An Enhanced Computer Vision Based Hand Movement Capturing System with Stereo Vision

Manaram Gnanasekera *manaramcv@gmail.com*
Department of Electrical and Computer Engineering
Sri Lanka Institute of Information Technology
Malabe, 10115, Sri Lanka

Nalan Karunanayake *nalan91karu@gmail.com*
Department of Computer Science
University of Sri Jayawardenapura
Nugegoda, 10250, Sri Lanka

Abstract

This framework is a hand movement capturing method which could be done in three different depth levels. The algorithm has the capability of capturing and identifying when the hand is moving up, down, right and left. From these captured movements four signals could be generated. Moreover, when these hand movements are done, 15cm-75cm, 75cm-100cm, 100cm-200cm from the camera (3 depth levels), twelve different signals could be generated. These generated signals could be used for applications such as game controlling (gaming).The existing method uses an object area based method for depth analysis. The results of the proposed work shows it has high accuracy compared to the existing method when tested for depth analysis.

Keywords: Computer Vision, Object tracking, Meanshift, Stereo Vision.

1. INTRODUCTION

Human body movement recognition has been one of the main interested research areas among researchers. Out of many applications of human body movement recognition, gaming is one of the most benefitted industries out of this technology which had gained huge amounts of success. Xbox 360 is one of the best examples for this fact.

Among various studies done on human body movement recognition, in [1] a hand movement capturing system based on the Camshift algorithm has been developed. There are six different hand movements, namely, hand moving up, down, right, left, back and forth which are captured by this algorithm. Unlike the other movements a special technique is used to capture hand moving back and forth. The increase of the area of the hand (in the computer screen) when moving towards the camera is used to track the forward movement. Similarly the area of the hand decreases (in the computer screen) when it goes away from the camera. In essence, the area of the hand is used to find the depth. Although this method has reached success in this research ([1]) the best method for depth analysis in computer vision is Stereo Vision. By motivated in this fact the work undertaken enhances the algorithm given in [1] by combining Stereo Vision. Moreover, instead of the forward and backward movements the proposed algorithm introduces a novel method to capture the up, down, left and right movements in three depth levels (15cm-75cm, 75cm-100cm and 100cm-200cm). Figure1 explains the scenario.

