Detection of Nearby Obstacles with Monocular Vision for Earthmoving Operations

H. Son, H. Sung, H. Choi, S. Lee, and C. Kim*

Department of Architectural Engineering, Chung-Ang University, South Korea
E-mail: hjson0908@cau.ac.kr, gusdn7543@gmail.com, vinaj516@gmail.com, leesungwook@cau.ac.kr, changwan@cau.ac.kr (*corresponding author)

Abstract

The equipment used in earthmoving operations poses a significant threat to the safety of the equipment operator and construction workers due to the operator's inherently poor visibility of the surrounding environment. This study proposes a method of automated detection of nearby obstacles with monocular vision, with the goal of protecting the equipment operator and construction workers from potentially dangerous situations, such as collisions between earthmoving equipment and obstacles within a certain proximity. The proposed method consists of three steps: 1) correction of lens distortion prior to further processing, 2) shadow removal, and 3) detection of nearby obstacles with a predefined height level via perspective transformations. The proposed method was tested on video streams acquired from a camera installed on the side of the equipment body while an excavator executed excavating and moving tasks. The experimental results showed that the proposed method can provide the equipment operator with information about nearby obstacles during the excavator's manipulation and transportation. It is expected that the proposed method can be implemented in rearview monitoring systems and surrounding view monitoring systems for operator assistance and to achieve active safety.

Keywords –

Active safety; earthmoving operations; imaging sensors; intelligent earthmoving equipment; operator assistant

1 Introduction

Visibility is important for securing safety during operation of the construction equipment. Since humans depend on visual sense for 90% of the information received through sensory organs, if an equipment operator fails to secure sufficient visibility, this could cause serious casualties and reduce task efficiency. As demand for a higher visibility standard grows in the construction industry, the International Organization for Standardization (ISO) has developed a number of international standards for construction machinery (e.g., ISO 5006: 2006) [1]. Installing multiple mirrors can be helpful in ensuring visibility, according to the standard, but it has fundamental limitations in that the equipment operator must check multiple mirrors from time to time. To improve the visibility of the surrounding environment, earthmoving equipment manufacturers have adopted a rearview monitoring system, or a system that monitors further around the equipment by installing camera(s) on the rear or on every side (e.g., the rear, left, right, and front) of the equipment body and displaying the views on the operator's monitor. Although these systems provide improved visibility of the surrounding environment, detecting potential collisions between earthmoving equipment and obstacles (e.g., construction workers, facilities, and others) is still cognitively effortful and restricted by the operator's limited cognitive capacity while executing tasks. Therefore, it is necessary to develop a method to rapidly provide information about 3D environments from the images acquired from cameras installed on each side, which is important and helpful in assisting operators.

Modeling of the surrounding environment is highly desirable for construction automation. However, it is a difficult task to effectively and rapidly represent surroundings due to the complexity of construction workspaces and rapid variation in objects' locations. Toward this end, various 3D imaging sensors are being developed and tested for 3D modeling. Some researchers have proposed the use of two 2D laser rangefinders to enhance safety [2]. In order to acquire consistent 3D information with such sensors, multiple consequent acquisitions are performed and merged. The data acquisition is therefore time-consuming. 3D laser scanners can provide high-resolution 3D information with large field of view. However, because it takes tens of seconds to scan once, it is not suitable for modeling the dynamic environment.

Laser-based flash laser distance and ranging (LADAR) and vision-based stereovision systems are state-of-the-art developments among 3D imaging sensors. Researchers (e.g., [3–5]) have adopted flash LADAR to represent the work environment. This flash LADAR can acquire 3D information about the environment, which may correspond to obstacles at 30 frames per second [6]. Although previous studies have validated that flash LADAR could be used to provide 3D information in construction sites, the resolution of 3D information is limited, and the information captured are contaminated by noise, especially in outdoor environments [7]. Stereovision systems can also acquire 3D information with two or more cameras. Researchers (e.g., [8]) have adopted stereovision systems to represent the work environment. Although stereovision can provide texture and color information, not just 3D information, the applicability is limited in environments with few textures. Also, maximum depth and the precision of 3D information are limited.

