Artificial Intelligence with Stereo Vision Algorithms and its Methods

Sahil S.Thakare[#1], Rupesh P. Arbal[#2], Makarand R. Shahade[*3]

[#1,2]Student, Department of IT, Jawaharlal Darda Institute of Engineering & Technology, Yavatmal (MS) INDIA

[*]Third Author

[*]Asst. Professor, Department of IT, Jawaharlal Darda Institute of Engineering & Technology, Yavatmal (MS) INDIA

ABSTRACT

Stereo vision is the process which is similar to the human being vision i.e. stereopsis. As we all know that stereo vision is field of computer vision and this is related to artificial intelligence field the stereo vision can be used in many application in field of artificial intelligence like in video cameras for security purpose and in robot this requires much things to do like image segmentation and the motion detection and to calculate distance of any object so stereo vision can be made in use for this methods thus this paper contains the algorithm for detecting motion and also smoothing filters to make output of two images smooth and to measure distance of any object i.e. blob detection the stereo ranging method is given i.e. novel algorithm. We will use novel algorithm for efficient motion detection and tracking of blob. It involves dividing video into image frames, converting frames into gray scale images, subtraction and thresholding of image frames, blob detection and combining blobs. As we all know that speed is also the important factor while making the motion detection the paralleling algorithm is explained for it. This paper also gives relationship of human vision with stereo vision and as we use to cameras in stereo vision there are problems like correspondence problem, camera calibration are explained and the feature and correlation based methods to solve this problems is given and the triangulation method is also made in use to compute depth.

Keywords

Stereo vision; parallel algorithm; correspondence; camera calibration; triangulation method; blob detection

1. INTRODUCTION

Stereo vision refers to the ability of inferring information on the 3D structure and the distance of a scene from two or more images taken from different viewpoints. Stereo vision is the process which is similar to the human process of eye vision i.e. known as stereopsis. The term Stereopsis(stereo- meaning "solid" or "three dimensional",, and opsis means view or sight) is the process in visual perception leading to the perception of depth from the two slightly different projections of the world onto the retinas of two eyes[1].

The human sight is considered as the most important sensing factor, nowadays we are looking around to use computer vision to enable systems to better see, interpret and respond to real world event in real time. Computer vision (or machine vision) is the science and technology of machines that see. Here see means the machine is able to extract information from an image, to solve some task, or perhaps "understand" the scene in either a broad or limited sense [2].The contradiction between human stereo vision and computer stereo vision is almost the same. That's why stereo vision can be presented by computer. but they are not totally the same. Because the computer do not know what people 's brain do. So it is necessary to "tell" the computer, which methods human used.

Computer stereo vision is the extraction of 3D information from digital images, such as obtained by a CCD Camera [3]. By comparing information about a scene from two vantage points, 3D information can be extracted by examination of the relative positions of objects in the two panels. There are two primary factors to generate a 3D stereo vision, convergence and parallax. In real camera several steps are to be considered such as barel distortion, image rectification and disparity map. This is similar to the biological process stereopsis.

The concept of stereo vision is widely used in field of artificial intelligence also i.e. we will mainly use it in detecting real time motion and to detect parameters regarding the detected object like its size shape mainly. For example, stereo vision can be used in robots, the robot can use their stereo vision to detect the distance between itself and the target object. Stereo vision can be used in the security system, to join two or more images together, so that can increase the range of vision. In the another hand, how to build stereo vision in realtime becomes a big problem because it faces many difficulties like image segmentation and filtering the images taken. Whatever in the case of robot or the case of security system, speed becomes more and more important. With paralleling the algorithms for stereo vision, we can get 2-3 times even 20 times speed up and in robot use the blob detection is required so stereo ranging technique gives the distance of the object.

