Human-Robot Interaction with Smart Shopping Trolley using Sign Language: Data Collection

Dmitry Ryumin
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
St. Petersburg, Russia
dl_03.03.1991@mail.ru

Denis Ivanko
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
St. Petersburg, Russia
denis.ivanko11@gmail.com

Alexandr Axyonov
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
St. Petersburg, Russia
axyonov@iias.spb.su

Ildar Kagirov
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
St. Petersburg, Russia
kagirov@iias.spb.su

Alexey Karpov
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
St. Petersburg, Russia
karpov@iias.spb.su

Milos Zelezny
University of West Bohemia (UWB)
Pilsen, Czech Republic
zeleny@kky.zcu.cz

***Abstract*—the paper presents a concept of a smart robotic trolley for supermarkets with a multimodal user interface, including sign language and acoustic speech recognition, and equipped with a touchscreen. Considerable progress in hand gesture recognition and automatic speech recognition within the last years has brought to life many human-computer interaction systems. At the moment the level of voiced speech and isolated/ static hand gesture automatic recognition quality is quite high. However, continuous or dynamic sign language recognition still remains an unresolved challenge. There exists no automatic recognition system for Russian sign language nowadays. There are also no relevant data for model training. In the present research, we try to fill in this gap for the Russian sign language. We present a Kinect 2.0 based software-hardware complex for collection of multimodal sign language databases with an optical video camera, infrared camera and depth sensor. We describe the architecture of the developed software as well as some details of the collected database. The collected corpus is meant for further development of a Russian sign language recognition system, which will be embedded into a robotic trolley for supermarkets with gestural and speech interfaces. The architecture of the developed system is also presented in the paper.**

Keywords—sign language recognition, human-robot interaction, dialogue systems, gesture database, smart shopping trolley

I. INTRODUCTION

It is well known that wandering around a supermarket in search of products is not always an easy task for the customer. Sometimes only the search for the right department takes most of the time spent in the store. The development and implementation of a service robot that could quickly lead a user to a department of interest or hint where to look for a particular product, would help save time and energy of the customer. The idea of a robotic mobile platform or a trolley that carries the user's personal belongings is not new. E.g. the main task of the automated trolley EffiBOT [1], developed by French start-up Effidence, is to assist warehouse workers. EffiBOT takes goods and automatically goes with them to the point of discharge. In addition, EffiBOT can follow the user when the corresponding mode is activated. The Dash Robotic Shopping Cart [2] from the American company Five Elements Robotics is conceptually similar to the project described in this paper: a supermarket trolley that facilitates shopping and navigation in the store. The cart is equipped with a touchscreen for entering a list of products of interest to the client. At the command given from the touchscreen, Dash leads the customer from one location to another. Another example of a robotic trolley would be Gita by Piaggio [3] - a robotic trolley that follows the owner. The goal of the project is to make it easier for the user to wander around with luggage.

However, none of the interfaces of the aforementioned robotic carts are multimodal. The main (and often the only) channel for the exchange of information is a touchscreen. The robotic trolley concept proposed in the present paper has the following features: (1) understanding voice commands; (2) understanding Russian sign language commands; (3) escort the user to a certain place in the store; (4) speech synthesis, synthesis of answers in Russian sign language using a 3D avatar.

The general architecture of the developed system is presented in Fig 1. The robotic trolley is able to interact with the user via three different channels: acoustic voice input, sign language input, touch screen input. It contains some hardware and software technologies, which work simultaneously and synchronously. Most important of these modules are: (1) speaker-independent system of automatic continuous Russian speech recognition; (2) speaker-independent system of Russian sign language recognition with video processing using Kinect 2.0 device; (3) an interactive graphical user interface with a touchscreen; (4) a dialogue and data manager that access an application database, generates multimodal output and synchronizes input modalities fusion and output modalities fission; (5) modules for audio-visual speech synthesis to be applied for a talking avatar.

