# Automatic Gesture Recognition for Human-Machine Interaction: An Overview

**Konkina Nataliia**

Department of Automation of Power Processes and Systems Engineering
(APEPS), Faculty of heat power engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

**Summary**
With the increasing reliance of computing systems in our everyday life, there is always a constant need to improve the ways users can interact with such systems in a more natural, effective, and convenient way. In the initial computing revolution, the interaction between the humans and machines have been limited. The machines were not necessarily meant to be intelligent. This begged for the need to develop systems that could automatically identify and interpret our actions. Automatic gesture recognition is one of the popular methods users can control systems with their gestures. This includes various kinds of tracking including the whole body, hands, head, face, etc. We also touch upon a different line of work including Brain-Computer Interface (BCI), Electromyography (EMG) as potential additions to the gesture recognition regime. In this work, we present an overview of several applications of automated gesture recognition systems and a brief look at the popular methods employed.

***Key words:*** *Human-Machine Interaction, Gesture Recognition Systems, Hand Tracking, Activity Recognition, Brain Computer Interface (BCI)*

## 1. Introduction

We live in the age of information and data. In the last century or so, there has been rapid technological advancements especially in the field of computing systems. The early systems used connecting wires, punch cards, paper tapes, etc as input mediums and display lights, teletypes, paper, etc as output mediums. Although these may seem inconvenient to us in the present day, they were indeed the state-of-the-art technologies back in the days. Over the years, more convenient mediums of interaction were developed like the keyboard, mouse, monitors, etc. These are the most popular form of devices we use to interact with the machines and computers. However, there is always scope for improving the ways in which we communicate and interact with devices. Human Machine Interaction (HMI) / Human-Computer Interaction (HCI) systems are pivotal in shaping the interaction between users and machines. They form the mediatory link that tries to understand what the user needs and what the system does [3]. Interaction can happen through several different input and output mediums, notably (i) visual medium (ii) auditory medium (iii) movement and (iv) haptic medium [3]. The holy grail of human-machine interaction would be achieved when machines can understand us emotionally, anticipate our needs, and our intentions and support us in carrying out our tasks in the most intuitive manner possible. This requires our computing systems to be “intelligent”. Most established systems out there are not really “intelligent”. They are essentially systems that carry out a fixed set of predetermined actions. These systems expect and rely on users to provide instructions. For example, when a user is preparing a table with metrics from an image of table, the user looks at the image, identifies the required table entry and types it out on a spreadsheet. If the system was intelligent, it should have ideally been able to identify the user intentions and then carry out the task after a few such iterations. Increasingly with the advent of Artificial Intelligence (Machine Learning, Deep Learning, Speech Recognition, Computer Vision, etc) technologies, they are beginning to embed “intelligence” into these systems. Automatic gesture recognition is at the forefront of this “intelligence” revolution. Automatic gesture recognition can be described as the non-verbal interaction from a user to the system.

## 2. Input Sources for Gesture Recognition

In this section we start with a gentle introduction to gestures and dive into the various types of input devices and sources used in gesture recognition systems. In the next section we look at the established applications, and a brief overview of the popular methods in the literature.

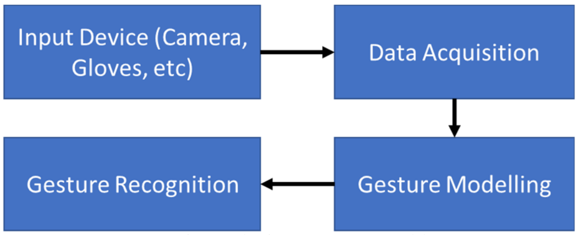


**Fig. 1** Stages of a Typical Gesture Recognition System.

Manuscript received December 5, 2021
Manuscript revised December 20, 2021
https://doi.org/10.22937/IJCSNS.2022.22.1.19

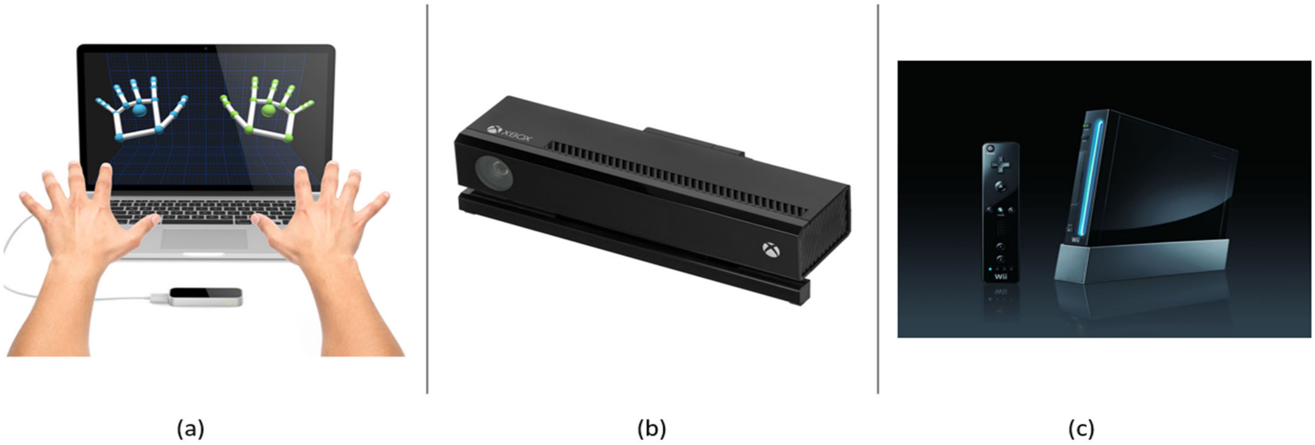


**Fig. 1** Few famous consumer electronic devices(a) Leap Motion [41] (b) Microsoft Kinect [40] (c) Nintendo Wii [42].

Traditional gestures can be broadly classified into (i) static, in which the user orients in a fixed pose or configuration called as posture in [5], (ii) dynamic, in which a sequence of actions or postures together constitutes a gesture and (iii) both static and dynamic nature as seen in sign languages [4].This definition of gestures can be limited. We therefore extend it by including the electrical signals from the brain (BCI) and the skeletal muscles (EMG) as these signals do contain an intent which can be interpreted. We can break down a traditional gesture recognition system into the following parts [5].

