2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Road Recognition from a Single Image using Prior Information

Kiyoshi Irie and Masahiro Tomono

***Abstract*— In this study, we present a novel road recognition method using a single image for mobile robot navigation. Vision-based road recognition in outdoor environments remains a significant challenge. Our approach exploits digital street maps, the robot position, and prior knowledge of the environment. We segment an input image into superpixels, which are grouped into various object classes such as *roadway*, *sidewalk*, *curb*, and *wall*. We formulate the classification problem as an energy minimization problem and employ graph cuts to estimate the optimal object classes in the image. Although prior information assists recognition, erroneous information can lead to false recognition. Therefore, we incorporate localization into our recognition method to correct errors in robot position. The effectiveness of our method was verified through experiments using real-world urban datasets.**

I. Introduction

Road recognition is an important function of autonomous navigation and has been studied extensively. To realize effective road recognition, numerous methods have been proposed using cameras [1], laser scanners [2], and sensor fusion [3]. We approach the problem of road recognition using a single camera image. In comparison to laser scanners and sensor fusion, cameras offer high frame rates, increased compactness, and low costs.

Many existing road recognition methods are designed for vehicle applications, such as automated cars and advanced driver-assistance systems. We aim to assist the navigation of relatively small robots, mainly on sidewalks. Such robots must distinguish between sidewalks and roadways. However, recognizing sidewalks using an image is difficult because sidewalks are similar to roadways and their appearance can vary with place (Fig. 1). In addition, apparent road boundaries, such as curbs and lane markings, are not always used.

We propose to improve recognition using prior information. During navigation tasks, a robot has a map, which can be used to track its position. Map information is also helpful for recognition. In particular, we use existing digital street maps, which contain semantic information (such as roadways and buildings). Previously, we proposed a probabilistic localization framework using digital street maps [4]. Here we extend the framework to estimate the road region from the map and the robot position.

The input for the proposed method is a single image. Prior information is obtained from digital street maps, knowledge of the environment (essentially object existence probabilities), and robot position. The image is segmented into superpixels, which are classified into several object classes to distinguish road and boundary regions. The object classes include *roadway*, *sidewalk*, *curb*, and *wall*. In contrast to conventional road recognition algorithms (road vs. non road binary classification), our method provides richer information that should be useful for navigation.

Kiyoshi Irie and Masahiro Tomono are with Future Robotics Technology Center, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba , Japan {irie, tomono}@furo.org

Fig. 1. Various sidewalks in urban environments

The proposed method consists of three steps. First, superpixels are individually classified by MAP estimation using observation likelihood, which is calculated by a support vector machine (SVM), and prior probability, which is calculated using prior information. Second, energy minimization method is applied so that interactions between neighboring superpixels are considered. Third, errors in lateral displacement and orientation are corrected by iterating classification and localization under the energy minimization scheme.

By testing the method on datasets collected in cluttered urban environments, we verified that prior information improves the recognition rates of ambiguous objects. Although inaccurate position information can degrade recognition quality, our localization method successfully corrected small position errors and recognized essential features.

The main contributions of the paper are twofold. First, we compute prior probabilities from digital street maps; second, we develop an energy minimization framework that simultaneously recognizes roads and localizes the robot. Although we use SVM as the baseline for our comparison, our method can be combined with other road classification methods.

II. Related Work

Thus far, most road detection studies have focused on detecting roadways. Those road detection methods are not proven to be effective in sidewalk environments. Some methods learn road colors on-the-fly [3] [5], while others probabilistically track road boundaries [6] [7]. These methods use sequences of sensory data. Road detection using

single images (i.e., in the absence of temporal information) has also been studied [8] [9]. These methods assume that both sides of the road boundaries are shown in the input image.

Many of these methods use road-shape models, which limit recognition to the given shapes. Alvares et al. addressed this issue by employing digital street maps as arbitrary road shapes, which could be expressed in a map [10]. Although our method also uses digital street maps, significant differences exist between our method and the earlier study. Alvares et al.'s method uses only road shapes, while we generalize digital street maps as the source of prior probabilities and recognize not only road shapes but also objects at road boundaries.

Many image labeling problems can be represented as energy minimization problems, which are widely used in the computer vision community [11] [12]. Several authors have already applied energy minimization to robotic mapping [13] [14]. In our method, energy minimization is based on Boykov's image segmentation [15].

We also incorporated localization to our recognition method. Other researchers have used digital street maps for localization. Hetschel et al. used laser scanners to match the walls of detected buildings with those in the OpenStreetMap [16]. Morales et al. detected the center of the road using a laser scanner and matched the obtained road information to the map [17]. However, these methods assume that boundary lines can be detected from height information, which is not always the case. Previously, we addressed this issue by introducing a probabilistic framework that simultaneously recognizes road boundaries and localizes the robot [4]. The main differences between the proposed method and our previous method are that a single image is input and that SVM and energy minimization are adopted for recognition.

