The 2016 AAAI Fall Symposium Series:
Artificial Intelligence for Human-Robot Interaction
Technical Report FS-16-01

SLIRS: Sign Language Interpreting System for Human-Robot Interaction

Nazgul Tazhigaliyeva
School of Informatics
College of Science & Engineering
University of Edinburgh
Edinburgh, UK
nazgul.tazhigaliyeva@gmail.com

Yerniyaz Nurgabylov
School of Science & Technology
Nazarbayev University
Astana, Kazakhstan
yerniyaz.nurgabylov@nu.edu.kz

German I. Parisi
Knowledge Technology Group
Department of Informatics
University of Hamburg
Hamburg, Germany
german.parisi@gmail.com

Anara Sandygulova
School of Science & Technology
Nazarbayev University
Astana, Kazakhstan
anara.sandygulova@nu.edu.kz

Abstract

Deaf-mute communities around the world experience a need in effective human-robot interaction system that would act as an interpreter in public places such as banks, hospitals, or police stations. The focus of this work is to address the challenges presented to hearing-impaired people by developing an interpreting robotic system required for effective communication in public places. To this end, we utilize a previously developed neural network-based learning architecture to recognize Cyrillic manual alphabet, which is used for finger spelling in Kazakhstan. In order to train and test the performance of the recognition system, we collected a depth data set of ten people and applied it to a learning-based method for gesture recognition by modeling motion data. We report our results that show an average accuracy of 77.2% for a complete alphabet recognition consisting of 33 letters.

Introduction

Hearing-impaired people around the world communicate via a sign language, which uses gestures to express meaning and intent that include hand-shapes, arms and body, facial expressions and lip-patterns (Tolba and Elons 2013). Similar to spoken languages, each country or region has its own sign language of varying grammar and rules, leading to a few hundreds of sign languages that exist today (Aran 2008). In addition, many deaf-mute people are not able to understand a written spoken language. Research and development on sign language analysis and recognition started with wearable devices such as gloves with sensors and trackers, colored gloves or colored fingers (Sahoo, Mishra, and Ravulakollu 2014). In contrast, vision-based systems provide a natural way of communicating for deaf people, however it still remains to be a challenging problem for effective hand detection, segmentation, and tracking (Aran 2008).

This paper describes the SLIRS project, which aims to develop an interpreting robotic system of a sign language tailored for Kazakhstan. Having consulted our local hearing-impaired community on their needs, SLIRS's first priority is to develop an interpreting system of the sign language vocabulary required for effective communication in Centers for Public Services (CPS). It is the place where our deaf-mute community has to constantly seek help from relatives or paid interpreting services to manage their identification documents, social payments and other usually urgent and highly significant matters.

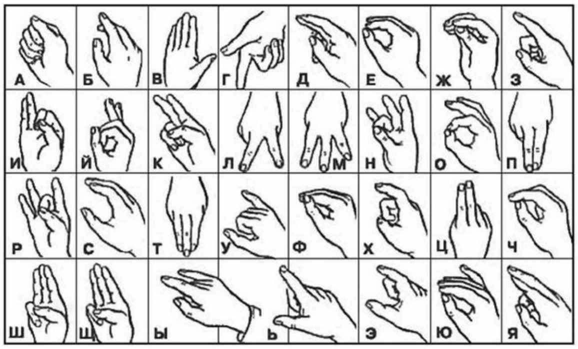

Figure 1: Sign language of Cyrillic alphabet

This research seeks to create such a robotic system tackling a few objectives: 1) to be able to automatically recognize sign language utilizing multi-sensory input from Leap Motion technology i.e. robot's depth sensor. The robot would then 2) vocalize the recognized text with a synthesized speech to be understood by the public services employees and gesticulate the response back to a hearing-impaired individual. And finally, SLIRS aims to achieve 3) action selection for social autonomy of the robot given perception input from the first objective.

In order to tackle the first objective, the focus of this paper is to report our method of fingerspelling recognition, which is mainly used to spell proper nouns, scientific and foreign borrowed terms, and other words lacking a sign representation. To this end, we utilized our previously developed method (Sandygulova et al. 2016) and applied it to a new application of recognizing 33 manual gestures of Cyrillic alphabet. Although the learning approach is an incremental modification of the learning algorithm proposed by Parisi, Weber, and Wermter (2015), where we just modify the labeling function to return the specific letter value, this paper reports fairly good recognition results of the Cyrillic manual alphabet, which serves as the first milestone in the progress of the SLIRS project.

