The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

# A Pure Vision-based Approach to Topological SLAM

Wen Lik Dennis Lui and Ray Jarvis

*Abstract*— **This paper describes a topological SLAM system using a purely vision-based approach. This robot utilizes a GPU-based omnidirectional catadioptric stereovision system to perceive and plan its path in the environment. Subsequently, the omnidirectional images generated are used to incrementally build a database of image signatures based on the standard 2D Haar Wavelet decomposition. In order to maintain a globally consistent topological map, a relaxation algorithm, which requires local metric information between nodes, is employed each time the appearance-based localization system revisits an existing node in the topological map. The relative transformation of the current position of the robot with respect to the actual position of the matched node is recovered by using a least squares estimation of the transformation parameters of two 3D point patterns generated by the stereovision system. In addition, local metric information is obtained by using the proposed visual odometry system which combines distance measurements calculated by using optical flow techniques which estimates the movement of a web camera relative to the ground being observed and bearing estimates from the omnidirectional catadioptric vision system. Experiments were conducted in a variety of environments ranging from indoor to outdoor environments which demonstrate the feasibility of this approach.**

## I. INTRODUCTION

The rapid growth and declining costs of computers and cameras in recent years have made vision-based robots more practical and affordable. In addition, the introduction of the Nvidia CUDA libraries [24] (which allow the Graphics Processing Unit (GPU) on a computer to be used for general purposes), facilitate the implementation of more sophisticated and parallelizable computer vision algorithms to satisfy the real time constraint in robotics. Since humans primarily rely on visual information to perform day to day tasks and are capable of using this information to explore and navigate unknown environments at ease, vision systems on mobile robots have become a norm in an attempt to develop more intelligent and robust systems.

Generally, robots without a priori knowledge of the environment will be required to explore, build and maintain a globally consistent map by identifying and tracking distinctive features or landmarks in the environment or by comparing the similarities between the current and reference sensor data (scan matching, appearance-based). On the other hand, maps built by the robot can either be in the form of metric maps which represent spatial geometry data in fixed or dynamic resolutions, topological maps which represent the explored environment in terms of a linked collection of waypoints based on some distinctive abstract feature or a combination of both.

For landmark/feature dependent vision systems, distinctive edges such as Harris corners or distinctive landmarks such as SURF [3] or SIFT features [20], [28] are identified and tracked. As for appearance-based vision systems, a database of image signatures is created from the original images based on the principal components of the image [18], image histograms [30], Haar wavelet coefficients [10], [27] or Fourier coefficients [33]. Recently, a new paradigm known as the bag-of-visual words [2], [6] has become an increasingly popular technique for appearance-based vision systems. This paradigm normally uses a combination of visual cues extracted from the image and builds a visual dictionary. Of course, for a complete localization and mapping system, the abovementioned techniques will normally be coupled with wheel odometry, visual odometry or GPS information and serve as inputs to a probabilistic framework.

Appearance-based localization systems using image histograms [30] have the advantage of being rotation invariant while linear PCA [18], [12] is rotation, scale and translation invariant. However, both of these techniques are sensitive to lighting variations and will not operate robustly in a semi-outdoor or outdoor environment. On the other hand, the use of Haar wavelet coefficients [10], Fourier coefficients [33] or the bags-of-visual words paradigm [2], [6] were experimentally proven to be much more robust to lighting variations and occlusions as compared to PCA or image histograms.

In this paper, an incremental appearance-based localization system based on the standard Haar wavelet decomposition is proposed. It was chosen over other methods due to it being algorithmically simple, efficient, scalable and yet robust. Although the underlying techniques used to create the image signature is similar to that described in [10], additional insight concerning the effects of image signature size with respect to matching accuracy is provided and the appearance-based system is tested in indoor, semi-outdoor and outdoor environments extensively. In addition, the system described in [10] is provided with a priori knowledge of the environment (extensive 3D model built using a laser scanner) and the database of image signatures was generated using synthetic images produced by using this 3D model, whereas the system described in this paper requires the robot to explore the environment and incrementally build and maintain the image database. Since a topological map structure is highly suitable for an appearance-based localization system, a relaxation algorithm proposed by Duckett et al. [7] is employed to maintain a globally consistent topological map. However, it

Wen Lik Dennis Lui and Ray Jarvis are with the Intelligent Robotics Research Centre, Department of Electrical and Computer Systems Engineering, Monash University, Clayton Campus, Australia dwllui@gmail.com, ray.jarvis@eng.monash.edu.au

will require local metric information between nodes to be available and the ability of the system to calculate the relative transformation of the current location of the robot with respect to the location of the matched node in the topological map. The matched node is provided by the appearance-based localization system when the robot revisits a previously explored location and the relative transformation is required since it is highly unlikely that the robot will be precisely at the originally observed position of the reference/matched node.

