Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

IoT based Pose detection of patients in Rehabilitation Centre by PoseNet Estimation Control

R. Asokan[1], T. Vijayakumar[2]

[1]Professor and Principal, Kongunadu College of Engineering and Technology, Tholurpatti, Tamil Nadu, India
[2]Professor, Department of CSE, Jai Shriram Engineering College, Avinashipalayam, Tamil Nadu, India

E-mail: [1]asokece@yahoo.com, [2]vishal_16278@yahoo.co.in

Abstract

Recently, Virtual rehabilitation has recently emerged as a contemporary option to treating chronic, handicapped, or mobility-impaired patients using virtual reality, augmented reality, and motion capture technology. Using a virtual environment, patients are able to work out in accordance with their treatment plan. This study provides a PoseNet-based in-home rehabilitation telemedicine system with integrated statistical computation allowing clinicians to assess a patient's recovery progress. Using a smartphone camera, patients may undertake rehabilitation activities at home. The angular motions of the patients' elbows and knees are detected and tracked using the PoseNet skeleton-tracking technology. The estimated elbow and other feature poses are recorded during the completion process of rehabilitation activities in front of the mobile camera. Finally, additional performance measurements are gathered and analysed in order to better understand how well the system works.

Keywords : Pose detection, rehabilitation, estimation, PoseNet, IoT, machine learning

1. Introduction

Medical treatment, health education, and health information are all examples of telehealth services that may be delivered across long distances using digital information and communication technology including laptops, desktops, and smartphones [1]. For patients and caregivers in remote locations, telehealth was developed as a means of overcoming the physical hurdles of providing healthcare services without in-person visits [2]. As a result, it is critical in providing patients with high-quality low-cost remote healthcare services. COVID-

DOI: https://doi.org/10.36548/jiip.2022.2.001
Received: 21.04.2022, received in revised form: 16.05.2022, accepted: 30 05.2022, published: 09.06.2022

19 has also pushed telemedicine to new heights and caused major changes in the way healthcare services are delivered [2]. During the COVID-19 epidemic, 90 percent of patients and parents preferred virtual appointments over in-person, according to research [3-5]. The telehealth system's capability and breadth are fast expanding, and both patients and professionals embrace it.

At the past, patients had to do their rehabilitation activities in a medical facility. Both governmental and non-governmental organisations in Malaysia provide rehabilitation services [6-9]. In remote locations, patients who lack access to reliable transportation or whose medical conditions prevent them from driving face, significant service hurdles as a result of the lack of coordination among various organisations. However, not many patients are able to pay the increased expense of hiring a home therapist to be treated at the patient's home. A telehealth system may be used to provide rehabilitative services in the comfort of one's own home. Prior to the advent of wearable sensors, a patient's movements are tracked by putting markings in many places on the body [10-12]. Figure 1 shows the block diagram of home rehabilitation process.

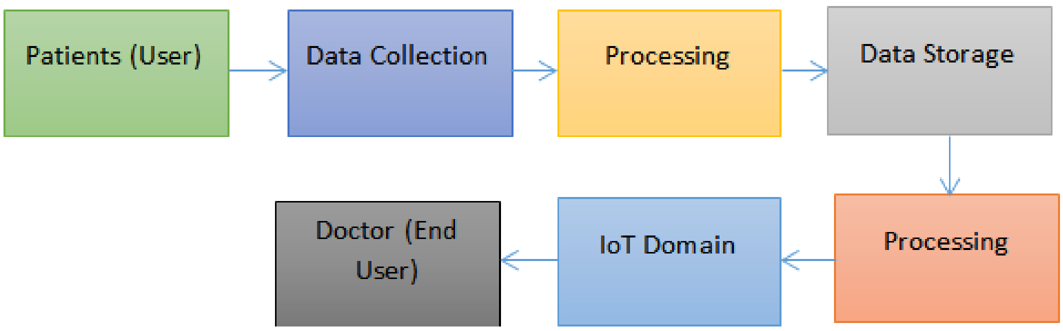

Figure 1. Home Rehabilitation Process

When using PoseNet architecture, the posture and angular motions of the body such as elbow and knee are recorded. The skeletal data may be obtained once the patient completes the rehabilitation activities for assessment, so that physicians can keep a close eye on how their patients' rehabilitation postures and angular motions are progressing. The physicians may then provide them with guidance or suggestions. Based on the doctor's instructions, the patient is able to enhance his posture by continuing to execute his proper movements for the specified period.. Thus, the PoseNet-based telemedicine system may be used to conduct an in-home rehabilitation approach, particularly for low-income patients in industrialised nations [13].

