ICIC Express Letters
Part B: Applications
Volume **7**, Number **2**, February **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 483–490

CORNER-BASED ROBOTIC STEREO VISUAL NAVIGATION OPTIMIZED BY REGION-OF-INTEREST PYRAMID AND MAXIMUM-UNCERTAINTY COMPARABILITY MEASUREMENT

Chung-Lin Li and Chian C. Ho*

Department of Electrical Engineering
National Yunlin University of Science and Technology
No. 123, University Road, Sec. 3, Douliou, Yunlin 64002, Taiwan
*Corresponding author: futureho@yuntech.edu.tw

Received July 2015; accepted September 2015

Abstract. *Scale-Invariant Feature Transform (SIFT) is the most common feature detection and matching algorithm in binocular images for robotic Simultaneous Localization and Mapping (SLAM). This paper develops a SIFT-based robotic stereo visual navigation. Besides, this paper proposes region-of-interest pyramid to increase a substantial number of feasible features, and adopts maximum-uncertainty comparability measurement to reduce the significant computation time of searching candidate matching features. Experimental results indicate the proposed methods can improve original SIFT-based robotic stereo visual navigation by over 400% increment of feature density at the similar cost of computational time.*

Keywords: Robot, SIFT, Stereo visual navigation, Uncertainty

1. **Introduction.** Stereo vision cameras have recently become the fundamental equipment in intelligent robots, since it can provide richer contextual information in a complex environment and broader intelligent applications than low-resolution sonar sensor or laser range finder [1,2]. Through elaborate computer vision techniques, intelligent robots equipped with stereo vision cameras can achieve extensive artificial intelligence like visual creatures, such as visual navigation, surveillance recognition, autonomous patrol, emergency rescue, and unmanned vehicles.

In the methodology of the robotic stereo visual navigation, the feature detection and matching algorithm to obtain enough representative features is the primary step followed by Simultaneous Localization and Mapping (SLAM) based on Extended Kalman Filter (EKF), path planning, and motion control. Against lots of feature detection and matching algorithms for the robotic stereo visual navigation, such as SUSAN corner detector [3,4] and Harris corner detector [5,6], Scale-Invariant Feature Transform (SIFT) [7,8] is the most popularly-adopted one because of its characteristic invariant to image scale, illumination, rotation, partial occlusion, clutter, distortion. However, amount of features extracted by all aforementioned feature detection and matching algorithms, including SIFT, are not many enough and not even enough for finer 3D environmental modeling and mapping. Especially, SIFT is more computationally-expensive and time-consuming such that it is difficult to accomplish high-refreshing visual navigation or real-time intelligent applications.

Therefore, this paper proposes two methods to optimize the feature density and distribution of SIFT-based robotic stereo visual navigation with a little extra computational time. The organization of this paper is as follows. In the next section, SIFT-based robotic stereo visual navigation is developed by MATLAB toolkit and OpenCV library, and its methodology is illustrated step by step. Section 3 and Section 4 present the region-of-interest pyramid and maximum-uncertainty comparability measurement, respectively, to optimize the developed stereo visual navigation system in Section 2. Section 5 compares and analyzes the experimental results. Finally, Section 6 draws conclusions.

2. **SIFT-Based Robotic Stereo Visual Navigation.** Figure 1 illustrates the flowchart of SIFT-based robotic stereo visual navigation developed by this paper. The flow in Figure 1 must be continuously cycled. In the flowchart of Figure 1, the first stage is to take two tiny-disparity images from binocular cameras. Here, the binocular cameras have to be calibrated in advance for acquiring the intrinsic and extrinsic parameters of the binocular cameras. This is because these intrinsic and extrinsic parameters, such as focal length, principle point, rotation vector, and translation vector, are necessarily used and substituted into the binocular stereo vision algorithm of the third stage to evaluate the depth of all matched features. Then, it is possible to backproject the matched features into 3D points for building a stereo depth model and map [9,10]. In this self-developed system, this paper makes use of "Camera Calibration Toolbox for MATLAB" [11] for camera calibration and camera parameters. The detailed calibration procedure is shown in Figure 2.

