Microsoft Kinect and IMU Devices ~~operation~~ characteristics and their limitation compensation method.

Grzegorz Glonek and Adam Wojciechowski

Institute of Computer Science
Lodz University of Technology, Lodz, Poland,
grzegorz@glonek.net.pl, adam.wojciechowski@p.lodz.pl

Abstract. Microsoft Kinect v.1 and inertial measurement units (IMU) became very popular devices nowadays, what allows home users to detect and track human motion. Due to their working characteristics both of these devices are sufficient for casual scenarios, where precision is not a crucial factor. In the following article a review of characteristics, verified by experiments of both devices, is presented, as well as the proposition of their limitations compensation method.

1 Introduction

Since Microsoft Kinect has been released in 2010 and inertial devices have become an integral and almost mandatory part of every smartphone, motion tracking and motion detection became very popular and easily available for the home usage. Kinect is mainly used in the field that it was created for – games and entertainment. However, due to the limitations, mostly caused by the way it was built, only relatively simple casual games were developed for this controller. On the other hand, Microsoft Kinect became a popular subject for researchers, who want to find out how this device might be applied in more advanced scenarios. The second of the mentioned devices – inertial measurement units (IMU) – can be easily found in almost every modern smartphone. From home user's point of view, the most noticeable functionality implemented thanks to these devices is a screen rotation. Also all applications that measure number of pedestrian's steps basis on them. Of course, these devices also have some flaws and limitations that need to be taken into consideration in order to achieve accurate and stable results. In further part of this article measurement and working characteristics of both devices have been described, as well as a proposed method of their limitations compensation.

2 Kinect characteristics

Microsoft Kinect version 1 is an RGB-D camera built from two CMOS cameras and integrated infrared (IR) projector. One of these CMOS is responsible for an RGB signal and the second one is calibrated to record IR beam's view. However, the most important part is the main chip created by PrimeSense company, which is responsible for the motion tracking and the gesture recognition. The simplified device schema is presented in figure 2. According to the official specification, an operation range of Kinect is between $0.8m$ and $4m$ in the field of view 57° horizontally (static) and 43° vertically (adjustable ±21°). The same specification doesn't include any information on the possible variety of measurements accuracy in this area. The range defined in the official specification is presented in figure 4. However, some users [9] and researchers [13] observed that different device series have a slightly different ranges where they operate, so above values should be treated as an average. An important characteristic of Microsoft Kinect – object distance measurement – is directly related to the device design and used algorithm – an example of a structured light. Microsoft hasn't published any document describing how the algorithm actually works, but basing on original patent forms [10,12,11] and independent research, some rough description can be created, [8] as well as an image of the used light pattern (fig. 3). Basing on the image of a lighted scene, Kinect analyzes the distortion of IR dots pattern to estimate objects' distance. Two techniques are used in parallel to compute such estimation:

- Dots blurriness analysis
- Stereo-vision based on a single IR camera and a projector

A detailed description of both techniques can be found in [14,**?**]. A human skeleton estimation is based on the previously estimated depth map and pattern recognition built on about 100000 predefined samples. It is worth noticing that a signal recorded from an RGB camera is not used at all in the human skeleton estimation process.

