Inter-vehicle Distance Detection Based on Keypoint Matching for Stereo Images

Yoshihiro Shima
School of Science and Engineering
Meisei University
Hino, Tokyo, Japan 191-8506

Abstract— **An algorithm to detect car distance from a pair of stereo images is presented. It is useful for drivers to avoid collisions and ensure safety to keep the car at a constant distance from the car ahead. The conventional distance detection method is based on image matching; the proposed algorithm is based on key-point matching. Key points are extracted from a stereo image pair by using Speeded Up Robust Features (SURF) . The distance is calculated from 3D binocular disparity, the difference of position at the object.**

Keywords- distcnce detection; 3D image; key-point matching; stereo vision; car control

I. Introduction

In vehicle recognition [1][2], the ability to find similar parts in a pair of stereo images is a highly relevant problem [3][4][5]. A popular method is to estimate features that best describe a chunk of data and that can be used to efficiently perform comparisons between different data regions [6]. In 3D perception, such features are usually local around a point in the sense that for a given point in the scene its neighborhood is used to determine the corresponding feature.

The entire task is typically subdivided into two subtasks, namely the identification of appropriate points, often referred to as interest points or key points, and the way in which the information in the vicinity of the points is encoded in a descriptor or description vector [7][8][9]. An important advantage of interest points is that they substantially reduce the search space and computation time required for finding correspondences between two scenes.

As a stereo algorithm, our approach contains three novelties. First, the image sampling problem is overcome by using a feature vector of key points that is insensitive to sampling. Second, the algorithm handles the optical geometry of non-parallel cameras, which presents an explicit equation for binocular disparity and distance. Finally, a distance histogram of interest points is used to reduce the effects of the background image of the car.

II. Distance Detection Method

Human ability to perceive depth comes from viewing an object along two different lines of sight. This phenomenon is called parallax. A two-lens camera, also known as a stereo camera, is called a 3D camera. A 3D digital camera is fitted with dual lenses to capture left and right images [10].

Fig. 1 shows the flowchart of detecting distance. First, a pair of left and right stereo images is obtained by the 3D digital camera. Next, interest (key) points are selected at distinctive locations in the image, such as corners, blobs, and T-junctions by using the Speeded up Robust Features (SURF) method [9]. Corresponding points in a stereo pair are selected by key-point matching. Finally, the 3D binocular disparity, the difference of location for corresponding points , is detected. The depth of the points is calculated from the parallax. To discriminate the car rear area from the background image, the mode from the histogram of the depth of points is selected as the distance between them.

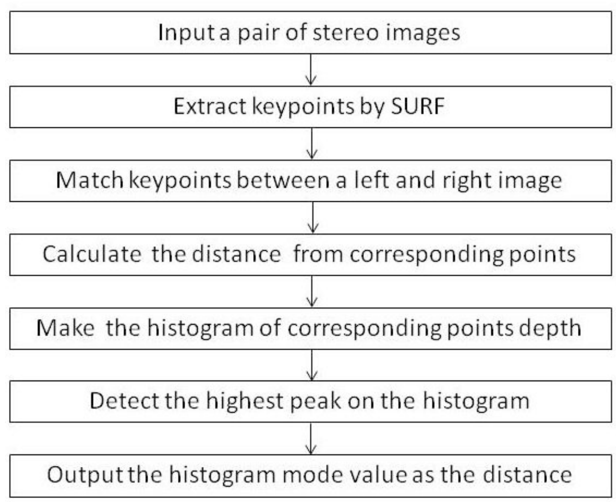

Figure 1. Flowchart of detecting distance.

A. Extraction of Key Points in Images by SURF

SURF is a scale- and rotation-invariant interest point detector and descriptor. It can be computed much faster than the previously proposed Scale Invariant Feature Transform (SIFT) method [8]. SURF is achieved by relying on integral images for image convolutions. It is also achieved by using a Hessian matrix-based measure for the detector, and a distribution-based descriptor. Fig. 2 shows an example of key points (shown as red dots) extracted by the SURF detector.

978-1-5386-1937-7/17/$31.00 ©2017 IEEE

Figure 2. Example of key points (shown as red dots) extracted by the SURF detector.

B. *Extraction of Corresponding Points by Matching Key Points between Stereo Image Pairs*

Key-point matching is based on minimum value detection of the Euclidean distance in the feature space between a pair of left and right stereo images. Key points with the minimum distance between them in a stereo image pair are called corresponding points. Fig. 3 shows key-point matching for a stereo image pair. End points of each red line show the corresponding points. These images are captured at a distance of 0.5 m from the car ahead. Fig. 4 shows an example of corresponding points.