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

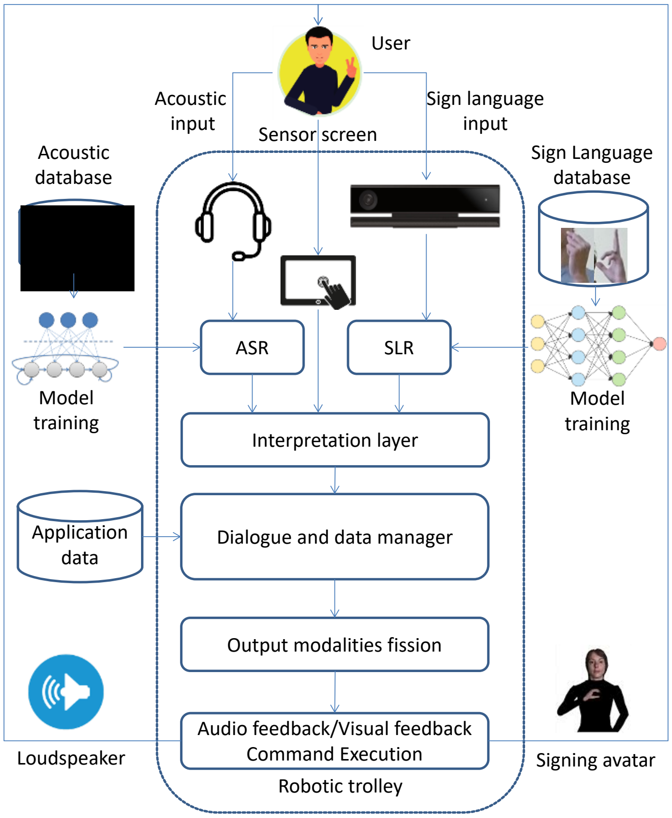

Fig. 1. The general architecture of the robotic trolley

According to World Health Organization statistics [4] over 5% of the world population – or 466 million people – has disabling hearing loss (432 million adults and 34 million children). It is estimated that by 2050 over 900 million people – or one in every ten people – will have disabling hearing loss. For the deaf and speech-impaired community, sign language (SL) serves as useful tools for daily interaction. Official data report about 500 000 people in the USA, 125 000 in the UK, 121 000 in France, 50 000 in Germany, 2 700 000 in India, 3 000 000 in Brazil, 120 000 in Russia, up to 20 000 000 people in China, etc. [5] using sign languages as their main way of communication. In real life, this number may be even greater.

Sign language is a structured form of hand gestures involving visual motions and signs, which is used as a communication system. SL involves the use of different parts of body (mostly fingers, hand and facial expressions) to deliver information [6]. However, no reliable automatic sign language recognition (SLR) system exists at the moment, in comparison with many state-of-the-art audio-based speech recognition systems (ASR) for many of the world's languages.

Multimodal SLR provides those people who cannot use voice input with convenient contactless human-machine interaction (HMI). Nowadays SLR is not as robust as ASR or standard keyboard input. Issues such as sensitivity to size and speed variations, poor performance under varying lighting conditions and complex background have limited the use of SLR in modern dialogue systems [7]. Another big challenge is to create a robust SLR system, taking into account that there is a lack of available representative databases for model training. Such databases are vital for the training of models.

The accuracy and reliability of the ASR and SLR systems of a robotic trolley directly depends on the quality of the training data. The focus of this work is directed towards resolving this task by collecting a representative Russian sign language database needed to create robust SLR system for supermarkets.

This paper presents a description of software architecture meant for recording sign language databases with the use of a Kinect 2.0 device, as well as a representation of the multimodal Thesaurus of Russian Sign Language (TheRusLan) collected using this software. Our main goal is to create an assistive robotic trolley for supermarkets with both audio-visual voice input and sign language input, based on recent advances in SLR investigation, pervasive computing and artificial intelligence. Collected database is meant for creating a robust sign language recognition system for the robotic trolley. Previous stages of research regarding audio-visual continuous Russian speech recognition are given in [8-10], Russian sign language (RSL) recognition described in [11, 12], and robotic trolley concept discussed in [13].

