

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Robust Omnidirectional Vision based Mobile Robot Hierarchical Localization and Autonomous Navigation

^{1,2}Li Maohai, ¹Sun Lining, ²Huang Qingcheng, ²Cai Zesu and ²Piao Songhao

¹School of Mechanical and Electric Engineering, Soochow University, Suzhou 215021, China

²School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Abstract: Robust hierarchical method is studied for implementing indoor mobile robot localization and autonomous navigation with omnidirectional vision. The localization is purely vision-based and uses a sequence of prerecorded image frames as reference topological map. For finding the image which best fits the current image, the system consists of two hierarchical components: global coarse localization and local accurate localization. The global color histograms are used to confirm fast the coarse robot position, which is not accurate, whereas the more accurate position can be obtained slowly through local scale invariant interest keypoints, which can be robustly matched with nearest neighbor search based on KD-Tree. The navigation system is composed of on-line and off-line stages. During the off-line learning stage, the robot performs paths and records a set of ordered key images. From this sequence a topological map is built. Each topological node provides a set of omnidirectional images characterized by geometrical scale invariant keypoints. In the on-line navigation stage, the robot recognizes the most likely node through robust hierarchical algorithm and then controlled by a vision based control law adapted to omnidirectional cameras to follow the visual path. To evaluate this method, actual experiment is carried out in an indoor environment. Results show the robustness and efficiency of the proposed method.

Key words: Mobile robot, localization, navigation, visual features, topological map, experiment

INTRODUCTION

Localization is one of the fundamental problems of mobile robots. The knowledge about its position allows a mobile robot to efficiently fulfill different useful tasks. In the past, a variety of approaches for mobile robot localization has been developed. They mainly differ in the techniques used to represent the belief of the robot position and the type of sensor. In this study we consider the problem of vision-based mobile robot localization. This study is focused on the localization with an omnidirectional vision. Omnidirectional vision is useful in many robotic applications because it provides a large field-of-view of the environment (Luke and Stefan, 2004; Tamimi, 2006; Jogan and Leonardis, 2003; Barfoot, 2005).

Over the past years, several vision-based localization systems have been developed. They mainly differ in the extracted features. For example, Hayet *et al.* (2003) extracted and recognized planar quadrangular landmarks from images. Kosecka *et al.* (2003) employed gradient orientation histograms to capture the essential object appearance information. Sim and Dudek (1999) learnt

natural visual features and used Principal Component Analysis (PCA) to match landmarks for pose estimation. Basri and Rivlin (1995) extract lines and edges from images and use this information to assign a geometric model. Dudek and Zhang (1996) use a neural network to learn the position of the robot given a reference image. Paletta *et al.* (2001) consider trajectories in the Eigenspaces of features. Se *et al.* (2001) uses scale-invariant features to estimate the position of the robot within a small operational range. Furthermore, there are different approaches that use techniques also applied to identify the current position of the robot (Fengda *et al.*, 2008; Ekvall *et al.*, 2007; Bradley *et al.*, 2005; Siggelkow, 2002; Wolf *et al.*, 2005; Weiss *et al.*, 2007).

In general, the most popular localization approach is based on the Scale Invariant Feature Transform (SIFT) (Lowe, 2004), which is rotation invariant and relatively robust to illumination changes. However, the number of interest points per image is usually large, extracting the features and matching two images is a high computational cost. Another group of vision-based localization algorithms uses global image features. Well-known

methods for robot localization are based on PCA (Tamimi, 2006; Bradley *et al.*, 2005; Jogan and Leonardis, 2003) or on Integral Invariants (Siggelkow, 2002; Wolf *et al.*, 2005). Bradley *et al.* (2005) presented an approach that uses Weighted G radient Orientation Histograms. Although, the global methods are faster, the global features are also more sensitive to illumination changes than local features.

In this study, we propose a hierarchical approach that combines the advantages of both global and local image features. In a first step, the global feature descriptors based on color histogram allow to select only some possible images. And then the coarse localization results are taken as the candidates for the following accurate localization. Each candidate in the map database is matched with the current view using the SIFT which gives accurate results. The experimental results show that the best compromise between computational efficiency and accuracy is obtained with this proposed method.

Vision based mobile robot navigation falls into three broad groups: (1) Map based navigation-systems that depend on user created environment maps; (2) Map building based navigation-systems that use sensors to construct their own environment maps and then use these maps for navigation; (3) Mapless navigation-systems that use no explicit representation at all about the space in which navigation is to take place, but rather resort to recognizing objects found in the environment or to tracking those objects by generating motions based on visual observations. Popular choices for the environment maps constructed by robot itself include metrical mapping (Schultz and Adams, 1998) topological mapping (Choset and Nagatani, 2001) and a hybrid approach (Estrada *et al.*, 2005). Metric mapping can be easy to build and maintain, however, the problem of large-scale structural ambiguity can be introduced, data association only has to be handled locally and accepting false negative place matches (Bosse *et al.*, 2003). A topological map is a concise description of the large-scale structure of the environment and compactly describes the environment as a collection of places linked by paths Choset and Nagatani (2001) The navigation system proposed in this paper works with natural environments. That means that the environment does not have to be modified for navigation in any way. Indeed, adding artificial markers to every room in a house doesn't seem feasible nor desirable. We choose a topological representation of the environment, rather than a metrical one, because of its resemblance to the intuitive system humans use for navigation, its flexibility, wide usability, memory-efficiency and ease for map building and path planning. Moreover, keeping features in the field of view of an omnidirectional camera is easier than in the case of a perspective camera.

