Obstacle avoidance for small UAVs using monocular vision

Jeong-Oog Lee and Keun-Hwan Lee
Aerospace Information Engineering, Konkuk University, Seoul, Republic of Korea

Sang-Heon Park
Pohang Institute of Intelligent Robotics, Gyeongbuk, Republic of Korea, and

Sung-Gyu Im and Jungkeun Park
Aerospace Information Engineering, Konkuk University, Seoul, Republic of Korea

Abstract

Purpose – The purpose of this paper is to propose that the three-dimensional information of obstacles should be identified to allow unmanned aerial vehicles (UAVs) to detect and avoid obstacles existing in their flight path.

Design/methodology/approach – First, the approximate outline of obstacles was detected using multi-scale-oriented patches (MOPS). At the same time, the spatial coordinates of feature points that exist in the internal outline of the obstacles were calculated through the scale-invariant feature transform (SIFT) algorithm. Finally, the results from MOPS and the results from the SIFT algorithm were merged to show the three-dimensional information of the obstacles.

Findings – As the method proposed in this paper reconstructs only the approximate outline of obstacles, a quick calculation can be done. Moreover, as the outline information is combined through SIFT feature points, detailed three-dimensional information pertaining to the obstacles can be obtained.

Practical implications – The proposed approach can be used efficiently in GPS-denied environments such as certain indoor environments.

Originality/value – For the autonomous flight of small UAVs having a payload limit, this paper suggests a means of forming three-dimensional information about obstacles with images obtained from a monocular camera.

Keywords Small UAVs, SIFT algorithm, MOPS, Monocular vision, Obstacle avoidance, Image processing, Collisions

Paper type Research paper

Introduction

Recently, there have been many studies on unmanned aerial vehicles (UAVs), among which small UAVs that can be utilized in many areas have come to the fore. Small UAVs are systems that perform unmanned missions via remote control on the ground through manual, semi-automatic, or automatic program methods. This type of system has been studied for mainly military purposes, though it can also be utilized in other areas, such as for meteorological observations, remote detection, communication relaying, and aircraft dusting and spraying. For the autonomous flight of UAVs, a function for the detection and avoidance of obstacles is needed, and research in this area is actively being performed. The size of UAVs has become increasingly smaller, and research on small UAVs that fly both indoors and outdoors has commenced (Finio *et al.*, 2009; Markus *et al.*, 2008; Mondragon *et al.*, 2007). The development of micro-robotic fly (Finio *et al.*, 2009) is an example of research on the topic of micro-air vehicles. Because such vehicles have a minimal size and weight, they can be utilized in various situations, e.g. they can easily access nuclear reactors or cable tunnels that are difficult to access by human personnel for observation purposes in these close areas. The feasibility of the autonomous flight of small UAVs, having a payload limit, requires them to be able to detect the location and shape of obstacles so as to avoid these obstacles on their own with minimal on-board sensors. Currently, studies of the recognition of three-dimensional obstacles are undertaken usually by means of stereo vision or using laser sensors in parallel.

In a distance calculation method using stereo vision, distance information is extracted through an arithmetic operation using the position difference of the same pixel between the left and the right images, the distance between two cameras, the focal length, and other factors (Garcia and Solanas, 2004; Park *et al.*, 2009). Calculating the distance to the obstacle in such a manner as done using stereo vision is straightforward to the point that it reduces the calculation time. However, the disadvantage is that the sensor is large and heavy, as this method needs to maintain a certain distance between the two cameras. This method is therefore usually used with robots that move on the ground, for which a heavy weight is not a critical issue.

Laser sensors have the advantage of minimizing interference through their superior monochromatic features and straightness. As compared to other sources of light, they also

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1748-8842.htm

Aircraft Engineering and Aerospace Technology: An International Journal
83/6 (2011) 397–406

[DOI 10.1108/00022661111173270]

This paper was supported by Konkuk University in 2010.

Aircraft Engineering and Aerospace Technology: An International Journal
Volume 83 · Number 6 · 2011 · 397–406

have relatively strong brightness (Markus *et al.*, 2008). The accuracy of laser sensors can express not only three-dimensional coordinate data but also the surface material quality of a structure. They have the disadvantage of weighing 80 g at a minimum, which is on the heavy side compared to other sensors. Moreover, the cost factor is high, at more than several thousand dollars.

The distance information extraction method using monocular vision matches information from sequentially taken images, identifies the same obstacles from every image, and calculates the distance to the obstacles. This method, using monocular vision, is commonly used in the study of small UAVs having a payload limit because it is both moderately priced and lightweight (Markus *et al.*, 2008; Mondragon *et al.*, 2007).

