International Journal of Applied Research 2018; 4(5): 147-154

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2018; 4(5): 147-154
www.allresearchjournal.com
Received: 08-03-2018
Accepted: 12-04-2018

David Rajkumar JayaKumar
M. Tech in Remote Sensing
and GIS, Intern-Proprocure,
London, England

Correspondence
David Rajkumar Jayakumar
M. Tech in Remote Sensing
and GIS, Intern-Proprocure,
London, England

Real-time multi-resolution range map generation using variable template matching and enhanced sum of absolute differences algorithm

David Rajkumar Jaya Kumar

Abstract
This paper presents multi-resolution real-time range map generation for stereo vision-based autonomous navigation system using template matching and the enhanced sum of absolute difference algorithm. Stereo vision sensor captures images from two different viewpoints to reconstruct the 3-dimensional information and navigate the autonomous ground vehicle by visual perception of the environment. The multi-resolution range map for the two images captured by stereo vision camera is generated using the high performance enhanced sum of absolute differences (SAD) algorithm, thereby computing the range of the object in the image. The depth of the object in the image is a vital parameter which is used by the autonomous navigation system.

Keywords: multi-resolution range map, template matching, enhanced sum

1. Introduction

An accurate and detailed 3D representation of the environment around a vehicle, with a passive sensor at relatively low cost, can be perceived using stereovision system. Capturing a scene (left and right images) from two points of view at the same time, stereo techniques aim at defining conjugate pairs of pixels, one in each image, that correspond to the same spatial point in the scene. The difference between positions of conjugate pixels, called the disparity, yields the depth of the point in the 3D scene. The range map is generated using Sum of Absolute Difference (SAD) algorithm.

2. Stereo Vision

Stereo vision is the process of extracting 3-Dimensional information from multiple 2-Dimensional views of a scene. The three-dimensional depth information can be reconstructed from two images using a computer by the corresponding the pixels in the left and right images captured using stereoscopic imaging sensors. Solving the Correspondence problem in the field of Computer Vision aims to create meaningful depth information from two images.

Artificial stereo vision is based on the same principles as biological stereo vision. A perfect example of stereo vision is the human visual system. Each person has two eyes that see two slightly different views of the observer's environment. An object seen by the right eye is in a slightly different position in the observer's field of view than an object seen by the left eye. Brain calculate the disparity and the approximate range of the object in the scene which is done by computer in a stereo vision system.

2.1 Geometry

A stereo vision system uses two cameras at two known positions know as viewpoints and captures a picture of the scene at the same time. The closer the object is to the cameras, the greater its difference in position in the two pictures taken with those cameras. The measure of that distance between the corresponding points from the optical axis is called the disparity from which the range of the object is estimated.

The purpose of the study was to find out the parental attitude towards female participation in sports. A self-made questionnaire was designed so as to get the relevant information that can be used for various purposes.

