Multi Cost Function Fuzzy Stereo Matching Algorithm for Object Detection and Robot Motion Control

Akhil Appu Shetty [1], Navya Thirumaleshwar Hegde [2*], Aldrin Claytus Vaz [3], C R Srinivasan [4]

[1, 4] Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

[2] Department of Aeronautical and Automobile Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

[3] Department of Electronics and Communication Engineering, St. Joseph Engineering College, Mangalore, India

Email: [1] akhilshetty88@gmail.com, [2] navya.hegde@manipal.edu, [3] aldrinv@sjec.ac.in, [4] cr.srinivasan@manipal.edu

*Corresponding Author

Abstract—Stereo matching algorithms work with multiple images of a scene, taken from two viewpoints, to generate depth information. Authors usually use a single matching function to generate similarity between corresponding regions in the images. In the present research, the authors have considered a combination of multiple data costs for disparity generation. Disparity maps generated from stereo images tend to have noisy sections. The presented research work is related to a methodology to refine such disparity maps such that they can be further processed to detect obstacle regions. A novel entropy based selective refinement (ESR) technique is proposed to refine the initial disparity map. The information from both the left disparity and right disparity maps are used for this refinement technique. For every disparity map, block wise entropy is calculated. The average entropy values of the corresponding positions in the disparity maps are compared. If the variation between these entropy values exceeds a threshold, then the corresponding disparity value is replaced with the mean disparity of the block with lower entropy. The results of this refinement are compared with similar methods and was observed to be better. Furthermore, in this research work, the v-disparity values are used to highlight the road surface in the disparity map. The regions belonging to the sky are removed through HSV based segmentation. The remaining regions which are our ROIs, are refined through a u-disparity area-based technique. Based on this, the closest obstacles are detected through the use of k-means segmentation. The segmented regions are further refined through a u-disparity image information-based technique and used as masks to highlight obstacle regions in the disparity maps. This information is used in conjunction with a kalman filter based path planning algorithm to guide a mobile robot from a source location to a destination location while also avoiding any obstacle detected in its path. A stereo camera setup was built and the performance of the algorithm on local real-life images, captured through the cameras, was observed. The evaluation of the proposed methodologies was carried out using real life out door images obtained from KITTI dataset and images with radiometric variations from Middlebury stereo dataset.

Keywords—Disparity Refinement; Kalman Filter; KITTI Dataset; K-means Segmentation; Obstacle Detection; Robot Navigation; Stereo Disparity Map.

I. INTRODUCTION

Computer vision applications have an immense use for stereo matching in situations which require 3-D reconstruction, view synthesis and autonomous navigation. For obtaining this 3-D information, the similarities between the image pairs are considered [1].

The main goal of stereo matching algorithms is the extraction of 3-D information from a scene by considering the similarities at different viewing positions, between the given image pairs acquired. Categorizing them based on the density of the results, these algorithms can be mainly split into dense stereo matching and sparse stereo matching. The aim of dense stereo matching algorithms is to determine the correspondences of every pixel in images, whereas sparse stereo matching algorithms mainly consider only the features within the scene and not all the pixels. This topic was exhaustively reviewed in [2, 3]. Apart from the above-mentioned classification, stereo algorithms can also be categorized based on the complexity of calculation into local and global methods. Small patches of pixels or windows of the reference image are used in local stereo matching approaches. The disparity of the pixels is determined by comparing this information in the window region with the patches in the target image. Global approaches, obtain results through a more complicated manner in which all the disparities are concurrently calculated by applying energy minimization methods. Generally, local stereo algorithms are faster than their global counterparts and exhibit a smaller memory requirement. Unfortunately, the results obtained through these methods are usually inclined to generate results with a lower accuracy when compared to their global counterparts. Numerous local algorithms based on adaptive weight [4, 5] can achieve results similar to those obtained using global stereo matching methods. One of the major drawbacks of these local algorithms is the high computational complexity that varies quadratically when compared to the window size used to aggregate the matching costs.

Obtaining dense results (disparity maps) from stereo matching algorithms would usually include the following four steps: (1) matching cost computation (2) cost aggregation (3)

disparity cost computation/optimization and (4) disparity refinement. As has been previously mentioned, these algorithms can be categorized in to two broad clusters namely local stereo matching algorithms and global stereo matching algorithms [6]. The aggregation of matching costs for local algorithms would be decided based on factors such as colour values or their image pixel intensities. During the selection of these support windows, it is implicitly assumed that values within this region would usually be smooth. The decision of the best disparity is taken through a winner takes all (WTA) approach. Unlike the local approaches, the global techniques assume explicit smoothness. These algorithms use complex optimization functions that decide the disparity of the present pixel based on the calculated disparities of other pixels [7]. Local algorithms generally compute disparities faster but have a tendency of generating less reliable matches in texture less areas and also where there are sudden discontinuities in depth. Comparatively, global algorithms generate a much more accurate result but, due to the computation complexity, are more expensive. The methodology presented in this work uses local algorithms.

The generation of the disparity maps from stereo image pairs would present us with viable results. But even though the results seem to be accurate, it would always be advisable to perform a post processing technique to further enhance the obtained results. A majority of the authors have proposed a pixel based refining technique to further reduce the errors of their methods. But in our case, since we are working with a super pixel-based aggregation technique for disparity calculation, the disparity values of a patch will be similar [8-10]. In such a situation, where disparity maps are generated in terms of patches rather that pixel-wise, it would be inappropriate to go ahead with a pixel-based approach as majority of the refinement techniques do. We would need to perform operations on the disparity maps in terms of patches as well. Rather than taking only a single cost function to evaluate the level of similarity, the authors have used a combination of normalized mutual information (NMI), normalized cross-correlation (NCC), and absolute differences of gradients on the image pairs. Along with this, the disparity selection is made through a fuzzy logic-based approach. The main contribution of this research work appears in the form of a refinement technique for the fuzzy- generated disparity map. An entropy-based refinement approach is utilized for this. The obtained disparity map had been further refined through a v-disparity and u-disparity based technique for detection of objects present in the reference images. These objects are further categorized into three groups based on its distance from the camera. This information is then provided to a path planning algorithm and implemented on a robot. The authors adopted feature-based matching algorithm over deep learning methodologies to overcome higher processing power [11-13]. In application which work on limited resources, this would prove to be quite a limitation.

In the rest of the paper, Section II discusses the literature survey of similar work, Section III presents the methodology followed and the algorithms developed to carry out this research work. The results and discussions are also presented in the respective sections. The conclusion is stated in Section IV.

II. RELATED WORK

Stereo vision is considered as a complex and complete perception due to its inherent nature to calculate distance (depth) information of the objects in an image as compared to monocular vision. Multiple experimentations on autonomous ground vehicles with this methodology have presented notable results. Most of the algorithms require a few assumptions related to the approximated free space on the ground to achieve considerable performance when it comes to long-range and complicated road conditions.

The disparity maps generated from stereo images would present feasible results to advance to the next subsequent stages of the system. While the results seem to be sufficient to proceed, it is always advisable to include a post processing stage to further enhance the obtained results. A majority of the authors have proposed a pixel-by-pixel based refining technique to further reduce the errors of their methods. But in the presented research, a super pixel-based [14-15] aggregation technique for disparity calculation is considered. Hence, the disparity values within a patch will be similar. In such a situation, where disparity maps are generated in terms of patches rather that pixel wise, it would be unfitting to go ahead with a pixel-based approach as majority of the refinement techniques do. In such situations, it becomes necessary to perform operations on the disparity maps in terms of patches too. Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) stereo dataset was used for the development of the refinement technique as it presented a variety of real-life outdoor situations which are more suited for this application. With the disparity maps now being refined and enhanced, the next step would be the isolation of the obstacles. For this it becomes mandatory to segment and highlight the regions of interest in the images. This has been accomplished by extracting certain information from the disparity maps themselves. This is an extremely critical and fundamental task in autonomous vehicles as they are able to perceive the environment through it.

In a dynamic environment, detection of obstacles is a major task of computer vision systems. The sensor measurements of the objective world are generated by the environmental measurement model. Due to reflections of light from multiple sources, noise affects the visual sensors. Hence, obstacle detection and its isolation in close proximity of the robot is given paramount importance [16]. In [17] a technique for the estimation depth data from a rectified image pair is presented. A convolution neural network (CNN) trained on small patches of input images was able to learn similarity between them. This result was used for the initialization of the cost function. The datasets from both KITTI as well as Middlebury were used for evaluation. During the training the ground truth disparity values of the image patches were made available to the network.

An algorithm based on random walk with restart for a dense stereo disparity map generation is proposed in [18]. Pixels were grouped into super pixels, which were then used for matching cost aggregation by combining image gradient matching and census transform. As compared to other stereo algorithms, such as belief propagation, graph cut or semi global matching, the algorithm estimates the best possible

disparity value for each pixel matching in the stage where updating the cost takes place [19].

A modified version of the Census Transform (CT) is presented in [20] which the authors claim to be comparatively more robust in situations pertaining to radiometric variations in real life outdoor road scenarios. The modification was made to the cost aggregation method in that it also considers the edge information present in the image pairs. They also observed that their proposed cost function generated better results as compared to the orthodox cost functions when they applied it on the KITTI stereo benchmark [21]. The authors in [22] propose an algorithm that uses the cues related to the scene data which include: identification of arrangements of surface normals and edges, detection of connected and coplanar section in images such as walls. A seed growing algorithms were also considered for determining the scene flow in a stereo setup [23]. The parameters used in the algorithm include a growing threshold, the temporal consistency enforcement parameter and an optical flow parameter. The correlations are based on the MNCC based statistic with the window size being 5x5 pixels.

In [24], a comparatively simple and fast algorithm was exhibited which is claimed to improve the number of valid disparity values through the use of local methods all the while reducing the number of pixels which have erroneous values. Through this technique, the initial disparity map was generated from a local stereo matching method which was repeatedly improved through depth interpolation method and image warping depending on the estimated depth. From the KITTI stereo data set, it had been observed that, on an average there was a rise in depth density after a single epoch, for a minor increase in computation. This technique also initially uses a small number of correspondence seeds. The proposed algorithm is capable of handling complicated issues, such as repetitive patterns and complex scenes, in a far better manner than that of the algorithms in the same category. Moreover, the proposed algorithm is also claimed to recover from erroneous initial seeds [25]. A novel approach for stereo matching problem which involves the use of a binary cost estimation and aggregation method was proposed [26]. The construction of the cost volume is based on bitwise operations on a set of binary strings. The results demonstrate that the algorithm is robust when compared to previous methods under these radiometric variations.