**Input Source**. This is governed by the available input devices like the type of cameras, sensors, data gloves, etc. We'll take a closer look at these in the later part of this section.

**Data Acquisition**. It defines the protocol in which the data collection happens. This is application dependent. This may also vary based on if the data is collected for offline training or real time gesture recognition. For example, in automated workout monitoring tools, a person may lose his form after going into the data collection process without fully rested. Hence it is crucial to establish the setup and protocol before acquiring data.

**Gesture Modelling**. Then we carefully lay out what gestures need to identified, the exact specification for each of the gestures, and the metrics for evaluating the performance of the system.

**Gesture Recognition**. Once all the above are defined, we need to identify the appropriate techniques and algorithms for gesture recognition. In the pre deep learning era, this includes feature extraction followed by a classification model. With deep learning, we can extract features jointly with the classification model. This has shown to do well over the traditional hand-crafted mechanisms. Because of the ubiquitous nature of gestures, the modalities of data one gets to work with in gesture recognition is very diverse. Hence making this field rich in the problem space and present a lot of opportunities to further advance the technology. The most critical aspect of a gesture recognition system is the data acquisition step. This is determined by the type of sensors used, the availability of different kinds of modalities of input data. High quality data acquisition governs the result of the system. Therefore, it is crucial to identify the exact requirements and then diligently choose the sensors and devices. We take a brief look at the common devices used for gesture recognition.

**RGB Camera**. These are the most widely used devices for gesture recognition these days. These include 2D cameras for capturing images and videos. Various applications like hand gesture recognition, sign language, etc can be done. One of the drawbacks of this mode of cameras are the lack of depth perception which may be crucial for certain applications.

**Depth Aware Cameras**. These type of cameras like time-of-flight cameras or structured light have the ability to approximate an additional dimension of depth which is not easily obtained from 2D cameras [1, 33].

**Stereo Cameras.** The key idea is to work with two different camerasin tandem with predetermined relationship between them. Using appropriate tools, one can obtain and estimate additional information like depth, 3D contours etc. These approaches are being increasingly utilized in Machine Vision [30, 1] applications.

**Infrared Imaging Devices.** These cameras can operate in the absence of light. They provide a different view as compared to the regular 2D cameras. These are very useful for surveillance cameras as they can detect activity at night. There is an increasing trend to incorporate gesture recognition on infrared streams [29].

**WiFi sensing / Radio Signal Based Sensors**. It is also known WLAN sensing and uses WiFi signals for detection and interpretation of event and activities. These signals have an added benefit that they can traverse through walls but reflect of humans. This can be used for activity recognition under occlusions [2, 32].

**Touch Screens.** In the last several decades touch based systems have been of widespread use in the consumer electronics like smart phones, touch screen laptops, monitors, etc. There has been a lot of interest in Braille based typing for the specially-abled people [35]. There have been algorithmic improvements like Swype Typing [34] where the input was a path traversed by the user on the keypad. It builds on error correction algorithms and leverages a language model to identify the intended word. This enables users to type faster on edge devices.

**Data Gloves and Tracking Suits**. Besides visual and depth inputs available from a variety of cameras and sensors, a rather different approach to gesture recognition arises from data gloves and tracking suits. There are specially crafted gloves / suits with embedded sensors that are capable of transmitting information like the position, orientation, etc, of various joints using sensors like accelerometer, gyroscope, inertial measurement units (IMU) etc. Using the information from these sensory inputs, we could discern the gestures more accurately. These are direct form of inputs that have a lot of potential to improve the recognition systems [36].

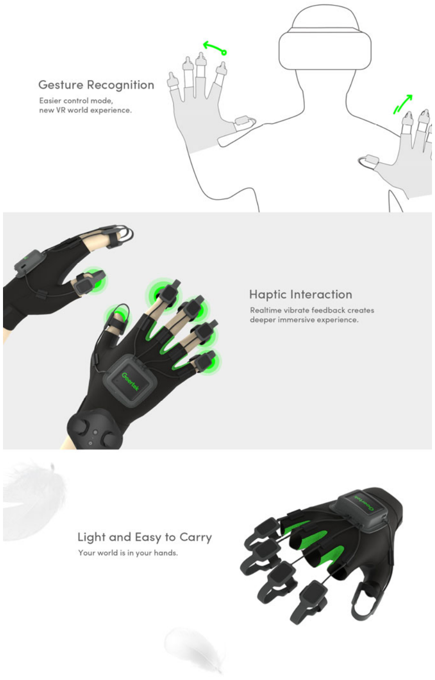

Gesture Recognition
Easier control mode, new VR world experience.
Haptic Interaction
Realtime vibrate feedback creates deeper immersive experience.
Light and Easy to Carry
Your world is in your hands.

**Fig. 3** Data Glove for Gesture Recognition [43].

**Gaming Controllers.** Traditionally, these controllers were meant as a way to communicate with the gaming system in a convenient way. However, there is an increasing body of research on how to make these controllers more natural and consequently enhance the realism of the games. Some of the popular controllers are Nintendo Wii, Sony PlayStation Eye, Microsoft Kinect, Leap Motion, etc.

**Brain Computer Interface (BCI).** BCI [37] has been one of the latest developments and a hot research area in the field of Human-Computer Interaction. The working principle of BCI is that there are distinguishable patterns in the neural responses based on one's thoughts and actions. By picking up these electrical signals from various regions of the brain, we can detect and interpret them to take meaningful actions as required by the application. The most common source of input is the Electroencephalogram (EEG) which are the electrical signals picked up on the scalp. They can be worn as electrode caps, passive electrodes, etc.

