Vision-Based Human Posture Detection from a Virtual Home-Care Unmanned Aerial Vehicle

Andrés Bustamante[1], Lidia M. Belmonte[1,2], António Pereira[3,4], Pascual González[1,5,6], Antonio Fernández-Caballero[1,2,6], and Rafael Morales[1,2(✉)]

[1] Instituto de Investigación en Informática de Albacete, Unidad Multidisciplinar en Neurocognición y Emoción, 02071 Albacete, Spain
[2] Universidad de Castilla-La Mancha, E.T.S. Ingenieros Industriales de Albacete, 02071 Albacete, Spain
Rafael.Morales@uclm.es
[3] Polytechnic Institute of Leiria, Computer Science and Communications Research Centre, School of Technology and Management, 2411-901 Leiria, Portugal
[4] INOV INESC INOVAÇÃO, Institute of New Technologies—Leiria Office, 2411-901 Leiria, Portugal
[5] Universidad de Castilla-La Mancha, Escula Superior de Ingeniería Informática de Albacete, 02071 Albacete, Spain
[6] Biomedical Research Networking Centre in Mental Health (CIBERSAM), 28016 Madrid, Spain

Abstract. Monitoring is essential to provide assistance to people who require home care due to their age or health condition. This paper presents the vision-based detection of three postures of a person (standing, sitting and laying down) from an unmanned aerial vehicle. The proposal uses the MediaPipe Pose Python module, considering only seven skeleton points and a set of trigonometric calculations. The work is evaluated in a Unity virtual reality (VR) environment that simulates the monitoring process of an assistant UAV. The images acquired by the UAV's on-board camera are sent from the VR visualiser to the Python module via the Message Queue Telemetry Transport (MQTT) protocol. The simulation shows very promising results for the detection of a person's postures.

Keywords: Unmanned aerial vehicle · Home assistance · Computer vision · Human posture · Virtual reality

1 Introduction

The elderly represent the population group with the highest level of dependency and need for care. They are prone to physical and mental disability or deterioration [7]. Moreover, they need adequate supervision in case of an accident or any other need. However, continuous supervision leads to work overload for carers, both in specialised centres and at home. In addition, due to the high costs of specialised care, family members often have to take care of the dependent persons

J. M. Ferrández Vicente et al. (Eds.): IWINAC 2022, LNCS 13259, pp. 482–491, 2022.
https://doi.org/10.1007/978-3-031-06527-9_48

themselves. All this has led to research of strategies to reduce the work overload of caregivers [8].

Several research projects have focused on supporting caregivers through the use of various technologies for monitoring dependent persons. Among them is the use of computer vision in real time. One such project was to detect whether a person was eating, observing or taking their medication, and to notify the caregiver of the occurred events [21]. This proposal demonstrates the usefulness of image processing where objects and actions are distinguished through the use of colour-based computer vision algorithms.

Other works have focused on the search for human activity recognition using different devices, from smartphones, wearables, video and electronic components to more innovative systems based on WiFi or assistance robots [19]. Interestingly, 60% of the technological monitoring solutions are based on computer vision through different camera types. Although visual monitoring has proven to be a viable and popular option, its implementation within a home would require a large number of cameras. Innovative approaches are therefore emerging in which unmanned aerial vehicles (UAVs) using robust trajectory planning to fly safely in indoor environments are able to monitor dependent people via an on-board camera [5].

Conducting these kinds of experiments indoors with drones can be dangerous in real environments. For this reason, research has initially focused on 3D environments through the development of a virtual reality (VR) platform [2,3]. Thanks to this approach, the benefits of using drones as assistant robots for monitoring dependent people in a realistic and safe virtual environment can be evaluated. The platform is based on real-time communication through the Message Queue Telemetry Transport (MQTT) protocol of the various modules implemented to recreate the behaviour of an autonomous vision-based UAV for monitoring dependent people. One of the main modules is the computer vision module in charge of processing the images captured by the UAV's on-board camera to detect various states of the person [12]. This article complements previous research, focusing on image processing to detect postures of the monitored person. We also use the MQTT protocol to transfer the images from the virtual environment to image processing in Python.

