Safe Operation of Autonomous Machines

Raj Aggarwal
Adjunct Professor, ECpE
Iowa State University
Ames, Iowa, USA
rka@iastate.edu

Haoyuan Lin
Graduate Student, ECpE
Iowa State University
Ames, Iowa, USA
linhaoyuan@gmail.com

***Abstract*—Reliable detection of foreign objects is a key requirement for safe operation of autonomous machines. In farming applications, an object is considered a foreign object if it can damage the machine or be damaged by it. The traditional machine vision approaches rely on detecting and classifying each type as a separate class of thereby increasing the computational load and compounding the machine vision problem since these machines have to operate in real-time and the foreign objects can appear in any orientation thereby increasing complexity. Moreover, it is an over kill since the safe operation of an autonomous machine does not require that we classify these objects as long as we can reliably detect them and direct the machine to take appropriate maneuvering action. In our application, the object of interest is bales and everything else is a foreign object. The foreign objects include humans, animals, vehicles and standing crop. We use disparity information from the two cameras in a stereo configuration and use the camera model to calculate the distance to the object. This object detection framework based on distance and size has proved to be more efficient and robust compared to traditional machine vision approaches.**

Keywords-machine vision, autonomous operation, safe operation, stereo configuration, disparity

I. INTRODUCTION

Most object detection research focuses on how to design algorithms which are both accurate and fast and treat each type of object as a separate class [1, 2, 3]. However, in our application, the task is to detect foreign objects with arbitrary shape and size and reliably differentiate them from the bales. A straight forward method is to simply divide the foreign object into multiple categories and use traditional multi-class object detection & classification algorithms. However, this is not computationally efficient or robust because of infinite variations in size and shape. Moreover, the deformation of the object may also degrade the performance of this approach. Since our problem is to reliably detect in real-time and not classify, we used two cameras in a stereo configuration to generate a distance map and find blocks of certain size. The merit of this method is to consider a foreign object as a block and not care which category it belongs to as long as its size is within the range. We designed a detection framework and associated algorithms to detect foreign objects and their locations that are within certain volume at a given distance.

The rest of the paper is structured as follows. In Section II, we describe the framework and algorithm using depth map to detect foreign objects. In Section III, we present the experimental results. Finally, Section IV concludes the paper.

II. STEREO CONFIGURATION APPROACH

A. 3-D Reconstruction

The goal of stereo vision is mainly to recover the 3D structure of the scene using two or more images acquired by cameras in a stereo configuration. With known camera configuration, a disparity map can be generated by calculating disparities of all the pixels in an image. One method to calculate the disparity is using the feature matching [4, 5], such as edges. The edge features can be derived for both the left and right images by using Gaussian filters. The features are then matched by comparing their orientations and strength. In the disparity map, the value for each pixel is the distance between the pixels which has the highest match score.

The camera model is shown is Fig. 1 and its parameters are described in Table 1.

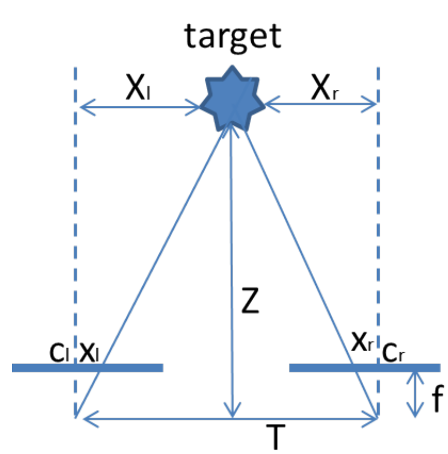

Fig. 1 The camera model

TABLE I. CAMERA MODEL PARAMETERS

Distance between the two cameras	T
Focus length of the cameras	f
Distance between target and camera	Z
Position of middle point of camera film	c
Position of object in camera file	x
Disparity of the target	d
Displacement between target and camera	X

Based on the triangulation principle, we have

$$\frac{x_l}{X_l} = \frac{f}{Z} \quad (1)$$

$$\frac{x_r}{X_r} = \frac{f}{Z} \quad (2)$$

This implies $$X_l = \frac{Z}{f} x_l \quad (3)$$

$$X_r = \frac{Z}{f} x_r \quad (4)$$

$$X_l + X_r = T \quad (5)$$

And $$\frac{Z}{f}\left(x_l + x_r\right) = T \quad (6)$$

Where $x_l + x_r$ is the disparity "d". So, the distance between the target and the camera is given by $Z = \frac{Tf}{d}$

B. Foreign Object Detection Framework

Disparity map is widely used in the computer vision applications to recover the 3D structure of the scene [6, 7]. The framework of the foreign object detection system using disparity map is shown is Fig. 2. The following sections will describe the flow chart in detail.

