2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Multi-Hypothesis Outdoor Localization using Multiple Visual Features with a Rough Map

Jooseop Yun and Jun Miura

Abstract—**We describe a method of mobile robot localization based on a rough map using stereo vision, which uses multiple visual features to detect and segment the buildings in the robot's field of view. The rough map is an inaccurate map with large uncertainties in the shapes, the dimensions and the locations of objects so that it can be built easily. The robot fuses odometry and vision information using extended Kalman filters to update the robot pose and the associated uncertainty based on the recognition of buildings in the map. We use multi-hypothesis Kalman filter to generate and track Gaussian pose hypotheses. An experimental result shows the feasibility of our localization method in an outdoor environment.**

I. Introduction

This paper presents an approach to determining the pose (position and orientation) of a mobile robot in an urban area using a set of stereo image pairs. Understanding the surrounding scene and identifying man-made structures are important tasks for the localization of a mobile robot in outdoor environments. For outdoor robot localization, GPS-based approaches are often accompanied with odometry. While GPS can provide more accurate pose information in open spaces, GPS signals are susceptible to various forms of interference and can be quite unreliable in urban areas [1].

Computer vision can provide both accurate localization and robustness against these environmental influences (see [2] for a survey). Vision-based approaches are attractive because they are self-contained in the sense that they require no external infrastructures such as beacons or satellites. The knowledge of having buildings in the environment allows us to exploit their typical characteristics: horizontal or vertical principal directions and abundant parallel or orthogonal relationships between lines and surfaces in the buildings. The vision system used in this paper attempts to capture these highly structured configurations in the buildings.

A. Approach

In this paper, we will guide the robot with a rough map which represents an environment as a set of 2D line segments and can thus be built easily. The map approximates the outlines of buildings except detailed features to be used as landmarks. We propose a method to robustly estimate the robot pose in the map using multiple visual features: low-contrast regions, nonvertical borders, vertical borders and disparity regions. Low-contrast regions include the sidewalls of buildings and the sky in outdoor scenes. Nonvertical and vertical borders are detected from the building structures such as windows, doors, corners, a roof and so forth. And the disparity regions are extracted for matching with the walls of buildings. Multiple visual features are matched to the given map and the results are integrated into the odometry for the estimation of robot pose using Extended Kalman Filter.

The unreliable data-association problems in vision-based localization motivate the development of method that can make delayed decision, i.e. a multi-hypothesis approach. The approach allows maintaining when and where to place pose hypotheses as many as necessary and as few as possible. This property is provided by using a constraint-based search in an interpretation tree. This tree is spanned by all possible local-to-global data associations, given a local map of observed features and a global map of model features [3, 4].

In our work, we explore a localization problem using a rough map in real outdoor environments. To solve this localization problem, we use a novel combination of efficient map-matching scheme and multi-hypothesis technique based on multiple visual features. For the efficient map matching, we use the ordering and the priority constraints of multiple visual features extracted robustly using stereo and low-contrast region. In addition, hypothesis generation is combined with the EKF framework to do the multi-hypothesis localization, followed by heuristic hypothesis management techniques. As far as we know, there are no works that previously explored about this kind of localization problem.

B. Related Works

Recent research in vision-based localization in urban environments focuses on the recognition and matching of building facades. Georgiev and Allen [5] have used vision-based techniques to supplement GPS and odometry. Their system requires detailed geometric models and they have only localized views in the vicinity of only a single building. Johansson and Cipolla [6] determined the relative pose of a camera by computing the transformation required to match the rectified facade view from a single image. Reitmayr and Drummond [7] presented a model-based hybrid tracking system for outdoor augmented reality. The system employs a 3D model capturing the overall shape of buildings as large planar surfaces with highly detailed textures. In these systems, creating such detailed models for large outdoor environments becomes a troublesome task.

Since it is hard to build the exact map of outdoor environments, using inaccurate maps is another easy method of giving the environmental information to the robot.

Jooseop Yun and Jun Miura are with the Department of Mechanical Engineering, Osaka University, Suita, Osaka, 565-0871, Japan (e-mail: {jsyun, jun}@cv.mech.eng.osaka-u.ac.jp).