We primarily focus on investigating real-time, omnidirectional vision based autonomous navigation for indoor environments. The main contributions of this paper are:

- A method to construct a topological map, which is robust to self-similarities in the environment
- A robust hierarchical localization algorithm is proposed to localize robot in topological map
- A visual servoing algorithm which is robust to moving objects or environment changes
- The integration of all these components in an autonomous topological navigation system

SYSTEM OVERVIEW

A framework of the localization system is presented. The system can be divided in three stages: map building, global coarse localization and local accurate localization.

Map building: The map is built only once, to build the map for a new environment. The mobile system is lead through all parts of the environment by hand, simultaneously it records images and odometry value at a constant rate. Later, this large set of omnidirectional images is automatically analyzed and converted into a topological map of the environment.

Global coarse localization: When the system is started up somewhere in the environment, it captures a new image with its camera. This current view is compared with the environment map and a topological node area is located through the coarse localization algorithm.

Local accurate localization: When the current topological node of the robot is known, distinctive scale invariant features combined with nearest neighbor search based on KD-Tree to localize accurately robot.

ENVIRONMENT MAP

In our approach, the environment map is built hierarchically during the off-line learning stage as shown in Fig. 1. The fundamental building blocks of our map representation at the local level are the local SIFT features based local maps of different regions of the environment, guaranteed to be of limited size and mutually independent in the statistical sense at all times. The descriptive power of the SIFT descriptor can improve the quality of data association. Each SIFT descriptor may have similar appearance but are spatially distinct. The implementation

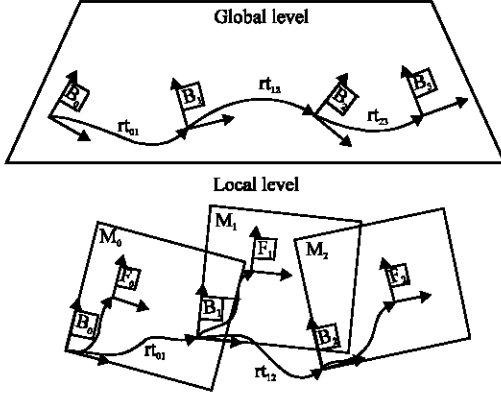


Fig. 1: Hierarchical map

of the map structure is based on KD-tree. All these features are with respect to a base reference B , which is the initial current robot location when the map is initialized. Furthermore, local maps are required to be mutually independent at all times for this representation to be consistent.

At the global level, the topology of the environment is represented by a graph in which each node i corresponds to the global features and a local map M_i of the local level. An arc rt_j in the graph represents a known topological relation between local maps M_i and M_j detected during the mapping process. These topological relations are the relative transformations between the base reference of both maps, $rt_{ij} = rt(B_i, B_j) = (x_{ij}, z_{ij}, \theta_{ij})$, which is calculated through the global minimization of loop closure alignment for odometry measurements.

Partitioning for the local map: For the images are acquired in sequence during the robot exploration and that the distance traveled between each image is not too great. In our approach, the following criterion for determining boundaries between the local maps is used: The images captured at the same part of the environment (a room, a corridor) that should be considered similar enough. The second condition is necessary, as the first condition can be satisfied even when someone is blocking the camera view. Another consideration is that the number of features in the current local map reaches a maximum, or the uncertainty of the current map reaches a limit. We make the decision to close the local map M_i and start a new map M_{i+1} whenever the above criterion is detected and M_{i+1} will have the current robot position as a base reference, which corresponds to the last robot location in M_i . This provides the initial relative transformation $rt_{i(i+1)}$, corresponding to the topological relation between the adjacent maps M_i and M_{i+1} , we will obtain the final robust relation through the global minimization of loop closure alignment.

Loop closure alignment: Each node of the map has an associated 2D position and orientation (x_i, z_i, θ_i) initialized to the robot odometry position when the node is created. A variance ω_i is also associated to each node and is initialized with the variance of the previous node. When a node is added or recognized in the map, a new edge is created. The relative metrical position of the two nodes obtained through the odometry measurements is memorized in this link $rt_{ij} = rt(B_i, B_j) = (x_{ij}, z_{ij}, \theta_{ij})$, a variance ω_{ij} is also associated to the edge.

After loop closure detection, as a consequence of the cumulative noise of odometry, the map is not alignment after loop closing. Thus, a relaxation algorithm is employed to estimate the position of each node that best satisfies the constraints imposed by the relative odometric information. As the maps only have a relatively small number of nodes and as the nodes have three freedom degrees. We use the iterative algorithm (Duckett *et al.*, 2000) to each node of the map:

- **Step 1:** Estimate the position of node i from each neighbouring node j :

$$\begin{aligned} (x'_i)_j &= x_j + r_{ij} \cos(\theta_{ji} + \theta_j) \\ (z'_i)_j &= z_j + r_{ij} \sin(\theta_{ji} + \theta_j) \\ (\theta'_i)_j &= \theta_j + \omega_{ij} \\ (\omega'_i)_j &= \omega_j + \omega_{ij} \end{aligned} \quad (1)$$

- **Step 2:** Estimate the variance of node i using harmonic mean of the estimates from the neighbours:

$$\omega_i = n_i / \sum_j \frac{1}{(\omega'_i)_j} \quad (2)$$

where, n_i is the number of neighbors of node i .

