

A Survey on Superpixel-Based Hand Gesture Recognition Using Static Dataset

Ankita Pramod Thakur¹, Prof. S.P.Magadam²

PG Student [VLSI], Electronics and Telecommunication, R.M.D Sinhgad school of Engineering, Pune, India ¹

Assistant Professor, Electronics and Telecommunication, R.M.D Sinhgad school of Engineering, Pune, India ²

ABSTRACT: Hand gesture recognition framework got incredible consideration in the current couple of years in light of its complexness applications and the capacity to connect with machine productively through human computer interaction. With the large use of computers Human computer Interaction has become an important part of our daily life. The hardware devices like keyboard, mouse are used for the Human Computer Interaction. This paper exhibits another new superpixel-based hand gesture recognition system based on a novel superpixel earth mover's distance metric, together with images taken from Kinect depth camera. The profundity and skeleton data from Kinect are successfully used to deliver marker less hand extraction. The hand shapes, relating surfaces and profundities are spoken to as superpixels, which successfully hold the general shapes and shade of the signals to be perceived. In view of this portrayal, a novel separate metric, superpixel earth mover's distance (SP-EMD), is proposed to quantify the disparity between the hand motions.

KEYWORDS: Hand gesture recognition, Human computer interaction, Superpixel earth mover's distance (SP-EMD)

I. INTRODUCTION

The basic point of building hand gesture recognition system is to create a natural interaction amongst human and PC where the perceived signals can be utilized for controlling a robot or, then again passing on important data. The most effective method to shape that came about hand motions to be comprehended and very much translated by the PC considered as the issue of gesture interaction. Human computer interaction (HCI) likewise named Man-Machine Interaction (MMI) alludes the connection between the human and the PC and since the machine is inconsequential without reasonable use by the human. There are two fundamental qualities ought to be regarded when outlining a HCI framework as said in: usefulness what's more, convenience. Framework usefulness alluded to the arrangement of capacities or administrations that the framework prepares to the clients, while framework ease of use alluded to the level and degree that the framework can work and perform particular client purposes proficiently. The framework that accomplishes an appropriate adjusts between these ideas considered as persuasive execution and intense framework. Gestures used for communicating between human and machines as well as between people using sign language.

II. RELATED WORK

In literature, the problem and the previous techniques are as described,

Y. R. Wang et.al [2] the paper proposes a novel real time hand detection algorithm based of skin-shading. The hands can be identified with no preparation or wrist portioning. The discovery separation can be up to 3 meters. The proposed technique, which is reasonable for executing on implanted framework, can perform hand detection fast with high exactness under element and complex conditions i.e. dynamic and complex environments.[2]

M. Kolsch et.al [3] Computer Vision strategies for hand gesture interfaces must outperform current execution as far as heartiness and speed to accomplish intuitiveness and ease of use. Late advances in example acknowledgment have made location at casing rate conceivable. This researched the appropriateness of different hand stances for safeguard location before discretionary foundations in dark level pictures. Our commitments are as per the following. In the first

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

place, we illustrate the reasonableness of the essential picture way to deal with the errand of identifying hand appearances. Second, a subjective measure is displayed that sums to form the earlier gauge of detect ability, easing the requirement for figure escalated preparing. Third, parameters of the identification technique are upgraded, accomplishing significant speed and precision changes. Generally speaking, this review demonstrates how the Viola-Jones finder can accomplish brilliant location rates for hand stances. These outcomes give an essential stride towards effortlessly deployable the, robust vision-based hand gesture interfaces.[3]

Object Removal by Region Based Filling Inpainting [4]: In this paper another calculation is presented for removing big objects from digital images .Best first algorithm is introduced to fill the gap that is left. It is a piece based testing handle, as opposed to an area based process. Block based testing procedure is utilized to fill the extensive question. The chosen substantial question is expelled and loaded with foundation of the first picture. This strategy takes after an approach of model based surface union for filling of the objective area with a balanced bound together plan. This strategy is capable of engendering two-dimensional surface and straight structure into the objective locale. The results say that the selection of the fill order should be done carefully and is necessary to deal with the task.

Brilliant Real-Time Video Inpainting with PixMix[5]: High magnificence constant fit for image capture and video inpainting is presented in this paper. They propose a calculation known as PixMix which works for the changing of the live video streams, and notwithstanding when applying conditions continuously. This works for the genuine decreased reality applications. A PixMix calculation is a pixel based approach as opposed to a fix based or area based. New following approach and casing –frame intelligence is consolidated for the video control progressively. To accomplish a high lucidness for turning and translational cameras homographic based approach is utilized. To give best in painting comes about this approach permits steadiness between the cost and spatial capacity. This in painting technique utilizes just the past and the current outline for constant controls.

