Available online at www.sciencedirect.com
SCIENCE DIRECT
Trans. Nonferrous Met. Soc. China 18(2008) 432−437
Transactions of Nonferrous Metals Society of China
www.csu.edu.cn/ysxb/

Scene recognition for mine rescue robot localization based on vision

CUI Yi-an(崔益安)[1,2], CAI Zi-xing(蔡自兴)[1], WANG Lu(王 璐)[1]

1. School of Information Science and Engineering, Central South University, Changsha 410083, China;
2. School of Info-Physics Engineering, Central South University, Changsha 410083, China

Received 18 April 2007; accepted 13 September 2007

Abstract: A new scene recognition system was presented based on fuzzy logic and hidden Markov model(HMM) that can be applied in mine rescue robot localization during emergencies. The system uses monocular camera to acquire omni-directional images of the mine environment where the robot locates. By adopting center-surround difference method, the salient local image regions are extracted from the images as natural landmarks. These landmarks are organized by using HMM to represent the scene where the robot is, and fuzzy logic strategy is used to match the scene and landmark. By this way, the localization problem, which is the scene recognition problem in the system, can be converted into the evaluation problem of HMM. The contributions of these skills make the system have the ability to deal with changes in scale, 2D rotation and viewpoint. The results of experiments also prove that the system has higher ratio of recognition and localization in both static and dynamic mine environments.

Key words: robot location; scene recognition; salient image; matching strategy; fuzzy logic; hidden Markov model

1 Introduction

Search and rescue in disaster area in the domain of robot is a burgeoning and challenging subject[1]. Mine rescue robot was developed to enter mines during emergencies to locate possible escape routes for those trapped inside and determine whether it is safe for human to enter or not. Localization is a fundamental problem in this field. Localization methods based on camera can be mainly classified into geometric, topological or hybrid ones[2]. With its feasibility and effectiveness, scene recognition becomes one of the important technologies of topological localization.

Currently most scene recognition methods are based on global image features and have two distinct stages: training offline and matching online.

During the training stage, robot collects the images of the environment where it works and processes the images to extract global features that represent the scene. Some approaches were used to analyze the data-set of image directly and some primary features were found, such as the PCA method[3]. However, the PCA method is not effective in distinguishing the classes of features. Another type of approach uses appearance features including color, texture and edge density to represent the image. For example, ZHOU et al[4] used multi-dimensional histograms to describe global appearance features. This method is simple but sensitive to scale and illumination changes. In fact, all kinds of global image features are suffered from the change of environment.

LOWE[5] presented a SIFT method that uses similarity invariant descriptors formed by characteristic scale and orientation at interest points to obtain the features. The features are invariant to image scaling, translation, rotation and partially invariant to illumination changes. But SIFT may generate 1 000 or more interest points, which may slow down the processor dramatically.

During the matching stage, nearest neighbor strategy(NN) is widely adopted for its facility and intelligibility[6]. But it cannot capture the contribution of individual feature for scene recognition. In experiments, the NN is not good enough to express the similarity between two patterns. Furthermore, the selected features can not represent the scene thoroughly according to the state-of-art pattern recognition, which makes recognition not reliable[7].

So in this work a new recognition system is presented, which is more reliable and effective if it is used

Foundation item: Project(60234030) supported by the National Natural Science Foundation of China; Project(A1420060159) supported by the Basic Research Program of the 11th Five-Year-Plan of China
Corresponding author: CUI Yi-an; Tel: +86-731-8877075; E-mail: csu-iag@mail.csu.edu.cn

in a complex mine environment. In this system, we improve the invariance by extracting salient local image regions as landmarks to replace the whole image to deal with large changes in scale, 2D rotation and viewpoint. And the number of interest points is reduced effectively, which makes the processing easier. Fuzzy recognition strategy is designed to recognize the landmarks in place of NN, which can strengthen the contribution of individual feature for scene recognition. Because of its partial information resuming ability, hidden Markov model is adopted to organize those landmarks, which can capture the structure or relationship among them. So scene recognition can be transformed to the evaluation problem of HMM, which makes recognition robust.

2 Salient local image regions detection

Researches on biological vision system indicate that organism (like drosophila) often pays attention to certain special regions in the scene for their behavioral relevance or local image cues while observing surroundings[8]. These regions can be taken as natural landmarks to effectively represent and distinguish different environments. Inspired by those, we use center-surround difference method to detect salient regions in multi-scale image spaces. The opponencies of color and texture are computed to create the saliency map.