The second stage is to detect and match representative features between two binocular images through SIFT algorithm. However, the features detected by original SIFT are often not distributed densely and evenly in high-contrast scenes as shown in Figure 3(a). In Figure 3(a), there are no detected features at all in the floor region in an indoor scene. Few and uneven features may be used to perform some object recognition or 2D computer vision, but cannot be sufficiently used to realize 3D EKF-based SLAM. Thus, this paper proposes a region-of-interest pyramid to overcome this issue rather than conventional active vision, uncertainty measurement, or growing and pruning criterion [5,12].

In addition, as shown in Figure 3(b), some of features detected and matched by SIFT in the second stage are actually not correct because the feature matching procedure depends only upon the distance of SIFT descriptor vector of features without consideration of

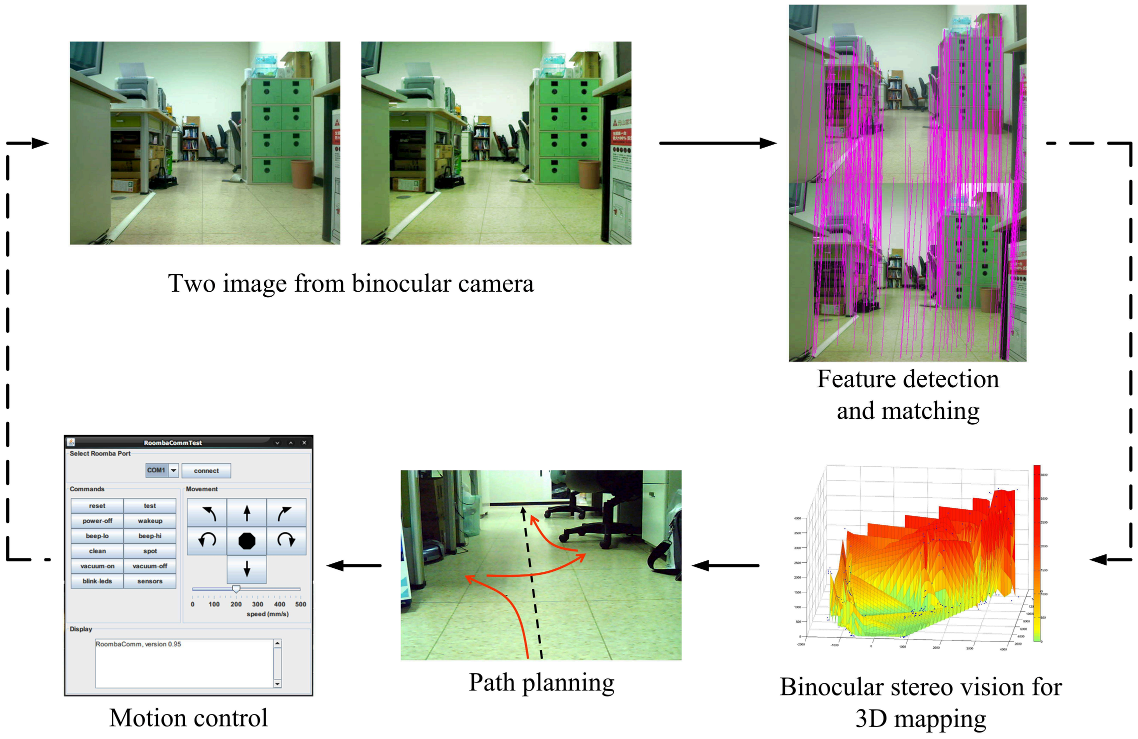
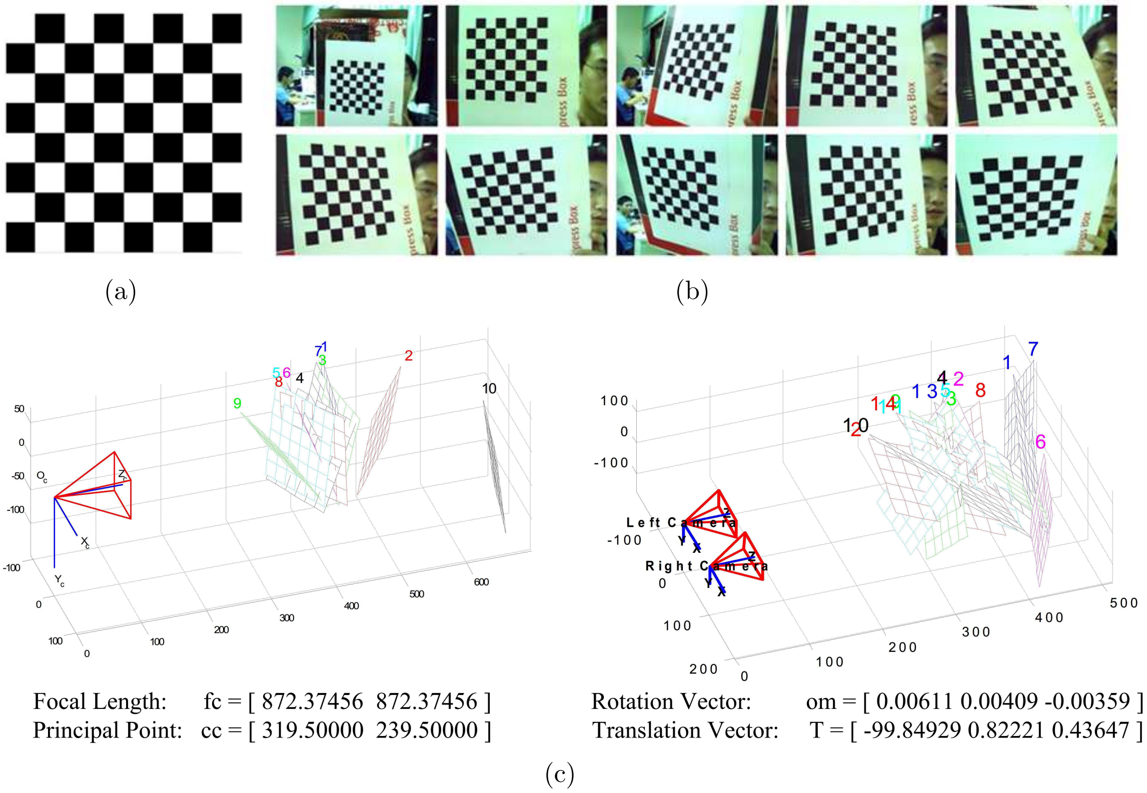