II. Sign Language Databases State-of-the-Art

Research in sign language recognition goes hand by hand with hand gesture recognition research, for sign language is a form of communicative hand gestures [14]. Hand gestures are classified according to temporal relationships into two types: static and dynamic gestures. Most of the "words" of a sign language are sequences of dynamic hand gestures. In dynamic hand gestures, the hand position changes continuously with respect to time. They rely on the hand trajectories, orientations, shape and fingers' flex angles. Fig. 2 shows the block diagram of a typical contactless gesture recognition system.

Nowadays there are various hand gesture databases, which are collected for different purposes and with different means. Standard hand gesture databases are necessary for the reliable testing and comparison of hand gesture recognition algorithms.

The authors of [15] investigated how to develop best representative gesture databases for training machine learning algorithms based on two measures, correctness and coverage. These metrics help to evaluate how good the dataset is in representing real world data. The researchers also are investigating the best way to instruct participants before recording sessions. Their valuable findings helped us in creating our own database of Russian sign language.

A list of 26 publicly available hand gesture datasets is given in review [7]. The most widely used in research work databases are Cambridge hand gesture dataset [16]: the target task for this dataset is to classify hand shapes and motions at the same time; NATOPS aircraft handling signals database [17]: RGB-D database includes 24 body and hand gestures, selected from NATOPS aircraft handling signals; Gesture dataset by Shen et al. [18]: 10 classes of dynamic hand gestures with 7 different hand poses (70 gesture samples per subject); ChaLearn Gesture data and ChaLearn multi-modal gesture data [19]: consists of a total 62000 samples; Sheffield Kinect

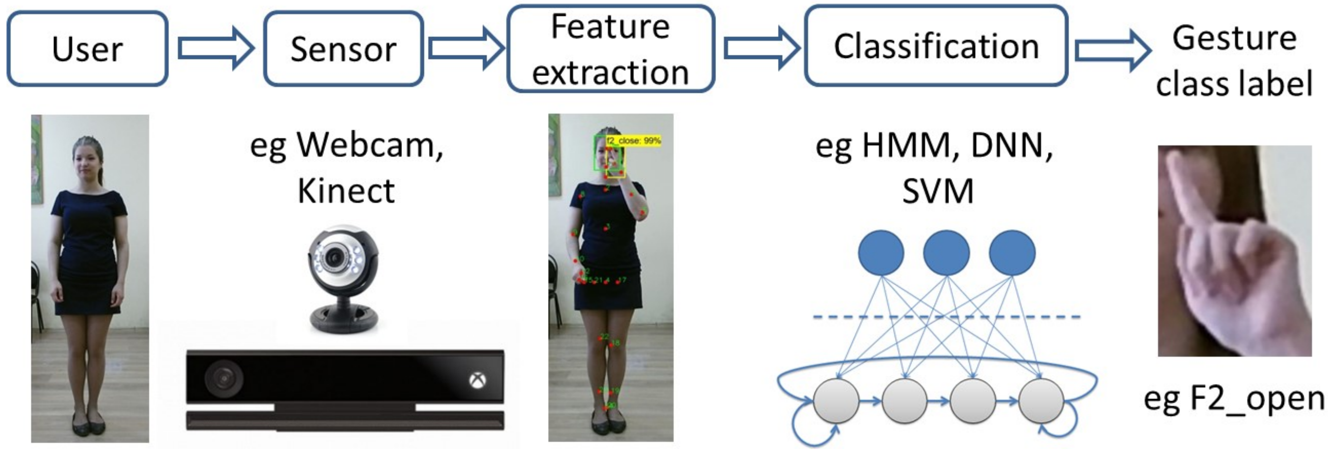

Fig. 2. Block diagram of a typical contactless gesture recognition system

gesture (SKIG) dataset [20]: 6 subjects, 10 categories of gestures, 3 different background conditions; MSRC-12 Kinect gesture dataset [15]: 12 class dynamic gesture dataset; NUS hand posture dataset [21]; NYU hand pose dataset [22]: has 72757 and 8252 numbers of frames in the training and test sets respectively; General-HANDS dataset: dataset contains a variety of 22 sequences, demonstrating different viewpoints, scales, poses, occlusions, and camera technologies; VPU hand gesture dataset [23]: 12 class data, 11 subjects; ASL Finger Spelling Dataset [24]: English letters from *a* to *y*, 14 subjects in total.