Input is provided in the form of static color image named as G_0. Multi-scale image spaces G_1−G_4 (1:1 to 1:64) are created by Eqns.(1) and (2).

$$G_{n0}=w*G_{n-1} \tag{1}$$

$$G_n=\text{Subsampled}\quad G_{n0}\quad n\in[1,4] \tag{2}$$

where w is a Gaussion low-pass filter, and "*" denotes convolution operation. Let centers are $\{G_1, G_2\}$ and surroundings are $\{G_3, G_4\}$, the definition of opponency among scales is the feature difference between centers and surroundings denoted by "Θ", which means that the surroundings are interpolated and then subtract the centers pixel by pixel.

To compute the desired color opponencies, it is necessary to convert the RGB space into RGBY space for emphasizing the opponencies of red/green and blue/yellow[9]. The space is calculated by

$$R=r-(g+b)/2$$

$$G=g-(r+b)/2$$

$$B=b-(r+g)/2$$

$$Y=(r+g)/2-|r-g|/2-b$$

So, the color opponencies are computed by

$$RG(c, s)=|(R(c)-G(c))\,\Theta\,(G(s)-R(s))|$$

$$BY(c, s)=|(B(c)-Y(c))\,\Theta\,(Y(s)-B(s))|$$

where $c\in$Centers, $s\in$Surroundings. $RG(c, s)$ denotes the opponency between red and green; $BY(c, s)$ denotes the opponency between blue and yellow.

To compute the texture opponencies, Gabor filter is selected because of its ability of acquiring local optimum either in the time domain or in the frequency domain. Researches on human psychophysics and vision physiology show that it resembles human attention mechanism[10].

Gabor filter is defined by $h(x, y)=g(x, y)e^{2\pi j(ux+vy)}$. Because Gabor filter is polar symmetric in the frequency domain, the orientation 0−π can cover the whole frequency domain. Generally only 4 orientations 0°, 45°, 90° and 135° are considered.

The texture of 4 orientations are computed by

$$T_\theta(x, y)=|G_n(x, y)*h_\theta(x, y)|$$

Now we can compute the texture opponencies by

$$T(c, s, \theta)=|T_{c\theta}(x, y)\,\Theta\,T_{s\theta}(x, y)|$$

where $c\in$Centers, $s\in$Surroundings, $\theta\in\{0°, 45°, 90°, 135°\}$.

Then all opponencies are combined according to Eqns.(3)−(5) to create the saliency map S. The definition of normalizing operator $N(\ )$ can be found in Ref.[9].

$$\overline{C}=\bigoplus_{c=1}^{2}\bigoplus_{s=3}^{4}[N(RG(c,s))+N(BY(c,s))] \tag{3}$$

$$\overline{T}=\sum_{\theta\in\{0°,45°,90°,135°\}}N\Big(\bigoplus_{c=1}^{2}\bigoplus_{s=3}^{4}T(c,s,\theta)\Big) \tag{4}$$

$$S=w_1\overline{C}+w_2\overline{T} \tag{5}$$

where w_1 and w_2 are weights that denote the significance of color and texture. We also design an algorithm using LMS to learn w_1 and w_2 offline. Fig.1(b) shows the saliency obtained landmark regions where the position with more lightness has more salience.

Follow-up, sub-image centered at the salient position in S is taken as the landmark region. The size of the landmark region can be decided adaptively according to the changes of gradient orientation of the local image[11].

Mobile robot navigation requires that natural landmarks should be detected stably when environments change to some extent. To validate the repeatability on landmark detection of our approach, we have done some experiments on the cases of scale, 2D rotation and viewpoint changes etc. Fig.2 shows that the door is detected for its saliency when viewpoint changes. More detailed analysis and results about scale and rotation can be found in our previous works[12].

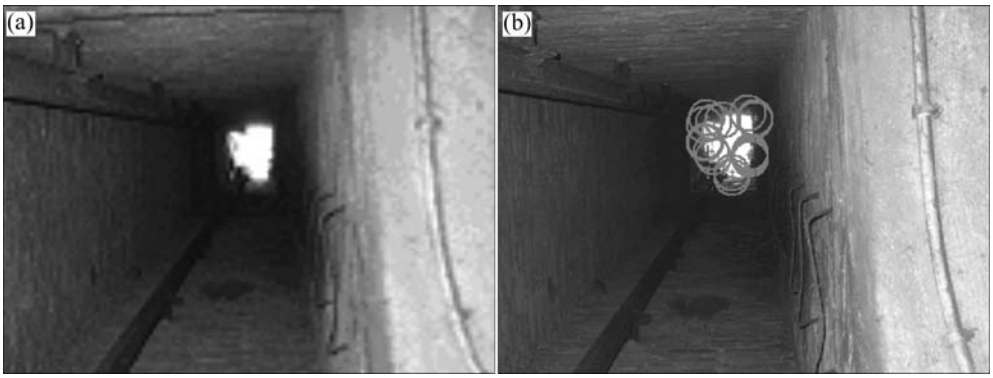

Fig.1 Saliency detection on real mine images: (a) Original image, (b) Obtained landmark regions

