International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 4, Special Issue 1 (September '2018)
Special Issue on 2nd National Conference on Advanced Computing and Communication Technologies,
Conference held at Siri PSG Arts and Science College for Women, Sankari,Tamilnadu,India.

MICROSOFT KINECT SENSOR "FUTURE TRENDS AND LATEST RESEARCH CHALLENGES"

Dr.T.Miranda Lakshmi,
Assistant Professor,
P.G and Research Dept. of Computer Science,
St.Joseph's College OF Arts and Science (Autonomous),
Cuddalore, Tamilnadu, India.

V.Muralidharan,
Dept.of Computer Science,
St.Joseph's College OF Arts and Science (Autonomous),
Cuddalore, Tamilnadu, India.

Abstract: In recent years, Kinect has gained more popularity as a portable, low cost, high-resolution depth, visual sensing and markerless human motion capture device . As a result of these advantages and the advanced skeletal tracking capabilities, it has become an important tool for clinical assessment, physiotherapy and rehabilitation. This paper contains an overview of evolution of different versions of Kinect and also it highlights their key features.

Keywords: Markerless human motion capture device, depth sensing, information fusion, RGB-D Camera ,Kinect sensor.

I.INTRODUCTION

Saving three-dimensional information about geometry of objects or scenes tends to be increasingly applied in the conventional workflow for documentation and analysis, of cultural heritage and archaeological objects or sites. In this particular field of study, the needs in terms of restoration, conservation, digital documentation, reconstruction or museum exhibitions can be mentioned [1,2]. The digitization process is nowadays greatly simplified thanks to several techniques available that provide 3D data [3]. In the case of large spaces or objects, terrestrial laser scanners (TLS) are preferred because this technology allows collecting a large amount of accurate data very quickly. While trying to reduce costs and working on smaller pieces, on the contrary, digital cameras are commonly used. They have the advantage of being rather easy to use, through image-based 3D reconstruction techniques [4]. Besides, both methodologies can also be merged in order to overcome their respective limitations and to provide more complete models [5,6]. Microsoft Kinect is a device originally designed for sensing human motion and developed as an controller for Xbox game console that is being sold since 2010. It did not take too long for researchers to notice that its applicability goes beyond playing video games, but to be used as a depth sensor that facilitates interaction using gestures and body motion. In 2013, a new Kinect device is introduced with the new game console called as Kinect v2 or Kinect for Xbox One. The new Kinect replaced the older technologies and brought many advancements to the quality and performance of the system. The older Kinect named as Kinect v1 or Kinect for Xbox 360 after new Kinect's arrival. Although it is categorized as a depth camera, the Kinect sensor is more than that. It has several advanced sensing hardware containing a color camera, a depth sensor, and a four-microphone array. These sensors ensure different opportunities at 3D motion capture, face and voice recognition areas [5]. While Kinect for Xbox 360 uses a structured light model to get a depth map of a scene, Kinect for Xbox One uses a faster and more accurate TOF sensor. Skeleton tracking features of Kinect are used to analyse human body movement for applications related to human computer interaction, motion capture, human activity recognition and more areas. Moreover, it makes a great use for studies especially in physical therapy and rehabilitation. An economical time of flight (TOF) technology with potential for application to patient positioning verification in radiotherapy. In radiotherapy the patient is initially positioned during the simulation computed tomography (CT) scan, which is then used to create a treatment plan. The treatment plan is designed to deliver tumoricidal dose to a planning target volume (PTV), which encompasses the gross disease with an added margin to account for setup uncertainties. Once a treatment plan is approved, patients return for multiple treatment fractions over a period of days or weeks. Replicating precise patient positioning between fractions is critical to ensure accurate and effective delivery of the approved treatment plan. The motivation of this survey is to provide a comprehensive and systematic description of popular RGB-D datasets for the convenience of other researchers in this field.

