I J C T A, 8(5), 2015, pp. 2071-2076
© International Science Press

Microsoft Kinect in Gesture Recognition: A Short Review

Soumi Paul*, Subhadip Basu*, Mita Nasipuri*

Abstract: Microsoft Kinect sensor is a low cost, high-resolution, depth and visual (RGB) sensing device. It became popular in a very short span for widespread use. The depth and visual information (RGB-D) together provided by the Kinect sensor opens up new opportunities for computer vision. This paper contains an overview of evolution of different versions of Kinect and highlights the differences of their key features. It also reviews the use of Kinect v1 and v2 in gesture recognition, an important field of computer vision. Finally, a summary of the challenges in this field and future research trends is provided.

Index Terms: Computer Vision, Gesture Recognition, Kinect v1 and v2, Sign Language.

1. INTRODUCTION

Microsoft Kinect is an RGB-D sensor providing synchronized colour and depth images. It was initially launched by Microsoft as an input device for the Xbox game console. Recently, computer vision research community has discovered that the depth sensing technology of Kinect can be extended far beyond gaming and at a much lower cost than traditional 3-D cameras, such as stereo cameras (https://en.wikipedia.org/wiki/Stereo_camera) and time-of-flight (TOF) cameras[1]. In just two years after Kinect was released, a large number of scientific papers, technical demonstrations have started appearing in diverse publication venues.

There exist a few review papers related to Kinect-based research [2]–[4]. The objective of the paper [2] is to reveal the smart technologies encoded in Kinect, such as sensor calibration, human skeletal tracking and facial-expression tracking. It also exhibit a prototype system that employs multiple Kinects in an immersive teleconferencing application. Another long paper [3] tries to give insights on how researchers exploit and improve computer vision algorithms using Kinect. The work of Zenaro et al. [4] aims at comparing the performance of Kinect v1 and v2 and their effectiveness in different applications exploiting depth data.

However, there does not exist any review on the differences of the general technical features of Kinect v1 and v2 and their suitability in different application domains. In this paper, we highlight this important aspect. Further, we present a review of state-of-the-art research using Kinect in gesture recognition, a prime research area of computer vision. Note that there exist a few very good survey articles covering prominent research on gesture recognition [5]–[8], however, all of these predate the era of Kinect.

Note that the paper [3] is a large review article containing diverse research in the vast field of computer vision. Though it touches some works on gesture recognition, it does not cover the recent state-of-the-art results. On the other hand, the paper [4] focuses only on 3D reconstruction and people tracking. Thus, the key contribution of our work is that it is an up-to-date review paper concentrating on gesture recognition.

The rest of the paper is organized as follows. First in Section 2, we discuss the components of the Kinect sensor *v*1 vs. Kinect *v*2 taking both hardware and software into account. The idea is to find out what

* Department of Computer Science & Engineering, Jadavpur University, India, *Email: soumipaul.work@gmail.com, subhadip@cse.jdvu.ac.in, mnasipuri@cse.jdvu.ac.in*

technology Kinect use to capture the depth image and what advantages the Kinect have compared to other conventional RGB cameras in the market. We also discuss how they have even improved from *v*1 to *v*2 over the years. In Section 3, we outline the research trends in gesture recognition using depth sensor. Section 4, abbreviated the recent works and future scopes using the new version of depth sensor Kinect *v*2. Finally, in Section 5, we have drawn our conclusion.

2. EVOLUTION OF KINECT TECHNOLOGY

In this section, we go through the technical details of Kinect and highlight the differences between version 1 and 2.

2.1. Kinect Hardware

The Kinect sensor, the first low cost depth camera, was introduced by Microsoft in November 2010. Firstly, it was typically a motion controlled game playing device. Then it was extended a new version for windows. Here in this section, we will discuss the evolution of Kinect from v1 to the recent version v2.

2.1.1. Kinect v1

Microsoft Kinect v1 (Fig. 1) was released in February 2012 and started competing with several other motion controllers available in the market. The hardware of Kinect consists of a sensor bar that comprises of 3D depth sensors, an RGB camera, a multi-array microphone and a motorized pivot. The sensor provides full body 3D motion capture, facial recognition and voice recognition. The depth sensor consists of an IR projector and an IR camera, which is a monochrome complementary metal-oxide semiconductor (CMOS) sensor. The depth-sensing technology is from PrimeSense, an Israeli company.

