vision-based SLAM. The non-Gaussian nature of the depth observations, particularly due to the short base-line of the stereo setup used is perhaps the most significant factor contributing to this error. Significant unmodelled variations in the disparity measurement due to poor texture, spacial discontinuities that are not properly captured by the validation based on a $3x3$ pixel patch and errors in the stereo calibration could also be contributing factors. We are currently in the process of examining these possibilities.

VII. Conclusions

A solution to the simultaneous localisation and mapping problem in an unmodified indoor environment has been presented. The approach, which uses an extended Kalman filter, assumes the availability of simultaneous visual information from two cameras, from which depth information is extracted. A measurement model that separates the information contained in the disparity and the bearing measurement is utilised, making it straightforward to exploit the bearing measurements when depth information is not available or seems unreliable. Data association based on a combination of SIFT descriptors and a Bayesian innovation gate has been exploited to enable loop closure even when the feature density is high and nearest neighbour data association on its own is impractical. A map management strategy to eliminate features that do not significantly contribute information to the estimator has also been implemented.

Results have been presented which demonstrate the viability of the innovations proposed to obtain reasonable estimates of robot locations and maps in two distinctively different small indoor environments, an office and a search and rescue arena, based on visual cues. Although location estimates that may be adequate for autonomous navigation were also obtained for longer runs, these were shown to be statistically inconsistent.

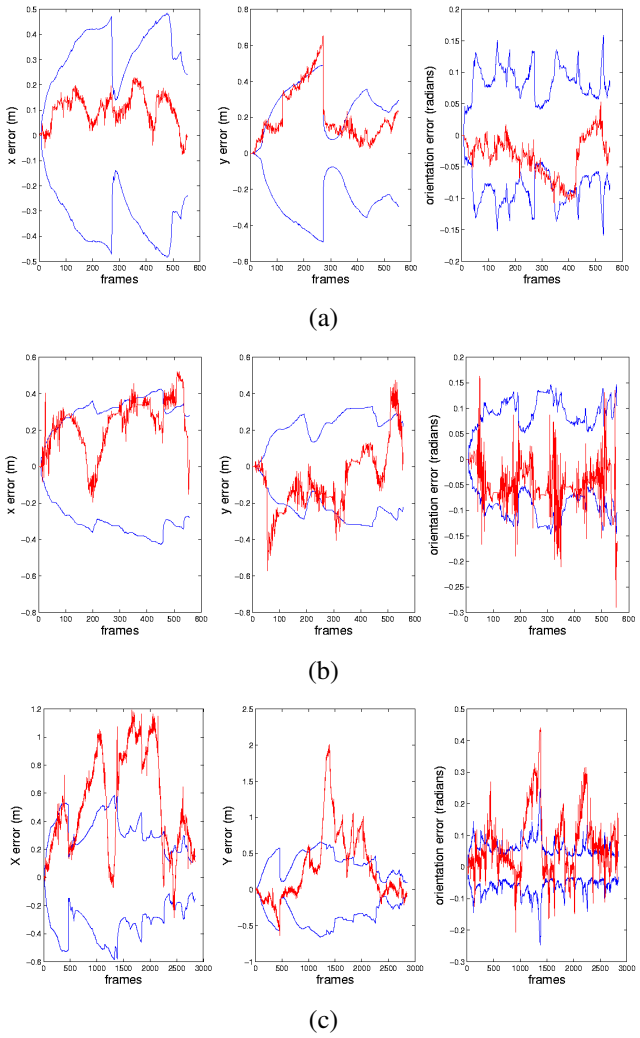

Fig. 9. 2σ error plots for the short office (a), rescue arena (b) and long office environment (c).

Further work is required to investigate the reasons for these errors. We believe that the statistical inconsistency was not apparent in the published literature because the experiments reported in previous publications either did not compare the location estimates generated with the ground truth or were conducted in small areas. Therefore, reliable generation of accurate geometric maps for larger indoor environments using vision only sensing still poses a significant challenge.

Further work is currently underway to test the algorithm on-line in a search and rescue mobile robot, and to naturally extend the estimation problem to 3D pose estimation. Furthermore, studying the 2D/3D SLAM problem when the odometry from wheel encoders is completely unreliable, or non-existent, is also planned.

REFERENCES

[1] G. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem", *IEEE Trans. on Robotics and Automation*, vol. 17, pp. 229–241, 2001.

[2] J. E. Guivant, E. M. Nebot, "Optimization of the simultaneous localization and map building (SLAM) algorithm for real time implementation", *IEEE Trans. on Robotics and Automation*, vol. 17, pp. 242–257, 2001.

