Human Gesture Recognition in Computer Vision Research

Zepei Zheng[1*]

[1]College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310023, China

ABSTRACT: Human gesture recognition is a popular issue in the studies of computer vision, since it provides technological expertise required to advance the interaction between people and computers, virtual environments, smart surveillance, motion tracking, as well as other domains. Extraction of the human skeleton is a rather typical gesture recognition approach using existing technologies based on two-dimensional human gesture detection. Likewise, I t cannot be overlooked that objects in the surrounding environment give some information about human gestures. To semantically recognize the posture of the human body, the logic system presented in this research integrates the components recognized in the visual environment alongside the human skeletal position. In principle, it can improve the precision of recognizing postures and semantically represent peoples' actions. As such, the paper suggests a potential and notion for recognizing human gestures, as well as increasing the quantity of information offered through analysis of images to enhance interaction between humans and computers.

1. INTRODUCTION

Gesture recognition is an essential aspect of computer science that aims to interpret human gestures using algorithms. Gesture recognition based on computer vision allows users to interact more naturally with technology [1]. It has the benefit of being less impacted by the surroundings. Users may connect with computers anywhere, without being constrained, and effectively allowing computers to grasp human instructions precisely and quickly. The procedures do not necessitate the use of any mechanical devices [2]. Throughout the interaction between computers and people, not only are gestures instant, but also intuitive and vivid, adaptable, as well as visible. This allows for completion of interactions seamlessly to cover the deficits betwixt virtual and reality [3].

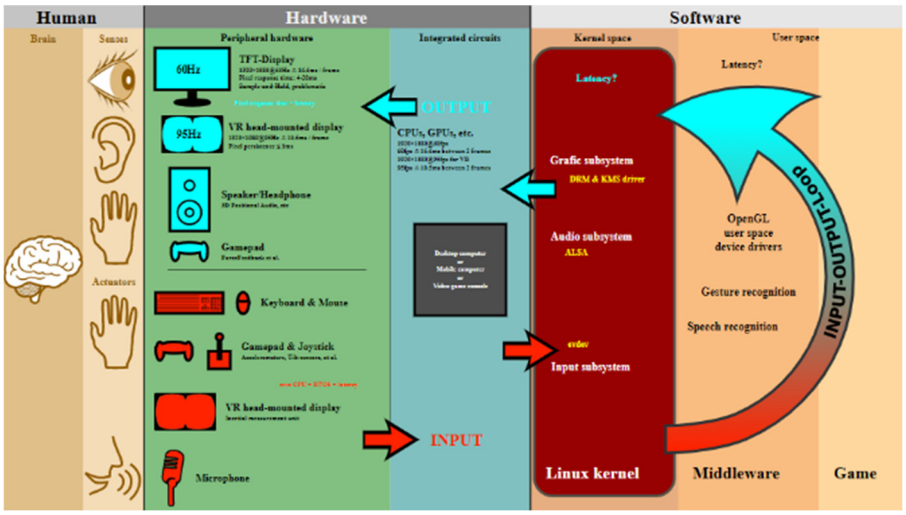

Figure 1. Gesture detection process.

Initial gesture recognition straightforwardly recognized the place of the hand's individual joints through wearable gadgets, including relaying the data to the PC via wired transmission, precisely putting away the client's hand movement data. For instance, data gloves among other gear, albeit the discovery impact is excellent, yet they remain costly and badly designed to utilize.

Accordingly, the optical checking technique utilizes infrared light for distinguishing the positioning plus hand development, which dislodged the data glove. Likewise, it has a great impact, yet requires more convoluted hardware.

*Corresponding author. Email: zzp18857109315@163.com

Albeit higher precision can be acquired by the assistance of outer gadgets, it is costly and influences the client's activities somewhat. The motion acknowledgment in light of PC vision alludes to the handling of the video information gathered using the camera via the computation for identifying gestures, which accomplishes the motivation behind detection of gestures besides having turned into an exploration area of interest lately.

2. STATE OF THE ART IN GESTURE RECOGNITION

Gesture detection using computer vision has recently emerged as a study focus in the community of computer vision applications. De Oliveira Junior [4] employed Hidden Markov Model or HMM to recognize single motions in video, achieving 94 percent identification accuracy across 262 movements. Reyes [5] devised a Dynamic Time Warping or DTW gesture detection approach that accomplishes the detection of depth gesture pictures for video. SimoSerra and others [6] completed gesture detection by imposing physical limitations upon the joint sites of the hands. Sinha [7] suggested a technique based on Matrix Completion [8] that may be used for vast-scaled, real-time gesture posture prediction without the need of a GPU.

