Two Handed Gesture Recognition System based on Indian Sign Languages

Prof. Sumitra Sadhukhan[1] Mr. Aakash Ghodke[2] Mr. Sanket Kadam[3]
[1,2,3]Department of Computer Engineering
[1,2,3]Rajiv Gandhi Institute of Technology, Mumbai, India

Abstract— This paper is based primarily on the recognition of sign language in which deaf and dumb people can communicate with normal people. It contains image recognition, color segmentation, epipolar image and some technique for 3D image mapping. Also in this paper of Hand gesture recognition of Indian sign language software we come up with an advance system for human hand gesture recognition by 3D image recognition technique typically focuses on the idea to help deaf and dumb people to communicate with normal people. The extended potential of this system can be implemented at public places where deaf dumb people need to connect with normal people to convey there message or as simple as to order a coffee.

Keywords: Contour Detection, Epipolar Detection, Camera Calibration, ISL, 3D Modelling

I. INTRODUCTION

The language of signs is widely used by deaf-dumb people. The sign language recognition system is a system for the interaction of human machines in which machines help to understand the sign language. With the help of this system, deaf and dumb people communicate and socialize in stores, shops, etc. This system will attempt to bridge the gab by replacing the need for a third person as a translator and establishing direct communication between deaf-dumb people. Hand Gestures are natural way in which human convey or emphasize there normal communication. Whereas vast majority of people face with the problem of hearing loss or unable to speak refer as deaf dumb people has no choice but to communicate with the help of sign language. Indian Sign Language (ISL) is a dominant sign language in south Asia. Creating an system software in order to detect and recognize ISL so this deaf dumb people can communicate with normal people .To recognize Hand Gesture specifically in ISL which has two hand gesturing simultaneously creating thousands of different gestures is challenging but exciting task, using computer vision techniques and implementing 3d object detection, skin color segmentation ,depth detection and recognition and creating a learning library of commonly use gesture will enable deaf and dumb people to one-to-one communicate to normal people

II. COUNTER DETECTION

First, pre-processing images such as background image difference are given to the sign language gesture. The method of extracting edge and contour features is then introduced. Finally, the weighted features match method with the values of the sample image feature is given. In order to sign language gestures, create a library of samples and validate the method with many experiments. The results show that this method can efficiently classify 30 sign language gestures and reach 93 percent of the recognition rate. The main basis for image analysis and recognition is the geometric characteristics of the image. The method of recognition based on geometric characteristics can reduce the algorithm's complexity, save the required storage space and save time. Because binary images are easily accessible, stored and processed, a convenient extraction of the geometric image features usually requires a binary image process.

III. EPIPOLAR IMAGING

Epipolar geometry is stereo vision geometry. When two cameras view a 3D scene from two different positions, there are a number of geometric relationships between the 3D points and their projections onto the 2D images, which lead to limitations between the points of the image.

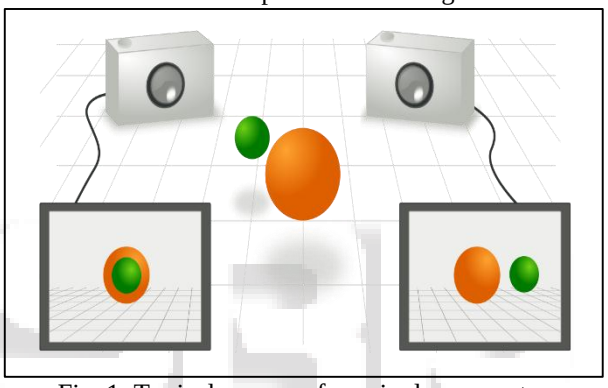

Fig. 1: Typical use case for epipolar geometry

These relationships are derived on the assumption that the pinhole camera model can approximate the cameras. As in fig 1 The image plane is actually behind the focal center in real cameras and produces an image that symmetries the focal center of the lens. However, the problem is simplified here by placing a virtual image plane in front of the focal centre, i.e. the optical center of each camera lens to produce an image that is not transformed by the symmetry.