In recent years, researchers working on autonomous robot navigation have been interested in representing the surrounding environment of a robot by relying on monocular camera inputs only (e.g., [9]). It has been proven that the monocular vision-based approach enables obstacle detection and environment representation at short range under certain environmental assumptions. In particular, it is possible to create a robust obstacle detector for low-speed maneuvers using monocular vision algorithms only in limited conditions. Earthmoving equipment needs to comply with speeds below 20 km/h within construction sites. With this point in mind, this study proposes a method for modeling the surrounding environment based on information from a single camera attached to the rear or all sides of the earthmoving equipment body.

The aim of this study is to propose an automated method for detecting nearby obstacles based on the use of a single camera, with the goal of protecting the equipment operator and construction workers from potentially dangerous situations, such as collisions between earthmoving equipment and obstacles within a certain proximity. This paper is organized as follows: in Section 2, the proposed approach is presented. Section 3 describes the proposed method in detail with experimental results. Finally, conclusions and suggestions for future research are given in Section 4.

2 Detection of Nearby Obstacles with Monocular Vision

In order to avoid collision, an equipment operator needs to pay attention to nearby areas with obstacles in meaningful dimensions while manipulating the equipment in a complex environment. Although it has been recognized in general that 3D representation is possible through more than two vision systems, the following prior information and assumptions allow distance estimation like that in the human vision between the camera and obstacles in 3D representation. Assume that the earthmoving equipment is lying on a plane and the height of the camera origin and azimuth (in degrees) are known. Then, the range of height of the obstacles and the range of distance between the camera origin and the obstacles are limited. Through projection transformation, we can obtain a point in 3D space and convert it to a 2D point in the image in pixels along with an optional *z*-value corresponding to the depth of each pixel.

In the case of an excavator, the lower caterpillar and the upper equipment body rotate separately around the axis of rotation. Figure 1 illustrates a top view of how the equipment body (represented by the yellow-orange color) at the top rotates while the caterpillar (represented by in black) is stationary. The dimensions of the caterpillar and equipment body were illustrated based on a 21-ton excavator. There are two areas: the area inside the inner circle and the area between the inner circle and the outer circle. The outer circle is a 12-meter circle around. The outer circle shows the extent to which obstacles should be detected if intersecting obstacles are set within 3 m. The inner circle represents the circle around the rotary long axis of the equipment body. This study focuses on obstacles within this outer circle by considering the speed of earthmoving equipment at the construction site. Then, for the purpose of preventing collisions with static or moving obstacles when the upper body of the equipment rotates, only objects with a height of 1 m or more are regarded as obstacles.

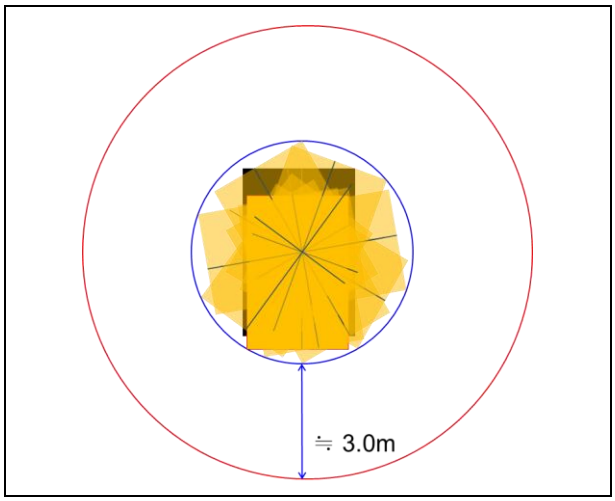

Figure 1. Example of the body rotation of the excavator from a top view

Unstructured environments are usually more complex than structured environment and have fewer features, which are obviously distinguishable [10]. Also, it is difficult to detect obstacles directly as obstacles have often irregular shapes, especially in work environment. Thus, this study proposes a different method of detecting obstacles by discarding irrelevant areas from the environment that are outside the specific range of the camera. The proposed method performs detection of nearby obstacles in a process, which is divided into three key steps presented in detail in the following subsections.