An approach for on-road distance estimation and vehicle detection, given disparity maps generated from a stereo matching algorithm is proposed in [27]. They have utilized available stereo matching method to estimate a dense disparity map. The obstacles space and free space are the two different spaces segmented by the dense disparity map. The KITTI datasets are used for the evaluation of the proposed algorithms.

In [28], vehicle detection in images from a classification system is presented. This system comprises of a machine learning classifier and an expert system preselecting relevant sections in an image by making use of the designed features. The authors validated the classification algorithms on real traffic videos and used it for vehicle detection. A real-time in-vehicle detection method that presents the idea that of vehicles

existing in rectangular sub-areas by a combination of multiple morphological vehicle features based on the classification of the detected features is presented in [29]. Authors in [30] presented that the observation of driving location is an essential and critical job of self-driving and Advanced Driver Assistance System (ADAS) technology. The proposed method is compared to basic methods that use KITTI stereo datasets. The observation from the results have revealed that the algorithms are robust to the measurement errors in the sequences of a long video. The authors used a stereo-vision based obstacle detection technique for intelligent ground vehicles and estimated the motion of obstacles in consecutive frames.

A methodology for vision based local navigation inspired by a model representing human navigation is proposed in [31], where the heading of the robot and potential field is derived by calculating its distance from the goal and the angular width of the obstacles. This field controls the steering of the robot and its angular acceleration. This method of directly controlling the steering is preferred for local navigation. A ceiling-mounted camera wheeled robot using image-based visual servoing algorithm was described in [32]. Visual servoing uses visual information to control the movement of the robot. Image-based visual servoing and position-based visual servoing are the main classifications in this method. The image error is used in image-based servoing method to control the robot. Simultaneous localization and mapping (SLAM) construct an environment map by parallel tracking the location of the robot. This method finds application in navigation, odometry and robotic mapping for augmented environments [33, 34]. Some of prominent methods like Extended Kalman Filter (EKF), particle filter, Graph-SLAM, and covariance intersection can be used to achieve this approach. EKF-SLAM uses feature extraction algorithms to extract features such as lines and corners to build the map of environment, which can be also used to identify the robot location [35]. The extracted features and the pose of the robot are positioned at a global coordinate system. A secondary map which contains the location of start and destination points of each segment is built using the extracted features.

The most basic issue with robot motion planning in a dynamic and complex environment is to find a path from the source position to the destination without colliding with obstacles in an optimal manner. A comparative compilation of various such approaches related to motion calculation in dynamic environments the simulation of few of the models is presented in [36 - 39]. As opposed to most of the existing monocular systems, our approach concentrates of a stereo camera setup to overcome the issues of shadow interference, illumination variation and object occlusion. A sparse set of features are selected for every frame, which are then projected into the 2D ground plane. A kernel-based approach for clustering is adopted for this research [40-44].

In order for the robot to reach its destination without collision, a recurrent fuzzy neural network (RFNN) approach is considered. RFNN combines the learning ability of the neural network and the inference capability of a fuzzy logic system. The side feedback loops in an RFNN helps improve the robustness and stability of the system. The weights of the network are trained with the help of an extended Kalman filter

[45-47]. By enforcing spatial dependencies among disparity estimates and scene characteristics, the authors were able to improve estimated disparity values. To attest the accuracy and viability of the approach, the results were evaluated through the Middlebury benchmark stereo datasets and the evaluation dataset version 3.0, where the presented approach was compared with the state of the art. It was observed that the proposed algorithm generated disparity values with higher accuracy when compared with non-learning and learning based algorithms [48-50].

III. PROPOSED ALGORITHMS AND RESULTS

The block diagram of the proposed methodology is presented in Fig. 1. As a first step, the left and right images are subjected to segmentation. Each segment obtained as a result, would contain pixels of similar intensities. Following this, similarity matching functions 'Data cost 1' and 'Data Cost 2' are used to match similar regions in the two images. The results generated through these similarity matching metrics are presented to the fuzzy disparity generator which provides the appropriate disparity values based on the results of the similarity metrics (results of 'Data cost 1' and 'Data cost 2'). This generates two disparity maps corresponding to the left and right images. Post this step, the proposed entropy-based refinement technique is applied to the generated disparity maps and a single refined disparity map is generated. Further, the obstacles and road segments are detected through u and v disparity methods.
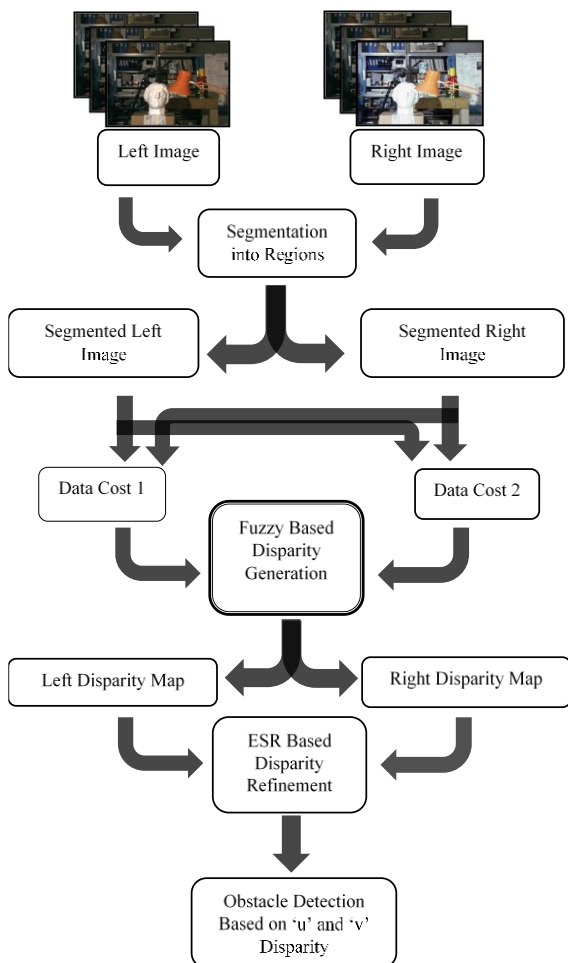

Fig. 1. Proposed methodology for disparity map-based obstacle detection

It has been observed from the literature review section that using multiple data costs provides a good solution to the problem at hand. Hence, we adopted a similar approach in our work and used the cost functions advocated in the literature, namely Normalized Cross Correlation (NCC) and Normalized Mutual Information (NMI), which are denoted by (1) to (4).

$$NCC(x_l, x_r) = \frac{\sum_i \left((x_l(i) - \bar{x_l}) \times (x_r(i) - \bar{x_r}) \right)}{\sqrt{\sum_i (x_l(i) - \bar{x_l})^2 \times \sum_i (x_r(i) - \bar{x_r})^2}}$$

(1)

$$NMI(X, Y) = \frac{H(X) + H(Y) - H(X, Y)}{\sqrt{H(X) \times H(Y)}}$$

(2)

where,

$$H(A) = -\sum_{i=0}^{n} p(i) \log p(i)$$

(3)

$$H(A, B) = -\sum_{i=0}^{n} \sum_{j=0}^{m} p(i, j) \log p(i, j)$$

(4)

Given a pair of stereo images, many traditional stereo matching methods use a single data cost metric to determine the disparity map such as absolute intensity difference and squared intensity differences, Sum of Absolute Differences (SAD), sum of squared differences and NCC. In real world scenarios, a small variation in radiometric conditions is observed between the two stereo images under consideration.

In such conditions, the performance of the traditional stereo matching algorithms tends to deteriorate. Over the years, it has been observed from literature [51-55] that using multiple data costs provides a better solution to the problem of computing accurate disparities. Inspired by these findings, the authors propose a stereo matching metric comprising of multiple data costs to overcome the abovementioned problem. The research work in the field of stereo matching in [56] compares NMI and NCC metrics, and claims that MI is suitable for this field. Both MI and NCC are capable of handling radiometric differences, but their performances deteriorates at a cost of window size. SAD is another prominent metric that can be used in stereo matching, but its results in the radiometric variations shows large error. Even though large error is generated in SAD, it is capable to generate better disparity maps while used on the gradient of images. In this work, NMI, NCC, and SAD gradient of images, as given by equation (5) to (7), and a novel metric comprising of linear combination of mentioned three data costs is considered.

$$I_g = \left[\left(\frac{\partial I}{\partial x} \right)^2 + \left(\frac{\partial I}{\partial y} \right)^2 \right]^{\frac{1}{2}}$$

(5)

$$GSAD(x_{lg}, x_{rg}) = \sum_i \left| x_{lg}(i) - x_{rg}(i) \right| \qquad (6)$$

$$IGSAD(x_{lg}, x_{rg}) = 1 - \frac{GSAD(x_{lg}, x_{rg})}{\max(GSAD)} \qquad (7)$$

where x_l and x_r are the left image and right image sections, respectively.

In Cost aggregation, it has been observed that it is not a good choice to use a fixed window size, as they include regions with depth discontinuities. The presence of uneven depths within the window generates erroneous disparities. An alternate method would be to use a window whose size is not fixed. Such a window should be able to change its shape depending on the properties of the pixels under observation. One such method is based on segmentation. Here, the reference image is initially segmented and matched with the target image.

