

Approach to Hand Tracking and Gesture Recognition Based on Depth-Sensing Cameras and EMG Monitoring

Ondrej Kainz¹, František Jakab¹

¹ Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Košice
Letná 9, 042 00 Košice, Slovakia

{ondrej.kainz, frantisek.jakab}@tuke.sk

Abstract: In this paper, a new approach for hand tracking and gesture recognition based on the Leap Motion device and surface electromyography (SEMG) is presented. The system is about to process the depth image information and the electrical activity produced by skeletal muscles on forearm. The purpose of such combination is enhancement in the gesture recognition rate. As a first we analyse the conventional approaches toward hand tracking and gesture recognition and present the results of various researches. Successive topic gives brief overview of depth-sensing cameras with focus on Leap motion device where we test its accuracy of fingers recognition. The vision-SEMG-based system is to be potentially applicable to many areas of human computer interaction.

Keywords: Depth-sensing camera, electromyography, gesture recognition, hand tracking, Leap Motion.

1 Introduction

Contemporary technology allows us to transfer almost any component of the real world into the digital form. Nature of such a component can be not only static but also dynamic, e.g. whole body, arms or legs movement. Transfer to the digital form is only one essence of the whole story. Another important step for machine (in our case personal computer) is to recognize and identify an object. For purposes of motion tracking, device must be able to sense the changes in real time; this is enabled by specific sensors dependent on used technology.

Motion tracking offers the ability to control the device. Approaches towards control and the advancement of their research have had virtually exponential nature. Modern sensors and technological advancements enable us to create sophisticated sensors that are able to precisely sense the environment and its changes. As we are part of this environment we are also subject of sensor monitoring. Human limbs such as hands or whole body have been used as a tool for control of machine through the recognition of specific patterns, as learnt by machine.

In the following chapters we present standard approaches towards hand tracking and gesture recognition and propose a new approach realized by means of image processing and body signals monitoring.

2 Hand tracking and gesture recognition

Gestures as such arise from a person's mental concept and are expressed through the motion of arms and hands. These expressions are then observed and recognized by the spectator (Pavlovic, Sharma, Huang, 1997). In our case the spectator is computational device that is able to recognize gestures through specific pre-learnt models.

Two conventional approaches towards gesture recognition are free-hand tracking and glove-based tracking. Both of these approaches have in common the need for specific mathematical model. First approach, *free-hand tracking*, enables control without any mechanical devices directly attached to user; the only input here is based on image processing (Mazumdar, Talukdar, Sarma, 2013). Second approach, *glove-based tracking*, uses mechanical device fixed on hand with sensors that track the movement. This kind of tracking is not affected by surroundings and allows the usage of multiple sensors such as force sensor, accelerometer, gyroscope or bend sensor.

As for the free-hand tracking Pavlovic, Sharma & Huang (1997) further define two main models. First, *3D hand-based models* are based on 3D description of hand. Further divided into *volumetric model*, where hand/arm postures are analysed by synthesizing of 3D model of human hand while its parameters are modulated until compliance with real human hand is found, and *skeletal model* - comparable to volumetric model but this model deals with reduced set of equivalent joint angle parameters together with segment lengths. 3D-based model was used for sign language recognition by Mehrez & Jemni (2012) with recognition rate of 98.5 using 900 ASL. Second, *appearance-based models*, are not directly derived from 3D description of hand but are rather based on display of hands in sequence of images. Great numbers of such models use parameters that are derived from images in the templates such as contours and edges, image moments or image eigenvectors.

3 Depth-sensing cameras

Hand gesture recognition has become widely implemented as one of the features offered to customers for computer control. Software tools that come with these devices have predefined

tracking patterns for specific operating system. Specific tools that offer customization are generally available for development.

Kinect by Microsoft is one of the most popular devices available on market among developers. This device was primarily intended for X-Box 360 video game console, few month after its release Microsoft announced *Kinect for Windows*. Kinect for Windows is a depth sensing and RGB camera – specifically it consists of depth camera, RGB camera, IR sensor, tilting motor and four microphones. Several tracking options are possible using Kinect: skeleton, face and hand tracking (Jana, 2012). For further specifications see Tab. 1. Due to its availability Kinect has become very popular in the area of research, mainly in the field of robotics (Hernández-López et al., 2012) and human-computer interaction. Other sorts of utilization were proposed as well – Changa, Chenb & Huang (2011) and González-Ortega et al. (2014) pointed out possibility of its usage for physical rehabilitation to be successful. A different – and for us more interesting – research is the one of Dominio, Donadeo & Zanuttigh (2013) which resulted in hand gestures recognition using the extracted depth data. Another analogous area of gesture recognition is sign language recognition – several research papers dealing with this type of recognition provided effective and useful results (Sun C. et al., 2013), (Verma, Aggarwal, Chandra, 2013).