Currently, most studies in the area of obstacle recognition using monocular vision either compare objects existing in a database for recognition (Davison and Kita, 2003) or recognize the approximate outline of an obstacle through an expression of the components of three-dimensional points only (Lowe, 2004). The method of using information stored in a database to identify and distinguish objects has disadvantage in that the recognition of new objects not existing in the database is impossible. The method of recognizing obstacles through an expression with point components has a disadvantage in that it cannot determine the exact shape of an object. Research related to the three-dimensional restructuring of new objects not existing in a database (Se *et al.*, 2002) is underway for the correct recognition of obstacles, but thus far the calculation time is high.

In this paper, a solution to detect and avoid obstacles in real time is proposed for the autonomous flight of small UAVs. In order to recognize an obstacle not existing in a database quickly and accurately, the approximate outline of the image must first be extracted, after which the shape is determined. The internal information of an object is expressed with the distance information of the scale-invariant feature transform (SIFT) feature points to be combined with outline information for the reconstruction of the three-dimensional information of the obstacles. This paper is composed as follows. Multi-scale-oriented patches (MOPS) and the SIFT algorithm are briefly introduced in second section. The concept of vision-based autonomous navigation by UAVs is introduced in third section. A method to obtain three-dimensional information using MOPS and SIFT is explained in fourth section. The introduction of the experiment done here and evaluation of the findings follows in fifth section, and the conclusion and further study are discussed in sixth section.

Basics

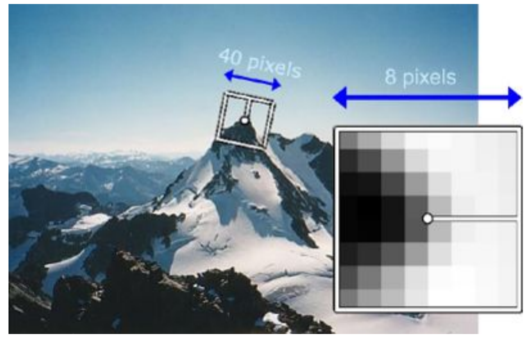
MOPS (Brown *et al.*, 2004) is introduced in next section followed by the introduction of the SIFT algorithm (Lowe, 2004) in subsequent section.

Multi-scale-oriented patches

MOPS is a system and process that identifies the same points among multiple images. Features are located at Harris corners in a discrete scale space. The feature descriptor in Figure 1 consists of an 8×8 patch of bias/gain normalization intensity values that are calculated by equation (1):

$$d_i = \frac{(d'_i - \mu)}{\sigma} \tag{1}$$

Figure 1 MOPS feature descriptor



Here, $d'_i, \quad i \in \{1 \ldots d^2\}$ are the elements of the descriptor vector, with: $\mu = (1/d^2)\sum_{i=1}^{d^2} d'_i$ and:

$$\mu = \sqrt{\frac{1}{d^2}\sum_{i=1}^{d^2}(d'_i - \mu)^2}.$$

Using MOPS, quick and accurate corner point matching is enabled. With these matched feature points, the spatial coordinates of the feature points are calculated using equation (2), which expresses the relationship between the global and camera coordinates:

$$Z_2 = \frac{f^*(x_1^* R_3 - f^* R_1)^* T_c}{(x_1^* R_3 - f^* R_1)^* P_{2c}} \tag{2}$$

For a quick recognition of obstacles, knowledge of the shape of each obstacle and the related distance information is extracted in real time. As the distance information of the corner point can be calculated with MOPS, the position of each obstacle can be captured. The three-dimensional information of the obstacles can be reconstructed when the outline information obtained by MOPS is combined with the information of the internal outline. SIFT feature points provide the information of the internal outline.

Scale-invariant feature transform

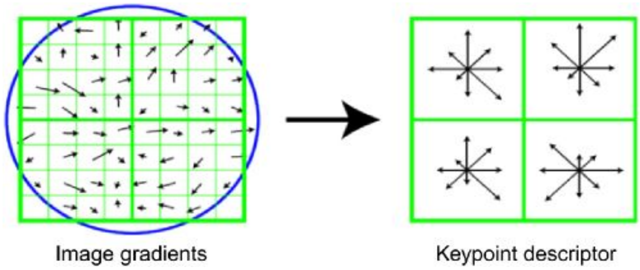
The SIFT (Lowe, 2004) is an image process technique. This algorithm is relatively unaffected by size change, as it creates an image pyramid and extracts feature points. It is also relatively unaffected by rotation changes because it can extract the orientation of the feature points and generate feature vectors.