Sign language databases are different from hand gesture databases. In work [25] all the existing SL databases are divided into 3 types: (1) lexical data sets; (2) linguistic data sets; (3) large data sets for pattern recognition purposes. Some examples of the most widely used SL databases: the American Sign Language Lexicon Video Dataset [26] forms a lexicon for American sign language, containing more than 3000 signs in multiple video views. Corpus [27] contains 12 hours of signing in upper-body and front view-talking 64,000 annotated glosses. Rutkowskietal [28] created a corpus for Polish sign language containing about 300 h of video footage of 80 deaf signers performing predefined language tasks. The SIGNUM corpus [29] consists of 25 recordings of signers and nearly 14,000 running glosses.

Berman et al. [30] reviewed different sensors used in gesture recognition systems and provided a deep analysis of integration of sensors into gesture recognition systems as well as their impact on the system performance. A modern trend in sign language recognition is the use of RGB-D sensor-based methods. The release of low-cost color-depth camera Kinect [31] by Microsoft has revolutionized gesture recognition by providing high quality depth images. Kinect calculates a 3-d map of the scene using a combination of RGB and infrared camera. The skeletal data from these RGB-D sensors is to be converted to more meaningful features for the robust classification of gestures. The influence of depth information in the gesture recognition process has been evaluated in the works [32, 7]. The authors concluded that use of depth data increases the recognition accuracy significantly.

However, only few databases are available for the Russian sign language. The largest (and only) publicly available corpus is the "Corpus of the Russian sign language" [33], created by the Novosibirsk State Technical University. The corpus contains over 230 texts from 43 native sign language speakers and is provided with translation and annotation. There are also some multimedia dictionaries of RSL, such as "Dictionary of Russian Sign Language - RuSLED" and video dictionary "Thematic Dictionary of Russian Sign Language". These databases are an effective tool for learning purposes. However, all of the abovementioned databases are not suitable for machine learning, because they do not have a proper labeling and recorded under different conditions with only a single repetition of phrases.

Based on the above information, it was decided to record our own database of RSL. For this purpose, a Kinect 2.0 based software-hardware complex with a video camera, infrared camera and depth sensor has been developed.

III. Software Architecture for Recording Sign Language Databases

Han at al. [34] provided a review of how Kinect is useful in addressing fundamental problems in gesture recognition. A typical modern gesture recognition system fuses several visual representations, such as optical, depth and infrared data. Our goal was to develop and implement an efficient software-

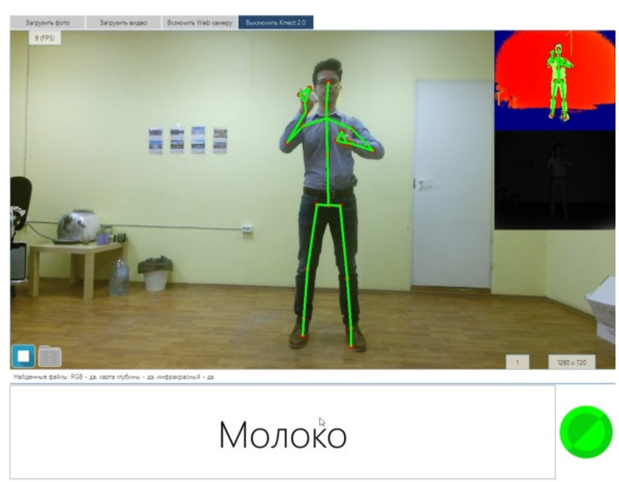

Fig. 3. The graphical user interface of the software for the recording sign language databases