FIGURE 1. Flowchart of self-developed SIFT-based robotic stereo visual navigation



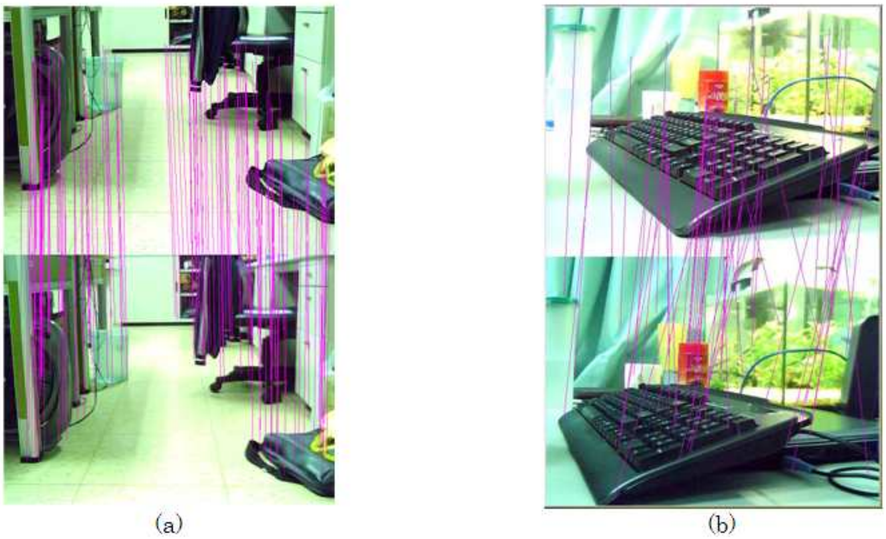
(a) (b)

Focal Length: fc = [872.37456 872.37456]
Principal Point: cc = [319.50000 239.50000]

Rotation Vector: om = [0.00611 0.00409 -0.00359]
Translation Vector: T = [-99.84929 0.82221 0.43647]

(c)

Figure 2. Camera calibration procedure. (a) 8 × 8 planar checkerboard image for calibration. (b) Loading calibration images for extracting the grid corner and main calibration scheme. (c) Intrinsic and extrinsic parameters exported from calibration.