Topological SLAM, specifically the combination of an image retrieval/place recognition system and topological mapping with metric information, is a well studied area [2], [23], [11]. Nevertheless, there are major differences between the underlying methods used in the proposed system to perform topological SLAM. For example, Haar wavelet is used for the image retrieval system as compared to the bag-of-visual words in [2] and radial features in [23], [11], an innovative visual odometry system is used instead of wheel odometry in [2] and data from an omndirectional stereovision system is used to recover the relative position between views via a least squares estimation technique instead of the 1D trifocal tensor in [23], [11]. There are also major differences in the localization framework and in the overall algorithm of the topological SLAM system.

The rest of this paper is organized as follows: In Section II, an overview of the system is provided. This is then followed by the description of the innovative visual odometry technique in Section III which provides the local metric information between nodes required to build a consistent topological map. Subsequently, the appearance-based localization and mapping system is detailed in Section IV which includes the exploration strategy and description of the technique used to calculate the relative transformation of the current location of the robot with respect to the location of the matched node using 3D information returned by the omnidirectional stereovision system. Experimental results are presented in Section V followed by a brief discussion in Section VI. Finally, conclusions are presented in Section VII.

## II. System Overview

The components of the robot are as shown in Fig. 1. A differential drive wheelchair motor/gear set is used to power the main research platform. It is equipped with a variable multibaseline omnidirectional catadioptric (mirror and camera combination) stereovision system whereby each individual catadioptric system is made up of a Canon Powershot S3 IS camera looking vertically upwards to an equiangular mirror designed by Chahl and Srinivasan [5]. The camera is capable of producing still images at 6MP resolution and a live video stream at 30Hz with a resolution of 320 x 240. Unfortunately, the epipolar geometry could not be derived for this mirror and camera combination since it does not possess the attractive single effective viewpoint property in central catadioptric systems. However, by vertically stacking of two catadioptric systems on top of one another as illustrated in Fig. 1, the search for the corresponding epipolar line becomes

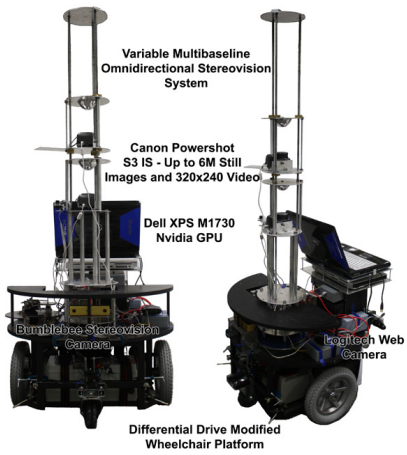



Fig. 1. The Eye Full Tower

a trivial task. Of course, a multi-camera rig such as [16] provides the flexibility in sensor placement and the stronger localization constraints provided by omnidirectional sensors. However, it is harder to setup (e.g. synchronization and calibration), may result in too much or no overlap between camera views and multiple optical centers although it has a much higher effective image resolution. On the other hand, fixed omndirectional multi-camera rigs such as the Pointgrey Ladybug alleviates this problem but generally comes with a heftier price tag.

This system was described in a previous work which can be found in [21], [22], which includes detailed explanation on the techniques used to establish stereo correspondences from a single stereo pair using local area-based matching techniques and extract 3D information using the proposed camera calibration technique specifically for equiangular mirrors on a GPU. In addition, it includes description of the multibaseline stereovision system and the technique to the automatic selection of baseline(s). Last but not least, the robot is equipped with a Bumbleebee [25] stereovision system for real time reactive obstacle avoidance and a Logitech web camera estimating its motion by observing the ground surface.

## III. Visual Odometry

Visual odometry can be achieved in many ways. In literature, it can be achieved by means of Structure-from-Motion (SfM) techniques [29], optical flow techniques [8], [17], [4], or by combining any of those with a GPS [1] for more robust tracking of the current position of the robot in outdoor environments. Generally, this involves the initialization of certain feature points in the image and tracking it through successive frames of the image sequence whilst detecting new points of interest as it progresses.

Our system performs real time visual odometry by combining the estimated distance calculated by using a pseudo optical flow algorithm described in [8] with bearing estimates obtained by using an appearance-based method for

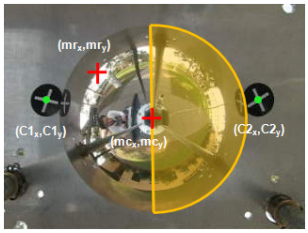



Fig. 2. FOV Utilized for Appearance-based Bearing Estimate Technique

omnidirectional vision systems described in [19]. A standard Logitech web camera is mounted on the rear of the robot such that its field of view (FOV) covers mostly the ground surface with its image plane parallel to the ground plane in order to provide distance travelled estimates. Although bearing estimates can be calculated as well, it is not as robust compared to the bearing estimates provided by the appearance-based method. Since it is very important to obtain accurate bearing estimates to reduce accumulated errors in visual odometry, the bearing estimates from the appearance-based method are used instead. Although the proposed visual odometry algorithm is based on the combination of [8] and [19], a number of modifications are made.