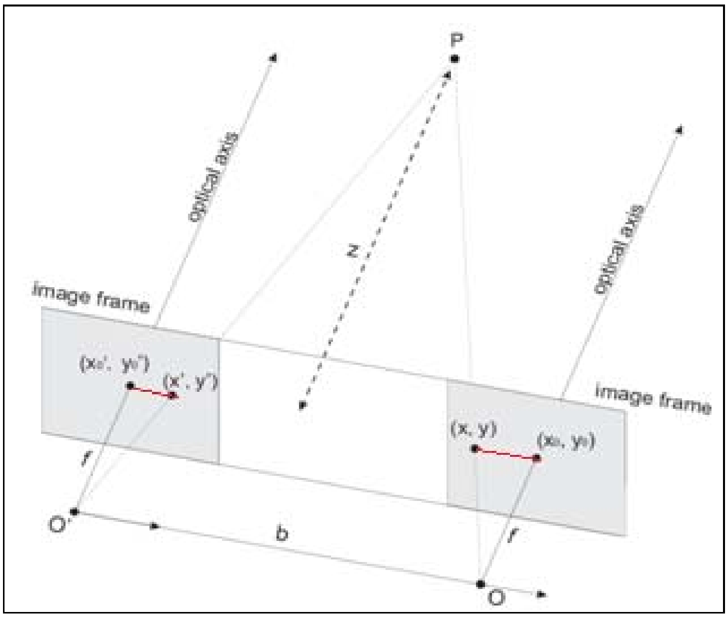


Fig 1: The geometry of stereo vision

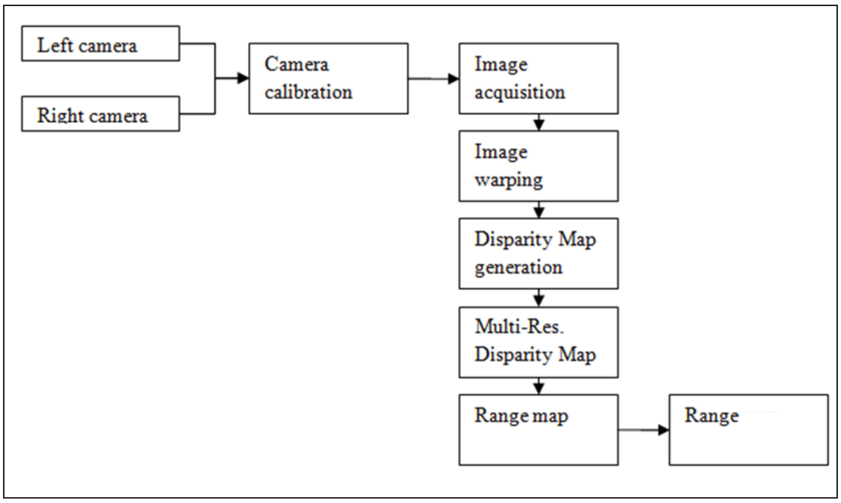
The distance 'f' is the perpendicular distance from each focal point to its corresponding image plane. Point 'P' is the object captured by these cameras. Point P has coordinates (x, y, z) measured with respect to a reference frame that is fixed to the two cameras and whose origin is at the midpoint of the line connecting the focal points. The projection of point P is shown as P_r in the right image and P_l in the left image and the coordinates of these points are written as (x_r, y_r) and (x_l, y_l) in terms of the image plane coordinate systems shown in the figure. The disparity defined above is x_l-x_r.

The distance is inversely proportional to disparity and that disparity is directly proportional to the baseline. When cameras are aligned horizontally, each image shows a horizontal difference, x_l-x_r, in the location of P_r and P_l, but no vertical difference. Each horizontal line in one image has a corresponding horizontal line in the other image. These two matching lines have the same pixels, with a disparity in the location of the pixels. The process of stereo correlation finds the matching pixels so that the disparity of each point can be known.

The objects at a great distance will appear to have no disparity. Since disparity and baseline are proportional, increasing the baseline will make it possible to detect a disparity in objects that are farther away. However, it is not always advantageous to increase the baseline because objects that are closer will disappear from the view of one or both cameras.

2.2 Methodologies

The process involved in the range estimation involves:



Stereo image sensors capture two images of the scene from two different viewpoints left and a right camera called stereo image pair. A stereo-pair image contains two views of a scene side by side.

Camera calibration is the process of estimating intrinsic and/or extrinsic parameters. Intrinsic parameters deal with the camera's internal characteristics, such as its focal length, skew, distortion, and image center. Extrinsic parameters describe its position and orientation in the world.

Image acquisition is the creation of photographic images, such as of a physical scene or of the interior structure of an object. The term is often assumed to imply or include the processing, compression, storage, printing, and display of such images. It's also called as Digital Imaging.

Image warping is the process of digitally manipulating an image such that any shapes portrayed in the image have been significantly distorted. Warping may be used for correcting image distortion as well as for creative purposes (e.g. morphing). The same techniques are equally applicable to video.

The disparity map is an array constructed by calculating the absolute difference in pixel values for each element in the left and the right image array which is constructed by finding pixel-to-pixel correspondences between the left and the right image array.

Multi-resolution representations have been used as part of a variety of visual algorithms ranging from image segmentation. Image compression algorithms transform image data from one representation to a new one that requires less storage space. Image pyramids are multi-resolution image representations.

The range is the relative distance of the obstacle from the stereo rig. The depth at which the obstacle is located is calculated using the disparity value, the focal length of the camera and the base line.

2.3 Image Rectification

Image rectification is a transformation process which is used to project multiple images onto a common image plane. It is used to rectify the distorted image into a standard coordinate system.

1. It is used in computer stereo vision to simplify the problem of finding matching points between images.
2. It is used in geographic information systems to merge images taken from multiple perspectives into a common map coordinate system.