China has also made some tremendous growth as far as computer vision-based gesture recognition goes [9], including the advancement of a technique of gesture detection based on perceived changes in picture transformation, and employed a discretization parameter model of visual picture in motion for identification of 120 movements. Also, a solution was presented for solving dynamic gesture identification via self-discovering sparse representation. The approach processes the raw picture directly, minus extraction of features processing in real-time. Thereafter, a technique for detecting static motions with the finger was devised. HOG extracted the features in order to generate a binary image [10].

The number of fingers was determined using a flexible logical procedure before being categorized using Support Vector Machine or SVM. The approach integrates the binary image with gray image methods, along with the number of fingers can be lowered to minimize the spectrum of gestures that can be identified [11]. The precision of 25 motions on that set of data was around 99 percent, however set of data was rather basic and not used, therefore sophisticated gestures were not used for validation. De Oliveira Junior introduced a recognition technique for identifying gestures in real-time that combines matching of optical flow with the AdaBoost algorithm with the sole involvement being the reading of 2D video clips to provide precise findings.

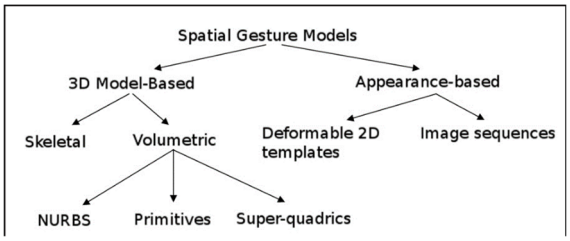

Figure 2. Different gesture detection techniques.

Many studies on gesture recognition have been conducted through various frameworks to tackle gesture difficulties. Research has been carried out, for instance, on aspects like static, dynamic, continuous, static, as well as isolated gestures. However, there are still issues with limited adaptability as well as a heavy reliance on sets of data.

3. SIGNIFICANT GESTURE RECOGNITION TECHNOLOGIES BASED ON COMPUTER VISION

Computer-based vision-oriented recognition of gestures captures an image of the hand using cameras, before using image processing as well as machine learning (ML) to add to the assessment and identification of the gesture in a complete manner. Presently, computer graphics-based gesture recognition is divided into three phases, including picture gathering, detection of the body or hand and segmentation, and gesture identification and categorization [11].

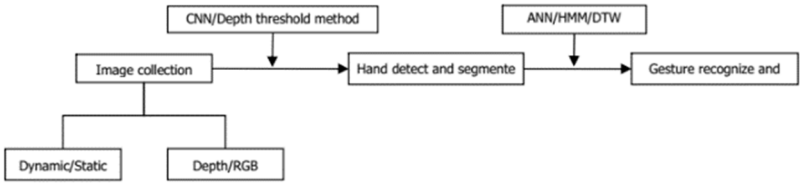

Figure 3. Gesture recognition process.

3.1. Gesture detection and categorization

The collection of images follows RBG images and depth images, where the former are obtained with a standard camera. Depth images, on the other hand, are obtained using depth camera like Leap Motion and Kinect; enables simultaneous collection of depth- plus RBG-images; depth images facilitate the detection of a segment of the space-based data, useful for identification and categorization of gesture. Challenges faced while gathering pictures include occlusion, varying light intensities and varying light directions, raises the bar for the algorithm's resilience. The increased practicability in the field of gesture recognition has seen a spike in algorithms looking to overcome invariances of light whilst arresting occlusion issues.

3.1.1.CNN-Based gesture categorization

Convolutional neural network-based gesture categorization involves optimizing using convolutional neural networks (CNN), such as Full Convolutional Neural Networks (FCN) or SegNet, with the two approaches substituting a deconvolution layer for the convolutional neural networks' last layer, whilst the picture gets returned to its initial size through sampling up, where individual pixels are anticipated. In comparison to CNN, both SegNet and FCN may take various sized pictures, eliminating the need for a standard size of all pictures and avoiding the difficulty of recurrent storage as well as convolution computations. Kumar [12] improved categorization accuracy by combining the complete CNN with context semantics. As such, the CNN categorization approach is adaptable, along with the capacity to onboard a range of diverse models to execute segmentation of gestures.