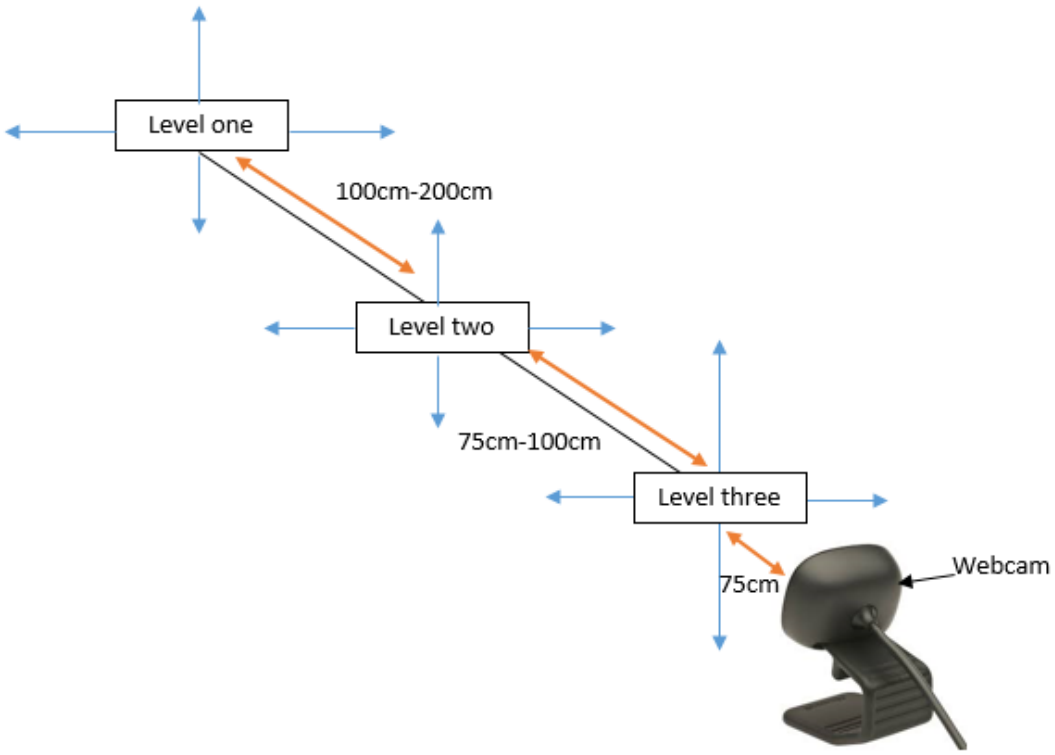

FIGURE 1: Four movements in three different depth levels.

All together there will be 12 different movements which could be captured. With the aid of these 12 different movements there will be 12 different signals generated. These generated signals could be used for different applications such as gaming.

In section 2 of the paper, the algorithms which are used to capture the movements of the hand is discussed. Then in section 3 the depth analyzing method is presented. Finally, in section 4 three experiments which were conducted in [1] are performed and hence, it is proved that the proposed method performs well in different depth levels compared to the area based method used in [1].

2. HAND MOVEMENT CAPTURING

The same approach as in [1] will be taken for the object recognition, which is a histogram based method. The histogram of the hand will be taken and a contour will be drawn around. This histogram will be unique for the hand (the environment around the hand will have different histograms). Therefore a contour could be perfectly drawn along the margin of the hand. After recognizing the object the next task will be to capture the movement of it. An extended version of the Meanshift tracker [2] named as Camshift [3] has been used for this task. Camshift or "Continuously Adaptive Meanshift" adjusts the search window size accordingly. At first the Camshift algorithm closes the initial location of the search window. Then the Meanshift operation is performed and the zeroth moment is stored. Finally, the search window size will set equal to a function of the stored zeroth moment. By the aid of the Camshift algorithm, up, down, right and left movement capturing has been tackled.

3. STEREO VISION DEPTH ANALYSIS

Stereo imaging is a technique which is used for estimating a 3D model to extract the depth information of the 2D coordinates of a 2D scene by taking two images using two cameras which are separated horizontally [4]. Furthermore, using the two 2D images can determine the three dimensional locations of the image points which corresponds to the exact physical points in space

[5]. When considering the two images from the left and right cameras, the horizontal points of the right image is lower than the horizontal coordinate points in the left image. This obtrusive shift is called disparity. The disparity obtained from the two images could be used to calculate the 3D location and the depth information of the particular points since the disparity is inversely proportional to the depth of the particular point [6].

3.1 Stereo Triangulation

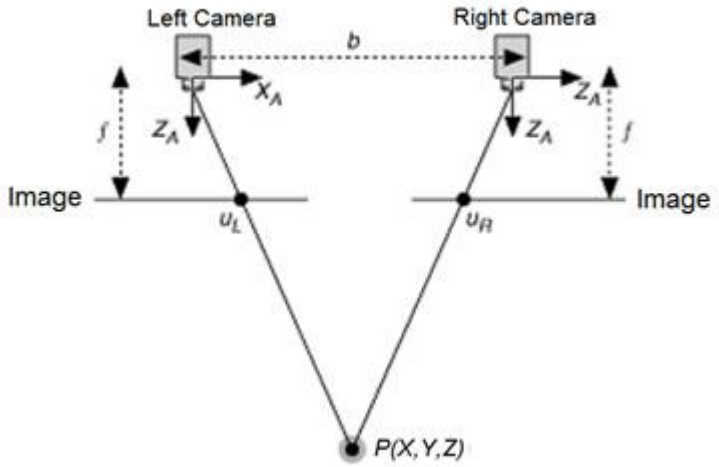

FIGURE 2: Stereo Camera model.

The disparity is,

$$d = U_L - U_R \quad (1)$$

Hence the depth of the point is,

$$Z = b * \frac{f}{d} \quad (2)$$

Where,
b is the baseline, or distance between the two cameras left and right.

f is the focal length of a cameras.