3 Methodology

3.1 Unwarping

Commercial rearview monitoring systems and surrounding view monitoring systems use wide-angle lenses in order to cover a larger area with fewer cameras. Although wide-angle lenses provide a large field of view, the images produced are severely distorted. Hence, a process called image unwarping is necessary to correct severe distortion of the images produced by the wide-angle lens. Image unwarping is used for obtaining a wide-angle view without strong aberration by both correcting lens distortion [11]. In order to exploit the large field of view of wide-angle lenses by correcting the lens distortion, this study adapted an unwarping approach proposed in Schulz et al. [12]. By generating the number of virtual views and merging them in a single image, we can re-project the original wide-angle image onto a semicylindrical surface. For this process, the semicylindrical surface is obtained for the camera with wide-angle lenses by exploiting the camera's extrinsic, intrinsic and distortion parameters [13]).

3.2 Shadow Removal

The outdoor environment in which the earthmoving equipment operates is affected by shadows such as those of obstacles, equipment, or surrounding objects or terrain features. Since image features such as color, texture, and intensity are used in the segmentation, the shadow is considered an irrelevant region because it is likely to be mistaken for the candidates of the obstacle. For this reason, once the correction of lens distortion is done prior to further processing, detection and removal of shadows is done to minimize the effect of shadows in the latter process. This study adapted the shadow detection and image restoration methods proposed by Sarabandi et al. [14] and Arévalo and Ambrosio [15]. In the study by Sarabandi et al. [15], color space transformation from red, green, and blue to $c_1c_2c_3$ was proposed since it shows the best results for detecting shadow regions in color images [16;14]. After color space transformation, a shadow region is defined as an area inside of the boundary that has a pixel value different from its surroundings. Therefore, the local variance of each pixel and its neighborhood is measured using the c_3 component to identify the shadow boundary by applying a 3-by-3 filter. The high variance value is the boundary between the shadow and non-shadow regions. After identifying this boundary, the shadow region can be detected by classifying pixels inside of the boundary as shadows. However, this method is limited for the following reasons. In an outdoor environment, c_3 may be noisy, which could cause a misleading finding of the boundary between the shadow and non-shadow regions. Also, the local variance becomes unstable for low saturation values (i.e. grey levels), which could also cause a misleading finding of the boundary between the shadow and non-shadow region (e.g., [15]). To overcome these limitations, Arevalo et al. [15] proposed an additional process to minimize the noise in c_3 and check the saturation and intensity values of pixels. Based on the shadow detection method proposed by Sarabandi et al. [14], this study adapted the advanced shadow detection method proposed by Arévalo and Ambrosio [15] and the linear-correlation method proposed by Sarabandi et al. [14] for image restoration.

3.3 Detection of Nearby Obstacles

3.3.1 Image segmentation

The ability to avoid obstacles depends on the image segmentation result, which is the process of separating obstacle candidates from the background. For this purpose, a rapid online segmentation method is proposed. Using the compact color and texture descriptor proposed by Blas et al. [17], this study integrates an intensity feature to reduce the effects of lighting changes during the segmentation process. In addition, a two-stage unsupervised online learning process is proposed. The integrated descriptors are first computed for each pixel and then clustered to find a small set of vectors or textons as the basis [18] for characterizing scene textures. In this process, each pixel is assigned to the closest texton. Then, the pixel classification result performed in the first step is refined by clustering the histograms of a small set of vectors over a window to find more coherent regions. The k-means clustering algorithm [19,20] was employed for the clustering.

3.3.2 Region of interest classification

For this purpose, the origins are defined for two different virtual cameras whose image plane are horizontal and aligned with the earth cardinal directions. The first virtual camera's origin is located in the bird's-

eye position, and the second virtual camera's origin is located below the original camera's origin. The first camera's field of view is looking at the floor vertically, and the second camera's field of view is parallel to the ground. The image plane of each virtual camera is projected onto the inertial planes using an adapted geometric method based on the concept of inverse perspective transformation.

