INDIAN JOURNAL OF
ENGINEERING

19(51), 2022

Design on FPGA of an obstacle detection module over stereo image for robotic learning

Mohamed Salah Salhi[1*], **Hamid Amiri**[2]

To Cite:
Salhi MS, Amiri H. Design on FPGA of an obstacle detection module over stereo image for robotic learning. *Indian Journal of Engineering*, 2022, 19(51), 72-84

Author Affiliation:
[1]University of Tunis El Manar, National Engineering School of Tunis, Tunisia
[2]Research Laboratory of Signal Image and Information Technology LR-SITI, Tunisia

***Corresponding Author:**
University of Tunis El Manar, National Engineering School of Tunis, Tunisia
Email: medsalah.salhi@enit.utm.tn

Peer-Review History
Received: 15 January 2022
Reviewed & Revised: 18/January/2022 to 26/February/2022
Accepted: 28 February 2022
Published: 01 March 2022

Peer-Review Model
External peer-review was done through double-blind method.

DISCOVERY
SCIENTIFIC SOCIETY

ABSTRACT

This paper presents an accelerated algorithm realized on field programming gate array FPGA to mimic the human vision system. The FPGA is chosen as support owing to their real time propriety RTP. The proposed approach aims to detect any natural image and obtain a "cyclopean image" that provides information on the objects depth with a better details resolution. This strategy uses the common advantages of hybrid stereo imaging models in processing two views of the same scene captured from different throwing angles. Then, it merges the two images to produce a final image containing the depth information of different objects. The adopted algorithm is applied to obstacles detection, for robotic learning, in which execution time is experimented and presented in results.

Keywords: Accelerated algorithm; Stereo imaging models; Cyclopean image; Real time propriety RTP.

1. INTRODUCTION

The stereo image history is linked to the human vision way, which exhibits many interesting properties. One of them is the ability to feel the depth of the scenes seen. The process behind this ability is called “stereopsis”. In fact, the brain processes the images from both eyes, the right and the left, and executes an image called a "cyclopean image". The cyclopean image then provides information on the objects depth with a better details resolution.

In order to mimic the human vision system, stereo imaging models have been created. The stereo vision system processes two views of the same scene captured from different throwing angles. Then, it merges the two images to produce a final image containing the depth information of different objects [1].

Stereo imaging has found its application in many fields such as engineering, architecture, science, education and the military. We can find systems based on stereo imagery in 3D computer graphics, simulators, training systems and robotics.

The primitives and dimensions established by a stereo image have been widely applied in home automation, robotic and other security measures [2].

A growing field of application is the autonomous driving of robots through guidance systems. The stereo imaging sensors used in this application type are the most often mounted in a "fish-eyes» topology. They, therefore, have good features such as the wide detection angle: 3/4 of a surrounding circle, with vertical and horizontal resolution, the ability to detect obstacles in a short time and the simultaneous detection of different obstacles with information about their positions and shapes.

In this paper, we try to benefit from stereo imagery to develop a ground obstacle detection module. This module will allow a robotic system to sense its environment to make better decisions while determining its orientation.

Indeed, ground is considered a reference when detecting obstacles and determining their real position [3].

Many approaches have been used to tackle the ground plane estimation, such as Least Squares LS, Eigen Value EV, and Expectation Maximization Methods EMM, etc [3].

Ground plane estimation is about distinguishing the ground pixels in an image. For an application, like pedestrian detection, ground estimation is important.

This paper is presented under different sections whose study and analyzes respectively a literal review on stereo image techniques, the adopted strategy and the experimented results.

2. EXISTING GROUND DETECTION TECHNIQUES

This section presents different ground detection methods. We will explain their principles and give an overview of their results.

2.1. Determination of the ground under the hypothesis of a flat road

The work conducted by Manolis and al. presents a method allowing a mobile vehicle to locate obstacles using a pair of images taken from its environment. Their method assumes that the vehicle is traveling on level ground. Their applied technique calculates a homography of the ground, which makes it possible to determine the movement of the ground and then allows the obstacles detection. Thus, the camera calibration is not necessary and the technique applies to both single and stereo images, so it does not rely on the disparity map calculation nor go through 3D reconstruction [3].

Their proposed technique starts by selecting and matching a set of corners from a pair of images, then, using the Least Squares Algorithm LSA based on its Median calculation, it identifies which of these corners belong to the ground. The following figure illustrates the processing flow of this technique.