II.LITRATURE SURVEY

Motion capture and depth sensing are two emerging areas of research in recent years. With the launch of Kinect in 2010, Microsoft opened doors for researchers to develop, test and optimize the algorithms for these two areas. Leyvand T [2] discussed about the Kinect technology. His work throws light on how the Identity of a person is tracked by the Kinect for XBox 360 sensor. Also a bit of information about how the changes are happening in the technology over the time is presented. With the launch of Kinect, expects a sea change in the identification and tracking techniques. They discussed the possible challenges over the next few years in the domain of gaming and Kinect sensor identification and

tracking. Kinect identification is done by two ways: Biometric sign-in and session tracking. They considered the face that players do not change their cloths or rearrange their hairstyle but they do change their facial expressions, gives different poses etc. He considers the biggest challenge in success of Kinect is the accuracy factor, both in terms of measuring and regressing. Key prospect of the method is they are considering a single depth image and are using an object recognition approach. From a single input depth image, they inferred a per pixel body part distribution.
Depth imaging refers to calculating depth of every pixel along with RGB image data. The Kinect sensor provides real-time depth data in isochronous mode[18]. Thus in order to track the movement correctly, every depth stream must be processed. Depth camera provides a lot of advantages over traditional camera. It can work in low light and is color invariant [1] the depth sensing can be performed either via time-of-flight laser sensing or structured light patterns combined with stereo sensing [9]. The proposed system uses the stereo sensing technique provided by PrimeSense [21]. Kinect depth sensing works in real-time with greater accuracy than any other currently available depth sensing camera. The Kinect depth sensing camera uses laser beam to predict the distance between object and sensor. The technology behind This system is that the CMOS image sensor is directly connected to Socket-on-chip [21]. Also, a sophisticated deciphering algorithm (not released by PrimeSense) is used to decipher the input depth data.

III.RESEARCH PROPOSAL

Because of their attractiveness and imaging capacities, lots of works have been dedicated to RGB-D cameras during the last decade. The aim of this section is to outline the state-of-the-art related to this technology, considering aspects such as fields of application, calibration methods or metrological approaches.

Fields of application of RGB-D cameras is a wide range of applications can be explored while considering RGB-D cameras. The main advantages are the cost, which is low for most of them compared to laser scanners, but also their high portability which enables a use on board of mobile platforms.

Towards 3D modeling of objects with a RGB-D camera is the creation of 3D models represents a common and interesting solution for the documentation and visualization of heritage and archaeological materials. Because of its remarkable results and its affordability, the probably most used technique by the archaeological community remains photogrammetry.

Error sources and calibration methods is the main problem while working with ToF cameras is due to the fact that the measurements realized are distorted by several phenomena. For guarantying the reliability of the acquired point clouds, especially for an accurate 3D modeling purpose, a prior removal of these distortions must be carried out. To do that, a good knowledge of the multiple error sources that affect the measurements is useful.

IV.OUTLOOK FOR THE FUTURE

By analyzing above papers, we believe that there are certainly many future works in this research community. Here, we discuss potential ideas for each of main vision topics separately.

Object tracking and recognition is from the background subtraction based on depth images can easily solve practical problems that have hindered object tracking and recognition for a long time. It will not be surprising if tiny devices equipped with Kinect-like RGB and depth cameras appear in normal office environments in the near future. However, the limited range of the depth camera may not allow it to be used for standard in- door surveillance applications. To address this problem, the combination of multiple Kinects may be a potential solution. This will of course require the communication between the Kinects and object reidentification across different views.

Human activity analysis is achieving a reliable algorithm that can estimate complex human poses (such as gymnastic or acrobatic poses) and the poses of tightly interacting people will definitely be active topics in the future. For activity recognition, further investigations for low-latency systems, such as the system described in , may become the trend in this filed, as more and more practical applications demand online recognition.

Hand gesture analysis is it can be seen that many approaches avoid the problem of detecting hands from a realistic situation by assuming that the hands are the closest objects to the camera. These methods are experimental and their use is limited to laboratory environments. In the future, methods that can handle arbitrary, high degree of freedom hand motions in realistic situations may attract more attention. Moreover, there is a dilemma between shape based and 3-D model based methods. The former allows high speed operation with a loss of generality while the latter provides generality at a higher cost of computational power. Therefore, the balance and trade-off between them will become an active topic.

Indoor 3D mapping is according to the evaluation results from the most current approaches fail when erroneous edges are created during the mapping. Hence, the methods that are able to detect erroneous edges and repair them autonomously will be very useful in the future . In sparse feature-based approaches, there might be a need to optimize the key point matching scheme, by either adding a feature look-up table or eliminating non-matched features. In dense point-matching approaches, it is worth trying to reconstruct larger scenes such as the interior of a whole building. Here, more memory efficient representations will be needed.