2. BACKGROUND

Around the year 1600, Giovanni Battista della Porta produced the first artificial 3-D drawing based on Euclid's notions on how 3-D perception by humans works. This was followed in 1611 when Kepler's *Dioptrice* was published which included a detailed description of the projection theory of human stereo vision i.e. stereopsis. The applications that have driven the development of computer stereo vision have varied greatly since its inception. The first major use was for mapping the topography of the land by performing calculation disparity in satellite imagery (Barnard & Fischler 1982). Stereo vision later saw applications in human motion capture and allowed a computer to better animate humanoid models by capturing 3D human motion.

For the biometric authentication face recognition is used mostly i.e to detect the face which has been kept as password for any security reason in field of artificial intelligence. Earlier the 2-D images technique was used but the two dimensional (2-D) images are affected strongly by variation in pose and illumination. Recently the use of 3-D information has gained much attention, since 3-D data is not affected by translation and rotation and is

immune to the effect of illumination variation. Automatic face recognition has wide areas of application, including human-machine interaction, personalization of devises, data encryption, security, virtual reality, computer games, surveillance systems, or electronic commerce [4].

Within the past decade gesture, fingertip and arm tracking have reached a point where real time interactions are possible. Various applications have since developed. These range from using 3D arm tracking as input to a robotic arm, achieving an uninterrupted and natural remote interface, to using stereo vision for rehabilitation and human motion analysis.

3. ANALYSIS

As we are using two cameras it is necessary that both the cameras should work in coordination. But as we are using the number of cameras and capturing images of many objects there may be problem in matching image frames taken from multiple cameras. This problem in stereo vision is known as correspondence problem. Also in stereo vision we need to recover the depth information of any 3D structure , it can be recovered with two images and triangulation method. And as there is a need of speed in stereo vision for implementing it in realtime there is need to parallel it and so parallel method regarding hardware i.e. cluster architecture is made in use. There are some basic steps to get the stereo vision in computer vision: Set up the environment, Calibrate on the camera, finding correspondence, compute depth, and there are also some algorithms which are made in use in computer vision field for motion detection and distance of blob can be calculated by novel algorithm .

3.1 Environment

As we are using two cameras it is necessary that both the cameras should work in coordination. But as we are using the number of cameras and capturing images of many objects there may be problem in matching image frames taken from multiple cameras. This problem in stereo vision are known as correspondence problem and reconstruction problems. Also in stereo vision we need to recover the depth information of any 3D structure, it can be recovered with two images and triangulation method. And as there is a need of speed in stereo vision for implementing it in realtime there is need to parallel it and so parallel algorithm is made in use. There are some basic steps to get the stereo vision in computer vision: Set up the environment, Calibrate on the camera, finding correspondence, reconstruction problem and compute depth.

To get the stereo vision, it's necessary to have two cameras and one computer. Instead of using a single camera two cameras are more beneficial. Benefits are as it helps distinguish interactive parts of captured images and more accurate and reliable 3D position data and also some variables of stereo vision can be accurately calculated.

3.2 Calibrate on the Camera

The common method to calibrate a camera is to take a image of a chessboard then do Canny edge detection , after the detection, fitting the straight line to detect linked edges, then intersecting the lines to obtain the image corners, match image corners and 3D target chessboard corners which will be visible in image of chess board. Then we get the pairs of match point [5].

3.3 Finding Correspondence

As we are using two cameras it is necessary that both the cameras should work at once and both images should contain same matching points. This problem in stereo vision are known as correspondence problem. In correspondence problem there is one basic problem in searching elements, i.e.element on left image should search for corresponding element in right image. For this elements chosen should be matching with each other and there should be similarity measure to compare those elements. Depth is inversely proportional to disparity.

There are two classes of algorithms proposed that solve the correspondence problem.

3.3.1 Correlation based algorithm

This produces the dense set of correspondences. For example image in Figure 1.

3.3.2 Feature based algorithm

This produce the sparse set of correspondences. This feature based algorithm is conceptually similar to the correlation based method in this they only search for the correspondences of a sparse set of image features in fig 2.