The following lists the main modifications made to the algorithm.

For distance travelled estimates,

- KLT good features to track are used instead of calculation of Sum of Absolute Differences (SAD) over local regions.

For bearing estimates,

- The front $180^o$ FOV of the robot in the omnidirectional image is used instead of the combined $60^o$ FOV of the front and back of the robot. This modification is made due to prolonged periods of my presence in that region while monitoring the robot.
- SAD instead of Euclidean distance is used to reduce computational requirements.
- Tracking of ($c1_x$,$c1_y$) and ($c2_x$,$c2_y$) is used to compensate for movements of the omnidirectional system due to vibration while robot is moving.

The final algorithm is summarized in Algorithm 1. A video demo of the modified bearing estimate technique in an outdoor environment can be found at

`http://www.youtube.com/watch?v=k6Qu98TrKQI`

**Algorithm 1** Visual Odometry - Fusion of Optical Flow and Appearance based Techniques

**Distance Travelled Estimates**

**Input:** Features initialized and tracked by Kanade-Lucas-Tomasi (KLT) feature tracker available in OpenCV.

1: **for** every two image point pairs returned by KLT **do**
2: Calculate the translation motion vector using pseudo optical flow algorithm in [8]
3: Filter and exclude vectors that are not achievable by the robot (i.e. amount of translation per frame)
4: **end for**
5: Average the resultant translation motion vectors

**Bearing Estimates**

**Input:** Coordinates of mirror centre ($mc_x$,$mc_y$), mirror rim ($mr_x$,$mr_y$) and centre of the two crosses ($c1_x$,$c1_y$) and ($c2_x$,$c2_y$) in the omnidirectional image shown in Fig. 2 (manually initialized)

**Parameters:** $T_A$ - Normalized amplitude threshold

**Require:** Reference image $I_{ref}$ = $I_{t=0}$ (image at time = 0)

1: **for** t=1 to $\infty$ **do**
2: Track and update($c1_x$,$c1_y$) and ($c2_x$,$c2_y$) using KLT and average the differences between the current and previous coordinates in the x and y directions
3: Use average differences to update ($mc_x$,$mc_y$) and ($mr_x$,$mr_y$)
4: Unwarp $I_{ref}$ and $I_t$ using the coordinates ($mc_x$,$mc_y$) and ($mr_x$,$mr_y$)
5: **for** i=0 to width of $I_t$ **do**
6: Column-wise shift the unwarped image of $I_t$
7: Compute SAD (for the front $180^o$ FOV of the robot) between unwarped $I_t$ with unwarped $I_{ref}$ and store score into array
8: **end for**
9: Find minimum score using interpolation/extrapolation
10: Calculate normalized amplitude, $A_n$, described in [19]
11: **if** $A_n < T_A$ **then**
12: $I_{ref}$ = $I_t$
13: **end if**
14: **end for**

Finally combine distance and bearing estimates to track the position of the robot

## IV. Appearance-based Localization and Mapping

### A. Image Retrieval using Haar Wavelets

The image retrieval system is based on [10] and is originally proposed by Jacobs et al. [13]. Although the Haar wavelet has been successfully applied in many different applications, it has not been used extensively for the localization of mobile robots. Ho and Jarvis [10] have adapted this algorithmically simple, efficient and yet robust framework into mobile robot localization and illustrated its robustness against lighting variation and occlusion in a semi-outdoor environment. In the original system [13], RGB images are converted into YIQ color space, decomposed using the standard 2D Haar decomposition technique, and the top 60 coefficients (quantized magnitudes and locations) are retained as the image signature. Subsequently, whenever a query image is presented to the system, it will be decomposed, quantized and a weighted score (depending on the location of the coefficient) is calculated.

Ho and Jarvis [10] adapted this for panoramic images by downsampling the original unwarped image to a size of 512 x 128 and retaining the coefficients within a bounding box of size 64 x 16 originating from the (0,0) coordinate of the decomposed image. The magnitude of these coefficients are

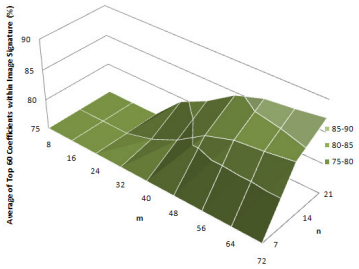



Fig. 3. Average of top 60 coefficients within bounding box of size $m$ by $n$

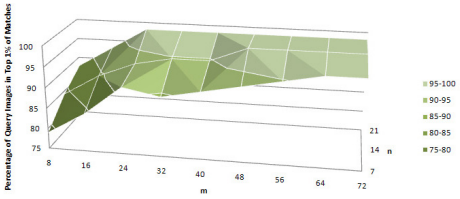



Fig. 4. Average top 1% matches using bounding box size $m$ by $n$

quantized and conveniently stored into a bit array, which significantly reduces the memory footprint for each image signature (location of coefficient is not required). Since the Haar wavelets are rotation variant, the unwarped panoramic image are column-wise shifted every 10 degrees equivalent in pixels and decomposed, quantized and stored in the database. In Fig. 3, a total of 335 panoramic images were used to find the average number of top 60 coefficients within the bounding box of size $m$ by $n$ and Fig. 4 shows the effect of matching accuracy with different bounding box sizes using 57 query images for a database with 2052 image signatures. With these quantitative results, an image signature of size 56 x 14 was chosen instead which contains an average of 82.8% of the top 60 coefficients and performing at an average of 98.2% to rank the correct image signature in the database in the top 1% of all returned matches for the 57 query images.

