

Image Processing based Human Gesture Recognition using Kinect Camera

Srinidhi.A S

Student, MTech

Microelectronics and Control Systems
Dayananda Sagar College Of Engineering
Bangalore, India

Rajashekar.J S

HOD & Associate Professor

Dept. of Electronics and Instrumentation Engineering
Dayananda Sagar College Of Engineering
Bangalore, India

Abstract- This paper introduces a novel way of the depth map and RGB image alignment for Kinect. Rule base algorithm is employed to estimate the joint co-ordinates from both depth and RGB images, and this technique is robust for the uncertainties of the particular extracted joint co-ordinates. Although still very difficult in uncontrolled scenarios, it has been successful in more restricted settings (e.g., fixed viewpoint, no occlusions) with recognition rates approaching 100%. The experimental results demonstrated which the proposed method is competent at precisely aligning the particular depth and RGB images, contributes a excellent improvement for body gesture segmentation, and may potentially improve the particular performance of body gesture recognition in Human-Machine Interaction (HMI).

Keywords- Kinect Camera, Matlab, Hand Gesture, RGB and Depth data, Image Processing

I. INTRODUCTION

The research on HMI is to create easy-to-use interfaces by directly employing natural communication and manipulation skills of humans. Among different human body parts, the hand is the most effective general-purpose interaction tool, due to its dexterity. Adopting hand gesture as an interface in HMI will not only allow the deployment of a wide range of applications in sophisticated computing environments such as virtual reality systems and interactive gaming platforms, but also benefit our daily life such as providing aids for the hearing impaired, and maintaining absolute sterility in health care environment using touchless interfaces via gestures [1].

Despite a significant research effort in the last decade, recognizing human actions from motion is still a challenging problem on account of scene complexity (e.g., occlusions, clutter, interacting objects, illumination changes), acquisition difficulties (camera movement, viewpoint changes), and the complexity of human actions (non-rigid objects, action variability).

At the moment, the most beneficial tools for capturing hands gesture usually are electro-mechanical or even magnetic realizing devices (data gloves) [2]. These procedures employ sensors attached to a glove that transduces finger flexions into electrical alerts to determine the hands gesture. They deliver essentially the most complete, application-independent pair of real-time measurements from the hand with HMI. Nonetheless, they

include several negatives (i) they are expensive for laid-back use, (ii) they hinder the naturalness of hand gesture, and (iii) needed complex calibration and startup procedures for getting precise proportions.

Vision-based hands gesture identification serves being a promising alternative to them for the potential to provide more natural, unencumbered, non-contact conversation. However, despite a lot of previous work [3], [5], [6], regular vision-based hands gesture identification methods continue to be far by satisfactory pertaining to real-life applications.

In Nov 2010 Microsoft launched Kinect, an infraredlight range-sensing camera, and in the process added raw depth being a new video-data get modality. Throughout June 2011 Microsoft released a Kinect Software package Development Equipment (SDK) consisting of a list of powerful algorithms pertaining to extracting landscape depth and object masks, and subsequently creating a skeleton type of a person in front of the camera in real time.

Although Kinect's concentrate on application is video gaming, the 3- D get and robustness associated with Kinect as well as SDK get helped spawn numerous research and "hacking" projects in the field associated with human-Machine interfaces. Most approaches, nevertheless, rely upon tracking not really recognition. Usually a collection of parameters, e. g., thresholds upon joint areas, velocities and accelerations, are generally specified to localize and track motions.

In this paper, we propose a Novel hand gesture recognition algorithm called Rule Base Algorithm to detect the location and orientation of an object in three-dimensional space ,real time using MATLAB and the X-box Kinect camera. The solution employed uses the MATLAB Image Acquisition toolbox and Image Processing Toolbox to retrieve and analyze the Kinect's RGB and Depth data. The system authenticates & actuates by means of several procedural actions such as abstract formation for human body & necessary actions to be taken out to predict the skeleton, point generation based on 3d posture based point value of an image predicted for human body skeleton, segmentation of human body, generating value for each segmentation and generating code for the value obtained from the predicted posture.

II. PROPOSED ALGORITHM

In this subsection Fig.1 shows the flowchart of Rule base algorithm which was developed for skeleton tracking and for recognition process.

A. Movement Capturing

Movement capturing is the process that consists on the translation of real movements into digital representations by tracking a set of interest points on scene during a determined time [7]. It is important to indicate that movement capturing can be done on any individual which possesses movement.

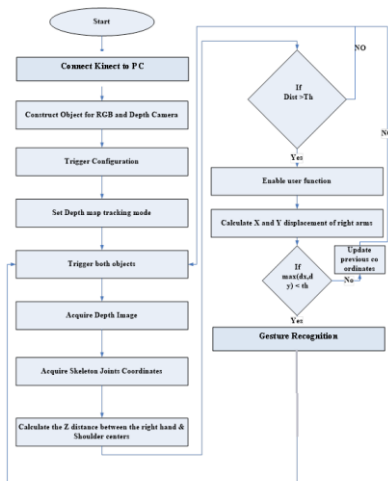


Figure 1: Flowchart of Proposed Algorithm

B. Depth Map

Distinctive from traditional methods of which use color-markers intended for hand detection, [4] uses both the depth map along with color image acquired by Kinect sensor in order to segment the hand shapes, which ensures its robustness in order to cluttered background, shown in Fig-2.

