Journal of Computer Networks, Architecture and High Performance Computing
Volume 3, Number 2, July 2021
https://doi.org/10.47709/cnahpc.v3i2.1067

Motion capture in humanoid model with Unity engine using Kinect V2

Amir Mahmud Husein[1)*], Jimmy Ciawi[2)]
[1)2)] Universitas Prima Medan, Indonesia
[1)] amirmahmud@unprimdn.ac.id [2)] jimmyciawly@gmail.com

ABSTRACT

Animation is a technique to create illusion of motion, which is created by displaying a series of motionless pictures in sequence or using a spine/bones to create motion that looks real. All this time, the process of creating an animation still using traditional technique which requires special skills and take some time to finish a complicated animation which is used in movies or video games. Motion capture is an animation-making technique by tracking every part of body in order to find position and rotation of human's joints which is generated by the image from the sensor. Motion capture has lots of method, such as marker base technique which use mark to track any motions. Motion capture markerless method that can capture or track motions without using any marks. Motion capture with markerless technique can be done by using RGB-Depth's camera censor, which is by using Microsoft Kinect V2 with Kinect V2 SDK in order to make Kinect connected with computer, and using Unity Engine, a game engine that has already provided animation timeline and supports any 3D format which contains animation file that can be used in mostly other models which is using humanoid. In order to obtain motion data which will be changed into skeleton joint data, we will use connector OpenNI and 3D Collada model with (.dae) format because Collada is 3D format Open Source which is built using XML-based so that it can easily read and written back into 3D file as an output.

Keywords: Animation, Collada, Video Game, Motion Capture, Motion Capture Markerless, OpenNI, Unity Engine.

INTRODUCTION

Animation is a technique for creating the illusion of movement, which is produced by displaying a series of still images in sequence or using the spine / bones to create a movement that looks real (Ariyati & Misriati, 2016). Animation is often used to make movies or video games. There are various kinds of techniques or methods on the basis of 2D or 3D such as Cel Animation Technique, 2D Computing Technique, 3D Computing Technique. However, the development of technology, many films and video games that use 3D-based images, has led to the emergence of motion capture techniques. Motion Capture is a technique of capturing the movement of real objects such as the anatomy of the Human body to create a movement that is more natural and precise. Using motion capture is also very easy compared to traditional animation techniques so that it can simplify and speed up the animator's work (Özdem, 2016).

Because using motion capture is very easy, with very high detail and is widely used in making animation in the film industry, video games, virtual / augmented reality or mixed reality (Chung et al., 2018), biomedical (Procházka et al., 2016), biomechanics (Napoli et al., 2017) and others so that the use of motion capture technology is expensive (Patrizi et al., 2016), and not all the film and video game industries can use it. Motion Capture which is usually used is called Optical Motion Capture (Hegarini et al., 2017) which is marker-based or mark-based using several optical sensors and passive markers coated with retro-reflective material (Chatzitofis et al., 2019) to reflect light or active has electricity or energy to emit light and infrared emitters on special clothes placed on the subject's body to be captured (Hegarini et al., 2017). Some examples of the use of motion capture in the film industry are The Lord Of The Ring, Transformer, Avatar, while in the video game industry are Street Fighter V, Tekken 7, Monster Hunter World and others.

Motion Capture Marker Based that uses tools such as clothes, transmitters and also uses a camera will make users feel uncomfortable and have difficulty doing motion capture. In this study, the markerless method is applied, namely

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

the motion capture method without using markers to capture or trace motion which can simplify and save work costs (Özdem, 2016) by utilizing a depth sensor camera, as for several products, namely, Intel Realsense (4). al., 2017), Orbbec (Călin & Coroiu, 2018), Nvidia Jetson stereo camera (Saini et al., 2019), and Microsoft Kinect which are cheaper (Patrizi et al., 2016) than Optical Motion Captur. Therefore, this study will use an RGB-Depth Sensor Camera, namely Microsoft Kinect V2 or Microsoft Kinect Xbox One which is used in game development technology or game engine, namely Unity Engine (Capece et al., 2018), (Borromeo, 2020), by using 3D data with collada format (.dae) as a Skeletal model (Win & Thein, 2020) which will be a preview and the output of this research.