2 Monitoring Dependent People at Home from UAVs

Our research revolves around the use of small vision-based UAVs to assist dependent people at home [2–6,12]. The main objective is to monitor the person in order to determine their condition and possible required assistance. Another alternative for monitoring would be the use of static cameras. However, this would require deploying multiple cameras in the home to avoid dead spots and, in addition, the ability to detect the person would be reduced as the person moves away from the camera. Therefore, a moving aerial robot has the potential to cover a larger area and monitor more closely and efficiently than a number of static cameras.

Fig. 1. The three postures to be identified (standing, sitting and laying down).

UAVs are useful tools that have been used in conjunction with image processing in various research areas. For instance, UAVs have been used with image processing based on colour detection algorithms in outdoor environments for human body detection [18]. However, in indoor scenarios people monitoring is problematic when using only colour-based algorithms due to the complexity and large number of objects and colours found in a house. On the other hand, through image processing it is possible to detect objects based on colour [21] or recognise mood on faces [12], but the detection of human posture is more complex.

In addition to cameras to effectively monitor a person, some hardware devices incorporating different senors have been used. For example, the Microsoft Kinect has a depth sensor that allows a more optimised tracking of the person human skeleton [13,20]. Some UAVs currently carry depth sensors to avoid obstacles and even track people, as is the case of the DJI drones [22]. However, for the detection and estimation of human poses solutions solely using conventional cameras have also been proposed. This is the case of a model for suspicious movements detection in people through posture estimation. The algorithm takes 3.4 s for the detection of a single person, without considering the extra time required to process and estimate the postures [15].

For an efficient monitoring of dependent people less time is required. This is why our proposal focuses on using a promising algorithm for human detection and pose estimation that was born from a very recent research [23]. The ultimate objective is to determine if a person is standing, sitting or laying down though only processing colour images captured by a UAV's camera. Figure 1 illustrates the three postures to be identified during the monitoring process in the VR platform. It should be noticed that dependent people are sitting or laying down most of the time. Having a system that allows the person to be detected efficiently and differentiated from other objects and recognise their current posture is an advance in monitoring, supervision and alarm of elderly or dependent people through affordable devices.

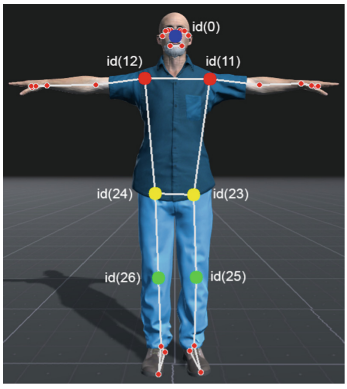

Fig. 2. Landmarks highlighting the relevant joints: nose = id(0); left shoulder = id(11); right shoulder = id(12); left hip = id(23); right hip = id(24); left knee = id(25); right knee = id(26).

3 Computer Vision Algorithms

This section describes the computer vision algorithms implemented to detect whether the person is standing, sitting or laying down. First, the MediaPipe framework used to obtain the required key points of the human skeleton is introduced. Then, the key points are used to detect the human avatar in the images obtained from the virtual scene in Unity. Finally, we describe how the three different postures are identified.

3.1 MediaPipe

MediaPipe is a framework for building multimodal applied machine learning pipelines. It provides solutions for different kinds of applications like face detection [11], hand bones detection and tracking [10] and so on. In this paper, the Pose library of MediaPipe is used to map the 3D pose landmarks which estimate the joints of the human skeleton. The library generates up to 33 landmarks, each one with a unique id. But, in our solution only seven relevant joints are used to determine one of the three searched postures of the human being (standing, sitting and laying down). The landmarks used are illustrated on the human avatar of Fig. 2.

3.2 Human Detection

The VR visualiser module developed in Unity as part of the VR simulation platform of the assistant UAV transmits the images captured by the UAV's on-board camera via MQTT to the new computer vision module. This module programmed in Python processes each image with MediaPipe and obtains the relevant points on the skeleton of the human avatar. The recognition of the

Fig. 3. Skeleton landmarks in the three postures (sitting, standing, and laying down).

avatar skeleton is shown in Fig. 3 for the three different positions to be estimated. It should be noted that the algorithm perfectly scans and generates the avatar skeleton in each of the positions evaluated in this work in a room that has several background colours. This would represent an obstacle for any algorithm where false positives could be generated [9], requiring additional processing to reduce these false alarms [16]. The efficiency of the visual computation in the MediaPipe algorithm is remarkable, as complex systems or multi-camera scanning are usually utilised to achieve this type of complete human mapping [17].

