Depth Cameras and their Usefulness in Social Places

Beáta Szendrey, István Járomi, Gergely Nagy, Balázs Hartal, Ádám Rozgonyi, Éva Hajnal
Óbuda University Alba Regia Technical Faculty, Székesfehérvár, Hungary
hajnal.eva@amk.uni-obuda.hu

***Abstract*—Nowadays the Kinect sensor and its analogs can be said very important innovations in creation of smart places, because they combine the 2D video image and the depth data to 3D and with time parameter to 4D information. These inspired the development a lot of applications at first in the field of game industry and entertainment, later in the field of medicine, health care and finally in the industry. In this paper a review is given of the pros and cons of this type of sensor and focuses onto the usage of it in social places. Some proposals are provided to use it in these places to become smarter.**

I. Introduction

After 2000 an extraordinary innovation raised up in the field of depths sensing complex cameras. The development was primarily inspired by the entertainments and game industry. The result is products with affordable price and enough robust market demand to produce and use these devices in large volume.

The sensing technology of depth cameras mainly based on two principles. One of them is the analogue of human depth sensing by comparing the two different pictures of two RGB cameras makes possible to compute the depth information. The other strategy use infrared (IR) sensors for supplement the 2D information of the RGB cameras video stream to gain 3D information and with the time parameter finally 4D information. In this paper the focus is on the IR based depth sensing cameras and mainly on Kinect sensors, because the experiments were taken with these devices (Fig 1).

Several examples of such types of devices are: Kinect for Xbox 360, Kinect v2, Intel SR300, Intel SR305 Intel R200, Orbbec Astra, Orbbec Persee.

However these cameras were primarily developed for entertainments, they implicated the scientific applications and connected researches and further the industrial applications [2]. Depth sensing is used for medical[3],

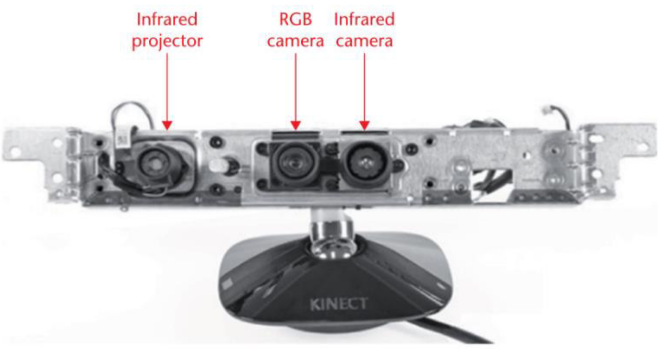

Figure 1Kinect sensor XBOX 360 [1]

rehabilitation and elderly care systems [4], even more in industrial usage is a target of important R&D [3]. Now the game industry is giving up this field, because the accuracy and speed of these devices do not give appropriate user experience but the scientific applications and the number of the connected publications is growing by exponential trend (Fig 2).

First Kinect version was launched on 4 November 2010. On 1 February 2012, and more than one year later Microsoft released the Kinect Software Development Kit (SDK) for Windows, which helps the software development.

The Kinect sensor contains several advanced sensing hardware. Most notably, it contains a depth sensor, an RGB camera, and a four-microphone array. These provide full-body 3D motion capture, facial recognition, and voice recognition capabilities [1]. Microsoft and its Kinect systems have been among the earliest affordable depth sensors that use different technologies for the respective iterations. The first version had a comparatively low resolution of only 320 × 240, whereas the next edition updated this to 512 × 424, amongst other improvements in RGB camera resolution and microphone arrays[5].

Table 1 Comparison of different depth cameras

	RGB camera resolution	Depth Camera resolution	Frame rate	Min. depth	Max. depth	Angle
Kinect 360 [6]	640x480	320x240	30fps	0.85	4m	58,44
Kinect v2 [6][7]	1920x1080	512x424	30fps	0.85	4,5	70.6,60
Intel SR 300[8]	1920x1080	640x480	60fps	0.2m	1.2m	71,55,88
Orbbec Persee [9]	1280x720	640x480	30fps	0.6m	8m	60,49,73

Intel has developed its RealSense technology for several years later. The cameras are demonstration sets for a desired future integration in end-user devices, such as note-books, smart phones, or tablets. The two presented systems are the back-facing R200 system, for installation at the back of a device, and the SR300 system for installation in the front of a device. Both cameras are capable of a depth resolution of 640 × 480 at high frame rates.