III. System Overview

Fig. 2 presents an overview of our proposed method. We recognize road regions and road boundary objects in an image under the assumption that a map and the approximate robot position is given. An input image is segmented into superpixels [18], denoted as $S = \{s_1, ..., s_n\}$. Image features extracted in superpixels, such as colors and textures, are denoted by $C = \{c_1, ..., c_n\}$. For road recognition, the superpixels are assigned to estimated object class labels $L = \{l_1, ..., l_n\}$. Objects in the input image are classified into eight classes (defined in Table I). The image features are shown in Table II. In this paper, rather than seeking the best possible feature set, we focus on improving recognition quality by exploiting prior information obtained from digital street maps, robot position, and knowledge of the environment. We construct grid maps containing semantic labels such as roadways and buildings, which are available in many existing digital street maps. So far, we manually convert existing maps into grid maps; however, it may be automated.

Our proposed method is divided into three steps. First, we classify each superpixel individually using MAP estimation based on prior information and observation likelihood calculated by an SVM. Objects of similar appearance, such as roads and gray walls, are not easily distinguished by SVM. Recognition of such objects is greatly improved by exploiting prior information. Second, we introduce an energy minimization method that globally estimates optimal labels from interactions between neighboring superpixels. Energy minimization improves recognition quality by smoothing the labels, thereby reducing noise. We employ graph cuts to minimize the energy [19]. Third, we extend energy minimization to estimate the errors in robot position. We iterate labeling and localization to simultaneously estimate optimal labels and robot position through energy minimization. When the robot is navigating a street, the longitudinal position is not readily estimated from a single image. Therefore, we estimate only the lateral displacement and orientation with respect to the road. Correcting errors in robot position is important because incorrect prior information can significantly degrade the recognition quality.

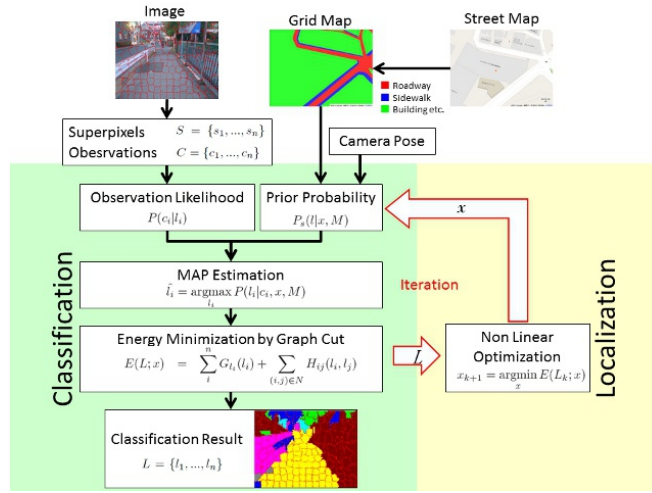

Fig. 2. Flow of proposed method

TABLE I
List of object classes $\mathcal{L}$

Class (Abbrev.)	Examples
Wall (WA)	Buildings, walls, fences
Curb (CU)	CurbsCsmall steps
Line (LI)	Lane markings
Vegetation (VE)	Bushes, trees
Guard Rail (GR)	Guard rails, poles
Roadway (RO)	Roadways
Sidewalk (SI)	Sidewalks
Open spaces (OP)	Parking lots, open spaces

IV. Individual MAP Estimation

The first step of our method estimates the labels of individual superpixels on the basis of MAP estimation. Each superpixel s_i is assigned an estimated label l_i given robot position x and map M by

$$\hat{l}_i = \underset{l_i}{\operatorname{argmax}}\, P(l_i|c_i, x, M). \tag{1}$$

TABLE II
FEATURES USED IN CLASSIFICATION

Category	Features
Color (HSI)	HS 2D-histogram
	I histogram
Textures and edges	Edge density
	Edge direction histogram
	Edge strength histogram
	Step edge density

Applying Bayes' theorem, eq. (1) is calculated as follows:

$$P(l_i|c_i, x, M) = \frac{P(c_i|l_i)P_{s_i}(l_i|x, M)}{P(c_i|x, M)} \quad (2)$$

$$\approx \frac{P(c_i|l_i)P_{s_i}(l_i|x, M)}{\sum_{l\in\mathcal{L}} P(c_i|l)P(l|x, M)}. \quad (3)$$

Assuming that the object classes of feature observations are independent of the robot position and map, we approximate $P(c_i|l_i, x, M) \approx P(c_i|l_i)$. Here $P(c_i|l_i)$ is the observation likelihood of superpixel s_i and $P_{s_i}(l_i|x, M)$ is the prior probability that object l_i appears at the location of s_i given prior information.