Stereo vision uses triangulation based on epipolar geometry to determine the distance of an object. Between two cameras there is a problem of finding a corresponding point viewed by one camera in the image of the other camera which is called the correspondence problem. In most camera configurations, finding correspondences requires a search in two dimensions. However, if the two cameras are aligned to have a common image plane, the search is simplified to one dimension-a line that is parallel to the line between the cameras (the baseline). Image rectification is an equivalent (and more often used) alternative to this precise camera alignment. It transforms the images to make the epipolar lines (epipolar geometry) align horizontally.

2.4 Disparity

The measure of the difference between positions of pixels between the left and right image of the stereo image pair is called the disparity. There are two types of disparity. They are Horizontal Disparity and Vertical Disparity. If the two cameras are mounted on a vertically aligned frame then there will be horizontal disparity alone and no vertical disparity. If the two cameras are mounted on a horizontally aligned frame then there will be vertical disparity and no horizontal disparity.

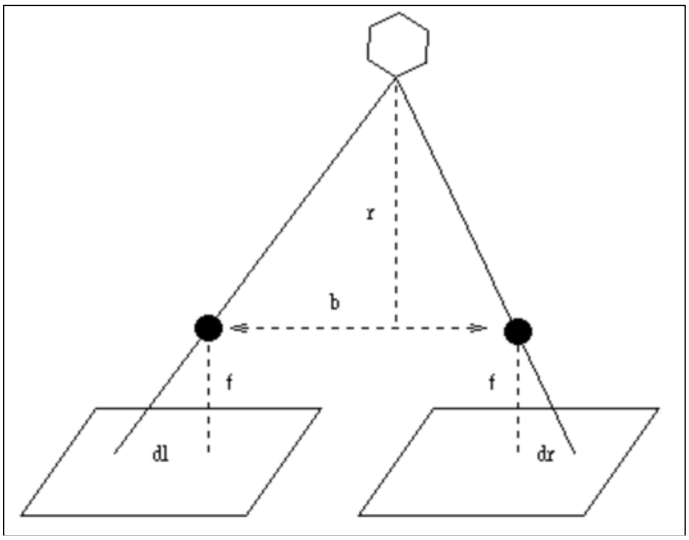


Fig 2: The offset of the image

Two images of the same object are taken from different viewpoints. The distance between the viewpoints is called the baseline (b). The focal length of the lenses is f. The horizontal distance from the image center to the object image is dl for the left image and d_r for the right image.

Normally, the stereo cameras are set up so that their image planes are embedded within the same plane. Under this condition, the difference between dl and dr is called the disparity and is directly related to the distance r of the object normal to the image plane. The relationship is

r = b * f/d,

Where d = d_l-d_r.

2.5 Range Estimation

A pixel in the disparity image represents the range of an object. This range, together with the position of the pixel in the image, determines the 3D position of the object. The coordinate system for the 3D image is taken from the optic center of the left camera of the stereo rig. Z is along the optic axis, with positive Z in front of the camera. X is along the camera scan lines, positive values to the right when looking along the Z-axis. Y is vertical, perpendicular to the scan lines, with positive values down. Finally, the viewpoint can be rotated around a point in the image, to allow a good assessment of the 3D quality of the stereo processing. The rotation point is selected automatically by finding the point closest to the left camera, near the optic ray of that camera.

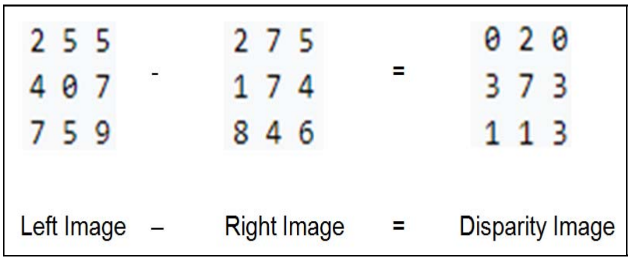
2.6 3D Reconstruction

The image coordinates of the mouse are given by the x, y values. The values in square brackets are the pixel values of the left and right images. If the right image is displaying stereo disparities, then the right value is the disparity value. Finally, the X, Y, Z values are the real-world coordinates of the image point, in meters.