Figure 4 shows a block diagram of the solution implemented in Python to detect the position of the avatar from an image received from the UAV camera monitoring the avatar at the virtual house. Once the image captured in Unity reaches Python, it is scanned by the MediaPipe algorithm where the skeleton reference points are obtained. Of these, only seven relevant points are considered to estimate the posture of the avatar corresponding to shoulders, hips, knees and nose. From the appropriate identifiers, the (x, y) position of these points in the 2D image plane is calculated. These coordinates are used to determine the posture of the avatar, as will be detailed in the next section.

3.3 Posture Detection

Once the coordinates of each point in the 2D plane are obtained, it is possible to estimate the avatar's posture from the position of the selected points and the angles generated among all the points by using simple trigonometry, which represents an enormous simplicity compared to the use of more advanced algorithms or classifiers [1,14].

Firstly, the height of the shoulder points, id(11) and id(12), are used to determine the direction of the avatar, since with a slight rotation both shoulders differ in height in the 2D plane. The direction of the avatar determines the points in the left or right side of the avatar that are used in the remaining calculations: left points (ids 11, 23, 25) when the avatar is turned to its right (left of the

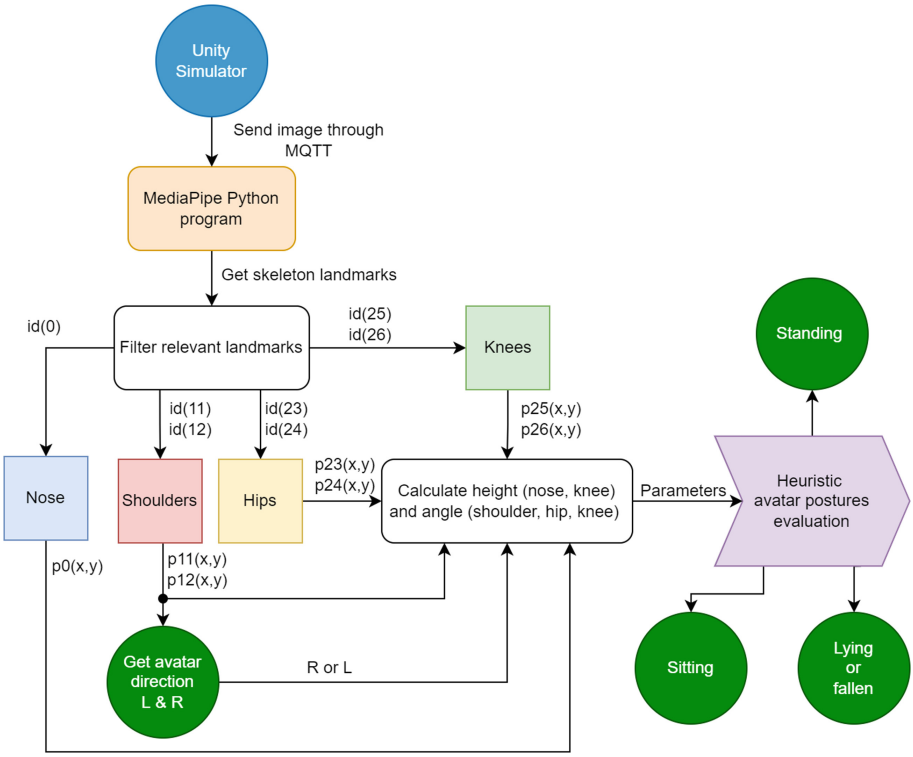

Fig. 4. Block diagram of the implemented solution.

image), or right points (ids 12, 24, 26) when the avatar is turned to its left (right of the image).