A. Prior Probability from Digital Street Map

Prior information consists of a map M, environmental knowledge, and the robot position x. From this information, we calculate the prior probabilities $P_s(l|x, M)$.

Existing digital street maps are input as two-dimensional grid maps. We assume that the maps contain semantic annotations. Each cell in the grid map is assigned one of the following labels: *roadway*, *sidewalk*, *building and others*, and boundaries between them.

Objects in the environment are nonuniformly distributed; for example, at boundaries between sidewalks and roadways, curbs are more common than walls. This type of knowledge is built into probability. The probabilities used in the current experiments are listed in Table. III. This table is based on the frequencies of objects in each map symbol in the training data set (described in section IV-B).

The probabilities of object presence in map M are projected onto an image plane using the robot position x. We assume that the robot navigates on a flat surface and that the pose of the camera relative to the robot is given. The map cell corresponding to an image pixel p is denoted by $M_{p,x}$. The probability that object l appears at the location of pixel p is given by

$$P_p(l|x, M) = P(l|M_{p,x}), \quad (4)$$

and is retrieved from the abovementioned probability table.

We can now calculate the probability that object l exists at the location of superpixel s by averaging the pixel-wise probabilities

$$P_s(l|x, M) = \frac{1}{|s|}\sum_{p\in s} P_p(l|x, M). \quad (5)$$

Here $|s|$ is the number of pixels in the superpixel s. An example of a prior probability is shown in Fig. 3. As seen in the figure, the probability of *curb* is high at the roadway-sidewalk boundary, while that of *wall* is high at the sidewalk-building boundary.

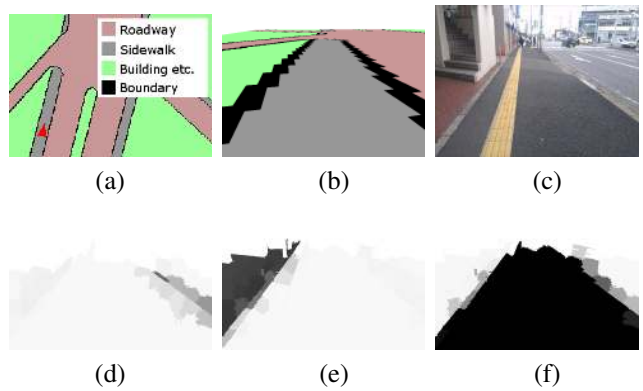

Fig. 3. Example of prior probabilities. (a) Map showing robot position (red triangle). (b) Map projected onto the image plane. (c) Actual images captured at the position. Calculated prior probabilities of curb (d), wall (e) and sidewalk (f). Darker colors indicate higher probability.

B. Observation likelihood using SVM

The observation likelihood $P(c|l)$ is calculated using an SVM that is pretrained using training data sets. To train the SVM, superpixels are extracted from images and manually labeled.

For each superpixel in the test images, multiclass classification is performed and the classification probabilities $P(l|c)$ are estimated [20]. By Bayes' theorem, we obtain

$$P(c|l) \propto \frac{P(l|c)}{P(l)}, \quad (6)$$

which is used to evaluate eq. (1). The prior probability $P(l)$ is based on the frequency of the object class in the training data sets.

V. ENERGY MINIMIZATION BY GRAPH CUTS

Image segmentation can be formulated as energy minimization [19]. In the previous section, labels of superpixels were assumed to be independent and were individually estimated. Energy minimization accounts for the interactions between neighboring superpixels and identifies globally optimal labels. Two benefits are expected: noise reduction by a smoothing factor and boundary detection. The energy is defined as follows:

$$E(L; x) = \sum_{i}^{n} G_i(l_i) + \lambda \sum_{(i,j)\in N} H_{ij}(l_i, l_j). \quad (7)$$

Here N is a set of neighboring superpixel pairs. The data term G_i is calculated from observation likelihood and the prior probabilities calculated using prior information. The interaction term H_{ij} is calculated by feature differences between neighboring superpixels and edge strengths at the boundary of the superpixels.