Another successful product on the market that enables depth recognition is called *Asus Xtion PRO LIVE*, depth-sensing and RGB camera. Jiang, Liu & Matsumaru (2012) proved the ability of this device to recognize background and also hand in the real-time. Research showed the inability to recognize proper multi-touch functionality. Haggag et al. (2013) proved Kinect to be superior when compared to Xtion PRO. Identically to Kinect, it is a low cost device, for further specifications see Tab. 1.

Designers of device called *Leap Motion* took another approach to depth sensing. Same as above mentioned is Leap Motion also equipped with depth sensing camera; however no additional features such as RGB camera or microphone are available. This device is also not focused on skeletal or face tracking; its principal use is aimed towards hand tracking and gesture recognition. Sabir et al. (2013) combined in their research Leap Motion with Kinect and were able to provide enhanced tracking and with six degrees of freedom - three axes for translation and the same for rotation, this research was aim to provide control for graphics system. Many other depth-sensing cameras solution are available, such as Creative Senz3D or SoftKinetic DS325, however both of these provide only small depth image size of QVGA (320x240).

	Microsoft Kinect	Asus Xtion PRO	Leap Motion
Distance (cm)	40 - 350	80 - 350	7 - 25
Field of view (horizontal, vertical, diagonal)	57°, 43°	58°, 45°, 70°	140°
Sensor	RGB, Depth, Microphone	RGB, Depth, Microphone	Depth
Depth image size	VGA (640x480)	VGA (640x480) : 30 fps, QVGA (320x240): 60 fps	VGA (640x480)
Resolution	SXGA– (1280 x 960)	SXGA (1280x1024)	N/A
Programming language support	C++, C#	C++, C#, Java	C++, C#, Java, JavaScript, Python

Dimensions (cm)	28 x 8 x 8	18 x 3.5 x 5	8 x 3 x 1
-----------------	------------	--------------	-----------

Tab. 1. Depth-sensing cameras comparison.

3.1 Hand tracking using Leap Motion

Mobility, price and distance interaction range were factors that led us to select Leap Motion as a device to be tested for hand tracking purposes. Our focus was placed primary on accuracy of the Leap Motion device - first for one hand and latter for both hands. Tests were conducted in the laboratory environment under regular lightning conditions, such conditions are expected for regular usage. No smudges or any other obstacles that might disrupt the measurement were observed. For measurement was used official web tool *Motion Dump* from the developer. Hands were in horizontal position in height of approx. 16 centimetres from device, with interaction height set to just declared value. The goal of testing was to estimate the accuracy of the device. Results were collected from over 1500 samples per one measurement. As we can see from a Fig. 1 the best accuracy was achieved when one finger was used tracking with accuracy rate of 96.34%. Minor change in accuracy was observed when thumb was used for tracking. Minimum accuracy was achieved for the case of three fingers tracking with thumb. Tracking of two hands had poorer tracking for all the fingers. It should be noted that distance between fingers and position of hand and fingers, i.e. adduction or abduction/flexion or extension, had great significance in recognition.

As demonstrated by Sabir el al. (2013) Leap Motion in combination with Kinect is suitable for usage in a 3D environment however its long-term usage as a replacement for mouse or keyboard is highly questionable. This device allows simple gesture recognition in default state it is not suitable for complex tasks. Further research on specific types of its utilization is to be conducted.