3. Disparity map generation

3.1 Disparity Estimation

Finding pixel-to-pixel correspondences between the left and the right image array. In the example below, for each pixel in the left image, search for the most similar pixel in the right image. The disparity map will have greater accuracy by using neighborhood windows.



$$\begin{matrix}2&5&5\\4&0&7\\7&5&9\end{matrix} \quad - \quad \begin{matrix}2&7&5\\1&7&4\\8&4&6\end{matrix} \quad = \quad \begin{matrix}0&2&0\\3&7&3\\1&1&3\end{matrix}$$

Left Image – Right Image = Disparity Image

Fig 3: Disparity estimation

3.2 Epipolarity

The epipolar line is an imaginary line considered across the left and right image to compare the pixels in the image array and to calculate disparity. If the two cameras are horizontally aligned and warped, the epipolar line is taken horizontally. If the two cameras are vertically aligned and warped, the epipolar line is taken vertically.

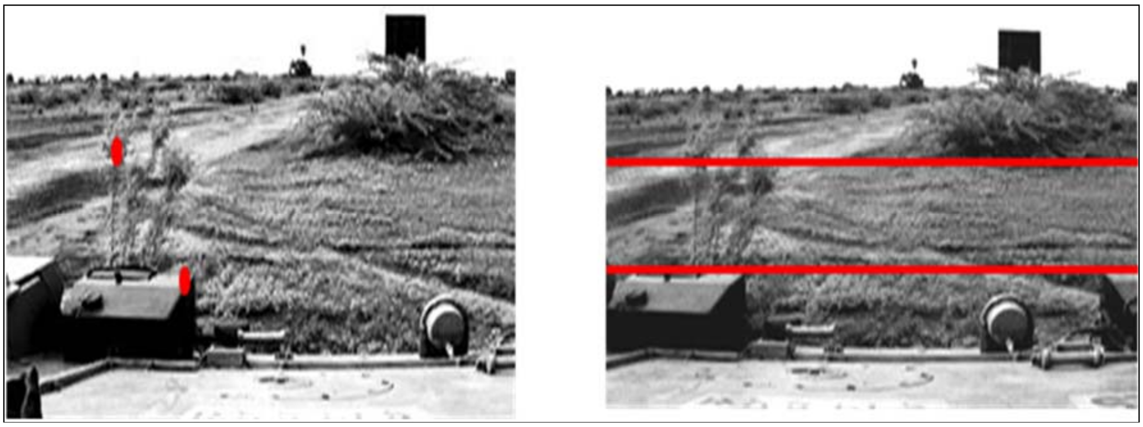


Fig 4: Epipolar line

3.3 Variable Template matching algorithm

The template matching algorithm is based on Pearson’s correlation. The application uses multiple templates which increases the tracking ability, but significantly increases the computation: a correlation must be computed for each template for each possible location of the template within the region of interest. Pearson’s correlation, represented by corr (2).

The template size doesn’t need to be the same to the target object on the image. Scan in various of size ratios of the image and the template to find the best match.

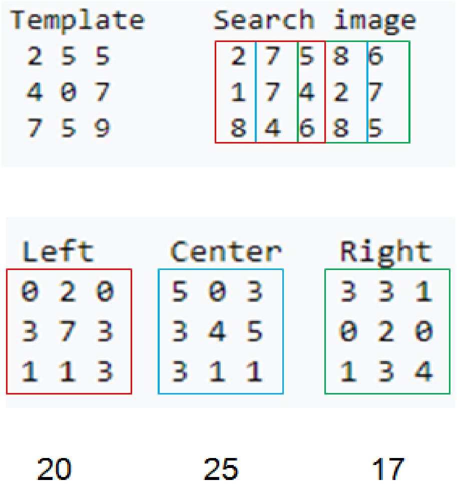
$$corr2(A,B)=\frac{\sum_M \sum_N (A_{MN}-\bar{A})(B_{MN}-\bar{B})}{\sqrt{\left(\sum_M \sum_N (A_{MN}-\bar{A})^2\right)\left(\sum_M \sum_N (B_{MN}-\bar{B})^2\right)}}$$

3.4 Sum of Absolute Differences

Disparity map generation algorithm uses a simple technique, SAD. SAD stands for Sum of Absolute Differences.

$$C(i,j,d)=\sum_{k}\sum_{l}((A_l(i+k,j+k)-B_r(i+k-d,j+l))$$

Where,
C (i, j, d) = Disparity Image array
A_l = Left Image array
B_r = Right Image array.