### B. Exploration Strategy

The robot has no a priori information of the environment and each node in the evolving topological map contains information of its global 2D location, heading and an estimation variance (initialized with variance of previous node plus 6% of distance travelled). As such, the initial position of the robot is assumed to be the origin of the global coordinate system with the current heading of the robot initialized as $0^o$ and with a variance of 0. A pair of still images is taken using the omnidirectional stereovision system at this initial position and the system returns 3D coordinates of all correspondences established with successive images taken at $T_D$ intervals (tracked by visual odometry). The 3D point clouds are then voxelized and clipped before being compressed into a 2D local grid map. Regions of the 2D local grid map are then

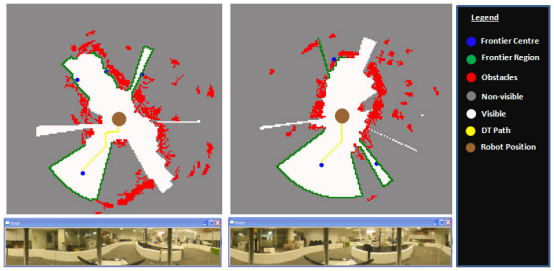



Fig. 5. Local 2d grid maps with Frontier regions and centres and segmented regions with Distance Transform path

segmented into visible, non-visible or obstacle regions by ray tracing from the centre of the robot in all directions. As long as this ray is not blocked by an obstacle, all grids traversed by this ray will be labeled as being visible to robot. Once it is terminated by an obstacle, any other grid location that lies on the same direction of this ray, which will eventually be located radially further away from this obstacle, is labeled non-visible to the robot or remain as being an obstacle if it is originally labeled as an obstacle.

Similar to frontier region detection [32] which detects regions between explored and unexplored areas, our system detect frontier regions between visible and non-visible areas and finds the centre of these detected regions. Obstacles will then be dilated to create a safety margin between the perceived obstacles and the robot. Assuming that these frontiers are starting positions for a 2D Distance Transform (DT) [15] algorithm, the position of the robot being the goal position, DT paths are planned (planned paths will be reversed once a path has been decided). In fact, the DT map can be calculated just by knowing the goal position (current position of the robot). Subsequently, with the starting positions initialized, the DT paths are traced and the selected path is reversed (in effect making the current goal position as the starting position and the starting position as the goal position). The reason for this is due to it being algorithmically simpler to have a single goal position with multiple starting positions.

Based on the topological map built so far, the robot will then decide which frontier region has not been explored. Since there might be more than one possible frontier for exploration and the robot can only go to one at a time, the robot will store this information into the current node of the topological map so that it can return to this node and explore if required. If all regions for the current position of the robot have been explored, the robot will use a nodal propagation technique described in [14] to return to the closest node which have been registered to have unexplored regions. Otherwise, it will return to its initial position. Fig. 5 illustrates the planned DT path, frontier regions and segmented regions.

### C. Loop Closure Detection

The key to maintaining a globally consistent topological map using the relaxation algorithm in [29] is to detect loop closure/previously visited nodes. The current method to detect this is to discriminate the matches based on the

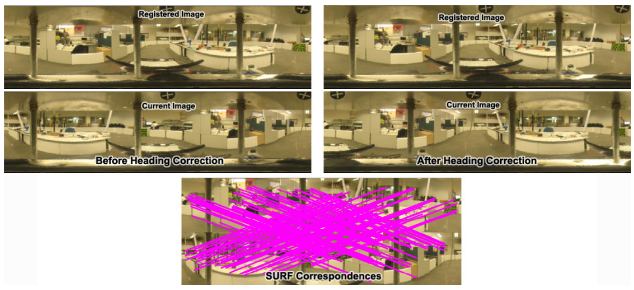



Fig. 6. Recovering relative heading and distance using SURF correspondence during loop closure