X_A is the X-axis of a camera.

Z_A is the optical axis of a camera.

P is the physical point defined by the coordinates X, Y, and Z.

U_L is the projection of the real-world physical point P in an image acquired by the left camera.

U_R is the projection of the real-world physical point P in an image acquired by the right camera

There are three stereo correspondence algorithms in the current literature as pixel based methods, area based methods and feature based methods [7, 8].

In the proposed method the sum of square differences (SSD) block matching method is used as the correspondence cost function due to it's simplicity than the other block matching cost functions [9].

The sum of square difference (SSD) function can be expressed as,

$$\sum_{x=1}^{x=n}\sum_{y=1}^{y=n} abs\left(I_{l(x,y)} - I_{r(x,y)}\right)^2 \quad 3$$

Where,

$I_{l(x,y)}$ and $I_{r(x,y)}$ are n x n symmetric matrices of the left and right images.

Before starting the stereo matching process, stereo calibration and stereo rectification should be done for better results [10]. Using the stereo vision technique the depth to the center of the captured hand could be calculated using (1). As a result the given three rangers (15cm-75cm, 75cm-100cm, 100cm-200cm) could be successfully identified.

4. RESULTS & DISCUSSION

The aim of the proposed work is to enhance the method given in [1] with stereo vision. Therefore some tests which were conducted in [1] will be performed again in order for comparison with [1]. The tracking accuracy has been taken as a percentage by considering the number of miss-tracks and correct tracks occurred when tested under different conditions.

4.1 Performance In Different Light Conditions

When the algorithm was tested under different light conditions (different LUX levels) the following results were observed.

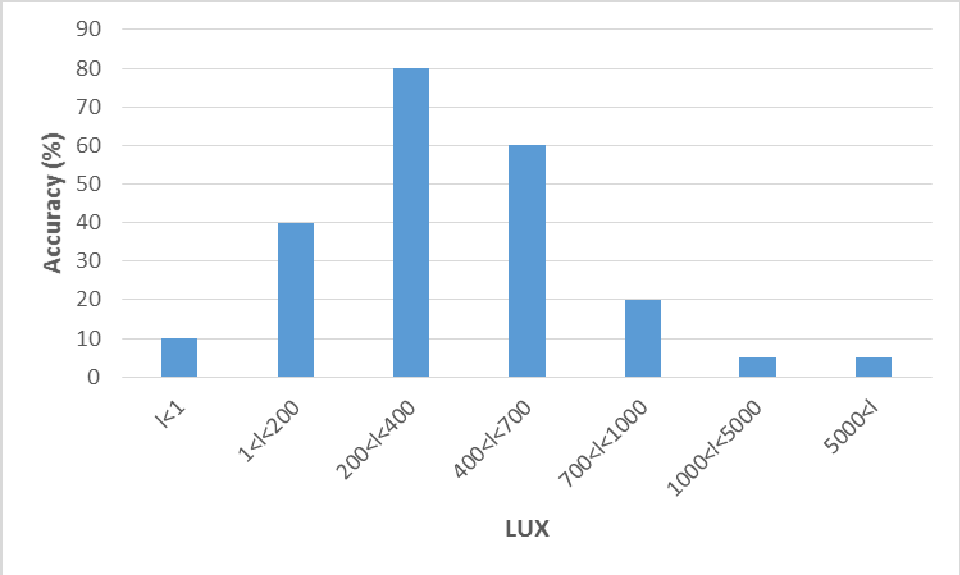

FIGURE 3: LUX level vs Accuracy.

The above observed results are very close to the light condition test's results in [1], which means the addition of stereo vision hasn't been able to increase the performance under different light conditions.

4.2 Usability for a Hand Amputated Person

Algorithm in [1] has extended it's usability for hand amputated users. A sock is inserted to the amputated arm and the sock is tracked instead of the hand. Therefore a test is done to check the accuracy of this method as well. The socked arm will be first kept parallel to the ground (perpendicular to the camera) and the accuracy is observed. Then the hand will be kept perpendicular to the ground (parallel to the camera) and the testing will be done. The same tests will be done for the proposed algorithm to check the usability.