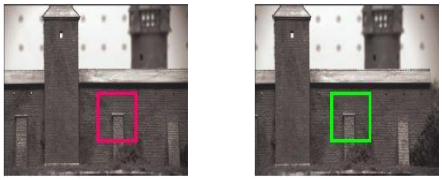

Fig 1. Correlation matching points

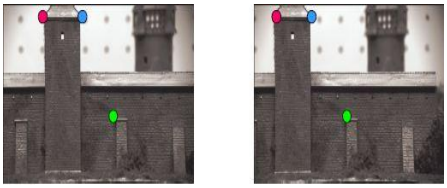

Fig 2. Feature based matching points

Features commonly used in this are regarding corners in this similarity should be in terms of surrounding gray values and there locations. Secondly, regarding edges and lines in this similarity measured should be in terms of orientation, contrast, length of lines. Thus we conclude that, the correlation based method requires textured images as they consider the dense points they are good for surface reconstruction they are sensitive to illumination variations but are inadequate for very different viewpoints. Whereas, in the featured based method they require prior knowledge of type of scence and has to find the features first but as they consider sparse points they are good for navigation.

3.4 Compute Depth (Triangulation)

The methodology used for compute depth is show as below and we can calculate depth from below fig 3 and in [6].

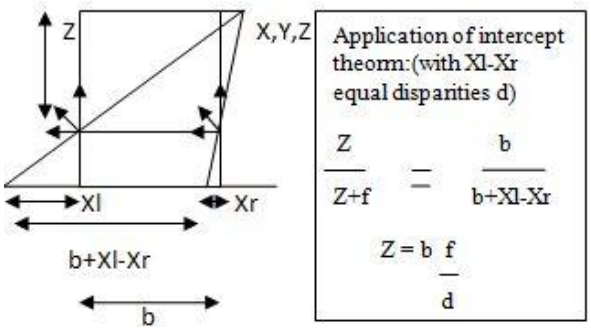

Fig. 3 Compute depth

4. ALGORITHMS USED FOR MOTION DETECTION

There are number of algorithms have been proposed keeping particular applications in mind. The algorithm explained here is the frame differencing algorithm .Following given are the popular algorithms used in computer vision field for motion detection.

4.1 Two Frames Difference Motion Detector

This type of motion detector is the simplest one and the quickest one. The idea of this detector is based on finding amount of difference in two consequent frames of video stream. The greater is difference, the greater is motion level. As it can be seen from the picture below, it does not suite very well those tasks, where it is required to precisely highlight moving object. However it has recommended itself very well for those tasks, which just require motion detection. The advantages of this method is that its easy to implement and faster than other methods and as we use previous frame as reference frame the background model changes over time there are some disadvantages because accuracy of frames depends on object speed and frame rate and there is one global threshold Th for all pixels in the image threshold Th is not a function of time. So this approach will not give good results in following conditions like if the background is not static, if the scene contains slowly moving objects, if the objects are fast and frame rate is slow, if general lighting conditions in the scene change with time.

4.2 Simple Background Modeling Motion Detector

In contrast to the above mentioned motion detector, this motion detector is based on finding difference between current video frame and a frame representing background. This motion detector tries to use simple techniques of modeling scene's background and updating it through time to get into account scene's changes. The background modeling feature of this motion detector gives the ability of more precise highlighting of motion regions. The model presented by Stauffer and Grimson [10].The advantages of this is a different threshold is selected for each pixel the objects are allowed to become a part of background without destroying the existing background model and it provides fast recovery. The only thing is this complexity gets higher.