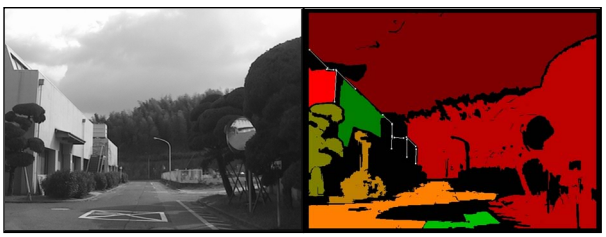
Inaccurate maps may include hand-drawn map or topological-geometrical map, where the relative poses among object models can be uncertain. Hand-drawn map is an interface for a sketch-based navigation [8], and topological-geometrical map is a hybrid map for navigation in large-scale environment [9]. The hand-drawn map is, however, hard to use for navigation tasks, because it has no metric information. And the topological-geometrical map needs a lot of cost to build and update the metric local models for object recognition in outdoor environments.

II. Feature Detection

In this paper, we shall consider self-localization by means of only vision and odometry. We are interested in navigation around urban environments such as our university campus. Since views of trees, cars and bicycles, however, differ from time to time, we use multiple visual features observed from buildings by using stereo vision with the angle of elevation more than 10°. Their multiple visual features are relatively large and static as landmarks to be used for localization as follows: low-contrast regions for identifying nonvertical and vertical borders, nonvertical borders for the vanishing points to calculate the wall directions of buildings, vertical borders corresponding to the corners of buildings, and disparity regions for matching with the walls of buildings.

A. Low-contrast Regions

Many real world scenes contain regions of low contrast. Typical regions include the sidewalls of buildings and the sky in outdoor scenes. Such low-contrast regions are very hard to estimate intensity-based depth because they lack any distinctive texture. We can, however, exploit the existence of low-contrast regions instead of their limited texture for detecting multiple visual features. In this case, the multiple visual features should line up in the intermediate regions of different low-contrast regions. The low-contrast region processing involves segmenting each image into no overlapping regions based on intensity. A simple linking algorithm starts at every pixel in the image and recursively grows out regions of similar intensity [10]. An input image at the first frame and its resultant low-contrast regions are shown in fig. 1. We recognize the sky regions (the top region shown in fig. 1 (b)) from the extracted low-contrast regions [11].



(a) (b)

Fig. 1. An input image (a) and its low-contrast regions with detected borders (b).

B. Border Features

We extract couples of nonvertical and vertical line segments fitted to the edge pixels and attain coupled borders from the line segments near enough to each other both in the image and the disparity space. When identifying the borders, we consider the height of their end points above ground using a segment-based stereo algorithm [12]. A nonvertical border can be coupled with up to two vertical borders which may be on its left and right sides, respectively, and vice versa. Each end of a nonvertical border should be coupled with the upside end of its coupled vertical border. Isolated borders are then detected when they are adjoining to the sky regions.

B.1. Nonvertical Borders

Nonvertical borders can be extracted from the building structures and can provide the relative orientation between the robot and the building. What is necessary for estimating the relative orientation in this case a vanishing point. The vanishing points (VPs) of nonvertical borders exist on the horizon in the image. We can then estimate the angles between the image plane and the lines from the camera center to vanishing points (VP1 or VP2 in fig. 2). The lines are parallel to the respective directions of visible walls with respect to X_G of the global coordinate system.

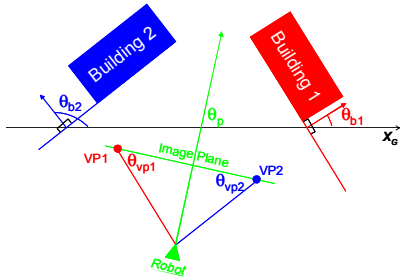


Fig. 2. Estimation of a robot orientation from vanishing points.

From fig. 2, we can deduce a following relation among the robot's orientation, θ_p with a range of [-π, π], the inward direction of a wall, θ_b with a range of [-π, π], and the angle from a vanishing point, θ_{vp} with a range of [-π/2, π/2]:

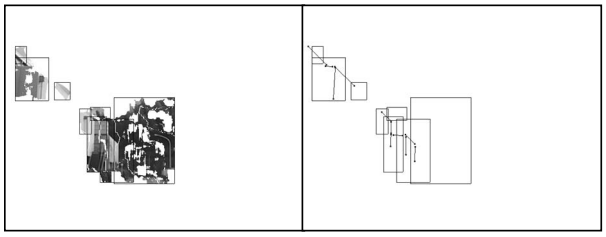
$$\theta_p - \theta_b + \theta_{vp} = 0. \quad (1)$$

B.2. Vertical Borders

Vertical borders correspond to the corners of buildings. Vertical borders are coupled with nonvertical borders as described in the previous section. Isolated vertical borders are detected also considering their heights when they are adjoining to the sky regions. The detected borders must be in the intermediate regions of different low-contrast regions as shown in fig. 1 (b) because there are many tall trees and streetlights resulting in similar features. The black regions in the figure are not low-contrast regions.