Template
$$\begin{matrix}2&5&5\\4&0&7\\7&5&9\end{matrix}$$

Search image
$$\begin{matrix}2&7&5&8&6\\1&7&4&2&7\\8&4&6&8&5\end{matrix}$$

Left
$$\begin{matrix}0&2&0\\3&7&3\\1&1&3\end{matrix}$$

Center
$$\begin{matrix}5&0&3\\3&4&5\\3&1&1\end{matrix}$$

Right
$$\begin{matrix}3&3&1\\0&2&0\\1&3&4\end{matrix}$$

20 25 17

Fig 5: Sum of Absolute Differences

3.5 Steps

The steps involved in disparity dap generation are

1. Read left and right image.
2. Store the images in an image array.
3. Calculate the absolute differences of each pixel in the image array using 5x5 neighborhood windows.
4. Find the sum of absolute differences.
5. Calculate the sum of absolute differences by linearly searching along 32 pixels in the right image for each

pixel in the left image. Here we search for 32 pixels because we use 32-bit disparity.

6. Compare and find the minimum sum of absolute difference.
7. Plot the minimum sum of the absolute difference in a new disparity image array.
8. Write the image array and display the disparity image.

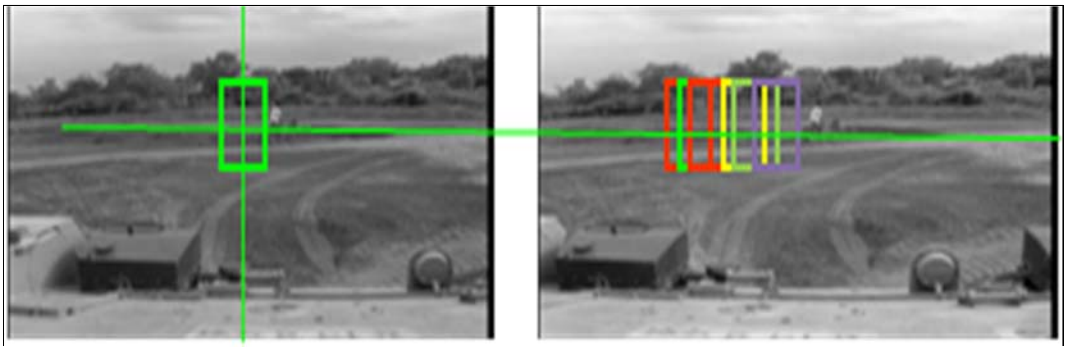


Fig 6: Linear Search

3.6 Multi-Resolution

Decomposing the images into a set of spatial frequency band pass component images is called multi resolution analysis. This is achieved by decomposing the image into a set of spatial frequency band pass component images. Individual samples of a component image represent image pattern information that is appropriately localized, while the band passed image as a whole represents information about a particular fineness of detail or scale. There is evidence that the human visual system uses such a representation, and multi resolution schemes are becoming increasingly popular in machine vision and in image processing in general.

The importance of analyzing images at many scales arises from the nature of image themselves. Scenes in the world contain objects of many sizes, and these objects contain features of many sizes. Moreover, objects can be at various distances from the viewer. As a result, any analysis procedure that is applied only at a single scale may miss information at other scales. The solution is to carry out analyses at all scales simultaneously.

3.7 Gaussian Pyramids

The bottom, or zero level of the pyramid, G0, is equal to the original image. This is low pass- filtered and sub sampled by a factor of two to obtain the next pyramid level, G1. G1 is then filtered in the same way and sub sampled to obtain G2. Further repetitions of the filter/subsample steps generate the remaining pyramid levels. To be precise, the levels of the pyramid are obtained iteratively as follows. For $0 < l < N$.

However, it is convenient to refer to this process as a standard REDUCE operation, and simply write

G_l = REDUCE [G_{l-1}]

The weighting function w (m, n) is called the "generating kernel." For reasons of computational efficiency, this should be small and separable. A five-tap filter was used to generate the pyramid.