Secondly, the difference in height between the points of the nose and the knee on the side of the skeleton as determined by the avatar's direction is analysed. After trial and error adjustment, it has been determined that if the distance is less than 20 pixels, the person will be laying down (see Fig. 5a). However, if this distance is greater or equal than 20 pixels, the person will be in a standing or sitting posture. It should be noted that the tests have been carried out when the person is laying down with a horizontal direction of the body (the imaginary line that would join the head with a foot). When this direction changes and approaches vertically, the laying posture looks very similar to the standing posture in a 2D image (see Fig. 5b), leading to estimation errors that will need to be resolved in future work. Finally, between the standing and sitting positions, the angles formed among the shoulder-hip-knee points are very noticeable (see Fig. 5b and 5c). Therefore, the differentiation of these postures is possible by simply measuring this angle. If the angle formed is higher than 140 °C, the avatar's posture is standing while if the angle is less than 140 °C, the avatar's posture is sitting.

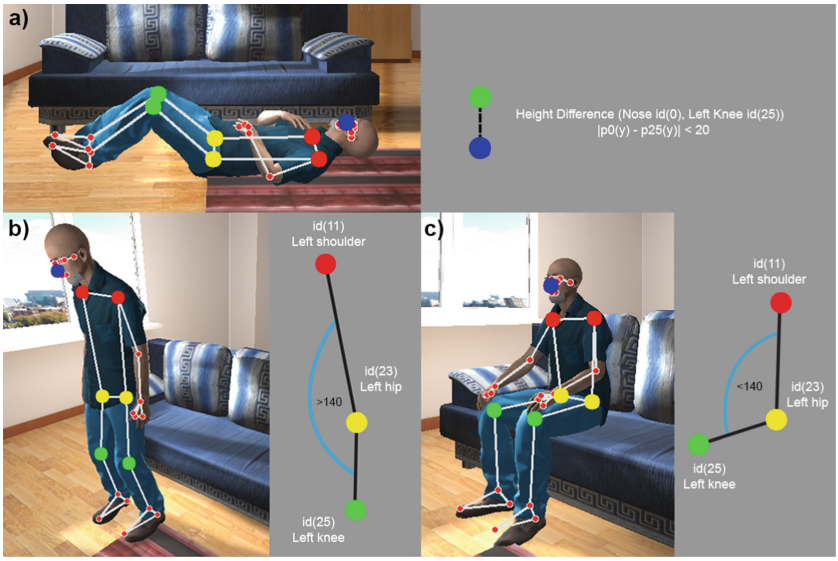

Fig. 5. a) Height condition in laying down posture. b) Angles in the standing posture. c) Angles in the sitting posture.

4 Preliminary Results

This section introduces some preliminary results of the solution implemented to detect the human pose from the monitoring process of the assistant UAV. The tests have been performed on the VR platform, where the images from the UAV camera in the virtual scenario in Unity are sent via MQTT to the computer vision module programmed in Python. Here, the objective is to first determine the direction of the avatar in the 2D image captured by the camera, meaning the side to which the avatar's body is turned, either to the left or to the right. Then, considering this direction and the angles formed by the three points of the shoulder, knee and hip joints, as well as the height of the nose in relation to the knee, the aim is to determine and differentiate the avatar's posture.

In order to evaluate the performance of the computer vision solution, different tests have been carried out considering the three possible positions of the avatar, and also considering that the avatar is turned and placed on opposite sides of the room, where its direction and skeletal points change. The results in all cases have been positive, as it has been possible to correctly determine the direction and posture of the avatar, as shown in Fig. 6. This figure shows the windows generated by the OpenGL library from the Python program. In the upper part you can see the result of the avatar in the three positions in which it is slightly turned to its left, and in the lower images the same result is shown, but in another side of the room in which it is turned to its right.