V.DESIGN METHODOLOGY

Our system implements Augmented Reality using processing capabilities of Kinect. The system consists of 4 major components as Tracking Device, Processing Device, Input Device and Display Device. We use Kinect as a Tracking device . It contains three sensors for processing of depth images, RGB images and voice. Depth camera and Multi-Array Mic of Kinect are used to capture Real-Time image stream and audio data respectively. Depth sensor is used to obtain the distance between sensor and tracking object. The input device to our set-up is a high definition camera which is used to get input image stream and run as the background to all Augmented Reality components. On this background stream, we superimpose event-specific 3D models to provide virtual reality experience. The processing Device, consisting of Data Processing Unit, Audio Unit and software associated with it takes care of which model to superimpose at which time. Processing Unit passes the input video stream and the 3D model to display device for visualization purpose.

The Kinect system plays an important role in working of overall system. This system works as tracking unit for the Augmented Reality System. This system uses some of most exciting functionalities of Kinect such as skeletal tracking, joint estimation and Speech recognition for a human body. Skeletal tracking is useful for determining the user's position from Kinect, when user is in frame, which will be used for guiding him through assembly procedure. Also, it helps in gesture recognition. This system guides the user through complete assembly of product using speech and gesture recognition. The assembly of product includes bringing together individual constituent parts and assembling them as a product.

There are two assembly modes for this system, Full Assembly and Part Assembly. In Full Assembly mode, Kinect will guide technician on how to assemble a whole product sequentially. This mode will be useful when whole product has to be assembled. In Part Assembly mode, technician has to select a part to be assembled and then Kinect will guide him on how to assemble a selected part. When assembly of that part is completed, technician can select another part or quit. This mode will be useful when a part/parts needs to be assembled.

The system has been developed to work in 2 modes, Speech Mode and Gesture mode. The choice to select a mode has been given to user based on his familiarity to system and convenience to use it. If user has opted for speech mode, he has to use voice commands to interact with the system and system will guide him through voice commands. On the other hand, if user has opted for gesture mode, he has to use gesture to interact with the system and system will guide him through voice commands. The 'START' command is used in both modes to initiate the system. After system initiation, user will select a speech mode or gesture mode and will continue working in the same.

SOFTWARE

VI.DOT NET TECHNOLOGY

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.

Kinect v1

Microsoft Kinect v1 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 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 .

Kinect v2

Microsoft Kinect v1 got an upgradation to v2 in November 2013. The second generation Kinect v2 is completely different based on its ToF technology . 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 .

Kinect Software

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 is only available for Windows whereas OpenNI is a multiplatform and open-source tool. Microsoft Kinect includes free downloadable software, which is Kinect development library tool.

VII.PRACTICAL EXPERIMENTS

Kinect, in this paper, refers to both the advanced RGB/depth sensing hardware and the software-based technology that interprets the RGB/depth signals. The hardware contains a normal RGB camera, a depth sensor and a four-microphone array, which are able to provide depth signals, RGB images, and audio signals simultaneously. With respect to the soft-ware, several tools are available, allowing users to develop products for various applications. These tools provide facilities to synchronize image signals, capture human 3-D motion, identify human faces, and recognize human voice, and others. Here, recognizing human voice is achieved by a distant speech recognition technique, thanks to the recent progresses on the surround sound echo cancelation and the microphone array processing. More details about Kinect audio processing can be found in [5] and [6]. In this paper, we focus on techniques relevant to computer vision, and so leave out the discussion of the audio component.

RGB Camera is to delivers three basic color components of the video. The camera operates at 30 Hz, and can offer images at 640×480 pixels with 8-bit per channel. Kinect also has the option to produce higher resolution images, running at 10 frames/s at the resolution of 1280×1024 pixels.

3-D Depth Sensor is consists of an IR laser projector and an IR camera. Together, the projector and the camera create a depth map, which provides the distance infor- mation between an object and the camera. The sensor has a practical ranging limit of 0.8m−3.5m distance, and outputs video at a frame rate of 30 frames/s with the resolution of 640×480 pixels.

Microsoft Kinect v1 got an upgradation to v2 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].

First of all, a central matrix of 10 × 10 pixels is considered in the input images acquired during the experiment. This enables to compute mean measured distances from the sensor for each position. Then, the deviations between real and measured distances are plotted on a graph as a function of the range. Each of the 50 deviations obtained from the 50 depthmaps acquired per station is represented as a point. As depicted in a B-spline function is estimated within these values. Since the sensor was accurately placed on the tripod with respect to its fixing screw, a systematic offset occurs on raw measurements because the reference point for the measurement does not correspond to the optical center of the lens. The influence of this offset corresponding to the constant distance between fixing point and lens (approximately 2 cm) is removed on this graph. It appears that the distortions for the averaged central area vary from −1.5 cm to 7 mm, which is rather low regarding the technology investigated. At 4.5 m range, a substantial variation is observed. Under 4.5 m range, the deviations are rather included within an interval of variation of almost 1 cm (from −1.5 cm to 7 mm).