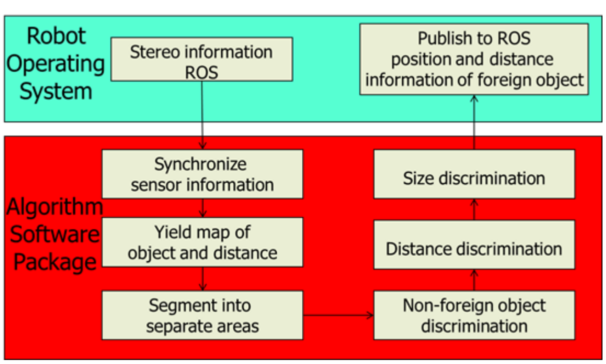

Fig. 2 The framework of the foreign object detection

The left camera and the right camera are of the same configuration and calibration. The pointing direction of both the cameras is the same and can use ground as a reference. The calibration step is done once offline before the system is used.

Since the distance similarity is the only feature used for detecting objects, the accuracy of the depth map [8, 9] is critical for the success of the algorithm. In order to make the distance calculation robust, the area of each frame is divided into sets of blocks instead of using each pixel. The edge feature within a block in the left frame can be used as the pattern to search in the right frame. For the purpose of the foreign object detection task instead of 3D construction, the depth calculation can be done coarsely. From our experiment, if the area of the block is too small, some holes will be shown is the depth map. Meanwhile, if the area of the block is too big, the object may not be detected, especially when the object is far away from the camera.

Fig. 3 shows the depth map when there is no foreign object using the 50*50 pixel block. In the figure below, the positive number means the distance between the background and camera in meters. The negative value means one of two things. One reason is that there is no matching block from the left camera frame to the right camera frame. A portion of the scene captured by the left camera may not be captured by the right camera. Since we use feature block in the left camera to find a match in the right camera, the distance along the left vertical line is negative. The second reason is the distance is too far and outside the range of interest. The disparity for such a block may be zero since the feature block is the same when watching from a long distance. The area in the sky is too far and the feature block looks the same. In either case, no special operation is required by the machine.

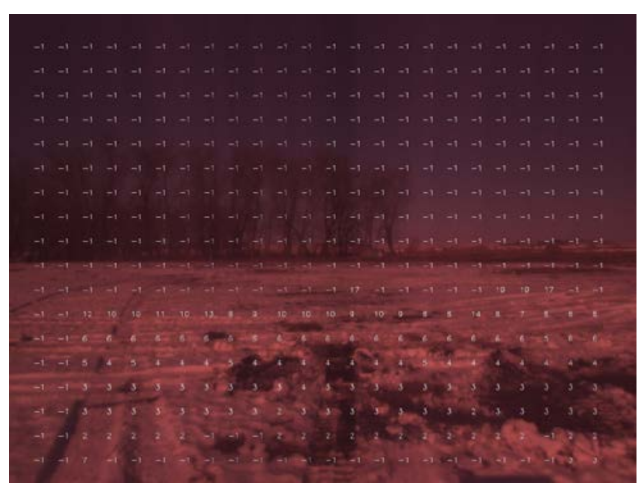

Fig. 3 The depth map when there is no foreign object

Once the depth map for the initial scene is generated, the system is ready for autonomous operation. The distance of the new frame is calculated and compared with the initial depth map. The distance filter us then used to separate the background object. Any feature block may be ignored if it is outside the critical range. A feature block is denoted if the distance is within the critical range. In our work, the ignored feature block is denoted as a negative sign and the blocks of interest are denoted with a positive sign, as shown in Fig. 4. After the processing of the filter, block fill algorithm is used to connect the neighbor blocks into one integrated block. For all blocks that are denoted by positive sign, the breadth first search algorithm with the neighbor rule is used to find all positive sign blocks and to mark these positions as a group. Each group represents one object. In our experiment setup, we use 8-neighbor rule to recognize the neighbor candidate around one block. We consider that the foreign object can be shown as any kind of shape; all 8 directions are considered as extension of the foreign object.

The size discrimination is used to create a decision table with the distance and object size information. When a small object is too close to the camera, the size of the pixel block is shown as a big block. By using the decision table we store the low-bound level of the size that has high confidence.

Fig. 4 shows the output of the filter based on the depth map when a person is walking in front of the camera.