To extract feature points, extreme points are found on a scale space. These extreme points are the position most likely to be re-detected after an image scale change or an image transformation. This feature can be very useful during the image-matching process.

Most stable extreme points are finally selected through a post-processing method, and the orientation of feature points is determined through the surrounding gradient values of such extreme points. Orientation histograms for the points located near the feature point are generated. The SIFT description vector is shown in Figure 2.

Those pixels with distinctive features are known as keypoints. The stability of feature points is critical for matching based on feature points, but these feature points are points that can be observed even if there is some distortion

Figure 2 SIFT description vector



or future change in the image of concern in terms of its image data.

After the extraction and description of the complete feature points, image matching is done. A simple form of image matching occurs in a similarity comparison between keypoints, where, when the numbers of the feature points for two images are N and M, the total N*M feature points matching is processed.

The Euclidean distance is used as the measure of similarity during process matching. If the ratio between the nearest and the second nearest Euclidean distances to the feature points within a database is 0.6 or less, it is considered as a match. If ten or more of the feature points extracted from the input images are matched to the feature points of the obstacles stored in the database, the obstacle is considered as recognized.

Figure 3 shows the matching image after the extraction of every SIFT feature point from two images. One of the advantages of SIFT is that it can recognize an obstacle regardless of any change in light or size. The recognition ratio is high as well, which is why it is commonly used for recognition and detection purposes. Because SIFT feature points exist not only at the edge of each object, but throughout the entire image, however, determining the size and shape of the obstacles can be difficult, which represents a disadvantage of this algorithm.

Vision-based autonomous navigation of UAVs

Autonomous navigation is an important and essential ability for all types of UAVs, including small UAVs. For autonomous navigation by small UAVs, practical and effective technologies such as attitude control, localization, environment modeling (map building), path planning, and obstacle avoidance are necessary. Earlier research worked on the development of vision-based localization and map building for small UAVs. In previous research by the authors (Lee *et al.*, 2010), we attempted to achieve natural landmark-based localization using a vision system for indoor navigation by small UAVs. Our method first extracts feature points (landmarks) from the image data taken by a monocular camera using the SIFT algorithm. The locations of the landmarks are then calculated and stored in a map database. Based on the landmark information, the current UAV position can be estimated.

In order to build a map of a flight environment, we calculate the position of the identified feature points using projection geometry. Figure 4 shows the projection of a landmark feature point.

Via triangle proportionality, we can easily derive the location of P_p, as shown below:

$$P_p(x_p, y_p) = \frac{f}{\vec{n}_x \cdot \overrightarrow{RP}_F} \left(\vec{n}_y \cdot \overrightarrow{RP}_F,\ \vec{n}_z \cdot \overrightarrow{RP}_F\right) \tag{3}$$

We need three independent equations to calculate the location of the feature point $P_F(x_F, y_F, z_F)$. This can be obtained using two different images that are taken two different locations but contain the same feature point, as shown in Figure 5.

Once the map of the flight environment is built, the current position of the UAV can be estimated using feature points in an image taken by the UAV. The estimation of the UAV position is similar to the mapping of feature points, which based on equation (3).

We conducted a number of experiments to validate the effectiveness of our method. We predefined 23 positions as UAV locations. We calculated all the pre-defined UAV locations from the calculated natural landmarks obtained during a mapping process, after which we compared the calculated UAV locations with the actual UAV locations as measured manually. Figure 6 shows the difference between the calculated UAV locations and the actual UAV locations. As shown in the figure, the method works moderately well.

For the next step for autonomous navigation, obstacle detection and avoidance, we sought to develop a method capable of obtaining the three-dimensional information of obstacles.

Constructing three-dimensional information of obstacles using MOPS and SIFT

The method proposed in this paper seeks to perform a data process with MOPS and SIFT in parallel, merge the results

Figure 3 SIFT feature point matching

Figure 4 Projection of a landmark feature point (PF)

