Towards Digitally-Mediated Sign Language Communication

Kalin Stefanov
University of Southern California
Los Angeles, USA
kstefanov@ict.usc.edu

Mayumi Bono
National Institute of Informatics
Tokyo, Japan
bono@nii.ac.jp

ABSTRACT

This paper presents our efforts towards building an architecture for digitally-mediated sign language communication. The architecture is based on a client-server model and enables a near real-time recognition of sign language signs on a mobile device. The paper describes the two main components of the architecture, a recognition engine (server-side) and a mobile application (client-side), and outlines directions for future work.

CCS CONCEPTS

• **Human-centered computing** → Natural language interfaces; Gestural input; • **Computing methodologies** → Supervised learning by classification; Cross-validation; • **Computer systems organization** → Client-server architectures.

KEYWORDS

sign language; key word signing; recognition;

ACM Reference Format:
Kalin Stefanov and Mayumi Bono. 2019. Towards Digitally-Mediated Sign Language Communication. In *Proceedings of the 7th International Conference on Human-Agent Interaction (HAI ’19), October 6–10, 2019, Kyoto, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3349537.3352794

1 INTRODUCTION

Sign languages and other forms of sign-based communication are important to large groups in society. In addition to members of the deaf community, that often have a sign language as their first language, there is a large group of people who use verbal communication, but rely on signing as a complement. Sign languages include thousands of signs that differ from each other only by minor changes in the hand shape and motion profile. Signing consists of either, manual components that are gestures involving the hands, where hand shape and motion convey the meaning, or finger spelling, used to spell out words. Non-manual components, like facial expressions and body posture, can also provide information during signing.

Collecting and annotating an extensive and diverse sign language dataset is an essential step in the process of building an automated sign language recognition system. Several projects are creating such datasets, for example, the DGS-Korpus dictionary project [12] in Germany, the BSL Corpus Project [11] in the UK, the SSL Corpus Project [13] in Sweden, the American Sign Language Linguistic Research Project [10] in the US, and the Corpus Project in Colloquial Japanese Sign Language [7] in Japan.

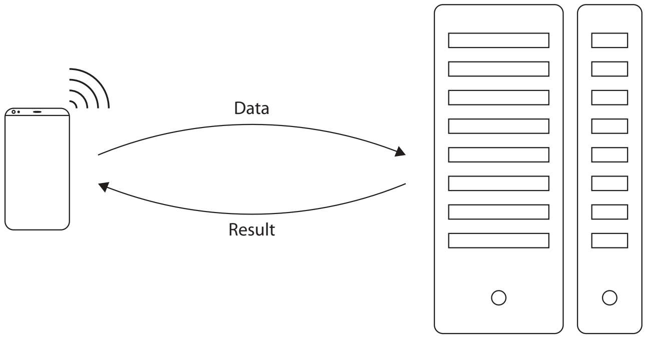

Figure 1: Client-server architecture for recognition of sign language signs on a mobile device.

Automatic sign language recognition inherits some of the difficulties of automatic speech recognition. For example, co-articulation between signs, meaning that a sign will be modified by the signs on either side of it, and large differences between signers - signer-specific styles (pronunciation in speech), both contribute to increased variation in signing. Dicta-Sign [6] and SIGNSPEAK [15] are examples of European Union projects aiming at recognition, generation and modeling of sign languages. A comprehensive review of the research on sign language recognition and the main challenges is provided in [5].

Our previous work on recognition of sign language signs was carried out in the context of the TIVOLI project [1, 2]. TIVOLI is a multimodal game and training application for Swedish key word signing, targeted at children with communicative disorders. The game has a built-in real-time recognizer, that attempts to model and recognize manual components of Swedish Sign Language signs, allowing the children to interact with the game through signing. As mentioned previously, manual components of signing are in general hand shapes/orientations and movement trajectories which are similar to gestures. A comprehensive survey on gesture recognition was performed in [8, 14]. The applicability of the TIVOLI game as a learning tool has been tested on a group of 38 children (ages 10-11) with no prior sign language skills. The result showed that computer games that employ signing as an interaction medium are successful learning environment that can support the acquisition of sign language skills [9].