C. Disparity Regions

We use an area-based stereo matching in order to extract a disparity image [12]. The depth is less sensitive to changes of illumination than the previous visual features provided by using a single intensity image [13]. In this study, disparity regions are detected as connected regions with a disparity in the disparity image.



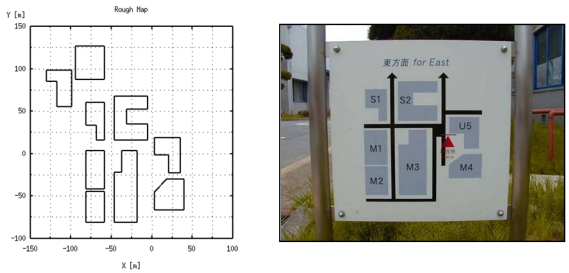
(a) (b)

Fig. 3. Detected building regions (a) and all multiple visual features (b).

One of the advantages of stereo vision is that it provides a more informative 2D depth map. With a priori knowledge about the minimum height of buildings in an urban environment, we can extract the regions of buildings from the disparity regions using the height calculated from the disparity image. The histograms of intensity gradient orientations in the regions weighted by their gradient magnitudes seem to be well-suited for discrimination between urban structures and natural environments. Intuitively, the histograms for the building regions tend to be unimodal or bimodal and most of whose peaks tend to be separated by approximately right angles to each other. But, the histograms at the tree regions, for example, tend to be more uniformly distributed and the peak values have lower maximum values than those of the building regions [11]. The textured boxes in fig. 3 (a) show the resultant building regions from in fig. 1 (a). The detection result of all multiple visual features is shown in fig. 3 (b), where the rectangles represent recognized building regions.

III. ROUGH MAP

Much of research efforts in the robot navigation have been directed towards the object representation on the map and the object recognition using the map. Although an accurate map provides accurate and efficient localization, it needs a lot of cost to build and update [5, 9]. A solution to this problem would be to allow a map to be defined roughly since a rough map is much easier to build [8]. We assume that the buildings in the rough map have planar walls and that these planes have both horizontal and vertical edges. This is often the case for buildings as they have windows and doors. We also assume a flat polygon on the top of a building as a roof since roof details on a tall building cannot be seen from the ground level.



(a) (b)

Fig. 4. An example of rough map (a) and a guide map of our campus (b).

The characteristics of rough map can be summarized as follows: The exact model of map uncertainty is unknown. The uncertainty may be not uniform across the map. The geometric details such as exact outlines, exact dimensions and exact poses of buildings are not available. The map also lacks information about exact models of the building structures. Fig. 4 shows a guide map for visitors to our university campus and an example of rough map built from the map. Human can use this kind of map to navigate efficiently, but it is difficult for the robot to use it, because of the deficiency of accurate metric and geometric information.

Relative poses between landmarks in a rough map are allowed to be uncertain. The uncertainty of rough map might cause the robot pose to be inconsistent if it is represented in the global coordinate system of reference. To address this problem, we represent the robot pose in a local coordinate system attached to a landmark which the robot has recognized recently. When the robot finds a new landmark, the robot changes the local coordinate system from the old landmark to the new one with coordinate transformation of its pose based on the relative pose between the old and new landmarks. We refer to the landmark as a local origin. As the robot moves, it changes the local origin. More specifically, we define the robot pose as a pair of a local origin and the pose in a local coordinate system attached to the local origin. Landmarks in the building with the local origin would have smaller positional uncertainty in the local coordinate system than in the global one thereby becoming easier to recognize.

To handle the uncertainty, the relative pose between two local coordinate systems is defined using a Gaussian random variable. Let $\boldsymbol{d}_{jk} = (x_{jk}, y_{jk}, \theta_{jk})^T$ be the relative pose of local coordinate system L_k with respect to local coordinate system L_j. The transformation of robot pose $\boldsymbol{X}_R = (x, y, \theta)^T$ from local coordinate system L_j to L_k can be calculated as follows:

$$\boldsymbol{X}_{Rk} = \mathbf{T}^{-1}(\theta_{jk})(\boldsymbol{X}_{Rj} - \boldsymbol{d}_{jk}), \tag{2}$$

$$\mathbf{T}(\theta_{jk}) = \begin{bmatrix} \cos(\theta_{jk}) & -\sin(\theta_{jk}) & 0 \\ \sin(\theta_{jk}) & \cos(\theta_{jk}) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where $\boldsymbol{X}_{Rj}$ and $\boldsymbol{X}_{Rk}$ are the robot poses with respect to the local coordinate systems L_j and L_k, as shown in fig. 5.