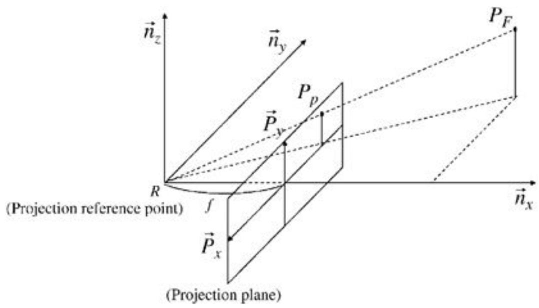
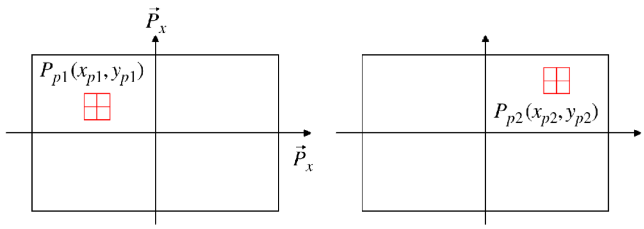


Figure 5 Two different images having the same feature point



from these two processes, and reconstruct the three-dimensional information of an obstacle. Figure 7 shows the manner of constructing the three-dimensional information of obstacles using MOPS and SIFT. The process of quickly identifying the approximate outline of an obstacle using MOPS is explained in section "extracting the outline information of an object through MOPS", and the process of extracting the SIFT feature points and generating the three-dimensional SIFT feature points is explained in section "grasping the internal outline information of an object through SIFT". These two processes act simultaneously. The construction of the three-dimensional information of the obstacles by merging the two sets of results is discussed in section "merging MOPS and SIFT process results and construction of the three-dimensional information of the object".

Extracting the outline information of an object through MOPS

The first step is the rapid extraction of the outline information of an object in the image using MOPS. The MOPS feature points for the corner point of the object are obtained first. The MOPS feature points of the image are matched, and the three-dimensional MOPS feature points, including the three-dimensional spatial information, is then extracted. Figure 8 shows the extracted result of the three-dimensional edge spatial information after matching the feature points of the two images.

Because the internal outline information is not included in the object outline information obtained, it cannot reconstruct the three-dimensional information of the object. The information needed for an internal outline is therefore extracted according to the method explained in the next section.

Grasping the internal outline information of an object through SIFT

The second process involves the use of the SIFT feature points to obtain the internal outline spatial information. The SIFT algorithm is applied to the two images and the feature points are extracted. The SIFT feature points can be extracted not only from the objects themselves, but also from a wall or a floor where changes in the color or light are clear. As such, because SIFT feature points can be extracted throughout the entire area of the image, the three-dimensional spatial information of the entire image can be expressed with point components. By matching the SIFT feature points of the two images obtained in this way, the three-dimensional SIFT feature points, including the three-dimensional spatial information, can be extracted.

Figure 9 shows the extracted three-dimensional spatial information after the matching process of the SIFT feature points of the two images. The SIFT three-dimensional spatial information obtained as such is merged with the outline information extracted by MOPS, and the three-dimensional information of the object is then reconstructed (Figure 10).

Figure 6 Differences between the calculated UAV locations and the actual UAV locations

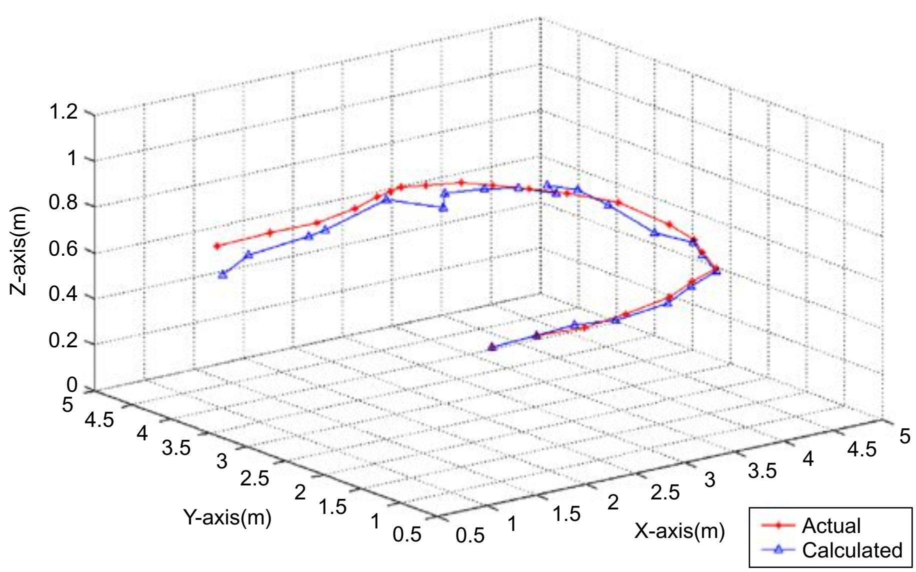


Figure 7 Constructing three-dimensional information of obstacles using MOPS and SIFT