Fig.2 Experiment on viewpoint changes

3 Scene recognition and localization

Different from other scene recognition systems, our system doesn't need training offline. In other words, our scenes are not classified in advance. When robot wanders, scenes captured at intervals of fixed time are used to build the vertex of a topological map, which represents the place where robot locates. Although the map's geometric layout is ignored by the localization system, it is useful for visualization and debugging[13] and beneficial to path planning. So localization means searching the best match of current scene on the map. In this paper hidden Markov model is used to organize the extracted landmarks from current scene and create the vertex of topological map for its partial information resuming ability.

Resembled by panoramic vision system, robot looks around to get omni-images. From each image, salient local regions are detected and formed to be a sequence, named as landmark sequence whose order is the same as the image sequence. Then a hidden Markov model is created based on the landmark sequence involving k salient local image regions, which is taken as the description of the place where the robot locates. In our system EVI-D70 camera has a view field of $\pm 170^\circ$. Considering the overlap effect, we sample environment every 45° to get 8 images.

Let the 8 images as hidden state S_i ($1 \leqslant i \leqslant 8$), the created HMM can be illustrated by Fig.3. The parameters of HMM, a_{ij} and b_{jk}, are achieved by learning, using Baulm-Welch algorithm[14]. The threshold of convergence is set as 0.001.

As for the edge of topological map, we assign it

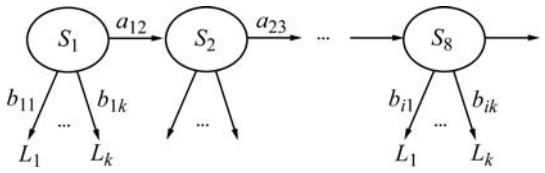

Fig.3 HMM of environment

with distance information between two vertices. The distances can be computed according to odometry readings.

To locate itself on the topological map, robot must run its 'eye' on environment and extract a landmark sequence $L'_1 - L'_k$, then search the map for the best matched vertex (scene). Different from traditional probabilistic localization[15], in our system localization problem can be converted to the evaluation problem of HMM. The vertex with the greatest evaluation value, which must also be greater than a threshold, is taken as the best matched vertex, which indicates the most possible place where the robot is.

For evaluation, firstly we must prepare an observation sequence V^{T} from $L'_1 - L'_k$ as below.

$V^{\mathrm{T}}=\{L_i|L_i \in$ vertex to be recognized and matches one of $L'_1 - L'_k\}$

To compute the similarity between L_i and L'_i, we design a new match strategy based on fuzzy logic, which can be found in the following section. Once the observation sequence is generated, evaluation process is to compute a posterior probabilistic value $P(V^{\mathrm{T}})$ according to Eqn.(6), where ω_r^{T} is a set of hidden state sequence and r is the index of a special hidden state sequence[3]. The detailed analysis and algorithm can be found in Ref.[4].

$$P(V^{\mathrm{T}}) = \sum_{r=1}^{r_{\max}} P(V^{\mathrm{T}} \left| \omega_r^{\mathrm{T}} \right.) P(\omega_r^{\mathrm{T}}) \tag{6}$$

4 Match strategy based on fuzzy logic

One of the key issues in image match problem is to choose the most effective features or descriptors to represent the original image. Due to robot movement, those extracted landmark regions will change at pixel level. So, the descriptors or features chosen should be invariant to some extent according to the changes of scale, rotation and viewpoint etc. In this paper, we use 4 features commonly adopted in the community that are briefly described as follows.

GO: Gradient orientation. It has been proved that illumination and rotation changes are likely to have less influence on it[5].

ASM and ENT: Angular second moment and entropy, which are two texture descriptors.

H: Hue, which is used to describe the fundamental information of the image.