2.1 Limitations

One of the basic limitations of this device is the sun light sensitivity. It is a result of relying only on the IR light to estimate scene depth. The sun light contains a full colors spectrum in both, visible and an invisible range and this causes a noise in the light pattern. The great example of the sun light impact on measurements can be found in [16]. That makes Kinect useless in outdoor scenarios and makes it difficult to use two or more Kinects simultaneously. However, despite Microsoft's recommendation to not use multiple Kinect controllers in one room, scientists work on algorithms of how to combine signals from few devices [17,18,19]. Other significant limitation is related to completeness of information gathered from the device. A skeleton estimated by the device contains information about position of 20 joints and rotations of bones between these joints. However, every joint is reduced to a single point and bones rotations are estimated basing on these points, what results in the lack of information about rotation along the bone (roll). Two other limitations are treated as the significant: occlusion and variety of the depth measurement accuracy in the field of observation. The first one occurs when a part user's body is covered behind an other object or is hidden behind any other body part (self-occlusion). When the occlusion happens, Kinect tries to "guess" the location of the covered joint or stops tracking it, when it is not able to provide any rough estimation. The occlusion by an external object seems to be intuitive and doesn't require any additional explanation, but self-occlusion is connected to Kinect's sensitivity to user's rotation to the camera. The official specification mentions that Kinect is designed to work in a face off pose. However, this document doesn't define what "face off" means and what is the exact angle between the human and the device when the occlusion occurs. Self experiments allow to observe how measurements change when the user rotates in front of the camera. The angle α (fig. 6) represents such rotation and has been calculated with eq. (1). Figures 5(a) and 5(b) show measured joints tracking state and elbow angle change during rotation. Of course, in this second case, the measured elbow belongs to the hand which has been visible to the camera all the time and its angle hasn't changed during the movement (rotation in "T-pose"). As we can see, when angle is greater than 50° measurements turn to be unstable and unreliable. Fluctuations in depth measurement accuracy has been observed in experiment with a professional MoCap system created by Vicon company. During that experiment, the user had to move back in the distance from $0.8m$ to $4m$ from the camera and such distance has been measured by Kinect and Vicon simultaneously. The result shows that in the close range Kinect slightly underestimates the distance and in the far range it overestimates it. The optimum is located in the distance of $2m - 2.3m$ from the camera. In fig. 1 the distance accuracy function chart can be found.

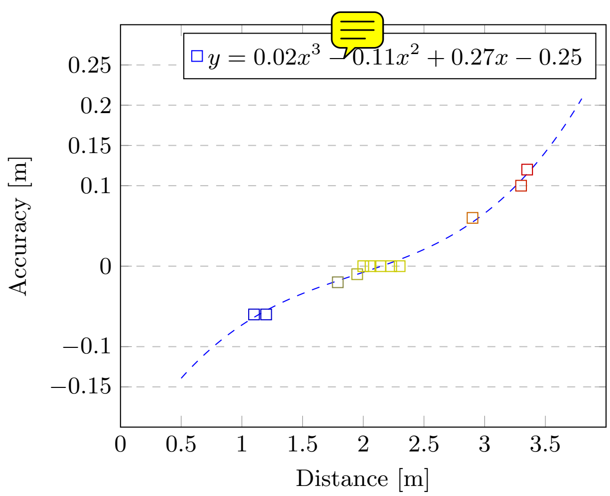


Fig. 1. Microsoft Kinect depth measurement accuracy by user distance

$$
\begin{aligned}
&P_{Dev}.Z = P_{Dev_1}.Z - P_{Dev_2}.Z\\
&P_{Dev}.X = P_{Dev_1}.X - P_{Dev_2}.X\\
&P_{Sh}.Z = P_{Sh_L}.Z - P_{Sh_R}.Z\\
&P_{Sh}.X = P_{Sh_L}.X - P_{Sh_R}.X\\
&\alpha = \begin{cases}
|atan(\frac{P_{Dev}.Z}{P_{Dev}.X}) - atan(\frac{P_{Sh}.Z}{P_{Sh}.X})| & ,P_{Dev}.X \neq 0 \& P_{Sh}.X \neq 0\\
|atan(\frac{P_{Dev}.Z}{P_{Dev}.X}) - \frac{\Pi}{2})| & ,P_{Dev}.X \neq 0 \& P_{Sh}.X = 0\\
|\frac{\Pi}{2} - atan(\frac{P_{Sh}.Z}{P_{Sh}.X})| & ,P_{Dev}.X = 0 \& P_{Sh}.X \neq 0\\
0 & ,P_{Dev}.X = 0 \& P_{Sh}.X = 0
\end{cases}
\end{aligned}
\tag{1}
$$

where:

- $P_{Dev_1}.X, P_{Dev_1}.Z, P_{Dev_2}.X, P_{Dev_2}.Z$ – X and Z axes coordinates of Kinect camera. These values are fixed to $((0,0),(1,0))$ respectively.
- $P_{Sh_L}.X, P_{Sh_L}.Z$ – X and Z axes coordinates of user's left shoulder.
- $P_{Sh_R}.X, P_{Sh_R}.Z$ – X and Z axes coordinates of user's right shoulder.