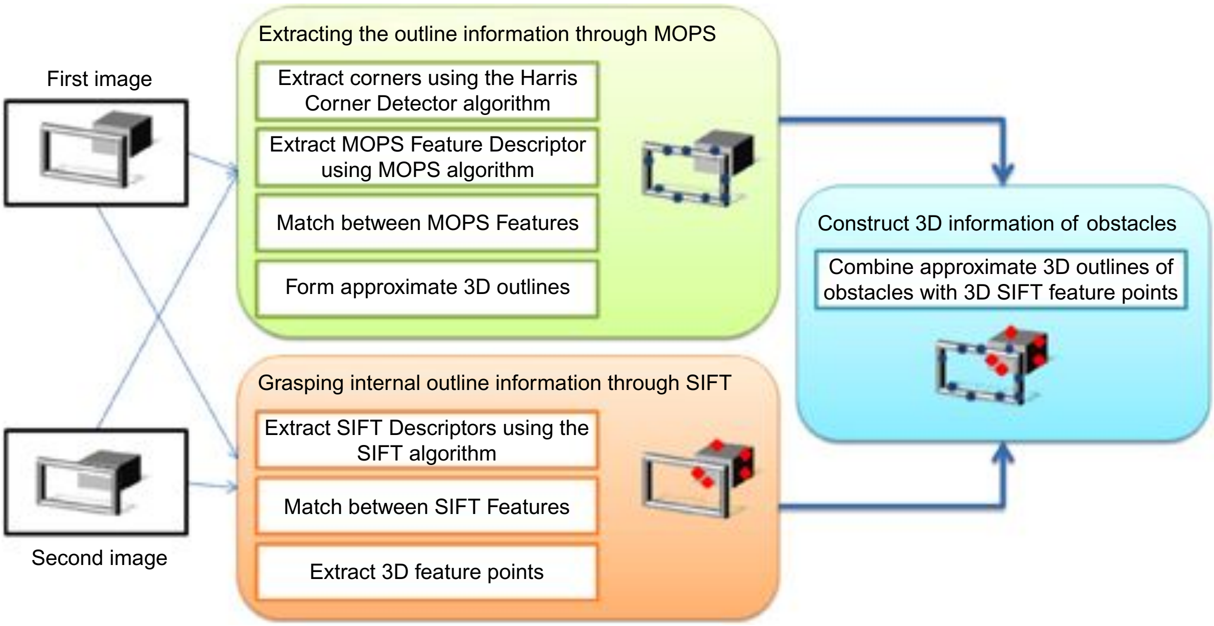
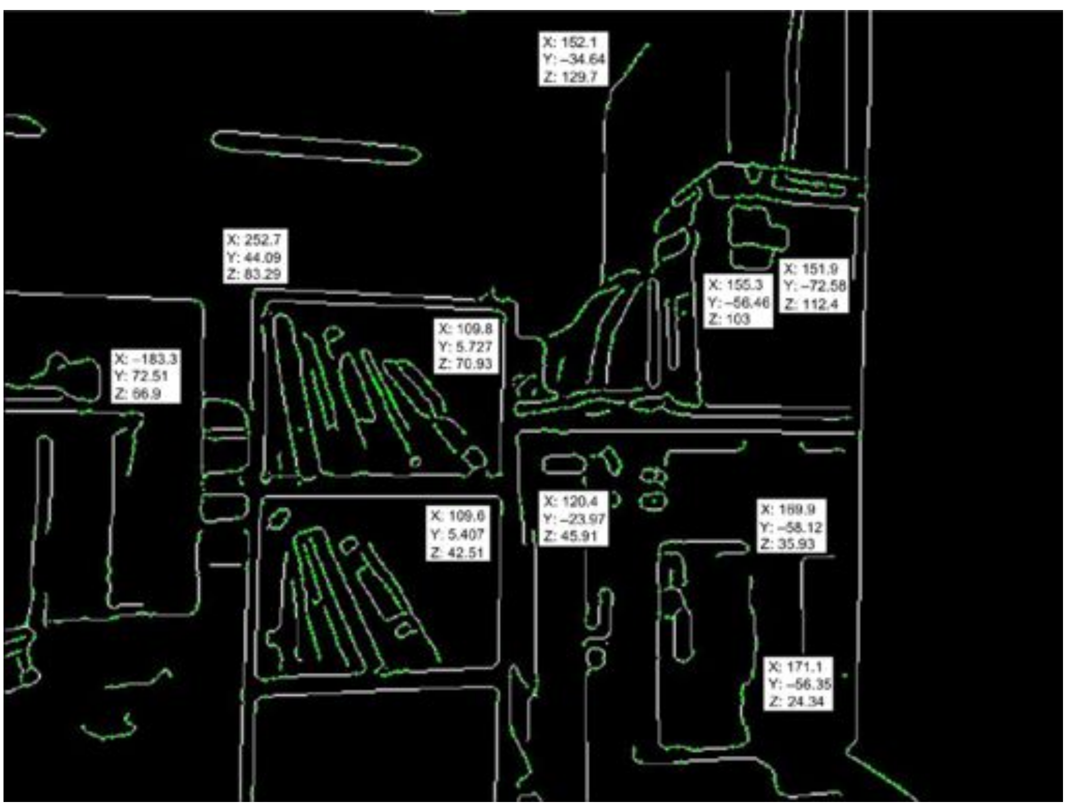