In this work, the authors have segmented the input images into patches [57] consisting of pixels belonging to the same object (region). To overcome the problem of windows having pixels of different depths, these segmented regions are considered as windows and stereo matching is performed using these newly obtained variable sized windows. Fig. 2 shows the segmentation result for one of the images in the dataset. As can be observed, pixels in each segment belong to the same region and would hence have similar depth. The disparity location that has the optimal cost is found after calculating the cost values for each disparity. WTA scheme is used to select the disparity. From all possible disparity locations, the one with the best cost is selected as the true disparity value. To overcome the possibility of ambiguity, an "intelligence" based disparity selection method is used wherein, instead of directly applying WTA on the results of the datacosts, they are fed into the fuzzy logic module to generate a value which is obtained through the combination of the two mentioned datacosts and apply the WTA on the result of this fuzzy logic module. Fig. 3 represents the work conducted in this paper using a Fuzzy Interference System (FIS). A rule base "if-then" approach is formulated to decide the output [58].
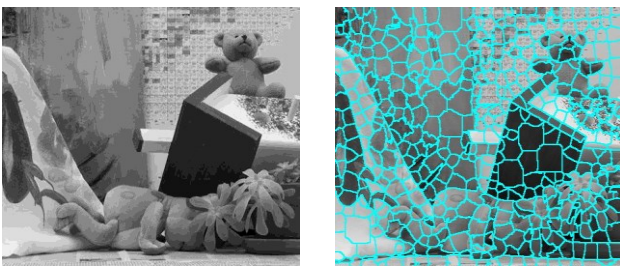

Fig. 2. Original image (left) and result of segmenting image into patches (right)

The two inputs to the FIS are the Cost function 1 and 2, defined by equation (8) and (9), respectively. The data cost metrics use for this research are NCC, NMI, and SAD of gradients as described in [59].

For deciding the best combination of the datacosts, the positions of the considered three data costs were interchanged in equation (1) and (2), the results were observed. Equations (10), (11), and (12), defines the membership functions for Cost function 1, Cost function 2, and the Output, respectively. The rules of the FIS are mentioned in Table I. The membership functions of input 1 and input 2 are shown in Fig. 4 and Fig. 5 respectively while the membership function of output is shown in Fig. 6.
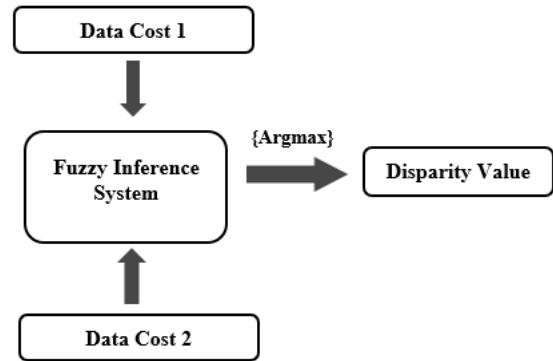

Fig. 3. Selection of disparities through a fuzzy based approach

$$Data\ Cost\ 1 = NCC \qquad (8)$$

$$Data\ Cost\ 2 = \alpha \times NMI + (1 - \alpha) \times SAD\ of\ Image\ Gradient \qquad (9)$$

$$Data\ Cost\ 1 : \{Poor, Very\ bad, Bad, Average, Good, Very\ good\} \qquad (10)$$

$$Data\ Cost\ 2 : \{Poor, Very\ Bad, Bad, Average, Good, Very\ Good\} \qquad (11)$$

$$Output : \{Worst, Poor, Very\ Bad, Bad, Below\ Average, Acceptable, Average, Above\ Average, Good, Very\ Good, Extremely\ Good, Great, Superb, Excellent, Best\} \qquad (12)$$

Both left and right disparity maps are subjected to the fuzzy-based approach as shown in Fig. 3. These resulting left and right disparity maps are used for left-right consistency checking to obtain the final refined disparity map. For all considered disparities, the data cost for every segment in the reference image are calculated using the cost functions and are fed into the FIS. This process is iterated for both the left and right images to obtain the disparity maps corresponding to the left and right images, respectively. Results are further improved by subjecting these images to left-right consistency check [60]. A point is said to be consistent, if the condition in equation (13) is met.

TABLE I. RULES OF FUZZY INFERENCE SYSTEM

Rule number	Data cost 1	Data cost 2	Output
1	Poor	Poor	Worst
2	Poor	Very bad	Worst
3	Poor	Bad	Very bad
4	Poor	Average	Bad
5	Poor	Good	Below average
6	Poor	Very good	Below average
7	Very bad	Poor	Poor
8	Very bad	Very bad	Very bad
9	Very bad	Bad	Very bad
10	Very bad	Average	Bad
11	Very bad	Good	Acceptable
12	Very bad	Very good	Average
13	Bad	Poor	Poor
14	Bad	Very bad	Very bad
15	Bad	Bad	Bad
16	Bad	Average	Below average
17	Bad	Good	Average
18	Bad	Very good	Above average
19	Average	Poor	Very bad
20	Average	Very bad	Very bad
21	Average	Bad	Below average
22	Average	Average	Average
23	Average	Good	Good
24	Average	Very good	Very good
25	Good	Poor	Bad
26	Good	Very bad	Below average
26	Good	Bad	Acceptable
28	Good	Average	Above average
29	Good	Good	Extremely good
30	Good	Very good	Superb
31	Very good	Poor	Below average
32	Very good	Very bad	Acceptable
33	Very good	Bad	Acceptable
34	Very good	Average	Great
35	Very good	Good	Excellent
36	Very good	Very good	Best

$$d_l(x, y) = d_r(x, y - d_l(x, y)) \qquad (13)$$

where, d_l and d_r are the left and right disparity maps.

The methodology for generation of disparity maps, given stereo image pairs, have been adopted from [14]. Another problem observed was the selection of the right disparity value for a pixel. Most of the methods in the literature concentrate mainly on the selection and development of the cost function.
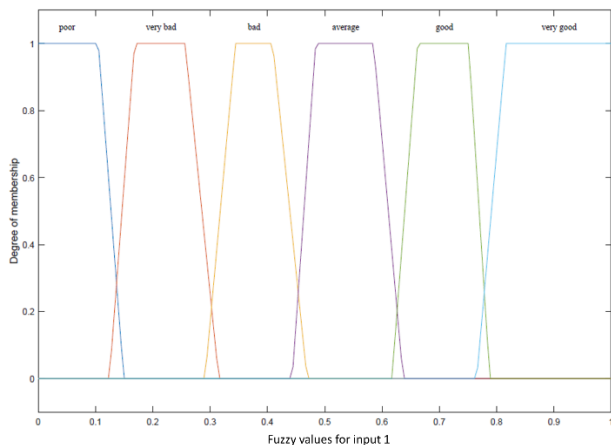
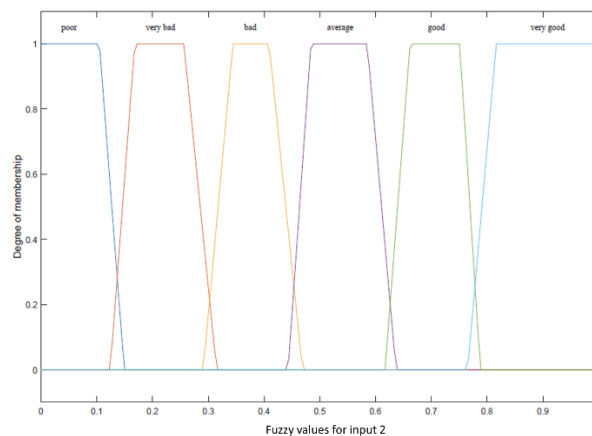Fig. 4. Membership functions of input 1
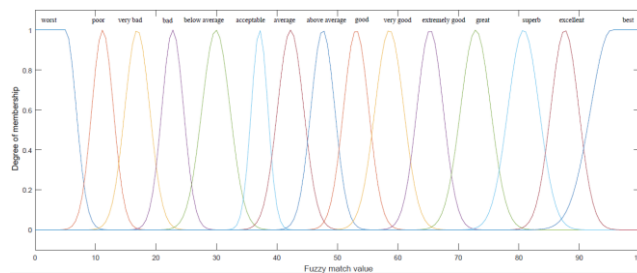
Fig. 5. Membership functions of input 2

Fig. 6. Membership functions of output

The selection of the disparity value is almost always done through a WTA-based method [61]. While calculating the cost, there is chance of multiple maximums/minimums to be produced by a cost function, which leads to ambiguity. A WTA approach in the presence of such ambiguities, might lead to erroneous results. An "intelligence" based disparity selection approach is proposed to tackle this problem.

To reduce the computation time, it is assumed that pixels which are close together and similar belong to the same object. These patches of similar and close pixels are considered as a patch of super pixels. Hence these similar pixels would have the same disparity.

A. Entropy based Selective Refinement of Disparity Maps

To reduce the computation time, it is assumed that pixels which are close together and similar belong to the same object. These patches of similar and close pixels are considered as a patch of super pixels. Hence these similar pixels would have the same disparity. This eliminates the need to calculate the disparity of every single pixel in the stereo pair. Instead, calculate the disparity of the whole patch altogether and assign the same disparity to all the pixels that belong to it. Though this drastically reduces the computation time, values that have been calculated erroneously would be seen as patches in the disparity map as well. The tradition methods of refinement [62] can be used only if the disparity map was generated using pixel wise computation as the errors would also be present in the form of uneven pixels. But in the proposed methodology, whole patches of erroneous values can be present wherein the traditional methods of refining single pixel values would not be suitable. For this purpose, a new entropy based selective refinement methodology (ESR) for refining the generated disparity map has been proposed.

The block wise representation of the proposed methodology for the refinement of the disparity maps is shown in Fig. 7.