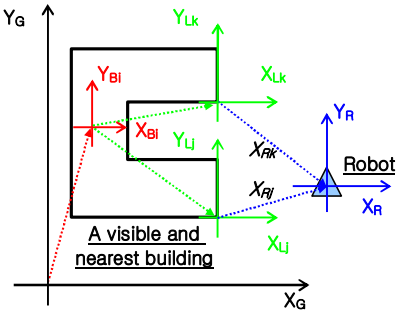
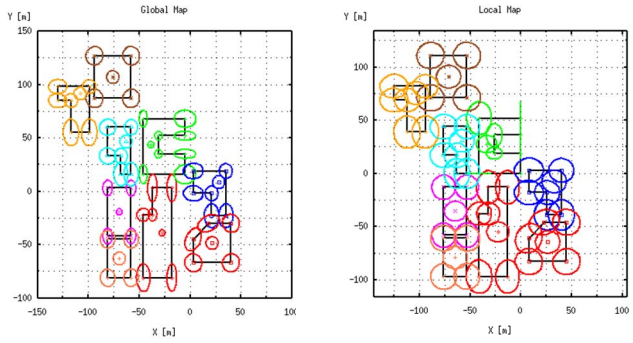


Fig. 5. A relationship among global (G), building (B), local (L) and robot (R) coordinate systems.

Fig. 5 shows representative transformations among global (X_G-Y_G), building (X_B-Y_B), local (X_L-Y_L), and robot (X_R-Y_R)

coordinate systems. A building coordinate system has its origin at the center of building and parallel to the principal directions of building. A local coordinate system is selected such that its origin is at a visible and nearest corner of a certain building from the robot. The local coordinate system is also parallel to the building coordinate system or one of its axes placed on the visible and nearest wall of the building. The uncertainty of local coordinate system is calculated from the uncertainty of building coordinate system with respect to the global coordinate system.



(a) (b)

Fig. 6. The uncertainty model of the global map (a) and a local map at the first frame (b).

We approximate the buildings present in an environment to polygonal objects on the map and compute the uncertainties of their poses and dimensions for estimating the robot pose. Fig. 6 shows examples of modeling the uncertainties of rough map in the global and local coordinate systems, where the buildings are drawn with their mean poses and mean dimensions. The global map uncertainty is roughly assumed with respect to the origin of the map and transformed to the local uncertainty model with respect to its local origin by a coordinate transformation using (2) and (3). The local map has greater uncertainty than the global map except the walls and the corners on the axes of the local coordinate system. The coordinates of (0, 0) on the both maps are their respective origins and the local origin has no error. Assuming Gaussian error models, the colored ellipses in each map mark 3σ uncertainty regions of the centers and corners of different buildings by their relative pose errors with respect to the respective origin.

IV. Map Matching

One of the most important and challenging aspects of map-based localization is map matching. In the matching process, we use the depth information: a segment-based stereo algorithm for the borders and an area-based one for the disparity regions corresponding to the walls of buildings. The multiple visual features should be used in order to match the sensory data to the environment map reliably.

When visual pose estimation is attempted, an approximate estimate of the pose is available from the odometry. This estimate is used to search the map for the most appropriate buildings for visual localization. The buildings within a certain distance range from the robot are selected by scanning through the given map. Only the features of buildings that are viewable under the orientation uncertainty of the robot pose are considered. We also eliminate the features of buildings that are visible at a too low angle to produce a stable match with the image.

Given the candidate features of buildings that successfully passed this selection process, the robot matches the detected features to those candidates with the Mahalanobis distance criterion using the depth. The resolution of the depth data is, however, not constant; the farther from the stereo camera, the larger the error of depth. In order to simplify the computation, we use disparity space [14] which keeps the error constant.

At this point, we are ready to match up the multiple visual features with the map. Instead of generating and testing all of the data associations using the multiple visual features simultaneously, we build the data associations according to the priority of nonvertical borders, vertical borders, and disparity regions in this order (see section V). Since this priority constraint can reduce more effectively the uncertainty of robot orientation, we can thus narrow the map candidates for the data association more efficiently.

The data association of nonvertical borders with outlines of buildings is possible when satisfying the following criteria:

1) The Mahalanobis distance d_{vp} between their vanishing points should be close enough to each other.

$$d_{vp} = (x_{vp} - X_{vp})(\sigma^2_{xvp} + \sigma^2_{Xvp})^{-1}(x_{vp} - X_{vp}), \qquad (4)$$

where x_{vp} and X_{vp} are the vanishing points of a nonvertical border and a building outline, and σ^2_{xvp} and σ^2_{Xvp} are their uncertainties, respectively. If d_{vp} is small, two vanishing points are considered to be consistent.

