IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Design of Vision based Gesture Controlled Interface for UAV

Avinash Kumar Bussa

Department of Aerospace Engineering, University of Petroleum and Energy Studies, Dehradun, India

Abstract: In the present scenario drones are buzzing in the skies all over the world. This paper presents the new interface apart from all familiar controlling methods of drones as Computer keyboard, smart phone, gaming controller or a RC transmitter. Hand Gesture Signals sensed by an open source SDK consists of stereo cameras which uses infrared emitters as positional references controls the drone which gives a hands-free flying experience. We also propose a vision based hand gesture driven UI control system that enables the user an intuitive way of working with the system with on-air hand movement and gesture. The system can also be used in a wide range of applications like UAV control, operating ground robots and Assembly lines in production lines of factories.

Keywords: Intel® RealSense™; Machine Vision; Gesture Control; UAV.

I. INTRODUCTION

Gesture Recognition is the technical advancement with the goal of developing human – machine interface by using mathematical algorithms. In the field of Computer Science many researches are going on to interpret the sign language by using cameras and computers. Gesture recognition enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant.

Hand gesture recognition, aims to design systems which can identify human hand gesture as an input and uses these gestures in order to control a device through mapping of command as an output. Hand gesture recognition is divided into two major part: 1-contact based, 2-vision based Contact based devices are based on physical interaction of user with the interfacing devices which usually are based on technologies like data glove, accelerometers, multi-touch screen etc. which use sensors and detectors. Vision-based system relies on camera, able to capturing video sequences by one or several camera to interpret and analyze the gesture. Vision based gesture technologies are cheaper, user friendly and easily available in many computer devices and there is no need of physical contact with the equipment under control.

II. RELATED WORK

The book "Intelligent Technologies for Interactive Entertainment" which has the proceedings of 4th INTETAIN. One of the paper selected in this conference explains about the idea of a gesture controlled drone by using Kinect controller which is generally used in the Microsoft Xbox to imitate the gesture of the users who uses it for playing the virtual reality games. Which also states that to overcome the problems of speed and interface of wearable devices vision based controllers are introduced.

Xiamong [6] proposed recognition techniques for application of human robot interaction. The author extracted topological features of the hand from the binary image of segmented hand. Some rules were defined to recognize gesture based on extracted features. The main idea of recognition is based on search circles and finger intersection points. The acquisition device was stereo camera and 8 postures were used. The application of proposed system is for humanoid robot control. The postured recognized with the accuracy of 95%.

Mr.Gobadi in his paper [5] used 3d range camera as an acquisition device and used adaboost classifier for hand detection. The idea of classification is the technique of drawing a circle at centre of mass with a radius depending on the distance of the hand from the camera. Then, the circle is tracked and the intersection of the circle with the hand is extracted as a binary function. The number of 1-to-0 transitions in binary function is counted and used as a feature to classify the hand. The recognition accuracy was 94.4%.

Lamberti [3] proposed a real time recognition techniques using colour glove. The proposed technique considers 9 features extracted from colour glove. The author calculated the lengths of the line connecting the centroids of the palm and each finger as well as fingers relevant angels. classification technique used for algorithm was Learning Vector Quantization (LVQ) and the recognition rate of the technique was around 98%.the average time taken for entire process was noted as 140 CPU msec. Webcam was used for the system and total 13 number of gesture was recognized.

Hand-driven 3D manipulation and editing of virtual objects is employed in [4], in the context of a virtual environment for molecular biologists. A hand-gesture interface is proposed that allows the manipulation of objects in a

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

virtual 3D environment by recognizing a few simple gestures and tracking hand motion

The Mean Shift algorithm which is discussed in the paper [11] is an iterative procedure that detects local maxima of a density function by shifting a kernel towards the average of data points in its neighborhood. The algorithm is significantly faster than exhaustive search, but requires appropriate initialization. The Mean Shift algorithm has been utilized in the tracking of moving objects in image sequences. It's utilization is not restricted to hand tracking, but can be used to track any moving object. It characterizes the object of interest through its color distribution as this appears in the acquired image sequence and utilizes the spatial gradient of the statistical measurement towards the most similar (in terms of color distribution similarity) image region.