An entropy-based disparity refinement technique is proposed to improve the quality of the generated disparity maps. The entropy of the block is computed through the Shannon's entropy [63], which is represented in equation (14).

$$E = -\sum_{i=0}^{n} p(i) \log p(i) \qquad (14)$$

The disparity maps are obtained through the methodology mentioned for generation of disparity maps. A window around each pixel is considered and the entropy of this block is calculated. Similar procedure is carried out for the corresponding position in the other disparity map. If the variation in the entropies between the two segments being considered exceeds a set threshold, then the center pixel of the block of left disparity map is replaced.
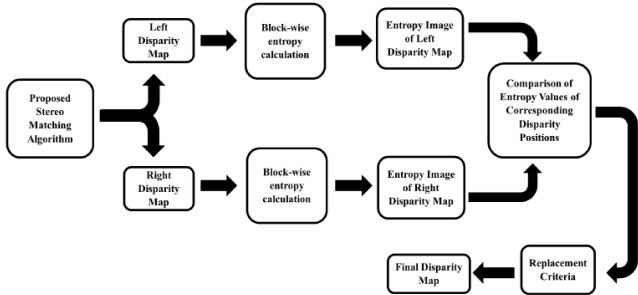
Fig. 7. Diagrammatic representation of ESR methodology for disparity map refinement

For the replacement value, find out which block has lower entropy (i.e., left block or right block). The mean of the disparity block with lower entropy is selected as the replacement. The replacement criteria can be represented as given in (15).

$$d = \begin{cases} \overline{d_l}, & if \ |\overline{E_l} - \overline{E_r}| > \tau_{ent} \ and \ \overline{E_l} < \overline{E_r} \\ \overline{d_r}, & if \ |\overline{E_l} - \overline{E_r}| > \tau_{ent} \ and \ \overline{E_l} > \overline{E_r} \\ d, & otherwise \end{cases}$$

$$(15)$$

where $\overline{d_l}$ and $\overline{E_l}$ are the mean and the mean entropy of the disparities of the left disparity map block while $\overline{d_r}$ and $\overline{E_r}$ are the mean and the mean entropy of the disparities of the right disparity map block. τ_{ent} is the entropy difference threshold value which was selected through experiments. The stereo matching algorithm would generate two disparity maps (one with reference as left image and another with reference as right image as reference) as shown in Fig. 8.
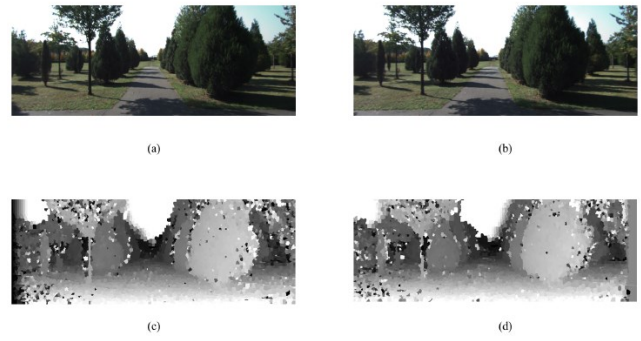

Fig. 8. Left and right stereo images along with the respective disparity maps

A comparison of the proposed methodology on KITTI dataset with that of other similar methods are presented in Table II. Fig. 9 and Fig. 10 present the entropy images of the left disparity map and the right disparity map obtained from the proposed disparity generation method. Fig. 11 depicts the effect of the size of the window and threshold value on the average error and the percentage of bad pixels of disparity maps.

TABLE II. COMPARISON OF PROPOSED METHODOLOGY USING KITTI DATASET

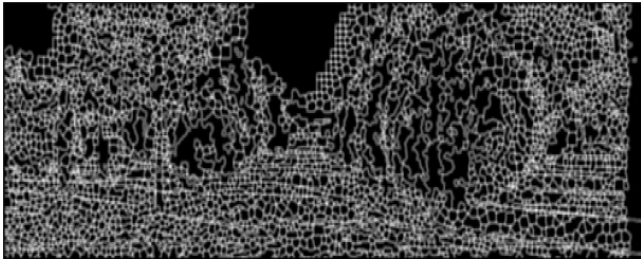| Methods	Average error (per pixel)	Percentage of bad pixel (>2)	Methodology followed
Competing method [64]	0.7	3.9	Convolution neural network-based method
Competing method [65]	1.2	7.91	Iterative optimization
Proposed method	1.33522	14.327	Non-iterative (Fuzzy logic-based disparity generation)
Competing method [66]	1.6	10.83	Non-iterative (Slanted support window with AD and CT based matching)
Competing method [67]	2.1	13.13	Non-iterative (Modified variant of CT)
Competing method [68]	2.0	15.26	Non-iterative (Adaptive window-based grouping and matching)
Competing method [69]	2.3	17.22	Non-iterative (top-down and bottom-up visual cues in terms of oriented planar primitives)
Competing method [70]	1.9	17.41	Non-iterative (Seed based disparity growing)
Competing method [71]	3.3	14.31	Iterative (Depth map interpolated and warped iteratively)
Competing method [72]	2.1	19.03	Non-iterative (Seed based disparity growing)
Competing method [73]	2.2	19.51	Non-Iterative (Binary cost computation)
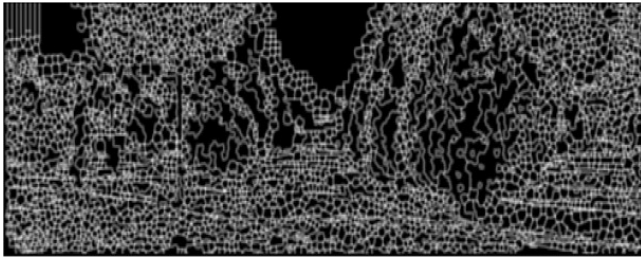
Fig. 9. Entropy image of left disparity map
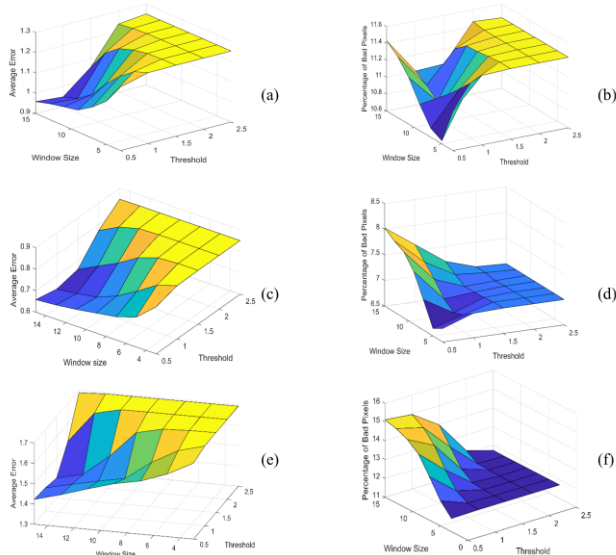

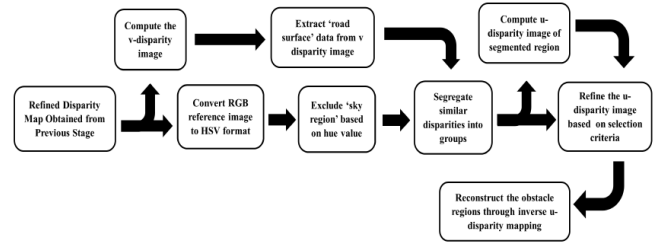Fig. 10. Entropy image of right disparity map

Fig. 12. Block diagram of methodology for obstacle detection

A sample left and right disparity maps (Fig. 13(a) and Fig. 13(b)) along with the refined disparity map obtained from the previous stage (Fig. 13(c)) is presented in Fig. 13.

For the selection of obstacle regions, first detect the locations that belong to the road and sky regions. For this purpose, calculate the histogram of every row in the image (also known as v-disparity map) as mentioned in (16) and (17).
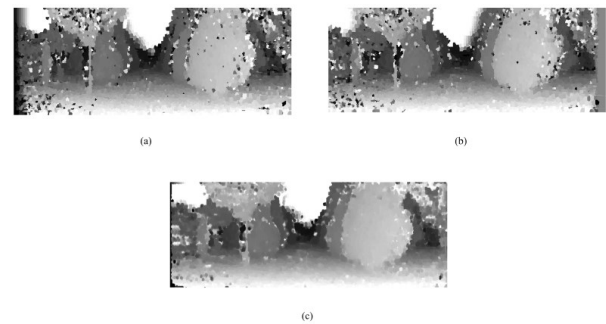

Fig. 13. Refined disparity maps

$$\partial_{i,j} = \begin{cases} 1, & disp_{i.j} = d \\ 0, & otherwise \end{cases} \qquad (16)$$

$$v_{d,j} = \sum_{j=1}^{rows} \partial_{i,j} \qquad (17)$$

It is observed in the row-histogram representation of the disparity map that the road region can be represented as a straight slanted line segment. Finding the equation of this segment yields the pixels regions that represent the road surface. This line segment is highlighted in Fig. 14. Once the equation for the highlighted line is found, extract the road surface regions from the disparity map as shown in Fig. 15. Fig. 15a represents the modified version of the disparity map without the road surface and an image which has only the road surface shown in Fig. 15 (b) which represent the points on the line segment. The results of applying morphological closing operation on the above images to remove tiny specs and smooth out the regions are shown in Fig. 16.
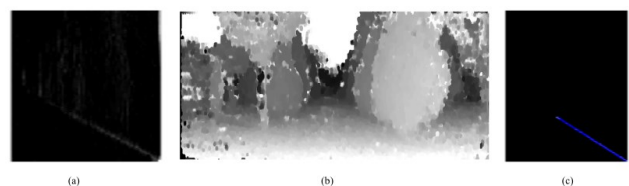

Fig. 14. Converting the disparity map in to v disparity image (a) and detecting the line that would represent the road surface (c) from the refined disparity map (b)

Fig. 11. Effect of window size and threshold value on average error and percentage of bad pixels

B. Segmentation of Obstacle Regions and Path Planning

In [25] the authors proposed a strategy to detect obstacles from planetary images in unfavorable imaging conditions. They had used a combination of 2D grayscale information along with 3D point statistical information from an image pair to extract a region-based map of obstacles. They used mean shift segmentation and seed points extracted from reprojection of the 3D points in the left image [74]. Proposes a technique for stereo vision-based obstacle detection that searches the depth map along the vertical direction for pixels possessing the same disparity. They also used a monocular camera model along with stereo vision to detect all the possible obstacles present. The authors claim that their method performs obstacle detection reliably under various traffic conditions.