returned weighted score and the current position of the robot. The top S matches that are smaller than a threshold $T_R$ and located within the boundaries of a circle centred on the current position of the robot with a radius defined by the current variance are considered as possible candidates. For each candidate, there are a pair of omnidirectional stereo images registered with it. As such, 3D information can be retrieved from these images. SURF correspondences will then be established between the bottom image of the current omnidirectional stereo pair with the bottom image of the omnidirectional stereo pair registered to the candidate. These correspondences will be associated to the 3D information obtained from the stereo images. Using a RANSAC [9] procedure and a least squares estimation of transformation parameters for two 3D point patterns proposed by Umeyama [31], the best transformation matrix is obtained. Relative translation vectors from the current position of the robot with respect to the position of the matched node is recovered using this procedure and relative bearing can be robustly recovered using the average difference between the horizontal position of the SURF correspondences in the unwarped panoramic image. The main reason for not using the returned transformation matrix to obtain relative bearing information is that, even when stereo data is noisy, the system can still recover accurate relative bearing information and will be able to drive the robot in the correct direction . The final algorithm is summarized in Algorithm 2 and Fig. 6 shows SURF correspondences established between the unwarped panoramic images.

**Algorithm 2** Loop Closure Detection

**Input:**

Current stereo pair $I_b$ and $I_t$

Top S weighted scores w[$i_0$...$i_S$]

**Define:**

Stereo() - function that takes a stereo pair and returns 3D coordinates using stereovision techniques

TMatrix() - function that takes two 3D point patterns and returns the transformation matrix

**Parameters:**

$T_R$ - weighted score threshold

S - number of top matches to consider

maxIter - max iterations for RANSAC procedure

$T_D$ - Euclidean distance error threshold

C1 = Stereo($I_b$, $I_t$)
**for** i=0 to S **do**
  **if** w[i] $< T_R$ **then**
    Load the reference stereo pair $R_b^i$ and $R_t^i$
    C2 = Stereo($R_b^i$, $R_t^i$)
    Establish SURF correspondences between $I_b$ and $R_b^i$
    Associate SURF correspondences with 3D points in C1 and C2 and store in S1 and S2
    Calculate relative rotation $\theta$ using horizontal positions of SURF correspondences
    **for** j=0 to maxIter **do**
      Randomly select two corresponding 3D point patterns from S1 and S2 and store in P1 and P2
      h = TMatrix(P1,P2)
      $S1^{'}$ = h $\times$ S1
      Find number of points in $S1^{'}$ that fits its corresponding point in S2 within defined error threshold $T_D$
      Keep the best transformation matrix in M[i]
    **end for**
  **end if**
**end for**
Find Euclidean distance using translation vectors of the best transformation matrices in M
Select the node within shortest range
**return** Euclidean distance from selected node and relative bearing $\theta$

## V. Experimental Results

The appearance-based localization system was tested extensively in a semi-outdoor and outdoor environment. A weighting scheme similar to the one described in [13] is used and weights are trained using logistic regression on a set of training images independent of the following experimental data. Fig. 7 and 8 show the locations and some sample query and database images that were manually collected (approximately 1.2m apart for neighboring locations) on different days and times. Since Haar wavelets are rotation variant, the unwarped panoramic images are column-wise shifted for every 10 degrees equivalent in pixels and a total of 36 images used to represent any single location if it were to be included into the database. For the outdoor experiment, a total of 202 images were collected for each query and database set. Since each location will generate 36 images, the database will consist of 7272 images in total. For the semi-outdoor environment, query and database images for 61 locations were collected. To find out whether the system will still work if an offset is introduced since it is highly unlikely for the robot to return to the exact location of the nodes in the database, another set of 202 and 61 query images for the outdoor and semi-outdoor environment with an offset of 0.6m from its original location had been collected on different days and time. Table I summarizes the results (Top 1%, 3 and 5 refers to the correct image being ranked in the top 1%, 3 and 5 of all images in the database).

The proposed visual odometry system has also been extensively tested in an indoor lab environment covered

Fig. 7. Semi-outdoor environment where images are taken with sample query and database images

Fig. 8. Outdoor locations where images are taken with sample query and database images (red dots are locations where the system consistently fails)

with carpet and semi-outdoor environment with concrete slab paving. The robot was manually driven around the indoor lab environment for a total of 59 times and ground truth measured accurately by strategically placing markers which are picked up by the Logitech web camera. Similarly, the robot was driven in the semi-outdoor environment for 22 times and ground truth measured accurately by picking up the intersections of the concrete slab paving. In both environments, the robot was driven in a loop fashion from location L1 to L18 and then closing the loop at L19. The average drift (average distance traveled was 16.95m) for the indoor experiments before loop closing was 5.58% and dropped to 3.34% after loop closing with an average distance estimate error of 0.044 $\pm$ 0.06m and an average heading error of 0.713 $\pm$ $4^o$. For the semi-outdoor experiments, the average drift (average distance travelled was 20.17m) before loop before loop closing was 5.64% and dropped to 4.2% after loop closing with an average distance estimate error of 0.0147 $\pm$ 0.097m and an average heading error of 0.913 $\pm$ $4.9^o$. For more details, please refer to Fig. 9-15.