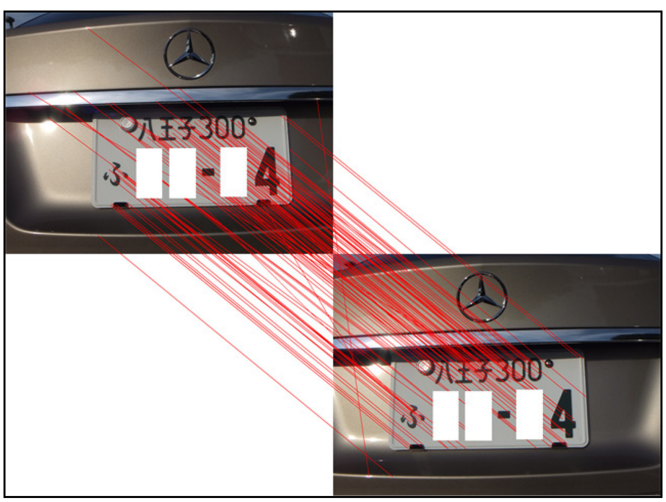

Figure 3. Example of key-point matching for a stereo image pair at 0.5 m.

Figure 4. Example of corresponding points for a left image at 5.0 m. Near points are shown as red dots and far points as green ones.

C. *Calculation of the Depth at Corresponding Points*

The 3D position of a point P is expressed as (Xp, Yp, Zp). The distance Zp can be calculated from 3D binocular disparity. Fig. 5 shows an optical system using stereo vision. Here, let us let OL and OR denote the lens center for the 3D camera, f denote the lens' focal length, d denote the distance between the two lenses, and C denote the convergence point distance. The convergence angle is denoted as α.

Equations (1) and (2) show the distance Zp. The equations are derived from the X positions of the points P, XL and XR in a pair of left and right images.

$$Z_p = \frac{d}{\frac{1}{\tan(\alpha + \beta_L)} + \frac{1}{\tan(\alpha + \beta_R)}} \tag{1}$$

$$Z_p = d \cdot \frac{(fc)^2 + fc\frac{d}{2}(x_L + x_R) + (\frac{d}{2})^2 x_L x_R}{2f^2\frac{d}{2} \cdot c - 2 \cdot \frac{d}{2}c \cdot x_L x_R + f((\frac{d}{2})^2 - c^2)(x_L + x_R)} \tag{2}$$

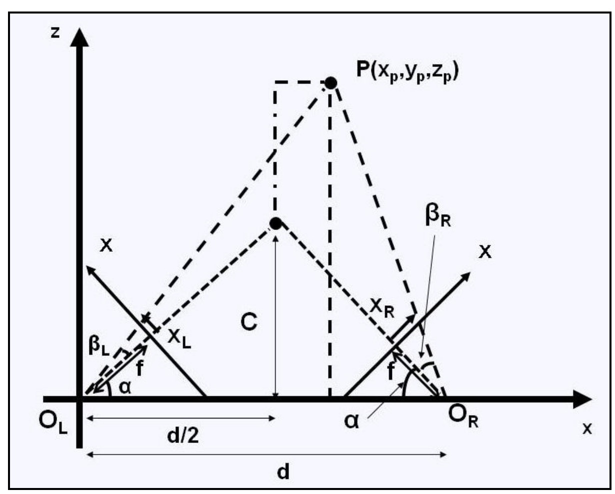

Figure 5. Optical geometry of non-parallel 3D digital cameras.

D. *Point Depth Histogram*

To select the corresponding points of the object area from the whole scenery image, it is necessary to discriminate the object area from the background area. The histogram is made from the depth of corresponding points. The mode, the highest peak in the form of the histogram, is detected as the distance of the object. Fig. 6 shows an example of the histogram of the corresponding point depth in stereo images at 4.0 m.

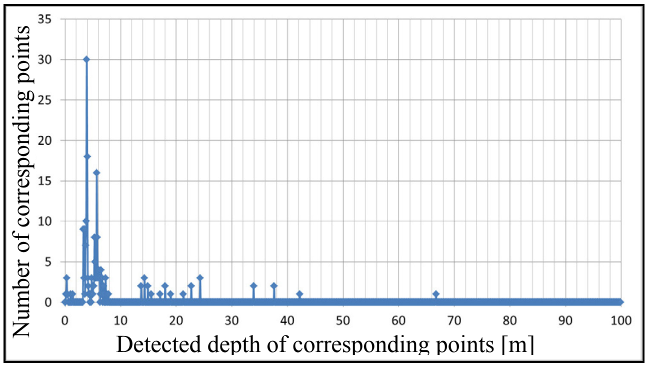

Figure 6. Example of the histogram of corresponding point depth in stereo images at 4.0 m.