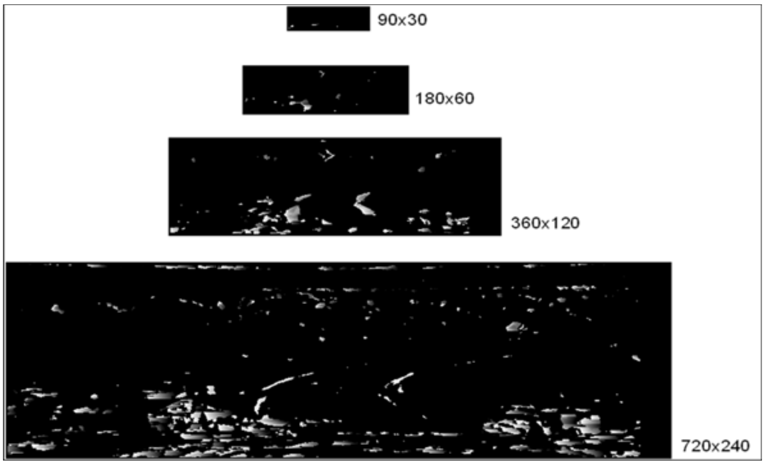


Fig 7: Gaussian Pyramid

Pyramid construction is equivalent to convolving the original image with a set of Gaussian-like weighting functions. Note that the functions double in width with each level. The convolution acts as a low pass filter with the band limit reduced correspondingly by one octave with each level. Because of this resemblance to the Gaussian density function we refer to the pyramid of low pass images as the "Gaussian pyramid". Band pass, rather than low pass, images are required for many purposes. These may be obtained by subtracting each Gaussian (low pass) pyramid level from the next lower level in the pyramid. Because these levels differ in their sample density it is necessary to interpolate new sample values between those in a given level before that level is subtracted from the next-lower level. Interpolation can be achieved by reversing the

REDUCE process. This is called as EXPAND operation. Let $G_{l,k}$ be the image obtained by expanding G_l k times.
Then G $_{l,k}$ = EXPAND [G Gl,k-1] or, to be precise, $G_{l,0} = G_l$, and for k>0,

$$G_{l,k}(i,j) = 4 \sum_{m} \sum_{n} G_{l,k\text{-}1} \left(\frac{2i+m}{2}, \frac{2j+n}{2} \right)$$

Here, only the terms for which (2i+m)/2 and (2j+n)/2 are integers contribute to the sum. The expand operation doubles the size of the image with each iteration, so that G $_{l,1}$, is the size of G $_{l,1}$, and G $_{l,1}$ is the same size as that of the original image.

3.8 Laplacian Pyramids

The levels of the band pass pyramid, L0, L1... LN, may now be specified in terms of the low pass pyramid levels as follows:
$L_l = G_l$-EXPAND [G_l+1] = G_l-G_l + 1
Gaussian pyramid could have been obtained directly by convolving a Gaussian-like equivalent weighting function with the original image and each value of this band pass pyramid could be obtained by convolving a difference of two Gaussians with the original image. These functions closely resemble the Laplacian operators commonly used in image processing. For this reason to the band pass pyramid is referred to as a "Laplacian pyramid”.

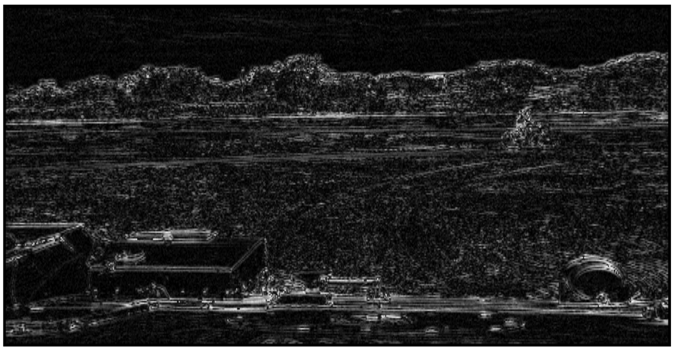


Fig 8: Laplacian pyramid

An important property of the Laplacian pyramid is that it is a complete image representation: the steps used to construct the pyramid may be reversed to recover the original image exactly. The top pyramid level, L_N, is first expanded and added to L_{N-1} to form G_{N-1} then this array is expanded and added to L_{N-2} to recover G_{N-2}, and so on. Alternatively, we may write $G_0 = L_{i,j}$. The pyramid has been introduced here as a data structure for supporting scaled image analysis. The same structure is well suited for a variety of other image processing tasks.

4. Autonomous navigation

On finding the range the autonomous navigation of the vehicle is performed based on the instructions given by the microcontroller.