(a) (b)

Figure 3. Binocular image issues with (a) too few and uneven SIFT-based features and (b) incorrect SIFT-based matched features

the relative location of features. In our developed system, this paper proposes 4 simple relative-distance constraints to verify and filter out a few incorrect matched features. If the matched features detected by SIFT conform to the following constraints as (1)-(4), they are incorrect and should be removed.

$$|leftimage.feature.y - rightimage.feature.y| > 3 \tag{1}$$

$$[leftimage.feature.x < (image.width)/2] \,\&\&\, [rightimage.feature.x > (image.width)/2] \tag{2}$$

$$leftimage.feature.x > rightimage.feature.x \tag{3}$$

$$|leftimage.feature.x - rightimage.feature.x| < 3 \tag{4}$$

where $leftimage.feature.y$ and $rightimage.feature.y$ mean the vertical coordinates of the matched features in left-eye and right-eye images, respectively. In the same way, $leftimage.feature.x$ and $rightimage.feature.x$ mean the horizontal coordinates of the matched features in left-eye and right-eye images, respectively. $image.width$ means the image width in left-eye or right-eye images. The unit of the value in (1)-(4) is pixel.

The epipolar lines of the binocular cameras are parallel to each other and both parallel to the ground. Equation (1) represents the vertical coordinate difference of some matched feature between the left-eye and right-eye images should not happen, but a small error margin (< 3) can be tolerant. Equation (2) means some matched feature on the left-half plane of the left-eye image should not appear on the right-half plane of the right-eye image. Equation (3) indicates some matched feature in the left-eye image should not appear more right in the right-eye image. In addition, the matched features far away from the binocular cameras should be ignored because of their inaccuracy and insignificance. Equation (4) implies the matched features with minor disparity (< 3) are not taken into consideration.

The third stage is to perform the binocular stereo vision algorithm. The binocular stereo vision algorithm is used to evaluate the depth of all matched features and estimate every feature's 3D coordinates where the coordinate system is originated at the location of the binocular cameras and the height of the binocular cameras is given. Then, 3D environmental modeling and mapping is finished.

Figures 4(a) and 4(b) illustrate the top view and side view of the binocular stereo vision algorithm scheme, respectively. In Figure 4(a), Z means the depth distance between the feature point and the binocular cameras, which is what the binocular stereo vision algorithm expects to estimate. L is the distance between two cameras, which is a constant given by the measurement. f is the focal length of the binocular cameras, which is also a constant acquired by "Camera Calibration Toolbox for MATLAB". As for dx_l and dx_r, they mean the horizontal distances between the principal point of the camera and the point projected from the feature point on the left-eye and right-eye image planes, respectively. Thus, through trigonometric functions, the distance of the feature point, Z, can be easily obtained by (5):

$$Z = \frac{f * L}{dx_l + dx_r} \tag{5}$$

Besides, according to Figure 4(b), the height of the feature point can further be evaluated by (6) and (7) if the height of the binocular cameras is given.

$$Y = \frac{dy * Z}{f} \tag{6}$$

$$H = Y + h \tag{7}$$

where Y means the vertical distance between the feature point and the epipolar line of the binocular cameras. dy is the vertical distance between the principal point of the camera and the point projected from the feature point on the binocular image planes. Z and f are known parameters resulting from (5). h is a constant about the height of the binocular cameras by measurement. H represents the height of the feature point from the ground.

Finally, after acquiring the horizontal distance, the depth distance, and the height of every feature through the binocular stereo vision algorithm, 3D model and map can be drawn. The resolution of 3D model and map depends heavily upon the density and distribution of the effective features. Subsequently, the optimal path planning in the fourth stage and the robotic motion control in the fifth stage can be easily worked out, as shown in bottom-left part of Figure 1.

