17354

Proceedings of ICAPP 2017
April 24-28, 2017 - Fukui and Kyoto (Japan)

MARKERLESS TRACKING IN NUCLEAR POWER PLANTS: A LINE SEGMENT-BASED APPROACH

Hirotake Ishii[*], Taro Kimura[*], Hiroki Tokumaru[*], Hiroshi Shimoda[*] and Yuya Koda[**]

[*]*Graduate School of Energy Science, Kyoto University, Yoshida Honmachi Sakyo, Kyoto, Kyoto, 606-8501*
{hirotake, kimura, tokumaru, shimoda}@ei.energy.kyoto-u.ac.jp
[**]*Fugen Decommissioning Engineering Center, Japan Atomic Energy Agency, Myojin-cho 3, Tsuruga, Fukui, 914-8510*
kouda.yuya@jaea.go.jp

To develop augmented reality-based support systems, a tracking method that measures the camera's position and orientation in real time is indispensable. A relocalization is one step that is used to (re)start the tracking. A line-segment-based relocalization method that uses a RGB-D camera and coarse-to-fine approach was developed and evaluated for this study. In the preparation stage, the target environment is scanned with a RGB-D camera. Line segments are recognized. Then three-dimensional positions of the line segments are calculated, and statistics of the line segments are calculated and stored in a database. In the relocalization stage, a few images that closely resemble the current RGB-D camera image are chosen from the database by comparing the statistics of the line segments. Then the most similar image is chosen using Normalized Cross-Correlation. This coarse-to-fine approach reduces the computational load to find the most similar image. The method was evaluated in the water purification room of the Fugen nuclear power plant. Results showed that the success rate of the relocalization is 93.6% and processing time is 45.7 ms per frame on average, which is promising for practical use.

I. INTRODUCTION

Nuclear power plants for which the designed lifetime has ended must be decommissioned carefully because of the risk of radioactive contamination. For that purpose, not only technical improvements of decommissioning but also effective support for workers is necessary. Augmented reality (AR) is a technology that superimposes three-dimensional (3D) computer graphics over a user's view[1]. Using AR, 3D information such as a distance between a wall and tank, the position of a dismantling target, and dangerous areas can be shown intuitively. To develop AR-based systems, a tracking method that measures 3D position and orientation (pose) of worker's view is indispensable. In most cases, 3D information is superimposed on a camera image. Therefore, the camera pose is measured instead of the worker's pose. Tracking methods of various kinds[2-4] have been developed in the past: marker-based, gyro-sensor-based, and magnetic-sensor-based. Among these methods, a natural-feature-based tracking method is the most applicable for use in nuclear power plants. To stabilize and increase the speed of the natural-feature-based tracking method, an assumption is made that the camera pose does not change much among sequential images. This assumption, however, does not hold when workers move quickly or when the camera lens is obstructed, as by a worker's hand. At the time of system startup, this assumption is not also valid. Therefore, the camera pose must be estimated using the current camera image only. A method that estimates the camera pose using only the current camera image is called "relocalization." A line-segment-based relocalization method that uses a RGB-D camera and coarse-to-fine approach was developed and evaluated for this study. Line segments are promising for use in filtering images because line segments are computationally light (6 degrees of freedom in 3D space) but very informative (they are distinguishable by its position, density, length, angle on both 2D image and 3D space) compared to point features (only position and density).

II. RELATED WORKS

To estimate the current camera pose, methods of several kinds are available. A camera pose can be estimated roughly using inertial sensors such as a gyro sensor and acceleration sensor attached to the camera. The estimated camera pose, however, includes much error accumulated during the passage of time. Therefore, some other method must be used simultaneously.

One desirable approach uses the camera image itself, which requires no additional sensor. Marker-based tracking, which estimates the relative pose between a camera and markers, is stable and accurate. However, pasting and measuring many markers in advance is onerous. Moreover, the markers might hinder work. Therefore, natural-feature-based tracking method is desirable. It uses visual features such as edges and corners existing in the work environment. The features are recognizable by image processing: a camera captures the work environment. Edges and corners on the image are recognized. Then the camera pose is estimated based on correspondence between 3D positions in the work environment and two-dimensional (2D) positions on the captured image[5,6]. To improve the natural-feature-based tracking performance (e.g., speed and accuracy), an assumption is made that the camera does not move quickly and that it always captures the environment. However, when this assumption does not hold, another method called "relocalization" must be used. The relocalization method estimates the camera pose using the current camera image only.