Orbbec is a company founded by one of the main developers of the OpenNI SDK and provides depth cameras using structured light technology. The main advantage of the Persee device is the integrated processing that enables the creation of depth maps and point clouds that can be directly sent over network. The comparison of these complex cameras in table 1 shows that by hardware solutions the Kinect camera is not an outstanding one.

Sensing of depth information can be based on two concepts. One possibility is to use the structured light (SL) concept, where the infrared laser light is passed over a diffraction grating, which sprinkles the light and creates a special pattern of infrared dots. This pattern is reflected on the surface and is changed. This pattern is sensed and used for the calculation of the depth information [1]. This is the working strategy of Kinect XBOX 360 sensor system. The Kinect v2 uses the TOF technology for sensing. In this concept the time difference is measured between the IR light emission and the reflected light sensing. If we know the speed of the light, we can calculate the depth information. Later investigations proved that that TOF method can produce more accurate results [1]. There is a quasi-compliance between experts that the Kinect cameras by hardware features are not outstanding in the market competitors, but more the marketing strength and mainly their software, the Microsoft Kinect SDK is really excellent. It is able to detect and follow in real time the human skeleton with many joints parallel for six persons, which makes possible to develop a lot of applications to this device. The other SDKs (OpenNI) later reached similar results [10].

In this paper we will focus on the human gesture and behaviour detection in social places. In this point of view the basis of the Kinect sensing is that the visual information supplemented with depth information is more effectively used for human body posture and movement detection than the only visual information [11]. This was proved in the literature and showed by Table 2.

Table 2 Accuracy of foreground recognition technics [10]

	True positive	False positive
Depth information	96.7%	0%
RGB information	45.3%	2.3%

Now the Kinect XBOX 360 creates depth maps with accuracy (0...40mm) and (0...18mm) in the newer version. Their accuracy depends on the temperature, light, and distance which are important parameters in application development [7][12][2].

Now the number of publications is growing dynamically respected to simple human gesture detection, which accuracy's improvement is an important research field up to now and the complex behaviour recognition research which will be the basis of the future applications (Figure 2).

Connected to this research several database developments were executed for the valid testing of different algorithms. Mainly the aim was the behavior and gesture type identification, and video and/or depth stream databases were collected in relation to different human activities [13]. These contain basic movements (sitting, standing, walking, lying), sport activities and later home activities etc. [13]. It was described that the adult and child motions are different, so there were databases for these activities as well. These databases were used not only for research purposes but moreover the film industry in animations [14]. Based on these databases, the movement identification methods become evaluable. Now the identification of the most simple postures, movements can be of near 90% accuracy and the more complex home activities (phone call, eating, clothing) can be detected with near 80% accuracy [13].

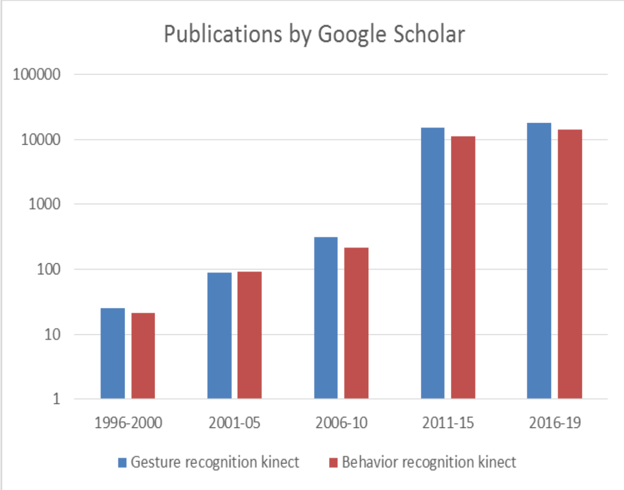

Figure 2 Number of Kinect related publications in the field of gesture and behavior recognition.

Now the applications of Kinect cameras are frequently limited by the sensors. There are strict limitations in depth and the number of identified persons. By the references the best method to place the camera to 3m height to the ceiling and it results 8,25m^2 valid sensed area/camera. [5]. Generally it can be summarized that the high accuracy sensing and gesture detection work now mostly in artificial laboratories, but the effects of a real environment (color patterns, different light) disturb these applications which are not used widespread yet[5].

In this paper we are focusing on the effective usage of depth sensing complex cameras in real applications. A database development is demonstrated, which aim is to help the effective storage, processing and querying of skeleton stream information. Furthermore we are discussing what type of developments is needed for this revolution mainly in the field of the human gesture and behavior detection in social places.