- **Step 3:** Estimate the position of the node as the mean of the estimates from its neighbours:

$$\begin{aligned} x_i &= \frac{1}{n_i} \sum_j \frac{(x'_i)_j \omega_i}{(\omega'_i)_j}, \quad z_i = \frac{1}{n_i} \sum_j \frac{(z'_i)_j \omega_i}{(\omega'_i)_j} \\ \theta_i &= \arctan \left(\frac{\sum_j \frac{\sin(\theta'_i)_j}{(\omega'_i)_j}}{\sum_j \frac{\cos(\theta'_i)_j}{(\omega'_i)_j}} \right) \end{aligned} \quad (3)$$

The above three steps are repeated until the total change in the nodes coordinates falls under a given threshold, or a maximum number of iterations is reached. The first node of the map is considered as the reference frame: its position is fixed at $[0, 0, 0]$ and its variance is

fixed at a small value. This algorithm was proven to converge and fast enough to be executed (Duckett *et al.*, 2000).

IMPLEMENTATION FOR OMNIDIRECTIONAL VISION

As shown in Fig. 2 our mobile robot system is built by ourself, Mounted upon this is an omnidirectional vision system. It contains a CCD camera pointed upwards looking at a spherical mirror, colorful images were captured at full resolution 640×480 pixels.

To be able to implement the robust localization, stable features are needed to be extracted from the image. The matching is done by comparing the features in the map with the features extracted from the current image. Ensure that the matching is invariant to image scale, rotation and translation, a good natural image features extraction and description method is of utmost importance. The hierarchical localization method based on visual appearance features proposed in this paper can finely solve this problem. In the higher level the global color histogram features are extracted, while the more distinctive SIFT features are extracted in the lower level.

Global image features: Color histograms have several attractive features especially for panoramic images. They represent the appearance of an image in a very compact manner and color histograms of panoramic images are invariant to rotations around the vertical axis. Thus, an image acquired at a particular position represents all images at this position with different orientation. In addition, color histograms usually vary smoothly as the field of vision sweeps the scene when a robot follows a path through the environment. Figure 3 shows two similar color histograms corresponding to two different panoramic images captured.



Fig. 2: (a, b) Robot omnidirectional vision system

The methods to evaluate the similarity of the two images are color histogram intersecting and Euclidean distance. Generally, high resolution graphics systems output their graphic images using red, green and blue primaries. But the HSI color space closely approximates the behavior of the human eye, so we use HIS space. Given the red, green and blue intensities of a pixel, the intensity I, the saturation S and the hue H is calculated directly as follows:

$$\begin{aligned} I &= (R + G + B)/3 \\ S &= \text{Max}(R,G,B) - \text{Min}(R,G,B) \\ H &= \cos^{-1} \left(\frac{(R - G) + (R - B)}{2\sqrt{(R - G)^2 + (R - B)(G - B)^2}} \right) \end{aligned} \quad (4)$$

In the HSI color space, the H plane stands for color information, so it is extracted from the current view to calculate 1D histogram, then the similarity implemented with Euclidean distance is evaluated. Given two color histogram H_1, H_2 , their Euclidean distance is calculated directly in following manner:

$$D(H_1, H_2) = \sqrt{\sum_{i=1}^m (h_1^i - h_2^i)^2} \quad (5)$$

Local image features: The local features extraction is based on SIFT. The features are invariant to image translation, scaling and rotation and partially invariant to illumination changes. These properties make SIFT very suitable for mobile robots.