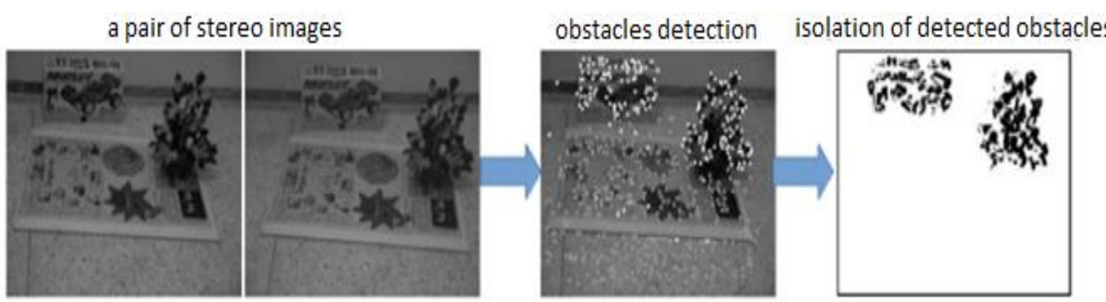


Figure 1: Ground sensing design flow based on plane geometry assumption

The picture in the middle represents the detection result of aberrations by the median of least squares when estimating the ground homography.

2.2. Role of field knowledge in obstacle detection

The work of Zhongfei Zhang and al. [4] investigated the role of a priori knowledge of the ground region in obstacle detection. Three types of algorithms were examined and a deduction was made that the algorithm, which continuously estimates the ground plane is the most robust and real. In fact, their research compared three types of algorithms: the first, which assumes a priori knowledge of the ground region, the second, which determines the ground region without prior knowledge and the third continuously, estimates the ground surface. The first two processes were sensitive to noise caused by ground surface irregularities, such as bumps and pits, but they were quick in determining the ground area. Their weakness is that they measure the deviation of data acquired from a known or unknown planar configuration, but it is not usually the case for the ground plane to be planar.

The ground surface estimation algorithm has been continuously proven to be robust to noise. The results given in the work of Zhongfei Zhang and al. show that the continuous estimation of the ground plane is better compared to other algorithms. Yet, its use

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

offers an advantage which is summed up in its ability to determine the height of obstacles although it requires partial calibration. However, it becomes ineffective in the absence of calibrated cameras.

2.3. Extraction of image primitives in obstacles detection

Dieter Koller and al. [5], in their research, proposed to use the features of track marks to control the ground surface. Their project used stereo imaging techniques to extract features from the road scene to help guide a vehicle while assuming the road level. Their developed model, using stereo analysis, continually predicts the position of lane markings that dramatically improves overall robustness and accuracy. Nevertheless, this technique depends closely on the quality of the acquired images, even if the track marks are not visible, which is a frequent case that limits the functionality of this technique.

Figure 2: Estimation of markings on a lane

2.4. Ground determination by monocular sequences

Ground can be determined by processing monocular images. However, special algorithms need to be developed. Jin Zhou [6] presented in this work, an approach to detect the ground plane based on homography. The images acquired are monocular with the advantage of being cost effective and the technique does not need calibration as in stereoscopic approaches. A ground plane homograph is determined from the derived constraints related to the positioning of the camera. Then, specially designed algorithms determine the ground surface. The algorithms presented are efficient, robust and precise.

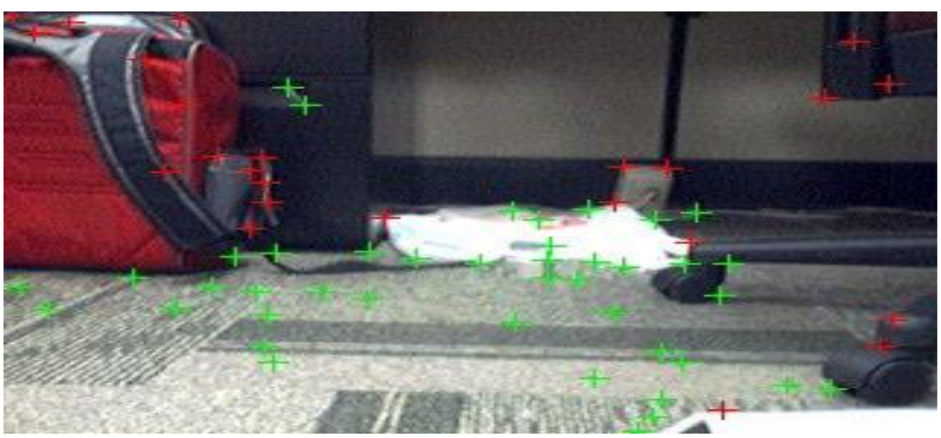


Figure 3: Illustration of a ground detection based on the monocular image

In this figure, a set of random corners is selected and the ground is determined by green crosses using appropriate algorithms.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

2.5. The ground detection by the "V-DISPARITY" image

Usually, the ground surface is not planar. It has its ups and downs due to its topography. Therefore, the often assumption based on planar ground is invalid. In fact, this idea causes various problems, such as imprecision and false detection of obstacle positions.