Image sequences from the vision systems required to perform visual odometry were saved on disk and 10 runs from each environment were compiled into the following videos,

Indoor - `http://www.youtube.com/watch?v=r8JKCSc5__g`
Semi-Out. - `http://www.youtube.com/watch?v=Y_rXRWD7eOI`

A video demo showing the robot autonomously performing topological SLAM in an indoor lab environment using the proposed visual odometry and appearance-based localization and mapping system combination is available at `http://www.youtube.com/watch?v=z077PZpPjnI` .The robot stops at intervals of 1m to capture digital still images from the omnidirectional stereovision system, performs path planning and builds a topological map associated to a database of image signatures. As the robot executes its planned path, it performs real time visual odometry using the system described in Section III and turns on the reactive obstacle avoidance system which performs a naive analysis of the disparity maps returned from the Bumblebee. The final topological map is illustrated in Fig. 16.

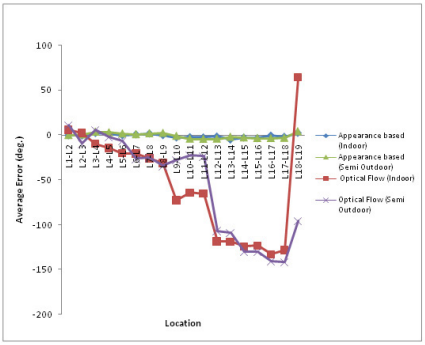



Fig. 9. Heading estimate error

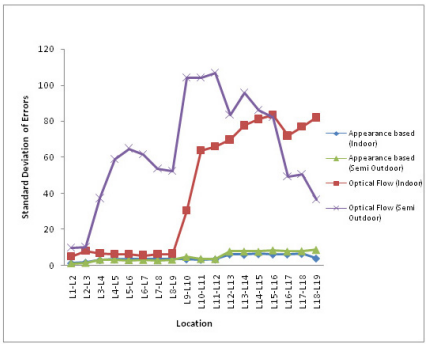



Fig. 10. Average standard deviation of heading errors

TABLE I
IMAGE RETRIEVAL MATCHING ACCURACY

| Dataset | Query Size | DB Size | Top 1% | Top 3 | Top 5 |
|---|---|---|---|---|---|
| Out | 202 | 7272 | 100 | 96.037 | - |
| Semi | 61 | 2196 | 100 | 100 | - |
| Both | 263 | 9468 | 100 | 96.958 | - |
| Out(Off) | 202 | 9468 | 100 | 89.552 | 95.532 |
| Semi(Off) | 202 | 9468 | 100 | 100 | 100 |

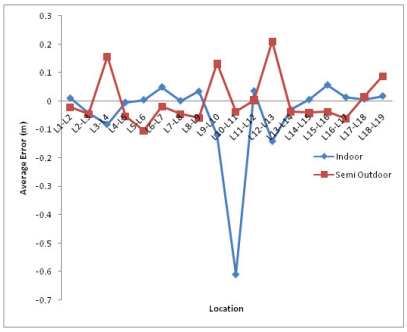



Fig. 11. Distance estimate error

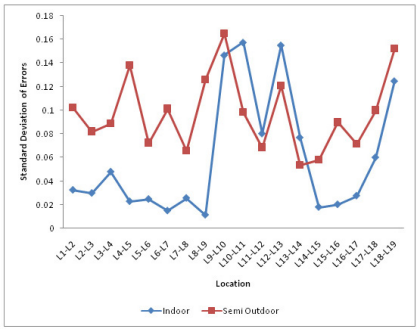



Fig. 12. Average standard deviation of distance errors

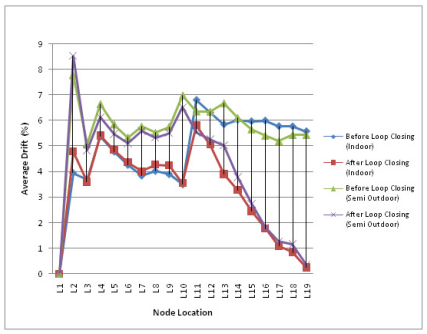
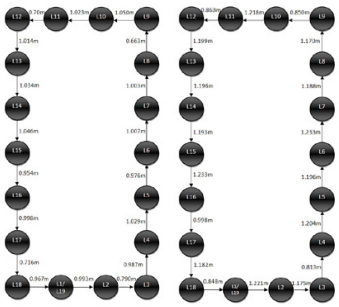
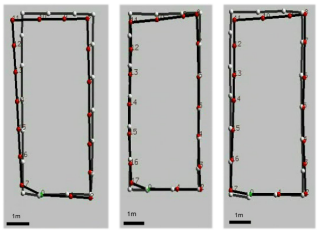