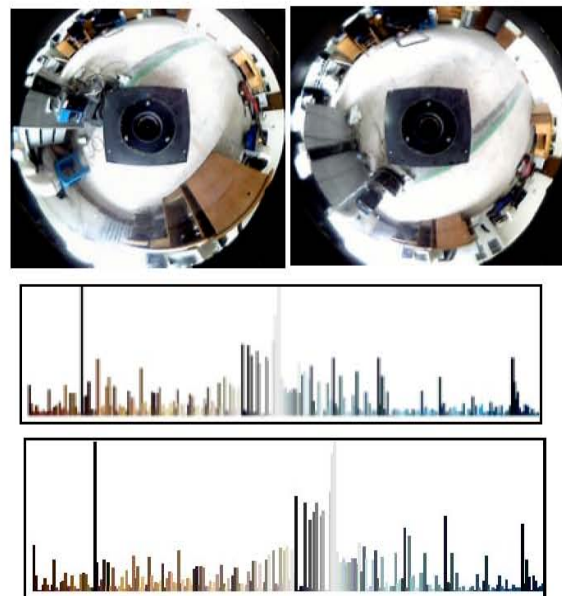


Fig. 3: Color histogram

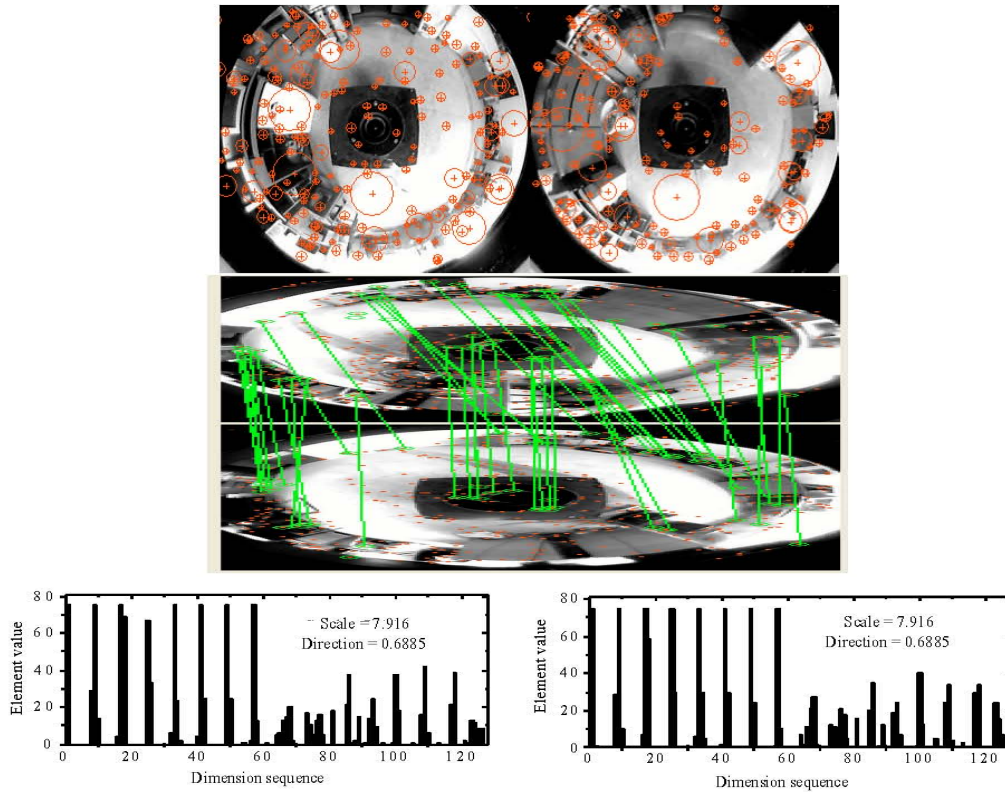


Fig. 4: SIFT feature: (a) Extraction, (b) Matching and (c) Descriptor histogram

The SIFT algorithm has four major stages: (1) Scale-space extrema detection-searches over scale space using a Difference of Gaussian function to identify potential interest points. (2) Key-point localization-location and scale of each candidate point is determined and key-points are selected based on measures of stability. (3) Orientation assignment-one or more orientations are assigned to each key-point based on local image gradients. (4) Key-point descriptor-a descriptor is generated for each key-point from local image gradients information at the scale found in stage 2.

However, the number of SIFT features generated is dependent on image size and content, we can obtain a lot of keypoints with expensive computation cost in a cluttered indoor scene, which can also aggravate the computation burden for feature matching. For the purpose of fast and efficient topological localization, we actually extract the landmarks extracted from the images with the modified Scale-Invariant Feature Transform. The images are obtained with an omnidirectional camera, which make a motion constraint for rotational invariance, yields a somewhat less complex feature extraction and more robust feature matching performance by eliminating of the

rotational normalization and rotational part of the descriptor. Matching of features in two images in then performed using squared Euclidean distance between the two corresponding histograms based on nearest neighbor search method with Kd-tree. An example of feature extraction and matching is shown in Fig. 4. Typical extracted SIFT features with their locations represented by '+'. The radius of the circle represents their scales. During the whole experiment, the matched accurate rate is higher than 80%, which can satisfy the localization need. Fig. 4c shows the key-point descriptor histograms of one matching pair, which illustrates the robust matching algorithm.

HIERARCHICAL LOCALIZATION

To estimate the current position of the robot, we propose a hierarchical framework to enhance the efficiency and accuracy of localization. The localization process includes two stages: coarse localization and accurate localization. The integration of coarse and fine stages realizes a fast and reliable localization system, which is shown in Fig. 5.

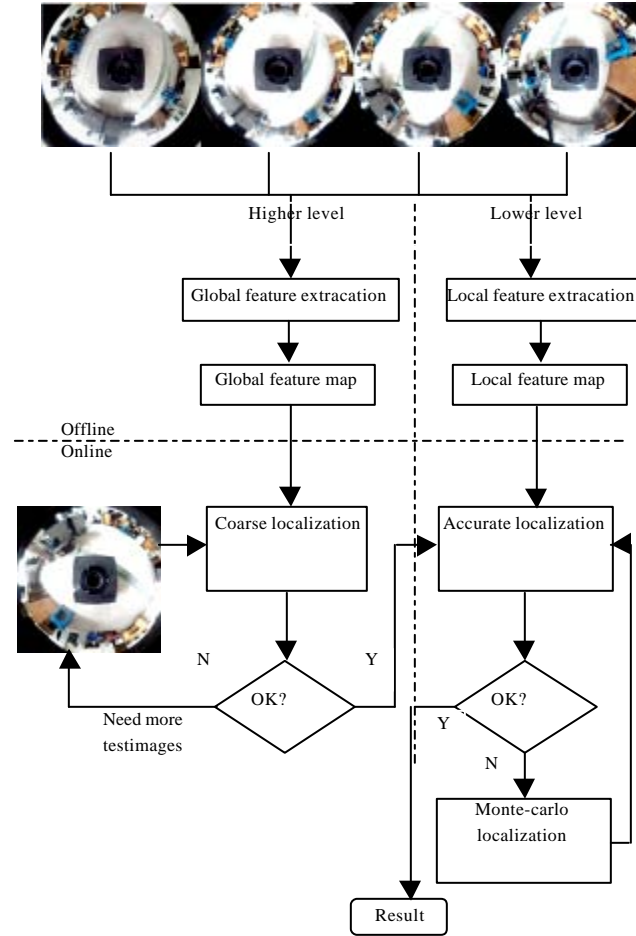


Fig. 5: Flowchart of the localization system

When a mobile robot moves in the environment, it obtains its coarse position by the coarse localization and these localization results are taken as the candidates for the following accurate localization. Each candidate in the map database is matched with the image for localization and the correct position is the one getting the largest number of matches. If the localization result is still ambiguous, epipolar geometry constraints and Monte-Carlo localization strategy are employed to verify the result.