This paper builds upon our previous effort on recognition of sign language signs in three directions: 1) we remove the depth sensor dependency from the recognizer, 2) we introduce an architecture for recognition on a mobile device, and 3) we explore the support for recognition of Japanese Sign Language signs.

HAI ’19, October 6–10, 2019, Kyoto, Japan

ACM ISBN 978-1-4503-6922-0/19/10.
https://doi.org/10.1145/3349537.3352794

2 SYSTEM

The developed architecture for digitally-mediated sign language communication is illustrated in Figure 1. The architecture is based on a client-server model and enables a near real-time recognition of sign language signs on a mobile device. The architecture has two main components, a mobile application (client-side) and a recognition engine (server-side). In the next subsections we describe the two components in more details.

2.1 Mobile Application

The mobile application (Android) has three main functions: 1) to acquire video data from the back smartphone camera, 2) to establish connection with the recognition engine (Section 2.2), and 3) to send the video data to the recognition engine and display the returned result. Future work includes extending this last functionality with text-to-speech, i.e., the result returned from the recognition engine is also spoken while it is displayed as text within the mobile application.

2.2 Recognition Engine

The recognition engine is deployed on a dedicated machine that performs all CPU- and GPU-intensive calculations. The recognition engine consists of several components: 1) a feature extractor, 2) an optimizer for building sign models, 3) a real-time recognizer, and 4) a parser for the recognition results. Additionally, it includes previously built sign models, routines for communication with the mobile application (Section 2.1), and routines for data management.

The feature extractor uses the video data sent from the mobile application and comprises routines for extracting and normalizing $2\mathbb{D}$ joint positions detected with OpenPose [4]. Visualization of the joints used during feature extraction is shown in Figure 2.

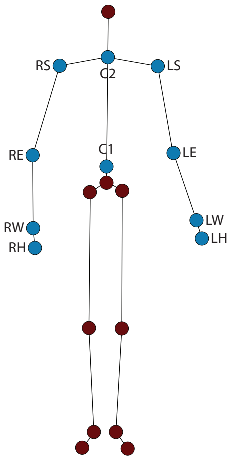

Figure 2: Joints used during feature extraction.

The optimizer and the real-time recognizer are based on Hidden Markov Models. For the the specific problem of recognition of sign language signs we modified the speech recognition tools from the Hidden Markov Model Toolkit [19] to operate on spatial data (the features extracted from the $2\mathbb{D}$ joint positions). The previously built sign models are based on our RGB-D dataset [16] of Swedish Sign Language signs. Figure 3 illustrates the topology of the developed models. On 51-class (sign) recognition problem and evaluated with leave-one-out cross-validation procedure, the recognizer reaches average recognition rate of ~90% and ~65% in signer-dependent and signer-independent settings, respectively [17, 18]. The role of the parser is to convert the raw recognition results into text (the sign meaning) which is sent back to the mobile application.

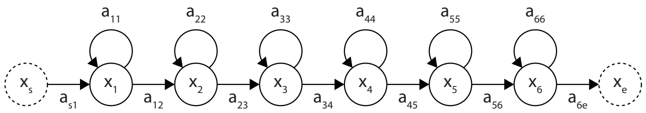

Figure 3: 8-state left-to-right HMM developed for modeling and recognition of Swedish Sign Language signs.

3 FUTURE WORK

There are numerous directions for future work. We plan to extend the feature extractor with a color-based hand tracker. The hand tracker is based on adaptive skin color modeling of the current signer by taking the face color as reference. We expect that the hand tracker will increase the tracking robustness when combined with the joint tracker. Furthermore, through the hand tracker we can introduce hand shape features in the modeling pipeline.

On the modeling side, we plan to investigate more advanced modeling methods for the sign models, for example Recurrent Neural Networks, in order to improve the overall recognition accuracy. Further recognition improvements are expected by introducing model adaptation step, based on a small set of signs from the target signer that could be collected during an enrollment/training phase.