Labayrade and al. [7] have proposed a new model, which can increase the reliability of the obstacle detection process. His method could detect obstacles without taking plane geometry; it was also original, fast and robust.

The advantage of this strategy is its ability to adapt to the up and down slopes. It is based on the construction and processing of the v-disparity image, which includes the projection of different features forming the scene being watched. The 'v-disparity' image provides a semi-global match. Obstacle detection is robust to partial occlusion or errors made during its construction. In addition, it is not necessary to extract external structures like lane markings or road edges to control the ground region from a pair of stereo images.

The construction of the v-disparity image is relatively simple. Based on it, it is possible to determine an obstacle detection method in both cases of plane and non-plane geometry. After estimating the road profile, objects located on the ground surface will be considered potential obstacles. Therefore, an accurate calculation of the points of contact between the ground and obstacles can be determined.

Figure 4: Representation of the V-disparity approach

2.6. Comparison between existing ground detection techniques

The following table compares the different algorithms presented in ground detection. Obviously, the image processing v-disparity technique is more advantageous than the other methods. However, the passage through the construction of the disparity map complicates the implementation of this algorithm on an FPGA.

Table 1: Comparison between different approaches in ground detection

Techniques	Inconveniently	Advantages
Flat road assumption	• Complex algorithm. • Need a textured floor. • The algorithm can be faulty if the	• No camera calibration. • Stereo or monocular application. • Efficient and robust.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

	data are wrongly chosen	
Prior knowledge of the ground	• Sensitive to noise. • There are cases where the system is unresolvable. • Complex algorithm.	• Robust in continuous ground estimation. • Fast when you have prior knowledge of the ground surface.
Lane marking method	• Depends on the clarity of the ground characteristics • Real-time issue	• Good performance in clear scenes • Simple and robust
Monocular Sequence Methods	• Complex algorithm • Poor results if the ground is not dominant in the scene	• Efficient and robust • Profitable • No calibration required
V-disparity method	• Using the disparity map slows down the process • The disparity map needs more processing such as smoothing and filtering	• No plane geometry assumption • Fast, reliable and robust against noise • Precise • Simple algorithms • Rich information included in the v-disparity image

Currently, a new approach is introduced in the sense of ignoring the calculation of the disparity map and directly constructing the v-disparity from a pair of stereo images while maintaining the same quality as that proposed by Labayrade [8].

3. SELECTED MODEL FOR IMPLEMENTATION

The process of detecting traversable regions is of great importance especially for navigation systems [9], [10], so it needs a proficient system that provides rich information to accomplish the task efficiently. For this reason we have chosen to use stereo imagery because it provides more information compared to other systems based on lasers, radar or monocular systems, etc.

Stereo vision calculates disparity by measuring distances between objects. Two ways to deal with the resulting image are built according to the disparity.

The first way is the 3D reconstruction according to the point cloud derived from a disparity map. Then, obstacles and the ground areas can be detected using edge detection, local safety map, level adjustment, etc.

However, this method is recognized by its high computational value, which makes it difficult to meet real-time requirements in robot navigation.

The second method is based on the building of a v-disparity image. Labayrade introduced this technique to detect obstacles either on a flat or non-flat road. The advantage of this method, as mentioned earlier, is that it does not consider the flat road assumption or the need to extract specific structures like road edges. It consists of accumulating pixels having the same disparity along each row of a disparity map image. The resulting image includes significant patterns that reflect different objects constructing the actual scene.

Obstacles, arranged perpendicular to the robot, are represented by vertical lines of width indicated by the pixel intensity. In addition, the road which is modeled as a succession of flat surfaces will be mapped as an oblique line segment. This line is called the ground correlation line. It represents the traversable region that contains information on the coverage angle of the cameras [11], [12].

Considering the lack of hardware resources on an FPGA platform, the v-disparity-based processing system is more appropriate even if the computation of the disparity map looks like an inhibiting factor. The work of Benenson showed that the construction of the v-disparity image can be done directly so that this method became the best candidate for our case. Additionally, inspecting the image required an algorithm to adjust the lines. Hough's transform was an option, but the 'Iteratively Reweighted Least Squares' IRLS algorithm presented a simpler and more robust solution to implement [13].