Based on the change in the pixels of the segmented region from the original image to the two different virtual images, the proposed method distinguishes whether an object has height or not and then estimates the maximum height difference. The maximum difference of the height value between the ground and each transformed pixel within the obstacle is defined as the region's height, which is used to classify whether the region of interest has a height of at least 1 m.

4 Results and Discussion

The steps from the correction of lens distortion prior to further processing, shadow removal, and detection of nearby obstacles with a predefined height level via perspective transformations were applied on video streams acquired from a camera installed on the side of the equipment body while an excavator executed excavating and moving tasks. Figure 2 shows an example of an original image acquired using a wide-angle lens with a 135◦ horizontal field of view. As shown in the figure, it can be seen that the worker standing in the left area is distorted as if he were tilted compared to the worker in the middle.

Figure 2. Example of an original wide-angle image

Figure 3 shows an example of the resulted unwarped image. In Figures 2 and 3(a), it can be seen that the distortion of the leftmost worker and the second worker from the left has been corrected. The adapted shadow removal process was tested with the unwarped image in Figure 3(a). Detected shadow regions were shown in white pixels in Figure 3(b).

(a)

(b)

Figure 3. (a) Unwarped image of (b) Shadow detection result

Once the shadow regions are detected, the brightness of shadow pixels to the first order can be restored by a linear function. Figure 4 illustrates the image segmentation result. In the image, the dotted white dotted line represents the distance (5 m) from the camera. For this purpose, the segmentation is performed from the bottom, and the area above the white dotted line will no longer be segmented, except the areas segmented with the pixels below the white dotted line. This is also effective in reducing the time required for segmentation of unnecessary areas, such as blue areas. Once the image segmentation is performed, the area of the magenta color separated from the ground or terrain where the equipment is placed is excluded from the candidate of obstacles. Then, there are four candidates of obstacles, as shown in Figure 4.

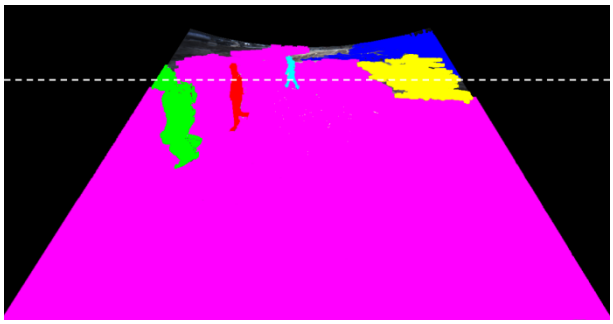

Figure 4. Image segmentation result

Among the regions of the candidates for obstacles, the regions that are determined to have height, whose maximum height difference was 1 m or more, and that were identified as obstacles were marked by boxes in Figure 5.

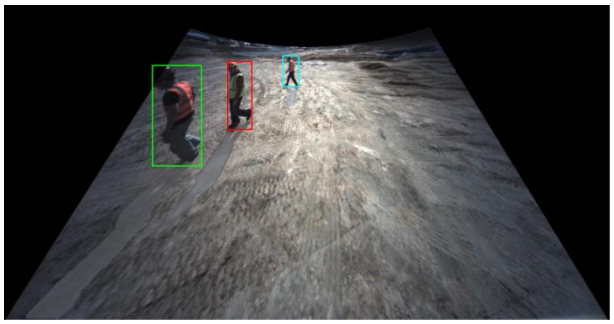

Figure 5. Region of interest classification result

5 Conclusion

This study presented a method for automated detection of nearby objects with monocular vision, with the goal of protecting the equipment operator and construction workers from potentially dangerous situations, such as collisions between earthmoving equipment and obstacles within a certain proximity. The experimental results showed that the proposed method could provide the equipment operator with information about nearby obstacles during the excavator's manipulation and transportation. It is expected that the proposed method could be implemented in rearview monitoring systems and surrounding view monitoring systems for operator assistance and to achieve active safety. Future works will focus on the estimation of collision-free area between earthmoving equipment and the obstacles. In addiition, obstacle detection is performed with planar ground assumption. Future work focuses on developing a method for detection of obstacles with no planar ground assumption by introducing an inertial sensor for acquiring 3D orientation and sensor network.