TABLE III

EXAMPLE OF PRIOR PROBABILITY

Map symbols	Object classes							
	SI	RO	OP	LI	CU	WA	GR	VE
Sidewalk	0.87931	0.017241	0.017241	0.017241	0.017241	0.017241	0.017241	0.017241
Roadway	0.017241	0.87931	0.017241	0.017241	0.017241	0.017241	0.017241	0.017241
Buildings	0.019231	0.019231	0.115385	0.019231	0.019231	0.480769	0.019231	0.307692
Sidewalk-roadway boundary	0.016667	0.016667	0.016667	0.183333	0.433333	0.033333	0.216667	0.083333
Building-sidewalk boundary	0.023256	0.023256	0.023256	0.046512	0.046512	0.511628	0.023256	0.302326
Roadway-building boundary	0.028571	0.028571	0.028571	0.114286	0.085714	0.514286	0.057143	0.142857

A. Data Term

The data term is defined as follows:

$$G_i = -\log P(l_i|c_i, x, M). \tag{8}$$

This term penalizes the energy based on the consistency between observation and prior information. The calculation is described in section IV.

B. Interaction Term

The interaction term penalizes the energy by the likelihood that neighboring superpixels belong to different classes, based on two cues. One is the feature difference f between the superpixels. Features in the same class should be similar, while those in different classes should be dissimilar. The other factor is edge strength at the boundary of neighboring superpixels. Different classes are often delineated by a strong edge. In our method, edges are detected on the image using a Canny detector. The average edge strength of the pixels constructing the boundary of superpixels i and j is denoted by e_{ij}. The strength is normalized to the range $(0, 1)$ by a sigmoidal function ς. Using these definitions, H_{ij} is calculated as follows:

$$H_{ij}(l_i, l_j) = \delta_{i,j} \cdot f(c_i, c_j) \cdot \varsigma(e_{ij}) \tag{9}$$

$$\delta_{i,j} = 1 \text{ (if } l_i \neq l_j), \text{otherwise } 0 \tag{10}$$

$$f(a, b) = \exp\{-\frac{||a-b||^2}{2\sigma^2}\} \tag{11}$$

$$\varsigma(x) = \frac{1}{1+e^{-a(-x+x_0)}}, \tag{12}$$

where the parameters σ, a, and x_0 are determined experimentally.

C. Labeling Using Graph Cuts

In previous studies, energy minimization problems have been solved by simulated annealing and ICM [21]. In the past few years, graph cuts have become very popular because it is computationally efficient and it can return exact solutions to certain energy minimization problems. Although our multilabeling problem is considered as NP-hard, therefore cannot be solved easily, several methods based on graph cuts provide approximate solutions [22]. We employ $\alpha - \beta$ swap to find a set of labels that approximately minimize the energy. The initial labels are estimated as described in section IV. Next, we extract superpixels from two classes $\alpha, \beta \in \mathcal{L}$ and use graph cuts to decide whether the labels of these superpixels should be swapped. The swap is repeated for all combinations of α and β. sD

VI. ITERATIVE PROCEDURE FOR SIMULTANEOUS CLASSIFICATION AND ROBOT LOCALIZATION

While prior information largely benefits recognition, erroneous prior information can lead to false recognition. Robot position contributes significantly to the information, but accurate robot position is not always known. Therefore, we correct the position errors of the robot. We extend the energy minimization procedure to simultaneously estimate the robot position and object classes.

Fig. 4 shows an example of energy distribution with respect to position errors. The energy is calculated by eq. (7), using the classification results by the method described in section V. In the graph, the global minimum is found around the position error of 0 m and 0 °, and several local minima appear. In seeking the global minimum, naive searching over space is computationally expensive because numerous graph cuts operations are required. Therefore, for this purpose we have developed a more efficient search algorithm.

Our method iterates the following two steps:

a. Labeling step

Given the current robot position x_k, estimate labels via graph cuts that minimize the energy.

$$L_k = \underset{L}{\operatorname{argmin}}\, E(L; x_k) \tag{13}$$

b. Localization step

For fixed labels L_k, estimate x that minimizes the energy using nonlinear optimization.

$$x_{k+1} = \underset{x}{\operatorname{argmin}}\, E(L_k; x) \tag{14}$$

In the labeling step, labels are estimated as described in section V using the current position x_k, to obtain the labeling result L_k. In the localization step, labels remain fixed and the position x is updated to minimize the energy. The result x_{k+1} is used in the next labeling step. When the robot is navigating a street, the map is nondiscriminative in the longitudinal direction (except at intersections). Therefore, we estimate only the lateral displacement and orientation. In the localization step, the optimal x is determined by the Nelder-Mead method. By iterating these two steps, we can obtain (at least locally) energy-minimized labeling and the robot position. However, our method is prone to trapping in local minima, depending on the initial robot position.