The computation of least squares weighted by the weights is performed iteratively to solve linear regression problems. The algorithm produces weights for the dataset that group observations in a tight linear relationship. The IRLS is used to calculate Maximum Likelihood Estimates' MLE. It is also popular for calculating robust estimates. It can offer a quick solution for difficult problems. Thus, the use of the IRLS algorithm is more practical in performing a linear regression that will allow the determination of the ground representative line in the constructed image.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

4. ADOPTED STRATEGY

Although the presented techniques give robust and fast results, they require General Purpose Processors GPPs and Graphic Processor Units GPUs implementation. Therefore, these solutions are not suitable for applications that lack power and space resources like small robots. Therefore, implementation on an FPGA platform is necessary to have low power consumption and to preserve the same processing quality.

Due to the considerable speed gain and the simplicity of the calculation, we will adopt the concepts developed by R.Benenson and make them suitable for implementation on an FPGA.

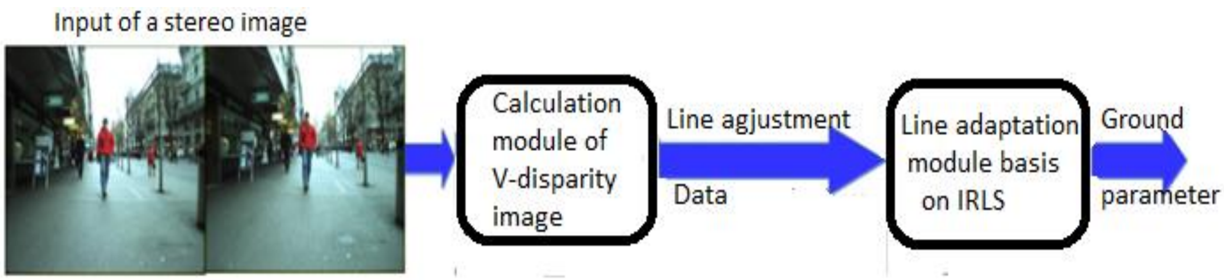


Figure 5: The stages of soil detection

In this work, we propose to implement on FPGA a ground detection module. The developed module could meet real-time execution requirements. It should allow an autonomous driving system, such as a robot, to integrate a low power, high performance and effective cost processing system.

It could therefore enable the vehicle to process the stereoscopic images in flow and to determine the ground region in a scene, making it possible to detect obstacles and to avoid them.

The best suited method to implement on the FPGA was chosen from other techniques. This method is based on building a v-disparity image and determining its ground region. The verification of the chosen algorithms was conducted by implementing them on Matlab and studying its feasibility by dividing it into smaller modules.

This approach leads to the design and implementation of specialized state machines, which calculate from a given pair of stereo images an image of v disparity. The latter will allow the system to determine the ground and any obstacles in the scene being watched.

4.1. V-disparity image computation

Given a stereo-image (left and right images), we take only the lower half part to compute the V-disparity image.

The algorithm can be written as follows:

```
-------------------------------------------------------------------------------------------------------------------
For each row in the left image
{
get left_row_iterator_beginning;
get left_row_iterator_ending;
get right_row_iterator_beginning;
for (disparity = 0; disparity <disparity-max; disparity++);
{
left_row_iterator = left_row_iterator_beginning + disparity + offset;
right_row_iterator = right_row_iterator_beginning;
V-disparity_cost = (Σmin(cost, cost_sum_saturation ))/3;
v-disparity-image_row[d] = V-disparity_cost;
}
}
```


ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

- **The cost function is computed as follows:**

cost = |left_r − right_r| + |left_b − right_b| + |left_g − right_g|;
_r = red pixel;
_b = blue pixel;
_g = green pixel;

- **Line parameter extraction:**

The line parameter extraction is based on the Iteratively Reweighted Least Squares (IRLS).
This algorithm is made to minimize the system: $w(i) * (b{-}A{*}x)$
"w" stands for the weights chosen for computation. The algorithm for this part can be written as follows:

- **Select points for line fitting:**

```
for ( disparity = 0; disparity <disparity-max; disparity++); {
if (V-disparity_row <= threshold); {
store the point of coordinates " x= disparity " and "y= row_number";
}
}
```

Initialization of A,b,x and w:

$A = x11. .. .. .. .xn1;$
$B = y1....yn;$
$X = ab;$
$W = w1....wn;$

"a" stands for the "slope of the line to be detected " .
"b" stands for "the coordinate at the origin".

- **Main algorithm**

```
for(I = 0; I < number_of_iterations, i++); {
recompute weights
w(i) = w(i)*w(i−1);
solve w(i)*A*x = w(i)*b;
}
```

- **Recomputed weights:**

The error vector can be written as:

$$e = b - A{*}x \tag{1}$$

We can compute weights according to "Tukey" as follows:

$$w(i)=(1-(e\ (i)/c\ )2)2 \tag{2}$$

4.2. Hardware implementation

In this section, we will present our solution for the implementation of ground detection module on an ARM-FPGA platform. The processing frequency of our module is 100 MHz.