Some relocalization methods use a database that includes pairs of a shrunken camera image and a camera pose when the image is captured is built in advance. We can assume that the camera poses are close when captured images are almost identical. Therefore, the camera pose when a similar image was captured is useful as an estimate of the current camera pose. The representative methods that follow this strategy are randomized fern[7] and FAB-MAP[8]. The method using the randomized fern compares intensities at randomly chosen pixels to thresholds that are also generated randomly. Because the pixels used for comparison are much fewer than when all pixels are compared, the calculation speed is extremely fast. In contrast, a method using the FAB-MAP calculates the probability of the same feature points appearing in both the candidate image and the current image. These methods can roughly filter out candidate images from the database. However, to achieve accurate elimination, parameters of these methods must be adjusted drastically. In that case, the time spent tends to be large, especially in cases for which vastly numerous images are stored in the database.

III. PROPOSAL OF A RELOCALIZATION METHOD USING LINE SEGMENTS

III.A. Configuration of the proposed method

Nuclear power plants have many artificial objects such as tanks and pipes, with edges recognized by image processing as a collection of line segments. The sizes and angles of the pipes and tanks, as well as their combinations, differ depending on their location. Therefore, the possibility exists that the relocalization speed and accuracy can be improved using line segments: statistical indexes of the line segments are used to reduce the number of camera pose candidates. The proposal is categorized into a coarse-to-fine approach: a filter for which the computational load is very small but which can eliminate candidate images only roughly is applied in the early stage. Then another computationally expensive filter that can choose a small number of candidate images accurately is applied. The final camera pose is decided using the small number of remaining candidate images.

The proposed relocalization method consists of two stages as presented in Fig. 1: a preparation stage and a relocalization stage. For this study, a RGB-D camera is used, which can capture not only a color image (RGB image) but also a depth image. The depth image represents distances between the camera focal point and environment at each pixel of the RGB image. In the preparation stage, a database is developed for use in relocalization. In the relocalization stage, the current camera pose is estimated by choosing the most similar image from the database using a coarse-to-fine approach.

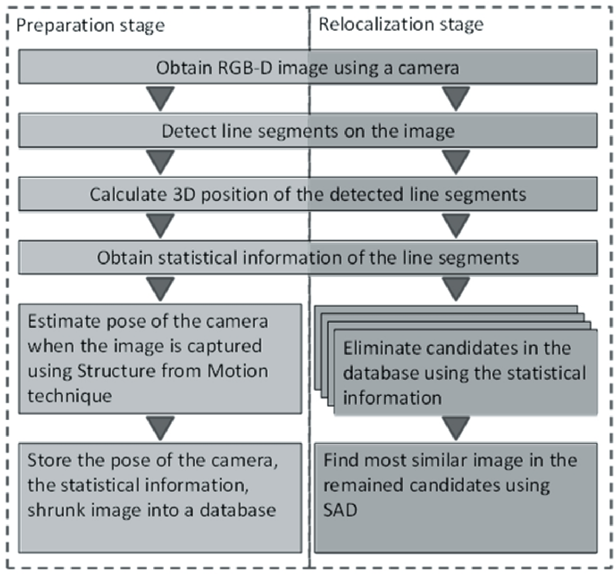

Fig. 1. Configuration of the proposed relocalization method.

III.B. Preparation stage

The left part of Fig. 1 shows the flow of database development in the preparation stage. First, RGB-D images are captured while the camera is moved very slowly. Then line segments are recognized using the Line Segment Detection (LSD) algorithm[9]. Figure 2 presents an example of some detected line segments using LSD. The 3D positions of the detected line segments are calculated using the depth image; then line segments with length greater than is threshold are chosen to reduce the number of line segments and to eliminate line segments that are recognized unstably. In this study, the thresholds are 0.1 m for 3D line segments and 20 pixels for 2D line segments. These thresholds were chosen empirically. Then the following statistical indexes are calculated.