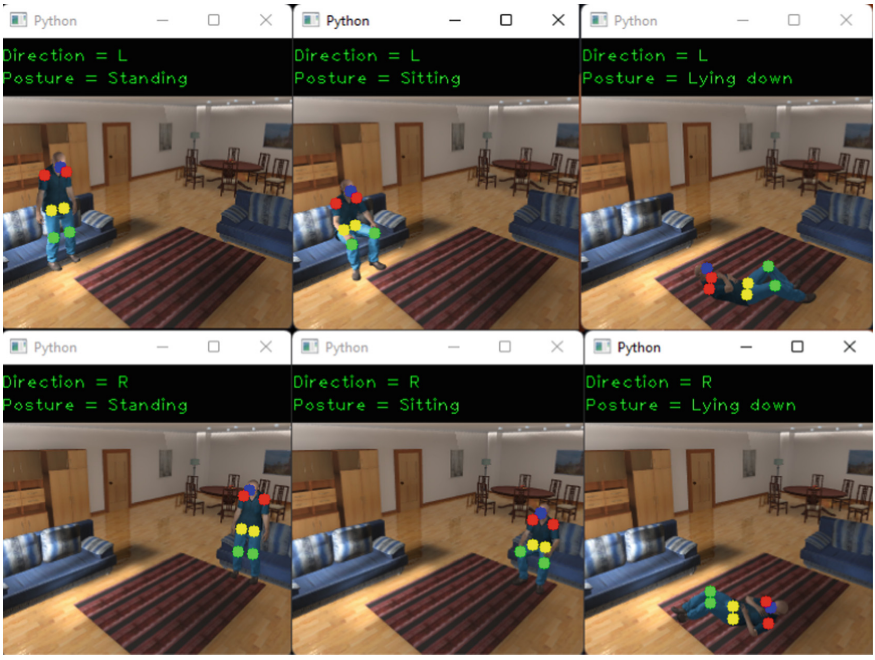

Fig. 6. Posture detection results

5 Conclusions

This article has introduced a computer vision solution for the detection of a person's posture monitored by a UAV for home care. The developed solution is based on the use of the Pose library of MediaPipe and allows differentiating between three possible postures: standing, sitting and laying down. It performs a series of trigonometric calculations by considering relevant reference points in the human skeleton. The solution has been implemented in the computer vision module programmed in Python for the VR platform, which simulates the process of monitoring a dependent person from a small drone in a virtual home.

The first evaluation results of the programmed solution are satisfactory. Furthermore, it should be noted that MediaPipe Pose library promises an optimal and fast recognition of the human body and can therefore be implemented in real-time systems. In addition, the skeleton is generated progressively as the human body is displayed on the camera, which prevents the algorithm from stopping due to incomplete visualisation of the body. The points obtained from the library cover the entire human skeleton and can be used in future work for more extensive posture recognition. One of the main areas for improvement is to extend the recognition of the laying position to other situations where the direction of the body is not horizontal. Another future work will be the estimation of the distance of the person with respect to the UAV's camera in order to use this information for the drone's trajectory planner.

Acknowledgements. Grants PID2020-115220RB-C21 and EQC2019-006063-P funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way to make Europe". This work was also partially supported by CIBERSAM of the Instituto de Salud Carlos III. This work has also been partially supported by Portuguese Fundação para a Ciência e a Tecnologia - FCT, I.P. under the project UIDB/04524/2020 and by Portuguese National funds through FITEC - Programa Interface, with reference CIT "INOV - INESC Inovação - Financiamento Base". This work has also been partially supported by Junta de Comunidades de Castilla-La Mancha/ESF (grant No. SBPLY/21/180501/000030).

References

1. Belagiannis, V., Zisserman, A.: Recurrent human pose estimation. In: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017, pp. 468–475 (2017). https://doi.org/10.1109/FG.2017.64
2. Belmonte, L.M., García, A.S., Morales, R., de la Vara, J.L., López de la Rosa, F., Fernández-Caballero, A.: Feeling of safety and comfort towards a socially assistive unmanned aerial vehicle that monitors people in a virtual home. Sensors **21**(3) (2021). https://doi.org/10.3390/s21030908
3. Belmonte, L.M., García, A.S., Segura, E., Novais, P., Morales, R., Fernández-Caballero, A.: Virtual reality simulation of a quadrotor to monitor dependent people at home. IEEE Trans. Emerg. Top. Comput. **9**(3), 1301–1315 (2021). https://doi.org/10.1109/TETC.2020.3000352
4. Belmonte, L.M., Morales, R., García, A.S., Segura, E., Novais, P., Fernández-Caballero, A.: Assisting dependent people at home through autonomous unmanned aerial vehicles. In: Novais, P., Lloret, J., Chamoso, P., Carneiro, D., Navarro, E., Omatu, S. (eds.) ISAmI 2019. AISC, vol. 1006, pp. 216–223. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-24097-4_26
5. Belmonte, L.M., Morales, R., García, A.S., Segura, E., Novais, P., Fernández-Caballero, A.: Trajectory planning of a quadrotor to monitor dependent people. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H. (eds.) IWINAC 2019. LNCS, vol. 11486, pp. 212–221. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19591-5_22
6. Belmonte, L.M., Morales, R., Fernández-Caballero, A.: Computer vision in autonomous unmanned aerial vehicles - a systematic mapping study. Appl. Sci. **9**(15) (2019). https://doi.org/10.3390/app9153196
7. Carretero, S., Garcés, J., Ródenas, F.: Evaluation of the home help service and its impact on the informal caregiver's burden of dependent elders. Int. J. Geriatr. Psychiatry **22**(8), 738–749 (2007). https://doi.org/10.1002/gps.1733
8. Carretero, S., Garcés, J., Ródenas, F., Sanjosé, V.: The informal caregiver's burden of dependent people: theory and empirical review. Arch. Gerontol. Geriatr. **49**(1), 74–79 (2009). https://doi.org/10.1016/j.archger.2008.05.004
9. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006). https://doi.org/10.1007/11744047_33
10. Halder, A., Tayade, A.: Real-time vernacular sign language recognition using MediaPipe and machine learning. Int. J. Res. Publ. Rev. **2**, 9–17 (2021)