4.3. Proposed hardware solution

We have chosen to implement the v-disparity image construction module on the programmable logic part of the FPGA. We decide to implement the line adaptation module on the processing system. Communication between the modules will be maintained in the AXI Bus interface.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

5. IMPLEMENTATION OF THE V-DISPARITY MODULE AND RESULTS

A state machine for building v-disparity images has been designed and implemented as shown in the following figure. However, the execution time was enormous and its image processing capacity was nowhere near 25 fps (frames per second). One row of the stereo image pair took a time of 3 ms, so we got a disparity image for 3ms * 240 rows = 720 ms, which is less than 2 fps.

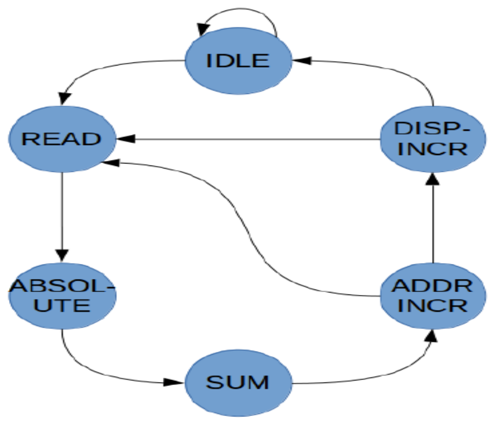


Figure 6: State machine for v-disparity image construction (first draft)

In the state "IDLE_State", the FSM for "First State Machine" is inactive.
In the 'READ_State' state, the FSM starts reading buffers.
In the 'ABSOLUTE_State' state, the FSM calculates the absolute difference between the right and left pixels.
In the 'SUM_State' state, the FSM adds the new difference to the cost.
In the "ADDRINCR_State" state, the FSM increments the address of the buffers from which it reads the different pixels values.

Then, it starts to read if it has not reached the end of the left buffer, otherwise it will increase the disparity.
In the 'DISPINCR_State' state, the FSM increments the disparity. Then, it starts reading if it does not reach the maximum disparity value; otherwise, it goes into the inactive state.

Figure 7: Runtime for the first draft implementation (3000us)

5.1. Fast track solution

To achieve higher performance, a new design of finite state machines has been proposed. Algorithm profiling has shown that the cost calculating for each disparity is the part which consumes the most time of the algorithm. Optimization involves the parallelism of this process and the processing of 8 pixels at a time.