Idx1. Number of line segments in the whole image

Idx2. Numbers of line segments in sub areas divided in advance

Idx3. Histogram of angles formed by pairs of 2D line segments in the image

Idx4. Histogram of angles formed by pairs of 3D line segments in the environment

Idx5. Histogram of lengths of 2D line segments in the image (unit is pixel)

Idx6. Histogram of lengths of 3D line segments in the environment (unit is millimeter)

Idx7. Histogram of distances between pairs of 2D line segments in the image (unit is pixel)

Idx8. Histogram of distances between pairs of 3D line segments in the environment (unit is millimeters)

Idx9. Number of line segments parallel to the vertical and horizontal axes

Here, distance between two line segments is defined as the shortest distance between infinite lines obtained by extending the line segments. The angle between 3D line segments is defined as the angle formed by two infinite lines: one infinite line is projected onto a plane defined by another line and a normal vector for which terminals are the shortest points between the two line segments. The poses of the camera, the statistical indexes, and shrunken images are stored in a database. In this proposal, it is assumed that the camera poses will be estimated using Structure from Motion methods which are very stable and accurate, but it requires much computation. It is therefore difficult to use for relocalization, which requires a quick response to avoid making the workers wait. However, it remains difficult to obtain camera poses that are sufficiently accurate to be used for this purpose. Further study is necessary.

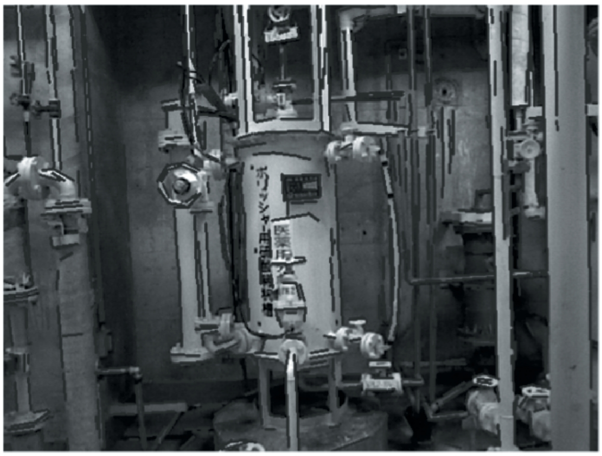

Fig. 2. Example of line segments detected using LSD.

III.C. Relocalization stage

The right part of Fig. 1 shows the relocalization flow. Line segments in the current camera image are recognized. Then its statistical indexes are calculated in the same way as in the preparation stage. Then the current camera image is compared to images stored in the database using the statistical indexes. First, all images stored in the database are nominated as candidate images. Then, candidates regarded as much different from the current camera image are filtered out. The remaining candidates are compared to the current image using another index.

For the indexes related to the number of line segments (Idx1, Idx2, and Idx9), if both the absolute difference of the numbers and the rate of the difference (absolute difference of the numbers / number of line segments in the current image) are greater than a threshold, the candidate image is regarded as much different from the current image. It is therefore eliminated from the candidate images.

For indexes related to the histogram (Idx3 – Idx8), the values of respective bins in the histogram are compared. If both the difference and the rate are larger than the threshold in at least one bin, then the candidate image is regarded as much different from the current image. It is eliminated from candidate images.

After filtering the candidate images using all indexes, the remaining candidate images are compared to the current image using the Sum of Absolute Differences (SAD). This comparison is conducted using shrunken images. Then the candidate image for which SAD is smallest is chosen as the most similar image.

In contrast to the case in which all images stored in the database are compared to the current image one by one using SAD (exhaustive comparison), the process to detect line segments in the current image is newly added. However, filtering using the statistical indexes is much faster than comparison using SAD. Therefore, the processing time for the proposed method is expected to be much less than that of exhaustive comparison when numerous images are stored in the database.

IV. EVALUATION

IV.A. Criteria and dataset used for evaluation

Ideally, 3D poses of a camera obtained by relocalization should be used to restart tracking. Then, the result should be compared with the ground truth obtained using some known reliable and accurate measurement method. However, it is difficult to use such a measurement method in nuclear power plants. Moreover, the results might differ according to the tracking method used. Therefore, we assume that the images for which the Normalized Sum of Squared Difference (NSSD) is smaller than a threshold are the correct images to be chosen using relocalization. Concretely, the relocalization methods were evaluated with two performance indexes: average time necessary for one relocalization and the success rate (successful trials / total trials). The relocalization is regarded as successful when the relocalization method chooses an image included in the right images chosen using NSSD. (This is the definition of "success" in this study.)