4.1 Range Estimation

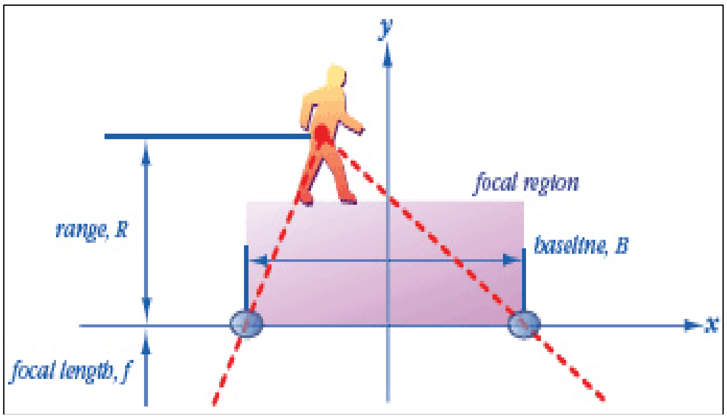


Fig 9: Range estimation

R = (f*B)/d = (f*B)/ (X_L-X_R)

Where,
R => range
B => baseline
F => focal length
D => disparity, (XL-XR)

4.2 Unmanned Ground Vehicles

Unmanned ground vehicles (UGV) are robotic platforms that are used as an extension of human capability. This type of robot is generally capable of operating outdoors and over a wide variety of terrain, functioning in place of humans.

4.3 Navigation

The navigation of the vehicle is performed as per the instructions were given to the microcontroller. The action to be performed is fed into the microcontroller for the various inputs perceived from the environment. The micro controller is interfaced with the system and the stereo cameras for choosing the navigation path.

5. Implementation

5.1 Bitmap Files

A bitmap is a file in which an image is stored in the specific format. The images obtained from the camera are stored in the memory in the form of bitmaps. The bitmap images are processed to produce a disparity map. The generated outputs disparity map and the range map are stored in the memory in the bitmap format.

5.2 WPF Application

A WPF application is a window based project. The source files are created implicitly on creating the workspace. The files are integrated and any modification of code is performed in the functions of the class files. It runs as windows based service watching the image files in the workspace location. As soon as the images are captured and saved by the camera in the workspace location the service picks up the files for processing and generates the disparity map.

6. Results

6.1 Disparity Map

The disparity map is generated which informs the system about the existence of an obstacle. Obstacle identification and classification is done using disparity map.

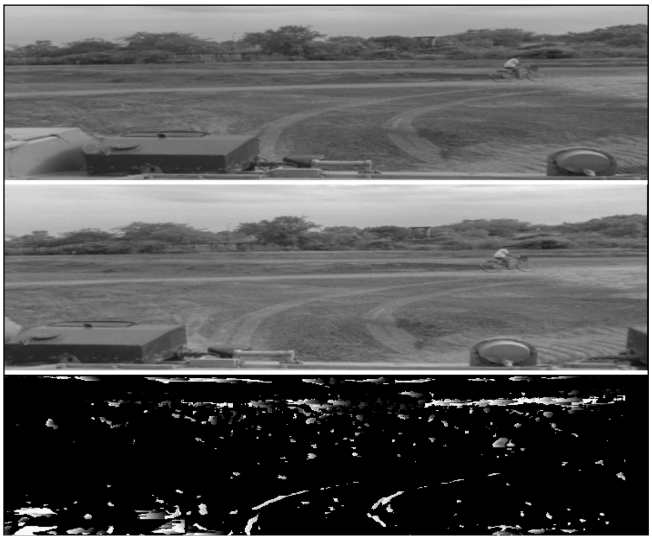


Fig 10: Disparity Map

6.2 Multi-Resolution Disparity Map

The multi-resolution disparity map is generated using the Gaussian pyramid. The pyramid is constructed. The disparity maps are generated by creating resized image array. The classification of the obstacle is performed using the disparity maps generated. The farther images are detected using disparity map of lower range and the nearer objects are easily identified using the disparity map of higher range. The generation of multi-resolution disparity map minimizes the search range and makes the obstacle detection task simpler.