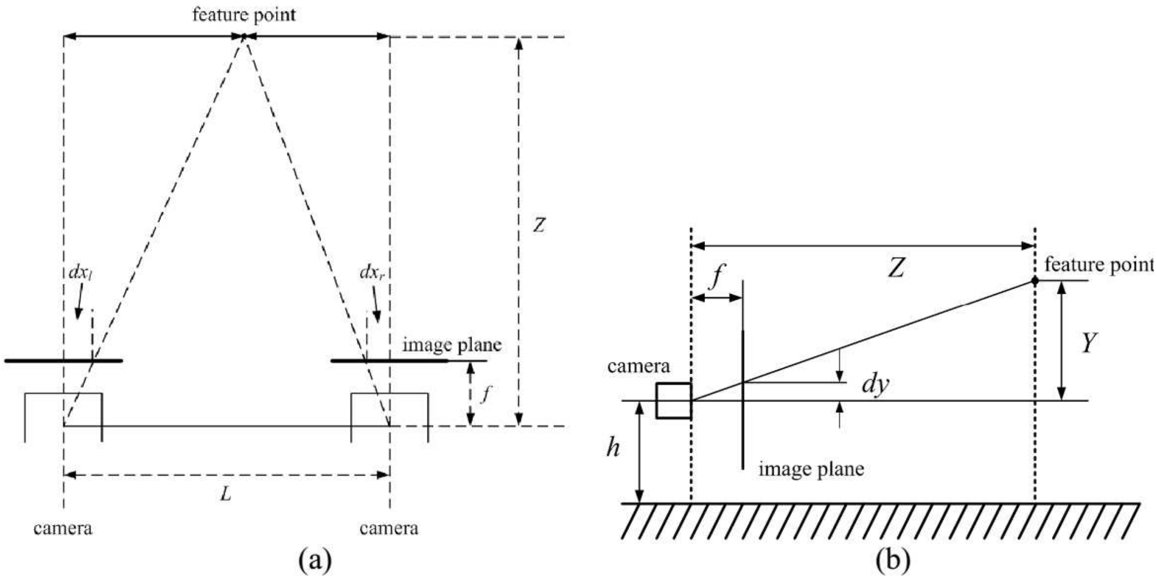


Figure 4. (a) top view and (b) side view of the binocular stereo vision algorithm scheme

However, the effective feature amount detected by the second stage of Figure 1 is usually required to appear as high as possible for accurate 3D model and map. And, the flow in Figure 1 is usually required to cycle itself as soon as possible for high-refreshing visual navigation or real-time intelligent applications. So this paper studies these two issues in the next two sections.

3. **Proposed Region-of-Interest Pyramid.** For too few and uneven SIFT-based features issue in high-contrast scenes as shown in Figure 3(a), this paper proposes to segment the image into several overlapped local regions as shown in Figure 5 and performs gray-scale histogram equalization onto these region-of-interest pyramid individually, as shown in Figure 6, before original SIFT-based feature detection and matching algorithm. In Figure 5(a), the original image is partitioned into 7 subimages. And, there are 6 local regions made up of some of these 7 subimages as shown in Figures 5(b)-(g). The index number at the bottom-right corner of every local region in Figures 5(b)-(g) means which portions of the original image in Figure 5(a) the local region is partitioned from. In fact, due to the duplicate characteristic of the local region in Figure 5(g) and the insignificance characteristic of the local region in Figure 5(b), the proposed local segmentation method can be simplified and do not take the two local regions into account. The features detected by original SIFT in the other 4 local regions are abundant and even enough to almost cover those in the local region in Figure 5(g). As for the features in the local region in Figure 5(b) they are usually far away from the binocular cameras, and not urgent and critical for SLAM.

4. **Maximum-Uncertainty Comparability Measurement.** In general, SIFT-based features are extracted by the following 4 steps: 1) scale-space extrema detection, 2) feature point localization, 3) feature orientation assignment, and 4) feature descriptor vector. However, the computational bottleneck of SIFT algorithm lies mostly in searching candidate matching features based on Euclidean distance of 128-dimensional feature descriptor vectors between binocular images. Euclidean distance (L_O) is much more complex than Cityblock distance (L_J) and Chessboard distance (L_Q). And, Euclidean distance (L_O) must be smaller than Cityblock distance (L_J), but larger than Chessbard distance (L_Q).

Thus, this paper adopts uncertain comparability measurement as (8) to break the computation bottleneck of searching candidate matching features. In short, this paper