III. Distance Detection Experiment

A stereo camera was used to take left and right images of the rear of the car. The distance between the camera and the car was manually measured. Fourteen images were taken at a distance from 0.5 m to 5.0 m. The image size was 3648 x 2736 pixels.

The source code for distance detection is written by C++. OpenCV2.1 was used for SURF and image functions. Processing time was 7.4 seconds for a stereo pair obtained with a CPU with Intel Core size of 3.1 GHz.

A. *3D Camera*

The FinePix REAL 3D W1 digital camera has two lenses, each capturing color images at 10 megapixel resolution [10]. Fig. 7 shows an external appearance of this camera. Focal length is f=7.8 mm. The crossing point distance is c=2.0 m. The distance between the two lenses is d=75.0 mm. Fig. 8 shows an external appearance of a camera tripod attached with the 3D digital camera. The camera altitude over the road surface is 885 mm.

Figure 7. External appearance of 3D digital camera. Two sets of lenses work to capture stereo images in the same way as human eyes.

Figure 8. External appearance of camera tripod attached with 3D digital camera.

B. *Binocular Disparity and Object Distance*

Binocular disparity refers to the difference in image location of an object captured by the left and right cameras, resulting from the horizontal separation of two lenses. Fig. 9 shows object distance and 3D binocular disparity, derived from Equation (2). Fig. 10 shows the 3D binocular disparity that was manually measured from a pair of left and right stereo images.

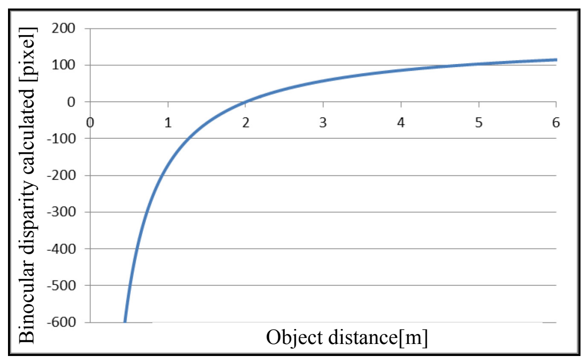

Figure 9. Object distance and 3D binocular disparity derived from Equation (2).

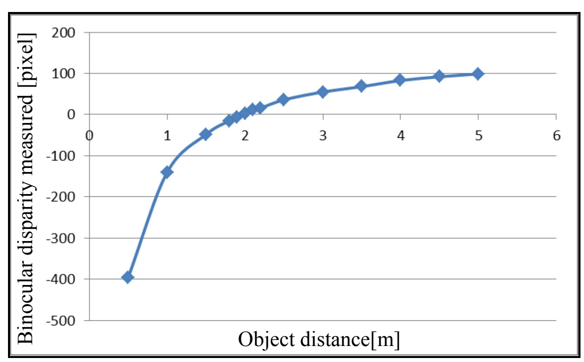

Figure 10. Object distance vs. manually measured 3D binocular disparity.

C. *Stereo Image Pairs*

Multiple pairs of left and right stereo images were captured by the 3D digital camera. The distance between the camera and the car was manually measured. Fourteen pairs of stereo images were taken at distances ranging from 0.5-5.0 m for the same car rear portion. Figs. 11, 12, 13 and 14 respectively show examples of a pair of left and right stereo images at 0.5, 2.0, 3.0 and 5.0 m.

(a) Left image

(b) Right image

Figure 11. Example of stereo image pair at 0.5 m.

(a) Left image

(b) Right image

Figure 12. Example of stereo image pair at 2.0 m.

(a) Left image

(b) Right image

Figure 13. Example of stereo image pair at 3.0 m.

(a) Left image

(b) Right image

Figure 14. Example of stereo image pair at 5.0 m.

D. *Corresponding Points*

The condition under which corresponding points are selected from key points is that a key point pair in the left and right images is the nearest and the Euclid distance between them is under 0.1 in the SURF feature space. Fig. 15 shows an example of corresponding points for a right image at 5.0 m. Near points are shown as red dots and far points as green dots. Fig. 16 shows an example of key-point matching for a stereo image pair at 5.0 m.

Fig. 17 shows the total number of key points and manually measured distance. Fig. 18 shows the total number of corresponding points and manually measured distance. Fig. 19 shows the number of corresponding points at the peak and manually measured distance.

Figure 15. Example of corresponding points for a right image at 5.0 m. Near points are shown as red dots and far points as green ones.