II. METHODS

A User Interface program and a database were developed for the storage of the skeleton data. The software is a simple .NET Framework WPF App, which offers a form for the user and collects and stores skeleton data into a database. Kinect handling which uses the Kinect SDK and data stream storage is executed in asynchrony thread.

In this project Kinect XBOX360 camera and its Microsoft Kinect SDK kit were used.

Now the UI collects data of one Kinect camera and stores them into a MongoDB database or can list onto the screen or avataring on the screen (Fig 4). Now the SDK can follow two persons and 20 joints of each person with the help of the Infrared sensor. The structure of the database is shown on figure 3. The 3D coordinates of each joints with a timestamp are stored to reach 4D data storage. The database contains the activity string, which is the ID of the current activity.

Each joint contains the x, y, z, timestamp data, the sensing frequency is 30 fps and storage frequency is the same. In the user interface the activity code can be typed till a good database will gathered for the valid classification of human activities.

The tests were made on a typical PC (Intel i7, 2GHz, (8GB RAM 64 bit Microsoft Windows 10 Operating System)

The MongoDB database was stored at the official site and used the free, minimal storage capacity.

```
{
  "id": "string",
  "activity": "string",
  "skeletonData": [{
    "jointsData1": {
      "jointType": "number",
      "x": "number",
      "y": "number",
      "z": "number"
    },
    "timestamp": "date"
  },{
    "jointsData2": {
      "jointType": "number",
      "x": "number",
      "y": "number",
      "z": "number"
    },
    "timestamp": "date"
  }
    ...
  }, ...]
}
```

Figure 3 Database structure

III. Results

The developed database and the connected UI was tested in a typical environment. The question is that a given PC with MongoDB database can execute real time data processing and storage. It was proved that it is inside the normal operational range. The storage capacity of the stream is 1MB/minute. In the later development the storage capacity can be reduced until 350KB/minute (10 Byte/joint*20joints*30fps*60second) without any calculated compression algorithm and to 100KB with compression. However it is important to focus on the computing capacity to the most important calculations, namely to the real time activity classification, which can be used for labeling the skeleton data. This label will make possible the later semantic querying of the database which makes possible the proposed application and research.

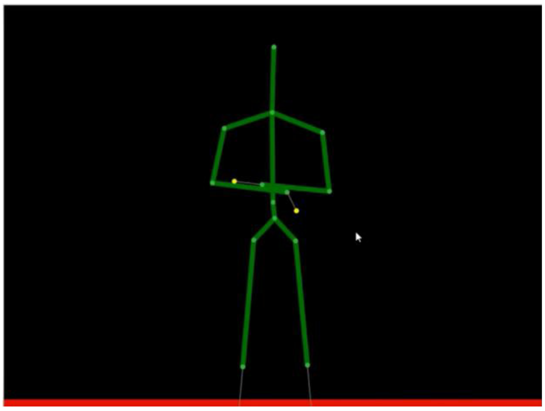

Figure 4 Skeleton picture by the detected 20 joints of a person

IV. Proposals

Our proposal and aim to research is how could be used depth sensor based complex cameras in the investigation of human population and people's activities in social places. It would be very interesting for the quantitative psychology and in practically it could be useful in safety services.

The advantages of these devices in safety that could be expected to detect, store and query data of human activities without the violation of human rights. The later developments will make possible the semantic querying of the database (a human is running, phoning, faints etc.).

The disadvantages of this concept are the restricted spatial scope of these devices, and the other difficulty is the restricted number of the followed human skeleton streams.

The safety services will be possible if these two restrictions will be defended. Now there are signs to this trend. The first sensors could handle 1-2 persons but the newer Kinect v2 can increase it to six and may be increased in the near future depending on the software developments. The scope of the Orbbec Persee has already reached the 8 m and later may will further increase to the scope of RGB cameras.

Nowadays the Kinect sensing is more accurate in optimal light conditions, the human detection algorithms make possible some improvements in this field in the near future (see Fig 5) [10]. The light will not effect it, so it can be used in dark places (cinemas, theatres etc.), furthermore the flashing light also will not limit the device, so it can be used in pop concerts and other night entertainment places.

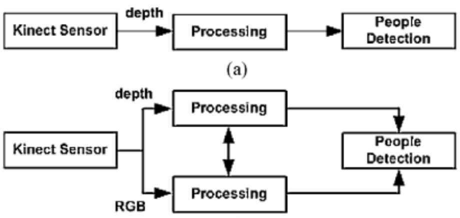

Figure 5 The schemes that Kinect use for human detection
source: J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, 2013.