- S.H.P.G – Socked Hand Parallel to the Ground
- S.H.Pe.G – Socked Hand Perpendicular to the Ground

The observed accuracy plots are as follows,

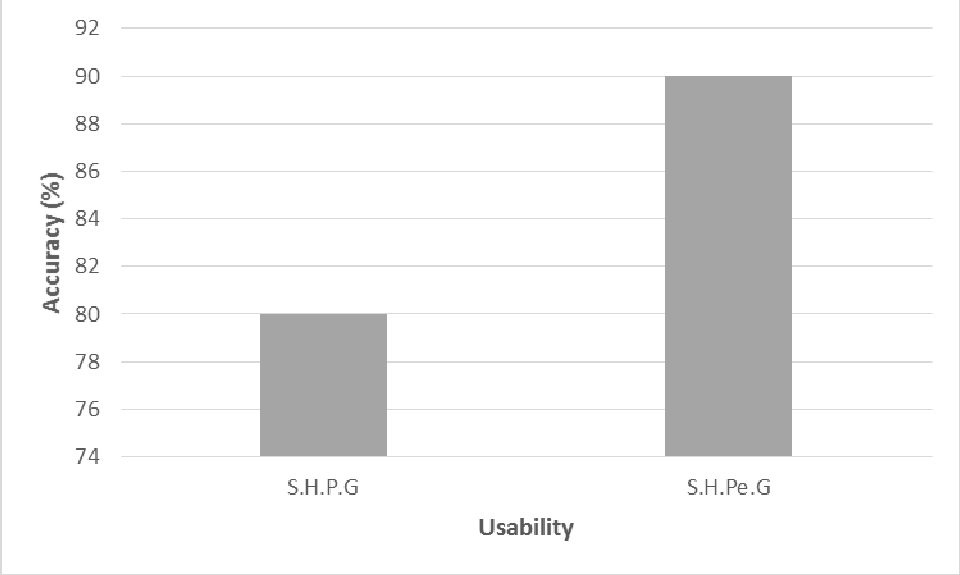

FIGURE 4: Usability vs Accuracy.

Similar to 4.1 the results of this test is almost equal to the usability test done in [1]. In essence, the stereo vision based method hasn't done any contribution towards developing the usability.

4.3 Distance of Tracking

The distance between the hand and the camera is varied or in other words, the hand is kept in different depth levels from the camera. This test is done to check the sensitiveness of the algorithm when the object (hand) moves away from the camera.

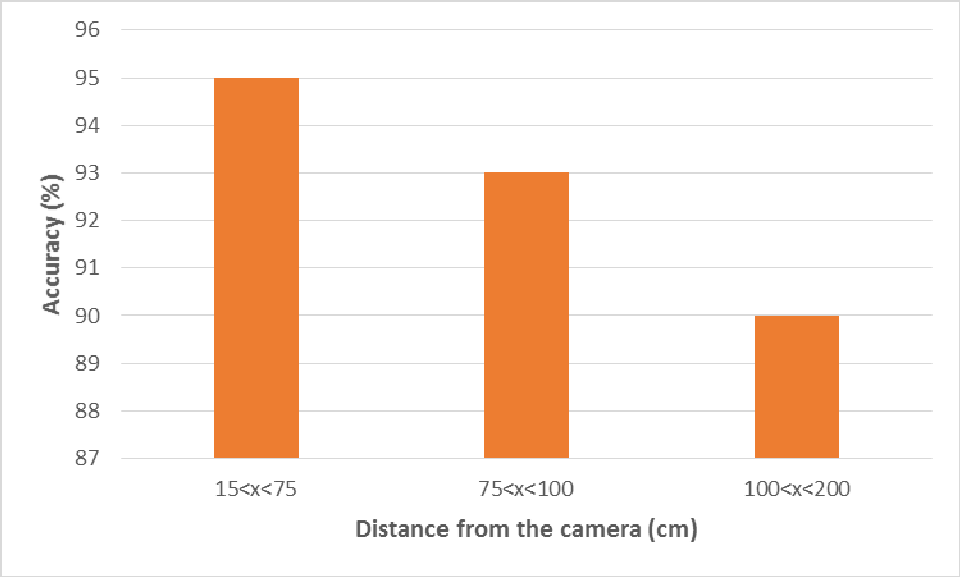

FIGURE 5: Depth vs Accuracy.