2) The Mahalanobis distance d_{rt} between their parameters in the disparity space should be small enough, because there are many buildings having parallel walls in urban environments.

$$d_{rt} = (\mathbf{x}_{rt} - \mathbf{X}_{rt})^T (\mathbf{\Sigma}_{\mathbf{x}rt} + \mathbf{\Sigma}_{\mathbf{X}rt})^{-1}(\mathbf{x}_{rt} - \mathbf{X}_{rt}), \qquad (5)$$

where $\mathbf{x}_{rt}$ and $\mathbf{X}_{rt}$ are the Hough parameters (r, t) in the disparity space of a nonvertical border and a building outline, and $\mathbf{\Sigma}_{\mathbf{x}rt}$ and $\mathbf{\Sigma}_{\mathbf{X}rt}$ are their error covariance matrices, respectively. The judgment of data association depends on the threshold d_{thresh} for each distance.

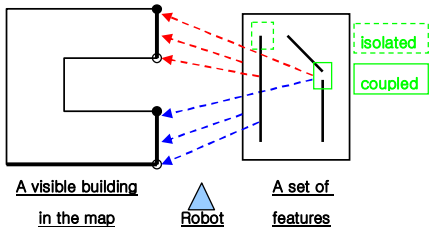


Fig. 7. Correspondences between a set of multiple visual features and a visible building.

The vertical borders are associated with the corners of

buildings using the Mahalanobis distance criterion in the disparity space. The coupled borders are associated with the coupled features of building outlines and corners. We also consider the ordering constraint in their associated building.

In the case of disparity data, the disparity regions recognized as building regions are associated with the walls of buildings using the Mahalanobis distance criterion. Neighboring disparity data corresponding to a wall are grouped from the result of data association. By using a plane fitting procedure, each group is fitted to a plane. The results are the plane models as the environment representation.

Fig. 7 shows a building in the map with visible walls and corners pictured as thick segments, and black dots and circles, respectively, on the left side. Coupled borders, an isolated vertical border, and a rectangular disparity region bounding them are shown on the right side of the figure. A set of the arrows of same color depicts that coupled borders, an isolated vertical border, and a disparity region are consistently matched to the outline, the corners, and the wall of same building, respectively. Considering the coupled borders and their ordering constraint, we can generate consistent sets of the data associations of coupled nonvertical and vertical borders, an isolated vertical border and a disparity region.

V. Multi-hypothesis Localization

The Kalman filter acts as a pose tracker in this paper. But, a false matching of the observed features to the model features can lead to an irrecoverable lost situation if only a single distribution is maintained. Although the most credible estimation at one time turns out to be totally wrong, a Multiple Hypothesis Localization allows alternative pose estimates to be maintained instead of tracking only the most credible hypothesis. The MHL has been widely used to solve global localization problems, in which a robot has no knowledge of its initial pose. In our work, however, instead of starting with an empty hypothesis, we start with a highly reliable hypothesis of robot pose. A starting robot pose around the true pose and its uncertainty of random size must be supplied by an operator.

This method can be achieved by explicitly tracking multiple pose hypotheses, via multiple Kalman filters using the priority data associations of multiple visual features (refer to our work [15] for detailed implementation). Direct association of all observations to all targets is, however, not practical as the number of possible hypotheses may be huge with frame steps. This is the reason various heuristics are introduced to keep the algorithm practicable in the method.

Step 1: Hypothesis evolution by robot motion. When the algorithm starts, it takes as input the prior set of pose hypotheses from the previous cycle. Each of the current hypotheses evolves to take into account the uncertainty of robot motion according to the odometry. Then, the local origin of each evolved pose hypothesis is changed when a new one is found. The global map is transformed to a local map with respect to the new local coordinate system and then to the disparity space for map matching. When no features are detected from the current input image or no detected features are matched with the map, the evolved pose hypotheses become the input set of current pose hypotheses in next cycle.

Step 2: Hypothesis generation by nonvertical borders. Nonvertical borders are first used for hypothesis generation. For each combination of possible associations between nonvertical borders and building outlines, a new pose hypothesis is generated using EKF. In this hypothesis generation, the robot orientation is mainly adjusted.

When each pose hypothesis violates one of the following four constraints, the hypothesis is considered to be infeasible:

- The pose hypothesis should satisfy the relative ordering constraint of borders in a single building.
- Matched borders in each data association must be in the predicted field of view.
- Each association should be possible in the sense of Mahalanobis distance check.
- A generated pose must not be largely far away from the evolved pose in the Mahalanobis distance sense.