For this study, a dataset was built to include numerous images obtained using an RGB-D camera (XTion; ASUS Corp.) in a water purification room in the Fugen nuclear power plant. The image capture resolution was 320×240. Figure 3 shows a 22.4 m × 9.7 m room in which images were captured while moving the camera slowly (about 5 cm/s) and repeatedly to avoid image blur and to allow the camera to capture room information as densely as possible. In all, 236,552 images were captured.

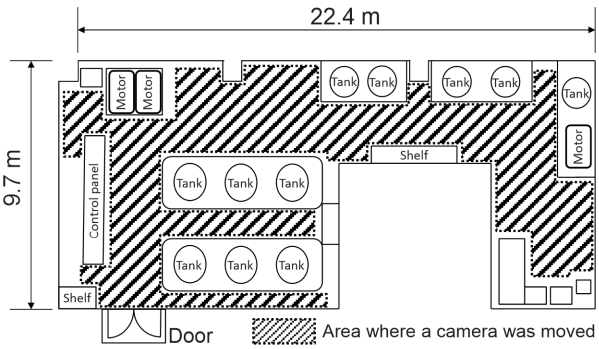

Fig. 3. Bird eye's view of a water purification room where images for the dataset were obtained.

As test images (which correspond to the on-line camera images), 1,000 images were chosen randomly from the dataset. These images were removed from the database so that the same images do not exist in the database, thereby reflecting the rarity of a camera capturing a completely identical image to one stored in the database. The right images (ground truth) for these 1,000 test images were chosen using NSSD (the threshold was set to 0.1).

IV.B. Order of filtering

As described above, nine indexes are used to filter out candidate images from the database. That performance will vary according to the order of the filtering because the computational load and accuracy differ among the indexes. The appropriate order of the filtering can be predicted to differ according to the target environment. To achieve the best performance of the proposed relocalization method, the order should be decided by application of the method to the entire target environment. It is not, however, realistic to choose the order using all images captured in the target nuclear power plant because the number of images is expected to be huge. Sometimes, it is difficult to conduct such preparation in advance. Therefore, the filtering order was decided for this study using a subset of images captured in the Fugen nuclear power plant. Concretely, we chose 5,000 RGB-D images randomly from the images to produce the subset and 100 ground truths. We used them to decide the filtering order. All filtering was applied to the subset images separately with their threshold parameters varied. The best filtering algorithm and a parameter which performed best (fastest under a condition that the remaining candidate images include at least one ground truth image) was chosen. Then the other filtering was applied to the remaining subset images. The best filtering index was chosen. The ranking of the filtering for each test image was recorded. Then the average ranking was calculated.

Table I presents the average ranking of the indexes. Filtering using the 2D angle histogram (Idx3) performed best, followed by the one using the 3D angle histogram (Idx4). However, filtering using the number of line segments (Idx1, Idx2 and Idx9) and distances between line segments (Idx7 and Idx8) performed worse. The performance is expected to be affected not only by the uniqueness of the index among different poses but also by the stability of the index when images are captured in similar poses. For this study, LSD, which is known as the state-of-the art method to detect line segments, was used to detect 2D line segments. However, even with this method, an edge that was detected as a single line segment in a certain frame was sometimes detected as two or more shorter line segments in other frames. This instability will adversely affect the performance of the indexes using the numbers and lengths of line segments.

TABLE I. Ranking of candidate image filtering performance

Index	Average ranking
(Idx3) 2D angle histogram	1.0
(Idx4) 3D angle histogram	2.0
(Idx6) 3D length histogram	3.2
(Idx5) 2D length histogram	4.1
(Idx9) Number of parallel lines	4.8
(Idx2) Number of segments in areas	6.3
(Idx8) 3D distance histogram	7.6
(Idx7) 2D distance histogram	7.7
(Idx1) Number of segments	8.3

However, the histogram of the angles will not be affected strongly by the instability because the angles are calculated between infinite lines obtained by extending the line segments, and it is expected the angles will not change much even if the line detection is not so stable.

IV.C. Evaluation conditions and results

Table II presents specifications of the PC and libraries used for the evaluation. Relocalization methods were implemented using OpenCV Version 2.4.9 to detect line segments using LSD. The CPU on the PC has physical multi-cores, but the relocalization methods were executed in a single thread. Enabling multithreads can certainly increase the speed because filtering can be parallelized easily. Enabling multithreads is a future work.