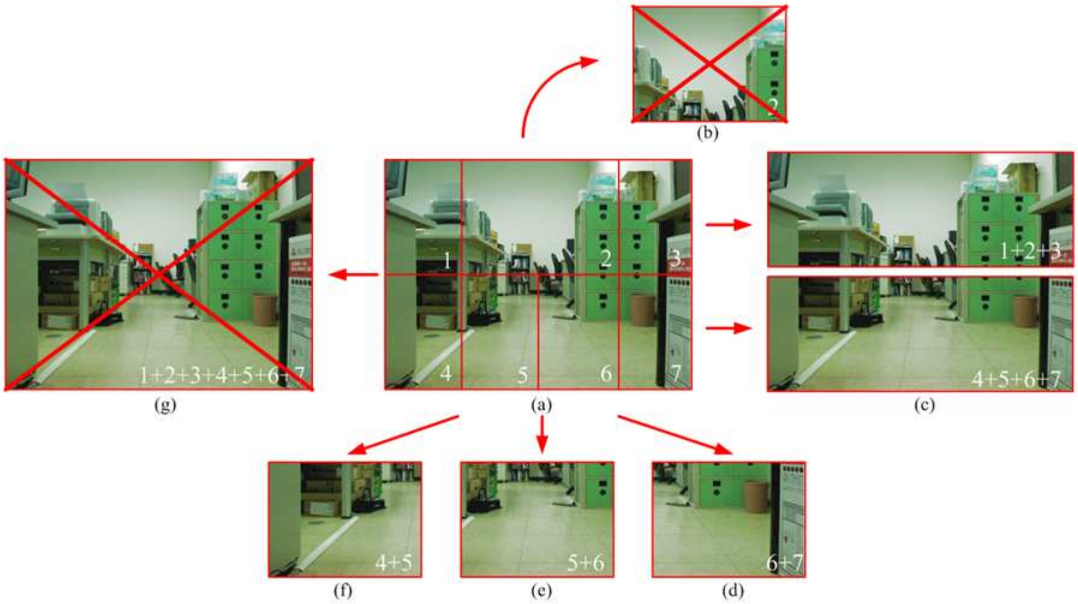


Figure 5. (a) Original image, (b)-(g) local regions partitioned from the original image

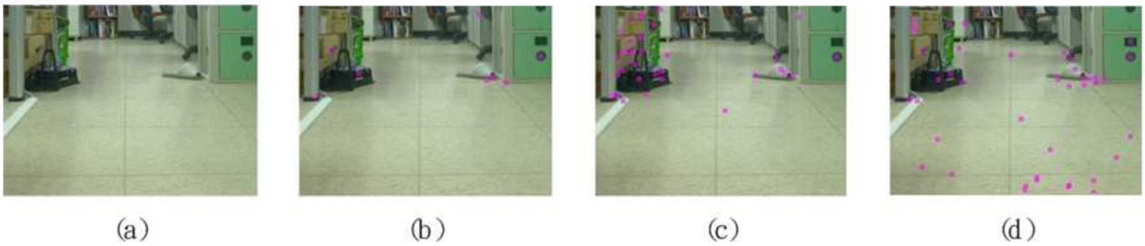


Figure 6. (a) Original color image, (b) SIFT-based features in (a), (c) SIFT-based features in histogram-equalized (a), (d) SIFT-based features in histogram-equalized and gray-scale (a)

replaces the computation of Euclidean distance with (8) to further simplify the proposed region-of-interest pyramid.

$$L_O = \alpha L_J + \beta L_Q \tag{8}$$

Here, α and β are constants by the empirical rule. However, this paper directly chooses 0.5 for both α and β in view of maximum uncertainty principle [14]. Meanwhile, the ratio of the nearest neighbor distance to the second nearest neighbor distance for the rejection constraint of the matched features must be refined from 0.8 to 0.7, that is, the matched features are reserved only if the distance ratio is less than 0.7. This is because this reconfiguration can raise the robustness of SIFT, especially when maximum-uncertainty comparability measurement is applied.

5. **Experimental Results.** In this paper, most of the developed system and experimental results are implemented by C code and OpenCV library, except that the camera pre-calibration and 3D mapping are run by MATLAB. In the developed and optimized SIFT-based robotic stereo visual navigation implementation, the binocular cameras are totally the same type and specification. Resolution of all binocular images is 640 × 480. The scene in the experiments is only indoors with some obstacles.

Figures 7(a) and 7(b) show the feature scenes of some indoor space image detected by original SIFT and the proposed SIFT, respectively. From the comparison of Figures 7(a)