The methodology used in this research work for segmentation of the obstacle regions in the obtained disparity maps is presented in Fig. 12.
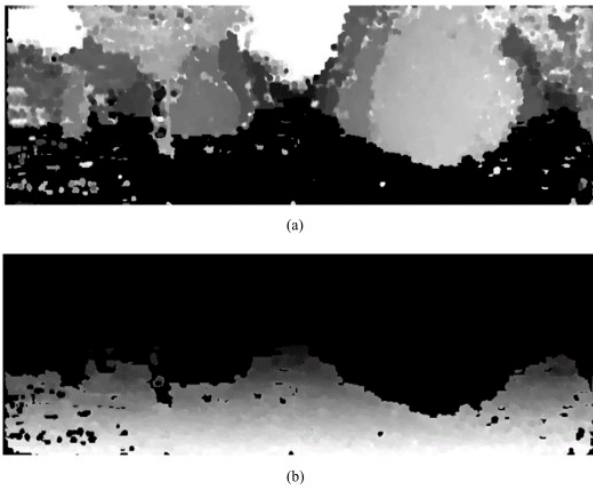
Fig. 15. Disparity map without road segment (a) and extracted road surface (b)
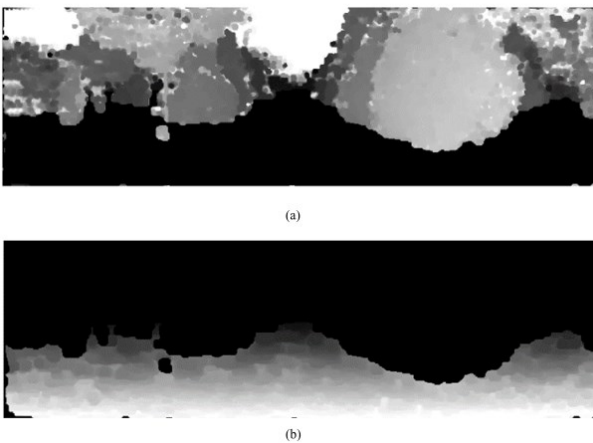

Fig. 16. Refined versions of the road (b) and no-road sections (a) of the disparity maps

In the next step, remove the sky regions in the modified disparity maps. The original RGB image was converted to HSV and the appropriate 'hue' value was used to eliminate the sky regions as presented in Fig. 17.
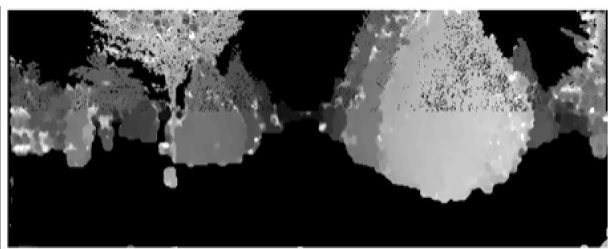

Fig. 17. Sky regions eliminated from the disparity map

The equations related to conversion from RGB to HSV [27] are as mentioned in (18). Let R, G, and B indicate red, green and blue bands of the images. We would first need to normalize these values as described in (18).

$$R' = \frac{R}{255} \qquad (18)$$

$$G' = \frac{G}{255}$$

$$B' = \frac{B}{255}$$

$$C_{max} = \max(R', G', B')$$
$$C_{min} = \min(R', G', B')$$
$$\Delta = C_{max} - C_{min}$$

The hue, H, can be calculated as described by (19).

$$d = \begin{cases} 0°, & \Delta = 0 \\ 60° \times \left(\frac{G' - B'}{\Delta} \, mod6\right), & C_{max} = R' \\ 60° \times \left(\frac{B' - R'}{\Delta} + 2\right), & C_{max} = G' \\ 60° \times \left(\frac{R' - G'}{\Delta} + 4\right), & C_{max} = B' \end{cases} \qquad (19)$$

The saturation, S, can be calculated as described by (20).

$$d = \begin{cases} 0, & C_{max} = 0 \\ \frac{\Delta}{C_{max}} \times 100, & otherwise \end{cases} \qquad (20)$$

The value V, is given by (21),

$$V = C_{max} \qquad (21)$$

As previously indicated, the correspondence operation on a stereo image pair results an image called as disparity map. In this image, the value of the disparity of a pixel is indicated by the brightness of each pixel. The depth of the object from the camera is inversely proportional to the disparity value, corresponding to the pixel. Since the obstacles which are within the close proximity are excluded from the classification process, it is difficult to keep an exact limit on the disparity for detecting obstacles. The disparities of the obstacles vary with their distances. Therefore, the grouping of obstacles based on their disparity values into different clusters is of vital importance. The categorization of image pixels into different clusters by dividing the data into set of groups is known as Image segmentation [75]. k-means clustering is one such popular unsupervised learning method to achieve this process, where the data collected are partitioned into k groups of data, i.e. the data is segregated into k-disjoint clusters. This objective is achieved in two separate phases by k-means algorithm. In the first stage, the centroids of each of the clusters are calculated. This is followed by assigning each point to the cluster it is closest to [76]. The process is repeated for all the data points with all the k centroid points. Euclidean distance is the most popular approach, by considering distance metric, in the calculation of the distance of the data point to the nearest centroid. This is represented by (22).

$$d = \|p(x, y) - c_k\| \qquad (22)$$

where p(x, y) is the pixel that needs to be grouped into one of the clusters, c_k is the center of the cluster and d is the calculated distance.

After the points are initially assigned to the cluster, the second phase operation of recalculating the new centroid for each cluster is performed. The new centroid is calculated as described by (23).

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} p(x, y) \qquad (23)$$

where k denotes the clusters. These steps are iterated until the tolerance or error value is reached.

The Euclidean distance between each point and every centre is recalculated depending on the new centroid, which will further adjust the data point's assignment into the clusters and refine the clusters iteratively. The parameters that define each of the clusters are its data points and the centroid. A new position is assigned as the centroid that has the least sum of the distance to every point in the cluster. Thus, the k-means algorithm, in an iterative manner minimizes the sum of distances of each point with its centroid over all clusters [77]. After this process is completed, each cluster will be consisting of similar disparity value pixels.

After the completion of the clustering stage, every cluster would contain pixels with similar disparity values. This would result in the disparity map getting segmented into separate regions. The next step is to group similar disparities together depending on the intensity values of the refined and modified disparity map. For this purpose, the k- means segmentation algorithm has been used and studied the accuracy of detection for varying sizes of clusters. A sample of one of the segments is presented in Fig. 18 which represents the group of pixels which are closest to the camera.

Fig. 18. An example of a segmented section of the disparity map

It is evident that the segmented regions contain noise induced during the stereo matching and the segmentation process. These results can be further refined to reduce the noisy speckles present. To rectify this, u- disparity image is used which represents every entity in the disparity map as a horizontal line. Similar to the v-disparity map discussed previously, generate the u-disparity image by calculating the histogram of all the pixels present in each column, as given in (24) and (25), and represent it as an image as shown in Fig. 19.
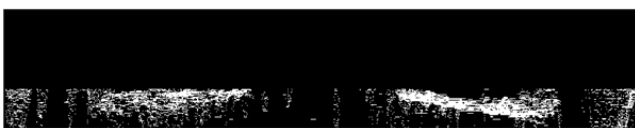

Fig. 19. The u-disparity image

$$\partial_{i,j} = \begin{cases} 1, & disp_{i,j} = d \\ 0, & otherwise \end{cases} \qquad (24)$$

$$u_{i,d} = \sum_{i=1}^{columns} \partial_{i,j} \qquad (25)$$

The u disparity image represents each object in the disparity map as a horizontal line. Hence by applying a restriction criterion on the length of the lines present in the u disparity image, a refined version of the input is obtained as shown in Fig. 20.

Fig. 20. The refined u disparity image

These refined u-disparity segments are used to reconstruct the refined version of the segmented disparity map by mapping it back to the original image while using only the points present in the refined u-disparity map. The result of this process is shown in Fig. 21.

Fig. 21. Obstacle disparities extracted from the u disparity image

In this manner, perform these operations on all the segments obtained from the k-means segmentation result to detect the obstacles present in the disparity map [78-79]. Based on this highlight and segment the obstacles in the images (for purpose of ease of identification, the closest obstacles are highlighted in red bounding boxes, slightly away are in green bounding boxes and farthest away in blue bounding boxes). The final set of obstacles detected in this manner is shown in Fig. 22.
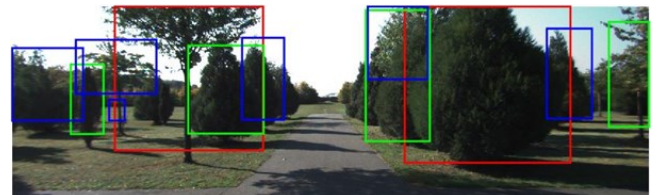

Fig. 22. Original image with obstacles indicated from the image based on proposed Methodology

Experiments on the threshold value are conducted, which the cluster center is set which in-turn decides how many pixels can be clustered in a group, and have set the value appropriately as shown in Fig. 23.
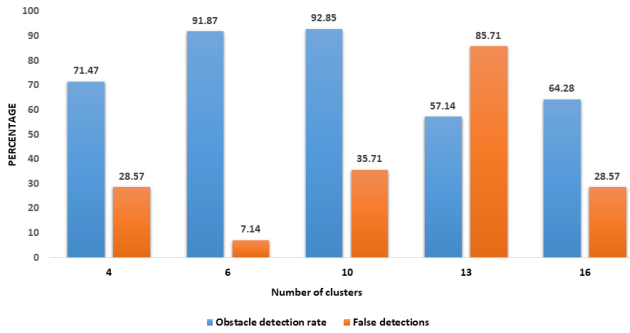
Fig. 23. Effect of number of clusters on obstacles detected

It was observed that for number of clusters values of 6 and 10, the highest obstacle detection rate is obtained but for the latter, the rate of false detections was also higher. For this reason, set the number of clusters as 6.

The proposed methodology was used on the KITTI dataset and its comparison with similar methods is represented in Table III and IV.

With the obstacles been clearly detected, verify the above proposed methodologies in real life images.