The IR projector projects IR laser which passes through a diffraction grating and turns into a set of IR dots. The projected dots into the 3D scene is invisible to the color camera but is visible to IR camera. The relative left-right translation of the dot pattern gives the depth of a point [9].

2.1.2. Kinect v2

Microsoft Kinect v1 got an upgradation to v2 (in Fig. 2) in November 2013. The second generation Kinect v2 is completely different based on its ToF technology [1]. Its basic principle is, an array of emitters send out a modulated signal that travels to the measured point, gets reflected and received by the CCD of the sensor. The sensor acquires a 512 × 424 depth map and a 1920 × 1080 RGB image at the rate of 15 to 30 frames per second [1][10].

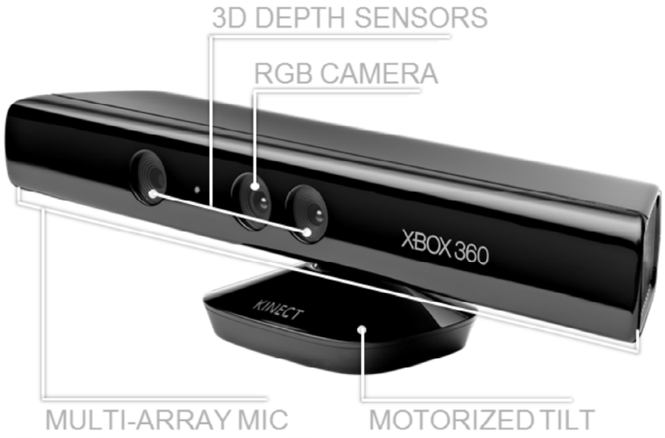

Figure 1: Microsoft Kinect v1

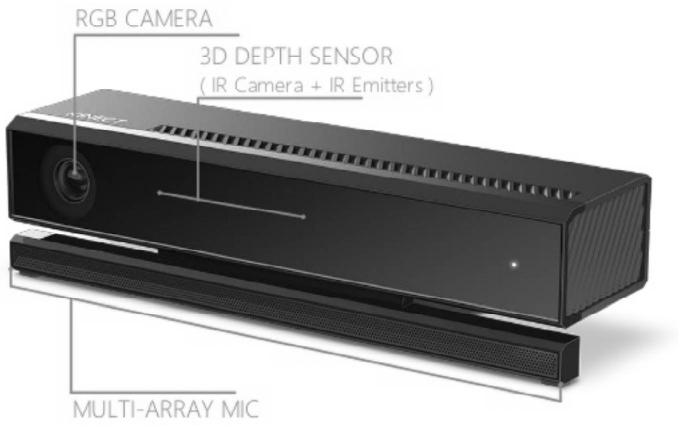

Figure 2: Microsoft Kinect v2

2.2. Kinect Software

Microsoft Kinect includes free downloadable software, which is Kinect development library tool. Currently, there are several available tools in the market, like OpenNI [8], Microsoft Kinect SDK [9] and OpenKinect (LibFreeNect) [10]. OpenNI comes with a middleware called NITE, and its highest version is 2.2. Microsoft Kinect SDK is released by Microsoft, and its current version is 2.0.

OpenKinect is a free, open source library maintained by an open community of Kinect people. Majority of users are uses first two libraries, which is OpenNI and Microsoft SDK. The Microsoft SDK (version 2.0) is only available for Windows whereas OpenNI (version 2.2) is a multiplatform and open-source tool.

With the help of above mentioned software's, Kinect is capable to capture mainly four type of images, RGB, depth, infrared and skeleton (Fig. 3).