VISUAL-SERVOING FOR NAVIGATION

Visual-servoing is often used to control robot follow the prerecorded trajectory and navigation can be considered as a sequence of visual-servoing tasks. The navigation problem is formulated as a path following to guide the nonholonomic mobile robot along the visual

path. This section describes a stabilization approach to control the robot motions from a key image to the next one.

The robot state consists of a pose vector $x = (x, y, \theta)$, describing its position and orientation. The navigation system can modify the robot's linear and angular velocities denoted by (v, ω) . The robot dynamic model is that of a wheeled unicycle mobile robot, with two degrees of freedom (linear and angular velocities):

$$\dot{x} = v \cos \theta, \quad \dot{y} = v \sin \theta, \quad \dot{\theta} = \omega \quad (6)$$

The tracking path P is defined as a set of points $xP = \{(x_0, y_0, \theta_0), \dots, (x_P, y_P, \theta_P)\}$, expressed in the same coordinate system and units as the robot state vector x . At each time instant, the path tracking module must determine a reference point on the trajectory, (x_p^{ref}, y_p^{ref}) which is then used to determine the position and orientation errors so as to correct the robot's motion:

$$(x_p^{ref}, y_p^{ref}) = \underset{(x_p^{ref}, y_p^{ref})}{\operatorname{argmin}} \left\{ \left\| (x_p^{ref}, y_p^{ref}) - (x, y) \right\|^2 \right\} \quad (7)$$

The dynamic controller used to generate the robot's angular velocity for path following:

$$\omega = -k_3 |v| \tilde{\theta} - k_2 v d \frac{\sin \tilde{\theta}}{\tilde{\theta}} + \frac{v \cos \tilde{\theta} c(s)}{1 - c(s)d}, \quad (8)$$

$$d = [x - x_p^{ref} \quad y - y_p^{ref}] [n_x \quad n_y]^T, \quad \tilde{\theta} = \theta - \theta_p^{ref}$$

where, d is distance to path error, $\tilde{\theta}$ is an orientation error and k_2, k_3 are constants to be tuned, s designates the path length and $c(s)$ is the local path curvature.

In order to tune the controller, usually set $k_2 = \alpha^2$, $k_3 = 2\epsilon\alpha$, where the damping coefficient ξ is set to $1/\sqrt{2}$ meaning a small overshoot and the natural frequency $\dot{\alpha}$ is left free to specify faster or slower systems. To describe more intuitively the control law of Eq. 8 it is interesting to analyze two particular cases of importance:

- When the reference path is a straight line, $c(s) = 0$, control reduces to the simple case of encompassing only two terms. One is proportional to $\tilde{\theta}$ and the other to d
- When the path is a circle, the curvature is a constant. If $\tilde{\theta}$ and d are close to zero, then the third term of the control law is constant which is according to the intuition that to describe a circle the robot must have a constant angular velocity

Mostly, the forward velocity v is equal to the maximum V_{max} but for safety reasons, we impose a maximum value on the angular velocity $|\omega| < W_{max}$. When this value is achieved, we saturate ω and reduce v to $V_{max} \cdot W_{max} / |\omega|$, in order to avoid large lags in narrow turns.

The curvature $c(s)$ is defined as $c(s) = d\theta/ds$. It is therefore important to have smooth trajectories as otherwise there will appear saturations on the control signal and very little ability of the controller to react to $\tilde{\theta}$ and d .

For the current control law, noise in self-localization measurements (x, y, θ) directly implies noise in control outputs (v, ω) . To prevent this direct noise transmission we include temporal integration of the measurements with an Extended Kalman Filter (EKF), whose inputs and outputs are then poses (x, y, θ) and the dynamics are that of a wheeled mobile robot, Eq. 8 with state vector augmented with velocities $x = (x, y, \theta, v, \omega)$.

EXPERIMENTAL RESULTS

The experiments are performed on a mobile robot system incorporating a 3.0 GHz Intel Pentium processor as shown in Fig. 1, which also provides motor control and the main processing power for vision processing and the navigation software. An omnidirectional color CCD camera mounted at the top of the robot is used for detecting the landmarks. The test environment is a robot laboratory composed of four rooms R_1, R_2, R_3, R_4 and a horizontal corridor, the model of which is shown in Fig. 6a, where the lower right room R_1 is shown in Fig. 6b. The experiments described here are designed to prove the performance of our method.

We firstly drive the robot to move along the path as shown in Fig. 6a, while the robot is moving ahead, the image frames are captured and recorded, from this sequence the topological map of this environment is built. Figure 7 shows partial image frames. As shown in Fig. 8, the points in the map denotes bird's-eye view of the landmarks, which represent occupied regions corresponding to objects such as chairs, shelves, desks and computers in the scene.