Another key issue in match problem is to choose a good match strategy or algorithm. Usually nearest neighbor strategy (NN) is used to measure the similarity between two patterns. But we have found in the experiments that NN can't adequately exhibit the individual descriptor or feature's contribution to similarity measurement. As indicated in Fig.4, the input image Fig.4(a) comes from different view of Fig.4(b). But the distance between Figs.4(a) and (b) computed by Jefferey divergence is larger than Fig.4(c).

To solve the problem, we design a new match algorithm based on fuzzy logic for exhibiting the subtle changes of each features. The algorithm is described as below.

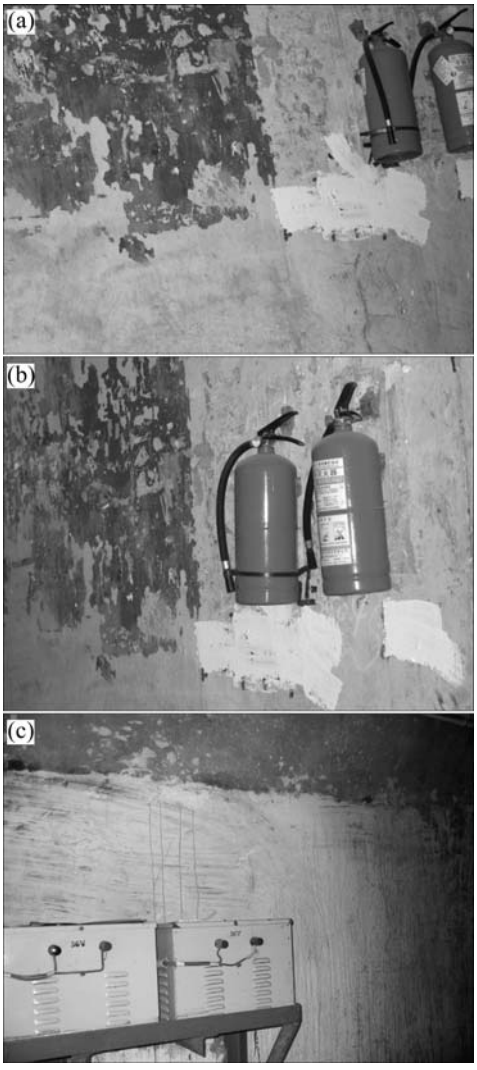

Fig.4 Similarity computed using Jefferey divergence: (a) Input image; (b) Landmark in database whose index is 6 and Jefferey divergence d=7.305 2; (c) Landmark in database whose index is 14 and its Jefferey divergence d=4.662 6

1) First all features are fuzzifyid as below.

$$\mu_{\mathrm{ASM}}=\sum_{k=0}^{k=1}P^2(k)\langle\mu_k\rangle$$

$$\mu_{\mathrm{ENT}}=-\sum_{k=0}^{k=1}P(k)\lg[P(k)]\langle\mu_k\rangle$$

where $P(k)=\dfrac{N_k}{N_{\mathrm{pixels}}}$, $\langle\mu_k\rangle=\dfrac{1}{N_k}\sum_{l=1}^{N_k}\mu_l$, and $\mu_l=e^{-|A_{ij}-k|}$.

$\mathrm{GO}_{ij}=\arctan(A_{ij}-A_{i+1,j}, A_{i,j+1}-A_{ij})$

$$\mu_{\mathrm{GO}}=\frac{1}{M}\sum_{k=0}^{M-1}mP_{\mathrm{GO}}(m)\langle\mu_m\rangle$$

$$H_{ij}=\arctan\left(\frac{\sqrt{3}(G_{ij}+B_{ij})}{2R_{ij}-G_{ij}-B_{ij}}\right)$$

$$\mu_{\mathrm{H}}=\frac{1}{M}\sum_{k=0}^{M-1}mP_{\mathrm{H}}(m)\langle\mu_m\rangle$$

where $P_{\mathrm{GO}}(m)=\dfrac{N_{m\_\mathrm{GO}}}{N_{\mathrm{pixels}}}$, $P_{\mathrm{H}}(m)=\dfrac{N_{m\_\mathrm{H}}}{N_{\mathrm{pixels}}}$, $\langle\mu_m\rangle=\dfrac{1}{N_m}\sum_{q=1}^{N_{m\_\mathrm{GO\_h}}}\mu_q$, and $\mu_q=\mathrm{e}^{-|\mathrm{GO}_{ij}(H_{ij}\text{ for }H)-m|}$.