**Electromyographic (EMG) Inputs.** EMG is a technique by which one can analyze the electrical activity and response of the skeletal muscles. When a brain sends signals to the end effectors, a series of chemical reactions transfers information in the form of neural activity. By tapping the electricity arising from these signals, we can interpret them as commands from the brain to carry out a particular function. This form of input is garnering interest in building prosthetics that can be controlled by brain signals [38]. Myo wristbands are an established brand of controllers that rely on EMG signals.

## 3. Gesture Recognition Systems

In this section we take a deeper look at the current trends of gesture recognition systems. We highlight certain established applications of these systems and then provide an overview of the popular approaches to solve these problems.

### 3.1 Hand Gesture Recognition

Hand Gesture Recognition is probably the most well studied application of gesture recognition. This is because they can be very expressive and solved using commonly available cameras. The scope of hand gesture recognition includes various application avenues like sign language detection for the differently a bled users to communicate with systems, Gesture based control systems for remote applications like remote surgery, human-robot-interaction with hand-controlled gestures etc. There are different kinds of gestures [4]. They can be listed as follows:

(i) **Gesticulation**. These are the spontaneous gestures that are accompanied by normal speech. The gestures are not necessarily required, but they can be looked at as add-ons that help improve the expressivity.

(ii) **Language-like gestures**. These are gestures that are similar to gesticulation but typically voluntary that can replace certain utterance.

(iii) **Pantomimes**. These are gestures of certain actions or objects that may or may not have accompanying speech.

(iv) **Emblems**. These are usually concretely.

`

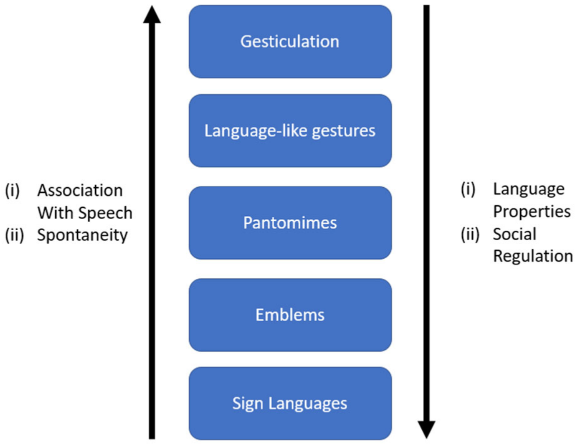


**Fig. 2** Various types of gestures with increasing and decreasing properties [4].

(i) **Sign Languages**. There are well defined rules and systems in place that enable people to communicate effectively without any accompanying speech.

The commonly used algorithms and models for gesture recognition include traditional machine learning models like.

**Hidden Markov Models (HMM)** have been successfully leveraged for applications [55, 56, 57, 58, 59] like activity recognition, gesture recognition, speech recognition, etc. HMM is a probabilistic graphical model defined over a sequence. It is "hidden" as it includes are states that are not observed.

**Condensation Algorithm** builds on the work of particle filtering. This has been shown to be an effective method for tracking motion of objects and can be applied for gesture tracking [59, 60].

**Finite State Machines (FSM)** are a computational abstraction that can be used to model the sequential logic of a system. They do find applications for gesture recognition [61, 62, 63] as a gesture can be viewed as a sequence of postures. Recently, with the popularity of deep learning models, the following architectures have also been leveraged successfully for gesture recognition.

**Convolutional Neural Networks (CNNs)** have been the powerhouse of computer vision tasks. In traditional image processing, one had to manually craft filters, however with CNNs, the filters are learnt jointly with the task which makes them very powerful. They have been successfully utilized in [64, 65, 68] for gesture recognition.

**Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** are the go-to approaches for modelling sequences in deep learning regime. They are powerful models with memory and have been shown to be useful for gesture recognition [66, 67, 68] along with CNNs.

**Hybrid Models** like a combination of CNN and RNN / LSTMs are powerful models that can work well with sequential image data as seen withmodels like [67, 68].

## 3. 2 Lipreading Systems

Lip reading is one of the essential cogs when it comes to visual speech recognition. Visual Speech recognition is at the intersection of Speech Recognition, Natural Language Processing and Computer Vision. It is the ability of the system to decipher the speech by attending to the motion of lips along with speech inputs.

However, lipreading can also be done in the absence of any other audio cues. Many a times, it may be difficult to decipher one's speech because of loss of information while collecting the audio signals, noise in the collection process, mumbling or stammering at the end of the speaker, etc. Combining these with lip reading can enhance the quality of speech recognition. [69] is an extensive survey on deep learning-based methods for lip reading applications.

It covers various methods starting from different types of CNN architectures to Attention-Transformers and Temporal CNNs. It dives deep into the details of the classification schema presented in Figure 6.

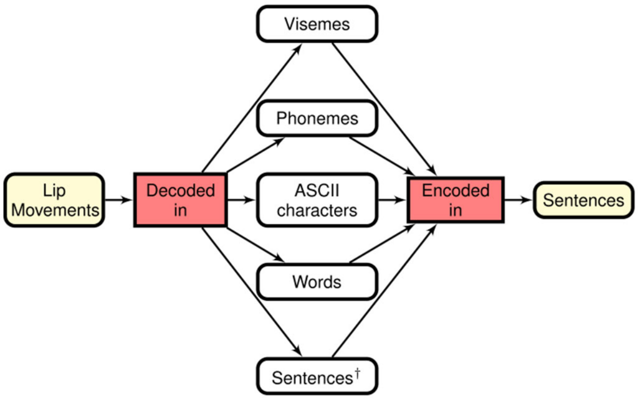


**Fig. 3** Classification Schema as stated in [69].

### 3.3 Eye Tracking Systems

Recently there has been an interest in eye tracking systems for various applications like drowsy driver detection [6, 7] where they propose to take preventive measures to avoid road accidents and promote driver well-being. [7] proposed InSight which is a cost-effective approach for driver fatigue and driver distraction detection that can work even in low light scenario.