LITERATURE REVIEW

Unity

Unity Engine, is a free game engine that can be used in the creation of 2D and 3D games as well as various industrial projects other than games such as filmmaking, animation, architecture and manufacturing. and also supports a variety of devices (cross-platform) written in the C ++ programming language (runtime) and development using the C # programming language (Borromeo, 2020), (Buyuksalih et al., 2017).

Unity also supports various 3D formats, such as Autodesk (.fbx), Blender (.blend), Obj, and even Collada (.dae) and Unity also provides a timeline for animation creation and animation review, making it very easy to create games (Buyuksalih et al., 2017). Unity also provides a prefab feature that can store various assets and conditions that were previously created and can be used at any time so there is no need to re-create a job.

Motion Capture

Motion Capture is a tool used to record a movement and displacement by recording various angular movements, rotation of the segments of the body parts in a living object to get motion data that will be processed by a computer and can be used by the animation process in a 3D model (Patrizi et al. al., 2016), (Hegarini et al., 2017).

In making animation and games, motion capture is usually used for the movement of Humanoid 3D objects whose body structures are like humans, by connecting the motion capture to the humanoid model, you will get animation or movement results that are very real life and not rigid.

Microsoft Kinect

Kinect is a depth sensor camera that combines an RGB and infrared camera that is used to play games on the Xbox console produced by Microsoft and developed by PrimeSense (Călin & Coroiu, 2018).

Apart from being a useful controller for playing video games on the Xbox console, now Microsoft has also provided an SDK (Software Development Kit) that can be used by various developers in making software or programs to use the Kinect hardware (Călin & Coroiu, 2018).

Collada

Collada is an open source file format used in 3D applications with the extension .dae (digital asset exchange) (Chelyshkov et al., 2021). Collada files are not like other 3D files whose contents have been compiled into binaries, Collada files are built using an XML core that can be read by various programming languages so that almost all 3D applications and game engines can use them.

OpenNI

OpenNI (Open Natural Interface) is an open source SDK used for the development of 3D sensing middleware libraries and applications for depth sensors such as Kinect (Călin & Coroiu, 2018). OpenNI to create a user interface between computer and human interactions using natural input. OpenNI requires a variety of artificial training to detect natural inputs provided by humans. Naturally, this means that humans can carry out an activity directly, such as moving their hands, body, singing, talking, and making faces, and computers can get data from these various activities.

A literature review is a critical, analytical summary and synthesis of the current knowledge of a topic. It should compare and relate different theories/research, findings, and so on, rather than just summarize them individually. It should also have a particular focus or theme to organize the review. In this section, the researcher can describe some

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

of the related previous studies. Researchers can review the gaps in the research, then it can be used as a basis for research to be carried out

METHOD

Development Model

Extreme Programming is a method that is included in the software development model approach that simplifies various stages in the development process so that it becomes more adaptive and flexible, with a minimalist design, gradual testing, and documentation that is not excessive (Fatoni & Dwi, 2016), (Hameed, 2016)). The Extreme Programming model is very suitable for the development process and the results that have been made in implementing motion capture in the 3D model used (Hameed, 2016), (LeMay, 2019).

Design

In order for Motion Capture to work and be able to connect with Unity, and produce usable animation, it must go through various stages such as installing the Kinect SDK, setting up the Kinect V2 hardware to be able to record the body properly, reading the Skeleton Joint Data generated by Kinect, then connecting it to 3D Humanoid model specified, and display design for easy use. The following is a flowchart of the input to output motion capture stages using Kinect V2 and Unity, as follows:

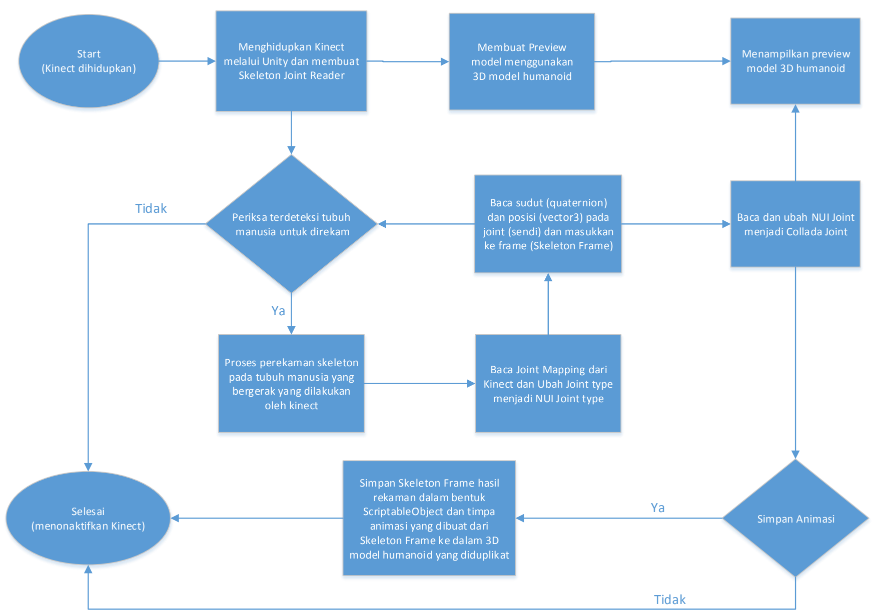

Fig. 1 Research Flowchart

Coding

The coding process is carried out using an object-oriented programming paradigm and using various abstract or derived classes because serialized data has almost the same functions and properties using the C # programming language (Taher, 2019).
Coding using Visual Studio Code with the extension C # for Visual Studio Code and the code will be compiled in the Unity Engine, coding can also use other text editors such as Visual Studio (Ritchie, 2016), MonoDevelop (Halpern, 2019), Rider (Borromeo, 2020) or Another text editor that supports the C # programming language. Unity Editor-based coding that uses the Unity GUI Layout (Kupiainen, 2018) which can be compiled immediately and run in Unity or commonly called the extension editor without having to build or in play mode, making it easier for users other than the author himself (Kupiainen, 2018).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Testing

The purpose of the testing is to find out how the results of performance, joint and mapping, the number of frames generated in the 3D animation data generated using Kinect V2 in the design that has been implemented.

Tools and Materials

Software Requirement

At this stage, the collection and analysis of needs are carried out to design and create a motion capture program prototype that will be made. Software, hardware and tools that will be used are as follows:

a. *Hardware*
   1. Laptop Asus TUF Gaming FX504GD Core i5-8300H (4 *core*, 8 *threat*) dengan pengolah grafis Nvidia GTX 1050 (*non* Ti).
   2. Kinect V2 dengan adaptor Windows USB 3.0 *Controller*.
   3. Tripod Takara ECO-173A.

b. *Software*
   1. Windows 10 (Sebagai Sistem Operasi)
   2. Unity Engine versi 2019.4.14f1
   3. Visual Studio Code
   4. Kinect v2 SDK untuk Windows

Kinect Spesification

The Kinect consists of several sensors, including a red-green-blue (RGB) digital camera, a microphone, and most importantly, a 3D range camera sensor, which can provide both range (X, Y, Z) and amplitude images (Kuan et al., 2019). In this study, the Kinect V2 will be used. The comparison of specifications between the Kinect V2 and the previous version of Kinect (Wasenmüller & Stricker, 2017) is as follows:

Feature	Kinect V1	Kinect V2
Color Camera	640p x 480p, 30 fps	1920p x 1080p, 30 fps
Infrared Camera (Depth Camera)	320p x 240p	512p x 424p
Max. Depth Distance	± 4 meter	± 4.5 meter
Min. Depth Distance	80 cm	50 cm
Horizontal Field of View	57 degrees	70 degrees
Vertical Field Of View	43 degrees	60 degrees
Skeleton Joint Defined	20 joint	25 joint
Full Skeleton Tracked	2	6

3D Data

3D data is included in the humanoid 3D model which serves as a preview as well as an animation output template produced in this study. The following is an image of the 3D model used in this research:

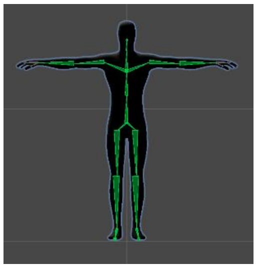

Fig. 2 Model 3D *Humanoid*

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

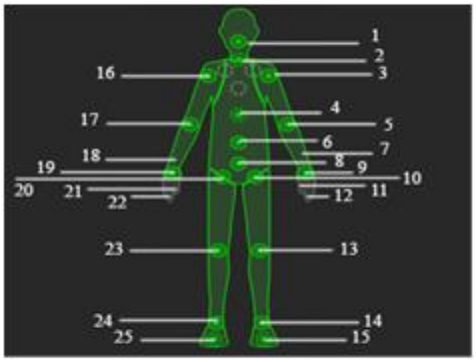

Fig. 3 *Skeletal tracking* atau letah *joint* pada tubuh

Movement Detection

Movement detection or skeletal tracking is the process of recording motion detection based on the rotation of the joint or the meeting point of the limb and the points or joints can be described in 3D coordinates such as X, Y, Z (Win & Thein, 2020). Some parts of the body that can be recognized by the Kinect V2 are:

a. Head
b. Neck
c. Shoulder Left
d. Spine Shoulder
e. Elbow Left
f. Spine Mid
g. Hand Left
h. Spine Base
i. Wrist Left
j. Hip Left
k. Thum Left
l. HandTip Left
m. Knee Left
n. Ankle Left
o. Foot Left
p. Shoulder Right
q. Elbow Right
r. Hand Right
s. Wrist Right
t. Hip Right
u. Thumb Right
v. HandTip Right
w. Knee Right
x. Ankle Right
y. Foot Right

RESULT

In this research, the Kinect V2 camera sensor will be tested to work properly so that the results of the research do not fail or error. Tests will be carried out on various lighting conditions run through Unity whether it can detect joints on the user's body and can be seen from the following image:

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

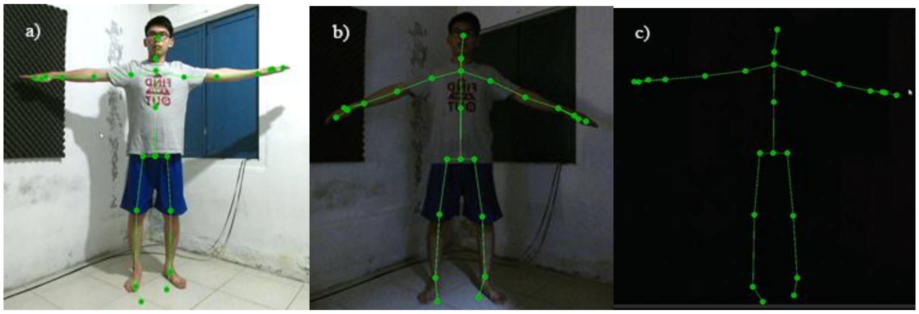

Fig. 5 Skeletal testing under various lighting conditions, a) Bright light, b) 13 watts of light, c) No light

In testing the jump motion, it can produce an animation with a number of 160 frames (60 frames per second), so the jump motion takes 2 seconds 40 milliseconds.

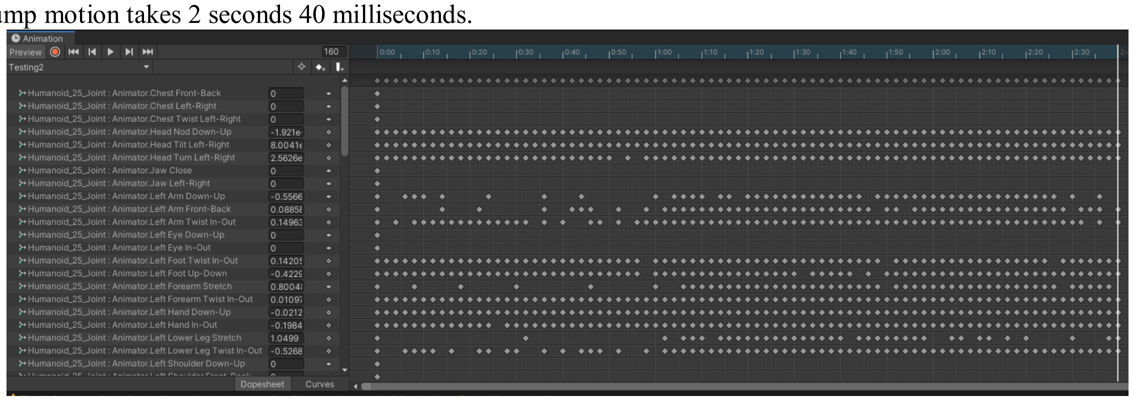

Fig. 6 Animated timeline result from research.