Fig. 13. Average Drift



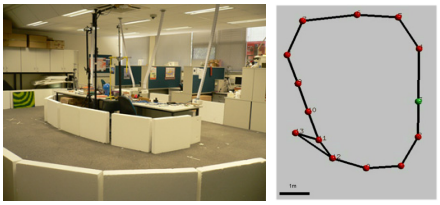
(a) Indoor Loop (b) Semi Outdoor Loop

Fig. 14. Distance between nodes for visual odometry tests



(a) Test 1 (b) Test 2 (c) Test 3

Fig. 15. Semi Outdoor Tests for Visual Odometry - Ground Truth (White Nodes), Robot's estimated trajectory (Red Nodes)



(a) Indoor Lab (b) Indoor Topological Map

Fig. 16. Indoor Experiment

## VI. DISCUSSION

From the experimental results, it can be seen that the image retrieval system is very robust with respect to lighting variation and occlusion as illustrated in the sample outdoor images where lighting variation is severe and objects such as cars, which are previously present, can be replaced by a different car or not be present in that space anymore. The experiment with the offset image dataset further reveals that the system can reliably localize itself even if it is offset by a certain distance from its original location (0.6m in this case) and yet being able to accurately differentiate between two locations separated by 1.2m. The effect on matching accuracy for different image signature sizes with respect to the percentage of top 60 coefficients within this bounding box has also been illustrated. In addition, although weights are trained using a totally independent dataset, it was reliably used to produce high matching accuracy for the semi-outdoor and outdoor datasets.

The proposed visual odometry system which combines optical flow and appearance based techniques, also yielded satisfactory results and was shown to work reliably in indoor and semi-outdoor environments. Although not tested in an outdoor environment, it is expected that this system will still work as long as the optical flows are not corrupted by moving shadows due to movement of trees, bushes or humans and when lighting conditions are not in either extremes (too dark or too bright) over a prolonged period of time. This is also assuming that perceptual aliasing is not too severe to affect the appearance-based bearing estimate technique and it is a fair assumption to make in general that, outdoor environments will have less problems with aliasing and more features and texture will be present relative to indoor and semi-outdoor environments.

The video demo showing the robot autonomously explor-

ing and mapping an indoor environment also shows that the exploration strategy is capable of directing the robot to unexplored environments using 3D information obtained from the omnidirectional stereovision system and autonomously recognize a previously visited node using the appearance-based localization system. Although the omnidirectional stereovision system can combine multiple stereo pairs together and is also equipped with an automatic baseline selection system, the baseline is fixed for these experiments due to it being in an indoor environment. However, for a larger semi-outdoor or outdoor environment, a loop closing mechanism must be included into the exploration strategy such that the robot can keep the errors bounded by closing the loop once the robot has traveled for a considerable amount of distance without revisiting nodes with low variances. In addition, a probabilistic framework suitable for the appearance-based localization system will be implemented in future based on the occurrences and scores of the matches which will subsequently lead to research in map-merging problems for large scale environments by combining structural information of the various individual topological maps.

## VII. CONCLUSIONS

Visual odometry is performed by combining distance information using optical flow techniques with bearing information from an appearance-based technique in order to overcome the shortcomings of the reliability of bearing estimates from a non-differential optical flow configuration. A GPU-based omnidirectional stereovision system was also successfully used to allow the robot to plan and explore an indoor environment. The combination of the abovementioned together with an appearance-based localization and mapping system equipped with loop closing detection and a relaxation algorithm has prove the feasibility of a purely vision-based mobile robot for topological SLAM.

## REFERENCES


[1] M. Agrawal and K. Konolige, "Rough Terrain Visual Odometry", *in Proceedings of the International Conference on Advanced Robotics*, 2007.

[2] A. Angeli, S. Doncieux, J.-A. Meyer and D. Filliat, "Visual Topological SLAM and Global Localization", *in IEEE International Conference on Robotics and Automation*, 2009.

[3] H. Bay, A. Ess, T. Tuytelaars, L Van Gool, "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding*, vol. 110, no. 3, 2008, pp 346-359.

[4] J. Campbell, R. Sukthankar, I. Nourbakhsh and A. Pahwa, "A Robust Visual Odometry and Precipice Detection System Using Consumer-grade Monocular Vision", *in IEEE International Conference on Robotics and Automation*, 2005.

[5] J.S. Chahl and M. Srinivasan, "Reflective Surfaces for Panoramic Imaging", *Applied Optics*, vol. 36, no. 31, 1997, pp 8275-8285.

[6] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance", *International Journal of Robotics Research*, vol. 27, no. 6, 2008, pp 647-665.

[7] T. Duckett, S. Marsland and J. Shapiro, "Learning Globally Consistent Maps by Relaxation", *in IEEE International Conference on Robotics and Automation*, 2000, pp 3841-3846.

[8] D. Fernandez and A. Price, "Visual Odometry for An Outdoor Mobile Robot", *in Proceedings of the 2004 IEEE Conference on Robotics, Automation and Mechatronics*, 2004, pp 816-821.

[9] M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of ACM*, vol. 24, no. 6, 1981, pp 381-395.