TABLE III. COMPARISON OF OBSTACLE DETECTION RATE OF PROPOSED METHODOLOGY (TRAFFIC CONDITIONS)

	KITTI 2015		KITTI 2012		TOTAL		COMPARISON	
	Present	Detected	Present	Detected	Present	Detected	Detection rate by proposed method	Detection rate by completing method [53]
Traffic Areas	644	612	397	356	1041	968	**92.9875**	89.26
Non-Traffic areas	248	215	201	186	449	401	**89.3095**	89.40

TABLE IV. COMPARISON OF OBSTACLE DETECTION RATE OF PROPOSED METHODOLOGY (URBAN CONDITIONS)

KITTI 2012		KITTI 2015		TOTAL		COMPARISON			
Present	Detected	Present	Detected	Present	Detected	Detection rate by proposed method	Detection rate by competing method [54]	Detection rate by competing method [55]	Detection rate by competing method [56]
598	542	892	827	1490	1369	**91.8791**	83.12	92.63	83.15

A Kalman filter [80], [81] based path planning algorithm was implemented to verify the use of the proposed algorithms and its ability to detect obstacles in real life environments. The flow chart of the path planning algorithm is presented in the Fig. 24.
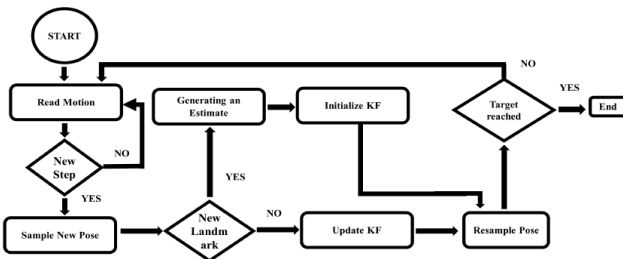

Fig. 24. Block diagram of path planning algorithm

The state estimation at time $t + \Delta t$ is acquired from the data from the previous instance of time t can be calculated as shown in (26) [82].

$$\hat{x}_{t+\Delta t|t} = F_{\Delta t}\hat{x}'_t + B_t\vec{u}_t$$
$$P_{t+\Delta t|t} = F_{\Delta t}P'_t F_t^T + Q_t \tag{26}$$

The variable $\hat{x}_t$ describes the estimated state at time t. It takes the form as mentioned in (27).

$$\hat{x}_t = \begin{bmatrix} \vec{p}_t \\ \vec{v}_t \end{bmatrix} \tag{27}$$

where $\vec{p}_t$ indicates the vector describing the position and $\vec{v}_t$ is the vector related to velocity. Here $\hat{x}_{t+\Delta t}$ is taken to be the representation of the previous state of the object at time t. The

matrix F_t is called the prediction matrix [83]. It uses the older position and velocity to estimate the next possible position and velocity as shown in (28).

$$F_{\Delta t} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{28}$$

The product of $F_{\Delta t}$ and $\hat{x}'_{t-\Delta t}$ is as described in (29).

$$F_{\Delta t}\hat{x}'_t = \begin{bmatrix} \vec{p}_t + \vec{v}_t\Delta t \\ \vec{v}_t \end{bmatrix} \tag{29}$$

In other words, it is assumed in the prediction matrix that there is no change in the velocity in the duration from t to $t + \Delta t$; this is related to the system dynamics linear assumption [84]. The varying magnitude and direction of motion is taken into consideration through the control vector $\vec{u}_t$ and the control matrix B_t. In the presented research work, the variation in velocity is treated as the control input. This control input is indicated as a random process which is a contributor to the noise in velocity [85]. For this reason, the control inputs are not made use of in the prediction, hence $\vec{u}_t$ and B_t are simply not considered and treated as 0 in the equations related to prediction. The matrix P_t is the covariance matrix and is presented in (30).

$$P_t = \begin{bmatrix} \sigma_{p,p} & 0 \\ 0 & \sigma_{v,v} \end{bmatrix} \tag{30}$$

The second term Q_t, in the equation related to the updation of the covariance between observed values, is termed as the covariance matrix that is responsible for noise in the process [86], which in this research work keeps track of the variations that would take place during the time interval Δt. The amount

of change to be incorporated from the prediction at time t, based on the observation at time t is decided by the Kalman gain function, which is described in (31).

$$K' = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1} \qquad (31)$$

The matrix H_t is accountable for the conversion to the measurement space from state space. H_t is taken to be the identity matrix since it can be assumed that the measurement space and the state space can be considered to be identical. The noise in the sensor reading can be represented as the matrix R_t. This is a diagonal matrix and consists of the variances of normal distributions for the respective dimensions [87]. The contribution of the matrix R_t is in modelling the uncertainty in the readings obtained by the sensor. The expected sensor readings for the covariance matrix P_t are indicated by the value of $H_t P_t H_t^T$.

A finer approximation of the covariance of the state and the current state can be obtained through the gain function used along with the two mentioned variables as described in (32).

$$\hat{x}'_t = \hat{x}_{t|t-\Delta t} + K'(\vec{z}_t - H_t \hat{x}_{t|t-\Delta t})$$
$$P'_t = P_{t|t-\Delta t} - K' H_t P_{t|t-\Delta t} \qquad (32)$$

The value represented by the vector $\vec{z}_t$ is the reading of the sensor at the time instance t. The vector $\hat{x}'_t$ represents an updated version of the estimate of $\hat{x}_t$ when the sensor readings are considered [88]. A similar concept can be applied to P'_t. The value indicated by $H_t \hat{x}_t$ is the representation of the expected sensor readings provided the estimated state $\hat{x}_{t|t-\Delta t}$ is given. Subtraction of this component from the actual sensor readings $\vec{z}_t$ generates the difference between the observed and the predicted. After this step, the gain function K' is incorporated in to the obtained difference, also known as the error, and this error is appended to $\hat{x}_{t|t-\Delta t}$ to generate $\hat{x}'_t$, which is the revised version of the estimate of $\hat{x}_{t|t-\Delta t}$. Lastly, when the KF is used again, the updated values of the estimates $\hat{x}'_t$ and P'_t as opposed to $\hat{x}_{t|t-\Delta t}$ and $P_{t|t-\Delta t}$ are taken as the values for the initial state equations [89]-[91].

At every frame capture iteration, the system detects the motion, i.e., whether any movement has been made by the robot [92]. This is estimated by generating the disparity map for each stereo image pair and carrying out checks to determine if there are any variations in the results. If it is concluded that there has been any form of motion, the latest pose of the robot is calculated, i.e., the co-ordinates after the detected motion and check for any new points of interest (landmarks) [93-95]. This is accomplished by making use of the calculated disparity map and identifying new points of interest. If any variations in obstacles are detected, the distance of these points are used to initialize and update the Kalman filter. The pose of the robot is recalculate based on this and the process is repeated. The simulation of the path planning algorithm is shown in Fig. 25. The sequence of actions performed by the robot using the mentioned methodologies is presented in Fig. 26 and the mobile robot used is shown in Fig. 27.
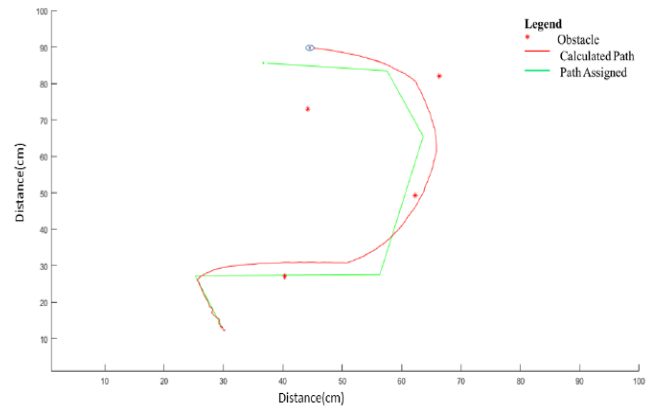

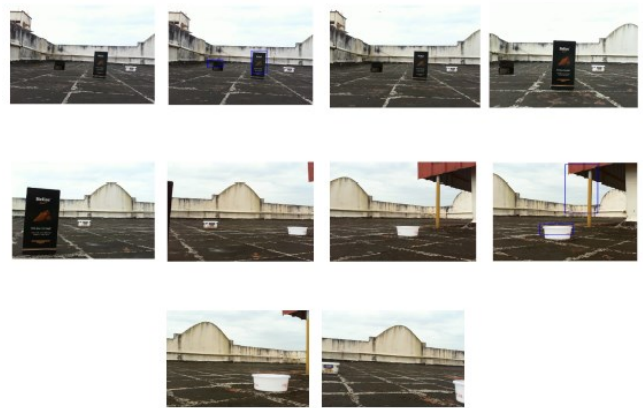Fig. 25. Simulation of path planning algorithm

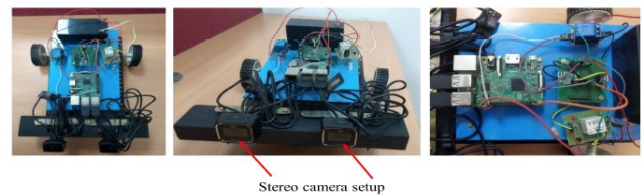Fig. 26. Sequence of actions taken by the robot to avoid obstacles

Fig. 27. Mobile robot used for the experiments

IV. CONCLUSION

A novel entropy-based technique is proposed to refine the generated disparity map. Based on the information obtained from these disparity maps, obstacles present in the images are detected and hence avoided. To detect these obstacles, the regions that are not obstacles need to be ignored. This is done with the help of v-disparity image information, which can provide information of the road regions. The regions belonging to the sky can be segmented using HSV information. Once this is carried out, what remains is the points that belong to obstacles. Among these points, the sections that belong to the closest obstacles are highlighted through k means segmentation. The segmented disparity map is further subjected to u-disparity information-based refinement that considers areas of the connected pixels to filter out any noise that was obtained from the k means segmentation. The proposed algorithm generated an average error of 1.33 and had an obstacle detection rate of 91.87% in the KITTI dataset. The proposed methodology was compared with similar methodologies and was found to be in par with them and in some of the cases provided better results.

Furthermore, a Kalman filter based path planning algorithm was used in conjunction with the mentioned algorithms. When implemented in real life scenario, the combination of proposed algorithms and the path planning algorithm was successfully able to guide a mobile robot from source location to a destination location while detecting and avoiding the obstacles present in its path.