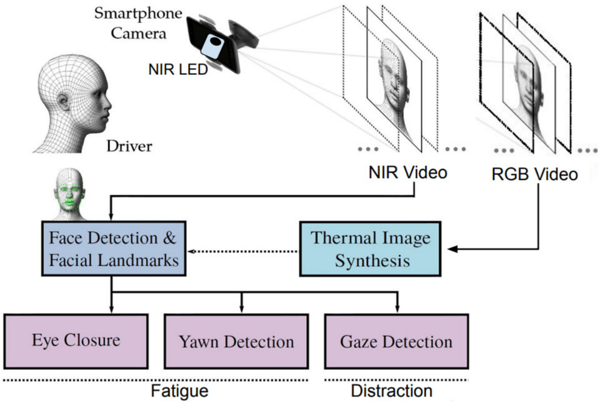
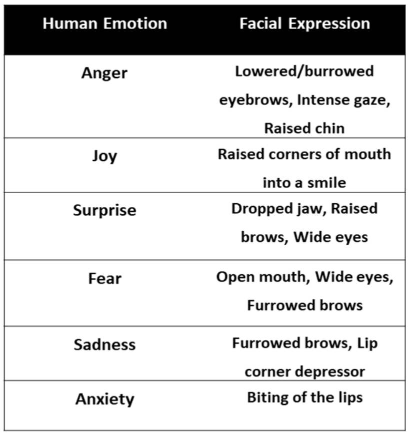


**Fig. 7** InSight Approach for Driver Fatigue and Distraction Detection [7].

Their system is summarized in Figure 6. They artificially synthesize thermal images using GANs (Generative Adversarial Network) and additionally uses NIR (Near-IR) LED to illuminate under low light conditions. There has been interest in tracking the user gaze for intent identification [8].This area of research could potentially be targetted by advertisers to effectiviely gauge the user focus [70].



| Human Emotion | Facial Expression |
|---|---|
| Anger | Lowered/burrowed eyebrows, Intense gaze, Raised chin |
| Joy | Raised corners of mouth into a smile |
| Surprise | Dropped jaw, Raised brows, Wide eyes |
| Fear | Open mouth, Wide eyes, Furrowed brows |
| Sadness | Furrowed brows, Lip corner depressor |
| Anxiety | Biting of the lips |

**Figure 8**: Manifestation of Human emotions into facial expressions [28, 26, 27]

### 3.4 Emotion Recognition Systems

Often when predicting and forecasting what humans need and require, machine intelligence may provide a rational solution, however, given that the humans are emotional beings, the solution required may not necessarily be rational. Almost every aspect of human life is governed by emotions, even when it may seem irrational to do so. Therefore, it is essential to identify the emotions and cater the solution accordingly. This is where, automated emotion recognition has a lot of scope. Emotion recognition can be done in various forms, like video, image, speech, text, dialogue, etc. Few of the applications (not limited to) are Security / Preventive Measures can be installed in institutions like school, etc. to prevent potential violent outbreaks.

Employee / Customer / User Satisfaction can be gauged by the emotion recognition systems. They help identify the response from the interacting with the systems, which in turn provides accurate feedback to improve. Additional support to differently abled people to interpret the feelings and emotions of the people they interact with. Entertainment industry – There are several avenues like video game, multimedia movies, Virtual Reality (VR) / Augmented Reality (AR) based simulations / movies to identify how the users perceive and respond to certain kinds of stimuli.

There are various kinds of emotions like Fear, Joy, Anger, Surprise, etc. Usually most of these emotions manifest as certain facial expressions. By detecting these differences in the facial expressions, we can reasonably detect human emotions. Refer Figure8 for additional details on the facial expressions and the associated human emotions.

### 3.5 Human Pose Estimation Systems

The goal of Human Pose Estimation is to identify and localize various parts of the body and construct a representation of the human body from various modalities of data. Images and videos were the primary sources of inputs. However, there has been an increasing trend in utilizing other sources of inputs like the data from Infrared Sensors, Radio Signals, Depth Sensors, etc.

One of the key aspects of human pose estimation is to

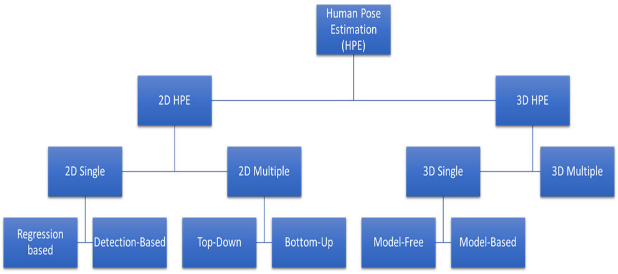


**Fig. 9** A taxonomy of HPE as proposed in [9].

identify and model the appropriate form of human body representation. There are three different kinds of human body representation [10].

**Kinematic model.** In this model, the primary joints like the limbs, hip, head, etc, are detected and localized along with their orientations. They are also referred to as skeleton-based model [9] or the kinematic chain model [11].