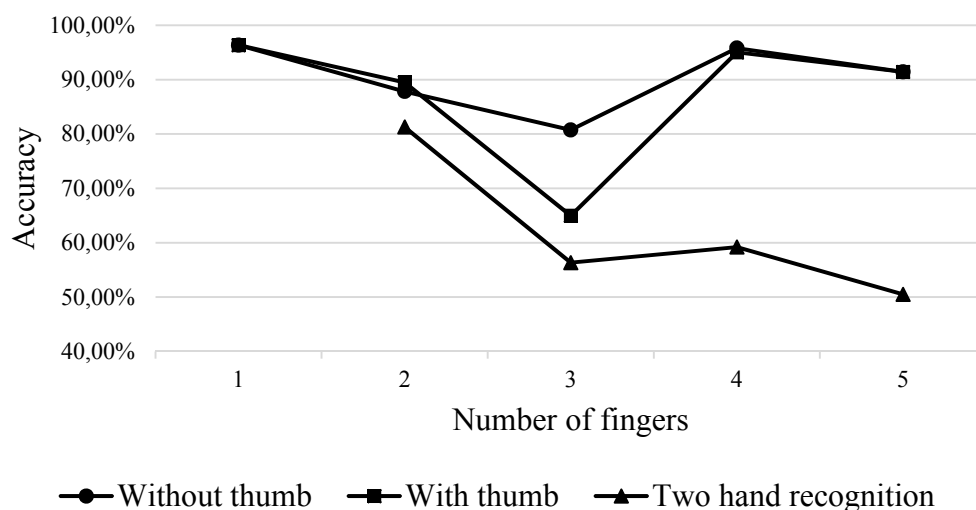


Fig. 1. One hand vs two hand recognition comparison.

4 Surface electromyography in hand gesture recognition

Usage of depth-sensing cameras for detailed gesture recognition might be in some cases troublesome or even impossible. Primary issue is recognition of specific hand positions and recognition in non-laboratory environment. Usage of glove is more invasive and may be

complex for handling. Bearing in mind all just stated we propose innovative approach, which combines depth-sensing cameras and EMG monitoring.

4.1 What is EMG?

A term EMG is known since the 1943 when introduced by Weddell et al. for the description of clinical application – in this process the needle was used for the examination of skeletal muscles (Katirji, 2007). Phenomena of electromechanical coupling in muscle is utilized, more specifically deep invagination of muscle membrane inside the muscle cells enable conduction of electrical impulses in skeletal and cardiac cells, these are called t-tubules. When muscle contraction occurs it can be then graphically depicted as electromyogram and recording is realized using electromyograph (Kumar, Mital, 1996). EMG enables diagnosis of some peripheral nervous system disorders. Traditional approach of EMG includes the usage of needle inserted into the muscle to enable recording. Less invasive or completely non-invasive approach is called surface electromyography (SEMG). Fig. 2 represents the difference between intramuscular and surface EMG. Sensor may be attached to specific muscles to monitor and provide capability for differentiation of the activity of the muscle. SEMG signal is non-linear and short-time stationary.

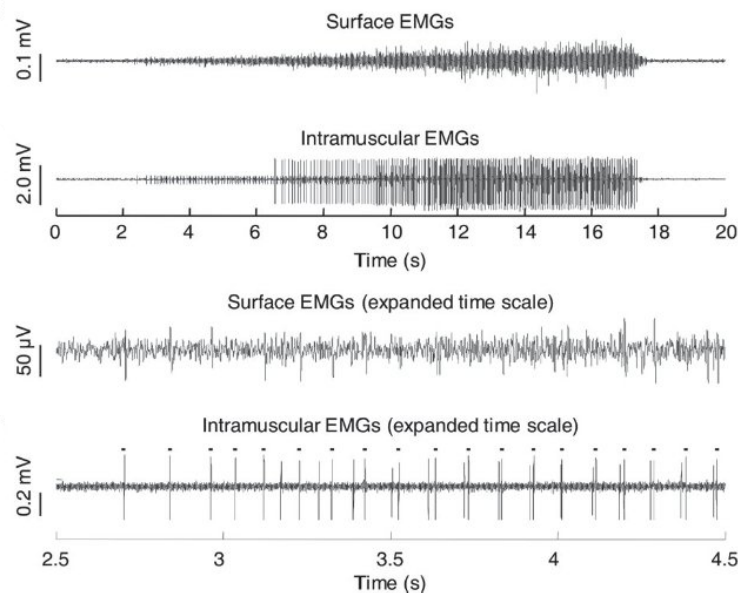


Fig. 2. Surface vs Intramuscular EMG signal of medial gastrocnemius muscle
Source: Garcia, Vieira (2011)