[10] N. Ho and R. Jarvis, "Vision Based Global Localisation Using a 3D Environmental Model Created by a Laser Range Scanner", *in IEEE International Conference on Intelligent Robots and Systems*, 2008, pp 2964-2969.

[11] J.J. Guerrero, A.C. Murillo and C. Sagüés, "Localization and Matching using Trifocal Tensor with Bearing-only Data", *IEEE Transactions on Robotics*, vol. 24, no. 2, 2008, pp 494-501.

[12] P.-C. Hsieh and P.-C. Tung, "A Novel Hybrid Approach Based on Sub-Pattern Technique and Whitened PCA for Face Recognition", *Pattern Recognition*, vol. 42, no. 5, 2009, pp. 978-984.

[13] C.E. Jacobs, A. Finkelstein, D.H. Salesin, "Fast Multiresolution Image Querying", *Computer Graphics*, 29 (Annual Conference Series), 1995, pp 277-286.

[14] R. Jarvis, "Optimal Pathways for Road Vehicle Navigation", *in IEEE TENCON*, vol. 2, 1992, pp 876-880.

[15] R. Jarvis, "Distance Transform Based Path Planning for Robot Navigation", *Recent Trends in Mobile Robots*, vol.11, 1993, Robotics and Intelligent Systems, ch. 1.

[16] M. Kaess and F. Dellaert, "Probabilistic Structure Matching for Visual SLAM with a Multi-Camera Rig", *Computer Vision and Image Understanding*, vol. 114, no. 2, 2010, pp 286-296.

[17] J. Kim and G. Brambley, "Dual Optic-flow Integrated Inertial Navigation", *Australasian Conference on Robotics and Automation*, 2005.

[18] B. Kröse, R. Bunschoten, N. Vlassis and Y. Motomura, "Appearance based Robot Localization", *in IJCAI 99 Workshop: Adaptive Spatial Representations of Dynamic Environments*, 1999, pp 53-58.

[19] F. Labrosse, "The Visual Compass Performance and Limitations of an Appearance Based Method", *Journal of Field Robotics*, vol. 23, no. 10, 2006, pp 913-941.

[20] D.G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, 2005, pp 91-110.

[21] W.L.D. Lui and R. Jarvis, "Eye-Full Tower: A GPU-based Variable Multibaseline Omnidirectional Stereovision System with Automatic Baseline Selection for Outdoor Mobile Robot Navigation", *Robotics and Autonomous Systems*, vol. 58, no. 6, 2010, pp 747-761.

[22] W.L.D. Lui and R. Jarvis, "Omnidirectional Vision System for Outdoor Mobile Robots", *Workshop on Omnidirectional Robot Vision (Workshop Proceedings of SIMPAR 2008)*, 2008, pp 273-284.

[23] A.C. Murillo, C. Sagüés, J.J. Guerrero, T. Goedemé, T. Tuytelaars and L. Van Gool, "From Omnidirectional Images to Hierarchical Localization", *Robotics and Autonomous Systems*, vol. 55, no. 5, 2007, pp 372-382.

[24] Nvidia Corporation - Nvidia CUDA, Online, 2009, http://www.nvidia.com/object/cuda_home.html

[25] Point Grey Research Inc. - Bumblebee Stereovision Camera, Online, 2009, http://www.ptgrey.com/products/stereo.asp.

[26] J.M. Porta, J.J. Verbeek and B. Kröse, "Active Appearance-based Robot Localization using Stereo Vision", *Autonomous Robots*, vol. 18, no. 1, pp 59-80.

[27] A. Pretto, E. Menegatti, E. Pagello, Y. Jitsukawa, R. Ueda and T. Arai, "Toward Image-based Localization for AIBO using Wavelet Transform", *in 10th Congress of the Italian Association for Artificial Intelligence*, 2007.

[28] S. Se, D.G. Lowe and J. Little, "Vision-based Mobile Robot Localization and Mapping using Scale-Invariant Features", *in IEEE International Conference on Robotics and Automation*, vol. 2, 2001, pp 2051-2058.

[29] M. Tomono, "3D Localization and Mapping using a Single Camera Based on Structure-from-Motion With Automatic Baseline Selection", *in IEEE International Conference on Robotics and Automation*, 2005, pp 3342-3347.

[30] I. Ulrich and I. Nourbakhsh, "Appearance-based Place Recognition For Topological Localization", *in IEEE International Conference on Robotics and Automation*, vol. 2, 2000, pp 1023-1029.

[31] S. Umeyama, "Least Squares Estimation of Transformation Parameters between Two Point Patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, 1991, pp 376-380.

[32] B. Yamauchi, "A Frontier-Based Approach for Autonomous Exploration", *in IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 1997, pp 146-151.

[33] A.M. Zhang and L. Kleeman, "Robust Appearance Based Visual Route Following for Navigation in Large-scale Outdoor Environments", *International Journal of Robotics Research*, vol. 28, no. 3, 2009, pp 331-356.