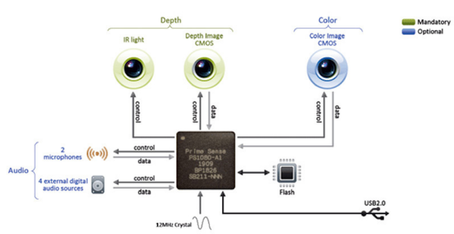
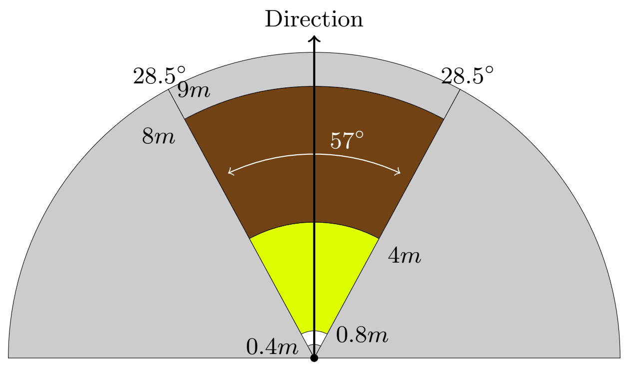


Fig. 2. Simplified Microsoft Kinect v.1 controler build schema [7]

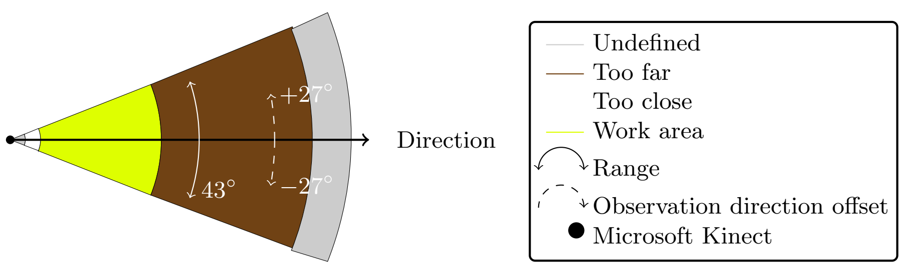
Fig. 3. IR scene lighting pattern[8]

3 IMU characteristics

IMU devices are in the professional usage for decades ie. gyroscopes in rockets during II World War or in navigation systems in the aircrafts. Such devices became popular in home devices since they were implemented in MEMS architecture. In this article, all experiments were performed with the usage of IvenSense MPU-6050 module, which integrates accelerometer, gyroscope and thermometer on single PCB. Both inertial sensors signals are affected by the external noise



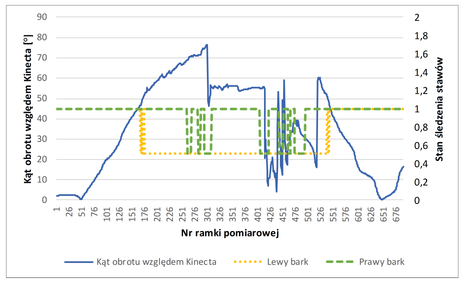
(a) Horizontal range



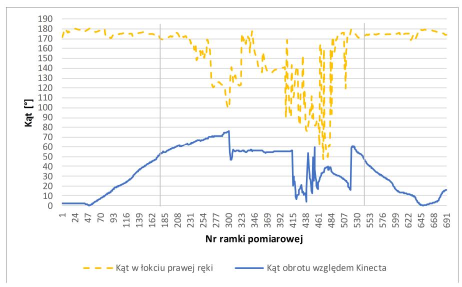
(b) Vertical range

Fig. 4. Microsoft Kinect v.1 work range

duża strata miejsca na samotny rysunek na środku strony. Trzeba będzie usunąć niektóóre lub połaczyć z innymi.