We use this map for the localization experiments. We choose 11 locations to test the localization performance as

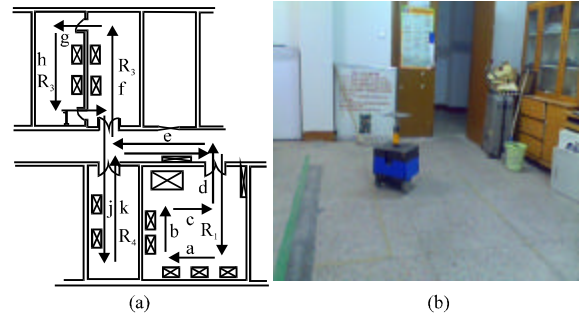


Fig. 6: (a) the environment model and (b) the room R_1

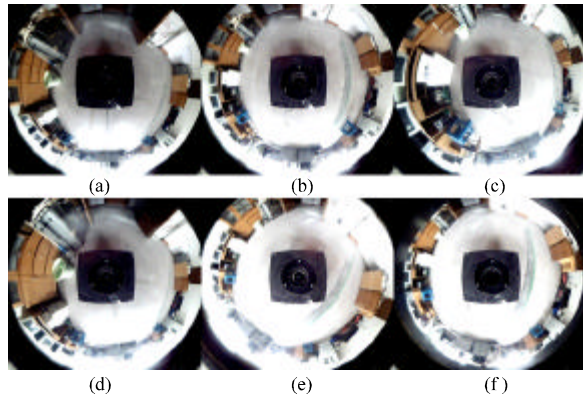


Fig. 7: (a-f) Partial image frames captured

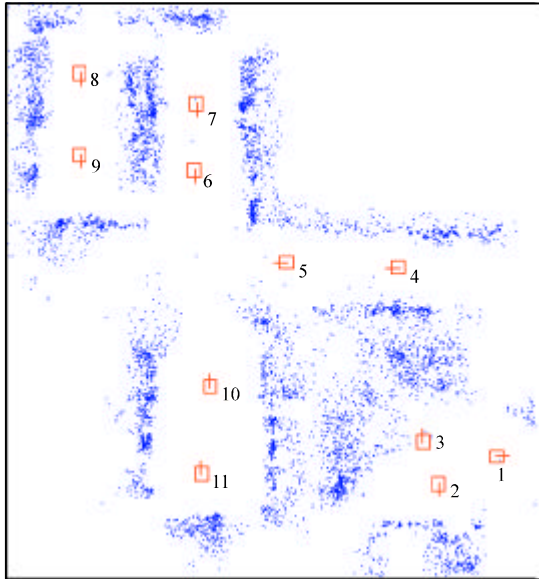


Fig. 8: Bird's-eye view of the map

shown in Fig. 8. The quantitative localization results are shown in Fig. 9. As shown in Fig. 9, the average Euclidean translation error is less than 10 cm in x direction and 15 cm in y direction for these 11 cases. The localization error is much larger and unstable when the robot locates at the corridor, such as case 4 and case 5, this because there are much less features in these locations. Figure 10 shows the localization error Gaussian distribution, which indicate further robustness of our presented method. These errors are sufficient for our requirement and then the proposed method has superior performance. Figure 11 show the detailed performance contradiction with the shows method.

For illustrating the robustness of our autonomous navigation method with an omnidirectional camera, as shown in Figure 8, we use this map for the navigation experiments.

In this map, we choose 17 key nodes to test the recognition performance. The quantitative recognition performances are shown in Fig. 12. The correct recognition rate based on the proposed hierarchical localization method is high when the robot moves from the starting to the destination. False recognition results are mostly obtained in node with very few features or in very ambiguous nodes.

Furthermore, we have tested control rule for the node converging performance as shown in Fig. 13. The initial robot poses are collected near the navigation path with the orientation similar to the desired paths. As a result, in most cases, the robot pose successfully converges to the recognized target node. The failed poses are caused by