2. Literature Survey

Computer vision researchers have spent a lot of time on Human Posture Estimation (HPE), which is the process of inferring human body part configurations from sensor data, such as photographs and videos. It is important to use a GAN network-based pose identification system to tell good body part movements apart. Chen et al., [14] built and proposed PoseNet, a conditional adversarial network that is structure-aware. Using two stacked hourglass networks as discriminator and generator, Chou et al., [15] constructed a confrontational learning-based network. Predicted and real-world heatmaps were distinguished by the generator, which guessed each joint's position. Peng et al., [16] proposed data augmentation techniques to avoid overfitting problems during the prediction of HPE. This provided better accuracy and a good prediction rate for analysis.

A CNN approach for HPE developed by Yang et al., [17] incorporated spatial and appearance consistency across human body components to detect hard negatives. For understanding the relationships between human body joints in HPE, a structured-based learning approach was developed [18], which collects deeper information about human body joints and enhances learning outcomes. The multi-scale structure-aware neural network developed by Ke et al., [19] integrated multiscale combinations with many key points for movement prediction. This training strategy was used to enhance the HPE in complicated environments. Using deep learning compositional models, Tang et al., [20] were able to learn body shape and its orientation based on human bodies. It is learned that not all characteristics are connected to each other. Thus, Tang and Wu [21] came up with the idea of using the part-based branch network instead of a single model for all parts.

The following are the main differences between this survey and the previous ones:

1. Modern HPE 2D and 3D approaches are studied by categorising them depending on the kind of situation they are for: 2D or 3D, single view or multiple views from mobile phone views.

2. HPE approaches in 2D and 3D have been thoroughly tested for performance. This paper summarises and compares the results of potential algorithms on standard datasets according to their categories. Research trends in HPE are shown by comparing the outcomes, which offers insight into the strengths and limitations of various approaches.

3. Reviews a variety of HPE-based industries and applications, including gaming and surveillance, as well as AR/VR and healthcare.

4. HPE 2D and 3D are discussed in terms of important difficulties, pointing to possible future research in HPE that might improve performance.

2.1 Research Gap

1. Camera pose regression datasets are generated using an automated approach of categorising data based on motion structure.

2. Transfer learning (PoseNet) uses picture recognition datasets to teach a posture regressor, which was previously learned as a classifier. It takes less time to get to a lower error rate than training from start, even with a limited training set.

3. Methodology

Managing the accounts of patients are under the new system suggested. It is possible for the patient to schedule virtual visits with the doctor or to engage in physical therapy exercises, at their convenient time. Initially, a patient can plan and execute many motion activities at their suitable time. As soon as the doctor confirms the appointment time, the patient will get an email notification. Figure 2 shows blocks of proposed algorithm.

3.1 Execution of an algorithm

The doctor can provide their maximum output comments during their video chatting with patients in the scheduled time. Patients can begin rehabilitation exercises after consulting with their doctor.

Step 1: The patient's angular motions are calculated using the suggested system.

Step 2: A note of them is prepared and put in the database.

Step 3: An angular movement can be obtained after finishing the rehabilitation activities.

Step 4: The angular motions are analysed using the graph.

Step 5: The physicians now have access to the patient’s records from the database.

Step 6: The skeletal outcomes are compared and contrasted at various times.

Step 7: The patient’s medical history and medications are given.

Step 8: It is concluded that the patient might review the prescriptions and adapt proper exercises.