Video abstraction based on the visual attention model and online clustering [6]: Video abstraction is utilized to quick peruse the video from the enormous video database and to get the quick get to and genuine substance. In light of the Visual consideration what's more, bunching a calculation is proposed. Numerous thoughts are placed in this paper. The principal thought is to choose the key casing; a disparate substance separate determination standard is utilized. Abnormal state same substance is recovered to speak to the principle content into video outline, this is finished by districts of concentrate extraction strategy, and no comparable substance is wiped out for the critical changes. Web based grouping is utilized decline similar redundancies, time cost which is utilized for continuous video preparing. In the first place shot limits are recognized, and then from this limits the key edges are taken. To accomplish this shot limits key edge choice calculation in view of content separation rule is utilized to get comparable key edges from the video. For both adequacy and execution the rundown calculations meet the necessities.

Multi-sensor background subtraction by fusing multiple region-based probabilistic classifiers [7]: Based on the combination of various district based classifiers which prepare the shading and profundity information which are caught by the RGB cameras, an able foundation subtraction technique is proposed. With a mix of locale based forefront expectation which depends on the profundity information, and with various profundity and shading models i.e. foundation models the frontal area articles are recognized in view of the scene at pixel and local level. The information gathered from these models is utilized by the blend of specialists to build the frontal area location exactness. This proposed demonstrate chiefly concentrates on the foreground and background area based models.

Model based Segmentation and recognition of dynamic gestures in continuous video stream [9]: In this paper the authors basically focused on the division and acknowledgment of continuous motions which impact from the spatiotemporal variations. The creators utilized two sorts of signals one is two arm developments for shape extraction, and another incorporates a solitary hand development hands form utilized as the element vector. They proposed a Multi scale Signal Model. This model proposes 3 approaches which primarily contrast in end point Localization. Initially proposed approach is utilized to discover the end focuses with Multi scale hunt and Motion recognition system. Second proposed approach is utilized to find the end purposes of the fingers generally with Dynamic Time wrapping. Third approach depends on Dynamic programming. Utilizing these three methodologies the acknowledgment of the hand

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

ranges from the 88% to 96%.The creator said that the third approach is the best for the acknowledgment of hand signal in continuous video streams.

III. MOTIVATION

Hand gesture recognition framework got extraordinary consideration in the current couple of years as a result of its complexness applications and the capacity to connect with machine proficiently through human PC interaction. Hand movement examination and blend has been a longstanding exploration subject in Computer vision due to their significance in human- Computer association (HCI) and film industry. human-computer interaction (HCI) additionally named Man-Machine Interaction (MMI) alludes to the connection between the human and the Computer or all the more accurately the machine, and since then it is irrelevant without appropriate use by the human. There are two principle attributes ought to be regarded when planning a HCI framework as specified: usefulness and ease of use. Framework usefulness alluded to the arrangement of capacities or administrations that the framework prepares to the clients, while framework ease of use alluded to the level and extension that the framework can operate and perform particular client purposes proficiently. The framework that achieves a reasonable harmony between these ideas considered as persuasive execution and capable framework. Motions utilized for conveying amongst human and machines and also between individuals utilizing gesture based communication. With the advancement in Computer Vision and Human Machine Interaction the Computer holds most vital part in our everyday life. Human Computer Interaction can give a few points of interest the presenting the distinctive common types of gadget free correspondence. Signal acknowledgment is one of the few sorts of them to collaborate with the people motions are the regular type of activity which we frequently utilized as a part of our everyday life. In any case, in Computer application to communicate people with machine the association with gadgets like console, mouse and so forth should be requires. As the different hand signals are every now and again utilized by people so the point of this venture is to lessen outside equipment cooperation which is required for Computer application, and subsequently this causes framework more solid for use effortlessly.

IV. OBJECTIVES

The objective of the proposed system is given as

1. To design a superpixel-based hand gesture recognition system
2. To proposed SP-EMD methods which naturally incorporates both the colour texture and depth map into the hand shapes