Matthew brand, Nauria Oliver and Alex Pentland's paper[12] on coupled Hidden markov models (HMM)s discusses about many algorithms which are used in Tracking and recognizing body and hand motion have been employed in personal training and movie making applications. This system infers the posture of the whole body by observing the trajectories of hands and the head, in constrained setups. Based upon this techniques a prototype system was designed for a virtual Personal Aerobics Trainer that recognizes stretching and aerobic movements and guides the user into a training program.

Temporal gesture Recognition requires temporal features along with spatial features. It's possible to recognize some gestures by 2D locations hands, even though it's not general and view dependant. The most fundamental feature is the 2D loction of interested blob view of particular object. Wren. Et. Al[13] uses a multi class static model of color and shape to obtain a 2D representation of the head of hand in a wide range of viewing conditions in their tracking system FFinder.

Rupam Das and Dr. K. B. Shivakumar's paper [15] on Augmented World: Real Time Gesture Based Image Processing Tool with Intel RealSense™ Technology. This technology supports real time background separation by combining the depth and RGB streams of RealSense™ camera. Many robust techniques for background separation are used to separate the background to augment segmented user photo in an artificial scene. They proposed a "Split grid" based hand gesture driven UI control system that enables the user an intuitive way of working with the system with on-air hand movement and gesture.

III. METHODOLOGY

The user interface and the corresponding interaction modalities play an essential role in the human-computer relationship. Advanced multimodal interfaces present yet another step in this equation, providing users with the freedom and flexibility to choose the best input modality for specific tasks. Users generally prefer multimodal interaction when it is available and intuitive to use. Gesture-based user interfaces, in combination with the latest technical advances that incorporate accurate yet affordable new types of input devices, provide realistic new opportunities for specific application areas (e.g., entertainment, learning, health, engineering), especially for users who are uncomfortable with more commonly used input devices and/or technology. Gesture input devices and sensors are of special interest.

Gesture acquisition methods can, in general, be divided into methods incorporating a specific device that the user must physically hold or have on his/her body and hands/body-free methods. The latter become more and more popular as the user becomes a controller rather than an operator. One of the first widespread, accurate, and commercially viable solutions was the Nintendo WiiMote controller, bundled with the Wii console, released in 2006. The WiiMote, besides its vocal and haptic modalities, incorporates an accelerometer that allows the acquisition of full 3D gestures. It can operate as a separate device and has been successfully used for many atypical applications. Another important milestone is the Microsoft Kinect sensor, an add-on for the Xbox 360 console, which was released in late 2010. The Kinect, among its visual and auditory inputs, includes a depth-sensing camera. In combination with an open SDK, it can be used to acquire and recognize full body gestures for multiple users at a time. The latest technological breakthrough in gesture-sensing devices has come in the form of a Intel® RealSense™ 3D camera which can recognize the gestures and track the objects. Intel® RealSense™ camera fits remarkable technology into a small package. There are three cameras that act like one—a 1080p HD camera, an infrared camera, and an infrared laser projector—they "see" like the human eye to sense depth and track human motion. Intel® RealSense™ technology[1] redefines how we interact with our devices for a more natural, intuitive and immersive experience.