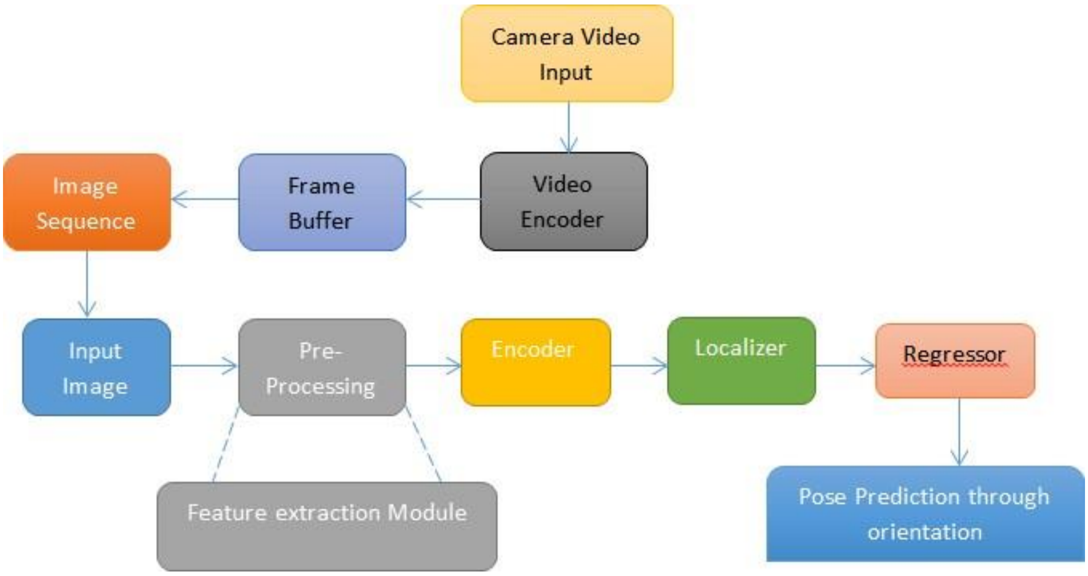

Figure 2. The proposed framework

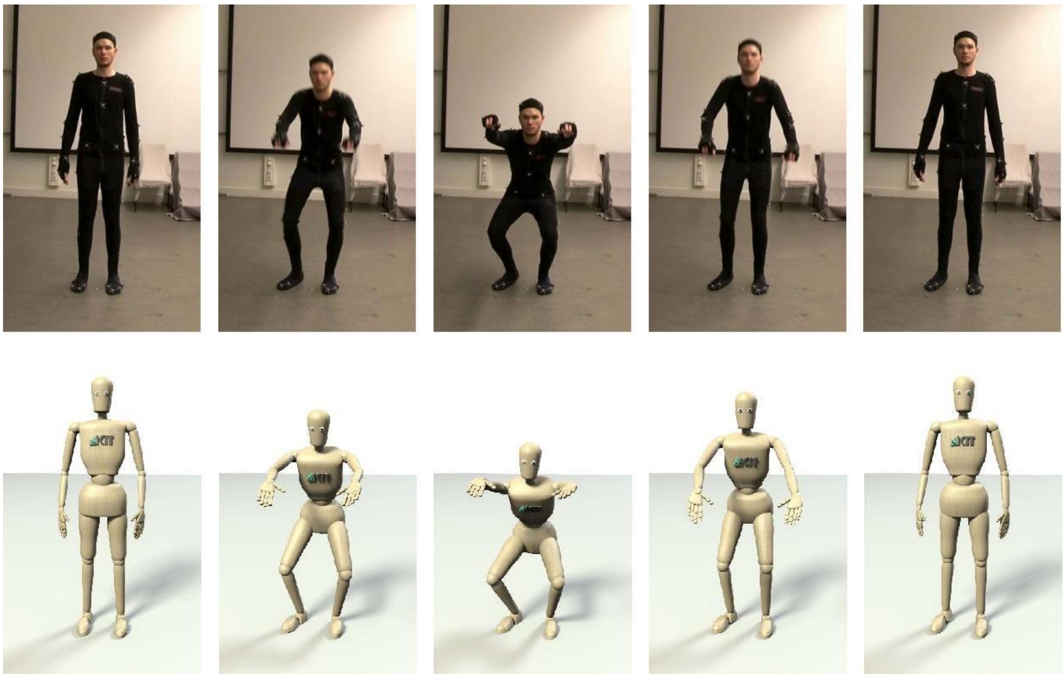

Figure 3. KIT dataset [13]

3.2 Implementation of Body Key points

The 17 key points on the human body are detected using PoseNet [22, 23]. The 2D coordinates of each significant point are used to represent it (x, y). The skeletal detection with 17 key locations is presented on the body during a rehabilitation session. When it comes to finding the patient's shoulder, elbow, wrist, hip, and ankle movements using the suggested approach, there are just 12 crucial points. Figure 3 shows some pose positions in KIT dataset.