Figure 16. Example of key-point matching for stereo image pair at 5.0 m.

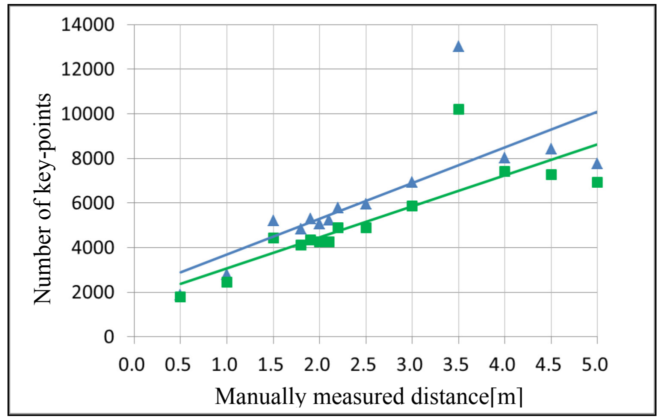

Figure 17. Number of key points and manually measured distance.

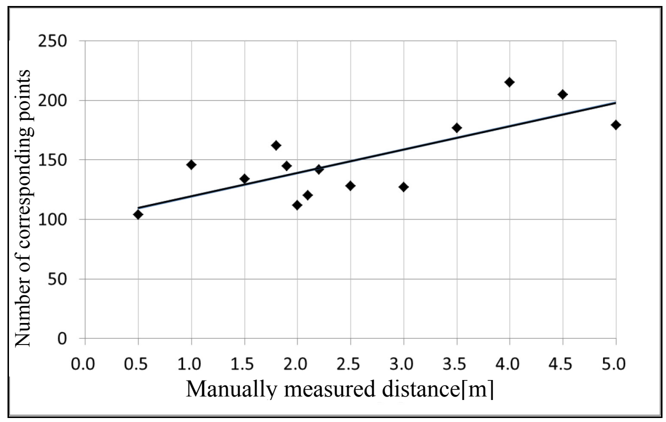

Figure 18. Number of corresponding points vs. manually measured distance.

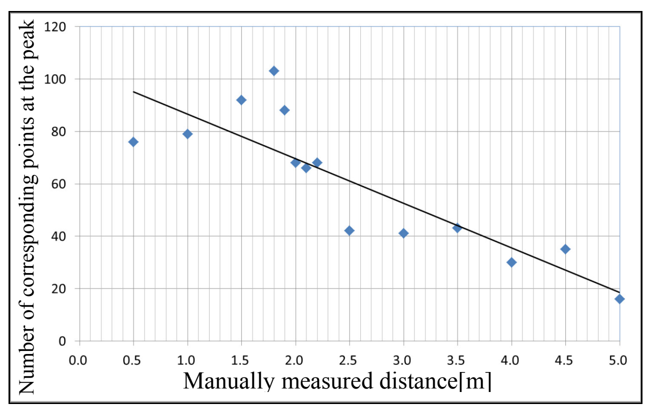

Figure 19. Number of corresponding points at peak vs. manually measured distance.

E. *Experimental Results of Distance Detection*

Fig. 20 shows the histogram of corresponding point distance in stereo images at 0.5 m. The peak point is 0.7 m as the detected distance. Fig. 21 shows the histogram of corresponding point depth in stereo images at 2.5 m. The peak point is 2.6 m. Fig. 22 shows the histogram of corresponding point distance in stereo images at 5.0 m. The peak point is 5.0 m.

Fig. 23 shows the relation among detected distance and manually measured distance. Table 1 shows the range error. The distance error is within the +0.2 m to -0.1m range. The error rate of distance is lower at the far position.

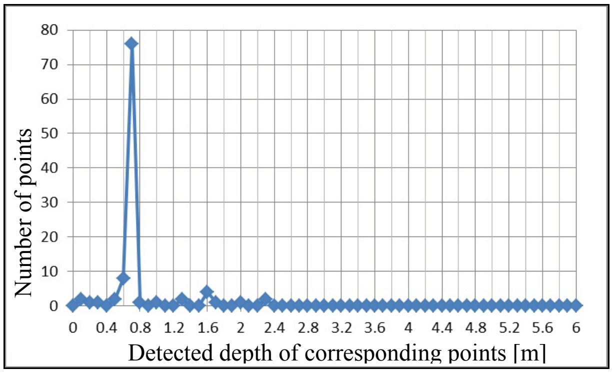

Figure 20. Histogram of corresponding point depth in stereo images at 0.5 m.