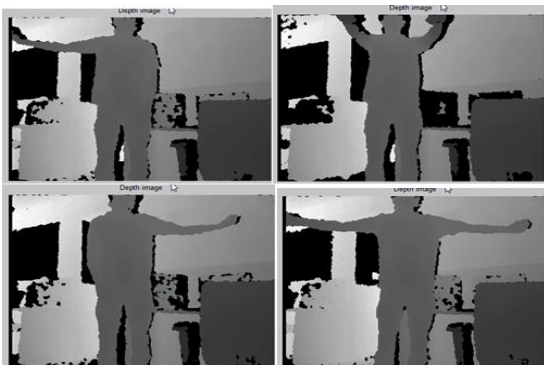


Figure 2: Depth Model Images

C. Skeleton Tracking

It is just a collection of human body joints to get the different parts from the body for which a tracking algorithm is applied. A Joint gives back the data on the entire body points.

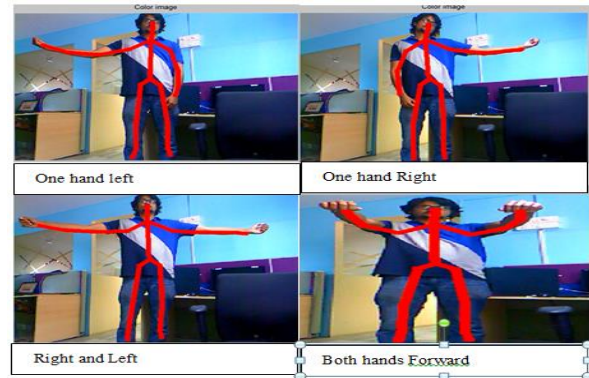


Figure 3: RGB and Skeleton Model Images

D. Gesture Recognition

The block diagram of proposed algorithm is shown in figure . The algorithm starts with interfacing kinect camera with MATLAB. For interface we have used MATLAB image acquisition toolbox. We ultimately get depth map and skeleton information. First of all we check whether gesture is enabled or not for that we first check z distance between right hand with shoulder. If the distance is more than predefined threshold then the gesture mode is enabled. After we calculate displacement of coordinates of hands if the either x or y displacement is higher than threshold then we go for decision. The final decision of recognised gesture can be given as an input to any application.

We find the Euclidean distance of the current frame and compare it to the previous frame from all the gestures. The Euclidean distance indicates how similar the two frames are. If frame are identical the distance will be zero.

We now have our two gestures, a prerecorded gesture(reference gesture) and newly performed gesture (input gesture), to do rule base algorithm calculations. The first thing we have to do is to calculate the cost between each reference and input frames. This can be visualized by skeleton joint matrix.

The Euclidean distance between points p and q is the length of the line segment connecting them .

In Cartesian coordinates, if $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n-space, then the distance (d) from p to q, or from q to p is given by the Pythagorean formula:

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \tag{1}$$

$$\text{Total Euclidean distance} = \sqrt{\sum_{i=1}^n ((q_i - p_i)^2)}$$

III. EXPERIMENT RESULTS

From our lab setup, we were able to identify gestures in more accurately. The whole system was implemented in real time and the results were very encouraging as shown in Table-1 which shows confusion matrix. There are no seriously confused categories. In this paper we have presented a Rule Base Algorithm approach to recognize our gesture.

	One hand left	One hand Right	Hands Right and Left	Both Hands Forward
One hand Left	49	1	0	0
One hand Right	2	48	0	0
Hands Right and Left	4	5	39	4
Both Hands Forward	0	1	8	41

TABLE 1: CONFUSION MATRIX

IV. CONCLUSION

In this paper, we connect the Kinect with the matlab tool for recognition of gestures. The recognition with the Kinect is with quite good performance. The gestures are recognized most of the times when the conditions are suitable and the user adheres to the assumptions made. This project is still in progress and it will make a significant change in applications. The Movement of user's finger will control the embedded system by moving hand in front of camera without wearing any gloves or markers.

The system does perform poorly under certain conditions. If another object is close to the hand, the hand point might move to the other object. It may happen that a hand point is lost. In general, very fast motions may cause tracking failure. Sometimes the swipes may not be recognized if there is too much variation in the depth of the hand that is being tracked. There might be false gestures because of bad tracking of the hand point. In some cases, overall tracking might be bad. Re-calibrating the user may resolve the problem. The depth map is unreliable when there is too much IR interference in the environment, for example in an open hall with sun light.

REFERENCES

- [1] J. P. Wachs, M. Klsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54:60–71, 2011.
- [2] E. Foxlin. Motion tracking requirements and technologies. *Handbook of Virtual Environment Technology*, pages 163–210, 2002.
- [3] C. Chua, H. Guan, and Y. Ho. Model-based 3d hand posture estimation from a single 2d image. *Image and Vision Computing*, 20:191 – 202, 2002.
- [4] Z. Ren, J. Yuan, and Z. Zhang. Robust hand gesture recognition based on finger-earth movers distance with a commodity depth camera. In *Proc. of ACM Multimeida*, 2011.
- [5] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura. Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints. In *Proc. of Third IEEE International Conf. on Face and Gesture Recognition*, 1998.
- [6] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla. Filtering using a tree-based estimator. In *Proc. of IEEE ICCV*, 2003.
- [7] Dover, "Complete human and animal locomotion": Eadweard Muybridge; New York, 1979.