Such infeasible pose hypotheses are all pruned.

Step 3: Hypothesis generation by vertical borders. Vertical borders are then used for hypothesis generation. For each pose hypothesis generated in the previous step, a set of combinations of consistent associations between vertical borders and building corners is generated. A new pose hypothesis is generated for each combination by using EKF. In the case of coupled vertical borders, only the hypotheses generated using the corresponding nonvertical borders are considered. The same pruning process is then applied to the generated pose hypotheses.

Step 4: Hypothesis refinement by disparity regions. For each pose hypothesis generated using nonvertical and/or vertical borders, the disparity data are clustered and matched with the map of the hypothesis. The pose hypothesis is then refined using the matching and EKF. The same pruning process is again applied to the refined pose hypotheses.

Step 5: Hypothesis merging. If multiple pose hypotheses have the same local coordinate system and are within a specified range of each other in the Mahalanobis distance sense, they are grouped and merged into a new pose hypothesis. Since the hypotheses in a group are considered to be equally plausible, the resultant pose estimate of merged hypothesis is the mean of their poses and its covariance estimated is determined to cover all their uncertainty regions. All pose hypotheses after this step constitute the input set of current pose hypotheses in next cycle.

VI. Experimental Result

Our implementation uses the multiple visual features and the state prediction is made by using only odometry data. The system was tested on about 200 x 200 m^2 site with 9 buildings on variable outlines (refer to fig. 4). In the test, a sequence of stereo images was obtained by driving the robot using the joystick interface to the steering control system all along the path and back to the start position. The visual localization routine proposed in this paper was then performed. It used the

accumulated error from odometry as an initial guess to determine the visible buildings on the given map and chose their multiple visual features observed for localization. Fig. 8 shows a robot path and sampled images of 30 frames counterclockwise used for our experiment, where the number in circle indicates the number of frame step.

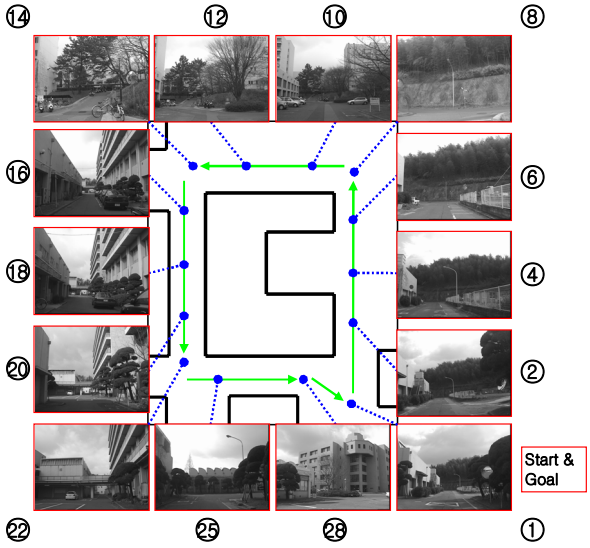


Fig. 8. A robot path and sampled images used for an experiment.

Fig. 9 shows the results of multi-hypothesis localization on the magnified local map using the multiple visual features at start position in fig. 8. The ellipses in the figure are the estimated 3σ uncertainty regions of the robot positions by matching the respective visual features to the map. A corner of a building linked to the centers of ellipses means the origin in the local map of the pose hypotheses. Multiple pose hypotheses are merged in order to keep the number of hypotheses low.

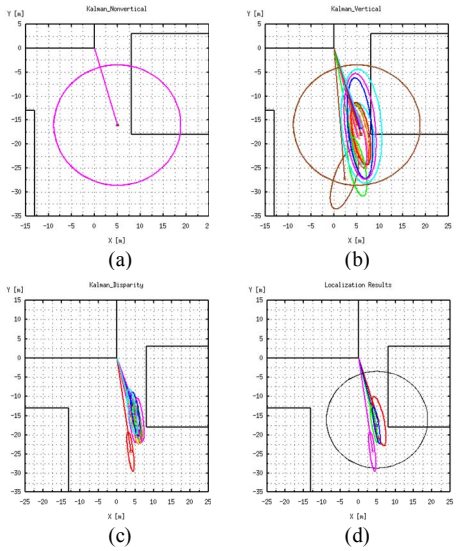


Fig. 9. Multiple pose hypotheses set at the first frame step using nonvertical borders (a), vertical borders (b), disparity regions (c) and resulting merged hypotheses (d) of (c).