(a) Joint tracking state



(b) Right elbow angle

Fig. 5. Measurements by orientation to Kinect

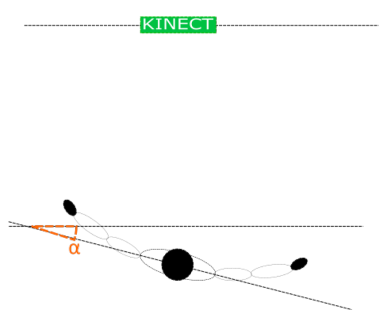


Fig. 6. Rotation angle α between user and Kinect

with different frequency characteristic and it need to be filtered out to make these signals usable. An accelerometer is a sensor responsible for a linear acceleration measurement in form of the g-force, so the device at rest measures the force of $1g$ ($1g = 9.81\frac{m}{s^2}$) in upwards direction. However, these devices measure every single temporary force that works on this device and this is treated as a high frequency noise. That means, the filter used to remove such incorrect data must be one of the low pass filters. Also the technology used to build such sensor – capacitive – results in the sensitivity for operating temperature changes. Such influence was the subject of some scientists' researches [6,?]. The same influence was also observed in the used device during the experiment, when the device measured g-force in the temperature range $10°C - 50°C$. The result of these measurements is presented in fig. 7. If the IMU device is placed on a human body, its temperature rises up to approx. $30°C$, so it requires some sort of the compensation. The second inertial sensor – a gyroscope – measures the angular velocity in $\frac{°}{s}$ units. Unlikely the accelerometer's short-term noise, the gyroscope suffers mostly from a constant, long-term bias that should be removed or limited. As this kind of noise is a low frequency signal, filter used to compensate it should be one of the high pass filters. This bias is also influenced by the operating temperature, however, there was no difference observed in filtered signal over the temperature range. It means that no additional compensation is required here. In these type of devices we can also notice the incomplete information problem. The accelerometer, even though it measures forces along all 3 axes, allows us to calculate orientation around two of them only. The orientation (or the rotation) around the gravity vector is unmeasurable for such device. Theoretically, it is possible to calculate the orientation around all 3 axes with the gyroscope signal and its integration over the time. However, even the filtered signal contains some noise that grows rapidly due to the numerical integration and it quickly results in the significant drift that makes such measurements useless.

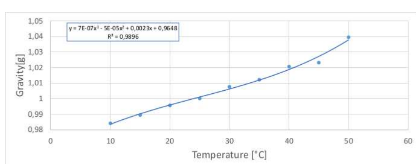


Fig. 7. Gravity measurement in temperature range $10°C - 50°C$

4 Compensation

4.1 Kinect

The compensation of Kinect measurements focuses on two aspects:

– To reduce the distance estimation inaccuracy according to the known model

– To filter unreliable joints positions estimations

The first aspect can be improved by the correction of the depth value ("z" in Kinect's coordination space) by function $f(z)$ according to eq. (2)

$$z' = f(z) = -0.02 * z^3 + 0.11 * z^2 - 0.27 * z + 0.25 \tag{2}$$

where: z' – joint distance correction value

This function is opposite to the function estimated from raw distance accuracy measurements, so adding it to the raw "z" value, reduces both: under- and overestimation.

The second problem can be reduced by the combination of few tools. First of all, joints positions should be filtered with the low pass filter (LPF). The Kinect camera works with the average frequency of $30Hz$ so rapid joint position changes between two consecutive frames is impossible. If such phenomena appears, LPF will remove such measurement. The other thing that should be taken into consideration is the user's orientation to the camera. From the experiment which results are presented in fig. 5 we can learn that the orientation greater than 50° cannot be considered as reliable measurement, so if Kinect is the only device used in the designed system, it is worth to consider some alert for the user to rotate back. Of course, the orientation angle calculation must be combined with ie. its variance in order to prevent "false" angles values (in fig. 5 the middle part of the chart shows values lower that 50°, but with the significant fluctuation, which is typical for unreliable measurements). For correct values the variance is close to 0, as there are no rapid changes in measurements.