4.3 Custom Frames Difference Motion Detector

This class implements motion detection algorithm, which is based on the difference of current video frame with predefined background frame, which puts it in-between of the two above classes. On the one hand this motion detector is based on simple differencing as the two frames difference motion detector, which makes it fast. On the other hand it does differencing of current video frame with background frame, which may allow finding moving objects, but not areas of changes (like in simple background modeling motion detector). However, user needs to specify background frame on his own (or the algorithm will take first video frame as a background frame) and the algorithm will never try to update it, which means no adaptation to scene changes. In all above three algorithms one thing common is that algorithm involves two frames, one for reference and another one is current frame. Reference frame is nothing but the frame chosen as a background which will be subtracted from current frame. So the equation from Frame Difference Motion Detection becomes

$$| frame_i – background_i | > Th$$

We will follow the first approach with some modification because, its the quickest one. Secondly, we require only motion detection of an object. Other algorithms are useful in applications like video surveillance where accuracy is more important. So for our approach the above equation can be rewritten as

$$| frame_{current} – frame_{previous} | > Th$$

But, this method doesnt give sufficient accuracy as discussed above. It also includes noise and hence requires some image processing before applying algorithm. Various filters can be used for minimizing noise and distortion of image and improving the quality of image. Then such image can be used for successfully motion detection purpose. Below some smoothing filters are presented for improving image quality.

4.4 Smoothing Filters

The smoothing filters play very important role in increasing efficiency of motion detection. Various smoothing filters mean, median, conservative smoothing and adaptive smoothing[11] are made in use but as mean filter is the best in them we have discussed it. This algorithms vary in efficiency in providing smoothnes in captured images.

4.4.1 Mean Filter

The median filter is normally used to reduce noise in an image, somewhat like the mean filter. However, it often does a better job than the mean filter of preserving useful detail in the image. The filter replaces each pixel of the original with the median of neighboring pixel values results in fig 4.

Fig 4. Mean filter results

5. PARALLEL ALGORITHM

Nowadays, stereo vision is not only use in mixing two images together but also used in many areas. For example, we use stereo vision on a robot, to simulate what human can see by their eyes. With this vision, robot can do some tasks for human. Such as pick up one object in the labyrinth, clean the inside of the pipe in some dangerous places. Speed is very important aspect for stereo vision process in realtime. It is necessary to parallel it. There are two ways to speed up the realtime stereo vision, the first way is use some hardware to speed up. The other way is to parallel the algorithms for stereo vision.

5.1 Hardware

There are some existing hardware architectures in the world. One of the hardware architectures is created [7].

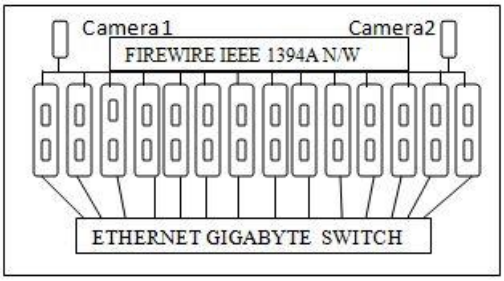

Fig. 5 Cluster architecture

The Figure 5 shows the cluster architecture. It includes 14 computing nodes. Each node has a dual processor Apple G5 running at 2GHz and 1Gb memory. All nodes connected to a gigabit Ethernet switch.The video streams is provided by two cameras via the FIREWIRE IEEE1394 NETWORK. It has a maximum speed at 400Mb per second.

The cluster is doing the job what people's brain do. The camera can capture images and provide the video streams enough for the processing. After each node received the images, they can work on the image, get the result and then send back to the master node. The result will present to the user on the output drive such as screen or printer.

On the other word, it must wait until each node finished the task on the current frame. In OpenMPI library we can use MPI_Barrier to make sure all the nodes received the images and work on the right images.From the result of the cluster Architecture as if the task process by using single processor. It will take 992.2ms. With 2 nodes it return the result 479.2ms. And it get the 38.2ms when using 28 nodes. That means it almost got 30 times speed up. A normal video got 30fps, and with the 28 nodes on processing the task, it almost got the quality of the realtime video.

