

Hand Posture to Text and its Implementation with Computer Vision

Satabda Das¹, Pranab Hazra², Debanjana Chall³, Subhadeep Karmakar⁴, Spandan Sarkar⁵,
Swastika Dey⁶, Ahana Chowdhury⁷, Mitadru Mukherjee⁸, Pankaj Ojha⁹,
Vishal Kumar Shaw¹⁰, Nishant Minz¹¹

^{1,3,4,5,6,7,8,9,10,11}Narula Institute of Technology, Kolkata, West Bengal, India.

²Assistant Professor, Narula Institute of Technology, Kolkata, West Bengal, India.

Email: ¹satabdadas3@gmail.com, ²pranabhazra2017@nit.ac.in,

³rajcumarchall3875@gmail.com, ⁴deep.subhak2y@gmail.com, ⁵ricksarkar70@gmail.com,

⁶swastikadey21@gmail.com, ⁷ahanachowdhury1506@gmail.com, ⁸mitadru727@gmail.com,

⁹pnkjojha4@gmail.com, ¹⁰vishalshaw666@gmail.com, ¹¹nishantminz2000@gmail.com

Abstract

One of the efficient ways of to avoid physical devices such as keyboards and mice are hand gesture controlled environment, where the user need to be located in a specific location to use these devices. The input feature plays a vital role in hand recognition, and the selection of good features representation. This paper deals with the hand posture and recognition method, as it is considered to be one of the challenging problem. Our main goal is to combine computer vision to our daily life to make optimal use of it. Some of the applications and system based on hand posture recognition are discussed in the explanation section hope it would help in succeeding the goal of this research area. In case of this we have tried to implement deep learning method called as convolution neural network(CNN) to make this project be more fruitful.

Index Terms—Computer vision, Convolution neural network(CNN), Hand gesture recognition, Human computer interaction (HCI).

I. INTRODUCTION

With the advancement of computer and computing experiences we currently interact with keyboard, mouse and pen are not natural enough for them. On PC platform, we get mixture of virtual and augmented reality which is more natural and efficient approach and requiring more natural and intuitive interface with lesser physical interference. For mobile or hand held devices, this freedom of experience is prohibited and user cannot exercise there interaction to the fullest as it is mostly embedded with tiny keyboard or touch screen. Hand gesture is frequently used in our daily life to convey message or to show our expression through body language. We are focusing on these part so that human can interact with computer with more efficiency and fluently.

Though terms like hand gesture and hand posture sounds similar in nature but there are some significant differences between them. Hand posture are simple hand signs in most

cases with specific meaning for example hand signs like victory sign, peace sign etc. [1] We can identify different movements of hand as dynamic movement and complex movement. Dynamic movement are recognized as simple hand movements like waving. Complex movement mainly contains wrist movement combined with fingers and hand movement and the change of hand orientations and positions. Example of this kind of movement is similar to sign language.

In this paper we are going to do a thorough search on the perspectives of hand movement and capturing there meanings through the computer vision enhanced by CNN and other features of deep learning.

II. LITERATURE REVIEW

Gesture is a body language through which human convey their feelings and express their emotions. [2] The different gestures of the five fingers may have their respective significant

meanings. Hand gesture recognition is a complex system that is composed of gesture modeling, gesture analysis and recognition, and machine learning. From previous works on hand gestures are- Hidden Markov Model (HMM) that was used to a real-time denotation level of Sign Language recognition system [9]. A gesture can be modeled as a HMM state sequence.

Using gesture denotation analysis is compatible for recognizing a sequence of gestures in doing a complex task, but it is insufficient to correctly identify gestures in a simple continuous motion. Recently deep learning is widely applied to many applications and CNN is one of the most popular and efficient option to handle image based deep learning implementations.