3.3 Key Point Estimation

The hand and leg’s right and left portion are the four bodily parts that have angles. The shoulders, elbows, wrists, and hips are referred to as A, B, C, and D, respectively. Thus, each hand and each leg has a total of twelve critical locations. A set of three 2D coordinates has been assigned to each of the four important locations in the body, with point A, point B, and point C [24]. Then the distances between each of the 12 critical points are determined using the Euclidean distance for each of the 12 video frames.

3.4 PoseNet estimate control

The PoseNet-based solution for at-home rehabilitation has passed the functional tests with the following validations:

1. In the first place, the patients' angular movements can be detected using PoseNet and an ordinary mobile camera (it's for IoT); in the second place, after the rehabilitation exercise, patients and doctors can immediately look at a computerised comparison of how the patients' joints moved on various days.
2. Doctors may analyse patients' recovery rates more thoroughly with the use of a visualisation graph and statistical calculation that are created automatically. There are two drawbacks in the suggested approach, even if it has been demonstrated to function based on the doctor's review.

3.5 Detection analysis

When using a camera for skeleton detection, the patient must make sure their visibility in front of the mobile camera. To begin with, the patient must make sure that their laptop is set up at the correct standing distance, which may be found in the system setup instructions. In order to do an accurate analysis, the camera should be able to observe the patient's whole body [25].

4. Results and Discussion

The standard poses are selected from various datasets for the series of benchmark events. But this project carries the KIT dataset for motion detection. This dataset contains whole-body movement, and it is a large dataset with various activities. Also, it focuses on whole-body motion estimation by multiple activities.

The obtained results are shown using the best feature engineering and the most robust response. This proposed work contains PoseNet architecture and provides good sensitive results from the texture features of the outdoor poses. This patch effect can be more informative for better prediction or detection accuracy. Also, the patch points provide the localization details of the patients' postures. Therefore, these texture and interest-based features provide better map value for higher accuracy than other models, and the results obtained are shown in Table 1.

The Mean Absolute Deviation (MAD) measures the current and previous image pixel comparison for each point noted in the human body. The Mean Square Error (MSE) is calculated as the difference between two poses in a sequence frame. Finally, the Mean Absolute Percentage Error (MAPE) is computed to identify accurate poses of the patients. It is a more sensitive measurement that measures little deviations in their poses. Because it can compare both the present and previous position of the skeleton results, these measurements are used to aid in measuring the forecasting performance. But, generally, MSE results are shown rather than MAD value.

Table 1. Results obtained by the proposed method (Overall Values)

Model	Deviation (Prediction error) rate			Accuracy	Sensitivity	IoT Response Time
	MAD	MAPE	MSE			
Transfer Learning	24.27	51.71%	1711.2	81.34%	78.34%	7.67 sec
Pre trained Neural networks	26.12	58.87%	1834.1	86.43%	80.34%	2.563 sec
Proposed Framework	21.81	34.99%	1403.6	95.12%	94.62%	1.567 sec

The results provide good sensitivity for dealing with the localized information (features) by the proposed PoseNet architecture. Figure 4 shows the overall performance measure graph.