Figure 8 Extracting the three-dimensional edge spatial information of obstacles



Merging MOPS and SIFT process results and construction of the three-dimensional information of the object

The final process is the merging of the approximate outline information obtained through MOPS in the first processing with the internal outline information obtained through the three-dimensional SIFT feature points in the second process, resulting in the construction of the three-dimensional information of the object. With three-dimensional SIFT feature points, which comprise the information of the internal

Aircraft Engineering and Aerospace Technology: An International Journal
Volume 83 · Number 6 · 2011 · 397–406

Figure 9 SIFT three-dimensional spatial information

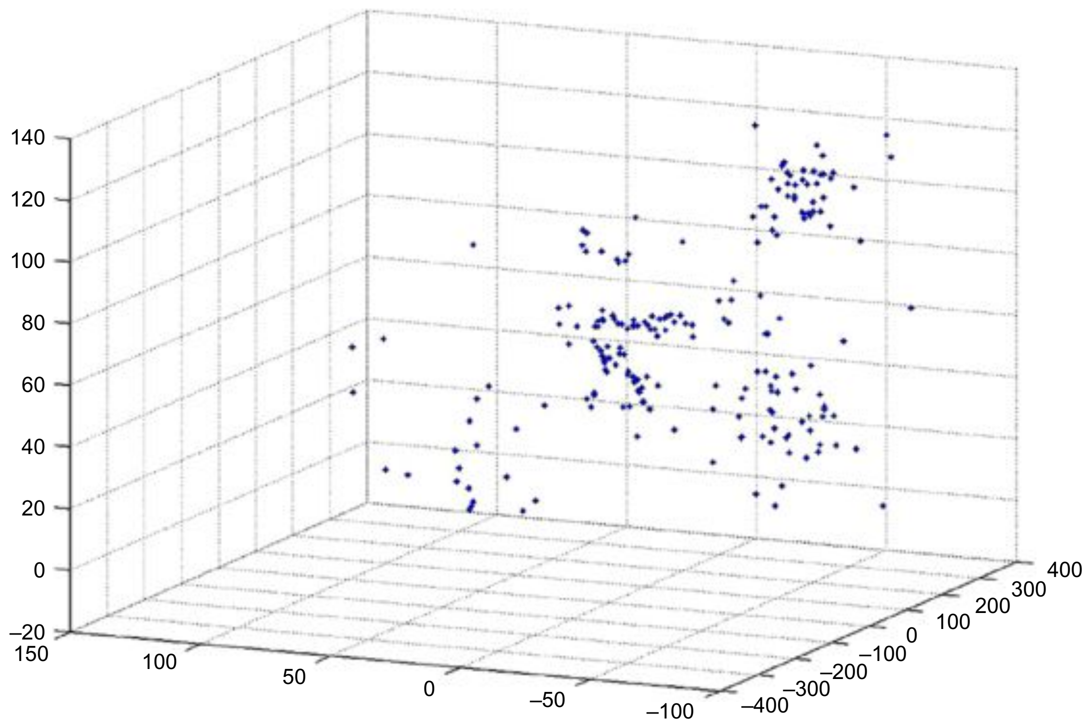


Figure 10 Adding of the SIFT three-dimensional spatial information to the obstacle outline information

outline of the obstacle, the distance information up to the outline and the distance information up to the SIFT feature points are compared to estimate the cubic information of the object. A random three-dimensional MOPS feature point existing in the outline of a random object is expressed in this case as M, and a random three-dimensional SIFT feature point existing in the internal outline of the object is expressed as S. If the distance from the small UAV to M is D_M, the distance information to S is D_S, with thresholds at T_1 and T_2. The relationship among the thresholds can then be presented as in the following three equations:

$$|D_M - D_S| \leq T_1 \tag{4}$$

$$T_1 < |D_M - D_S| \leq T_2 \tag{5}$$

$$T_2 < |D_M - D_S| \tag{6}$$

For threshold T_1, set the value having a high possibility that S and M are placed on the same object; for T_2, set the value having a high possibility that S and M are placed on a different object.