to trzeba zapisać wzorem jako reguła decyzyjna, fragment algorytmu

4.2 IMU

The accelerometer and the gyroscope measurements are usually fused together in order to compensate the noise that affects both signals and to estimate the device orientation in the space. The most popular methods to achieve such fusion are Kalman filter and, lately, Madgwick filter [4]. Both of them can be described as a sort of complementary filters where two signals with low and high frequency noises are fused together with proper the noise reduction and some fusion factors. That allows to join together the same information from both sources with different level of importance. As the accelerometer signal needs to be filtered with LPF and the measurement is sensitive to the temperature, before it is be fused with the gyroscope, the correction according to the temperature should be done. The regression analysis allows to estimate the formula of such accelerometer signal correction presented in eq. (3).

$$A' = \frac{A}{1 + \beta(T - T_0)} \tag{3}$$

where:

- A' – corrected accelerometer measurement
- A – accelerometer measurement

- T – temperature measurement
- T_0 – device reference operating temperature. For used device $T_0 = 25^{\circ}C$
- β – correction factor. For used device $\beta = 0.0011$

The result of such correction is presented in figure 8. Next, these signals can be used in the fusion filter. The accuracy comparison of the previously mentioned filters can be found in the literature ie. [4].

4.3 Information incompleteness

Compensation methods presented previously are able to improve the quality of data gathered from both types of devices, however, are not able to add the missing information about rotations (Kinect – roll – rotation along the bone; IMU – yaw – rotation around gravity vector). To achieve that, both signals: from Kinect and IMU must be fused together. Then, the combined data will include full set of information. In the literature there are presented several methods of Kinect and IMU fusion [2,**?**,**?**] and each of them has different approach to such fusion and the different accuracy. Authors of this article created the original method, based on the bones orientations fusion that takes the source data quality and motion context into consideration during adjusting fusion factors. As both devices suffer with lack of rotation information around different axes, the fusion is done separately for each axis, and then results are combined together and joined with bones fixed length model to calculate final joints positions. The overall method schema is presented in figure 9. The proposed method has been tested on a set of four upper limb moves (fig. 3) that brought out measurement devices limitations. Results were compared with the method described in [1], as it has the best accuracy among all methods described in the known literature, but doesn't contain compensation of device limitations described in this article. Results presented in fig. 11 and 12 show the improvement in the joints position and the elbow angle estimation, up to 25% for the method that compensate devices imperfections.

z artykułem można tez tak zagrać, ze jest on usprawnieniem poprzedniego tylko trzeba pokazać wyniki bez udoskonalania nieprawidłowości sprzętu i z udoskonalaniem

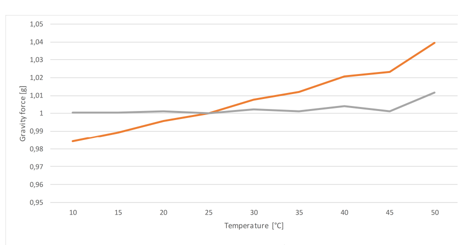


Fig. 8. Gravity measurement compensation due to the temperature

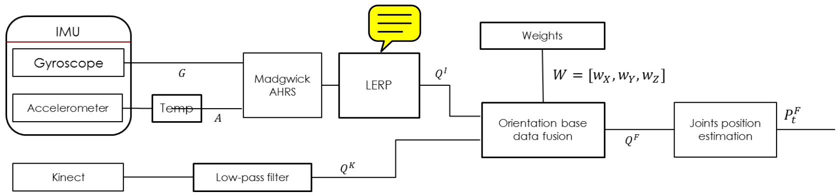


Fig. 9. Orientation based fusion method

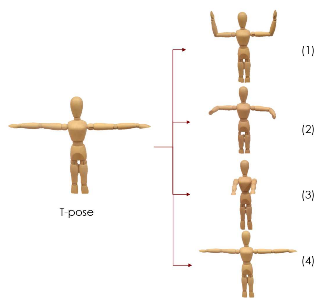


Fig. 10. Movement sequences performed during tests.