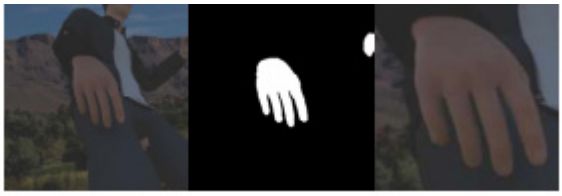

Figure 4. Hand identification and categorization.

3.1.2 Categorization of gestures via depth threshold technique

The depth threshold approach calculates the distance betwixt individual pixels with the camera based on the distance involved betwixt the target with the camera within the depth picture, before extracting a snapshot through a distance inside a preset range. For improved extraction of the body range, the body's depth range within the depth picture is defined. Alternatively, the body could be regarded as the nearest item to the camera. The above technique increases the pre-processing impact, acquires a far more accurate body area, whilst enhancing the accuracy of recognizing gestures. Whereas the depth threshold approach may execute picture categorization quickly and easily, it poses a significant limitation on user behavior and a limited expansion space.

3.2. Numbering Detection and Classification of Gestures

There are two types of gesture recognition, including dynamic recognition as well as static recognition. The latter involves motions from a single image, whereas the former entails variations in gesture motions through time, i.e. numerous successive stationary gestures [13]. The picture for gesture recognition gets separated into three parts, involving depth map, RGB map, and RGB-D map. For the depth map, it can indicate the space betwixt the camera with the item directly. The portrayal of the depth map is identical to the gray picture, save for the latter depicting the proximity of the item to the camera with each pixel. For the RGB-D picture, it includes RGB 3-channel RGB pictures as well as depth images. Whereas the two photos appear to be distinct, the pixels have a one-to-one connection [14].

Since recently, the majority of gesture recognition algorithms have relied on artificial neural networks (ANN) such as GAN, RNN, or ANN-CNN, along with DTW and HMM.

3.2.1 Neural network-based methods

According to Zhang and Deng [15], the popularity of neural network-based approaches, particularly those that rely on CNN algorithms have increased. The identification accuracy is great, the resilience is excellent, while still providing for usage in both static and dynamic settings, as well as depth, RGB picture, along with other variables. As a gesture detection algorithm that is still evolving, it is seeing increased application through cyclic neural network alongside its produced confrontation network.

3.2.2 The DTW algorithm

Not only is the approach simple, but also quick with the goal of locating the ideal path and then identifying the comparative sequence according to the least overhead based on the optimal path. HMM searches for concealed sequences within visible sequences so as to discover the meaning communicated through gestures. Initially, CNN was utilized for image categorization. It may be utilized to extract features as well as reducing dimensions during detection of gestures along with instant classification. As such, DTW makes for a a more adaptable approach, although it needs a huge amount of labeled data for training, making the training time-consuming besides requiring GPU acceleration.

3.2.3 HMM Algorithm

For a long time, HMM has been utilized as a standard approach in statistics and probability, despite its initial application being in the realm of voice detection. Moreover, HMM has recently made significant progress as far as gesture recognition goes. HMM-oriented gesture detection necessitates the creation of an HMM for individual gestures, a computationally-demanding ordeal which also comes with real-time performance effects.

4. TRENDS AND CHALLENGES

With over two decades of development, gesture recognition that relies on computer vision technology has achieved significant milestones, going through the optical flow technique, the SVM segmentation approach, the method for constraint addition, the HMM methodology, the DTW algorithm, as well as the ANN algorithm. Nevertheless, constraints still exist in gesture recognition along with some evolutionary obstacles. The issues are discussed below.

The backdrop has a significant impact. In complicated backdrops, excellent gesture categorization is still

necessary, much as traditional picture classification. During recognition of gestures, the ability for effectively segregating the gesture from the backdrop is critical to enhancing the precision of recognition. Gestures are likely to be obstructed by environmental objects during dynamic gesture recognition, making it more challenging to track gestures. Diverse motions are similar, while the same gesture can be distinct. The body has several degrees of flexibility. Several degrees of freedom exist within the posture activity space, making it hard for different algorithms to execute satisfactory calculations for individual freedom levels, besides being time-consuming to compute the multiple freedom levels. This increases the challenge of detection in real-time.