Although the proposed methods have generated results that achieve the objectives undertaken by this research work, they have also put forward numerous aspects for further investigation. One such are is to study the effect of multiple light sources. Another aspect would be into the development of cost functions that could be used to generate dense disparity maps from multi modal images for stereo correspondence. This can be useful when dealing in military situations where images from multiple sources (thermal, night vision, etc.) could be available. Work can also be carried out towards the reduction of execution time through the use of graphical processing units or other dedicated hardware for disparity map generation.

REFERENCES

[1] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust Stereo Matching Using Adaptive Normalized Cross-Correlation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 807-822, April 2011, doi: 10.1109/TPAMI.2010.136.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, pp. 7-42, 2002.

[3] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[4] A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann, "Local stereo matching using geodesic support weights," *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2093-2096, 2009, doi: 10.1109/ICIP.2009.5414478.

[5] K-. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650-656, April 2006, doi: 10.1109/TPAMI.2006.70.

[6] M. Poggi *et al*., "On the Confidence of Stereo Matching in a Deep-Learning Era: A Quantitative Evaluation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5293-5313, 1 Sept. 2022, doi: 10.1109/TPAMI.2021.3069706.

[7] Q. Yang, "A non-local cost aggregation method for stereo matching," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1402-1409, 2012, doi: 10.1109/CVPR.2012.6247827.

[8] G. Egnal, "Mutual Information as a Stereo Correspondence Measure," *Univ. Pennsylvania, USA*, Jan. 2000.

[9] H. Hirschmuller and D. Scharstein, "Evaluation of Stereo Matching Costs on Images with Radiometric Differences," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1582-1599, Sept. 2009, doi: 10.1109/TPAMI.2008.221.

[10] G. Yang, H. Zhao, J. Shi, Z. Deng, and J. Jia, "SegStereo: Exploiting Semantic Information for Disparity Estimation," in *Proc. European Conf. Comput. Vis. (ECCV)*, pp. 636-651, 2018.

[11] X. Cheng, P. Wang, and R. Yang, "Learning Depth with Convolutional Spatial Propagation Network," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2361-2379, 1 Oct. 2020, doi: 10.1109/TPAMI.2019.2947374.

[12] X. Song, X. Zhao, L. Fang, and H. Hu, "EdgeStereo: An Effective Multi-Task Learning Network for Stereo Matching and Edge Detection," *Int. J. Comput. Vis.*, vol. 128, no. 4, pp. 910-930, Apr. 2020.

[13] A. A. Shetty, R. Shetty, N. T. Hegde, A. C. Vaz, and C. R. Srinivasan, "A Study on the Effect of Radiometric Variations on A Fuzzy Stereo Matching Algorithm: A Statistical Analysis," *Engineered Science*, vol. 16, pp. 269-280, Dec. 2021.

[14] A. A. Shetty, V. I. George, C. G. Nayak, and R. Shetty, "Performance analysis of a fuzzy disparity selector for stereo matching of image segments under radiometric variations," *Turk. J. Elec. Eng. & Comp. Sci.*, vol. 28, no. 4, pp. 1965 – 1983, Mar. 2020.

[15] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52-57, Mar. 2002.

[16] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2287-2318, Jan. 2016.

[17] S. Lee, J. H. Lee, J. Lim, and I. H. Suh, "Robust stereo matching using adaptive random walk with restart algorithm," *Image and Vision Computing*, vol. 37, pp. 1-11, May 2015.

[18] X. Li, J. Liu, G. Chen, and H. Fu, "Efficient methods using slanted support windows for slanted surfaces," *IET Computer Vision*, vol. 10, no. 5, pp. 384-391, Aug. 2016.

[19] O. Zeglazi, M. Rziza, A. Amine, and C. Demonceaux, "Accurate dense stereo matching for road scenes," in *IEEE International Conference on Image Processing (ICIP)*, pp. 720-724, 2017.

[20] J. Kostkova and R. Sara, "Stratified Dense Matching for Stereopsis in Complex Scenes," in *British Machine Vision Conference*, pp. 1-10, 2003.

[21] S. Hadfield, K. Lebeda, and R. Bowden, "Stereo reconstruction using top-down cues," *Computer Vision and Image Understanding*, vol. 157, pp. 206-222, Apr. 2017.

[22] J. Čech, J. Sanchez-Riera, and R. Horaud, "Scene flow estimation by growing correspondence seeds," *CVPR 2011*, pp. 3129-3136, 2011, doi: 10.1109/CVPR.2011.5995442.

[23] A. Murarka and N. Einecke, "A Meta-Technique for Increasing Density of Local Stereo Methods through Iterative Interpolation and Warping," *2014 Canadian Conference on Computer and Robot Vision*, pp. 254-261, 2014, doi: 10.1109/CRV.2014.59.

[24] J. Cech and R. Sara, "Efficient Sampling of Disparity Space for Fast And Accurate Matching," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007, doi: 10.1109/CVPR.2007.383355.

[25] K. Zhang, J. Li, Y. Li, W. Hu, L. Sun, and S. Yang, "Binary stereo matching," *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 356-359, 2012.

[26] D. Dekkiche, B. Vincke, and A. M´erigot, "Vehicles Detection in Stereo Vision Based on Disparity Map Segmentation and Objects Classification," in *International Symposium on Visual Computing*, pp. 762-773, 2015.

[27] P. Bergmiller, M. Botsch, J. Speth, and U. Hofmann, "Vehicle rear detection in images with Generalized Radial-Basis-Function classifiers," *2008 IEEE Intelligent Vehicles Symposium*, pp. 226-233, 2008, doi: 10.1109/IVS.2008.4621273.

[28] D. Alonso, L. Salgado, and M. Nieto, "Robust Vehicle Detection Through Multidimensional Classification for on Board Video Based Systems," *2007 IEEE International Conference on Image Processing*, pp. 321-324, 2007, doi: 10.1109/ICIP.2007.4380019.

[29] H. Ding, L. Tian, Y. Liu, M. li, and B. Guan, "Stereovision Based Generic Obstacle Detection and Motion Estimation Using V-stxiel Algorithm," *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 903-908, 2018, doi: 10.1109/ITOEC.2018.8740535.

[30] K. Sabe, M. Fukuchi, J. S. Gutmann, T. Ohashi, K. Kawamoto, and T. Yoshigahara, "Obstacle avoidance and path planning for humanoid robots using stereo vision," in *IEEE Int. Conf. Robotics and Automation*, pp. 592-597, 2004.

[31] B. Xi, R. Guo, F. Sun, and Y. Huang, "Simulation research for active simultaneous localization and mapping based on extended kalman filter," in *IEEE Int. Conf. Automation and Logistics*, pp. 2443-2448, 2008.

[32] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Analysis and improvement of the consistency of extended Kalman filter based SLAM," *2008 IEEE International Conference on Robotics and Automation*, pp. 473-479, 2008, doi: 10.1109/ROBOT.2008.4543252.

[33] K. Kohara, N. Suganuma, T. Negishi, and T. Nanri, "Obstacle detection based on occupancy grid maps using stereovision system," *Int. J. Intelligent Transportation Systems Research*, vol. 8, no. 2, pp. 85-95, May 2010.

[34] A. Garulli, A. Giannitrapani, A. Rossi, and A. Vicino, "Mobile robot SLAM for line-based environment representation," in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 2041-2046, 2005.

[35] J. Jiao, R. Wang, W. Wang, S. Dong, Z. Wang, and W. Gao, "Local Stereo Matching with Improved Matching Cost and Disparity Refinement," in *IEEE MultiMedia*, vol. 21, no. 4, pp. 16-27, Oct.-Dec. 2014, doi: 10.1109/MMUL.2014.51.

[36] M. Tahmasebi, M. Gohari, and A. Emami, "An autonomous pesticide sprayer robot with a color-based vision system," *Int. J. Robotics and Control Sys.*, vol. 2, no. 1, pp. 115-123, 2022.

[37] J. Laarni, H. Koskinen, and A. Väätänen, "Concept of operations as a boundary object for knowledge sharing in the design of robotic swarms," *Int. J. Robotics and Control Sys.*, vol. 2, no. 4, pp. 692-708, 2022.

[38] M. G. Mohanan, and A. Salgaonkar, "Robotic Motion Planning in Dynamic Environments and its Applications," *Int. J. Robotics and Control Sys.*, vol. 2, no. 4, pp. 666-691, 2022.

[39] I. Hassani, I. Maalej, and C. Rekik, "Robot path planning with avoiding obstacles using free segments and turning points algorithms," *Mathematical problem in Engineering*, vol. 2018, no. 6, pp. 1-13, 2018.

[40] S. Sahloul, D. Ben Halima and C. Rekik, "An hybridization of global-local methods for autonomous mobile robot navigation in partially-known environments," *J. Robotics and Control*, vol.2, no. 4, pp. 351-355, 2021.

[41] L. Cai, L. He, Y. Xu, Y. Zhao, and X. Yang, "Multi-object detection and tracking by stereo vision," *Patt. Recogn.*, vol. 43, no.12, pp. 4028-4041, Dec. 2010.

[42] H. Liu, Y. Cai, S. Zhou, and J. Yang, "Stereo Matching with Multi-scale Based on Anisotropic Match Cost," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, Aug. 2020.

[43] M. V. Bobyr, N. A. Milostnaya, S. V. Gorbachev, S. Bhattacharyya, and J. Cao, "The Method of Depth Map Calculating Based on Soft Operators in Multi-Agent Robotic Stereo Vision Systems," *Research Transcripts in Computer, Electrical and Electronics Engineering*, vol. 2, pp. 83-98, 2021.

[44] T. Mahalingam and M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 2-18, Jan. 2021.

[45] W. Sun, "Robot Obstacle Recognition and Target Tracking Based on Binocular Vision," *Advances in Multimedia*, vol. 2022, pp. 1-12, Aug. 2022.

[46] M. T. Shahria, M. S. H. Sunny, M. I. I. Zarif, J. Ghommam, S. I. Ahamed, and M. H. Rahman, "A Comprehensive Review of Vision-Based Robotic Applications: Current State, Components, Approaches, Barriers, and Potential Solutions," *Robotics*, vol. 11, no.6, pp. 1-20, Dec. 2022.