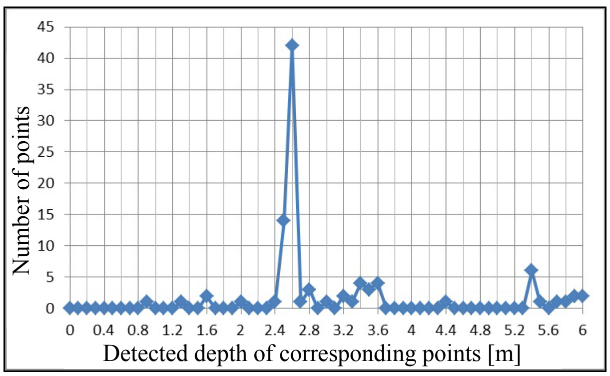

Figure 21. Histogram of corresponding point depth in stereo images at 2.5 m.

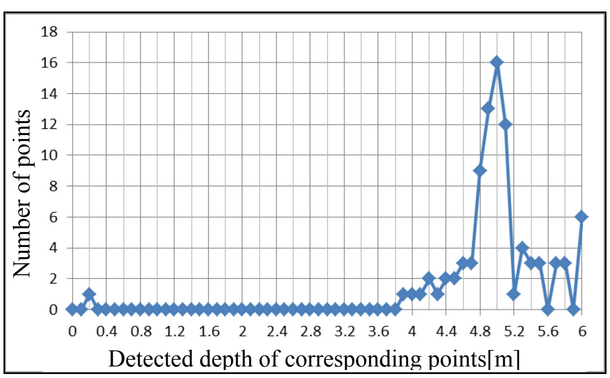

Figure 22. Histogram of corresponding point depth in stereo images at 5.0 m.

IV. CONCLUSION

To determine the distance between cars as a means to avoid accidents, we used key-point selection and matching between stereo pairs of images to detect the distance between a car and the car ahead of it. For this purpose we used a point depth histogram to discriminate the distance between a car and the background image behind it. In future work we will aim at long distance detection where the car is far over 100 m behind the car ahead.

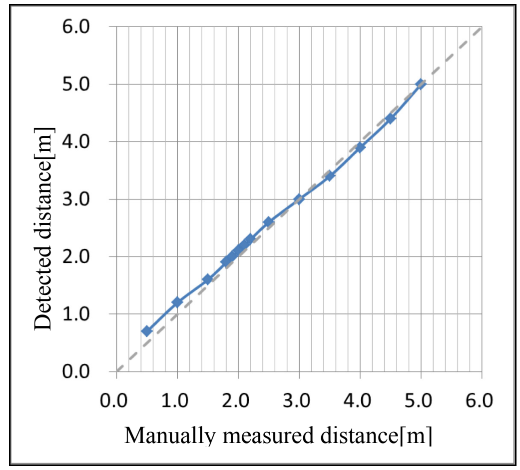

Figure 23. Detected distance vs. manually measured distance.

TABLE I. RANGE ERROR

Manually Measured Distance [m]	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Range Error [m]	0.2	0.2	0.1	0.1	0.1	0.0	-0.1	-0.1	-0.1	0.0

REFERENCES

[1] Z. Sun, G. Bebis and R. Miller, “On-road vehicle detection: a review, ” IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 28, Issue 5, pp. 694–711, 2006.

[2] M. Rezaei and M. Terauchi, “Robust Vehicle Detection and Distance Estimation Under Challenging Lighting Conditions, ” IEEE Transactions on Intelligent Transportation Systems, Volume 16, Issue 5, pp. 2723-2743, 2015.

[3] T. Kowsari, S. S. Beauchemin and J. Cho, “Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost, ” ITSC 2011 14th International IEEE Conference on Intelligent Transportation Systems, pp. 1255-1260, 2011.

[4] H. H. Baker and T. O. Binford, “Depth from edge and intensity based stereo, ”IJCAI'81 Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2 , pp. 631–636, 1981.

[5] Y. Ohta and T. Kanade. “Stereo by intra- and interscanline search using dynamic programming. ”IEEE Trans. on PAMI, 7(2),pp.139–154, 1985.

[6] S. Birchfield and C. Tomasi, “Depth Discontinuities by Pixel-to-Pixel Stereo, ”Proceedings of the Sixth IEEE International Conference on Computer Vision,Mumbai, India, pp. 1073-1080, January 1998.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. Journal of Computer Vision, 60, 2, pp. 91-110,2004.

[8] D. G. Lowe, "Local feature view clustering for 3D object recognition," IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, pp. 682-688, December 2001.

[9] H. Bay,A. Ess,T. Tuytelaars and L. V. Goola, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.

[10] FUJIFILM, 3D Image Products, http://www.fujifilm.com/products/3d/.
.