[3] P. Newman, On the Structure and Solution of the Simultaneous Localization and Map Building Problem. PhD thesis, Australian Centre of Field Robotics, University of Sydney, Sydney, 2000.

[4] J. A Castellanos, J. Neira, J. D. Tardos, "Multisensor fusion for simultaneous localization and map building", *IEEE Trans. on Robotics and Automation*, vol. 17, pp. 908–914, 2001.

[5] J. Leonard, P. Newman, "Consistent, convergent and constant time SLAM", in *Int. Joint Conf. on Artificial Intelligence*, Acapulco, Mexico, 2003, pp. 1143–1150.

[6] J. Folkesson, H. I. Christensen, "Graphical SLAM - a self-correcting map", in *Proc. IEEE Int. Conf. on Robotics and Automation*, New Orleans, LA, 2004, pp. 383-390.

[7] S. Thrun, Y. Liu, D. Koller,A. Y. Ng, Z. Ghahramani, H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters", *Int. J. of Robotics Research*, vol. 23, pp. 693–716, 2004.

[8] P. Newman, K. Ho, "SLAM - loop closing with visually salient features", in *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, 2005, pp. 644–651.

[9] A. J. Davison, D. Murray, "Simultaneous localization and map-building using active vision", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24(7), pp. 865–880, 2002.

[10] J. A Castellanos, J. M. M. Montiel, J. Neira, J. D. Tardós, "Sensor influence in the performance of simultaneous mobile robot localization and map building", in *6th Int. Symp. on Experimental Robotics*, Sydney, Australia, 26-28 March 1999, pp. 203–212.

[11] S. Se, D. G. Lowe, J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks ", *Int. J. of Robotics Research*, vol. 21(8), pp. 735–758, 2002.

[12] S. Se, D. G. Lowe, J. Little, "Vision-based global localization and mapping for mobile robots", *IEEE Trans. on Robotics*, vol. 21(3), pp. 364–375, 2005.

[13] T. Bailey, "Constrained initialization for bearing-only SLAM", in *Proc. IEEE Int. Conf. on Robotics and Automation*, Taipei, Taiwan, 2005, pp. 1966–1971.

[14] P. Newman, J. Leonard, R. Rikoski, M. Bosse, "Mapping partially observable features from multiple uncertain vantage points", *Int. J. of Robotics Research*, vol. 21(10–11), pp. 943–976, 2002.

[15] N. M. Kwok, G. Dissanayake, "An efficient multiple hypothesis filter for bearing-only SLAM", in *IEEE Int. Conf. on Intelligent Robot and Systems*, Alberta, Canada, 2004, pp. 736–741.

[16] J. Solà i Ortega, T. Lemaire, M. Devy, S. Lacroix, A. Monin, "Delayed vs Undelayed landmark initialization for bearing-only SLAM", presented at the IEEE Int. Conf. on Robotics and Automation workshop on SLAM, Barcelona, Spain, 2005.

[17] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, M. E. Munich, "The vSLAM algorithm for robust localization and mapping", in *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, 2005, pp. 24–29.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *Int. J. of Computer Vision*, vol. 60(2), pp. 91–110, 2004.

[19] M. Svedman, "3-D structure from stereo vision using unsynchronized cameras", Master's thesis, Royal Institute of Technology (KTH), 2005.

[20] R. Sim, P. Elinas, M. Griffin, J. J. Little, "Vision-based SLAM using Rao-Blackwellised particle filter", presented at the Int. Joint Conf. on Artificial Intelligence workshop on Reasoning with Uncertainty in Robotics, Edinburgh, Scotland, 2005.

[21] J. M. Saez, F. Escolano, "Entropy minimization SLAM using stereo vision", in *Proc. IEEE Int.Conf. on Robotics and Automation*, Barcelona, Spain, 2005, pp. 36–43.

[22] M. N. Dailey, M. Parnichkun, "Landmark-based simultaneous localization and mapping with stereo vision", in *Proc. IEEE Asian Conf. on Industrial Automation and Robotics*, Bangkok, Thailand, 2005, pp. 108–113.

[23] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors", in *Proc. of IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Madison, USA, pp. 2003, pp. 257-263.

[24] G. Dissanayake, S. B. Williams, H. Durrant-Whyte, T. Bailey, "Map management for efficient simultaneous localization and mapping (SLAM)", *Autonomous Robots*, vol. 12, pp. 267–286, 2002.