At the first frame (start position), the robot has a rough knowledge on its pose and observes 6 nonvertical borders, 5 vertical borders and 8 disparity regions (refer to fig. 3 (b)). This yields 20, 14 and 13 pose hypotheses for each visual feature, respectively (shown in fig. 9 (a) through (c)); only position information from the pose hypotheses is displayed. Fig. 9 (a) shows a 3σ uncertainty ellipse of superposed 20 pose hypotheses using nonvertical borders. The reason for the superposition is because we assumed no correlation between the position and orientation of the robot at start position. The vanishing point of a nonvertical border provides information about the robot's orientation only. 13 pose hypotheses generated using disparity data in fig. 9 (c) are merged to 4 hypotheses in fig. 9 (d) by the constraint of the same local coordinate system and the threshold of Mahalanobis distance. The large circle drawn in fig. 9 (d) denotes the uncertainty region of an initial robot position with respect to a local coordinate system. The uncertainty circle consists of the global uncertainty of an initial robot pose and the global uncertainty of a local origin by a coordinate transformation using (2) and (3).

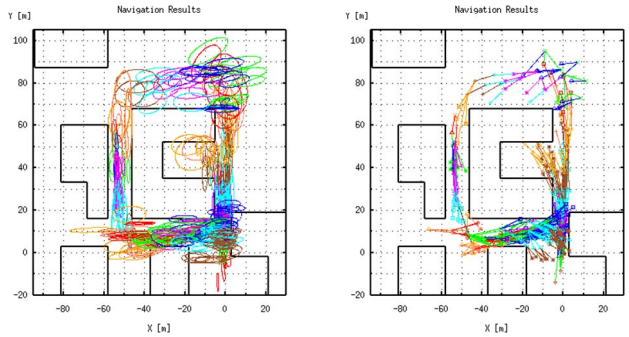


Fig. 10. The results of multi-hypothesis localization on the test run.

During this test run of 30 frame steps, the sensor data were recorded by the robot stopping at regular intervals to take a pair of stereo images. The average relative displacement between the observations of each frame was less than 10 m and 20° for translation, and less than 90° for rotation. The algorithm succeeded always in generating and tracking the pose hypotheses around the true poses of robot. The error ellipses in fig. 10 (a) denote the unmagnified 3σ uncertainty levels of the robot positions in the local coordinate system displayed on the global map. The linked tracks of the ellipse centers between current and descendent pose hypotheses in the figure are shown in fig. 10 (b) with line segments of same color for separating different frame steps. The terminated line segments in the figure represent the pruned hypotheses.

The robot stays localized in the presence of errors and sensing ambiguities where tracking of a single hypothesis would fail. This is a dramatic increase in robustness made possible with a small computational cost of pose hypotheses as shown in fig. 11. At each frame step, the average processing time of all hypotheses was lower than 5% of the total execution time including the visual processing. The minimum number of the pose hypotheses at each step was larger than 1,000 when the hypothesis management strategies

of validating, pruning and merging were not applied.

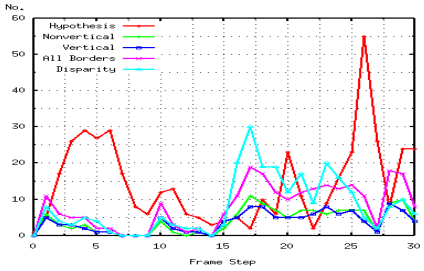
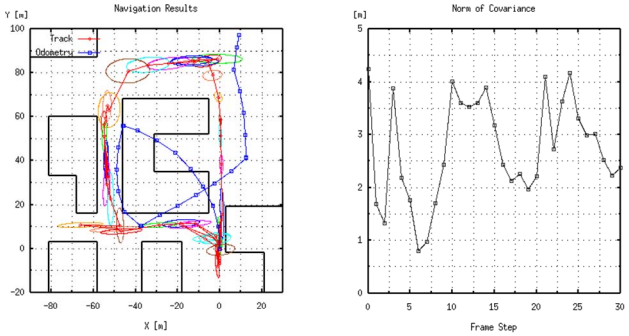


Fig. 11. Number of validated pose hypotheses and respective features.

The number of pose hypotheses generally increases as the number of observed features increases. Considering the large number of features around step 17, the small number of pose hypotheses at the step is, however, thought mainly due to that completely new buildings were obtained in the map when turning left into the byroad, and most of the observed features were extracted from and matched to the same building. The high number of pose hypotheses for the low number of features around step 26 is thought due to the place where there exist many buildings.