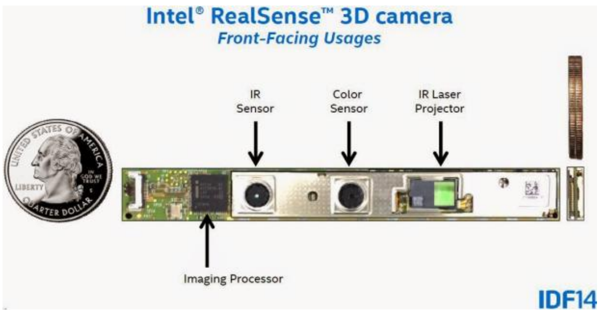

Figure 1. Intel® RealSense™ 3D camera component view

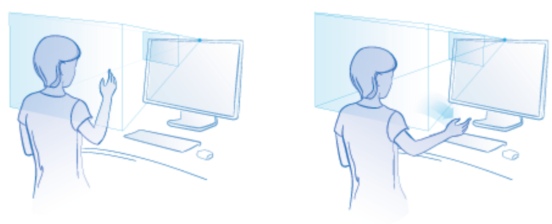

Figure 2. Range of Intel® RealSense™ 3D camera integrated PC

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Most of the complete hand interactive systems can be considered to be comprised of three layers: detection, tracking and recognition[7]. The detection layer is responsible for defining and extracting visual features that can be attributed to the presence of hands in the field of view of the camera(s). The tracking layer is responsible for performing temporal data association between successive image frames, so that, at each moment in time, the system may be aware of "what is where". Moreover, in model-based methods, tracking also provides a way to maintain estimates of model parameters, variables and features that are not directly observable at a certain moment in time. Last, the recognition layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated to particular classes of gestures.

A. Detection

The primary step in gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. This segmentation is crucial because it isolates the task-relevant data from the image background before passing them to the subsequent tracking and recognition stages. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. Such features are skin color, shape, motion and anatomical models of hands.

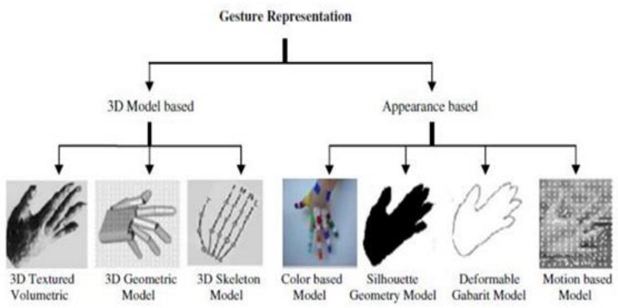

Figure 3. Vision based hand gesture detection representation

B. Tracking

Tracking, or the frame-to-frame correspondence of the segmented hand regions or features, is the second step in the process towards understanding the observed hand movements. The importance of robust tracking is twofold. First, it provides the inter-frame linking of hand/finger appearances, giving rise to trajectories of features in time. These trajectories convey essential information regarding the gesture and might be used either in a raw form (e.g. in certain control applications like virtual drawing the tracked hand trajectory directly guides the drawing operation) or after further analysis (e.g. recognition of a certain type of hand gesture). Second, in model-based methods, tracking also provides a way to maintain estimates of model parameters variables and features that are not directly observable at a certain moment in time.

C. Recognition

The overall goal of hand gesture recognition is the interpretation of the semantics that the hand(s) location, posture, or gesture conveys. Basically, there have been two types of interaction in which hands are employed in the user's communication with a computer.

The first is control applications such as drawing, where the user sketches a curve while the computer renders this curve on a 2D canvas. Methods that relate to hand-driven control focus on the detection and tracking of some feature (e.g. the fingertip, the centroid of the hand in the image etc.) and can be handled with the information extracted through the tracking of these features. The second type of interaction involves the recognition of hand postures, or signs, and gestures. Naturally, the vocabulary of signs or gestures is largely application dependent. Typically, the larger the vocabulary is, the hardest the recognition task becomes.

IV. GESTURE CHALLENGE

Gesture recognition systems suffer from some challenges which can be raised at any stage of recognition. Some of these challenges are easy to handle, like scaling problem or skin color but some like cluttered background or lighting variation can pose problems in recognition. However these challenges could be classified as follows.

A. Gesture Challenge

The most important challenges of recognition could be noted as translation, scaling, rotation and background. Translation challenge is related to hand object location in the image which might change during capturing process either by changing the location of hand or the camera. Scaling is the distance of the hand from the camera lens .it can affect the overall recognition if the hand is too far. Generally this problem could be taken care by histogram or scale matching techniques. Rotation of gesture or hand object in particular, is a difficulty which influences the recognition accuracy. This challenge could be tackled by only soft computing techniques and some classifier by proper training samples. The most important challenge of gesture is background problem. It is related to extraction of foreground object (hand) from background. In fact the hand has to be extracted for recognition decision stage from background. Some researchers overcome this problem by using uniform background but in real system which has complex or clutter background, use of robust segmentation techniques can overcome this issue.