TABLE II. Specifications of PC and libraries used for evaluation

Item	Values
Processor	Intel Core i7-4700S 3.10 GHz
RAM	24GB
OS	Windows7 Professional 64 bit
IDE	Visual Studio 2010 C++
Library	OpenCV2.4.9, Boost 1.5.9

TABLE III. Time spent for filtering using each index

Process	Time [ms]	Ratio [%]
Detect line segments	8.9	19.5
2D angle histogram	12.0	26.3
3D angle histogram	4.9	10.7
3D length histogram	3.0	6.6
2D length histogram	0.9	2.0
Number of parallel lines	0.4	0.9
Number of segments in areas	2.0	4.4
3D distance histogram	0.4	0.9
2D distance histogram	0.4	0.9
Number of segments	0.2	0.4
SAD	12.8	28.0
Total	45.7	100

Figure 4 shows the number of remaining candidate images after filtering using each index in series. Table III presents the time spent for filtering using each index. As the figure shows, filtering using 2D angle histogram eliminates about 95% of the candidate images. Filtering using 3D angles further eliminates about half of the remaining candidate images. Because the processing time for SAD is much greater than these filters, the total processing time can be reduced drastically.

Table IV presents a comparison between relocalization methods using randomized fern and the proposed method. The performance of these relocalization methods varies according to the parameters used. When the thresholds of the filters for eliminating candidate images are loosen, the success rate will increase because more candidate images will pass through the filters and reach at the last filter: SAD. The SAD is computationally expensive but very accurate: the SAD will always choose the correct image if the correct image is included in the remaining candidate images. But of course, the time spent for the relocalization will increase drastically. In this way, the accuracy can be adjusted by changing the threshold for the filters. In this evaluation, the parameters were adjusted so that the relocalization success rates were the same. Consequently, the processing time for the proposed method is slightly shorter than that of the randomized fern. Figure 5 portrays the estimated processing time for relocalization per frame using the randomized fern and the proposed method against the number of images stored in a database. Because the proposed method must detect line segments in an image for every frame, which entails a high computational cost, the total processing time is greater than that of the randomized fern when the number of images stored in a database is small. However, when the number of images is large (in this case, greater than 180k images), the total processing time for the proposed method is less than that of the randomized fern because the processing time for detecting line segments is constant even if the number of images stored in the database changes. Therefore, when the images are numerous, the processing time for detecting line segments becomes small compared to the total processing time for the filtering. In contrast, the total processing time for the randomized fern increases linearly with the number of images stored in the database. Moreover, the processing time for one image stored in the database is less in the case of the proposed method than in the case of the randomized fern, which leads to performance reversal at around 180k images.

TABLE IV. Evaluation results of randomized fern and proposed method

Method	Success rate [%]	Process time per frame [ms]
Randomized fern	93.6	49.6
Proposed method	93.6	45.7

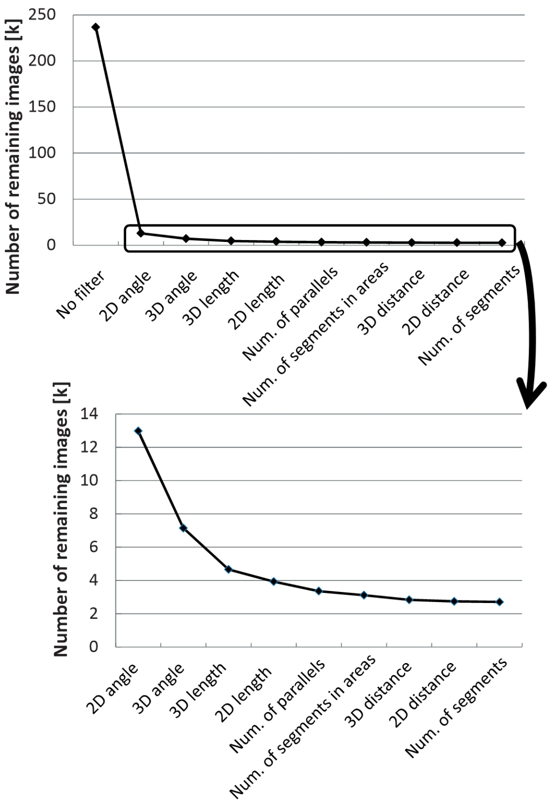

Fig. 4. Remaining candidate images after filtering with statistical indexes.