11. Lugaresi, C., et al.: MediaPipe: a framework for building perception pipelines (2019)
12. Martínez, A., Belmonte, L.M., García, A.S., Fernández-Caballero, A., Morales, R.: Facial emotion recognition from an unmanned flying social robot for home care of dependent people. Electronics **10**(7) (2021). https://doi.org/10.3390/electronics10070868
13. Moon, S., Park, Y., Ko, D.W., Suh, I.H.: Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering. Int. J. Adv. Rob. Syst. **13**(2), 65 (2016). https://doi.org/10.5772/62415
14. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29
15. Penmetsa, S., Minhuj, F., Singh, A., Omkar, S.: Autonomous UAV for suspicious action detection using pictorial human pose estimation and classification. ELCVIA Electron. Lett. Comput. Vis. Image Anal. **13**(1), 18 (2014). https://doi.org/10.5565/rev/elcvia.582
16. Pietraszek, T.: Using adaptive alert classification to reduce false positives in intrusion detection. In: Jonsson, E., Valdes, A., Almgren, M. (eds.) RAID 2004. LNCS, vol. 3224, pp. 102–124. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30143-1_6
17. Puwein, J., Ballan, L., Ziegler, R., Pollefeys, M.: Joint camera pose estimation and 3d human pose estimation in a multi-camera setup. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9004, pp. 473–487. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16808-1_32
18. Rudol, P., Doherty, P.: Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery. In: 2008 IEEE Aerospace Conference, pp. 1–8. IEEE (2008)
19. Sanchez-Comas, A., Synnes, K., Hallberg, J.: Hardware for recognition of human activities: a review of smart home and AAL related technologies. Sensors **20**(15), 4227 (2020). https://doi.org/10.3390/s20154227
20. Schwarz, L.A., Mkhitaryan, A., Mateus, D., Navab, N.: Human skeleton tracking from depth data using geodesic distances and optical flow. Image Vis. Comput. **30**(3), 217–226 (2012). Best of Automatic Face and Gesture Recognition 2011. https://doi.org/10.1016/j.imavis.2011.12.001
21. Seint, P., Zin, T., Tin, P.: Intelligent monitoring for elder care using vision-based technology. Int. J. Innov. Comput. Inf. Control **17**(3), 905–918 (2021). https://doi.org/10.24507/ijicic.17.03.905
22. Watkins, L., Fairbanks, K.D., Li, C., Yang, M., Robinson, W.H., Rubin, A.: A black box approach to inferring, characterizing, and breaking native device tracking autonomy. In: 2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), pp. 0303–0308 (2020). https://doi.org/10.1109/UEMCON51285.2020.9298163
23. Xu, H., Bazavan, E.G., Zanfir, A., Freeman, W.T., Sukthankar, R., Sminchisescu, C.: GHUM & GHUML: generative 3d human shape and articulated pose models. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6183–6192 (2020). https://doi.org/10.1109/CVPR42600.2020.00622