5.2 Software

Paralleling in programming takes a very important part in stereo vision. Here are some of the example codes which are working on the Cluster Architecture [8],[9].

The example code received the frame from the cameras via FIREWIRE Network, create a task on the cluster world then process the task on the cluster. After the node finish processing the data, the nodes will send back the data to the root which means the master, and the master will output the data.

Thus we can conclude that parallel algorithms for stereo vision is based on the stereo vision algorithms. It does the same thing what stereo vision algorithms do. Only difference is we will separate the whole task into some small independence task in the master processor, send those small tasks to each slave node and process in the node, get the result back from the slave node, then join them together. Build up the hardware is quite expensive, so parallel the algorithms is a good way to speed up.

6. NOVEL ALGORITHM

We will use novel algorithm for efficient motion detection and tracking of blob. It involves dividing video into image frames, converting frames into gray scale images, subtraction and thresholding of image frames, blob detection and combining blobs. To calculate the distance of blob we will use stereo ranging method.

6.1 Divide Video into Image Frames

Video technology is used for electronically capturing, recording, processing, storing, transmitting and reconstructing a sequence for still images representing scenes in motion. The number of still pictures per unit of time of video ranges from six or eight frames per second (frame/s) for old mechanical cameras to 120 or more frames per second for new professional cameras.

6.2 Converting Image into gray Scale

As we know that an image is comprised of millions of pixels and each pixel consists of 3 colours Red, Green and Blue. Any colour can be obtained by just adjusting these 3 colours. A hex triplet is a six-digit, three- byte hexadecimal number used in HTML, CSS, SVG and other computing applications, to represent colours [12]. The bytes represent the red, green and blue components of the colours. One byte represents a number in the range 00 to FF (in hexadecimal notation), or 0 to 255 in decimal notation. This represents the least (0) to the most (255) intensity of each of the colour components.

Red=Pixel value of red colour in decimal
Green=Pixel value of green colour in decimal
Blue=Pixel value of blue colour in decimal
Gray value = (Red + Green + Blue) / 3
Image Pixels (Blue, x, y) = Gray value
Image Pixels (Green, x, y) = Gray value
Image Pixels (Red, x, y) = Gray value

Where, Image Pixels (Colour, x-coordinate, y-coordinate) is an array used to store the RGB values of each pixel separately. 'x' and 'y' are the image coordinates, the maximum value of 'x' can goes up to the width of the image and maximum value of 'y' goes up to height of the image.

6.3 Subtraction and Thresholding of Image Frames

In this step RGB value of each pixel of one image from two successive images is subtracted from RGB values of the previous image.

Red = ImagePixels1 (Red, i, j) – ImagePixels2 (Red, i, j)
Green= ImagePixels1 (Green, i, j) – ImagePixels2 (Green, i, j)
Blue = ImagePixels1 (Blue, i, j) – ImagePixels2 (Blue, i, j)

Where Red, Green and Blue are the resultant colours after subtracting 2 images. ImagePixels1() and ImagePixelsl2() are arrays of image frame 1 and image Frame 2 respectively, which are used to store the RGB values of each pixel separately. In the resultant image (which is obtained after subtracting the image frames) objects which are in motion are highlighted in gray shades and rest all is black out.

6.4 Blob Detection

Blobs are compact objects of approximately the same intensity. The image contains number of white spot which are known as blobs. It is the most important step in motion detection. The white spot represent the moving bodies in the image frames where each white pixel is counted in the image and the pixels which are connected to each other are considered as a part of one object. This is useful to differentiate between various objects. Whole

image is scanned pixel and a unique label is assigned to each object (a group of connected components i.e. to each white pixels encountered).These labels are the unique values used to distinguish different objects. In an image different objects are classified by identifying groups of similar pixels.

6.5 Stereo ranging

Calculating the distance to objects by making a pair of observations at different location. For stereo vision calculating the range of an observed object or feature we use following formula.