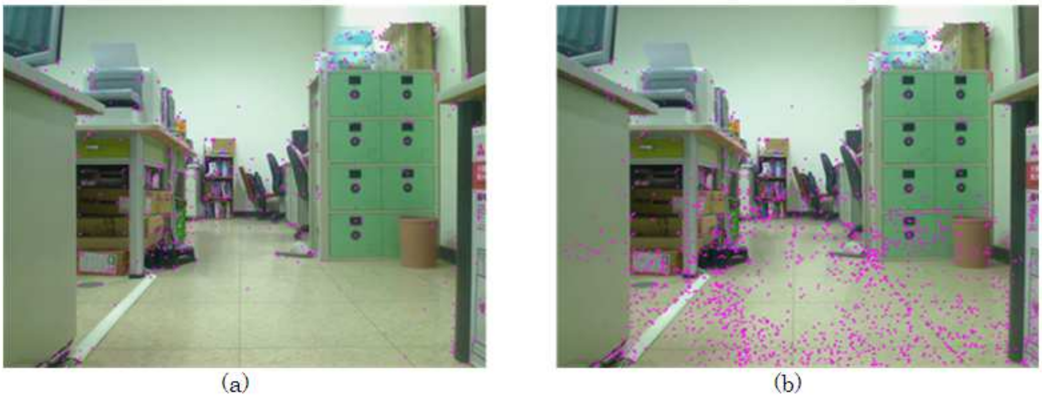


Figure 7. Feature scenes detected by (a) original SIFT and (b) proposed SIFT

Table 1. Feature amount comparison

	Scene 1	Scene 2	Scene 3
Original SIFT	693 (points)	1060 (points)	759 (points)
Proposed SIFT	4516 (points)	5809 (points)	4125 (points)
Increment ratio	552%	448%	443%

and 7(b), it is obvious that the proposed SIFT can generate the detected features more densely and evenly, especially in high-contrast scenes. Table 1 displays the experimental comparison of the detailed feature amount in various scene images detected by original SIFT and the proposed SIFT. From Table 1, it is seen that the proposed SIFT can improve original SIFT-based robotic stereo visual navigation by over 400% increment.

After the feature extraction step in Figure 7(b), the feature pairs of binocular images are matched and filtered by the proposed SIFT. It is obvious that the proposed SIFT can generate the matched feature pairs more densely and evenly, especially in high-contrast scenes. Finally, the front view and top view of the 3D model and map can be generated through the binocular stereo vision algorithm.

6. **Conclusions.** SIFT-based feature detection and matching algorithm is the key technology for robotic stereo visual navigation. Rather than complex or inefficient algorithms, this paper adopts region-of-interest pyramid and maximum-uncertainty comparability measurement to increase the SIFT-based feasible features and decrease the execution time of SIFT simultaneously. Thereby, the 3D model and map drawn by the proposed robotic stereo visual navigation is finer and better, and the subsequent stages of path planning and motion control can also be easily worked out.

REFERENCES

[1] S. Livatino, F. Banno and G. Muscato, 3-D integration of robot vision and laser data with semi-automatic calibration in augmented reality stereoscopic visual interface, *IEEE Trans. Industrial Informatics*, vol.8, pp.69-77, 2012.

[2] T. Lupton and S. Sukkarieh, Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions, *IEEE Trans. Robotics*, vol.28, pp.61-76, 2012.

[3] J. Lobo and J. Dias, Vision and inertial sensor cooperation using gravity as a vertical reference, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, pp.1597-1608, 2003.

[4] S. M. Smith and J. M. Brady, SUSAN – A new approach to low level image processing, *International Journal of Computer Vision*, vol.23, pp.45-78, 1997.

[5] A. J. Davison and D. W. Murray, Simultaneous localization and map-building using active vision, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, pp.865-880, 2002.

[6] C. G. Harris and M. Stephens, A combined corner and edge detector, *Proc. of the 4th Alvey Vision Conference*, pp.147-151, 1988.

[7] D. G. Lowe, Object recognition from local scale-invariant features, *Proc. of the International Conference on Computer Vision*, vol.2, pp.1150-1157, 1999.

[8] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, pp.91-110, 2004.

[9] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[11] J. Y. Bouguet, Camera calibration toolbox for MATLAB, *Computational Vision at the California Institute of Technology*, http://www.vision.caltech.edu/bouguetj/calib_ doc/, 2013.

[12] J. L. Blanco, J. A. Fernández-Madrigal and J. Gonzalez, A novel measure of uncertainty for mobile robot SLAM with Rao-Blackwellized particle filters, *The International Journal of Robotics Research*, vol.27, pp.73-89, 2008.

[13] D. Zhu, Binocular vision-SLAM using improved SIFT, *Proc. of International Conference on Information Security and Assurance*, pp.38-41, 2010.

[14] B. Liu, *Uncertainty Theory*, 4th Edition, www.orsc.edu.cn/liu/ut.pdf, 2015.