Although the tests conducted in section 4.1 and 4.2 are more likely to the corresponding tests done in [1], a vast improvement has been shown in the results of this experiment when the depth is increased. When the experiment is done between 15cm to 75cm the observed accuracy is about 95%. Whereas for the same experiment in [1] the accuracy is roughly around 90%. When it is done at 75cm to 100cm the accuracy of the proposed algorithm is around 93% and the accuracy in [1] is around 60%. When the distance is increased between 100cm to 200cm, an accuracy of 90% is still observed. However, a very low level of accuracy (20%) has been observed when the same test is done in [1]. Even though the distance between the hand and the camera is increased the proposed stereo vision based method has been able to maintain a tracking accuracy around 90% compared with the method proposed in [1]. The speciality is, even though the accuracy is significantly dropped in the method given in [1], a 90% accuracy is maintained by the proposed algorithm.

5. CONCLUSION

In conclusion, the work undertaken enhances the algorithm of "Computer Vision Based Hand Movement Capturing System" ([1]) by using Stereo Vision. The work presented proposes a hand movement capturing method which could be done in three depth levels. The captured hand movements are namely, moving up, down, right and left and this capturing could be done in 15cm-75cm, 75cm-100cm and 100cm-200cm from the camera. Camshift algorithm is used to track the basic movements of the hand. Three tests which were conducted in [1] were performed back for the purpose of comparison. The experiment done to investigate the performance in different light conditions had no difference between the proposed method and [1] when the results were compared. Similar results were observed when the algorithm was tested with an amputated hand. However, when the distance with the camera was varied (depth level) the proposed method has shown more than 90% accuracy. This is a vast improvement compared to the results observed in [1]. Finally there are six different movements captured in [1] which generates six different signals. The proposed method could generate twelve (twice) signals with the captured twelve movements.

6. FUTURE WORK

As future expansions this algorithm will be modified for a full body tracking in different depth levels. Furthermore it is also planned to increase the number of depth levels.

7. REFERENCES

[1] M. Gnanasekera, "Computer Vision Based Hand Movement Capturing System," in The 8th International Conference on Computer Science & Education (ICCSE 2013), Colombo, 2013.

[2] D. Comaniciu et al, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in Conference on Computer Vision and Pattern Recognition (CVPR), Hilton Head, 2000.

[3] G. R. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface," Microcomputer Research Lab, Santa Clara, 2002.

[4] Richard Szeliski. “Stereo Correspondence” in Computer Vision: Algorithms and Applications, 1st ed., vol. 1, Springer 2010, pp.537-571.

[5] Myron Z. Brown, Darius Burschka, “Advances in Computational Stereo” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp.993-1008, August, 2003.

[6] Samarath Brahmbhatt, “3D Geometry and Stereo Vision” in Practical OpenCV,1st ed., vol 1, Apress 2013, pp. 173-200.

[7] U. R Dhond., J. K. Aggarwal, “Structure from Stereo.” IEEE Transaction on Systems, Man and Cybernatics, vol. 19, pp. 1489-1510, November 1989.

[8] Vibin N. Valsan and C.Y Patil. “A System on Chip based Stereo Vision Approach for Disparity Measurement,” in Proc. 2015 International Conference on Industrial Instrumentation and Control (ICIC), 2015, pp. 1284-1287.

[9] Rong Xiang, Tao Hong, Ming Zhou. “Analysis of Depth Measurement Errors of Tomatoes Using Binocular Stereo Vision Based on Single Factor Experiments,” in Proc. 13th International Conference on Control, Automation, Robotics & Vision (ICARCV), December 2014, pp. 88-93.

[10] G. Balakrishnan, G. Sainarayanan, R. Nagarajan “Stereo Image Sonification for Blind Navigation”, Tamkang Journal of Science and Engineering, vol. 10, pp. 67_76, 2007.