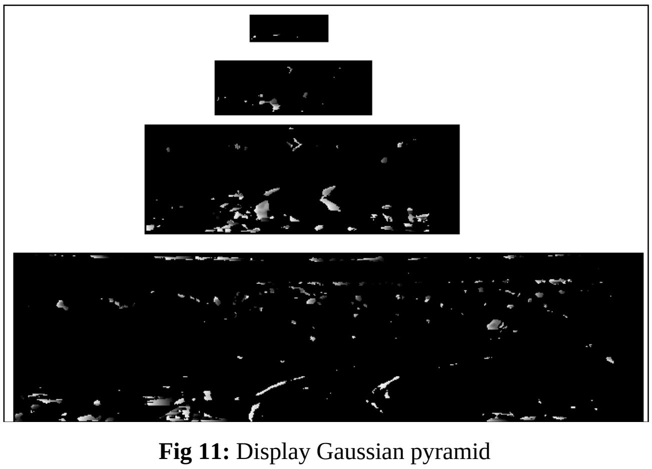


Fig 11: Display Gaussian pyramid

6.3 Range Map

Range map generated gives the approximate range of the obstacle

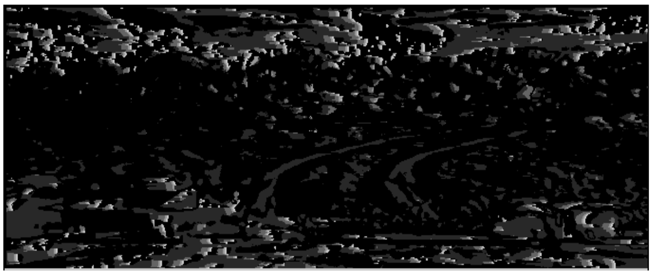


Fig 12: Range Map

6.4 Sample Range Calculation

Range calculation formula is
R = (f*B)/d

Assuming d=12,
R = (887*0.176)/12
= 13m

An obstacle at the range between 5 to 52 m could be detected.

6.5 Advantages

The advantages of using stereo vision system are

1. Highly reliable and effective.
2. Stereo cameras are a passive sensor.
3. The system can be easily integrated with other routines and hardware.
4. Economic when compared with RADAR systems.
5. No harmful radiation.
6. Uses natural energy source

6.7 Future Enhancements

There are possibilities for much advancement in the future.

1. The multi-resolution disparity image is fused to get the enhance disparity image.
2. The system can be integrated with GPS tracking device which autonomously drives the vehicle from the source to destination.
3. The set up can be modified to adjust its baseline automatically to identify the closer and longer obstacle.

6.8 Summary

A range map is generated using Sum of Absolute Differences (SAD) algorithm and the obstacle is identified efficiently. Stereo vision based obstacle detection system is a real-time application which could be applied to the autonomous navigation of transportation vehicle.

Graphics processing units (GPUs) offer higher peak performance than CPUs, but for a limited problem space. Even within this space, GPU solutions are often restricted to a set of specific problem instances or offer greatly varying performance for slightly different parameters.

7. References

1. Zhilenkov. The study of the process of the development of marine robotics, Vibroengineering Procedia. 2016; 8:17-21.
2. Scharstein D, Hirschmuller H, Kitajima Y, Krathwohl G, Nesi Nc, Wang X *et al*. High-resolution stereo datasets with the sub pixel-accurate ground truth. In German Conference on Pattern Recognition, Springer, 2014, 31-42.

3. Ryan Fanello S, Rhemann C, Tankovich V, Kowdle A, Orts Escolano S, Kim D *et al*. Hyperdepth: Learning depth from structured light without matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 5441-5450.
4. Shahbazi M, Sohn G, The J, end PM. Revisiting Intrinsic Curves for Efficient Dense Stereo Matching. ISPRS Annals of Photogrammetry, Remote Sensing, and Spatial Information Sciences, June, 2016, 123-130.
5. Oskam A, Hornung H, Bowles K, Mitchell M, Gross OSCAM. Optimized Stereoscopic Camera Control for Interactive 3D. ACM Trans Graph (SIGGRAPH). 2011; 30(189):1-8.
6. Tzung-Han Lin, Shang-Jen Hu. Perceived Depth Analysis for View Navigation of Stereoscopic Three-Dimensional Models. Journal of Electronic Imaging, Jul! Aug. 2014; 23(4):1-12.
7. Lisitsa D, Zhilenkov A. Comparative analysis of the classical and nonclassical artificial neural networks, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017.
8. Lisitsa D, Zhilenkov A. Prospects for the development and application of spiking neural networks, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017.
9. Zhilenkov, Denk D. Based on MEMS sensors man-machine interface for mechatronic objects control, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017.