III. METHODOLOGY

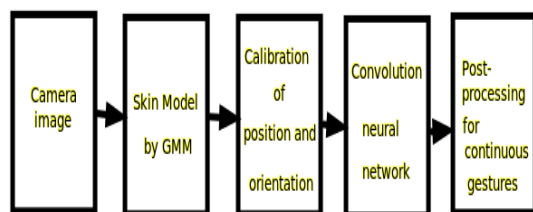


Fig.1 Framework of the purposed human gesture recognition

IV. SKIN SEGMENTATION

In computer vision, the source of light plays a vital role as it affects the system performance and its effectiveness.[22] Using color threshold to distinguish between skin and non skin parts are common practice, but color thresholds are not enough to describe the some properties of skin color under certain light conditions, though the YCbCrcolor space is less sensitive to the light condition than the RGB color space is used, the result is still defective. From Fig. 2, some of the pixels on the little fingers and the ring finger are seemed to be non-skin pixels.



Fig. 2 Histogram image of hand

V. HAND TRACKING

It is a challenging task to track the objects. Texture or appearance based methods have been improved to be more robust for the non-rigid objects.[6] Although shape based methods achieve better results for rigid objects, it is not suitable for the jointed objects, such as hand. Some approaches resort to background modeling with limitation of stationery camera. In case of tracking position of hand optical based detection is used as it is more effective.

VI. HAND SEGMENTATION

After detecting a hand ,the first performed by computer is capturing the color of hand from the neighborhood of features mean position. Considering trade-off between computational cost and accuracy of description, we Gaussian model to describe hand color in HSV color space. In Fig.3(a), [7] the bigger white dot denotes features' mean position. features within the circumference are used to get instant hand color models. Histogram method is typically used to detect skin as it excludes other body parts and objects. If histogram detects other objects then it will ignore it and cut the important sections from the image via image segmentation. In that case our method can get better results like Fig.3(b).

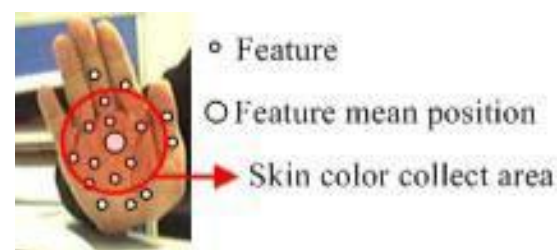


Fig .3(a) Mean position detection.



Fig . 3(b) Histogram detected pictures of hand.

VII. GESTURE AND POSTURE RECOGNITION

Gesture system can be classified as the following three states, 1>glove based vision 2>low level features and 3>vision based features. The most efficient one among these approaches is glove based vision but it comes with a drawback as it is quite costly, so other alternatives like vision recognition and gesture recognition mainly focuses on the algorithm and computer vision to identifying the shape of hand and work according to it to perform work suggested by the machine or robot.

A. 3D Hand Model based Approaches: There are many methods that have already been used to analysis, model and represent the shape and size of hand and a large data base .[26] In case of 3D hand modeling hand is observed as an articulated deformable object, therefore feature extraction cause a great deal of problem while running the algorithm.

B. Appearance based approaches: Known as View based approaches, it signifies that when ever the picture of a hand is taken using the intensity of 2d images it tries to define the gesture of views. Appearance based approach is much easier compared to 3D model based approaches, as a result there are many to search for alternative representation of the hands.

C. Low Level Features based Approaches: these approach is mainly based on mapping between the input video and the gestures. According to some researchers full hand reconstruction are not needed for gesture recognition system. [27] An example those features which can be extracted quickly are supposed to be corrupted with noise. Low level features include: the center of the hand region , an elliptical bounding parts of the hands, edges, fingers and depressions, moments, and histograms .



Fig. 4. Different hand positions captured.

VIII. APPLICATIONS OF HAND GESTURE RECOGNITION

Here are some of the well known fields of hand gesture recognition :

A. Sign Language Recognition

Sign language recognition is a problem that has been addressed in research for years. However, we are still far from finding a complete solution available in our society.

Among the works developed to address this problem, the majority of them have been based on basically two approaches: contact-based systems, such as sensor gloves; or vision-based systems, using only cameras.

This paper presents a prototype of a computer vision based translation system for sign language using convolution neural networks. The post is divided into three main parts: the system design, the data-set, and the deep learning model training and evaluation.

