

Wearable Sensors and Computer-Vision-Based Methods

Subjects: Computer Science, Artificial Intelligence

Contributor: zhidong xue

Real-time sensing and modeling of the human body, especially the hands, is an important research endeavor for various applicative purposes such as in natural human computer interactions. Hand pose estimation is a big academic and technical challenge due to the complex structure and dexterous movement of human hands. Boosted by advancements from both hardware and artificial intelligence, various prototypes of data gloves and computer-vision-based methods have been proposed for accurate and rapid hand pose estimation in recent years. However, existing reviews either focused on data gloves or on vision methods or were even based on a particular type of camera, such as the depth camera. The purpose of this survey is to conduct a comprehensive and timely review of recent research advances in sensor-based hand pose estimation, including wearable and vision-based solutions. Hand kinematic models are firstly discussed. An in-depth review is conducted on data gloves and vision-based sensor systems with corresponding modeling methods. Particularly, this review also discusses deep-learning-based methods, which are very promising in hand pose estimation. Moreover, the advantages and drawbacks of the current hand gesture estimation methods, the applicative scope, and related challenges are also discussed.

Keywords: human–computer interaction ; computer vision ; data gloves ; hand pose estimation ; deep learning ; wearable devices

1. Description

With the rapid growth of computer science and related fields, the way that humans interact with computers has evolved towards a more natural and ubiquitous form. Various technologies have been developed to capture users' facial expressions as well as body movements and postures to serve two types of applications: information captured becomes a “snapshot” of a user for computers to better understand users' intentions or emotional states; and users apply natural movements instead of using dedicated input devices to send commands for system control or to interact with digital content in a virtual environment.

Among all body parts, we depend heavily on our hands to manipulate objects and communicate with other people in daily life, since hands are dexterous and effective tools with highly developed sensory and motor structures. Therefore, the hand is a critical component for natural human–computer interactions, and many efforts have been made to integrate our hands in the interaction loop for more convenient and comfortable interactive experiences, especially in a multimodal context as demonstrated in the “put-that-there” demonstration [1].

Driven by applications like sign language interpretation and gesture-based system control, hand gesture recognition has been extensively studied from early on and there exist many comprehensive reviews [2][3][4][5][6]. Hand gestures, either static or dynamic, can now be successfully recognized if the gesture categories are well defined with proper inter-class distances. Many consumer-level applications, such as the gesture control on Microsoft HoloLens [7], can already provide robust recognition performance. Nevertheless, despite sharing some common points with gesture recognition, accurate hand pose estimation of all hand joint, remains a challenging problem.

With the emergence of low-cost depth sensors such as Microsoft Kinect [8] and Intel RealSense [9], and also the boost of machine learning methods, especially the rapid development of convolutional neural networks, there has been considerable progress in hand pose estimation, and state-of-the-art methods can now achieve good performance in a controlled environment. However, hand pose estimation has had much less attention in the literature compared to the recognition.

Hand pose estimation can be roughly put into two categories based on the corresponding sensing hardware: wearable sensors and vision-based sensors. While glove-shaped wearable sensors are mostly self-contained and portable, vision-based sensors are very popular since they are more affordable and allow unconstrained finger movements. Both types of devices find their usage under certain circumstances and are still in constant development.

Existing hand pose estimation systems can already accurately track the movement of the human hand in real time in a relatively controllable environment. However, hand pose estimation cannot yet be considered as a solved problem and still faces many challenges, especially in open and complex environments, where we should take the amount of computing resources needed into consideration.

2. Wearable Sensors

Wearable sensors, or data gloves, are promising for accurate and disturbance-free hand modeling since they generally have compact design and become lighter and less cumbersome for dexterous hand movements. However, there are three main challenges remaining to be solved. Most data gloves are still “in the lab” and there is no industrial standard on the design and fabrication of such devices, which leads to high costs of available commercial products, making them unaffordable for daily use. Second, except gloves that are based on stretch sensors, most gloves have fixed size and are difficult to match different users’ hands. Lastly, gloves are unsuitable to be used in certain cases, for example, some stroke patients have difficulties opening their hands to wear gloves designed for normal users, or in situations when the user needs to manipulate tiny objects, or put their hands into water, etc.

3. Vision-based Methods

Vision-based methods have overcome many difficulties faced by common computer vision tasks, such as rotation, scale and illumination invariance, and cluttered backgrounds. The high dimensional nature of hand pose representation, and even hand self-occlusion, are no longer obstacles in the way of achieving accurate hand pose estimation in real time. However, vision-based methods still face the following challenges:

First, occlusion is still the major problem. As the hands are extensively used to manipulate objects in daily life, they are very likely to be blocked or partially blocked by objects during interaction, which forms the hand–object–interaction (HOI) problem. There are already some efforts to deal with object occlusion. For example, Tekin et al. [11] proposed an end-to-end architecture to jointly estimate the 3D hand and object poses from egocentric RGB images. Myanganbayar et al. [12] proposed a challenging dataset consisting of hands interacting with 148 objects as a novel benchmark for HOI.