DISCUSSIONS

Testing of joint data on the humanoid 3D model generated based on the application of motion capture research using the Kinect V2 can be carried out precisely and the joint does not exchange or cannot be detected. Even though the names are different, the 3D rotation (X, Y, Z) and the translation of each joint are the same so that what is needed is a naming match so that the application of the motion capture animation results can work properly.

Table 1
Mapping Collada Joint dan NUI Joint

Collada Joint	**NUI Joint**
SpineBase	Waist
SpineMid	Torso
SpineShoulder	SpineShoulder
Neck	Neck
Head	Head
HipRight	RightHip
KneeRight	RightKnee
AnkleRight	RightAnkle
FootRight	RightFoot
HipLeft	LeftHip

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

KneeLeft	LeftKnee
AnkleLeft	LeftAnkle
FootLeft	LeftFoot
ShoulderRight	RightShoulder
ElbowRight	RightElbow
WristRight	RightWrist
HandRight	RightHand
HandTipRight	HandTipRight
ThumbRight	ThumbRight
ShoulderLeft	LeftShoulder
ElbowLeft	LeftElbow
WristLeft	LeftWrist
HandLeft	LeftHand
HandTipLeft	HandTipLeft
ThumbLeft	ThumbLeft

After matching the naming or mapping, the translation and rotation of each joint will be synchronized with the 3D model used. Here is an image of the 3D model and skeleton displayed by Kinect. The result of the motion capture is an animated 3D file that can be reused on all humanoid 3D models in the Unity supported file format. The movement used for the experiment is jumping with open arms.

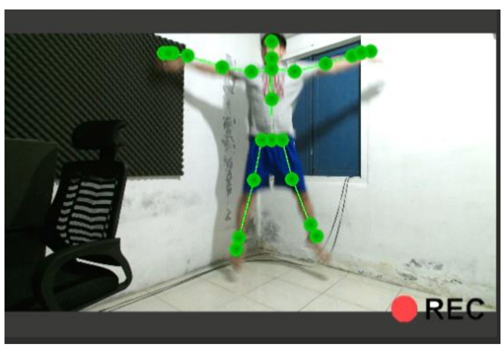

Fig.7 Users with skeletal joints perform jumping movements.

Fig. 8 3D model that follows the user's movements while performing a jumping motion

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnahpc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

CONCLUSION

In the research that has been done, the Kinect V2 can easily capture motion from the 25 joint points which allows for markerless motion capture and without lighting it can also detect limbs because the Kinect is a device consisting of various sensors such as a camera. RGB, a depth sensor (Depth Sensor) so that even in minimal lighting conditions, motion capture can be carried out smoothly.

By using the Unity Engine in this study, which has provided an animation timeline so that it can easily cut and modify the resulting animation according to the wishes and needs of the user. And using Collada's 3D model as a template can make it easier to read or rewrite the translation and rotation of the humanoid model.

REFERENCES

Ariyati, S., & Misriati, T. (2016). Perancangan Animasi Interaktif Pembelajaran Asmaul Husna. *Jurnal Teknik Komputer Amik Bsi*, *II*(1), 116–121.

Borromeo, N. A. (2020). *Hands-On Unity 2020 Game Development*.

Buyuksalih, I., Bayburt, S., Buyuksalih, G., Baskaraca, A. P., Karim, H., & Rahman, A. A. (2017). 3D MODELLING and VISUALIZATION BASED on the UNITY GAME ENGINE - ADVANTAGES and CHALLENGES. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(4W4), 161–166. https://doi.org/10.5194/isprs-annals-IV-4-W4-161-2017

Călin, A. D., & Coroiu, A. (2018). Interchangeability of kinect and orbbec sensors for gesture recognition. *Proceedings - 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing, ICCP 2018*, 309–315. https://doi.org/10.1109/ICCP.2018.8516586

Capece, N., Erra, U., & Romaniello, G. (2018). A Low-Cost Full Body Tracking System in Virtual Reality Based on Microsoft Kinect. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10851 LNCS*, 623–635. https://doi.org/10.1007/978-3-319-95282-6_44

Chatzitofis, A., Zarpalas, D., Kollias, S., & Daras, P. (2019). Deepmocap: Deep optical motion capture using multiple depth sensors and retro-reflectors. *Sensors (Switzerland)*, *19*(2), 1–26. https://doi.org/10.3390/s19020282