4.2 Related work

Following section introduces current progress in research related to SEMG, its results were taken into consideration during vision-SEMG-system development; focal point is classification of recorded data. Luo, Wang, Ma (2006) used autoregressive model to describe SEMG and utilized model coefficients in order to implement recognition of motion patterns – as a classifier high-order neural network called Pi-Sigma is utilized. Focus of research was on the movement recognition and as one of the conclusions authors pointed out that proper position of electrodes placement play a very important role in the whole measurement process. Zhaojie, Honghai (2014) utilized Fuzzy Gaussian Mixture Models (FGMMs) for hand grasps and manipulation. Results of study show that as for the recognition rate (92.75 percent) these models outperform Gaussian Mixture Models and as likewise vector machine (SVM). Combination of linear and nonlinear measurements while utilizing adaptive neuro-

fuzzy inference system (ANFIS) was introduced by Gaoxiang et al. (2014) with pretty good accuracy of 99.1 percent. Prototype for mobile device control was proposed and implemented by Zhiyuan et al. (2014), this approach results with the accuracy of 95 percent while utilizing Bayes linear classifier (BLC) and dynamic time wrappings (DTW).

5 Vision-SEMG-based system

Approach we take combines SEMG and depth-camera. Visual monitoring of hands realized, as proposed, by Leap Motion sensor device that is directly connected to the computer. Main drawback when using this device is requirement for a user to remain near the interaction zone, i.e. near computer. Hand and fingers position or general gestures were using functions available by Leap Motion API. Four disposable electrodes were utilized in SEMG monitoring, with Arduino microcontroller used to record the signals. Arduino enabled us to utilize its capabilities – we designed shield that was directly attached to microcontroller, this shield serves as an amplifier of EMG signals and connects Arduino board to electrodes via 3.5 mm jack. Electrodes were directly attached to a skin, conductive gel was used as to enhance connection. All the processing of signal is accomplished on data processing machine, i.e. personal computer.

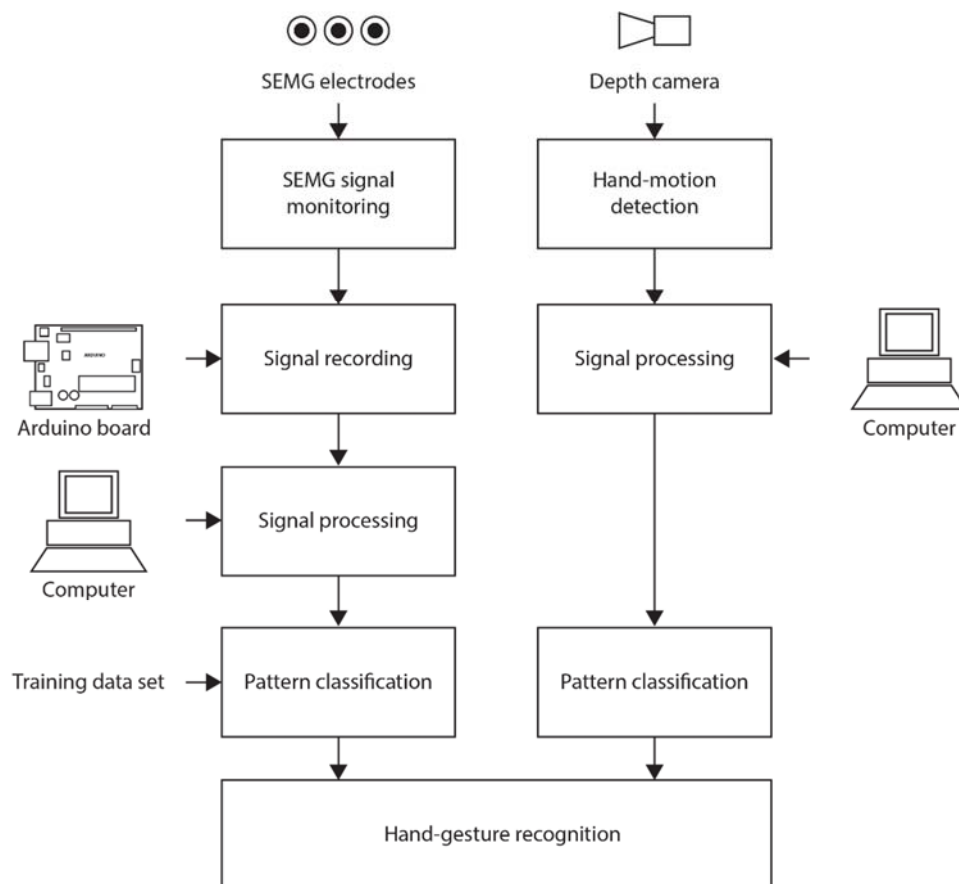


Fig. 3. SEMG and Leap Motion.