[47] Q. Zhu, Y. Han, P. Liu, Y. Xiao, P. Lu, and C. Cai, "Motion planning of autonomous mobile robot using recurrent fuzzy neural network trained by extended Kalman filter," *Computational intelligence and neuroscience*, vol. 2019, pp. 1-17, Jan. 2019.

[48] K. S. Sharma and P. V. Manivannan, "Stereo Image Partitioning Based Fuzzy Logic," *Int. J. Mechanical Engineering and Robotics Research*, vol. 9, no. 8, Aug. 2020.

[49] H. Shabanian and M. Balasubramanian, "A novel factor graph-based optimization technique for stereo correspondence estimation," *Scientific Reports*, vol. 12, no. 1, p. 15613, Sept. 2022.

[50] S. Hożyń and B. Żak, "Stereo Vision System for Vision-Based Control of Inspection-Class ROVs," *Remote Sensing*, vol. 13, no. 24, pp. 1-25, Dec. 2021.

[51] Y. Wang *et al.*, "Vision based obstacle detection using rover stereo images," in *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, pp. 1471-1477, 2019.

[52] S. Alqahtani, S. Taylor, I. Riley, R. Gamble, and R. Mailler, "Predictive path planning algorithm using kalman filters and mtl robustness," in *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 1-7, 2018,

[53] T. Yan, Y. Gan, Z. Xia, and Q. Zhao, "Segment-Based Disparity Refinement With Occlusion Handling for Stereo Matching," in *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3885-3897, Aug. 2019, doi: 10.1109/TIP.2019.2903318.

[54] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Near real-time stereo based on effective cost aggregation," *2008 19th International Conference on Pattern Recognition*, pp. 1-4, 2008, doi: 10.1109/ICPR.2008.4761024.

[55] Y. Xu, Y. Zhao, and M. Ji, "Local stereo matching with adaptive shape support window-based cost aggregation," *Appl. Opt.*, vol. 53, no. 29, pp. 6885-6892, Oct. 2014.

[56] J. Zhang *et al.*, "RGB-D saliency detection via cascaded mutual information minimization," In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4338-4347, 2021.

[57] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[58] T. J. Ross, *Fuzzy logic with engineering applications*. John Wiley & Sons, 2009.

[59] A. A. Shetty, V. I. George, C. G. Nayak, and R. Shetty, "Fuzzy logic-based disparity selection using multiple data costs for stereo correspondence," *Turk. J. Elec. Eng. & Comp. Sci.*, vol. 27, no. 1, pp. 377-391, Jan. 2019.

[60] F. Mroz and T. P. Breckon, "An empirical comparison of real-time dense stereo approaches for use in the automotive environment," *J. Image Video Proc.*, vol. 2012, no. 13, pp. 1-19, Aug. 2012.

[61] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," *Procedia Comp. Sci*., vol. 54, pp. 764-771, Jan. 2015.

[62] Y. Marom and D. Feldman, "k-Means clustering of lines for big data," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[63] S. Zhang, H. Wang, W. Huang, and Z. You, "Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG," *Optik*, vol. 157, pp. 866-872, Mar. 2018.

[64] S. Duggal, S. Wang, W. C. Ma, R. Hu, and R. Urtasun, "Deeppruner: Learning efficient stereo matching via differentiable patchmatch," In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4384-4393, 2019.

[65] S. Ma, X. Bai, Y. Wang, and R. Fang, "Robust stereo visual-inertial odometry using nonlinear optimization," *Sensors*, vol. 19, no. 17, p. 3747, 2019.

[66] H. Luo, C. Pape, and E. Reithmeier, "Robust RGBD Visual Odometry Using Windowed Direct Bundle Adjustment and Slanted Support Plane," in *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 350-357, Jan. 2022, doi: 10.1109/LRA.2021.3126347.

[67] Z. Rao, M. He, Y. Dai, Z. Zhu, B. Li, and R. He, "MSDC-Net: Multi-Scale Dense and Contextual Networks for Stereo Matching," *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 578-583, 2019, doi: 10.1109/APSIPAASC47483.2019.9023237.

[68] Y. P. Chandra and T. Matuska, "Stratification analysis of domestic hot water storage tanks: A comprehensive review," *Energy and Buildings*, vol. 187, pp. 110-131, 2019.

[69] Z. Su, L. Xu, Z. Zheng, T. Yu, Y. Liu, and L. Fang, "Robustfusion: Human volumetric capture with data-driven visual cues using a rgbd camera," In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pp. 246-264, 2020.

[70] R. Battrawy, R. Schuster, O. Wasenmüller, Q. Rao, and D. Stricker, "LiDAR-Flow: Dense Scene Flow Estimation from Sparse LiDAR and Stereo Images," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7762-7769, 2019, doi: 10.1109/IROS40897.2019.8967739.

[71] S. Lee, Y. Kim, and E. Eisemann, "Iterative depth warping," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 5, pp. 1-13, 2018.

[72] Q. Wang, S. Shi, S. Zheng, K. Zhao, and X. Chu, "FADNet: A Fast and Accurate Network for Disparity Estimation," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 101-107, 2020, doi: 10.1109/ICRA40945.2020.9197031.

[73] H. I. F. Ibrahim, H. Khaled, N. A. Seada, and H. M. Faheem, "Parallel Dense Binary Stereo Matching Using CUDA," *2020 15th International*

Conference on Computer Engineering and Systems (ICCES), pp. 1-6, 2020, doi: 10.1109/ICCES51560.2020.9334591.

[74] Yaman and S. Kalkan, "Performance evaluation of similarity measures for dense multimodal stereo vision," *J. Elect. Imag.*, vol. 25, no. 3, pp. 1-31, Jun. 2016.

[75] M. Baydoun and M. A. Al-Alaoui, "Enhancing stereo matching with varying illumination through histogram information and normalized cross correlation," in *IEEE International Conference on Systems, Signals and Image Processing*, pp. 5-9, 2013.

[76] Y. C. Tsai, K. H. Chen, Y. Chen, and J. H. Cheng, "Accurate and fast obstacle detection method for automotive applications based on stereo vision," in *2018 International Symposium on VLSI Design, Automation and Test*, pp. 1-4, 2018.

[77] S. D. Perkasa, P. Megantoro, and H. A. Winarno, "Implementation of a camera sensor pixy 2 CMUcam5 to a two wheeled robot to follow colored object," *Journal of Robotics and Control*, vol. 2, no. 6, pp. 469-501, 2021.

[78] R. Fan and M. Liu, "Road Damage Detection Based on Unsupervised Disparity Map Segmentation," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4906-4911, Nov. 2020, doi: 10.1109/TITS.2019.2947206.

[79] H. Tao and X. Lu, "Automatic smoky vehicle detection from traffic surveillance video based on vehicle rear detection and multi-feature fusion," *IET Intelligent Transport Systems*, vol. 13, no. 2, pp. 252-259, 2019.

[80] N. Y. Ershadi, J. M. Menéndez, and D. Jiménez, "Robust vehicle detection in different weather conditions: Using MIPM," *PloS one*, vol. 13, no. 3, p. e0191355, 2018.

[81] J. Lian *et al.*, "Study on Obstacle Detection and Recognition Method Based on Stereo Vision and Convolutional Neural Network," *2019 Chinese Control Conference (CCC)*, pp. 8766-8771, 2019, doi: 10.23919/ChiCC.2019.8866348.

[82] O. Amimi, A. Mansouri, S. D. Bennani, and Y. Ruichek, "Stereo vision based advanced driver assistance system," in *2017 International Conference on wireless Technologies, Embedded and Intelligent Systems*, pp. 1-5, 2017.

[83] T. O. Kvålseth, "On Normalized Mutual Information: Measure Derivations and Properties," *Entropy*, vol. 19, no. 11, p. 631, 2017.

[84] D. H. Lee, J. Y. Chang, and Y. S. Heo, "Stereo matching using cost volume fusion for high dynamic range scenes," *Electronics Letters*, vol. 53, no. 23, pp. 1522-1524, 2017.

[85] M. Lopez-Franco, E. N. Sanchez, A. Y. Alanis, and C. López-Franco, "Neural control for driving a mobile robot integrating stereo vision feedback," *Neural Processing Letters*, vol. 43, no. 2, pp. 425-444, 2016.

[86] L. Pérez, Í. Rodríguez, N. Rodríguez, R. Usamentiaga, and D. García, "Robot guidance using machine vision techniques in industrial environments: A comparative review," *Sensors*, vol. 16, no. 3, pp. 335, 2016.

[87] R. A. Hamzah, H. Ibrahim, and A. H. A. Hassan, "Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation," *Journal of Visual Communication and Image Representation*, vol. 42, pp. 145-160, 2017.

[88] L. Li, S. Zhang, X. Yu, and L. Zhang, "PMSC: Patch Match-based superpixel cut for accurate stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 3, pp. 679-692, 2016.

[89] M. Abdullah-Al-Noman, A. N. Eva, T. B. Yeahyea, and R. Khan, "Computer Vision-based Robotic Arm for Object Color, Shape, and Size Detection," *Journal of Robotics and Control*, vol. 3, no. 2, pp. 180-186, 2022.

[90] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3D object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3782-3795, 2019.

[91] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu, "Convolutional neural network-based detection and judgement of environmental obstacle in vehicle operation," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 2, pp. 80-91, 2019.

[92] Q. Dong and J. Feng, "Outlier detection and disparity refinement in stereo matching," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 380-390, 2019.

[93] N. Choi, J. Jang, and J. Paik, "Illuminant-invariant stereo matching using cost volume and confidence-based disparity refinement," *Journal of the Optical Society of America A*, vol. 36, no. 10, pp. 1768-1776, 2019.

[94] D. Falanga, K. Kleber, and D. Scaramuzza, "Dynamic obstacle avoidance for quadrotors with event cameras," *Science Robotics*, vol. 5, no. 40, pp. 1-51, 2020.

[95] V. Kramar, O. Kramar, and A. Kabanov, "Self-Collision Avoidance Control of Dual-Arm Multi-Link Robot Using Neural Network Approach," *Journal of Robotics and Control*, vol. 3, no. 3, pp. 309-319, 2022.