In person identification it can be successfully used when the basis of the identification are biometric features, so it can be used if the face of the person is not visible and with some restrictions if the person wears special clothes which makes the identification harder (chador, cloak etc.).

There are references of the investigations of person identification by anthropometric and motion identical features [15]. Present results show that the identification of anthropometric features can be made with 80-90% accuracy and the identification by the movement is a bit lower. These results are the consequences of the limited spatial scope and the depth sensing accuracy of the device. For example if the scope of the camera is 4m and the person is walking or running, he/she can do only 3-4 steps in the scope of the device which may not be enough to analyse the identical features [15].

If the spatial scope and the depth sensing accuracy can be increased, this causes the increase of person identification. The identical feature selection of the human motions is an interesting research area, much younger than the anthropometric identification and there are possibilities to reach remarkable development in it.

V. CONCLUSION

In this paper a summary was collected about the investigations with depth sensing complex cameras. The point of view was from our research question that with the present devices can be placed cameras in social areas and detect the human behaviour. It was clear that such research has importance in safety and in psychology. However the present hardware and software solutions make possible this type of applications with some limitations. Smaller areas, for example medical places, smaller educational places, bank vestibules and limited population can be monitored. Monitoring of larger areas and higher population needs further developments. Based on the literature we really can expect remarkable development in this field.

In this paper the development and testing of a real time NOSQL (MongoDb) database was presented, which will be applicable for the semantic querying. This is a missing link between the human gesture, activity detection, storage and quantitative behaviour analysis.

ACKNOWLEDGMENT

I would like to thank the help and the effort of colleagues and students. The financial support of this paper was from Arconic 2019 grant.

REFERENCES

[1] W. Zeng and Z. Zhang, "Multimedia at Work Microsoft Kinect Sensor and Its Effect," *IEEE Multimed.*, vol. 19, pp. 4–10, 2012.

[2] D. Bakos, "Gyártósori mozdulatelemzés Xbox One Kinect," Budapesti Műszaki és Gazdaságtudományi Egyetem, 2018.

[3] R. Lun and W. Zhao, "A Survey of Applications and Human Motion Recognition with Microsoft Kinect," in *Electrical Engineering & Computer Science Faculty Publications. 408.*, 2015.

[4] B. Lange, C. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool Using the Microsoft Kinect Sensor," *33rd Annu. Int. Conf. IEEE EMBS*, pp. 1831–1834, 2011.

[5] M. Ruben, M. Michael, A. Carsten, S. Andreas, M. R. Mettel, M. Alekseew, C. Stocklöw, and A. Braun, "Designing and evaluating safety services using depth cameras," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 2, pp. 747–759, Feb. 2019.

[6] "DEPTH SENSOR SHOOTOUT." [Online]. Available: https://stimulant.com/depth-sensor-shootout-2/#device4. [Accessed: 23-Aug-2019].

[7] "Kinect V1 and Kinect V2 fields of view compared." [Online]. Available: https://smeenk.com/kinect-field-of-view-comparison/. [Accessed: 23-Aug-2019].

[8] "Intel ® RealSense ™ Camera SR300 Embedded Coded Light 3D Imaging System with Full High Definition Color Camera," no. June, pp. 1–54, 2016.

[9] "Orbbec Persee." [Online]. Available: https://orbbec3d.com/product-persee/. [Accessed: 23-Aug-2019].

[10] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, 2013.

[11] H. Liu and L. Wang, "Gesture recognition for human-robot collaboration : A review," *Int. J. Ind. Ergon.*, pp. 1–13, 2017.

[12] O. Wasenmüller and D. Stricker, "Comparison of kinect v1 and v2 depth images in terms of accuracy and precision," in *Asian Conference on Computer Vision*, 2016, pp. 34–45.

[13] B. Ni and P. Moulin, "RGBD-HuDaAct : A Color-Depth Video Database For Human Daily Activity Recognition," in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, 2011, pp. 1147–1153.

[14] A. Aloba, G. Flores, J. Woodward, A. Shaw, A. Castonguay, Y. Dong, E. Jain, L. Anthony, I. Cuba, Y. Dong, E. Jain, and L. Anthony, "Kinder-Gator: The UF Kinect Database of Child and Adult Motion.," in *Eurographics (Short Papers)*, 2018, pp. 13–16.

[15] V. O. Andersson and R. M. Araujo, "Person Identification Using Anthropometric and Gait Data from Kinect Sensor," *Proc. Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 425–431, 2015.