B. System Challenge

The main system challenges are response time and cost of the system. Generally for real time application the execution time of any gesture algorithm has to be fast, so that there has not to be a noticeable time between the user gesturing and computer response. The next system challenge is the cost. According to application of gesture system, it requires special hardware like camera, appropriate sensor or data glove if necessary.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

C. Environmental Challenge

These challenges are classified into illumination & ethnic group. The causes of these problems are because of surrounding environment. These challenges are there by nature and affect the input image. Firstly is illumination which could be defined as changes in the light condition due to natural or artificial, indoor or outdoor factor. These challenge should be handles in the preprocessing stage either by segmentation techniques or color space model such as HSV, Lab, Ycbcr or both of segmentation and color space conversion techniques.

V. CONCLUSION AND FUTURE RECOMMENDATIONS

Given the current increase in popularity of personal drones, we need to create natural interaction techniques to best support users. The Intel® RealSense™ undoubtedly represents a revolutionary input device for gesture-based human-computer interaction. From this project, the evaluation is stating this controller can be a possible replacement for speed and accuracy loaded optical motion capture system in a controlled space and we are establishing an interface from these recognition system to drones by using Java and C++ coding.

The next step is the implementation of the best technical solution to support multimodal Input/output for HDI, taking into consideration the need for multiple modalities based on the context of use. The scenarios of use we described can then be implemented. We will also look at proxemics in the 3D space for HDI. As several users mentioned feeling attached to the drone, it would be interesting to further study human emotion towards drones and how they differ from interacting with ground robots.

REFERENCES

[1] Jessica R. Cauchard , Jane L.E, Kevin Y. Zhai, James A. Landay's study " Drone & Me – An exploration into natural human-drone interaction" , Stanford University, Department of Computer Science, Stanford.
[2] L. Lamberti and F. Camastra, "Real-time hand gesture recognition using a color glove," in Image Analysis and Processing–ICIAP 2011, Springer, 2011, pp. 365–373.
[3] Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence., July, 1997. Vol. 19(7), pp. 677 -695.
[4] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," Pattern Anal. Mach. Intell. IEEE Trans., vol. 15, no. 10, pp. 1042–1052, 1993.
[5] X. Yin and M. Xie, "Hand Posture Segmentation, Recognition and Application for Human-Robot Interaction," in International Conference on Advanced Robotics, 2007.
[6] S. X. Ju, M. J. Black, S. Minneman, and D. Kimber, "Analysis of gesture and action in technical talks for video indexing," in Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997.
[7] X. Zabulis , H. Baltzakis , A. Argyros "Vision-based Hand Gesture Recognition for Human-Computer Interaction" in Institute of Computer Science Foundation for Research and Technology - Hellas (FORTH) Heraklion, Crete, Greece.
[8] X. Yin and M. Xie, "Hand Posture Segmentation, Recognition and Application for Human-Robot Interaction," in International Conference on Advanced Robotics, 2007.
[9] Bo Liu, Guijin Wang*, Xinghao Chen, Bei He "Stable and real-time hand gesture recognition based on RGB-D data" Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China
[10] Sujeet D.Gawande , Prof. Nitin R. Chopde's "Human Computer Interaction using Hand Gesture Recognition with Neural Network: A Review", International Journal of Application or Innovation in Engineering & Management (IJAIEM).
[11] Y. Cheng. Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Analysis and Machine Intelligence,1995
[12] Matthew Brand, Nauria Oliver and Alex Pentland's "Coupled Hidden Markov Models for complex action Recognition", Vision and modeling group, MIT Media lab.
[13] Christopher R. Wren, Alex P. Pentland's "Dynamic Models of Human Motion", MIT Media Laboratory; 20 Ames Street; Cambridge MA 02139 USA.
[14] Jochen Triescha, Christoph von der Malsburgb "Classification of hand postures against complex backgrounds using elastic graph matching" Department of Cognitive Science, UC San Diego, 9500 Gilman Drive, 92093-0515 La Jolla, CA, USA.
[15] Rupam Das and Dr. K. B. Shivakumar "Augmented World: Real Time Gesture Based Image Processing Tool with Intel RealSense™ Technology" International Journal of Signal Processing, Image Processing and Pattern Recognition, 2016.