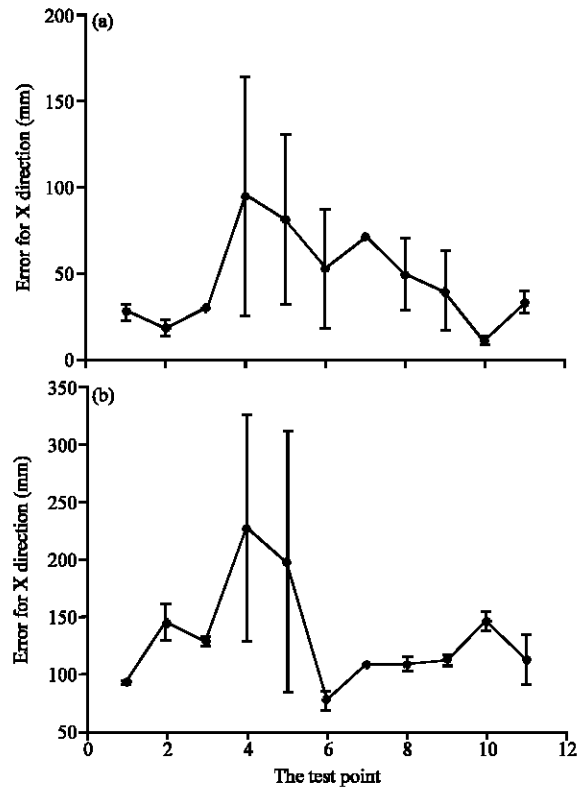


Fig. 9: (a, b) The localization results

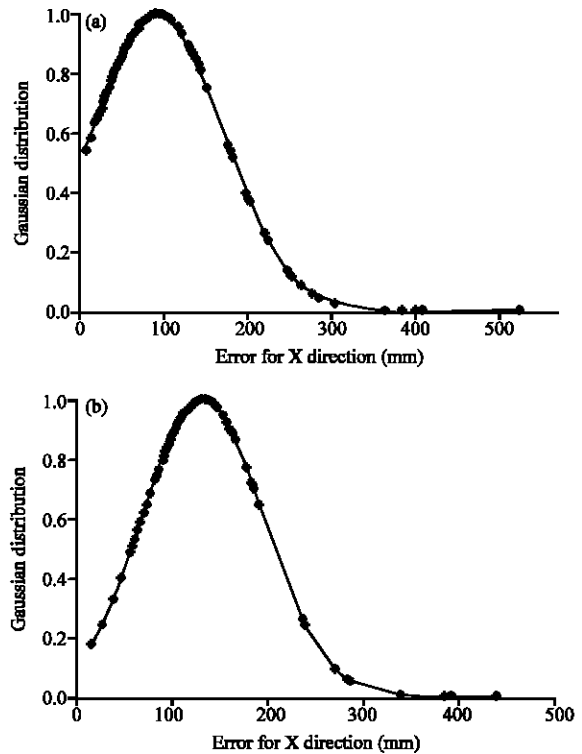


Fig. 10: (a, b) Localization error Gaussian distribution

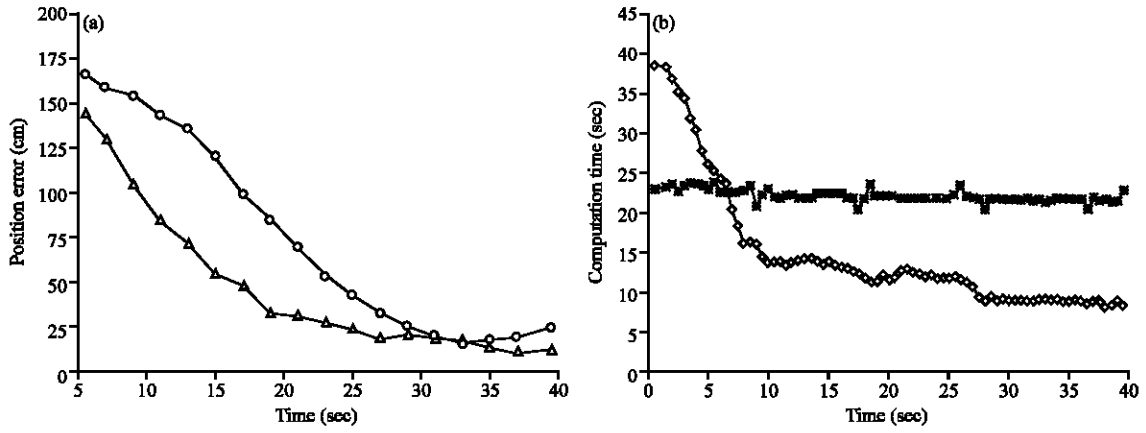


Fig. 11: Performance contradiction: (a) position error; (b) computation time

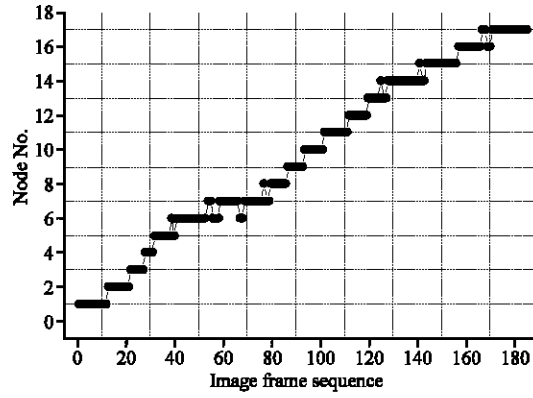


Fig. 12: Node recognition result

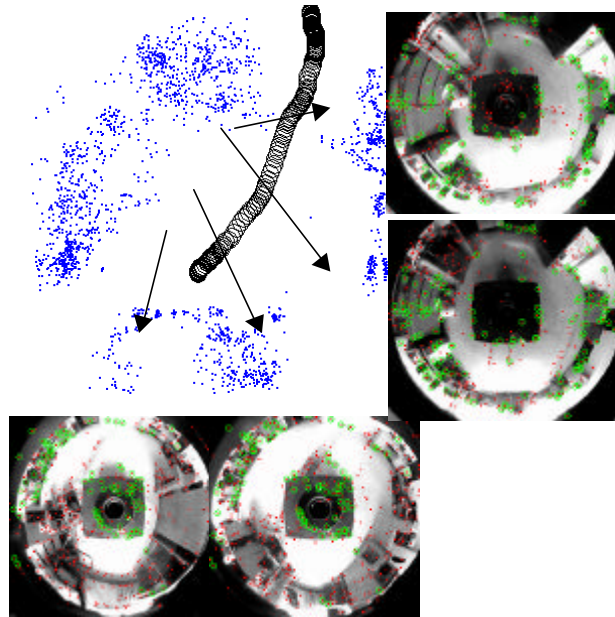


Fig. 13: Examples of iterative converging process

false recognition results, because they produce false relative node estimation and control rule which will eventually diverge the robot pose.