B. Robotics, Human Manipulation and Instruction

Its applicability is vast in health sectors as the can be used to build high quality prosthetic limbs capable of doing jobs resembling natural hands and legs. Various researches for robot control applications are implemented.

This display presents the template of a dual-cam first-person insight conveyance system for a sign language with the help of convolutional neural networks. The use of gestures to control the robots is correlated to the virtual reality interaction system.

C. Virtual Reality

Virtual Reality (VR) is computer generated world to give real world experience or beyond it, making the user feel they are immersed in their surroundings in other words it is perfect blend between physical and virtually created world. Based on the Wii controller hardware. The exact origins of virtual reality are disputed, partly because of how difficult it has been to formulate a definition for the concept of an alternative existence.[22] The development of perspective bought new era and created convincing depictions of spaces that did not exist.

D. Gesture-to-speech

gesture-to-speech is an application of modern days which help to convert any kind of body or hand gesture to speech. I can be very much useful in the medical field for hearing impaired

people, it would make their public interaction much more easier and efficient as communication will occur through computer devices surrounding them. Glove mapping devices to capture hand posture would be quite effective for this purpose and conversion gesture to speech with neural networks and algorithms.

E. Games

For computer games, are huge market based on entertainments. To add gesture recognition in these field will make it more enjoyable and give better real life experience to the player. Rautaray (2010) implemented earlier computer vision and gesture recognition in gaming world to enhance the process, and developed a low cost vision based vlc system.

IX. CONCLUSION

In this paper, we developed a CNN-based human hand gesture recognition system. The salient feature of the system is that there is no need to build a model for every gesture using hand features such as fingertips and contours. To have robust performance, we applied a GMM to learn the skin model and segment the hand area for recognition. Also, the calibration of the hand pose was used to rotate and shift the hand on the image to a neutral pose. Then, a CNN was trained to learn seven gesture types in this paper.

REFERENCES

1. A. Kojima, M. Izumi, T. Tamura, and K. Fukunaga, "Generating natural language description of human behavior from video images," in *Int. Conf. Pattern Recog.*, vol. 4. IEEE, 2000, pp. 728–731.
2. C. J. Cohen, F. Morelli, and K. A. Scott, "A surveillance system for the recognition of intent within individuals and crowds," in *IEEE. Conf. Technol. for Homeland Secur.* IEEE, 2008, pp. 559–565.
3. S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern. C*, vol. 37, no. 3, pp. 311–324, 2007.
4. C. Vogler and D. Metaxas, "ASL recognition based on a coupling between hmms and 3d motion analysis," in *Int. Conf. Computer Vision.* IEEE, 1998, pp. 363–369.
5. C. F. Bond Jr, A. Omar, A. Mahmoud, and R. N. Bonser, "Lie detection across cultures," *J. nonverbal behav.*, vol. 14, no. 3, pp. 189–204, 1990.
6. H. I. Lin, C. H. Cheng, and W. K. Chen, "Learning a pick-and-place robot task from human demonstration," in *Proc. Int. Conf. Automat. Control.* IEEE, 2013, pp. 312–317.
7. T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1371–1375, 1998.
8. J. Davis and M. Shah, "Visual gesture recognition," in *IEE Proc. Vision, Image and Signal Process.*, vol. 141, no. 2. IET, 1994, pp. 101–106.
9. M.-H. Yang and N. Ahuja, "Recognizing hand gestures using motion trajectories," in *Face Detection and Gesture Recognition for Human-Computer Interaction.* Springer, 2001, pp. 53–81.
10. K. Oka, Y. Sato, and H. Koike, "Real-time tracking of multiple finger-tips and gesture recognition for augmented desk interface systems," in *IEEE Int. Proc. Automat. Face and Gesture Recog.* IEEE, 2002, pp. 429–434.
11. Moni, M. A. & Ali, A. B. M. S. (2009). HMM based hand gesture recognition: A review on techniques and approaches. 2nd IEEE International Conference on Computer Science and Information Technology, (ICCSIT 2009). <http://dx.doi.org/10.1109/ICCSIT.2009.5234536>
12. Murakami, K., & Taguchi, H. (1999). Gesture recognition using recurrent neural networks. ACM, Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology CHI '91, pp. 237-242. <http://dx.doi.org/10.1145/108844.108900>
13. Murthy, G. R. S., & Jadon, R. S. (2009). A review of vision based hand gestures recognition. *International Journal of Information Technology and Knowledge Management*, 2(2), 405-410. Retrieved from http://www.csjournals.com/IJITKM/PDF/34-G.R.S.Murthy_R.S.Jadon.pdf

14. Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual interpretation of hand gestures for human -computer interaction: A review. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 19(7), 677- 695. <http://dx.doi.org/10.1109/34.598226>
15. Peter, H. P. (2011). *Image Segmentation*. (1sted.). India. (Part 2). *Image Segmentation Methods Image Segmentation through Clustering Based on Natural Computing Techniques*. Retrieved from www.scribd.com/doc/70993072/Image-Segmentation
16. Porta, M. (2002). Vision-based user interfaces: Methods and applications. Elsevier, *International Journal Human-Computer Studies*, 2002(57), 27-73. <http://dx.doi.org/10.1006/ijhc.1012>
17. Rautaray, S. S., & Agrawal, A. (2010). A vision based hand gesture interface for controlling VLC media player.
18. *International Journal of Computer Applications*, 10(7). Retrieved from www.ijcaonline.org/volume10/number7/pxc3872012.pdf
19. Rodriguez, O. P., Gonzalez, O. O., Avizzano, C. A., Ruffaldi, E., Vercelli, D., & Bergamasco, M. (2008). Development of a 3D real time gesture recognition methodology for virtual environment control. *IEEE Proceedings 17th International Symposium on Robot and Human Interactive Communication*, pp. 279-284. <http://dx.doi.org/10.1109/ROMAN.2008.4600679>
20. Stergiopoulou, E., & Papamarkos, N. (2009). Hand gesture recognition using a neural network shape fitting technique. *Elsevier Engineering Applications of Artificial Intelligence* 22, 1141-1158. <http://dx.doi.org/10.1016/j.engappai.2009.03.008>
21. Thomas, S., Benjamin, P., Niels, H., & Susanne, B. (2008). Gesture recognition with a Wii controller. *ACM, Proceedings of the Second International Conference on Tangible and Embedded Interaction (TEI'08)*, Germany. <http://dx.doi.org/10.1145/1347390.1347395>
22. Verma, R., & Dev, A. (2009). Vision based hand gesture recognition using finite state machines and fuzzy logic.
23. *IEEE International Conference on Ultra Modern Telecommunications & Workshops, ICUMT '09*, pp. 1-6,
24. Petersburg. <http://dx.doi.org/10.1109/ICUMT.2009.5345425>
25. Vogler, C., & Metaxas, D. (2001). A framework for recognizing the simultaneous aspects of American Sign Language. *Ideal Computer Vision and Image Understanding* 81, pp. 358-384. <http://dx.doi.org/10.1006/cviu.2000.0895>
26. Wang, H. H. X., Mandal, M. K., Meng, M., & Li, D. (2003). Efficient face and gesture recognition techniques for robot control. *IEEE Canadian Conference on Electrical and Computer Engineering, CCECE 2003*, 3, 1757-1762. <http://dx.doi.org/10.1109/CCECE.2003.1226250>
27. Xu, Z., Xiang, C., Lantz, V., & Kong-qiao, W. (2009). Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. *ACM Proceedings of the 14th international conference on Intelligent user interfaces IUI '09*. <http://dx.doi.org/doi:10.1145/1502650.1502708>
28. Yin, X., & Xie, M. (2007). Hand posture segmentation, Recognition and Application for Human-Robot Interaction. Retrieved from www.intechopen.com/download/pdf/pdfs_id/5276
29. Zhu, Y., Huang, Y., Xu, G., Ren, H., & Wcn, Z. (1998). Vision-based Interpretation of Hand Gestures by Modeling Appearance Changes in Image Sequences. *WA '98 IAPH Workshop on Machine Vision Applications. Makuhm. Japan*. Retrieved from www.mva-org.jp/Proceedings/.../1998/papers/1998573.pdf