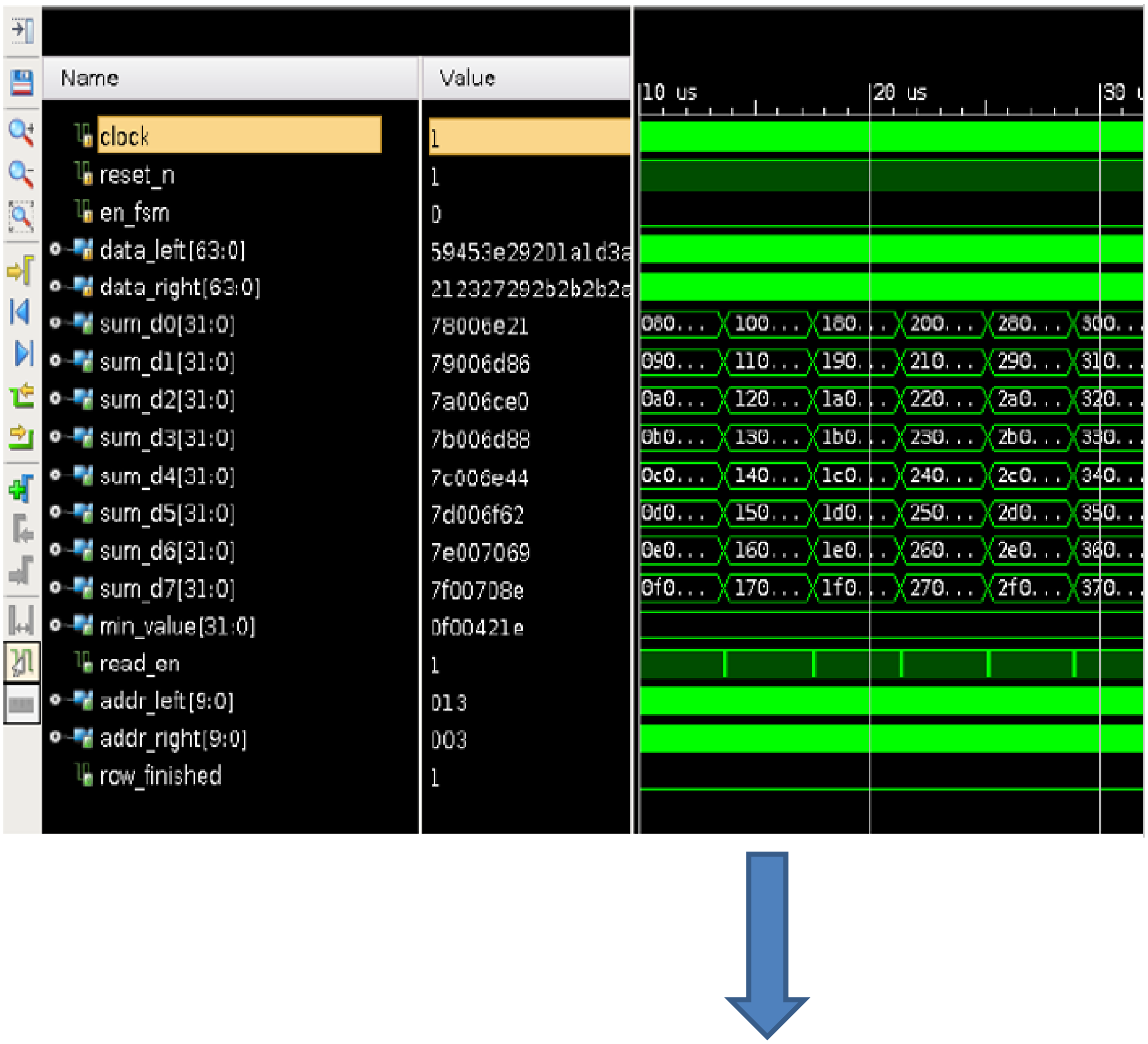
Because the luminescence value of the pixels is a positive integer, the cost can be written as follows.

$$cost = \sum_{n=1}^{n\_data} |a_n - b_n| = \sum_{n=1}^{n\_data} \max(a_n, b_n) - \sum_{n=1}^{n\_data} \min(a_n, b_n) \tag{3}$$

With (an) is a left pixel value and (bn) is a right pixel value. In this way we can focus only on the sum of the max and min values in parallel, for the difference.

So, by processing more data (8 pixels), the process could achieve more processing speed. It can process an image in 15.2ms and it can reach higher speeds. The acceleration rate was 740/15, 2 = 48 times.

The following figure shows the computational time taken to complete a row in the v-disparity construction. The state machine can calculate 8 disparities at a time that corresponds to the variables sum_dn.



(In follow the second part of the figure)

Figure 8: Execution time after acceleration (60us)

5.2. Implementation of the IRLS algorithm

The algorithm was developed and implemented using the C language. A runtime test was performed on an ARM (Advanced RISC Machine) Processor integrated on the FPGA. The results showed that it can run in enough 18ms for our system to get it up to 25fps of speed.

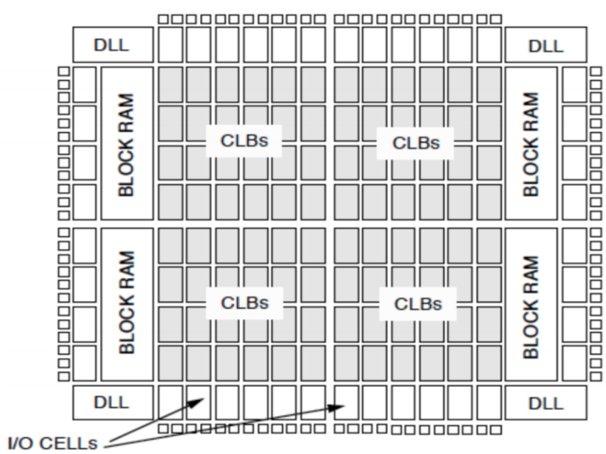


Figure 9: The FPGA architecture

Field-Programmable Gate Arrays FPGAs [14] are semiconductor chips incorporating Configurable Logic Blocks in a 2D array order. CLB's are connected via programmable interconnects. In fact, the interconnects make a network of vertical and horizontal wires that constitute links between the CLBs. Each intersection between the wires lodge a switcher which permits reconfiguration of CLBs. Modern FPGA's embed hundreds of thousands of CLBs with the inclusion of hardened functional units [14], [15]. The FPGA construction permits fast and efficient development of common functions. It is possible to program electronically the FPGA device

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS

by loading a bit stream configuration file into the device. The configuration is memorized in an SRAM memory, consequently an FPGA can be reprogrammed more than once.