In these equations N_k represents the number of pixels with gray level k, N_{pixels} the total number of pixels of the image, N_{m_GO} the number of pixels with angle degree m in $\{\mathrm{GO}_{ij}\}$, N_{m_H} in $\{H_{ij}\}$. A_{ij} represents the gray value of the pixel, and $\langle\mu_k\rangle$ the averaged degree attributed through the fuzzy classification to the gray level k, $\langle\mu_m\rangle$ to the angular degree m. k is equal to 256 and m is equal to 360.

2) The similarity between two landmarks is computed using individual feature, respectively. The similarity degree about the lth feature among the fuzzy set {ASM, ENT, GO, H} is defined as

$$r_{l\_ij}=\mathrm{e}^{\frac{-2|\mu_{\mathrm{feature\_}l}(i)-\mu_{\mathrm{feature\_}l}(j)|}{\mu_{\mathrm{feature\_}l}(i)+\mu_{\mathrm{feature\_}l}(j)}}$$

Then we compare the local image with every one in the database. $R_{\max}$ and r_{mean} are recorded.

3) All similarity degrees of each feature are fused to obtain a judgement, which can be formalized by Eqn.(7).

$$J=\bigcup_{l=1}^{4}r_{l\_ij}=\sum_{l=1}^{4}w_l r_{l\_ij} \tag{7}$$

The weights w_l are decided according to $r_{\max}-r_{\mathrm{mean}}$ of each feature. The deviations are sorted, then w_l is assigned to be 0.4, 0.3, 0.2, 0.1, respectively according to the order.

And the landmark in the database whose fused similarity degree is higher than any others is taken as the best match. The match results of Figs.4(b) and (c) are demonstrated by Fig.5. As indicated, this method can measure the similarity effectively between two patterns.

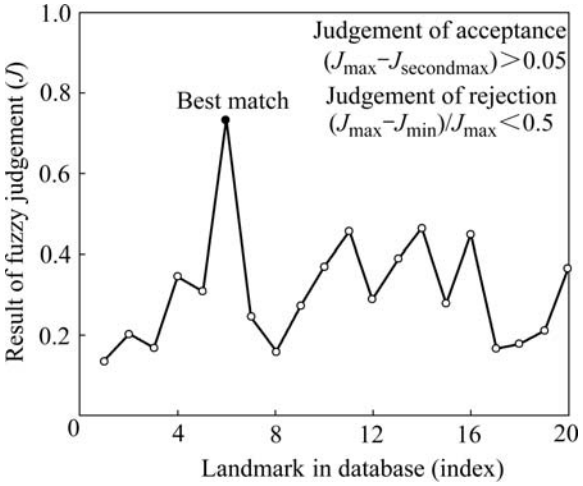

Fig.5 Similarity computed using fuzzy strategy

5 Experiments and analysis

The localization system has been implemented on a mobile robot, which is built by our laboratory. The vision system is composed of a CCD camera and a frame-grabber IVC-4200. The resolution of image is set to be 400×320 and the sample frequency is set to be 10 frames/s. The computer system is composed of 1 GHz processor and 512 M memory, which is carried by the robot. Presently the robot works in indoor environments.

Because HMM is adopted to represent and recognize the scene, our system has the ability to capture the discrimination about distribution of salient local image regions and distinguish similar scenes effectively. Table 1 shows the recognition result of static environments including 5 laneways and a silo. 10 scenes are selected from each environment and HMMs are created for each scene. Then 20 scenes are collected when the robot enters each environment subsequently to match the 60 HMMs above.

In the table, "truth" means that the scene to be localized matches with the right scene (the evaluation value of HMM is 30% greater than the second high evaluation). "Uncertainty" means that the evaluation value of HMM is greater than the second high evaluation under 10%. "Error match" means that the scene to be localized matches with the wrong scene. In the table, the ratio of error match is 0. But it is possible that the scene to be localized can't match any scenes and new vertexes are created. Furthermore, the "ratio of truth" about silo is

Table 1 Recognition results of static environments

Scene	Ratio of truth/%	Ratio of uncertainty/%	Ratio of error match/%
Laneway 1	90	10	0
Laneway 2	85	5	0
Laneway 3	95	5	0
Laneway 4	85	10	0
Laneway 5	90	10	0
Silo	70	15	0
Summation	85.83	9.17	0

lower because salient cues are fewer in this kind of environment.

In the period of automatic exploring, similar scenes can be combined. The process can be summarized as: when localization succeeds, the current landmark sequence is added to the accompanying observation sequence of the matched vertex un-repeatedly according to their orientation (including the angle of the image from which the salient local region and the heading of the robot come). The parameters of HMM are learned again.