V. PROPOSED SYSTEM

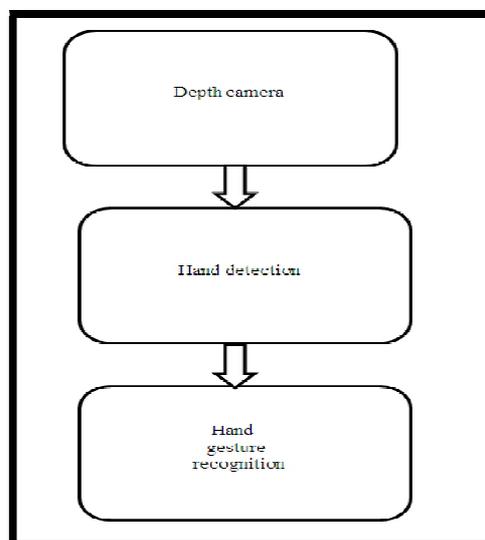


Figure 1 Block Diagram of proposed system

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

The description of the figure 1 is given below

1. **Depth camera:** The Depth Cam Sensor service provides information from a depth camera. At present the only supported depth camera is the Kinect, which means that the Depth Cam Sensor service is not really generic. The Depth Cam Sensor service provides the following operations.
2. **Hand detection:** The hand tracking is based on color recognition. The program is therefore initialized by sampling color from the hand. The hand is then extracted from the background by using a threshold using the sampled color profile. Each color in the profile produces a binary image which in turn is all summed together. A nonlinear median filter is then applied to get a smooth and noise free binary representation of the hand detection, including hand localization, segmentation and representation, is a non-trivial problem in gesture recognition. Previous studies reveal many challenges, such as low accuracy of predicting user's body movement and the weakness against cluttered backgrounds or various lighting conditions. In this work, we utilize the depth information and skeleton tracking provided by Kinect to address these problems. The Earth Mover's Distance (EMD) is a measure of the distance between two probability distributions over a region. It is widely used in image retrieval and pattern recognition. Inspired by the work of Finger-Earth Mover's Distance (FEMD) , we now propose a novel distance metric, Superpixel Earth Mover's Distance (SP-EMD), to recognize the hand gesture. The proposed SP-EMD jointly measures the distance between two hand gestures based on the shape, texture and depth information, while FEMD only uses the contour [1].
3. **Hand gesture recognition:** Hand gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction. Gestures can be static (posture or certain pose) which require less computational complexity or dynamic (sequence of postures) which are more complex but suitable for real time environments. Different methods have been proposed for acquiring information necessary for recognition gestures system. Some methods used additional hardware devices such as data glove devices and color markers to easily extract comprehensive description of gesture features. Some recent reviews explained gesture recognition system applications and its growing importance in our life especially for Human computer Interaction HCI, Robot control, games, and surveillance, using different tools and algorithms.

VI. CONCLUSION

Hand gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction. In the proposed system the SP-EMD and depth camera for contactless HCI are used. The system is based upon the representation in the form of super pixels, which efficiently capture the shape, texture and depth features of the gestures. Based on this representation, a novel distance metric, super pixel earth mover's distance (SP-EMD), is proposed as the dissimilarity measurement for gesture recognition. The proposed SP-EMD metric achieves better performance for hand gesture recognition. Moreover, it is very computationally efficient and thus suitable for real-life HCI applications.

REFERENCES

- [1] Chong Wang, Zhong Liu and Shing-Chow Chan "Superpixel-Based Hand Gesture Recognition with Kinect Depth Camera", IEEE Trans. Multimedia, Vol. 17, No. 1, Jan. 2015.
- [2] Y. R. Wang, W. H. Lin, and L. Yang, "A novel real time hand detection based on skin-color," in *Proc. ISCE*, Hsinchu, Taiwan, pp. 141–142, 2013
- [3] M. Kolsch and M. Turk, "Robust hand detection," in *Proc. FG*, Seoul, Korea, pp. 614–619, 2004.
- [4] N. Neelima , M. Arulvan, B. S. Abdur, "Object Removal by Region Based Filling Inpainting" , IEEE transaction 978-1-4673-5301/2013.
- [5] Jan Herling, and Wolfgang Broll, "High-Quality Real-Time Video Inpainting with PixMix", IEEE transactions on visualization and computer graphics, vol. 20, no. 6, june 2014
- [6] Qing-Ge Ji, Zhi-DangFang, Zhen-HuaXie, Zhe-MingLu, "Video abstraction based on the visual attention model and online clustering" ,Published at Elsevier, Signal Processing: Image Communication 28,241– 253 in 2013
- [7] Massimo Camplani , Carlos R. del Blanco , Luis Salgado , Fernando Jaureguizar , Narciso García, "Multisensor background subtraction by fusing multiple region-based probabilistic classifiers", published at Elsevier Pattern Recognition Letters 2013.
- [8] Hong Li, Michael Greenspan, "Model based Segmentation and recognition of dynamic gestures in continuous video streams",Springer publications in Pattern recognition, 1614-1628, 44(2011).
- [9] Blanca Miriam Lee-Cosioa, Carlos Delgado-Mataa, Jesus Ibanezb, "ANN for Gesture Recognition using Accelerometer Data". Elsevier Publications, Procedia Technology 3 109 – 120, 2012.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

- [10] Fabio Dominio, Mauro Donadeo, Pietro Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition", Elsevier, Pattern recognition Letters 2013.
- [11] Sheng-Yu Peng, Kanoksak Wattanachote, Hwei-Jen Lin, Kuan-Ching Li, "A Real-Time Hand Gesture Recognition System for Daily Information Retrieval from Internet", IEEE Fourth International Conference on Ubi-Media Computing, 978-0-7695-4493-9/11 2011.

BIOGRAPHY



Ms. Ankita Thakur, currently studying M.E in VLSI and Embedded system at RMD Sinhgad School of Engg, Pune. The author has her personal field of interest in the domain of image processing, VLSI and Embedded system.