Related Work

Robust gesture recognition is an essential objective for any sign language interpreting system. This section presents related research efforts that have been carried out to address this objective.

According to Suarez and Murphy (2012), the main components of any type of sign recognition system include image acquisition, hand localization, pose estimation and gesture classification. Once the image is acquired from a depth camera, it is processed using various *hand localization* methods. The problem of segmentation has been addressed by using depth thresholding for hand isolation (Mo and Neumann 2006; Breuer, Eckes, and Müller 2007; Liu and Fujimura 2004; Uebersax et al. 2011; Yoo et al. 2010; Du and To 2011; Frati and Prattichizzo 2011; Ren, Yuan, and Zhang 2011; Trigueiros, Ribeiro, and Lopes 2012) and by placing bounds on the number of pixels inspected in the area of detected hand (Li and Jarvis 2009; Klompmaker, Nebe, and Fast 2012). Some techniques involve predicting the hand location by relating it to other body parts (Fujimura and Liu 2006).

Temporal and spatial information of hands makes the *hand tracking* possible, which in turn leads to dynamic gesture recognition. The research community most often uses the NITE body tracking middleware in combination with the Kinect SDK (Ronchetti and Avancini 2011; Bellmore, Ptucha, and Savakis 2011; Chang, Chen, and Huang 2011; Frati and Prattichizzo 2011; Hassani et al. 2011; Lai 2011; Ramey, González-Pacheco, and Salichs 2011).

As the next step, various classification algorithms are used to categorize a particular sign or hand gesture. These algorithms take segmented hand images and their tracked trajectories as input and make a prediction. According to McNeill (2000), gestures can be categorized to several types: gesticulations (used to emphasize speech), emblems (universal signs, e.g. the "OK" sign) and sign language (used as a speech replacement). The most commonly used *gesture classification* algorithms include Hidden Markov Model (HMMs). HMMs are used for manual gestures with temporal information and are known to have high classification rates (Wang et al. 2008; Tang 2011; Yang et al. 2012; Hassani et al. 2011)). Another popular algorithm is k-Nearest Neighbours (k-NNs), which produces high classification rates for static poses in combination with some preprocessing (Van den Bergh et al. 2011; Feris et al. 2005)). Support Vector Machines (SVMs) are also commonly utilized (Biswas and Basu 2011; Tang 2011; Keskin et al. 2013). Finally, neural network algorithms have also been previously used (Keskin et al. 2013). In general, recent progress in the areas of Computer Vision and Machine Learning has helped advance HCI and HRI fields.

Malassiotis, Aifanti, and Strintzis (2002) developed gesture classification of 20 letters from the Greek Sign Language, primarily of numbers from 0-9. Feris et al. (2005) addressed the problem of finger occlusion that arise in fingerspelling by introducing a small modification to the capture setup. Keskin et al. (2013) utilized *the object recognition by parts* approach to recognize 10 digits of American Sign Language (ASL).

Apart from recognition, sign language and gesture recognition has been utilized for robot control and human-robot interaction. Singh, Jain, and Kumar (2012) proposed a new approach for robust automated real-time robot control tool using Indian Sign Language. Their technique combines feed forward back propagation neural network (FNN) and Hidden Markov Model (HMM) to deal with dynamic sign language recognition, learning and interpretation of continuous signals. The algorithm was integrated with the HOAP-2 humanoid robot generated in WEBOTS. The proposed system achieved 95.34% of recognition and interpretation accuracy of 21 gestures offline.

Sohn, Kim, and Oh (2013) presented a 2-Tier control for a human-robot collaborative tasks. The adult-sized humanoid robot, Hubo, whose lower and upper bodies are controlled separately using data from MoCap and sign language accordingly. The sign language gestures were recorded offline using the MoCap motion capturing system and evaluated by the Monte Carlo learning agent. As a result, the Hubo humanoid robot successfully assisted the human operator in object carrying task.

Luis-Péréz, Trujillo-Romero, and Martínez-Velazco (2011) utilized a set of Mexican Sign Language to make the robot perform specific tasks. The system recognizes and interprets 23 signs of the alphabet with the accuracy of 95.8%. Sugiuchi, Morino, and Terauchi (2002) exploited the sign language to control the multi-fingered Dual robot hand in human mimetic approach and to perform paper cutting and chopstick handling. However, the authors report that despite that the sign language interpretation software system could be effectively implemented, there were major limitations in hardware.