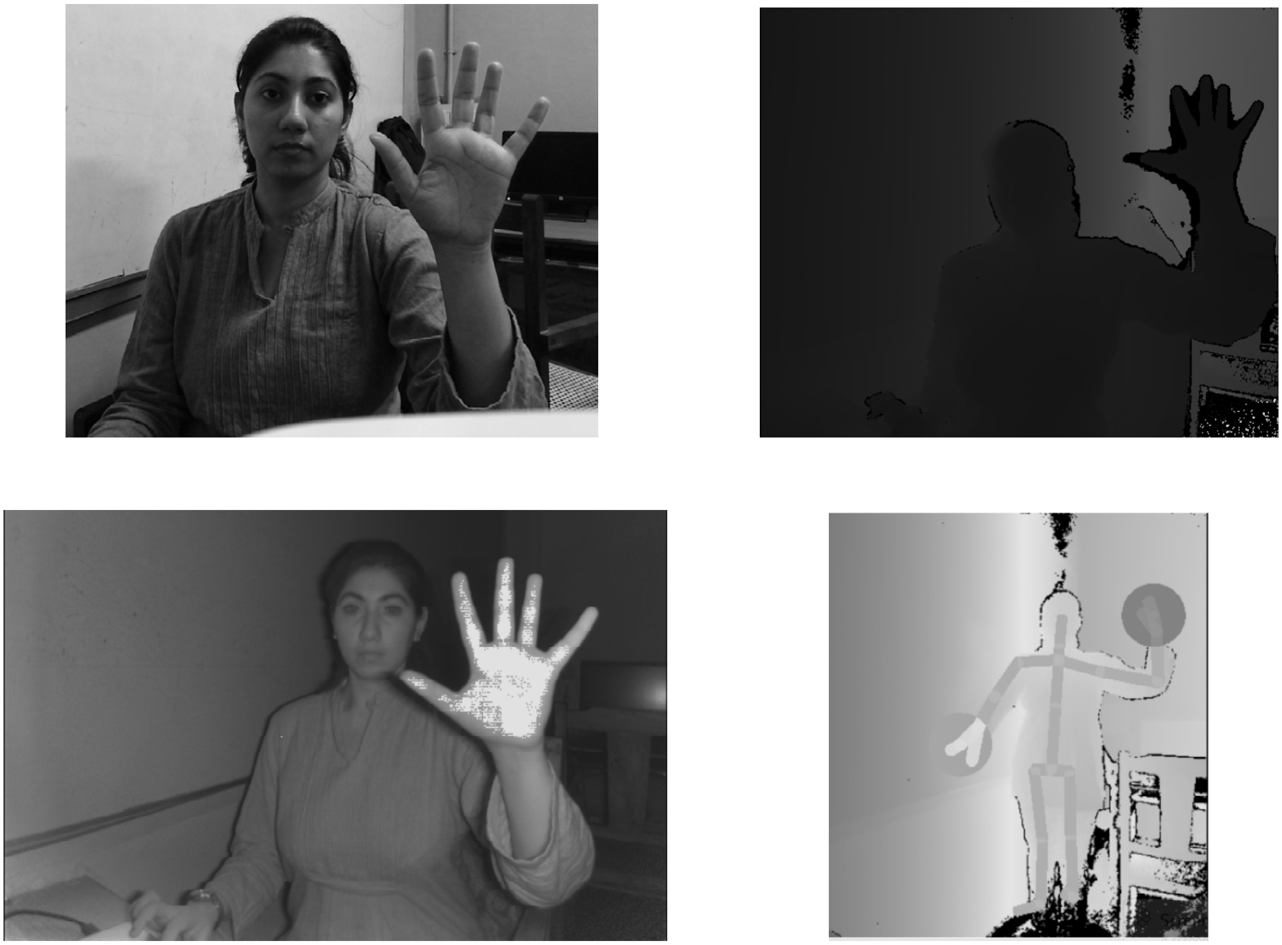

Figure 3: Microsoft Kinect captured RGB, Depth, Infrared, Skeleton (from left to right) images

2.3. Comparison between Version 1 and 2

In Table 1, we have compared the specification of Microsoft Kinect *v*1 and Kinect *v*2. From technical standpoint, we can see that Kinect *v*1 was a huge improvement over normal RGB cameras whereas, another big upgradation has been made from Kinect *v*1 to *v*2. The main shortcoming of Kinect v1 was its resolution 640 × 480 which got a boost in version 2 with 1920 × 1080. Not only that, the field of view has been expanded, skeleton joint point has been upgraded to 25 and most importantly with USB 3.0 the speed has been increased to get more support for real time applications.