Second, since many methods are data-driven, the quality and coverage of training datasets is of great importance. As discussed in Section 4.4, there are already many useful datasets with 2D/3D annotations. However, a larger portion of annotated data comes from synthetic simulations. Existing methods tried to employ weakly supervised learning, transfer learning, or different data augmentation approaches to better cope with insufficiency of real world data, but more data representing tremendous viewpoints, shapes, illumination, background variations, and objects in interaction are required to train deep learning-based architecture, or we must find a new way to incorporate the hand model for 3D pose recovery.

Moreover, most deep learning-based methods also require large amounts of computational resources during the training and inference stages. Many algorithms need to run on a graphics processing unit (GPU) to achieve a real-time frame rate, making it difficult to be deployed to portable devices such as mobile phones and tablets. Thus, it is important to find effective and efficient solutions on mobile platforms for ubiquitous applications.

4. Future Work

In the near future, expertise from material science and electronics is needed to build easy to wear and maintain, yet more affordable data gloves for accurate hand modeling. Regarding vision-based methods, data-efficient methods such as weakly supervised learning or hybrid methods are needed to minimize the dependency on large hand pose datasets and to improve the generalization ability to unseen situations. Moreover, we can already see the benefits of new sensors, e.g., the depth sensor, as they can largely reduce the computation complexity by using 2D data to deduce 3D poses; thus, novel accurate long-range 3D sensors will definitely contribute to contactless hand pose estimation.

References

1. Bolt, R.A.; “Put-That-There”: Voice and Gesture at the Graphics Interface. *Graph.* **1980**, *14*, 262–270, .

2. Haitham Sabah Hasan; Sameem Abdul Kareem; Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey. *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)* **2012**, , 55-60, [10.1109/acsat.2012.37](#).
3. Lingchen Chen; Feng Wang; Hui Deng; Kaifan Ji; A Survey on Hand Gesture Recognition. *2013 International Conference on Computer Sciences and Applications* **2013**, , 313-316, [10.1109/csa.2013.79](#).
4. Ahmad Sami Al-Shamayleh; Rodina Ahmad; Mohammad A. M. Abushariah; Khubaib Amjad Alam; Nazean Jomhari; A systematic literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications* **2018**, 77, 28121-28184, [10.1007/s11042-018-5971-z](#).
5. Ming Jin Cheok; Zaid Omar; Mohamed Hisham Jaward; A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics* **2017**, 10, 131-153, [10.1007/s13042-017-0705-5](#).
6. Jisun Park; Yong Jin; Seoungjae Cho; Yunsick Sung; Kyungeun Cho; Advanced Machine Learning for Gesture Learning and Recognition Based on Intelligent Big Data of Heterogeneous Sensors. *Symmetry* **2019**, 11, 929, [10.3390/sym11070929](#).
7. Hololens 2 From Microsoft. Available online: <https://www.microsoft.com/en-us/hololens/> (accessed on 2 February 2020).
8. Kinect V2, Microsoft. Available online: <http://www.k4w.cn/> (accessed on 2 February 2020).
9. Realsense Cameras, Intel. Available online: <https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html> (accessed on 2 February 2020).
10. D.J. Sturman; D. Zeltzer; A survey of glove-based input. *IEEE Engineering in Medicine and Biology Magazine* **1994**, 14, 30-39, [10.1109/38.250916](#).
11. Tekin, B.; Bogo, F.; Pollefeys, M. H+ O: Unified Egocentric Recognition of 3D Hand-Object Poses and Interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
12. Myanganbayar, B.; Mata, C.; Dekel, G.; Katz, B.; Ben-Yosef, G.; Barbu, A. Partially Occluded Hands: A Challenging New Dataset for Single-Image Hand Pose Estimation. In Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018), Perth, Australia, 2–6 December 2018.
13. James S. Supancic; Gregory Rogez; Yi Yang; Jamie Shotton; Deva Ramanan; Depth-Based Hand Pose Estimation: Data, Methods, and Challenges. *2015 IEEE International Conference on Computer Vision (ICCV)* **2015**, , 1868-1876, [10.1109/iccv.2015.217](#).
14. Rui Li; Zhenyu Liu; Jianrong Tan; A survey on 3D hand pose estimation: Cameras, methods, and datasets. *Pattern Recognition* **2019**, 93, 251-272, [10.1016/j.patcog.2019.04.026](#).
15. Tekin, B.; Bogo, F.; Pollefeys, M. H+ O: Unified Egocentric Recognition of 3D Hand-Object Poses and Interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
16. Myanganbayar, B.; Mata, C.; Dekel, G.; Katz, B.; Ben-Yosef, G.; Barbu, A. Partially Occluded Hands: A Challenging New Dataset for Single-Image Hand Pose Estimation. In Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018), Perth, Australia, 2–6 December 2018.