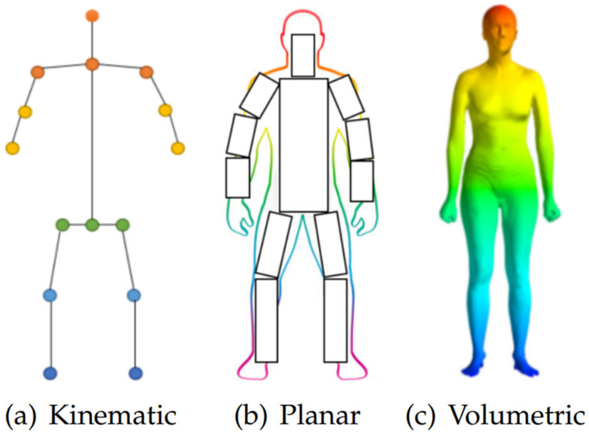


**Fig. 10** Three Models of Human Body Representation [10].

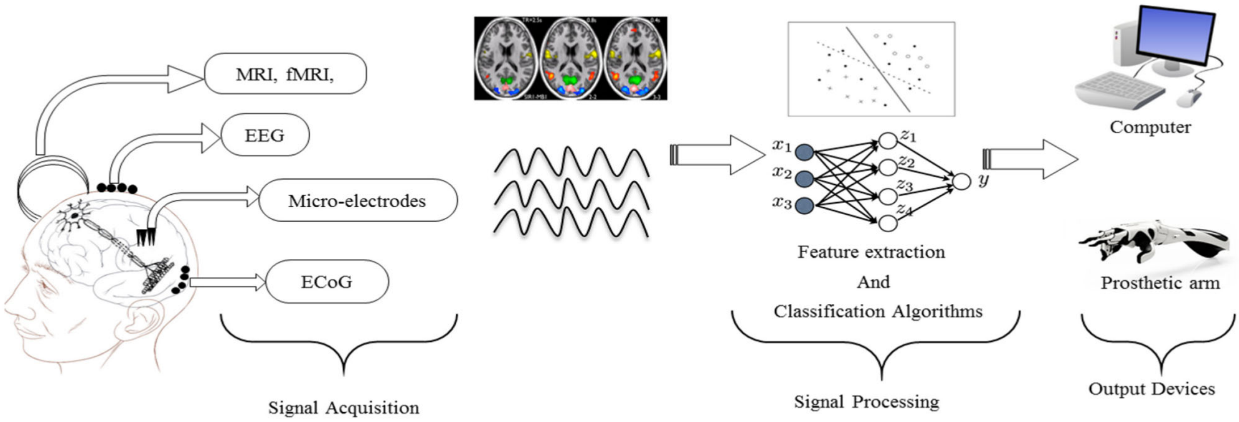


**Fig. 11** A Brain Computer Interface System [37].

**Planar Model.** Its an extension to the kinematic model in which there are bounding boxes along with the localization. These boxes are meant to serve as an approximation to the human body contours. They are also known as contour-based models.

**Volumetric Model**. It is the closest form of human body reconstruction. This representation captures the nuanced contours and can be easily extended to 3D. This form of representation can also capture depth information unlike the earlier two representations. Some of the popular models are Skinned Multi-Person Linear (SMPL) [12], Dynamic Human Shape in Motion (DYNA) [13], Stitched Puppet Model [14], Frankenstein & Adam: The Frankenstein model [15], GHUM and its light variant GHUML(ite) [16].

Chen et. al. [9] has summarized various approaches under a taxonomy as described in Fig. 2. They broadly classify the models under 2D HPE and 3D HPE. Under each of these categories, they make a distinction between estimation of single and multiple persons jointly. We summarize the popular 2D HPE [9] methods below.

**Regression based methods** primarily include (i) Direct Prediction [17] (ii) Supervision Improvement [22, 23,24,25] (iii) Multi-Task [18, 19,20,21].

**Detection based methods** include Patch Based [71], Temporal Constraint based [72], Network Compression based [73], Network design based [74], Body Structure Constraint [75]methods-based approaches.

**Top-Down methods** comprises of coarse-to-fine [76], bounding-box refinements [77] etc.

**Bottom-Up methods** consists of approaches like DeepCut [78], Deeper Cut [79], etc.

Recently, there has been an increasing interest in pose estimation using inputs beyond the traditional images and videos. [31] Rao et. al. proposes to use Microsoft Kinect Sensor to track 3D skeletal information for physiotherapy sessions. They leverage DWT (Discrete Wavelet Transform) coefficients in conjunction with LSTM (Long-Short Term Memory) units to identify the exercise. They also introduce a fidelity score to monitor the quality of the exercises performed with respect to a set of "ideal" instances. [2] Zhao et. al. presents a unique approach for human pose estimation. Their approach can detect human poses accurately throughs walls and occlusions using Radio Signals. They leverage the fact that WiFi signals traverse through walls and can reflect off humans.

## 3.6 Brain-Computer Interface (BCI) Systems

BCI is a way in which one can control systems using signals from the brain. This is akin to controlling the system with your thoughts. This may not be too far away in the future. Electroencephalogram (EEG) is a common way to quantify electrical activity of the brain. EEG is a non-invasive technique in which electrodes are placed on the scalp and does not require any form of surgery. Electrocorticography (ECoG) is an invasive form of EEG, in which electrodes are implanted in the exposed surface of the brain. This acquisition technique is also referred to as intracranial electroencephalography (iEEG). The most invasive form of acquisition is the deep brain recording in which a micro-electrode is placed deep inside the brain. The quality of recordings is commensurate with degree of invasiveness. Beyond BCI, EEG based predictive models are being used for healthcare. Sleep stage classification using automated algorithms help reducing the time required to manually evaluate the sleep stages by experts. The brain waves contain latent information about sleep patterns that often indicate underlying sleep disorders. This enables faster diagnosis of sleep disorders. EEG set ups are usually much cheaper than other forms of neuronal signal acquisition methods and does not require any form of surgery, they are a convenient way to analyze brain patterns. An electrode recording activity forms an EEG channel. Typical EEG systems comprises from a single channel to 256 channels. There are different kind of brain signals [37] (i) Event Related Potential (ii) Evoked Brain Potential (iii) Sensorimotor Rhythms (SMRs) (iv) Slow Corticol Potential (SCP). Quite often the signals are affected by other artifacts such as the blinking of the eye, arbitrary muscle moments, etc. These manifest as noise during the signal acquistion process. Hence feature extraction techniques should be able to minimize the effect of noise. The common feature extraction techniques for EEG based signals are :

- **Common Spatial Pattern (CSP)** [44] – It is filter that operates in spatial domain and tries to maximise the variance in the filtered brain signal.

- **Principal Component Analysis (PCA)** – It is one of the most popular dimension reduction algorithms. It has been shown to be useful for BCI based applications like in [45, 46].

-**Independent Component Analysis (ICA)** – ICA is one of the computational approaches to separate multivariate signal / mixed signal into independent sources or subcomponents. It is used for blind sources seperation. There are a lot of works that build on ICA for BCI applications [48, 49, 50, 51].

- **Wavelet Transform**. It is similar to Fourier transform (or much more to the windowed Fourier transform) with a completely different merit function. The main difference is that Fourier transform decomposes the signal into sines and cosines, i.e.