CONCLUSIONS

This study has proposed a hierarchical system for an indoor mobile robot based on omnidirectional vision. The localization is purely vision-based and uses a sequence of prerecorded image frames as reference topological map. For finding the image which best fits the current image, the system consists of two hierarchical components: global coarse localization and local accurate localization. The global color histograms are used to confirm fast the coarse robot position, which is not accurate, whereas the more accurate position can be obtained slowly through local scale invariant interest keypoints. These keypoints represented with multi-dimension descriptors, can be robustly matched despite changes in contrast, scale and viewpoint, demonstrated with nearest neighbor search based on KD-Tree. The navigation system is composed of on-line and off-line stages. During the off-line learning stage, the robot performs paths and records a set of ordered key images. From this sequence a topological map is built. Each topological node provides a set of omnidirectional images characterized by geometrical scale invariant keypoints. In the on-line navigation stage, the robot recognizes the most likely node through robust hierarchical algorithm and then controlled by a vision based control law adapted to omnidirectional cameras to follow the visual path. To evaluate this method, actual experiment is carried out in an indoor environment. Results show the robustness and efficiency of the proposed method.

ACKNOWLEDGMENT

My Research Project was partially sponsored by the 44th China Postdoctoral Science Foundation (period: 2008-2010, China) and the National Natural Science Foundation of China (No. 60905047) (period: 2010-2012, China).

REFERENCES

- Barfoot, T.D., 2005. Online visual motion estimation using fastslam with sift features. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, Aug. 2-6, Edmonton, Canada, pp: 579-585.
- Basri, R. and E. Rivlin, 1995. Localization and homing using combinations of model views. *Artificial Intel.*, 78: 327-354.
- Bosse, M., P. Newman, J. Leonard, M. Soika, W. Feiten and S. Teller, 2003. An atlas framework for scalable mapping. *Proceedings of the IEEE International Conference Robotics and Automation (IICRA'03)*, Taiwan, pp: 1899-1906.
- Bradley, D., R. Patel, N. Vandapel and S. Thayer, 2005. Real-time image based topological localization in large outdoor environments. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, August 2005, Edmonton, Canada, pp: 3670-3677.
- Choset, H. and K. Nagatani, 2001. Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization. *IEEE Trans. Robot. Automat.*, 17: 125-137.
- David, G.L., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60: 91-110.
- Duckett, T., S. Marsland and J. Shapiro, 2000. Learning globally consistent maps by relaxation. *IEEE Int. Conf. Robotics Automation*, 4: 3841-3846.
- Dudek, G. and C. Zhang, 1996. Vision-based robot localization without explicit object models. *Proc. Int. Conf. Robotics Automation.*, 1: 76-82.
- Ekvall, S., D. Kragic and P. Jensfelt, 2007. Object detection and mapping for service robot tasks. *Robotica*, 25: 175-187.
- Estrada, C., J. Neira and J.D. Tardos, 2005. Hierarchical SLAM: Real-time accurate mapping of large environments. *IEEE Int. Conf. Robotics Automation*, 21: 588-596.
- Fengda, Z.H., K. Lingfu and W. Peiliang, 2008. Mobile robot monte-carlo localization using image retrieval. *Proceedings of the 3rd IEEE Conference on Industrial Electronics and Applications*, June 3-5, Singapore, pp: 1362-1365.
- Hayet, J.B., F. Lerasle and M. Devy, 2003. Visual landmarks detection and recognition for mobile robot navigation. *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, June 18-20, Madison, WI, pp: 313-318.
- Jogan, M. and A. Leonardis, 2003. Robust localization using an omnidirectional appearance-based subspace model of environment. *Robotics Autonomous Syst.*, 45: 51-72.
- Kosecka, J., P. Barber and Z. Duric, 2003. Qualitative image based localization in indoors environments. *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition (CVPR'03)*, June 18-20, Madison, Wisconsin, pp: 3-10.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60: 91-110.

- Luke, L. and W. Stefan, 2004. Reduced sift features for image retrieval and indoor localisation. Proceedings of the Australian Conference on Robotics and Automation. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.9220>.
- Paletta, L., S. Frintrop and J. Hertzberg, 2001. Robust localization using context in omnidirectional imaging. Proc. Int. Conf. Robotics Automation, 2: 2072-2077.
- Schultz, A.C. and W. Adams, 1998. Continuous localization using evidence grids. Proceedings of the IEEE International Conference Robotics and Automation (ICRA), May 16-20, Leuven, Belgium, pp: 2833-2839.
- Se, S., D. Lowe and J. Little, 2001. Local and global localization for mobile robots using visual landmarks. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct. 29-Nov. 03, Maui, Hawaii, pp: 414-420.
- Siggelkow, S., 2002. Feature histograms for content-based image retrieval. Ph.D. Dissertation, Institute for Computer Science, University of Freiburg, Germany.
- Sim, R., G. Dudek, 1999. Learning and evaluating visual features for pose estimation. Proceedings of the 7th International Conference Computer Vision (ICCV), Sept. 1999, Kerkyra, Greece, pp: 1217-1222.
- Tamimi, H., 2006. Vision-based features for mobile robot localization. Ph.D. Dissertation, Department of Computer Science, University of Tübingen, Germany.
- Weiss, C., A. Masselli, H. Tamimi and A. Zell, 2007. Fast outdoor robot localization using integral invariants. Proceedings of the 5th International Conference on Computer Vision Systems, May 21-24, Bielefeld, Germany, pp: 24-24.
- Wolf, J., W. Burgard and H. Burkhardt, 2005. Robust vision-based localization by combining an image retrieval system with monte carlo localization. IEEE Trans. Robotics, 21: 208-216.