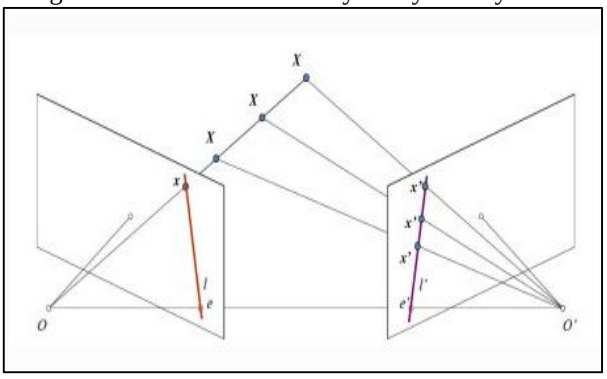

Fig. 2: Epipolar imaging

In fig 2. OL and OR represent the two cameras lens symmetry centers. In both cameras, X is the point of interest. Points xL and xR are the image planes projected by point X.

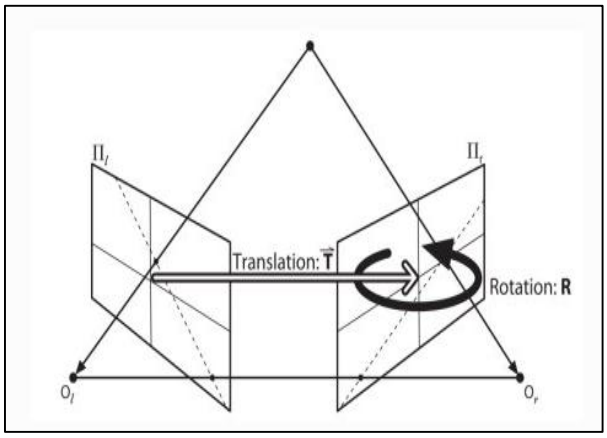

Fig. 3: Translation and rotation.

IV. CAMERA CALLIBERATION AND IMAGING

Pinhole camera introduces lots of distortion in the image to work with the real time video analysis this may cause the error in object detection. This distortions are result of some intrinsic as well as extrinsic properties of camera, radial distortion and tangential distortion are two major distortion factor to be corrected previously before template matching.
Radial Correction factor

$$x_{correction} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$y_{correction} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

Tangential Correction factor

$$x_{correction} = x + [2 p_1 xy + p_2 (r^2 + 2x^2)]$$

$$y_{correction} = y + [p_1 (r^2 + 2y^2) + 2 p_2 xy]$$

In general we need 5 factor for correction of distortion correction.

$$Distortion\_coefficients = (k_1, k_2, p_1, p_2, k_3)$$

and intrinsic parameter such as focal length and optical center of the camera. Represented as camera matrix

$$camera\_matrix = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

For our stereo camera setup we need to calibrate the camera to remove this camera distortion base on sample patterns.

Fig. 4: Camera matrix calibration

We saw basic concepts such as epipolar constraints and other related terms in the last session. We also saw that if we have two images of the same scene, we can intuitively obtain detailed information from it.Below is an image and some simple mathematical formulas which proves that intuition. The above diagram contains equivalent triangles. Writing their equivalent equations will yield us following result:

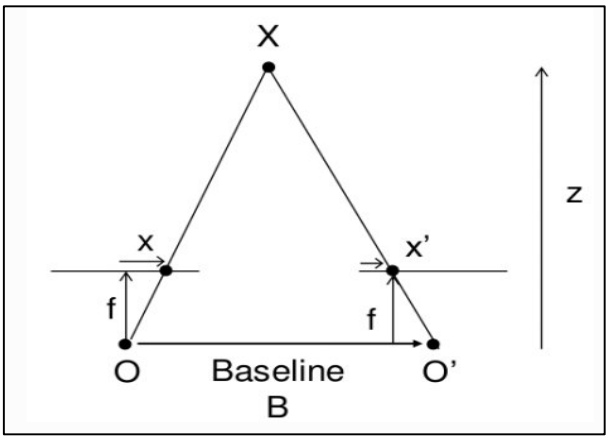

Fig. 5: Depth map for calibration

$$disparity = (x - x') = \frac{Bf}{Z}$$

Here x and x' are the distance between points in image plane corresponding to the scene point 3D and their camera center. Is the distance between two (which we know) cameras and the focal length of the camera (which is already known). In short, above the equation states that the depth of a point in a scene is inversely proportional to the distance difference between the corresponding image points and their camera centers. Thus, we can derive the depth of all pixels in an image with this information.