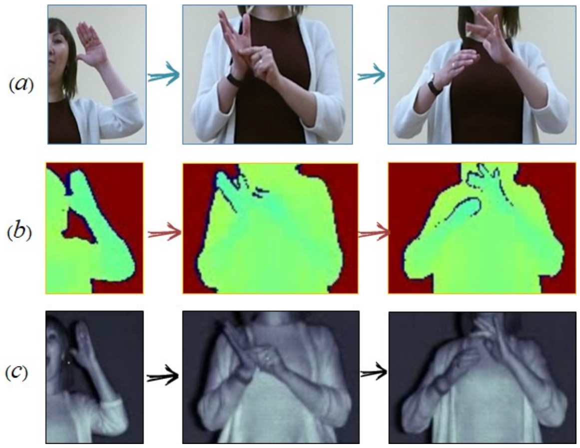

Fig. 4. An example of word “sparkling water” in Russian sign language: color (a), depth (b) and infrared (c) modes

hardware complex that would allow us to collect the sign language corpus for the purposes of SL speech recognition. Following the trend, we have chosen Kinect 2.0 as a recording device [35].

The GUI developed for the recording is presented in Fig. 3. It allows recording sessions in both manual and automatic modes. Phrases that need to be shown by the signer appear at the bottom of the screen (loaded from the dictionary). The number of required repetitions is also regulated by the system. The data obtained from all three sensors (optical, depth sensor and infrared camera) are stored in the database. Optical video format is FullHD (1920×1080), and depth and infrared sensors have the same resolution of 512×424 pixels. An example of how the phrase "sparkling water" looks in all three modalities is shown in Fig. 4.

The frequency of video frames for all sensors is 30 fps. The developed software allows working with a dictionary of any size, as well as embedding subsections for it. The process of creating a dictionary is described in detail in section 4, which focuses on TheRusLan database.

Fig. 5 shows the architecture of software for sign speech database recording. The software complex consists of four main modules:

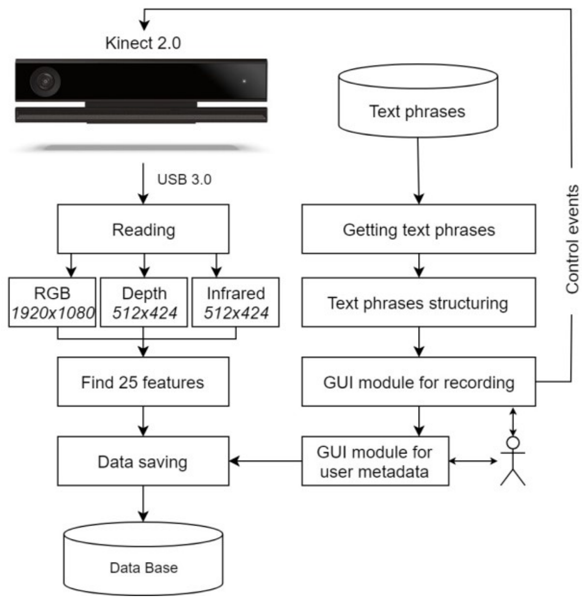

Fig. 5. The architecture of software for sign language database recording

(1) Data capturing module: uses optical, depth and infrared cameras of Kinect 2.0 device for raw data acquisition.

(2) Feature extraction module: saves three-dimensional coordinates of 25 reference dots of the human skeleton model as features.

(3) Text phrases structuring module: gets and sorts phrases from the dictionary, as well as displays them to the user via GUI.

(4) GUI for interaction with a user. The developed software has two GUI modules: for interaction with a user and for receiving his/her metadata.

The GUI modules provide two modes for data recording: (1) manual mode, when the user manages the start and the end points of each phrase; (2) automatic mode, when phrases automatically change after a predetermined time has passed. After the recording phase, all the data of the current speaker are saved into the database.

Fig. 6. Example of signers during a recording session

IV. Corpus description

The developed software was used for recording of "Thesaurus of Russian Sign Language" – TheRusLan. The corpus consists of color optical FullHD (1920×1080) video files, infrared video files (512×424) and depth video files (512×424), it also includes feature files (with 25 skeletal reference dots on each frame), text files of temporal annotation of phrases, words and gesture classes (was made manually by an expert). A brief summary of the contents of the database is presented in Table 1.