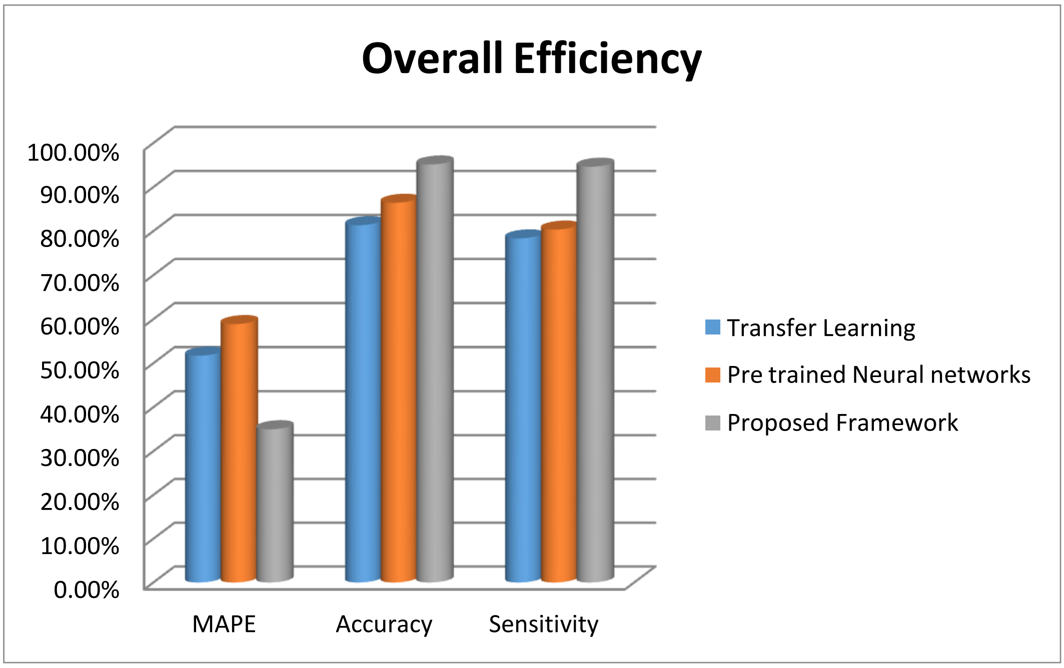

Figure 4. Overall performance measures

5. Conclusion

In this paper, the proposed PoseNet algorithm has proved its superiority in several performance metrics. The proposed learning framework from networks trained as classifiers may be used to avoid the requirement for millions of training photos, as proven. Although trained to generate posture-invariant outputs, it has been proven that these networks retain adequate pose information in their feature vectors. The total model's performance has been compared to that of many other methods. The suggested work has shown its superiority above others, as discussed before. Training datasets containing ground truth annotations such as calculating human postures from photographs of infants or artwork collections, lack for certain applications. In addition, the data distributions for these applications are distinct from those of conventional posing data. The HPE algorithms that have been trained on conventional datasets may not generalise effectively to new fields of application. To bridge the knowledge gap, GAN-based learning techniques have recently gained popularity as a solution. However, how to transfer human pose knowledge across domains is still an open question.

References

[1] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. "Learning and transferring mid-level image representations using convolutional neural networks" In Computer Vision and

Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1717–1724. IEEE, 2014.

[2] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. “CNN features off-the-shelf: an astounding baseline for recognition” In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, pages 512–519. IEEE, 2014.

[3] N. S¨underhauf, F. Dayoub, S. Shirazi, B. Upcroft, and M. Milford. “On the performance of convnet features for place recognition” arXiv preprint arXiv:1501.04158, 2015.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. “Going deeper with convolutions” arXiv preprint arXiv:1409.4842, 2014.

[5] Uddin, S.M.; Rahman, A.; Ansari, E.U. COMPARISON OF SOME STATISTICAL FORECASTING TECHNIQUES WITH GMDH PREDICTOR: A CASE STUDY. J. Mech. Eng. **2018**, 47, 16–21.

[6] Dengen, N.; Haviluddin; Andriyani, L.; Wati, M.; Budiman, E.; Alameka, F. Medicine Stock Forecasting Using Least Square Method. In Proceedings of the 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), Makassar, Indonesia, 6–7 November 2018. [CrossRef]

[7] Mandery, C.; Terlemez, O.; Do, M.; Vahrenkamp, N.; Asfour, T. The KIT Whole-Body Human Motion Database. In Proceedings of the 2015 International Conference on Advanced Robotics (ICAR), Istanbul, Turkey, 27–31 July 2015; pp. 329–336.

[8] The Chien Hoang; Ha Trang Dang; Viet Dung Nguyen. Kinect-Based Virtual Training System for Rehabilitation. In Proceedings of the 2017 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam, 21–23 July 2017; pp. 53–56.

[9] Shapi’i, A.; Bahari, N.N.; Arshad, H.; Zin, N.A.M.; Mahayuddin, Z.R. Rehabilitation Exercise Game Model for Post-Stroke Using Microsoft Kinect Camera. In Proceedings of the 2015 2nd International Conference on Biomedical Engineering (ICoBE), Penang, Malaysia, 30–31 March 2015.