Equation (4) represents the case in which the absolute value of the difference between the distance to the outline and the distance to the internal outline is within threshold T_1, which can be considered as the cross-section of an object with a closed interior (Figure 11(a)). Equation (5) represents the case in which the absolute value of the difference between the distance to the outline and the distance to the internal outline exists between T_1 and T_2. This case refers to an object with a closed interior with a convex shape (Figure 11(b)) or to an object with a closed interior (Figure 11(c)). Equation (6) represents the case in which the absolute value of the difference between the distance to the outline and the distance to the internal outline is larger than threshold T_2, where it can be considered that the corresponding surface is an open space (Figure 11(d)). In such a case, the feature points of another object are extracted behind the open space.

Experiment and evaluation

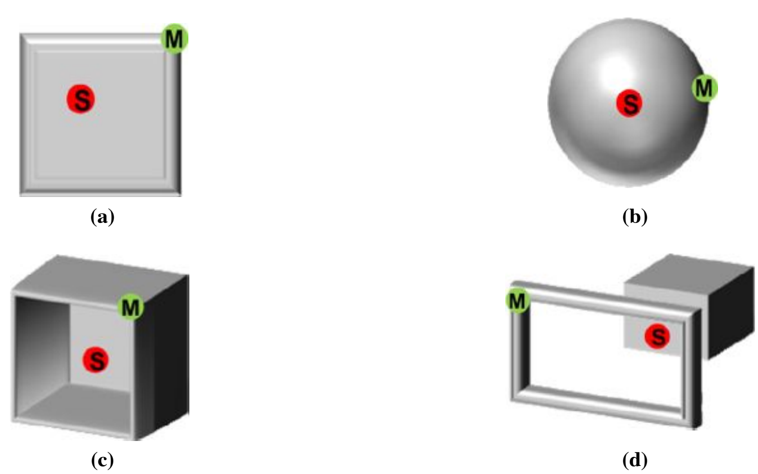
In order to evaluate the accuracy and performance speed of the proposed method, an experiment was performed and evaluated using MATLAB on an Intel Core 2 Quad CPU 2.40 GHz processor. In the experimental environment, there were two main obstacles. The first is rectangular with a closed interior and the second is square with an open space, as shown in Figure 12.

In each of the experiments, images were extracted from streaming image data. Table I shows some of the experimental results. The average time taken to extract the outline information of the object through MOPS was 0.577 s. Although the number of MOPS feature points was maximized in this experiment, the processing of two images per second was achieved, which shows that this function can be processed in real time during actual flight. To grasp the internal outline information of an object through SIFT, 0.993 s were used on average. As shown in the images, the maximum number of SIFT feature points was extracted for this experiment, demonstrating that the SIFT algorithm required more time than MOPS during the processing step. Finally, an average of 0.020 s was required to merge the MOPS and SIFT process results and compose the three-dimensional information of the object. This represents a very small portion of the overall process.

In our approach, the three-dimensional coordinates of each feature point (the MOPS feature point and the SIFT feature point) can be calculated using the camera position, camera direction, and feature point locations in the images. The camera direction was obtained from the attitude heading reference system:

- MOPS three-dimensional feature points located on the edges of the objects in an image were obtained and the approximate shape of the objects could be induced using the MOPS feature points. At this point, the small UAV cannot likely determine whether the object has a closed interior or an open space.
- In contrast to MOPS, SIFT three-dimensional feature points could be extracted throughout the entire area

Figure 11 The relationship between the distances from a small UAV to the object outline and to the internal feature



Notes: (a) When the distance information of S and M are similar; (b) the distance information of S is slightly closer than that of M; (c) the distance information of S is slightly further than that of M; (d) the distance information of S is much further than that of M

Aircraft Engineering and Aerospace Technology: An International Journal
Volume 83 · Number 6 · 2011 · 397–406

Figure 12 Experimental images of the MOPS and SIFT feature point extraction test

of the image. However, SIFT feature points when used by themselves do not grasp the approximate shapes of objects.