Range = (Focal length x Camera baseline) / Disparity

Where the disparity is the horizontal difference in pixels between the position of the feature in left and right images, the focal length is also expressed in pixels and the baseline distance is in millimeters.

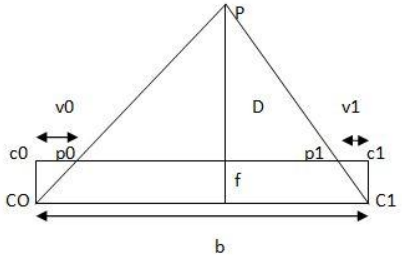

Fig 6.Certainity problem

C0,C1 - Left Camera , Right Camera
P -Observed feature point
F -focal length, B -baseline distance
D -distance to observed feature point
c0, c1 -Pixel center of camera images
v0, p1 -pixel position of observed feature point
v0, v1 -Pixel displacement of observed feature point
Disparity (D) = v1-v0
Distance (D) = bf/d

6.5.1 Uncertainity problem

Since there is uncertainty associated with the disparity measurement (which can be up to plus or minus half a pixel) the actual range lies within a spatial probability distribution, which can be represented using a sensor model.

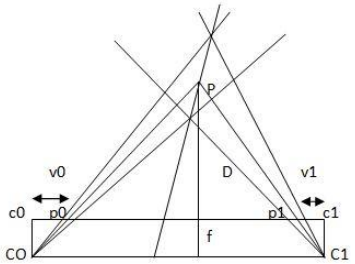

Fig 7.Uncertainity Problem

7. CONCLUSION

This paper explains the stereo vision and its relationship with the human vision and also its algorithms for using it in many field of the artificial intelligence. There are also basic steps to get the stereo vision in computer vision and its problems and also the triangulation method to compute depth is given in this paper. As the speed is an important factor for making the use of stereo vision in the real time the parallel algorithm and its architecture has been discussed in this paper. The algorithms for motion detections are also studied and the smoothing filters are made in use to get good output and the stereo ranging method i.e. novel algorithm is explained for distance calculation of the blob i.e.object.

8. REFERENCES

[1] Retrieved 15 MAY 2009. "Stereopsis," in *Wikipedia,The free Encyclopedia*, Available: http://en.wikipedia.org/wiki/Stereo_vision

[2] Dana H. Ballard and Christopher M. Brown (1982). "Computer Vision." *Prentice.Hall.ISBN*

[3] Bradski, Gary and Kaehler, Adrian. "Learning OpenCV: Computer Vision with the OpenCV Library". *O'Reilly.*

[4] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," *UMD CfaR Technical Report CAR-TR-948, 2000*

[5] R. Y. Tsai,"A Versatile Camera Calibration Technique for 3D Machine Vision", *IEEE J. Robotics & Automation, RA3, No. 4, August 1987, pp. 323344*

[6] Klaus D.Toennies., "5.Stereo Vision (Introduction)," *in 3D Computer Vision, University Magdeburg*

[7] J. Falcou, J. Serot, T. Chateau, F. Jurie, "A Parallel Implementation of a 3D Reconstruction Algorithm for RealTime Vision." *in Parallel Computing 2005.*

[8] R.Hartley and A. Zisserman, "Multiple View Geometry," in *Computer Vision*,Cambridge University Press, 2000, pp. 138183

[9] O. Faugeras, "ThreeDimensional Computer Vision: A Geometric Approach", MIT Press,1996, pp.3368

[10] C. Stauffer, W.E.L. Grimson, "Adaptive background mixture modelsfor real-time tracking" Proc. of CVPR 1999, pp. 246-252.

[11] Andrew Kirillov, Smoothing Filters (undated).[online]. Available:http://www.aforgenet.com/framework/features/smoothing_filters.html

[12] Mandeep Singh, "Improved Morphological Method in Motion Detection", *International Journal of Computer Applications(0975 – 8887) Volume 5– No.8,August 2010.*