Compared with the approaches using appearance features of the whole image (Method 2, M2), our system (M1) uses local salient regions to localize and map, which makes it have more tolerance of scale, viewpoint changes caused by robot's movement and higher ratio of recognition and fewer amount of vertices on the topological map. So, our system has better performance in dynamic environment. These can be seen in Table 2. Laneways 1, 2, 4, 5 are in operation where some miners are working, which puzzle the robot.

Table 2 Recognition results of dynamic environments

Scene	Ratio of truth/%		Ratio of error match/%	
	M1	M2	M1	M2
Laneway 1	80	80	0	20
Laneway 2	60	40	10	40
Laneway 3	90	85	0	10
Laneway 4	75	75	10	10
Laneway 5	80	70	0	10
Silo	70	90	0	0
Summation	75.8	73.3	3.3	15

6 Conclusions

1) Salient local image features are extracted to replace the whole image to participate in recognition, which improve the tolerance of changes in scale, 2D rotation and viewpoint of environment image.

2) Fuzzy logic is used to recognize the local image, and emphasize the individual feature's contribution to recognition, which improves the reliability of landmarks.

3) HMM is used to capture the structure or relationship of those local images, which converts the scene recognition problem into the evaluation problem of HMM.

4) The results from the above experiments demonstrate that the mine rescue robot scene recognition system has higher ratio of recognition and localization.

Future work will be focused on using HMM to deal with the uncertainty of localization.

References

[1] QIAN Shan-hua, GE Shi-rong, WANG Yong-sheng, LIU Chang-qing. Research status of the disaster rescue robot and its applications to the mine rescue [J]. Robot, 2006, 28(3): 350−354.

[2] ULRICH I, NOURBAKHSH I. Appearance based place recognition for topological localization [C]// KHATIB Q. Proc of the IEEE International Conference on Robotics and Automation. San Francisco, 2000: 1023−1029.

[3] ARTAC M, JOGAN M, LEONARDIS A. Mobile robot localization using an incremental eigenspace model [C]// DUBEY R V. Proc of the IEEE International Conference on Robotics and Automation. Washington, 2002: 1025−1030.

[4] ZHOU C, WEI Y C, TAN T N. Mobile robot self-localization based on global visual appearance features [C]// LUO R C. Proc of the IEEE International Conference on Robotics and Automation. Taipei, 2003: 1271−1276.

[5] LOWE D. Object recognition from local scale invariant features [C]// BOYER K L. Proc of the International Conference on Computer Vision. Greece, 1999: 1150−1157.

[6] CHEN Yong-sheng, HUNG Yi-ping, YEN Ting-fang, FUH Chiou-shan. Fast and versatile algorithm for nearest neighbor search based on a lower bound tree [J]. Pattern Recognition, 2007, 40(2): 360−375.

[7] RAPTIS S N, TZAFESTAS S G. Fuzzy-probabilistic object recognition based on edge and texture descriptors fusion [C]// JENG Mu-de. IEEE International Conference on Systems, Man, and Cybernetics. Nashville, USA. 2000: 1621−1626.

[8] HEISENBERG M, WOLF R. Studies of brain function [M]. Berlin, Germany: Springer-Verlag Press, 1984.

[9] EDUARDO T, TORRAS C. Detection of nature landmarks through multiscale opponent features [C]// LOOG M. Proc of International Conference on Pattern Recognition. Barcelona, 2000: 3988−3991.

[10] SHENG W, XIA B. Texture segmentation method based on Gabor ring filtering [J]. Infrared and Laser Engineering, 2003, 32(5): 484−488.

[11] MIKOLAJCZYK K, SCHMID C. Scale and affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, 60(1): 63−86.

[12] WANG L, CAI Z X. Saliency based natural landmarks detection under unknown environments [J]. Pattern Recognition and Artificial Intelligence, 2006, 1(1): 52−56.

[13] GUILHERME N D, AVINASH C K. Vision for mobile robot navigation: A survey [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(2): 1−31.

[14] DUDA R, HART P, STORK D. Pattern classification [M]. 2nd ed. Beijing: China Machine Press, 2004.

[15] VALE A, RIBEIRO M I. A probabilistic approach for the localization of mobile robots in topological maps [C]// KOVACIC Z. Proc of the 10th Mediterranean Conference on Control and Automation. Lisbon, Portugal, 2002: 1244−1249.

(Edited by YUAN Sai-qian)