Classification of specific patterns for SEMG monitoring is realized by utilization of neural networks. For our system we selected Deep belief network (DBN). This network is type of deep neural network with multi hidden layers connected together while using unsupervised algorithm that trains the network layer after layer. Layer is using a restricted Boltzmann

machine (RBM). Training itself is carried on in unsupervised way. Future development will focus on combination of DBN and hidden Markov model (HMM). As analysed by Zheng et al. (2014) utilization of DBN shows fine accuracy, however the very same research concluded better results while utilizing DBN-HMM. Research was focused on electroencephalography (EEG) reading, while in our case we use it for EMG – nevertheless accuracy was analogical as presented by authors. We have developed and tested rudimentary system with accuracy of 86 percent – hand gasp. When implemented together with depth camera accuracy of 95 percent was achieved. Both of these monitoring devices were first designed and tested separately. Hand as detected by depth camera and signal from SEMG electrodes were deployed as a whole system for hand gesture monitoring as depicted in Fig. 3. This system may be used for tracking of one hand only.

When monitoring specific muscles determined hand or even fingers movement is enabling us to track gestures, for instance for monitoring of forearm flexion or extension (Fig. 4). Together with Leap Motion recognition rate is sufficient enough for monitoring of virtually all gestures with high certainty. Monitoring of SEMG and visual tracking using Leap Motion can be further used for other purposes of non-invasive hand tracking and gesture recognition, e.g. use for sign language recognition or wireless machine control.

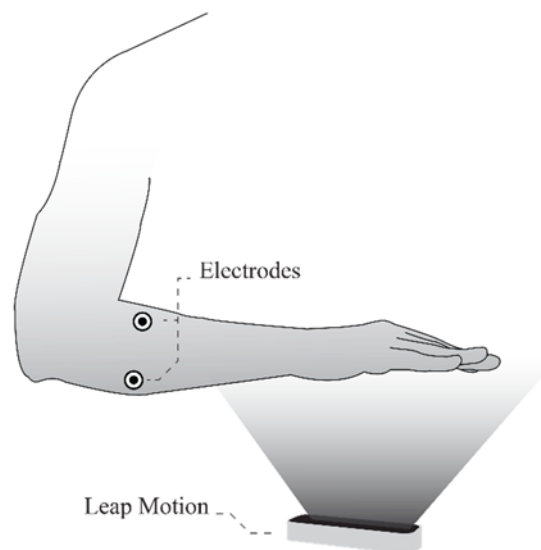


Fig. 4. Forearm flexor/extensor electrode placement with Leap Motion tracking.

6 Conclusion and future works

This study presented a new hand tracking approach for gesture recognition. This approach is indented for human computer interaction using SEMG technique in combination with visual-based tracking system with depth recognition. Monitoring is achieved through the hand muscles EMG signals tracking and visual image detection of hand movement. Leap Motion device proved ability to effectively recognize gestures, however issues with continuous finger tracking and certain hand positions were observed. This system is supposed to overcome shortcomings of solely visual-based system where tracking of hand is not always recognized due obstacles – e.g. hand covered by hand. We expect utilization of SEMG system in many areas. Next stage of project is to design a wireless armband that will include sensors for SEMG monitoring, this might allow usage of solely SEMG device without the need of utilizing Leap Motion. Another aim is to develop such device in accordance with WBAN standard.

Acknowledgement

Research described in the paper was supervised by Assoc. Prof. Frantisek Jakab, PhD., Computer Networks Laboratory at DCI FEEI TU in Košice. We support research activities in Slovakia/This project is being co-financed by the European Union. Paper is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF.