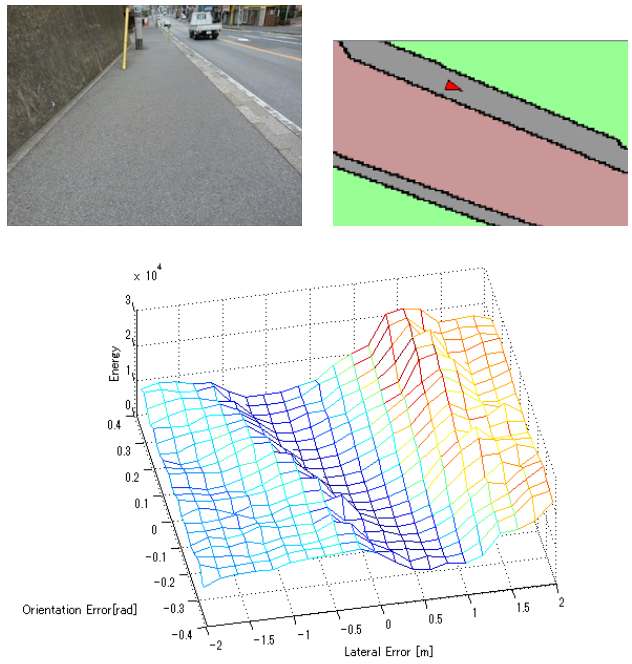

Fig. 4. Example of energy distribution. Top: input image and robot position on the map. Bottom: energy distribution with respect to position errors.

VII. Experiments

Our method was tested on the data set collected in crowded areas of Narashino city. A digital still camera (Panasonic FZ-100) was mounted onto an electric wheel-chair robot, as shown in Fig. 5. The robot was manually navigated along a 2 km pathway in Narashino city, capturing an image every 20 s. The system was evaluated on 82 images.

Superpixels were extracted from each image and manually labeled. Objects beyond the scope of our method (Table. I) such as pedestrians and cars were unlabeled and excluded from evaluation. The SVM was trained from 91 images that were captured on a path not overlapping the test set. The robot position was manually input by referring to the GPS logs. The grid map was generated using an image captured from Google Map (Fig. 5). The cell size (23.5cm) matched the distance per pixel of the Google Map.

A. Classification using Prior Information

The classification results at correct robot positions are shown in Fig. 6. Panel (a) shows misclassifications of a manhole cover and vegetation, which were corrected by the proposed method. Roadways in (b), which are visually similar to sidewalks, were correctly recognized by the proposed method. Vanishing point detection is not likely effective on (c) because no parallel road boundaries were found in the image. However, our method successfully estimated the boundary between the wall and the sidewalk. The parking lot, which appears on the right of the image in (d), appears very similar to the sidewalk. The position of the boundary line between these features was corrected by graph cuts. However, in some situations (e), graph cuts led to oversmoothing.

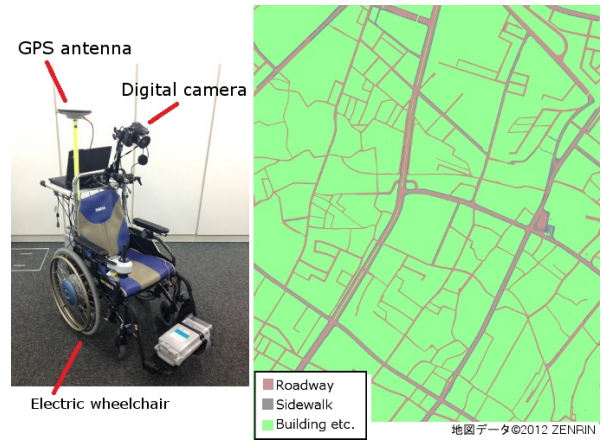

Fig. 5. Wheel chair robot and map used in the experiments.

TABLE IV
Classification accuracy

Method	Accuracy
SVM	45.6 %
Individual MAP	72.5 %
Graph Cuts	78.6 %

In (f), the detection of road boundary failed because of significant road width error in the map.

The classification accuracies are summarized in Table. IV. In the absence of prior information, SVM correctly classified 45.6% of superpixels. Using the method described in section IV, the classification accuracy was improved to 72.5%. The graph cuts introduced in section V improved the accuracy further to 78.6%.

Confusion matrices are shown in Fig. 7. Without prior information, significant confusion appears among *sidewalk*, *roadway*, and *wall* classes. The proposed method shows a large improvement in the ability to distinguish these classes.

B. Simultaneous Classification and Localization

We next evaluated the effectiveness and robustness of our method against position errors. We introduced numerous artificial errors to the position used as prior information and compared the classification performance with and without localization. Tests were conducted using translational errors of 50, 75, and 100 cm, orientation errors of 5°, 10°, and 15°, and combinations of these. The results are shown in Fig. 8. Although erroneous position information degrades the classification accuracy, localization provides large improvements in tests with larger errors.