Child-size humanoid robots have been exploited for demonstration of signs and other significant components of sign language such as facial expressions and mimicry (Uluer, Akalın, and Köse 2015). Screen avatars have also been used to interpret written English text into American Sign Language (ASL) (Huenerfauth 2004; Kipp, Heloir, and Nguyen 2011).

This paper does not focus on advancing the state of the art in hand segmentation, localization and/or tracking, but contributes with its approach of using previously developed gesture classification method to classify our dataset which was collected via inexpensive Leap Motion camera.

Data Collection

The depth data was collected on a regular day. Volunteers were brought to the specially allocated classroom to stand

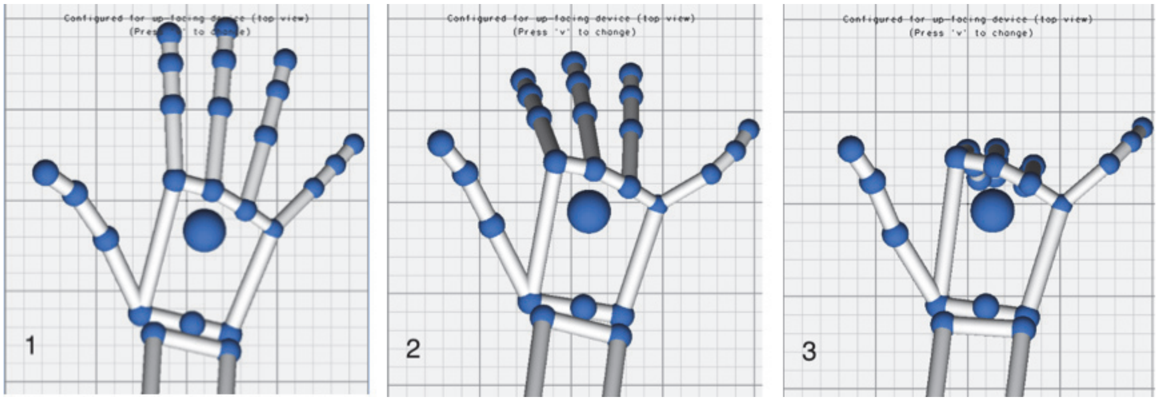

Figure 2: Finger spelling in process: 1. Palm initial position 2. Transition between the initial position and the letter 3. Actual letter

in front of the Leap Motion sensor in order to capture depth data. Each session involved one participant at a time. Each participant was asked to repeat the Cyrillic sign language alphabet letters one by one as it was shown on the video tutorial taken from the official website specifically designed for those who needs to learn the sign language alphabet [1].

The Leap Motion was used to track and estimate depth data from raw motion. The sensor was placed vertically as opposed to traditional horizontal position. The sensor was located in front of the subjects that had to stand and demonstrate gestures and letters from the sign language alphabet. The data was obtained by the method of defining coordinates of the participant's hand and movement of each finger of this hand, precisely the x, y, and z coordinates of each finger joint, orientation and direction of the palm, direction of fingers, frame and hand translation for dynamic gestures. The Leap Motion Visualizer was used for hand and finger tracking purposes. Each letter shown by the participant was recorded in the corresponding .csv file. Each .csv file contained 500 frames of data on average and held data about the transition from the initial position of the palm to the position when the letter could be clearly seen (Fig. 2). The initial position of the palm was needed to ensure robust tracking of fingers by the leap motion sensor.

A Cyrillic alphabet contains 33 letters. A depth dataset was collected for 10 people aged between 20 and 30 years old. Participants without physical disabilities were invited for data collection.

At the end of the experiment the entire data set was processed to contain an equal number of attributes for each frame.

Learning Architecture

The learning architecture consists of 2 hierarchically arranged self-organizing neural networks (Fig. 3). The use of hierarchical self-organization has been shown to be an efficient and effective method for recognizing human motion (Parisi, Weber, and Wermter 2015). This method is consistent with neurophysiological findings that have identified a specialized area for the visual processing of complex motion in the brain in a hierarchical fashion (Rolls and Caan 1982). From a computational perspective, self-organization is an unsupervised learning mechanism that allows to learn representations of the input by iteratively obtaining a nonlinear projection of the feature space (Kohonen 1989). Furthermore, it has been found that learning plays a crucial role in complex motion discrimination. Numerous studies have shown that the recognition speed and accuracy of humans have improved after a number of training sessions (Jastorff, Kourtzi, and Giese 2006).