We are working on the annotation and preparation of the largest to date dataset of Japanese Sign Language. The dataset consists of ~35 hours of video recorded with three high-definition cameras. The quality of the recordings is further enhanced by the careful arrangement of the recording environment. The participants in the dataset are seated in blue chairs, illuminated with four lighting devices and the recording space is covered with blue panels. Three approaches were used to collect data: interviews (for introductory purposes), dialogues, and lexical elicitation [3].

The underlying need for this research is to read and understand sign language signs. An example application is a person that does not know sign language and needs to communicate with a deaf person. The benefits of being able to engage in natural interaction with deaf or hard-of-hearing person on a daily basis are difficult to quantify. It is very much about a higher quality of life.

ACKNOWLEDGMENTS

We would like to thank A. Hedkrans, C. Dath, G. Schneider, J. Friedrich, J. Lilja, M. Östling, M. Nilsson, O. Florin, R. Berggren, T. Nielsen, and T. Möre for their work on the mobile application.

This work was supported in part by the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

[1] J. Beskow, S. Alexanderson, K. Stefanov, B. Claesson, S. Derbring, and M. Fredriksson. 2013. The Tivoli System - A Sign-driven Game for Children with Communicative Disorders. In *Proceedings of the 1st Symposium on Multimodal Communication*.

[2] J. Beskow, S. Alexanderson, K. Stefanov, B. Claesson, S. Derbring, M. Fredriksson, J. Starck, and E. Axelsson. 2014. Tivoli - Learning Signs Through Games and Interaction for Children with Communicative Disorders. In *Proceedings of the 6th Biennial Conference of the International Society for Augmentative and Alternative Communication*.

[3] M. Bono, K. Kikuchi, P. Cibulka, and Y. Osugi. 2014. A Colloquial Corpus of Japanese Sign Language: Linguistic Resources for Observing Sign Language Conversations. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. 1898–1904.

[4] Zhe C., Gines H., Tomas S., Shih-En W., and Yaser S. 2018. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

[5] H. Cooper, B. Holt, and R. Bowden. 2011. *Sign Language Recognition*. Springer London, 539–562.

[6] Dicta-Sign. 2012. http://www.dictasign.eu.

[7] Corpus Project in Colloquial Japanese Sign Language. 2011. http://research.nii.ac.jp/jsl-corpus/public/en/index.html.

[8] S. Mitra and T. Acharya. 2007. Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics* 37, 3 (2007), 311–324.

[9] D. Potrus. 2017. *Swedish Sign Language Skills Training and Assessment*. Master's thesis. KTH Royal Institute of Technology.

[10] American Sign Language Linguistic Research Project. 2006. http://www.bu.edu/asllrp/.

[11] BSL Corpus Project. 2010. http://www.bslcorpusproject.org/.

[12] DGS-Korpus Dictionary Project. 2010. http://www.sign-lang.uni-hamburg.de/dgs-korpus/.

[13] SSL Corpus Project. 2009. http://www.ling.su.se/english/research/research-projects/sign-language.

[14] S. S. Rautaray and A. Agrawal. 2015. Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey. *Artificial Intelligence Review* 43, 1 (2015), 1–54.

[15] SIGNSPEAK. 2012. http://www.signspeak.eu.

[16] K. Stefanov and J. Beskow. 2013. A Kinect Corpus of Swedish Sign Language Signs. In *Proceedings of the IVA Workshop Multimodal Corpora: Beyond Audio and Video*.

[17] K. Stefanov and J. Beskow. 2016. Gesture Recognition System for Isolated Sign Language Signs. In *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication*. 57–59.

[18] K. Stefanov and J. Beskow. 2017. A Real-Time Gesture Recognition System for Isolated Swedish Sign Language Signs. In *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication*. 18–27.

[19] S. Young. 1994. The HTK Hidden Markov Model Toolkit: Design and Philosophy. *Entropic Cambridge Research Laboratory* 2 (1994), 2–44.