Fig. 9 shows details of the test run under translational and orientation errors of 75 cm and 10°. Although position errors in (g), (h), (i), and (j) are successfully corrected. Typical failures are shown in (k) and (l). In (k), Braille blocks were wrongly detected as a road boundary. As evident in the graph, the position errors were slightly decreased but our method became trapped in a local minimum. This problem may be solved by a randomized initialization approach [23]. In (l), the orientation error remained large because small translations and rotations are not easily distinguished from

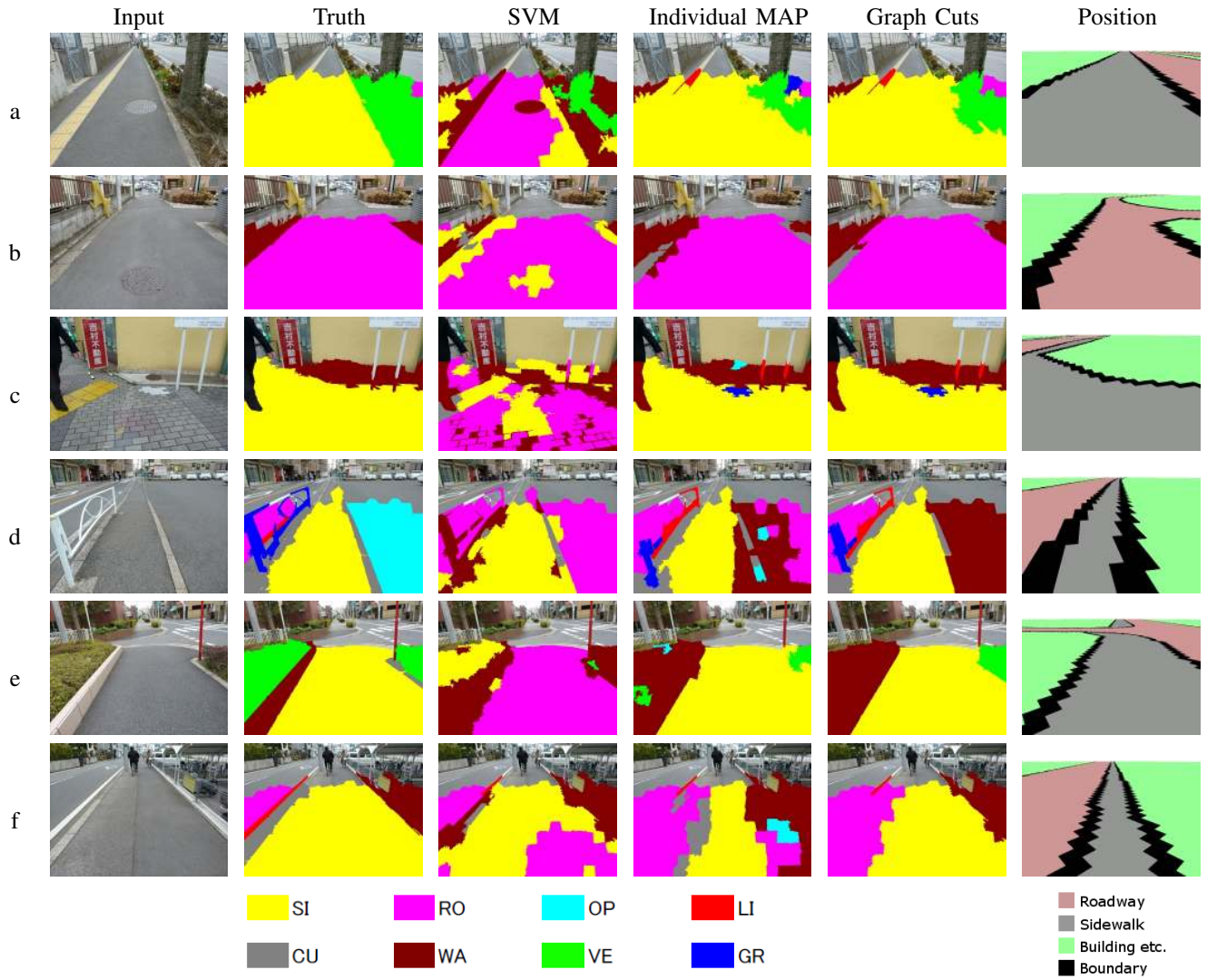

Fig. 6. Results of classification experiments. The rightmost column shows simulated view at the given position.

a single image. In this case, since the road boundaries were detected properly, the error is not be detrimental for navigation.