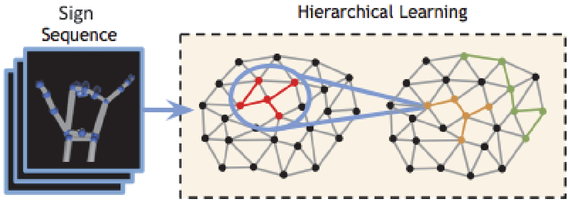

Figure 3: Hierarchical architecture with self-organizing neural learning (GWR networks). Learning is carried out by training a higher-level network with neuron activation trajectories from a lower level network trained on hand gesture sequences.

Hierarchical Self-Organizing Learning

The learning model consists of Growing When Required (GWR) networks (Marsland, Shapiro, and Nehmzow 2002) that iteratively obtain generalized representations of sensory inputs and learn inherent spatio-temporal dependencies. The GWR network is composed of a set of neurons and their associated weight vectors $\mathbf{w}_j$ linked by a set of edges. The activity of a neuron is computed as a function of the distance (usually the Euclidean distance) between the input and its weight vector. During the training, the network dynamically changes its topological structure to better match the input space following competitive Hebbian learning (Martinetz 1993).

[1] http://www.surdo.kz

Table 1: Recognition result for each letter. There are 33 letters in total.

Letter	А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й
Accuracy (%)	92	96	99	88	84.54	72.58	41.24	76.94	74.10	99.00	98.00
Letter	К	Л	М	Н	О	П	Р	С	Т	У	Ф
Accuracy (%)	35	69	98.93	41.67	84.05	97.11	92.09	39.27	68.99	99.27	70.51
Letter	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
Accuracy (%)	83.98	68.39	84.12	68.88	30.63	90.36	82.66	92.54	74.61	76.10	79.09
Average for all	77.2%										

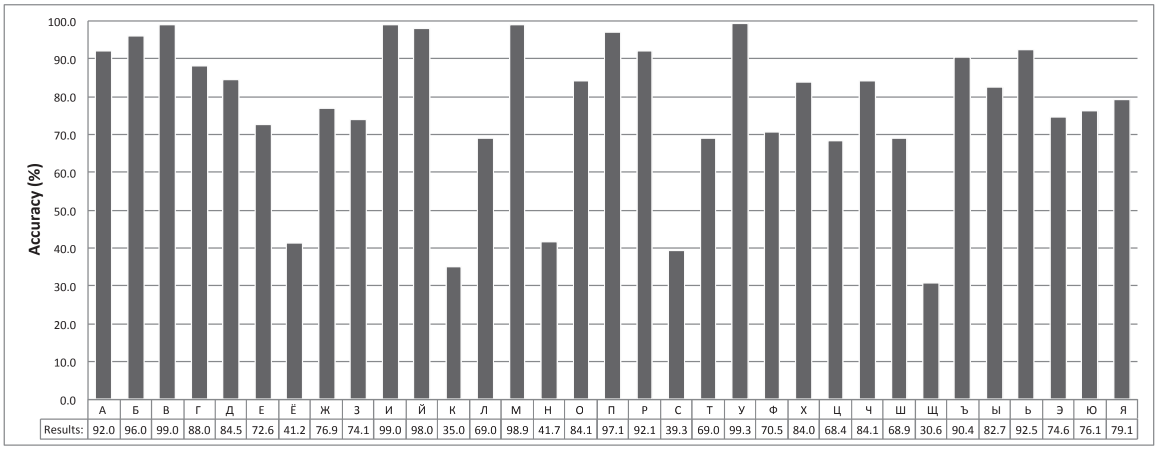

Figure 4: Experimental results: Classification accuracy for each letter.

Different from other models of incremental self-organization, GWR-based learning takes into account the number of times that a neuron has fired so that neurons that have fired frequently are trained less. For this purpose, the network implements a habituation counter to express how frequently a neuron has fired based on a simplified model of how the efficacy of an habituating synapse reduces over time.