13 native Russian sign language speakers (11 females and 2 males) from "Pavlovsk boarding school for the hearing impaired" or "Deaf-Mute school" (the oldest institution for the hearing impaired in Russia) participated in the recordings. Screenshots of the signers in the course of recording sessions are shown in Fig. 6. Each of the signers demonstrated 164 phrases for 5 times.

TABLE I. Contents of TheRusLan database

Parameter	Value
Number of signers	13
Phrases per signer	164
Number of repetitions	5
Number of samples per signer	820
Total number of samples	10 660
Recording device	MS Kinect V2
Resolution (color)	FullHD (1920×1080)
Resolution (depth sensor)	512×424
Resolution (infrared camera)	512×424
Distance to camera	1.0-2 m.
Duration (total)	7 hours 56 min.
Duration (per signer)	~36 min
Number of frames with gestures (per signer)	~30-35 000
Average age of signers	24
Total amount of data	~4 TB

Since the database is intended for the implementation of a Russian dialogue system for a robotic-assistant in supermarkets, a primary list of lexemes and their combinations in the subject area "food and household goods" was compiled to fill the dictionary. The primary list was formed by simple export of text files from the navigation menu of Internet sites of a number of large Russian hypermarkets. The resulting list of words was formed by screening out units containing specific names (brand name of the product, manufacturer). In addition, the resulting list does not include products that are not very popular among Russian consumers, according to the personal experience of the authors of this work (for example, "tiger shrimps" or "pomelo"). Finally, lexical units, for which fingerspelling is used due to the lack of generally accepted gestures, were excluded from the dictionary.

Command subsection of the dictionary includes vocabulary related to special orientation ("forward", "backward", "right", "left"), series of movement verbs ("go", "let's go") and their modifiers ("faster", "slower"), location requests ("where is…?", "show me…"), products requests ("I need…", "I want…").

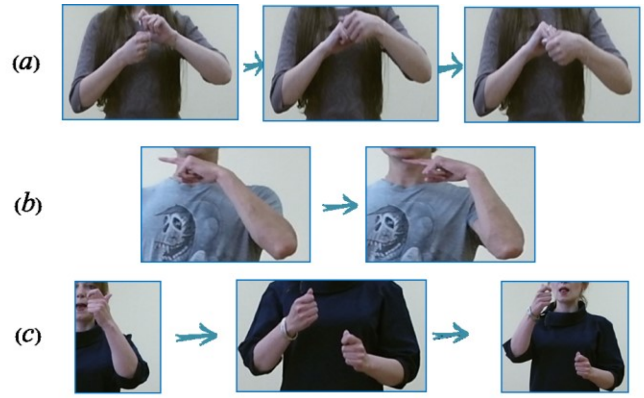

Fig. 7. Examples of one-handed and two-handed gestures: orange (a), meat (b), milk (c)

A brief summary of the subsections of the dictionary and phrases examples is presented in Table 2.

For the sign language interpretation of phrases, a sequence of hand gestures is used. They can be one-handed or two-handed, as shown in Fig. 7. In addition, hand gestures are characterized by such features as shape, location, trajectory of movement. Therefore, while the dictionary includes 164 phrases, the number of recognizable hand gestures in it is much larger. 44 unique items for one-handed gestures and 211 items for two-handed gestures were discovered in the TheRusLan. For them, time and frame labeling was done including timestamps, 3-d coordinates and gesture labels.

TABLE II. Dictionary subsections and phrases examples.