Table 1
Comparison of the key features of Microsoft Kinect *v*1 and *v*2

Features	*Kinect for Windows v1*	*Kinect for windows v2*
Color Camera	640 × 480 @ 30 fps	1920 x 1080 @30 fps
Depth Camera	320 × 240	512 × 424
Max Depth Distance	~ 4.5 M	~ 4.5 M
Min Depth Distance	40 cm in near mode	50 cm
Horizontal Field of View	57 degree	70 degree
Vertical Field of View	43 degree	60 degree
Tilt Motor	Yes	No
Skeleton Joints Defined	20 joints	25 joints
Full Skeletons Tracked	2	6
USB Standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8-8.1 (WSA)
Horizontal Field of View	57 Degree	70 degree

3. KINECT IN GESTURE RECOGNITION

Gesture recognition is an important task now a days. Under Human–Machine Interaction, lots of work is being done to understand gesture with the help of machine. Gesture recognition can be broadly categorized into Gesture Detection, Pose Estimation and Gesture Classification.

To recognize gesture, Wilson et al. has proposed an approach to extend the standard hidden Markov model method of gesture recognition by including a global parametric variation in the output probabilities of the HMM states. They have also formulated an expectation-maximization (EM) method for training the parametric HMM [11]. Hand gesture detection is even a critical subcategory comes under Gesture detection. As hand is a smaller part respect to the whole body, so detection and classification of hand gesture is even more complex. Ren et al. [12] proposed to use a novel part based hand gesture recognition system which uses Finger-Earth Mover's Distance as a distance metric to measure the dissimilarity between hand shapes. On the other hand, Wang et al. [13] proposed another new super-pixel based earth mover's distance, which is not only robust to distortion and articulation, but also invariant to scaling, translation and rotation with proper preprocessing. Another hand pose tracking is done by Liang [14], where he proposes a Superpixel-Markov Random Field (SMRF) parsing scheme to extract a high-level description of the hand from the depth image. Poularakis et al. [15] in his work, uses Fourier Descriptors to locate apex–shaped structures in a hand contour and deals with partially merged fingers. Another work by Yao [16] is a hand contour model which simplify the gesture matching process. They also propose a 14-patch hand partition scheme for color-based semiautomatic labeling. To recognize Sign language, which is one important and specific social application of gesture recognition, Chikkanna et al. [17] collected data using Kinect, applied k-means to extract features and used HCRF which gives them 95% of recognition rate.

4. RECENT WORKS AND FUTURE SCOPE USING KINECT V2

We have already discussed about Microsoft Kinect *v*2 which is a recent upgradation of Microsoft Kinect *v*1 with some different technologies inside. Now here we will find out what are the recent trends of work with this upgraded Microsoft Kinect *v*2. According to paper [18], Fürntratt has done an accuracy analysis using Kinect *v*2 as a pointing device based on 3D joint positions of the user's arm. Whereas, Lachat has experimented the ability of close range 3D modelling with Kinect v2 sensor [19]. Measuring the depth accuracy of the newly released Kinect v2 depth sensor by obtaining a cone model to illustrate its accuracy distribution in done by Yang [20]. Gaber et al. has worked on grading of facial paralysis with low cost system like Kinect *v*2. They are also claiming that the extended work has a fair chance to work as a virtual rehabilitation tool for facial paralysis[21]. Another paper by Zennaro, has done some research with achievements that have been obtained with the switch of technology from Kinect *v*1 to Kinect *v*2, in the context of 3D re-construction and people tracking [4].

Interestingly, most of the works on gesture recognition is based on Kinect v1, and very few on *v*2. Since Kinect *v*2 has many enhanced features than *v*1, use of Kinect *v*2 is likely to be more useful in gesture recognition. Thus, there is a lot of scope in future research on gesture recognition using Kinect *v*2 and our ongoing work is currently exploring this.

In this paper, we have focused on gesture recognition. It is to be noted that human activity is a sequence of gesture. Gaze is also a particular type of gesture. Thus, recognition of these fall into the extended domain of gesture recognition. Due to shortage of space, this conference version focuses on the review of the basic gesture recognition works. We plan to expand the current work significantly to include the extended domain of gesture recognition in a future journal extension.