This mechanism allows to create new neurons whenever it is required. The GWR algorithm will then iterate over the training set until a given stop criterion is met, in our case a maximum number of iterations. The standard procedure for GWR learning is described by Algorithm 1 (except for Steps 6.c and 7.c that are discussed in the following Section). For GWR learning, we used the following training parameters: insertion threshold $a_T = 0.70$, learning rates $\epsilon_b = 0.3$, and $\epsilon_n = 0.006$, $\kappa = 0.5$, maximum age $a_{max} = 50$, firing counter parameters $\eta_0 = 1$, $\tau_b = 0.3$, $\tau_n = 0.1$, firing threshold $\eta_T = 0.01$, 300 iterations for each network. A thorough discussion of training parameters was presented in (Marsland, Shapiro, and Nehmzow 2002).

The motivation for using hierarchical learning is to use trajectories of neuron activations from one network as input for the training for a subsequent network. This mechanism allows to obtain specialized neurons coding spatio-temporal dependencies of the input, consistent with the assumption that the recognition must be selective for temporal order. Hierarchical learning is carried out by training a higher-level network with neuron activation trajectories from a lower level network. These trajectories are obtained by computing the best-matching neuron of the input sequence with respect to the trained network with N neurons, so that a set of trajectories of length q is given by

$$\Omega^q(\mathbf{x}_i) = \{\mathbf{w}_{b(\mathbf{x}_i)}, \mathbf{w}_{b(\mathbf{x}_{i-1})}, ..., \mathbf{w}_{b(\mathbf{x}_{i-q+1})}\} \quad (1)$$

with $b(\mathbf{x}_i) = \arg\min_{i \in N} \|\mathbf{x}_i - \mathbf{w}_j\|$ and $q = 5$. After training of the higher level network is completed, each neuron will encode a sequence-selective action segment of 5 consecutive frames.

Classification

At recognition time, our goal is to process and classify action sequences in terms of the sign language alphabet. For this purpose, we extended the unsupervised GWR-based learning of the higher level network to attach labels to trained neurons (Algorithm 1, steps 6.c and 7.c). In this case, the network will be trained with the motion sequences in an unsupervised fashion while attaching the labels of the input $\lambda(\mathbf{x}_t)$, i.e.letter, to best-matching neurons. As a result of this process, each neuron in the higher level network encoding a motion segment will be associated to a label. Different from previous approaches using GWR-based associative learning (Parisi, Weber, and Wermter 2015), in our approach each label has a letter value, so that new samples can be processed through the hierarchy and return the label values of the best-matching sequence.

Results

Sequential minimal optimization (SMO) algorithm for training a support vector classifier was used as a baseline for the evaluation of the modified algorithm. As we deal with the sequence of frames for each of 33 classes, standard 2-class classification algorithms could not evaluate our dataset. Therefore, the Multi-Instance Learning Kit (MILK) [2] developed for Weka (Hall et al. 2009) was used for classification. Overall, multi-class SMO 10-fold cross validation showed an accuracy of 12.86%.

The classification accuracy of the modified algorithm for each letter is shown in Table 1 and Figure 4. Our system achieved an average classification accuracy of 77.2% for Cyrillic alphabet, which consists of 33 letters. It should be noted that we trained our method on half of the dataset (five people) and tested the accuracy performance on the other half of the dataset (five other people).

As can be seen from Table 1, the classification accuracy varies from 99% for easy letters such as В and У to 30.63% for a rather complicated for the camera letters such as Ё and Щ. Indeed, the latter mentioned letters differ from their closely related pairs quite slightly, in particular since these two signs are dynamic. As a result, these letters were recognized incorrectly. This problem will be investigated and could be solved once we collect a larger training dataset.

These results suggest that although some letters are harder to recognize than others, our method is suitable at achieving successful results. As reported in previous experiments with Leap Motion (Potter, Araullo, and Carter 2013), metrics extracted from depth information with this camera generally contains noisy samples that may have a negative influence on neural network learning (Parisi, Weber, and Wermter 2015).

On the other hand, although the accuracy of Leap Motion technology is not so precise for some letters, this approach is computationally efficient and allows to extract 3D information in real time, thereby enabling us to recognize fingerspelling with very low latency in a live scenario. This is in fact a very desirable property, since we aim for real-time performance of the SLIRS system in perceptually challenging environments, where slight delays in performance will result in dissatisfaction with the robot and negative experience.