	Subsections	Examples
1	Commands	Stop. Follow me. Faster. Slower.
2	Requests	Show me X. I want X. Guide me to X.
3	Answers	Goodbye. Yes. No. How can I help you.
4	Dairy produce	Milk. Cheese. Eggs. Butter. Sour cream
5	Confectionary	Cake. Pie. Cookie. Biscuit.
6	Bakery	Bread. Long loaf.
7	Beverages	Water. Sparkling water. Juice. Cold tea.
8	Alcohol	Beer. Wine. Vodka. Cognac.
9	Fruits and vegetables	Apples, Bananas, Tomatoes, Cucumbers.
10	Grocery	Tea. Coffee. Pasta. Honey.
11	Cereals	Rice. Oatmeal. Grain. Beans. Buckwheat.
12	Canned Goods	Tinned Goods. Pickles. Pickled tomatoes.
13	Spice	Salt. Sugar. Pepper.
14	Meat	Chicken. Beef. Pork. Sausages.
15	Fish and seafood	Frozen fish. Live fish. Shrimps. Mussels.
16	Oils & Sauces	Sunflower oil. Olive oil. Vinegar.
17	Frozen foods	Dumplings. Pizza. Vegetable mix.
18	Chocolate	Chocolate. Candies. Waffles.
19	Goods at the checkout	Gum. Batteries. Lighter.
20	Departments and locations	Exit. Restroom. Detergents. Toys. Books.

All the recordings were organized into a logically structured database that comprises a file with information about all the speakers and recording parameters.

V. Conclusions and Future Work

In this paper we presented the concept of a smart robotic trolley for supermarkets with a multimodal user interface, including sign language and acoustic speech recognition, and equipped with a touchscreen, as well as developed software framework for collection of sign language databases using Kinect 2.0 device. Currently, collected corpus TheRusLan comprises recordings of 13 native signers of the Russian sign language. Each signer demonstrated 164 phrases for 5 times. The total number of samples in the corpus is 10660. TheRusLan is meant for further research and experiments on Russian sign language recognition system, which will be embedded into the dialogue system of the smart shopping trolley.

Acknowledgment

This research is financially supported by the Ministry of Education and Science of the Russian Federation, agreement No. 14.616.21.0095 (reference RFMEFI61618X0095) and by the Ministry of Education of the Czech Republic, project No. LTARF18017.

References

[1] Découvrez EffiBOT, votre assistant logistique tout au long de l'année. Retreived November 26, 2018, from : https://www.effidence.com/effibot

[2] DASH Retail Robotic Shopping Cart. Retreived November 26, 2018, from : http://5elementsrobotics.com/dash-retail-robotic-shopping-cart/

[3] Introducing Gita. Retreived November 26, 2018, from : https://www.piaggiofastforward.com/gita

[4] WHO deafness and hearing loss statistic. Retreived November 26, 2018, from : http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[5] List of sign languages by number of native signers. Retreived November 26, 2018, from:: https://en.wikipedia.org/wiki /List_of_ sign_languages_by_number_of_native_signers

[6] M. T. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," International Journal of Machine Learning and Cydernetics, pp. 1-23, 2017

[7] P. K. Pisharady, M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," in Computer Vision and Image Understanding, vol. 14, pp. 152-165, 2015

[8] D. Ivanko, A. Karpov, D. Fedotov, I. Kipyatkova, D. Ryumin, Dm. Ivanko, W. Minker, M. Zelezny, "Multimodal speech recognition: increasing accuracy using high speed video data," Journal of Multimodal User Interfaces, vol. 12, no. 4, pp. 319-328, 2018

[9] L. Besacier, E. Barnard, A. Karpov, and T. Shultz, "Automatic speech recognition for under-resourced languages: A survey," Speech Communication, vol. 56, pp. 85-100, 2014

[10] V. Verkhodanova, A. Ronzhin, I. Kipyatkova, D. Ivanko, A. Karpov, and M. Zelezny, "HAVRUS corpus: high-speed recordings of audio-visual Russian speech," In Proc. Int. Conf. on Speech and Computer SPECOM-2016, Springer, LNCS, vol. 9811, pp. 338-345, 2016

[11] D. Ryumin, and A. Karpov, "Towards automatic recognition of sign language gestures using kinect 2.0," In Proc. 19th Int. Conf. on Human-Computer Interaction HCII-2017, Vancouver, Canada, Springer LNCS vol. 10278, pp. 89-104, 2017

[12] I. Gruber, D. Ryumin, M. Hruz, and A. Karpov, "Sign language numeral gestures recognition using convolutional neural network," In Proc. 3rd Int. Conf. on Interactive Collaborative Robotics ICR-2018, Leipzig, Germany, Springer, LNAI vol. 11097, pp. 70-77, 2018