The functions localized in Fourier space in contrary the wavelet transform uses functions that are localized in both the real and Fourier space. Some of the works that leverage wavelet tranforms are [52, 53, 54].

## 4. Conclusion

In this work, we covered a very broad set of the areas at the intersection of human-machine interaction and automated gesture recognition. It was meant to provide the reader with a broader picture of the current state of research and applications with a brief technical treatment of the popular methods. We extended the definition of gestures to increase the scope of applications in BCI and EMG based approaches. As described throughout this work, we see the tremendous potential of having systems that we as users can interact with in an intuitive manner. This can make the interaction between humans and machine more convenient and efficient. There is a lot of untapped potential on improving accessibility for the specially abled people like sign language understanding, lip reading, braille based touch systems, BCI and EMG based prosthetics etc. Many have forecasted that with the time to come, the humans and machines will work together in an entwined fashion. There is always the flipside to most technological advancement. As the presence of machines in our lives are going to be immense, it presents many challenges. How can we ensure that the devices we build are safe and does not affect the physical and mental wellbeing of the user? There are a lot of studies highlighting the negative effects on one’s mental health surrounding social media usage. The more we rely on machines, the higher the impact when something goes wrong. Can we ensure a secure way to communicate? Technologies like EMG can control the limbs of a user. What if this system is compromised to negative actors? As most of the methods and algorithms are based on Machine Learning and Deep Learning, they require a lot of data to train the model. Can there be a privacy related concerns with data. There will be many ethical and moral questions posed. Is there a place we can draw a line?

## References

[1] Khan, U.M., Kabir, Z., Hassan, S. A., Ahmed, S. H.: *A Deep Learning Framework Using Passive WiFi Sensing for Respiration Monitoring*. In: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, pp. 1-6, doi: 10.1109/GLOCOM.2017.8255027 (2017)

[2] Zhao, M., Li T., Alsheikh, M.A., Tian Y., Zhao H., Torralba A., Katabi D.: *Through-Wall Human Pose Estimation Using Radio Signals.* In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7356-7365, doi: 10.1109/CVPR.2018.00768 (2018)

[3] Dix, A.: *Human-Computer Interaction*. In: Encyclopedia of Database Systems, pp. 1734–1739. New York, NY: Springer New York (2018).

[4] Mitra, S., & Acharya, T.: *Gesture Recognition: A Survey*. In: IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews: A Publication of the IEEE Systems, Man, and Cybernetics Society, vol. 37(3), pp. 311–324. doi:10.1109/tsmcc.2007.893280 (2007).

[5] Sarkar, A.R., Sanyal, G., Majumder, S.: *Hand Gesture Recognition Systems: A Survey.* In: International Journal of Computer Applications (2013)

[6] Bansode, R., Pashte, S., Sawant, S., Sabnis, S.K.: *Drowsy Driver Detection System.* In: International Journal for Scientific Research & Development, vol. 5, no. 2, pp. 2134–2137 (2016)

[7] Janveja, I., Nambi, A., Bannur, S., Gupta, S., & Padmanabhan, V.: *InSight: Monitoring the state of the driver in low-light using smartphones*. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4(3), pp.1–29. doi:10.1145/3411819 (2020).

[8] Low, T., Bubalo, N., Gossen, T., Kotzyba, M., Brechmann, A, Huckauf, A., Nürnberger, A.: *Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking*. In: Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17). Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3020165.3022131 (2017).

[9] Chen, Y., Tian, Y., He, M.: *Monocular Human Pose Estimation: A Survey of Deep Learning-based Methods.* In: Computer Vision and Image Understanding (CVIU), vol. 192, https://doi.org/10.1016/j.cviu.2019.102897 (2020).

[10] Zheng, C., Wu, W., Yang, T., Zhu, S., Chen, C., Liu, ., Shen, J., Kehtarnavaz, N., Shah, M.: *Deep Learning-Based Human Pose Estimation: A Survey.* Arxiv Preprint (2021)

[11] Marinoiu, E., Papava, D., & Sminchisescu, C.: *Pictorial human spaces: How well do humans perceive a 3D articulated pose?* In: 2013 IEEE International Conference on Computer Vision (2013)

[12] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J.: SMPL: A skinned multi-person linear model. In: ACM Transactions on Graphics, vol. 34(6), pp. 1–16 (2015)

[13] Pons-Moll, G., Romero, J., Mahmood, N., & Black, M. J.: Dyna: A model of dynamic human shape in motion. In: ACM Transactions on Graphics, vol. 34(4), pp. 1–14 (2015)

[14] Zuffi, S., Black, M.J.: *The Stitched Puppet: A Graphical Model of 3D Human Shape and Pose.* In: Computer Vision and Pattern Recognition (CVPR) (2015).

[15] Joo, H., Simon, T., Sheikh, Y.: *Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies.* In: Computer Vision and Pattern Recognition (CVPR) (2018).

[16] 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: Computer Vision and Pattern Recognition (CVPR) (2020).

[17] Pfister T., Simonyan K., Charles J., Zisserman A.: *Deep Convolutional Neural Networks for Efficient Pose Estimation in Gesture Videos*. In: (eds) Computer Vision – ACCV 2014. ACCV 2014. Lecture Notes in Computer Science, vol. 9003. Springer, Cham. https://doi.org/10.1007/978-3-319-16865-4_35 (2015).

[18] Li, S., Liu, Z.Q., Chan, A.B.: *Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network*. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 482–489 (2014).

[19] Gkioxari, G., Hariharan, B., Girshick, R., Malik, J.: *R-cnns for pose estimation and action detection*. In: arXiv preprint arXiv:1406.5212 (2014).