Conclusion

In this paper we utilized our previously developed method and applied it to a new application i.e. recognition of Cyrillic fingerspelling consisting of 33 manual gestures. Although the learning-based method utilized here was previously used within a different application, these results motivate our future work to continue modeling the hand motion based on relevant metrics from each fingerspelling gesture. In addition, our contribution of classifying 33 gestures of the Cyrillic manual alphabet, is also in a larger number of classified gestures than previously reported research on fingerspelling recognition. Future work will include collecting a larger dataset for training and overall development of a real-time interpreting autonomy for the robots to be deployed in public spaces.

[2] http://www.cs.waikato.ac.nz/ ml/milk/

References

Aran, O. 2008. *Vision based sign language recognition: modeling and recognizing isolated signs with manual and non-manual components.* Ph.D. Dissertation, Citeseer.

Bellmore, C.; Ptucha, R.; and Savakis, A. 2011. Interactive display using depth and rgb sensors for face and gesture control. In *Image Processing Workshop (WNYIPW), 2011 IEEE Western New York*, 1–4. IEEE.

Biswas, K. K., and Basu, S. K. 2011. Gesture recognition using microsoft kinect®. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, 100–103. IEEE.

Breuer, P.; Eckes, C.; and Müller, S. 2007. Hand gesture recognition with a novel ir time-of-flight range camera–a pilot study. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, 247–260. Springer.

Chang, Y.-J.; Chen, S.-F.; and 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.

Du, H., and To, T. 2011. Hand gesture recognition using kinect. *Techical Report, Boston University.*

Feris, R.; Turk, M.; Raskar, R.; Tan, K.-H.; and Ohashi, G. 2005. Recognition of isolated fingerspelling gestures using depth edges. In *Real-Time Vision for Human-Computer Interaction*. Springer. 43–56.

Frati, V., and Prattichizzo, D. 2011. Using kinect for hand tracking and rendering in wearable haptics. In *World Haptics Conference (WHC), 2011 IEEE*, 317–321. IEEE.

Fujimura, K., and Liu, X. 2006. Sign recognition using depth image streams. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 381–386. IEEE.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 11(1):10–18.

Hassani, A. Z.; van Dijk, B.; Ludden, G.; and Eertink, H. 2011. Touch versus in-air hand gestures: evaluating the acceptance by seniors of human-robot interaction. In *International Joint Conference on Ambient Intelligence*, 309–313. Springer.

Huenerfauth, M. 2004. A multi-path architecture for machine translation of english text into american sign language animation. In *Proceedings of the Student Research Workshop at HLT-NAACL 2004*, 25–30. Association for Computational Linguistics.

Jastorff, J.; Kourtzi, Z.; and Giese, M. A. 2006. Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision* 6(8):3.

Keskin, C.; Kıraç, F.; Kara, Y. E.; and Akarun, L. 2013. Real time hand pose estimation using depth sensors. In *Consumer Depth Cameras for Computer Vision*. Springer. 119–137.

Kipp, M.; Heloir, A.; and Nguyen, Q. 2011. Sign language avatars: Animation and comprehensibility. In *International Workshop on Intelligent Virtual Agents*, 113–126. Springer.

Klompmaker, F.; Nebe, K.; and Fast, A. 2012. dsensingni: a framework for advanced tangible interaction using a depth camera. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, 217–224. ACM.

Kohonen, T. 1989. *Self-organization and Associative Memory: 3rd Edition*. New York, NY, USA: Springer-Verlag New York, Inc.

Lai, H. 2011. Using commodity visual gesture recognition technology to replace or to augment touch interfaces. In *15th Twente Student Conference on IT*.

Li, Z., and Jarvis, R. 2009. Real time hand gesture recognition using a range camera. In *Australasian Conference on Robotics and Automation*, 21–27.

Liu, X., and Fujimura, K. 2004. Hand gesture recognition using depth data. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 529–534. IEEE.

Luis-Pérez, F. E.; Trujillo-Romero, F.; and Martínez-Velazco, W. 2011. Control of a service robot using the mexican sign language. In *Mexican International Conference on Artificial Intelligence*, 419–430. Springer.

Malassiotis, S.; Aifanti, N.; and Strintzis, M. G. 2002. A gesture recognition system using 3d data. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, 190–193. IEEE.

Marsland, S.; Shapiro, J.; and Nehmzow, U. 2002. A self-organising network that grows when required. *Neural Networks* 15(8-9):1041–1058.

Martinetz, T. M. 1993. Competitive hebbian learning rule forms perfectly topology preserving maps. In *ICANN'93: International Conference on Artificial Neural Networks*, 427–434. Amsterdam: Springer.