- After combining both SIFT (to obtain the internal outline) and MOPS (to obtain the approximate shape) and calculating the difference in the distance between SIFT feature points and MOPS feature points, three-dimensional information of the obstacles could be reconstructed. The small UAV could determine at this point whether a detected object has a closed interior or an open space. As the small UAV identified the three-dimensional coordinates of the detected MOPS and SIFT feature points via a calculation, the size and position of the obstacles could be inferred.

Conclusion

This paper introduces a means of extracting the outline and internal outline information of an object from images obtained using a monocular camera on a small UAV along with how to use such information to form three-dimensional information pertaining to obstacles. The average time taken to extract the outline of an obstacle was 0.577 s. Because the system extracts the approximate outline information to the extent needed for detection rather than extracting detailed three-dimensional information regarding the obstacle, fewer calculations are required. In addition, the information of the internal object outline was expressed using three-dimensional SIFT feature points to reconstruct the three-dimensional information of the object by merging the SIFT feature points with the outline information of the object. Because the proposed method is based on monocular vision, it can enormously improve the obstacle detection and avoid functions of small UAVs that are sensitive to payload.

Further work

A future project will be to study the reconstruction of a three-dimensional map of the flight environment using the distance information up to MOPS feature points located on the outline of an object and SIFT feature points that exist internally in the object. This can enable the development of an efficient path planning algorithm applicable to small UAVs.

Table I Summary of the experimental results

	Experimental image									
	1&2	**2&3**	**3&4**	**4&5**	**5&6**	**6&7**	**7&8**	**8&9**	**9&10**	**10&11**
The number of matched SIFT feature points	751	797	715	700	607	530	500	582	538	570
The number of matched MOPS feature points	181	317	247	322	252	208	269	295	282	193
The time taken to extract the three-dimensional coordinates of the MOPS feature points	0.598	0.573	0.513	0.526	0.572	0.598	0.579	0.597	0.598	0.614
The time taken to extract the three-dimensional coordinates of the SIFT feature points	0.966	1.002	1.017	1.007	1.013	0.964	0.983	0.999	0.985	1.003
The total required performance time for all processes	1.584	1.595	1.550	1.553	1.605	1.572	1.592	1.606	1.593	1.637

References

Brown, M., Szeliski, R. and Winder, S. (2004), "Multi-image matching using multi-scale oriented patches", Technical Report MSR-TR-2004-133, Microsoft Research, Beijing, December.

Davison, A.J. and Kita, N. (2003), "3D simultaneous localisation and map-building using active vision for a robot moving on undulating terrain", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, December 2001*, pp. 384-91.

Finio, B.M., Shang, J.K. and Wood, R.J. (2009), "Body torque modulation for a microrobotic fly", *Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, May*, pp. 3449-56.

Garcia, M.A. and Solanas, A. (2004), "3D simultaneous localization and modeling from stereo vision", *Proceedings of the IEEE International Conference on Robotics and Automation, New Orleans, LA, USA*, pp. 847-53.

Lee, J.O., Kang, T.S., Lee, K.H., Im, S.K. and Park, J.K. (2010), "Vision-based indoor localization for unmanned aerial vehicles", *Journal of Aerospace Engineering*, June.

Lowe, D.G. (2004), "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, Vol. 60, pp. 91-110.

Markus, A., Abraham, B., Ruijie, H., Samuel, P. and Nicholas, R. (2008), "Autonomous navigation and exploration of a quadrotor helicopter in GPS-denied indoor environments", paper presented at Robotics: Science and Systems Conference, Zurich, June.

Mondragon, I.F., Campoy, P., Correa, J.F. and Mejias, L. (2007), "Visual model feature tracking for UAV control", *Proceedings of the IEEE International Symposium on Intelligent Signal Processing* (*WISP 2007*), *Alcala de Henares, Spain*, pp. 1-6.

Park, S.H., Lee, J.O. and Baik, D.K. (2009), *Vision-based Obstacle Detection and Avoidance for Small UAVs*, The KOREA Society for Simulation, Seoul, pp. 245-9.

Se, S., Lowe, D.G. and Little, J.J. (2002), "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks", *International Journal of Robotics Research*, Vol. 21, pp. 735-58.

Further reading

Agarwal, A. and Triggs, B. (2006), "Recovering 3D human pose from monocular images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28 No. 1, pp. 44-58.

Corresponding author

Jeong-Oog Lee can be contacted at: ljo7@konkuk.ac.kr