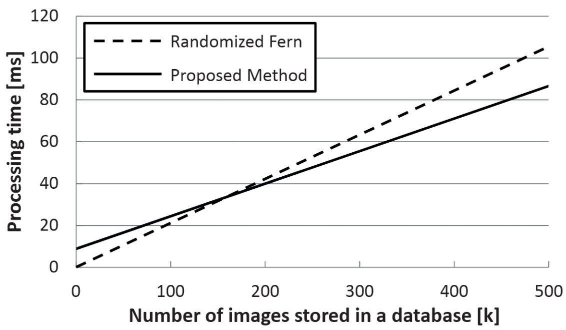

Fig. 5. Estimated processing time for relocalization per frame using randomized fern and the proposed method against the number of images stored in the database.

V. CONCLUSIONS

This study developed and evaluated line-segment-based relocalization using a coarse-to-fine approach. The method filters out candidate images stored in a database using nine filtering algorithms in the order that the computational load is smaller and more images can be eliminated in a shorter time. The order was chosen based on evaluation using images captured at the Fugen nuclear power plant. The first filtering algorithm is a 2D angle histogram comparison, which compares the distribution of angles between line segments detected in an image. The second filtering algorithm is a 3D angle histogram comparison. These results demonstrate that relative angles between line segments are obtainable more stably than the numbers or distances of line segments, which also implies that improvements of the line detection algorithm might alter the appropriate order of the filtering algorithms. Future studies might use some other index to improve the performance such as average depth and variation of image, combined with line segments. The combination of the line segment index and randomized fern might further improve the performance.

Performance evaluations in more complex and severe environments must be conducted in future work. For example, in some cases, semi-transparent sheets are put over the instruments, where line segments are hardly detected and the success rate will decrease even if the thresholds are set loosely.

Another future direction of this work is a development of a model based relocalization method by combining the relocalization method developed in this study and reconstruction models made by SfM (Structure from Motion) or scanned by a laser range finder. In this method, images are rendered using the reconstruction models from various points of view and its resultant images and their camera poses are stored into a database. This strategy will make it possible to conduct the relocalization even if the camera is lost where any images are not captured in advance, which means that the workers who use AR based system can move more freely than the case only using the images captured in advance.

REFERENCES

1. H. ISHII, "Augmented Reality: Fundamentals and Nuclear Related Applications", *International Journal of Nuclear Safety and Simulation*, Vol. 1, No.4, pp. 316–327 (2010).
2. H. Kato, M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system", *Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality*, pp. 85–94 (1999).
3. H. Ishii, W. Yan, S. Yang, H. Shimoda and M. Izumi, "Wide Area Tracking Method for Augmented Reality Supporting Nuclear Power Plant Maintenance Work", *International Journal of Nuclear Safety and Simulation*, Vol. 1, No. 1, pp. 45–51 (2010).
4. D. Han, D. G. Andersen, M. Kaminsky, K. Papagiannaki, S. Seshan, "Access Point Localization using Local Signal Strength Gradient", *Proceedings*

of International Conference on Passive and Active Network Measurement, pp. 99–108 (2009).
5. G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces”, *Proceedings of International Symposium on Mixed and Augmented Reality*, pp. 225–234 (2007).
6. A. J. DAVISON, I. D. REID, N. D. MOLTON, and O. STASSE, “MonoSLAM: Real-Time Single Camera SLAM”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 1052–1067 (2007).
7. B. GLOCKER, S. IZADI, J. SHOTTON, A. CRIMINISI, “Real-Time RGB-D Camera Relocalization”, *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 173–179 (2013).
8. M. CUMMINS and P. NEWMAN, “Appearance-only SLAM at Large Scale with FAB-MAP 2.0”, *The International Journal of Robotics Research*, Vol. 30, No. 9, pp. 1100–1123 (2010).
9. R. G. GIOI, J. JAKUBOWICZ, J. M. MOREL, G. RANDALL, “LSD: A Fast Line Segment Detector with a False Detection Control”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 4, pp. 722–732 (2010).