References

- Changa, Y.-J., Chenb, S.-F., Huang, J.-D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, 32(6), 2566–2570. doi: j.ridd.2011.07.002
- Dominio, F., Donadeo, M., Zanuttigh, P. (2013). Combining multiple depth-based descriptors for hand gesture recognition. *Pattern Recognition Letters*, in press, doi: j.patrec.2013.10.010
- Gaoxiang, O., Xiangyang, Z., Zhaojie, J., Honghai, L. (2014). Dynamical Characteristics of Surface EMG Signals of Hand Grasps via Recurrence Plot. *IEEE Journal of Biomedical and Health Informatics*, 18(1), 257–265. doi: 10.1109/JBHI.2013.2261311.
- Garcia M.A. C.,Vieira T.M. M. (2011) Surface electromyography: Why, when and how to use it. *Revista Andaluza de Medicina del Deporte*, 10(6), 17-28
- González-Ortega, D., Díaz-Pernas, F.J., Martínez-Zarzuela, M., Antón-Rodríguez, M. (2014). A Kinect-based system for cognitive rehabilitation exercises monitoring. *Computer Methods and Programs in Biomedicine*, 113(2), 620–631. doi: j.cmpb.2013.10.014
- Haggag, H., Hossny, M., Filippidis, D., Creighton, D., Nahavandi, S., Puri, V. (2013). Measuring depth accuracy in RGBD cameras. *7th International Conference on Signal Processing and Communication Systems* (pp. 1-7). doi: 10.1109/ICSPCS.2013.6723971
- Hernández-López, J.-J., Quintanilla-Olvera, A.-L., López-Ramírez, J.-L., Rangel-Butandaa, F.-J., Ibarra-Manzanao, M.-A., Almanza-Ojedab, D.-L. (2012). Detecting objects using color and depth segmentation with Kinect sensor. *Procedia Technology*, 3, 196-204. doi: j.protcy.2012.03.021
- Jana, A. (2012). *Kinect for Windows SDK Programming Guide*. Birmingham: Packt Publishing Ltd.
- Jiang, Y., Liu, Y., Matsumaru, T. (2012). Applying infrared radiation image sensor to step-on interface: Touched point detection and tracking, *IEEE/SICE International Symposium on System Integration* (pp. 752-757). doi: 10.1109/SII.2012.6427302
- Katirji B. (2007) *Electromyography in Clinical Practice: A Case Study Approach*. Philadelphia: Mosby Elsevier.
- Kumar S., Mital A. (1996). *Electromyography in Ergonomics*. London: T J Press.
- Luo Z., Wang F., Ma W. (2006). Pattern Classification of Surface Electromyography Based on AR Model and High-order Neural Network. *Proceedings of the 2nd IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications* (pp. 1-6), doi: 10.1109/MESA.2006.296982
- Mazumdar, D., Talukdar, A.K. & Sarma, K.K. (2013). Gloved and free hand tracking based hand gesture recognition. *1st International Conference on Emerging Trends and Applications in Computer Science* (pp. 197-202). doi: 10.1109/ICETACS.2013.6691422
- Mehrez, B., Jemni, M. (2012). 3D Motion Trajectory Analysis Approach to Improve Sign Language 3D-based Content Recognition. *Proceedings of the International Neural Network Society Winter Conference* (13 vol., pp. 133-143). doi: 10.1016/j.procs.2012.09.122
- Pavlovic, V.I., Sharma, R., & Huang, T.S. (1997). Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 677 – 695., doi: 10.1109/34.598226

- Sabir, K., Stolte, C., Tabor, B., O'Donoghue, S.I. (2013). The Molecular Control Toolkit: Controlling 3D molecular graphics via gesture and voice. *IEEE Symposium on Biological Data Visualization* (pp. 49-56), doi: 10.1109/BioVis.2013.6664346
- Sun, C., Zhang, T., Bao, B.-K., Xu, C., Mei, T. (2013). Discriminative exemplar coding for sign language recognition with Kinect. *IEEE Transactions on Cybernetics*, 43(5), 1418-1428. doi: 10.1109/TCYB.2013.2265337
- Verma, H.V., Aggarwal, E., Chandra, S. (2013). Gesture recognition using kinect for sign language translation. *IEEE Second International Conference on Image Information Processing* (pp. 96-100). doi: 10.1109/ICIIP.2013.6707563
- Zhaojie, J., Honghai L. (2014). Human Hand Motion Analysis With Multisensory Information. *IEEE/ASME Transactions on Mechatronics*, 19(2), 456–466. doi: 10.1109/TMECH.2013.2240312
- Zheng, W.-L., Zhu J.-Y., Peng Y., Lu, B.-L. (2014). EEG-Based Emotion Classification Using Deep Belief Networks. *IEEE International Conference on Multimedia & Expo (ICME)*
- Zhiyuan, L., Xiang, Ch., Qiang, L., Xu, Z., Ping, Z. (2014). A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices. *IEEE Transactions on Human-Machine Systems*, 44(2), 293 – 299. doi: 10.1109/THMS.2014.2302794