(a) (b)

Fig. 12. A longest track by the proposed method and an odometry-alone result (a) and norm of uncertainty covariance along the track (b).

Fig. 12 (a) shows a comparison of dead-reckoning estimates and a trajectory estimated by the MHL approach on the global map which drawn with the mean poses and mean dimensions of buildings. The total path length is about 250 meters and the image sequence consists of 30 frames. The dead-reckoning path is completely wrong after a short frame interval. A longest track of the tracks displayed in fig. 10 (b) is also plotted with its 3σ uncertainty ellipses at each step. It is backtracked from the end position nearest to the start position to demonstrate the feasibility and good performance of the proposed MHL approach in this paper. Fig. 12 (b) reports the change of the norm of position covariance along the track. When no matchable features are in view, the uncertainty of robot position becomes greater. The uncertainty of robot position decreases whenever matchable features are in the field of view.

VII. Conclusion and Future Work

This paper presents an approach to determining the robot pose in an urban area where GPS cannot work since the satellite signals are often blocked by the buildings. We tested the method with real data and the obtained results show that the method is potentially applicable even in the presence of errors in feature detection of the visual features and incomplete model description of the rough map.

To make use of the rough map, which is an incomplete description of the environment, we deploy a technique based on multi-hypothesis tracking in localization. The main disadvantage of the multiple hypothesis approach is, however, the very large number of hypotheses that may be generated. But, the hypothesis management techniques of validating, pruning and merging appear to constrain the hypothesis trees to manageable sizes.

This method is a part of our ongoing research aiming autonomous outdoor navigation of a mobile robot to follow the planned path to a user-chosen location on the rough map. Thus, we want to address the integration of our system with an autonomous navigation module in the near future.

References

[1] K. Ohno, T. Tsubouchi, B. Shigematsu, and S. Yuta, Differential GPS and odometry-based outdoor navigation of a mobile robot, in *Advanced Robotics*, 2004, Vol. 18, No. 6, pp. 611-635.

[2] G. N. DeSouza and A. C. Kak, Vision for mobile robot navigation: a survey, in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, Vol. 24, No. 2, pp. 237-267.

[3] P. Jensfelt and S. Kristensen, Active global localization for a mobile robot using multiple hypothesis tracking, in *IEEE Trans. on Robotics and Automation*, 2001, Vol. 17, No. 5, pp. 748-760.

[4] K. O. Arras, J. A. Castellanos, M. Schilt, and R. Siegwart, Feature-base multi-hypothesis localization and tracking using geometric constraints, in *Robotics and Autonomous Systems*, 2003, Vol. 44, No. 1, pp. 41-53.

[5] A. Georgiev and P. K. Allen, Vision for mobile robot localization in urban environments, in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2002, pp. 472-477.

[6] B. Johansson and R. Cipolla, A system for automatic pose-estimation from a single image in a city scene, in *Proc. of Int. Conf. on Signal Processing, Pattern Recognition, and Applications*, 2002.

[7] G. Reitmayr and T. Drummond, Going out: Robust model-based tracking for outdoor augmented reality, in *Proc. of IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, 2006.

[8] G. Chronis and M. Skubic, Sketch-based navigation for mobile robots, in *Proc. of IEEE Int. Conf. on Fuzzy Systems*, 2003, pp. 284-289.

[9] M. Tomono and S. Yuta, Mobile robot localization based on an inaccurate map, in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2001, pp. 399-405.

[10] N. L. Chang and A. Zakhor, Constructing a multivalued representation for view synthesis, in *International Journal of Computer Vision*, 2001, Vol. 45, No. 2, pp. 157-190.

[11] H. Katsura, J. Miura, M. Hild, and Y. Shirai, A view-based outdoor navigation using object recognition robust to changes of weather and seasons, in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2003, pp. 2974-2979.

[12] I. Moon, J, Miura, and Y. Shirai, On-line extraction of stable visual landmarks for a mobile robot with stereo vision, in *Advanced Robotics*, 2002, Vol. 16, No. 8, pp. 701-719.

[13] J. M. Porta, J. J. Verbeek, and B. J. A. Krose, Active appearance-based robot localization using stereo vision, in *Autonomous Robots*, 2005, Vol. 18, No. 1, pp. 59-80.

[14] A. Okamoto, Y. Shirai, and M. Asada, Integration of color and range data for Three-Dimensional scene description, in *IEICE Trans. on Information and Systems*, 1993, Vol. E76-D, No. 4, pp. 501-506.

[15] J, Yun and J. Miura, Multi-hypothesis localization with a rough map using multiple visual features for an outdoor navigation, in *Advanced Robotics*, to be published.