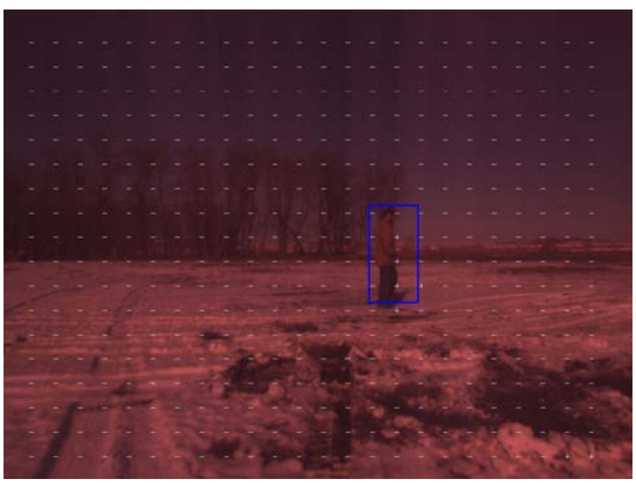

Fig. 4 The result after the block fill algorithm when one person is walking in front of the camera

III. EXPERIMENTAL RESULTS

In our experiment, we used nine data sets, shown in Table 2, to test the correctness of the algorithm. The data sets include two kinds of objects, a pedestrian and a vehicle. For each data set, the "# frames" means the number of frames in the video clips and the "# object" means the times the object appears in the video clip. We also note the moving direction of the object for the purpose of testing all corner cases.

The foreign object could move in any direction along a set route. It could move towards or away from the camera. It could move from left to right or right to left. The moving speed was slow to normal. The pose of the objects was changed from erect to leaning.

We compared our stereo-based algorithm based on the depth information with the previous work using the multi-classifier method based on the shape information [10]. From the results, the detection rate improved over the shape-based method. Besides that, the false alarm also decreased over the shape-based method. Sometimes, the foreground and background may mix together to make the frame area appear as a target of interest. However, the distance to the object is not always in the range. Such cases can be eliminated by using the depth information.

TABLE II. EXPERIMENT RESULT COMPARISON BETWEEN STEREO-BASED AND SHAPE-BASED METHOD

Datasets	Stereo-based			Shape-based		
	Detected	Miss	False alarm	Accuracy	Miss	False alarm
Dataset1 (42objects/ 116frames)	40	2	0	39	3	2
Dataset2 (15objects/ 34frames)	14	1	0	13	2	0
Dataset3 (12objects/ 56frames)	11	1	0	10	2	0
Dataset4 (22objects/ 60frames)	21	1	1	21	1	2
Dataset5 (12objects/ 46frames)	11	1	0	9	3	2
Dataset6 (27objects/ 51frames)	24	3	0	24	3	1
Dataset7 (8objects / 35frames)	7	1	0	6	2	0
Dataset8 (18objects/ 38frames)	13	5	0	7	11	1
Dataset9 (8objects / 37frames)	8	0	0	7	1	0

IV. CONCLUSION

In this paper, we describe a computer vision approach which is robust and efficient in detecting foreign objects with no pre-set shape, essential for safe operation of autonomous machines. We use a stereo configuration to generate a depth map. We then use a stereo matching algorithm to get the disparity information based on intensity images from stereo cameras and using the camera model to retrieve the distance information. From the result of our experiments, the proposed framework has a better performance with higher detection rate with lower false alarm.

The algorithm performed very well at short ranges (<10 meters); however, not as well object of interest is further away. One could investigate ways to improve the range accuracy by more rigorous modeling. The target classification accuracy can also be improved by incorporating shape information. The object tracking algorithm can also be improved by sequential frame processing.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pp 886–893, 2005.

[2] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):pp 743–761, 2012.

[3] S.Tang, M. Andriluka and B. Schiele, "Detection and tracking of occluded people," International Journal of Computer Vision, pp 1–12, 2012.

[4] T.D. Sanger, "Stereo disparity computation using Gabor filters", Biological Cybernetics, vol 59, no. 6, pp 405-418, Oct. 1988, doi: 10.1007/BF00336114

[5] P. Kauff, N. Atzpadin, C. Fehn, M. Muler, O. Schreer, A. Smolic and R. Tanger, "Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability", Signal Processing: Image Communication, Vol 22, No. 2, pp 217-234, Feb. 2007, doi: 10.1016/j.image.2006.11.013

[6] S. Izadi, et al., " Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera", Proc. of the 24th Annual Symposium on User Interface Software and Technology, pp 559-568, ACM 2011.

[7] M. Jenkin, A. D. Jepson and J. K. Tsotsos. "Techniques for disparity measurement." CVGIP: Image Understanding 53, no. 1 (1991): 14-30.

[8] S. Battiato, A. Capra, S. Curti, and Marco La Cascia. "3D stereoscopic image pairs by depth-map generation,". Proc. 2nd International Symposium on Data Processing, Visualization and Transmission, pp. 124-131. IEEE, 2004.

[9] K. Ikeuchi, "Constructing a depth map from images." In AI Memo AIM-744, Artificial Intelligence Laboratory, MIT. 1983.

[10] H. Lin, "Foreign object detection (FOD) using multi-class classifier with single camera vs. distance map with stereo configuration." (2015), Graduate thsis, Paper 14565.

 ISBN: 978-1-61208-451-0