A. Depth Mapping

Dynamic programming helps find a global optimum with a polynomial complexity, stereo methods that rely on it are not problem-free. The requirements of uniqueness and ordering constraints that are not always met in real scenes impose the most severe limitation. As a result, errors can occur and therefore appear in the depth map as a horizontal string in the disparity map. The appearance of horizontal noise lines in the depth map in stereo vision systems is a common problem

B. Two Hand Recognition and Tracking

In the case of detection of two hands, algorithm flow is the same as in detection of one hand. If two hands are extended the same length, the usual algorithm for gesture recognition will have little difference in step-5. We should select two suitable blobs after labeling. In the case of a hand detection algorithm, skin-color detection step may be skipped, but it becomes mandatory for a two-hand recognition system, since the second label may result in a noise peas with a larger area. All other steps will follow one hand algorithm after detecting two palm blobs. Two x,y,z co-ordinates will be received as a result. Some experimental results are available in the section on experiments and results If two hands are not extended in the same way, the above algorithm for the recognition of two hands cannot be applied without any changes. The second

hand can be skipped accidentally, it can happen at step-3 when the algorithm is checked and detected first hand, but the second hand will ignore which palms are located shortly after the first hand. The third step of the algorithm should be changed in order to avoid this case.

C. Experimental result and discussion

Minoru3D is used for stereo imaging with a low-cost 3D stereo camera. Its 30fps resolution is 320x240 for < 15fps and 640x480. Although this is at the bottom of our requirements, our experiments have a resolution of 320x240. Skin tone detection is carried out before the generation of the depth map. The results obtained from the detection of skin color based on multiple thresholds are shown in Figure For stereo calibration, pictures of an 8x5 chessboard were taken. A total of 160 different pairs of images were taken in different inclinations with the chessboard. For each of these pairs, rotation and translation matrices were calculated, resulting in a total of 160 matric pairs. In Open-CV, the function cvStereoCalibrate () was used to generate these matrices, which were then approximated using the iterative algorithm of Marquita into a pair (R, T) with minor changes. The performance evaluation was carried out using 20 test cases for each gesture and we achieved a total accuracy of 93.75% within the nominal distance of 1 feet from the Minor 3D camera. Since the field of view of the Minor 3D is 40 degrees, the minimum distance to make the gesture is 9 cm. The accuracy of the HGR reduces, however, as the distance between the user and the camera increases, as shown in Fig.8. The breakdown of computational loads for different algorithms is determined by They report that most of the computational complexity is due to the calculation of the depth/disparity map (41 percent) and the segmentation of objects (53 percent), tasks that can be pipe-lined to two different processors To obtain a depth map, an OpenCV function based on an efficient block matching stereo algorithm is used.cv.find Stereo Correspondence BM (). SAD window size 9 is used in our study. The stereo correspondence result from the algorithm corresponding to the block is shown in Fig.6 as a color depth map. The color intensity can be observed as the object gets closer.

V. INDIAN SIGN LANGUAGE

In South Asia, ISL is the predominant sign language used by at least several hundred thousand deaf signatories (2003). Like many sign languages, it is difficult to estimate figures with any certainty, since the Indian Census does not list sign languages and most studies have focused on the north and urban areas. The Indian deaf population of 1.1 million is 98 percent alphabet According to oralist philosophy, deaf schools try to intervene early with hearing aids, etc., but they are largely dysfunctional in a poor society. As of 1986, only 2% of deaf children attended school. Pakistan has a deaf population of 0.24 million, about 7.4 percent of the country's total disabled population.