[20] Fan, X., Zheng, K., Lin, Y., Wang, S.: *Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation*. In: arXiv preprint arXiv:1504.07159 (2015)

[21] Luvizon, D.C., Picard, D., Tabia, H.: *2D/3D pose estimation and action recognition using multitask deep learning*. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 5137–5146 (2018)

[22] Luvizon, D.C., Tabia, H., Picard, D.: *Human pose regression by combining indirect part detection and contextual information*. In: arXiv preprint arXiv:1710.02322 (2017)

[23] Nibali, A., He, Z., Morgan, S., Prendergast, L.: *Numerical coordinate regression with convolutional neural networks.* In: arXiv preprint arXiv:1801.07372 (2018)

[24] Carreira, J., Agrawal, P., Fragkiadaki, K., Malik, J.: *Human pose estimation with iterative error feedback*. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4733–4742 (2016)

[25] Sun, X., Shang, J., Liang, S., Wei, Y.: *Compositional human pose regression*. In: Proc. IEEE International Conference on Computer Vision, p. 7 (2017)

[26] Holland, A. C., O'Connell, G., & Dziobek, I.: *Facial mimicry, empathy, and emotion recognition: a meta-analysis of correlations*. In: Cognition & Emotion, vol. 35(1), pp.150–168. (2021). https://doi.org/10.1080/02699931.2020.1815655

[27] Cuncic, A.: *How to better understand facial expressions*. In: Verywell Mind. Retrieved December 19, 2021, from https://www.verywellmind.com/understanding-emotions-through-facial-expressions-3024851 (March 30, 2021)

[28] Cowen, A.S., Keltner, D., Schroff, F., Jou, B., Adam, H., Prasad G.: *Sixteen facial expressions occur in similar contexts worldwide*. In: Nature, vol.589(7841), pp. 251-257 (2021)

[29] Nogales, R.E., Benalcázar, M.E.: *Hand gesture recognition using machine learning and infrared information: a systematic literature review*. In: International Journal of Machine Learning and Cybernetics, vol. 12, pp. 2859–2886 (2021)

[30] Luo, W., Schwing, A. G., & Urtasun, R.: *Efficient deep learning for stereo matching*. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

[31] Hamid, M. S., Fajar, N., Manap, A., Hamzah R.A., Kadmin A.F.: *Stereo matching algorithm based on deep learning: A survey*. In: Journal of King Saud University - Computer and Information Sciences (2020)

[32] Rao, N., Surana, P.M, Ragesh, R., Srinivasa G.: *Analysis of Joints for Tracking Fitness and Monitoring Progress in Physiotherapy.* In: The Proceedings of IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA 2019), Malaysia (2019)

[33] Yang, W., Peng, Y., & Xie, H.: *Action Recognition Based on Kinect Deep Learning*. In: Journal of Frontiers of Society, Science and Technology, vol. 1(2), pp.11-15 (2021)

[34] Anson, D., Brandon, C., Davis, A., Hill, M., Michalik, B., & Sennett, C.: *Swype vrs. conventional on-screen keyboards: Efficacy compared*. In: RESNA Annual Conference (2012)

[35] Shokat, S., Riaz, R., Rizvi, S. S., Abbasi, A. M., Abbasi, A. A., & Kwon, S. J.: *Deep learning scheme for character prediction with position-free touch screen-based Braille input method*. In: Human-Centric Computing and Information Sciences, vol. 10(1), pp. 1-24 (2020)

[36] Lu, D., Yu, Y., & Liu, H.: *Gesture recognition using data glove: An extreme learning machine method.* In : 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, pp. 1349-1354 (2016) https://doi.org/10.1109/robio.2016.7866514

[37] Bablani, A., Edla, D. R., Tripathi, D., & Cheruku, R.: *Survey on brain-computer interface: An emerging computational intelligence paradigm*. In: ACM Computing Surveys, vol. 52(1), pp.1–32 (2019)

[38] Jaramillo, A. G., & Benalcazar, M. E.: *Real-time hand gesture recognition with EMG using machine learning.* In: 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM). pp. 1-5 (2017)

[39] Baldeon, K., Oñate, W., & Caiza, G.: *Augmented reality for learning sign language using Kinect tool*. In: Smart Innovation, Systems and Technologies, pp. 447–457. Springer Singapore (2021)

[40] Ergürel, D.: *Leap Motion announces $50 million in Series C funding*. In: Haptical. https://haptic.al/leap-motion-announces-50-million-in-series-c-funding-a1a1f8c0440a (2017, July 18)

[41] Hayward, A.: *Nintendo Wii U review*. In: TechRadar. https://www.techradar.com/reviews/gaming/games-consoles/nintendo-wii-u-1084120/review (2015, December 1)

[42] Grover, S.: *Myo gesture armband*. In: CyberGeeks. https://cybergeeks.in/myo-armband/ (2014, December 30)

[43] *Data Glove_Products & Solutions_Goertek*. In: Goertek.Com. Retrieved December 19, 2021, from https://www.goertek.com/en/content/details62_16718.html (n.d.)

[44] Koles, Z.J, Lazar, M.S, Zhou, S.Z.: *Spatial patterns underlying population differences in the background EEG*. In: Brain Topography, vol. 2(4), pp. 275–284 (1990)

[45] Boye, A. T., Kristiansen, U. Q., Billinger, M., Nascimento, O. F. do, & Farina, D.: *Identification of movement-related cortical potentials with optimized spatial filtering and principal component analysis*. In: Biomedical Signal Processing and Control, vol.3(4), pp.300–304 (2008)

[46] Andersen, A. H., Gash, D. M., & Avison, M. J.: *Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework*. In: Magnetic Resonance Imaging, vol.17(6), pp.795–815 (1999)

[47] Herault, J., Jutten, C., Denker, J.S.: *Space or time adaptive signal processing by neural network models*. In: AIP Conference Proceedings, vol. 151, pp. 206–211 (1986)

[48] Xu, N., Gao, X., Hong, B., Miao, X., Gao, S., Yang, F. *BCI competition 2003-dataset IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications*. In: IEEE Transactions on Biomedical Engineering, vol.51(6), pp.1067–1072 (2004)

[49] Bell, A.J, Sejnowski, T.J.: *An information-maximization approach to blind separation and blind deconvolution*. In: Neural Computation, vol.7(6), pp. 1129–1159 (1995)