Since a set of 50 successive depthmaps is acquired for each position of the sensor, a standard deviation can also be computed over each sample presents a separate graph showing the evolution of the computed standard deviations as a function of the range. As it can be seen, the standard deviation increases with the range. This means that the scattering of the measurements increases around the mean estimated distance when the sensor moves away from the scene. Moreover, for the nearest range (0.8 m), the standard deviation reported stands out among all other positions. As a matter of fact, measurements realized at the minimal announced range of 0.5 m would probably be still less precise. Since a clear degradation with depth is showed, it makes sense to bring a correction.

VIII.RESULT AND DISCUSSION

The dream of building a computer that can recognize and understand scenes like human has already brought many challenges for computer-vision researchers and engineers. The emergence of Microsoft Kinect (both hardware and software) and subsequent research efforts have brought us closer to this goal. In this review, we summarized the main methods that were explored for addressing various vision problems. The covered topics included object tracking and recognition, human activity analysis, hand gesture analysis, and indoor 3-D mapping. We also suggested several technical and intellectual challenges that need to be studied in the future.

IX.CONCLUSION

This paper surveys the studies that use Microsoft Kinect technology to develop applications and games in physical rehabilitation field. Kinect shows a great potential with its low-cost and portability and fast application development times against its rivals using markers for human motion sensing. There is an ongoing interest in developing Kinect-based systems for physical rehabilitation purposes, and the new studies improves as in the accuracy and performance, and reveals new usage areas for Kinect in the field of rehabilitation.

X.REFERENCES

[1]. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. Remote Sens. 2011, 3, 1104–1138.

[2]. Remondino, F.; Rizzi, A.; Agugiaro, G.; Girardi, S.; De Amicis, R.; Magliocchetti, D.; Girardi, G.; Baratti, G. Geomatics and geoinformatics for digital 3D documentation, fruition and valorization of cultural heritage. In Proceedings of the EUROMED 2010 Workshop "Museum Futures: Emerging Technological and Social Paradigms", Lemessos, Cyprus, 8–13 November 2010.

[3]. Sansoni, G.; Trebeschi, M.; Docchio, F. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. Sensors 2009, 9, 568–601.

[4]. Hullo, J.F.; Grussenmeyer, P.; Fares, S. Photogrammetry and dense stereo matching approach applied to the documentation of the cultural heritage site of Kilwa (Saudi Arabia). In Proceedings of CIPA Symposium, Kyoto, Japan, 10–15 October 2009.

[5]. Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, "Real-time human pose recognition in parts from single depth images," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference.

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

[7]. 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.

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

[9]. C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, "Time-of-Flight Cameras and Microsoft KinectTM," pp. 107–108, 2012.

[10]. 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.

[11]. 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.

[12]. Kahlmann, T.; Remondino, F.; Ingensand, H. Calibration for increased accuracy of the range imaging camera SwissrangerTM. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2006, XXXVI-5, 136–141.

[13]. Lindner, M.; Kolb, A. Calibration of the intensity-related distance error of the PMD ToF-camera. Proc. SPIE 2007, 6764, doi:10.1117/12.752808.

[14]. Chow, J.; Ang, K.; Lichti, D.; Teskey, W. Performance analysis of a low-cost triangulation-based 3D camera: Microsoft Kinect system. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2012, XXXIX-B5, 175–180.

[15]. Kinect for Windows. Available online: https://www.microsoft.com/en-us/kinectforwindows/default.aspx (accessed on 28 June 2015).

[16]. Structure Sensor. Available online: http://structure.io/ (accessed on 24 August 2015).

[17]. Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Fitzgibbon, A. KinectFusion: Real-time dense surface mapping and tracking. In Proceedings of 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 26–29 October 2011; pp. 127–136.

[18]. Kourakli, M., Altanis, I., Retalis, S., Boloudakis, M., Zbainos, D., & Antonopoulou, K. (2017). Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games. International Journal of Child-Computer Interaction, 11, 28-39.

[19]. P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3-D modeling of indoor environments," Int. J. Robot. Res., vol. 31, no. 5, pp. 647–663, Apr. 2012.