Acknowledgements

This research was supported by a grant (17SCIP-B079689-04) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA).

References

[1] ISO 5006: Earth-Moving Machinery – Operator's Field of View – Test Method and Performance Criteria. ISO, Geneva, 2006.

[2] Kim, C., Haas, C. T., and Liapi, K. A. Rapid, on-site spatial information acquisition and its use for infrastructure operation and maintenance. *Automation in Construction*, 14(5):666–684, 2005.

[3] Teizer, J., Caldas, C. H., and Haas, C. T. Real-time three-dimensional occupancy grid modeling for the detection and tracking of construction resources. *Journal of Construction Engineering and Management*, 133(11):880–888, 2007.

[4] Gong, J. and Caldas, C. Data processing for real-time construction site spatial modeling. *Automation in Construction*, 17(5):526–535, 2008.

[5] Son, H., Kim, C., and Choi, K. Rapid 3D object detection and modeling using range data from 3D range imaging camera for heavy equipment operation. *Automation in Construction*, 19(7):898–906, 2010.

[6] Chan, D., Buisman, H., Theobalt, C., and Thrun, S. A noise-aware filter for real-time depth upsampling. In *Proceedings of the ECCV Workshop on Multi-Camera and Multi-Modal Sensor Fusion Algorithms and Applications*, Marseille, France, 2008.

[7] Xiang, X., Li, G., Tong, J., and Pan, Z. Fast and simple super resolution for range data. In *Proceedings of the International Conference on Cyberworlds*, pages 319–324, Singapore, Singapore, 2010.

[8] Ishimoto, H. and Tsubouchi, T. Stereo vision based worker detection system for excavator. In *Proceedings of the International Symposium for Automation and Robotics in Construction (ISARC)*, pages 1004–1012, Montréal, Canada, 2013.

[9] Wybo, S., Tsishkou, D., Vestri, C., Abad, F., Bendahan, R. and Bougnoux, S. Obstacles avoidance by monocular multi-cue image analysis. In *Proceedings of the 15th World Congress on Intelligent Transport Systems*, pages 1–12, New York, NY, 2008.

[10] Yu, H., Zhu, J., Wang, Y., Jia, W., Sun, M., and Tang, Y. Obstacle classification and 3D measurement in unstructured environments based on ToF cameras. *Sensors*, 14(6):10753–10782, 2014.

[11] Bertozzi, M., Castangia, L., Cattani, S., Prioletti, A., and Versari, P. 360° Detection and tracking algorithm of both pedestrian and vehicle using fisheye images. In *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 132–137, Seoul, South Korea, 2015.

[12] Schulz, W., Enzweiler, M., and Ehlgen, T. Pedestrian recognition from a moving catadioptric camera. *Pattern Recognition. DAGM 2007. Lecture Notes in Computer Science*, 4713:456–465, 2007.

[13] Heng, L., Li, B., and Pollefeys, M. Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1793–1800, Tokyo, Japan, 2013.

[14] Sarabandi, P., Yamazaki, F., Matsuoka, M., and Kiremidjian, A. Shadow detection and radiometric restoration in satellite high resolution images. In *Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, AK, 2004.

[15] Arévalo, V., González, J., and Ambrosio, G. Shadow detection in colour high-resolution satellite images. *International Journal of Remote Sensing*, 29:1945–1963, 2008.

[16] Salvador, E., Cavallaro, A., and Ebrahimi, T. Shadow identification and classification using invariant color models. In *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1545–1548, Salt Lake City, UT, 2001.

[17] Blas, M. R., Agawal, M., Sundaresan, A., and Konolige, K. Fast color/texture segmentation for outdoor robots. In *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pages 4078–4085, Nice, France, 2008.

[18] Leung, T. and Malik, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[19] Duda, R. and Hart, P. *Pattern classification and scene analysis*, Wiley, New York, NY, 1973.

[20] Jain A. and Dubes, R. *Algorithms for clustering data*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1988.