VIII. Conclusions

In this study, we proposed a road recognition method utilizing prior information for mobile robot navigation in urban environments. We aimed to distinguish objects with similar appearances, such as roadways, sidewalks, and gray walls from a single image. To equip the robot for recognizing such poorly distinguishable objects, we used digital street maps and the robot position. While prior information significantly helps recognition, errors in prior information cause recognition failures. Our method estimated position errors in prior information and effectively corrected small errors, as verified in a series of experiments. Correction of road width errors, which also constitute prior information errors, will be handled in future work.

References

[1] J. Crisman and C. Thorpe, "SCARF: a color vision system that tracks roads and intersections," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 1, pp. 49–58, 1993.

[2] W. Wijesoma, K. Kodagoda, and A. Balasuriya, "Road-boundary detection and tracking using ladar sensing," *IEEE Trans. on Robotics and Automation*, vol. 20, no. 3, pp. 456–464, 2004.

[3] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. Bradski, "Self-supervised monocular road detection in desert terrain," in *Proceedings of Robotics: Science and Systems*, 2006.

[4] K. Irie and M. Tomono, "Localization and road boundary recognition in urban environments using digital street maps," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2012, pp. 4493–4499.

[5] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, "Road detection using support vector machine based on online learning and evaluation," in *IEEE Intelligent Vehicles Symposium (IV)*, 2010, pp. 256–261.

[6] R. Danescu and S. Nedevschi, "Probabilistic lane tracking in difficult road scenarios using stereovision," *IEEE Trans. on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 272–282, 2009.

[7] Y. Matsushita and J. Miura, "On-line road boundary modeling with multiple sensory features, flexible road model, and particle filter," *Robotics and Autonomous Systems*, vol. 59, no. 5, pp. 274–284, 2011.

[8] H. Kong, J. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Trans. on Image Processing*, vol. 19, no. 8, pp. 2211–2220, 2010.