5. CONCLUSION

Though computer science has progressed a lot but still building a computer which will be able to understand and interact with human is still in its infancy. So this is a big challenges to the researchers in the field of computer vision to make such a system which can be a good support system for our society. The important of gesture recognition range from sign language recognition to virtual reality to medical rehabilitation. Microsoft Kinect is a very innovative invention in this field. This gives extra information (e.g, depth) compare to normal basic RGB cameras and in a very affordable prices. So that all the researchers can be able to get the scope to explore the device and can get full benefit out of it. In this paper we have only covered gesture recognition part. The major tools surveyed in this purpose are DTW, FEMD, SP-EMD, SMRF, HCRF etc. These are some basic areas where already researchers are actively working, but there may be thousands of other areas which are unexplored yet.

Acknowledgements

The first author would like to thank the Department of Science and Technology (DST), Govt. of India, for funding our ongoing project (of which, the current work is only a part) through DST Women Scientists Scheme (WOS-A), Grant No. SR/WOS-A/ET-1001/2015 (G).

References

[1] S. Gokturk, H. Yalcin, and C. Bamji, “A time-of-flight depth sensor-system description, issues and solutions,” *Comput. Vis. Pattern …*, vol. 00, no. C, pp. 35–35, 2004.

[2] Z. Zhang, “Microsoft kinect sensor and its effect,” *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, 2012.

[3] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with microsoft kinect sensor: A review,” *Cybern. IEEE Trans.*, vol. 43, no. 5, pp. 1318–1334, 2013.

[4] S. Zennaro, M. Munaro, S. Milani, P. Zanuttigh, A. Bernardi, S. Ghidoni, and E. Menegatti, “Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications,” in *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, 2015, pp. 1–6.

[5] S. Mitra and T. Acharya, “Gesture recognition: A survey,” *Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans.*, vol. 37, no. 3, pp. 311–324, 2007.

[6] J. Daugman, “Face and Gesture Recognition: Overview,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 19, no. 7, pp. 675–676, 1997.

[7] D. M. Gavrila, “The Visual Analysis of Human Movement: A Survey,” *Sci. Direct*, vol. 73, no. 1, pp. 82–98, 1999.

[8] B. Fasel and J. Luettin, “Automatic facial expression analysis: A survey,” *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, 2002.

[9] J. Geng, “Structured-light 3D surface imaging: a tutorial,” *Adv. Opt. Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[10] C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, “Time-of-Flight Cameras and Microsoft Kinect™,” pp. 107–108, 2012.

[11] A. D. Wilson and A. F. Bobick, “Parametric hidden Markov models for gesture recognition,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 21, no. 9, pp. 884–900, 1999.

[12] Z. Ren, J. Yuan, and Z. Zhang, “Robust hand gesture recognition based on finger-earth mover’s distance with a commodity depth camera,” in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 1093–1096.

[13] C. Wang, Z. Liu, and S.C. Chan, “Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera,” *Trans. Multimed.*, vol. 17, no. 1, pp. 29–39, 2015.

[14] H. Liang, J. Yuan, and D. Thalmann, “Parsing the Hand in Depth Images,” *IEEE Trans. Multimed.*, vol. 16, no. 5, pp. 1241–1253, 2014.

[15] S. Poularakis and I. Katsavounidis, “Finger detection and hand posture recognition based on depth information,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 4329–4333.

[16] Y. Yao and Y. Fu, “Contour model-based hand-gesture recognition using the Kinect sensor,” *Circuits Syst. Video Technol. IEEE Trans.*, vol. 24, no. 11, pp. 1935–1944, 2014.

[17] M. Chikkanna and R. M. R. Guddeti, “Kinect based real-time gesture spotting using HCRF,” in *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, 2013, pp. 925–928.

[18] H. Fürntratt and H. Neuschmied, “Evaluating pointing accuracy on Kinect V2 sensor,” in *International Conference on Multimedia and Human-Computer Interaction (MHCI)*, 2014.

[19] E. Lachat, H. Macher, M.-A. A. Mittet, T. Landes, and P. Grussenmeyer, “First Experiences With Kinect V2 Sensor for Close Range 3D Modelling,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XL–5/W4, no. January 2016, pp. 93–100, 2015.

[20] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, “Evaluating and improving the depth accuracy of Kinect for Windows v2,” *IEEE Sens. J.*, vol. PP, no. 99, pp. 1–1, 2015.

[21] A. Gaber, M. F. Faher, and M. A. Waned, “Automated grading of facial paralysis using the Kinect v2: A proof of concept study,” in *Virtual Rehabilitation Proceedings (ICVR), 2015 International Conference on*, 2015, pp. 258–264.