[13] I. Kagirov, "Basic concept of a robotic shopping cart enabling speech and Russian hand gesture recognition," in Proc. Regional Informatics and Information security, issue 5, pp. 273-277, 2018

[14] Y. Wu, and T. S. Huang, "Human hand modelling, analysis and animation in the context of HCI," in Image Processing, ICIP, Proc. Int. Conf. IEEE, pp. 6-10, 1999

[15] F. Simon, M. M. Helena, K. Pushmeet, and N. Sebastian, "Instructing people for training gestural interactive systems," Proc. of International Conference on Human Factors in Computing Systems, CHI, ACM, pp. 1737-1746, 2012

[16] T. K. Kim, S. F. Wong, and R. Cipolla, "Tensor canonical correlation analysis for action classification," Proc. IEEE Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2007

[17] S. Yale, D. David, and D. Randall, "Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database," in Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 500-506, 2011

[18] X. H. Shen, G. Hua, L. Williams, and Y. Wu, "Dynamic hand gesture recognition: an exemplar-based approach from motion divergence fields," Image Vis.Comput., vol. 30, no. 3, pp. 227-235 (2012)

[19] U. Mahbub, H. Impiaz, T. Roy, M. Rahman, and M. Ahad, "A template matching approach of one-shot learning gesture recognition," Pattern Recognition Lett. 34, pp.1780-1788, 2013

[20] L. Liu, and L. Shao, "Learning discriminative representations from RGB-D video data," Proc. of International Joint Conference on Artificial Intelligence, 2013

[21] P. K. Pisharady, P. Vadakkepat, and A. P. Loh, "Hand posture and face recognition using a fuzzy-rough approach," International Journal Humanoid Robot, vol. 7, no. 3, pp. 331-356, 2010

[22] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," ACM Trans. Graph., vol. 33, pp. 169:01 – 169:10, 2014

[23] E. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," International Journal Intell. Syst. Technol. Appl., vol. 5, pp. 334-343, 2007

[24] N. Pugeault, and R. Bowden, "Spelling it out: real time ASL fingerspelling recognition," In: Proc. of International Conference on Computer Vision, ICCV, 2012

[25] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," in Journal on Computer Vision and Image Understanding, vol. 141, pp. 108-125, 2015

[26] C. Neidle, and C. Vogler, "A new web interface to facilitate access to corpora: development of the ASLLRP data access interface," in proc. of the international conference on Language Resources and Evaluation, ELRA, pp. 137-142, 2012

[27] O. Crasborn, and I. Zwitserlood, "The corpus NGT: an online corpus for professional and laymen," in: Construction and Exploitation of Sign Language Corpora," 3rd workshop on the representation and processing of sign languages, ELDA, Paris, pp. 44-49, 2008

[28] P. Rutkowski, S. Lozinska, J. Fillipczak, J. Lacheta, and P. Mostowski, "Polish sign language (PJM) corpus: assumptions procedures methodology," M. Sak (Ed.) deafness studies condition in Poland, 2013

[29] U. von Agris, M. Knorr, and K.F. Kraiss, "The significance of facial features for automatic sign language recognition," 8th IEEE International Conference on Automatic Face&Gesture Recognition, pp. 1-6, 2008

[30] S. Berman, and H. Stern, "Sensors for gesture recognition systems," IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 42 (3), pp. 277-290, 2012

[31] Z. Zhang. "Microsoft kinect sensor and its effect," IEEE MultiMed. 19 (2), pp. 4-10, 2012

[32] R. Munoz-Salinas, R. Medina-Carnicer, F. J. Madrid-Cuevas, and A. Carmona-Poyato, "Depth silhouettes for gesture recognition," Pattern Recognition, vol. 29, no. 3, pp.319-329, 2008

[33] Retreived November 26, 2018, from : http://rsl.nstu.ru/

[34] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: a review," IEEE Trans. Cybern., vol. 43, no. 5, pp. 1318-1334, 2013

[35] F. Domino, M. Donadeo, and P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition," Pattern Recognition, vol. 50, pp. 101-111, 2014