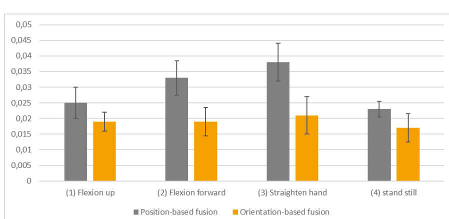


Fig. 11. Elbow positioning average accuracy

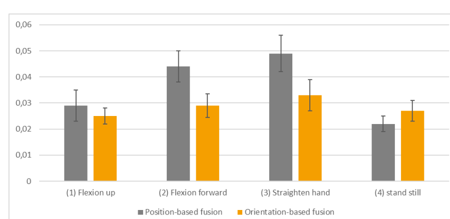


Fig. 12. Wrist positioning average accuracy

5 Summary

In this article authors presented the analysis of the popular motion tracking devices. Authors focused on the construction and characteristics od these devices, and how that affects their measurement abilities and limitations. This article also describes methods of the compensation of such flaws for each device individually, as well as for both fused together, to provide complete set of information that are missing in the individual devices. Also the original Kinect and IMU fusion method that includes compensation of pointed devices' imperfections has been presented.

References

1. Kalkbrenner, Christoph et al: Motion Capturing with Inertial Measurement Units and Kinect - Tracking of Limb Movement using Optical and Orientation Information. Proceedings of the International Conference on Biomedical Electronics and Devices (2014)
2. Tian, Yushuang et al: Upper limb motion tracking with the integration of IMU and Kinect Neurocomputing (2015)
3. Bo, Antonio Padilha Lanari et al: Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS/2011 (2011)
4. Madgwick, S.O.H.: An efficient orientation filter for inertial and inertial/magnetic sensor arrays. (2010)
5. M. Grigorie, C. de Raad, F. Krummenacher, and C. Enz: Analog Temperature Compensation for Capacitive Sensor Interfaces pp. 14, 1996.
6. S. Gebhardt, G. Scheinert, and F. H. Uhlmann: Temperature influence on capacitive sensor structures. INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING - DEVICES AND SYSTEMS, MATERIALS AND TECHNOLOGIES FOR THE FUTURE Startseite, 2006
7. iFixIt: Xbox 360 Kinect Teardown - iFixit. Available: `https://www.ifixit.com/Teardown/Xbox+360+Kinect+Teardown/4066` [Accessed: 12-Apr-2016]
8. Reichinger, A.: Kinect Pattern Uncovered. Available: `https://azttm.wordpress.com/2011/04/03/kinect-pattern-uncovered/` [Accessed: 12-Apr-2016]
9. Stackoverflow Community: Precision of the kinect depth camera. Available: `http://stackoverflow.com/questions/7696436/precision-of-the-kinect-depth-camera` [Accessed: 12-Apr-2016]
10. Shpunt, A. and Zalevsky, Z.: Depth-varying light fields for three dimensional sensing. Available: `http://www.freepatentsonline.com/y2008/0106746.html` [Accessed: 12-Apr-2016]
11. Freedman, B., et al.: Depth Mapping Using Projected Patterns. Available: `http://www.freepatentsonline.com/y2010/0118123.html` [Accessed: 12-Apr-2016]
12. Shpunt, A.: Depth Mapping Using Multi-Beam Illumination. Available: `http://www.freepatentsonline.com/y2010/0020078.html` [Accessed: 12-Apr-2016]
13. DiFilippo, N. M. and Jouaneh, M. K.: Characterization of Different Microsoft Kinect Sensor Models. IEEE Sensors Journal, vol 15, 8, 4554-4564 (2015)
14. Rzeszotarski, D and Strumiłło, P., et al.: System Obrazowania Stereoskopowego Sekwencji Scen Trójwymiarowych. Zesz. Nauk. Elektron. (2006)

15. Fofi, D and Sliwa, T and Voisin, Y: A comparative survey on invisible structured light. SPIE Electron. ImagingMachine Vis. Appl. Ind. Insp. XII San José USA (2004)
16. Suarez, J. and Murphy, R. R.: Using the Kinect for search and rescue robotics. 2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (2012)
17. Kitsikidis, A. , et al.: Dance Analysis using Multiple Kinect Sensors (2011).
18. Asteriadis, S. , et al.: Estimating human motion from multiple Kinect sensors. Proc. 6th Int. Conf. Comput. Vis. / Comput. Graph. Collab. Tech. Appl. - MIRAGE '13 (2013)
19. Schröder, Y. , et al.: Multiple Kinect Studies Technical Report (2011).