Various angles of observing alongside diverse light intensities during identification of gestures is more challenging to overcome.

The neural network approach tends to be slow, precise, and more data-dependent. As such, it needs a significant volume of tagged data along with a strong computational speed to make sure it does not satisfy real-time needs. The DTW approach is quicker compared to the HMM technique, although it lacks the precision plus model resilience when pitted against the neural network as illustrated in table 1 below.

Table 1. Comparing various gesture recognition approaches

Algorithm	Number of samples required	Calculate speed	Recognition accuracy	Number of papers recent years
ANN	Extremely large	Extremely slow	High	Many
DTW	Small	Faster	Low	A few
HMM	Large	Middle	Middle	A few

5. CONCLUSION

This research examines computer vision-based gesture recognition, having overcome the wearable restrictions after almost two decades of work. However, issues, including poor uniformity, sensitivity to lighting variations as well as occlusion, and unreliable performance in real-time still exist. Certainly, increasing computer computational speed can alleviate the challenge of unreliable performance in real-time detection. The impact of inadequate uniformity and lighting variations may be mitigated by algorithm progress, although the problem of obstruction requires long-term study.

REFERENCES

1. Lai, K., & Yanushkevich, S. N. (2018). CNN+RNN depth and skeleton based dynamic hand gesture recognition. 2018 24th International Conference on Pattern Recognition (ICPR). https://doi.org/10.1109/icpr.2018.8545718
2. He, X., & Zhang, J. (2020). Design and implementation of number gesture recognition system based on Kinect. 2020 39th Chinese Control Conference (CCC). https://doi.org/10.23919/ccc50068.2020.9189566
3. Yang, F., Sun, Q., Jin, H., & Zhou, Z. (2020). Superpixel segmentation with fully Convolutional networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.01398
4. De Oliveira Junior, L. A., Medeiros, H. R., Macedo, D., Zanchettin, C., Oliveira, A. L., & Ludermir, T. (2018). SegNetRes-CRF: A deep Convolutional encoder-decoder architecture for semantic image segmentation. 2018 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn.2018.8489376
5. Reyes, M., Dominguez, G., & Escalera, S. (2011). Featureweighting in dynamic timewarping for gesture recognition in depth data. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). https://doi.org/10.1109/iccvw.2011.6130384
6. Simo-Serra, E., Ramisa, A., Alenya, G., Torras, C., & Moreno-Noguer, F. (2012). Single image 3D human pose estimation from noisy observations. 2012 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2012.6247988
7. Sinha, A., Choi, C., & Ramani, K. (2016). DeepHand: Robust hand pose estimation by completing a matrix imputed with deep features. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2016.450
8. Zhang, X., Wang, J., Wang, X., & Ma, X. (2016). Improvement of dynamic hand gesture recognition based on HMM algorithm. 2016 International Conference on Information System and Artificial Intelligence (ISAI). https://doi.org/10.1109/isai.2016.0091
9. Cicirelli, G., & D'Orazio, T. (2017). Gesture recognition by using depth data: Comparison of different methodologies. Motion Tracking and Gesture Recognition. https://doi.org/10.5772/68118
10. Cui, H., & Wang, Y. (2020). Research on gesture recognition method based on computer vision technology. 2020 International Conference on Computer Information and Big Data Applications (CIBDA). https://doi.org/10.1109/cibda50819.2020.00087
11. Zhao, D., Liu, Y., & Li, G. (2018). Skeleton-based dynamic hand gesture recognition using 3D depth

data. Electronic Imaging, 30(18), 461-1-461-8. https://doi.org/10.2352/issn.2470-1173.2018.18.3dipm-461

12. Kumar, V., Namboodiri, A., Paluri, M., & Jawahar, C. V. (2017). Pose-aware person recognition. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2017.719
13. Martínez-Hinarejos, C., & Parcheta, Z. (2017). Spanish sign language recognition with different topology hidden Markov models. Interspeech 2017. https://doi.org/10.21437/interspeech.2017-275
14. Zhang, C., & Tian, Y. (2013). Edge enhanced depth motion map for dynamic hand gesture recognition. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/cvprw.2013.80
15. Zhang, Q., & Deng, F. (2017). Dynamic gesture recognition based on LeapMotion and HMM-CART model. Journal of Physics: Conference Series, 910, 012037. https://doi.org/10.1088/1742-6596/910/1/012037