Chelyshkov, P., Volkov, S. A., & Babushkin, E. S. (2021). Analysis of world and domestic experience in the use of XML schemas in the implementation of information interaction during maintaining the information model of a capital construction object. *IOP Conference Series: Materials Science and Engineering*, *1030*, 012067. https://doi.org/10.1088/1757-899x/1030/1/012067

Chung, S., Cho, S., & Kim, S. (2018). Interaction using Wearable Motion Capture for the Virtual Reality Game. *Journal of The Korean Society for Computer Game*, *31*(3), 81–89. https://doi.org/10.22819/kscg.2018.31.3.010

Fatoni, A., & Dwi, D. (2016). Rancang Bangun Sistem Extreme Programming Sebagai Metodologi Pengembangan Sistem. *Prosisko*, *3*(1), 1–4. http://e-jurnal.lppmunsera.org/index.php/PROSISKO/article/view/116

Halpern, J. (2019). Developing 2D Games with Unity. In *Developing 2D Games with Unity*. https://doi.org/10.1007/978-1-4842-3772-4

Hameed, A. (2016). Software Development Lifecycle for Extreme Programming. *International Journal of Information Technology and Electrical Engineering ITEE*, *5*(1), 7–13.

Hegarini, E., Mutiara, A. B., Suhendra, A., Iqbal, M., & Wardijono, B. A. (2017). Similarity analysis of motion based on motion capture technology. *2016 International Conference on Informatics and Computing, ICIC 2016*, 389–393. https://doi.org/10.1109/IAC.2016.7905750

Keselman, L., Woodfill, J. I., Grunnet-Jepsen, A., & Bhowmik, A. (2017). *Intel R RealSense TM Stereoscopic Depth Cameras*.

Kuan, Y. W., Ee, N. O., & Wei, L. S. (2019). Comparative study of intel R200, Kinect v2, and primesense RGB-D sensors performance outdoors. *IEEE Sensors Journal*, *19*(19), 8741–8750. https://doi.org/10.1109/JSEN.2019.2920976

Kupiainen, H. (2018). *EXTENDING UNITY GAME ENGINE THROUGH EDITOR SCRIPTING*. *October*, 4.

LeMay, M. (2019). *Agile for Everybody*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

Napoli, A., Glass, S., Ward, C., Tucker, C., & Obeid, I. (2017). Performance analysis of a generalized motion capture system using microsoft kinect 2.0. *Biomedical Signal Processing and Control*, *38*, 265–280.

* Corresponding author

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

https://doi.org/10.1016/j.bspc.2017.06.006
Özdem, A. (2016). Motion Capture. *Вестник Новых Медицинских Технологий*, *XXIII*(2), 1–9.
Patrizi, A., Pennestrì, E., & Valentini, P. P. (2016). Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics. *Ergonomics*, *59*(1), 155–162. https://doi.org/10.1080/00140139.2015.1057238
Procházka, A., Schätz, M., Vyšata, O., & Vališ, M. (2016). Microsoft Kinect visual and depth sensors for breathing and heart rate analysis. *Sensors (Switzerland)*, *16*(7), 1–11. https://doi.org/10.3390/s16070996
Ritchie, P. (2016). Practical Microsoft Visual Studio 2015. *Practical Microsoft Visual Studio 2015*, 1–25. https://doi.org/10.1007/978-1-4842-2313-0
Saini, N., Price, E., Tallamraju, R., Enficiaud, R., Ludwig, R., Martinovic, I., Ahmad, A., & Black, M. J. (2019). *Markerless Outdoor Human Motion Capture Using Multiple Autonomous Micro Aerial Vehicles*.
Taher, R. (2019). *Hands-On Object-Oriented Programming with C #*. Packt Publishing Ltd.
Wasenmüller, O., & Stricker, D. (2017). Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10117 LNCS*, 34–45. https://doi.org/10.1007/978-3-319-54427-4_3
Win, S., & Thein, T. L. L. (2020, February 1). Real-Time Human Motion Detection, Tracking and Activity Recognition with Skeletal Model. *2020 IEEE Conference on Computer Applications, ICCA 2020*. https://doi.org/10.1109/ICCA49400.2020.9022822

* Corresponding author