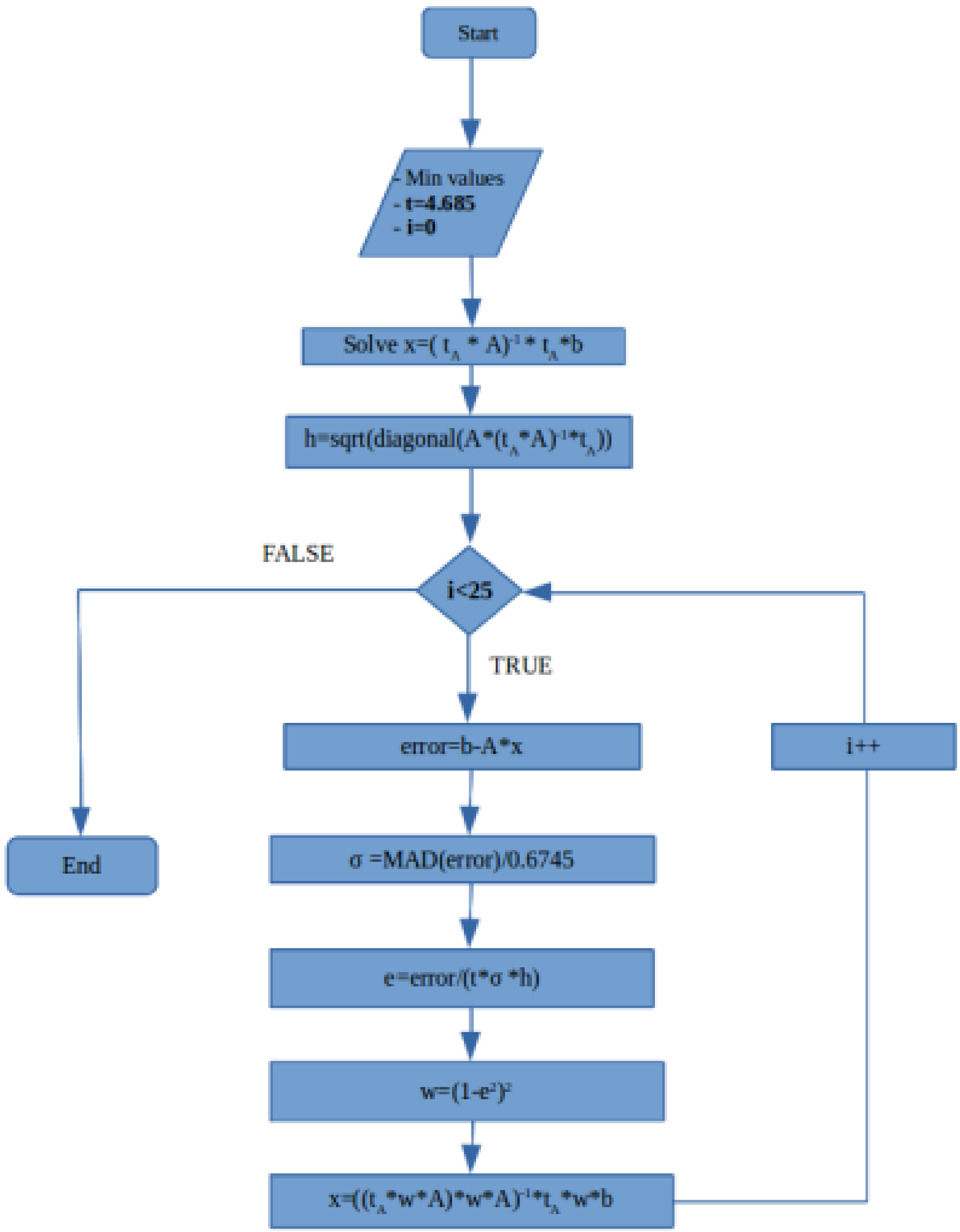


Figure 10: Flow chart for IRLS algorithm implementation

Validation of the hardware implementation of the v-disparity image construction module

The simulation of the v-disparity image construction module showed that the module could process one frame every 15.3ms. The quality of the resulting image was acceptable, but the destruction of the soil parameters had to be accurate due to the good quality of the resulting image.

Validation of the hardware implementation of the IRLS module

The IRLS calculation module was developed using C code and executed the ARM Processor embedded on an FPGA. The results showed that the module can process a frame in 18ms. The total time taken to process a frame is 15.3 ms + 18 ms = 33.3 ms. Therefore, the system could process: 1/33, 3 ms = 30 fps.