[50] Delorme, A., & Makeig, S.: *EEG changes accompanying learned regulation of 12-Hz EEG activity*. In: IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society, vol.11(2), pp.133–137 (2003)

[51] Kanoga, S., Nakanishi, M., & Mitsukura, Y.: *Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram.* In: Neurocomputing, vol.193, pp. 20–32 (2016)

[52] Ting, W., Guo-zheng, Y., Bang-hua, Y., & Hong, S.: *EEG feature extraction based on wavelet packet decomposition for brain computer interface*. In: Measurement: Journal of the International Measurement Confederation, vol.41(6), pp. 618–625 (2008)

[53] Yang, B.-H., Yan, G.-Z., Wu, T., & Yan, R.-G.: *Subject-based feature extraction using fuzzy wavelet packet in brain–computer interfaces*. In: Signal Processing, vol.87(7), pp. 1569–1574 (2007)

[54] Wang, X., Xia, M., Cai, H., Gao, Y., & Cattani, C.: *Hidden-Markov-Models-based dynamic hand gesture recognition*. In: Mathematical Problems in Engineering, pp. 1–11 (2012)

[55] Yamato, J., Ohya, J., Ishii, K.: *Recognizing human action in time sequential images using hidden Markov model*. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, IL, pp. 379–385 (1992)

[56] Starner, T., & Pentland, A. *Real-time american sign language recognition from video using hidden markov models*. In: Motion-based recognition. Springer, Dordrecht, pp. 227-243 (1997)

[57] Starner, T., Weaver, J., & Pentland, A.: *Real-time American sign language recognition using desk and wearable computer based video.* In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 1371–1378 (1998)

[58] Isard, M., Blake A.: *CONDENSATION -- conditional density propagation for visual tracking.* In: Int. J. Comput. Vis., vol. 1, pp. 5–28 (1998)

[59] Black, M. J., Jepson, A. D.: *A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions*. In: Proc. 5th Eur. Conf. Comput. Vis., vol. 1, pp. 909–924 (1998)

[60] Davis, J., Shah, M.: *Visual gesture recognition*. In: Vis., Image Signal Process., vol. 141, pp. 101–106 (1994)

[61] Hong, P., Turk M., Huang, T. S.: *Gesture modeling and recognition using finite state machines*. In: Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recogn., Grenoble, France, pp. 410–415 (2000)

[62] Yeasin, M., Chaudhuri, S.: *Visual understanding of dynamic hand gestures.* In: Pattern Recogn., vol. 33, pp. 1805–1817, (2000)

[63] Tur, A. O., & Keles, H. Y.: *Evaluation of hidden Markov models using deep CNN features in isolated sign recognition.* In: Multimedia Tools and Applications, vol. 80(13), pp. 19137–19155 (2021)

[64] Pigou, L., Dieleman, S., Kindermans, P.J., Schrauwen, B.: *Sign language recognition using convolutional neural networks*. In: Workshop at the European Conference on Computer Vision, pp. 572–578. Springer (2014)

[65] Nishida, N., Nakayama, H.: *Multimodal Gesture Recognition Using Multi-stream Recurrent Neural Network*. In: Image and Video Technology, Lecture Notes in Computer Science, pp. 682–694. Springer International Publishing, Cham (2016)

[66] Núñez, J. C., Cabido, R., Pantrigo, J. J., Montemayor, A. S., & Vélez, J. F.: Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognition*, vol. 76, pp.80–94 (2018)

[67] Zheng, Z., Chen, Z., Hu, F., Zhu, J., Tang, Q., Liang, Y.: *An Automatic Diagnosis of Arrhythmias Using a Combination of CNN and LSTM Technology*. In: Electronics, vol.9(1), p.121 (2020)

[68] Fenghour, S., Chen, D., Guo, K., Li, B., & Xiao, P. Deep learning-based automated lip-reading: A survey. *IEEE Access: Practical Innovations, Open Solutions*, vol. 9, pp. 121184–121205 (2021)

[69] Trachuk, T., Vdovichena, O., Andriushchenko, M., Semenda, O., Pashkevych, M.: *Branding and Advertising on Social Networks: Current Trends.* In: IJCSNS International Journal of Computer Science and Network Security, vol.21 no.4, pp. 178 -185 (2021)

[70] Jain, A., Tompson, J., Andriluka, M., Taylor, G.W., Bregler, C.: *Learning human pose estimation features with convolutional networks*. In: arXiv preprint arXiv:1312.7302 (2013)

[71] Jain, A., Tompson, J., LeCun, Y., Bregler, C. Modeep.: *A deep learning framework using motion features for human pose estimation*. In: Proc. Asian conference on computer vision, Springer. pp. 302–315 (2014)

[72] Tang, Z., Peng, X., Geng, S., Wu, L., Zhang, S., Metaxas, D.: *Quantized densely connected u-nets for efficient landmark localization*. In: Proc. European Conference on Computer Vision, pp. 339–354 (2018)

[73] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C.: *Efficient object localization using convolutional networks*. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 648–656c (2015)

[74] Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: *Joint training of a convolutional network and a graphical model for human pose estimation*. In: Advances in neural information processing systems, pp. 1799–1807 (2014)

[75] Iqbal, U., Milan, A., & Gall, J.: *PoseTrack: Joint Multi-person Pose Estimation and Tracking.* In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

[76] Fang, H., Xie, S., Tai, Y.W., Lu, C.: *Rmpe: Regional multi-person pose estimation*. In: Proc. IEEE International Conference on Computer Vision, pp. 2334–2343 (2017)

[77] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P.V., Schiele, B.: *Deepcut: Joint subset partition and labeling for multiperson pose estimation*. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4929–4937 (2016)

[78] Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., Schiele, B.: *Deepercut: A deeper, stronger, and faster multi-person pose estimation model.* In: Proc. European Conference on Computer Vision, Springer. pp. 34–50 (2016)

**Konkina Nataliia,** Master Department of Automation of Power Processes and Systems Engineering (APEPS), Faculty of heat power engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine, https://orcid.org/0000-0002-5325-507X