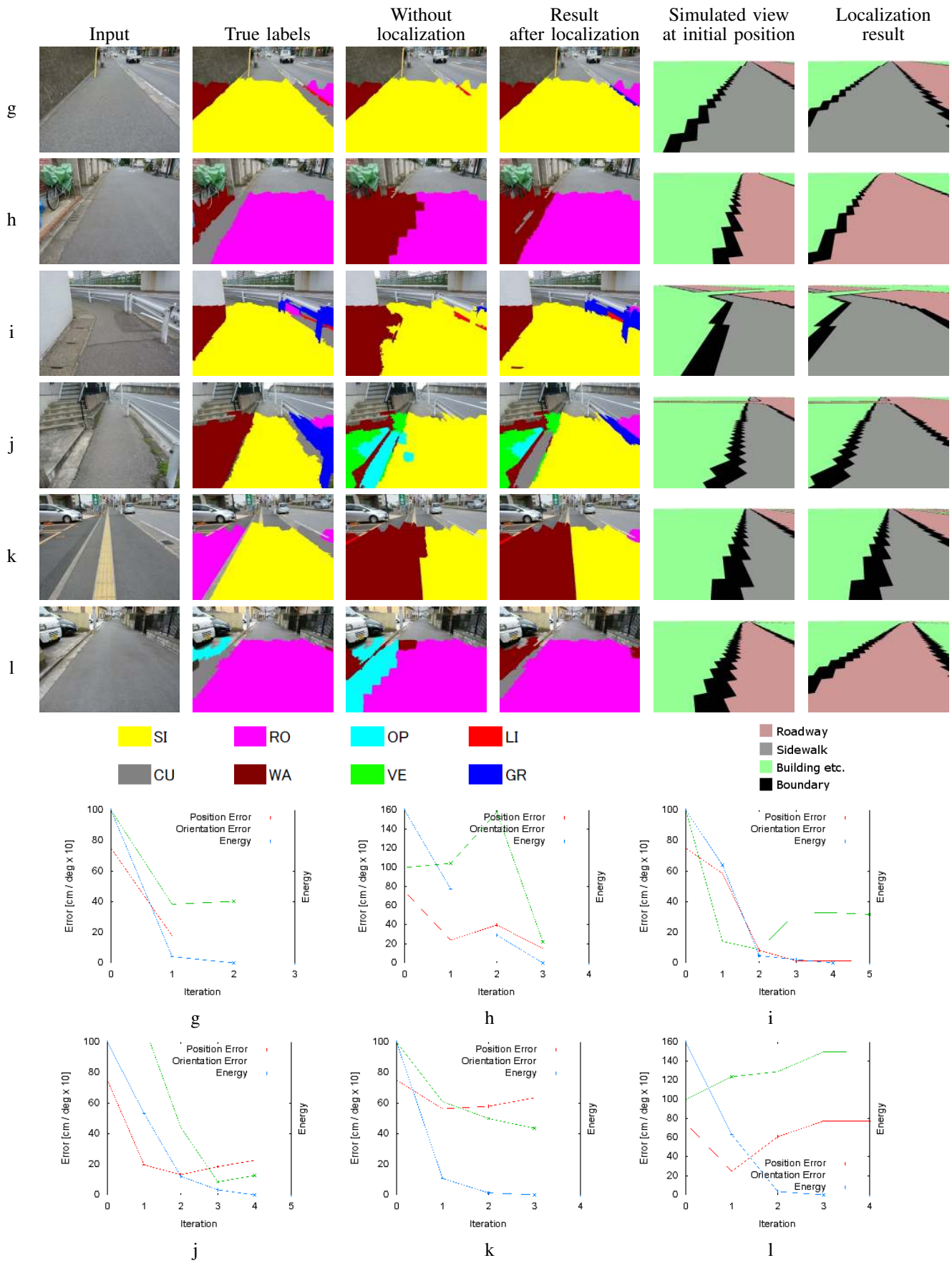

Fig. 9. Localization experiments with artificial position errors of 75 cm and 10°. Graphs in the bottom half of the figure plot the energy and position errors during localization.

Truth	Classification by SVM							
	SI	RO	OP	LI	CU	WA	GR	VE
SI	2594	2619	0	0	7	196	0	10
RO	476	1466	1	4	3	160	4	1
OP	132	146	0	2	4	46	0	4
LI	0	36	0	6	0	4	1	0
CU	408	71	0	1	14	153	0	0
WA	371	70	0	6	7	300	0	51
GR	13	81	0	0	3	85	13	0
VE	116	0	0	0	0	77	0	102

Truth	Classification by proposed method							
	SI	RO	OP	LI	CU	WA	GR	VE
SI	4853	276	23	12	74	122	6	60
RO	102	1912	0	20	24	35	21	1
OP	29	73	39	6	0	181	0	6
LI	0	21	0	21	0	0	5	0
CU	84	207	0	5	231	102	18	0
WA	84	21	50	12	3	515	15	105
GR	21	84	0	27	3	2	58	0
VE	52	4	1	0	25	83	1	129

Fig. 7. Confusion matrices

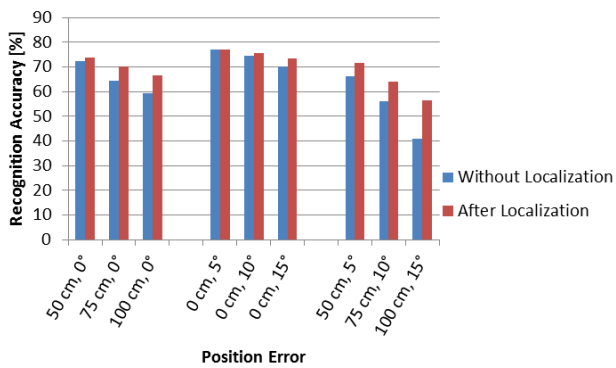

Fig. 8. Evaluation of classification accuracy with artificial position errors.

[9] B. Yu and A. Jain, "Lane boundary detection using a multiresolution hough transform," in *Int. Conf. on Image Processing*, vol. 2, 1997, pp. 748–751.

[10] J. Alvarez, F. Lumbreras, T. Gevers, and A. Lopez, "Geographic information for vision-based road detection," in *IEEE Intelligent Vehicles Symposium (IV)*, 2010, pp. 621–626.

[11] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive gmmrf model," in *Proc. of European Conference on Computer Vision*, 2004, pp. 428–441.

[12] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[13] B. Douillard, D. Fox, and D. Ramos, "Laser and vision based outdoor object mapping," in *Proc. of Robotics: Science and Systems IV*, 2008.

[14] I. Posner, M. Cummins, and P. Newman, "A generative framework for fast urban labeling using spatial and temporal context," *Autonomous Robots*, vol. 26, pp. 153–170, 2009.

[15] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2001, pp. 105–112.

[16] M. Hentschel and B. Wagner, "Autonomous robot navigation based on OpenStreetMap geodata," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems (ITSC)*, 2010, pp. 1645–1650.

[17] Y. Morales, T. Tsubouchi, and S. Yuta, "Vehicle 3D localization in mountainous woodland environments," in *Proc. of the IEEE Int. Conf. on Intelligent Robots & Systems (IROS)*, 2009, pp. 3588–3594.

[18] A. Radhakrishna, S. Appu, S. Kevin, L. Aurelien, F. Pascal, and S. Sabine, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[19] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1124–1137, 2004.

[20] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[21] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[22] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1222–1239, 2001.

[23] M. Tomono, "3D localization based on visual odometry and landmark recognition using image edge points," in *Proc. of the IEEE Int. Conf. on Intelligent Robots & Systems (IROS)*, 2010, pp. 5953–5959.