6. CONCLUSION

This topic presented the design, implementation and optimization of ground detection modules on FPGAs. We have chosen a method based on the construction of a v-disparity image, because; it is simpler to implement and robust. A direct method for constructing the v-disparity image made the implementation more eligible to try.

We studied the algorithms for constructing v-disparity images and the line adaptation used to detect the ground plane. For the implementation, we used Xilinx's Vivado development tool and zedboard development board.

The design included two main modules; the first allows the construction of v-disparity image. It was implemented on the programmable logic of the FPGA device. The second module aims at ground detection by the line alignment algorithm. It was implemented on the processing system. Communication between the two modules was provided by the AXI bus interface.

The implementation obtained reached a speed of 30 fps at a frequency of 100 MHZ but the communication between the modules is still to be optimized.

Funding

This study has not received any external funding.

Conflict of Interest

The author declares that there are no conflicts of interests.

Data and materials availability

All data associated with this study are present in the paper.

REFERENCES AND NOTES

1. Willian T. F. D. Silva, Filipe D. D. M. Borges. ‘Genetic Algoritms to determinate the optimal parameters of an ensemble local mean decomposition’, Acta Polytechnica journal, pp. 465-475, volume 61, No. 3-2021, DOI: https://doi.org/10.14311/AP.2021.61.0465.
2. Paulo Henrique Barbosa de Souza Pinheiro, Mayara Helena Moreira Nogueira Santos, Angelo Cesar Colombini, Bruno Wanderley França, Marcio Zamboti Fortes. ‘Detailed Modeling and Analysis of Digital MHO Distance Relay with single-pole operation', Acta Polytechnica journal, pp. 537-551, volume 61, No. 4-2021, DOI: https://doi.org/10.14311/AP.2021.61.0537.
3. KE Peiqi, MENG Cai, LI Jihao and LIU Ying. "Homography-based Ground Area Detection for Indoor Mobile Robot Using Binocular Cameras". 2011 IEEE 5th International Conference on Robotics, Automation and Mechatronics (RAM)
4. M.I.A. Lourakis and S.C. Orphanoudakis. "Visual detection of obstacles assuming a locally planar ground". In Proc. 3rd ACCV, volume II, pages 527-534, 1998.
5. Z. Zhang, R. Weiss, and R.A. Hanson. "Qualitative obstacle detection". In Proc. of International Conference on Computer Vision and Pattern Recognition, pages 554-559, 1994.
6. D. Koller, T. Luong, and J. Malik. "Binocular stereopsis and lane marker flow for vehicle navigation: lateral and longitudinal control". Technical Report UCB/CSD 94-804, University of California at Berkeley, Computer Science Division, 1994.
7. J. Zhou, B. Li, "Homography-based ground detection for a mobile robot platform using a single camera", in: Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, IEEE, 2006.
8. R. Labayrade, D. Aubert, and J.-P. Tarel. "Real time obstacle detection on non flat road geometry through v-disparity representation". In IVS, 2002.
9. King, R. D. (2013) "A brief history of stereoscopy". WIREs Comput Stat, 5: 334340. doi:10.1002/wics.1264
10. Labayrade, R., Aubert, D., Tarel, J.P.: "Real time obstacle detection on non flat road geometry through v-disparity representation". In: IVS. (2002)

11. Paul W. Holland & Roy E. Welsch (1977) "Robust regression using iteratively reweighted least-squares", Communications in Statistics - Theory and Methods, 6:9, 813-827, DOI: 10.1080/03610927708827533.
12. Mark Galassi, Jim Davies , James Theiler, Brian Gough, Gerard Jungman, Patrick Alken, Michael Booth, Fabrice Rossi,Rhys Ulerich:"GNU Scientific Library". Reference Manual, Edition 2.2.1, for GSL Version 2.2.1, 31 August 2016.
13. Mohamed Salah Salhi, Rahmouni M.H., Hamid amiri :' Evolutionary Deep-indicators Algorithm in Facial Recognition improvement over SoC, International Journal of Advanced Science and Technology Vol. 29, No. 5, (2020), pp. 5376 – 5387.
14. Petr Dlask: 'Simulation Model based on Regional Development and Virtual Changes', Acta Polytechnica journal, pp. 291-300(10), Volume 55, Number 5, 2015.
15. Anita Agárdi, Károly Nehéz. 'Parallel Machine Scheduling with Monte Carlo Tree search', Acta Polytechnica journal, pp. 307-312, volume 61, No. 2-2021, DOI: https://doi.org/10.14311/AP.2021.61.0307.

ISSN 2319–7757 EISSN 2319–7765 I OPEN ACCESS