McNeill, D. 2000. Language and gesture (vol. 2). *Language, culture, and cognition. Cambridge University Press. Pika, S., Nicoladis, E., & Marentette, PF (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer* 319–327.

Mo, Z., and Neumann, U. 2006. Real-time hand pose recognition using low-resolution depth images. In *CVPR (2)*, 1499–1505.

Parisi, G. I.; Weber, C.; and Wermter, S. 2015. Self-organizing neural integration of pose-motion features for human action recognition. *Frontiers in Neurorobotics* 9(3).

Potter, L. E.; Araullo, J.; and Carter, L. 2013. The leap motion controller: a view on sign language. In *Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, 175–178. ACM.

Ramey, A.; González-Pacheco, V.; and Salichs, M. A. 2011. Integration of a low-cost rgb-d sensor in a social robot for gesture recognition. In *Proceedings of the 6th international conference on Human-robot interaction*, 229–230. ACM.

Ren, Z.; Yuan, J.; and Zhang, Z. 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proceedings of the 19th ACM international conference on Multimedia*, 1093–1096. ACM.

Rolls, E. T., and Caan, W. 1982. Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47:329–342.

Ronchetti, M., and Avancini, M. 2011. Using kinect to emulate an interactive whiteboard. *MS in Computer Science, University of Trento*.

Sahoo, A. K.; Mishra, G. S.; and Ravulakollu, K. K. 2014. Sign language recognition: State of the art. *ARPN Journal of Engineering and Applied Sciences* 9(2):116–134.

Sandygulova, A.; Absattar, Y.; Doszhan, D.; and Parisi, G. I. 2016. Child-centred motion-based age and gender estimation with neural network learning. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*.

Singh, S.; Jain, A.; and Kumar, D. 2012. Recognizing and interpreting sign language gesture for human robot interaction. *International Journal of Computer Applications* 52(11).

Sohn, K.; Kim, Y.; and Oh, P. 2013. 2-tier control of a humanoid robot and use of sign language learned by monte carlo method. In *2013 IEEE RO-MAN*, 547–552. IEEE.

Suarez, J., and Murphy, R. R. 2012. Hand gesture recognition with depth images: A review. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 411–417. IEEE.

Sugiuchi, H.; Morino, T.; and Terauchi, M. 2002. Execution and description of dexterous hand task by using multi-finger dual robot hand system-realization of japanese sign language. In *Intelligent Control, 2002. Proceedings of the 2002 IEEE International Symposium on*, 544–548. IEEE.

Tang, M. 2011. Recognizing hand gestures with microsoft's kinect. *Palo Alto: Department of Electrical Engineering of Stanford University:[sn]*.

Tolba, M., and Elons, A. 2013. Recent developments in sign language recognition systems. In *Computer Engineering & Systems (ICCES), 2013 8th International Conference on*, xxxvi–xlii. IEEE.

Trigueiros, P.; Ribeiro, F.; and Lopes, G. 2012. Vision-based hand segmentation techniques for human-robot interaction for real-time applications.

Uebersax, D.; Gall, J.; Van den Bergh, M.; and Van Gool, L. 2011. Real-time sign language letter and word recognition from depth data. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 383–390. IEEE.

Uluer, P.; Akalın, N.; and Köse, H. 2015. A new robotic platform for sign language tutoring. *International Journal of Social Robotics* 7(5):571–585.

Van den Bergh, M.; Carton, D.; De Nijs, R.; Mitsou, N.; Landsiedel, C.; Kuehnlenz, K.; Wollherr, D.; Van Gool, L.; and Buss, M. 2011. Real-time 3d hand gesture interaction with a robot for understanding directions from humans. In *2011 Ro-Man*, 357–362. IEEE.

Wang, Y.; Yu, T.; Shi, L.; and Li, Z. 2008. Using human body gestures as inputs for gaming via depth analysis. In *2008 IEEE International Conference on Multimedia and Expo*, 993–996. IEEE.

Yang, C.; Jang, Y.; Beh, J.; Han, D.; and Ko, H. 2012. Gesture recognition using depth-based hand tracking for contactless controller application. In *2012 IEEE International Conference on Consumer Electronics, ICCE 2012*.

Yoo, B.; Han, J.-J.; Choi, C.; Yi, K.; Suh, S.; Park, D.; and Kim, C. 2010. 3d user interface combining gaze and hand gestures for large-scale display. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 3709–3714. ACM.