Fig. 6: Data-set1

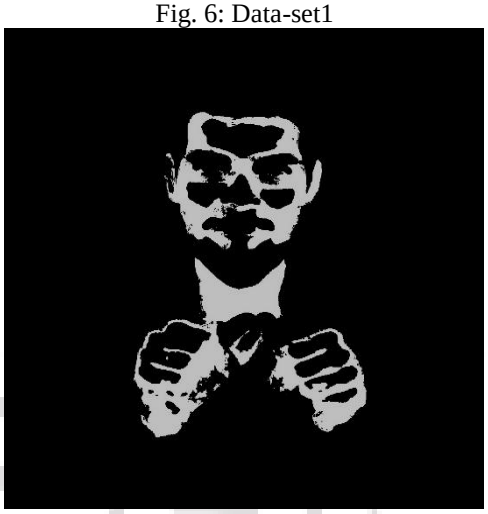

Fig. 7: Data-set2

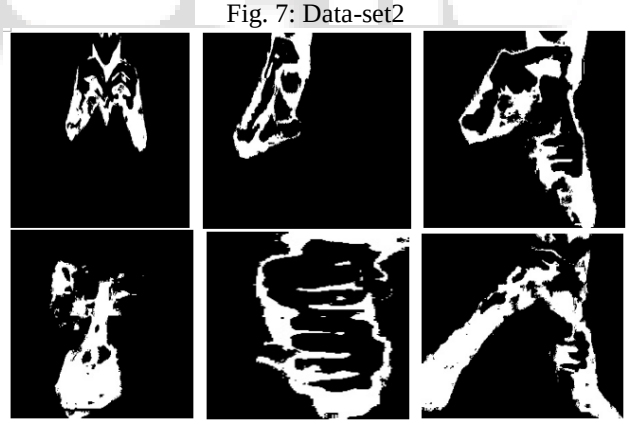

Fig. 8: Data-set 3

VI. 3D MODELLING

Computer vision allows us to reconstruct highly naturalistic computer models of 3D environments from camera images We may need to extract the geometry of the camera (calibration), the structure of the scene (surface geometry), as well as the visual appearance (color and texture) of the scene, or improve virtual actors.

Generating a solid object's 3D computer model is quite simple these days. Just put something on a platform, take the right picture series, and voila! However, dynamic models are much harder. You can't rely on a series of pictures, just take a shot, make a 3D model and update it in real time.

This is not an easy task. Fortunately, Toshiba CRL has a strong history of computer vision exploration. They helped to develop Toshiba's gesture recognition systems in its TV sets. The use of RGB lights is just as clever. Each part of your face reflects a different amount of each colored light, depending on its contours, allowing a computer to identify the surface angle by its overlapping colour. Since these colors can be identified as quickly as the camera refresh rate, 3D mapping is highly accurate in real time. The RGB light overlaps the contours of the human face in different colour.

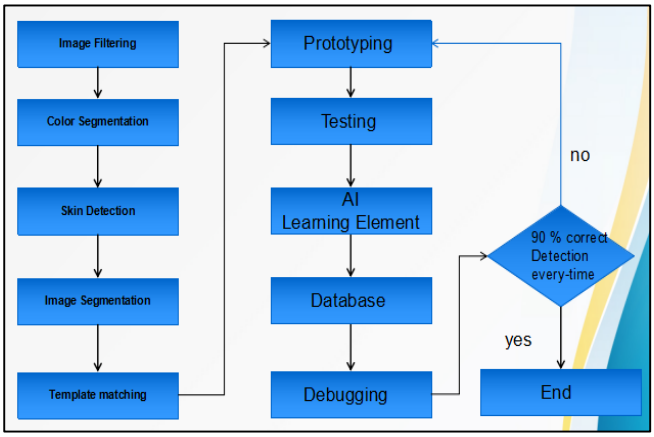

Fig. 9: Architecture of system

REFERENCES

[1] https://singularityhub.com/2010/04/06/generating-3d-models-of-your-face-in-real-time-with-rgb/#sm.0001387pwlbtvddm10lv6vzb0whtd
[2] Paulraj M. P. Sazali Yaacob, Mohd Shuhanaz bin Zanar Azalan, Rajkumar Palaniappan, "A Phoneme based sign language recognition system using skin color segmentation", Signal Processing and its Applications (CSPA) – pp: 1 – 5, 2010.
[3] Byong K. Ko and H. S Yang, "Finger mouse and gesture recognition system as a new human computer interface", pp: 555-561, 1997.
[4] Yang quan, "Chinese Sign Language Recognition Based on Video Swquence Appearance Modeling", ICIEA, the 5th IEEE Conference, pp: 1537 – 1542, 2010.
[5] P. Mekala, R. Salmeron, Jeffery Fan, A Davari, J Tan, "Occlusion Detection Using Motion-Position Analysis" IEEE 42nd Southeastern Symposium, on System Theory (SSST"10), pp: 197-201, 2010.
[6] Jae Y. Lee and Suk I. Yoo "An Elliptical Boundary Model for Skin Color Detection" pp: 2- 5, 2002.