[10] Shapi’i, A.; Arshad, H.; Baharuddin, M.S.; Mohd Sarim, H. Serious Games for Post-Stroke Rehabilitation Using Microsoft Kinect. Int. J. Adv. Sci. Eng. Inf. Technol. **2018**, 8, 1654–1661.

[11] Wosik, J.; Fudim, M.; Cameron, B.; Gellad, Z.F.; Cho, A.; Phinney, D.; Curtis, S.; Roman, M.; Poon, E.G.; Ferranti, J.; et al. Telehealth Transformation: COVID-19 and the Rise of Virtual Care. J. Am. Med Inform. Assoc. **2020**, 27, 957–962.

[12] Bate, N.J.; Xu, S.C.; Pacilli, M.; Roberts, L.J.; Kimber, C.; Nataraja, R.M. Effect of the COVID-19 Induced Phase of Massive Telehealth Uptake on End-User Satisfaction. Intern. Med. J. **2021**, 51, 206–214.

[13] Plappert, Matthias, Christian Mandery and Tamim Asfour. “The KIT Motion-Language Dataset.” *Big data* 4 4 (2016): 236-252.

[14] Chen, Yu, Chunhua Shen, Xiu-Shen Wei, Lingqiao Liu, and Jian Yang. "Adversarial posenet: A structure-aware convolutional network for human pose estimation." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1212-1221. 2017.

[15] Chou, Chia-Jung, Jui-Ting Chien, and Hwann-Tzong Chen. "Self adversarial training for human pose estimation." In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 17-30. IEEE, 2018.

[16] Peng, Xi, Zhiqiang Tang, Fei Yang, Rogerio S. Feris, and Dimitris Metaxas. "Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2226-2234. 2018.

[17] Yang, Wei, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3073-3082. 2016.

[18] Chu, Xiao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. "Structured feature learning for pose estimation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4715-4723. 2016.

[19] Ke, Lipeng, Ming-Ching Chang, Honggang Qi, and Siwei Lyu. "Multi-scale structure-aware network for human pose estimation." In *Proceedings of the european conference on computer vision (ECCV)*, pp. 713-728. 2018.

[20] Tang, Wei, Pei Yu, and Ying Wu. "Deeply learned compositional models for human pose estimation." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 190-206. 2018.

[21] Tang, Wei, and Ying Wu. "Does learning specific features for related parts help human pose estimation?." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1107-1116. 2019.

[22] Tay, E.L.; Wong, C.P. A Cross-Sectional Survey of Rehabilitation Service Provision for Children with Brain Injury in Selangor, Malaysia. Disabil. CBR Incl. Dev. **2018**, 29, 45–58.

[23] Nitta, Y.; Murayama, Y. Privacy-Aware Remote Monitoring System by Skeleton Recognition. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, 8–11 January 2019.

[24] Zhen, Jianan, Qi Fang, Jiaming Sun, Wentao Liu, Wei Jiang, Hujun Bao, and Xiaowei Zhou. "Smap: Single-shot multi-person absolute 3d pose estimation." In *European Conference on Computer Vision*, pp. 550-566. Springer, Cham, 2020.

[25] Ce Zheng, Matias Mendieta, Pu Wang, Aidong Lu, and Chen Chen. 2021. A Lightweight Graph Transformer Network for Human Mesh Reconstruction from 2D Human Pose. arXiv preprint arXiv:2111.12696 (2021).

Author's biography

R. Asokan received his B.E degree in electronics and communication from Bharathiar University and MS degree in electronics and control from Birla Institute of Technology. He obtained an M.Tech degree in electronics and communication from Pondicherry Engineering College, with distinction. He obtained PhD in information and communication engineering from Anna University, Chennai. He is currently the Principal, Kongunadu College of Engineering and Technology, Thottiyam, TamilNadu, India. He has published more than 65 papers in national and international journals and conferences. He has over 25 years of teaching experience. He is a member of various scientific and professional societies. His areas of interest include wireless networks, network security and image processing.

T. Vijayakumar is currently working as Professor in the Department of ECE at Jai Shriram Engineering College, Avinashipalayam, Tamil Nadu, India. His research includes Computer Vision, Motion Analysis, Stereo Vision, Object Recognition, computer graphics, photo interpretation, image retrieval, Embedded Image Processing and Real-time image and video processing applications.