SLRTP 2020: The Sign Language Recognition, Translation & Production Workshop

Necati Cihan Camgöz[1], Gül Varol[2], Samuel Albanie[2], Neil Fox[3],
Richard Bowden[1], Andrew Zisserman[2], and Kearsy Cormier[3]

[1] CVSSP, University of Surrey,
{n.camgoz, r.bowden}@surrey.ac.uk
[2] Visual Geometry Group, University of Oxford,
{gul, albanie, az}@robots.ox.ac.uk
[3] Deafness, Cognition and Language Research Centre, University College London,
{neil.fox, k.cormier}@ucl.ac.uk
https://www.slrtp.com/

Abstract. The objective of the "Sign Language Recognition, Translation & Production" (SLRTP 2020) Workshop was to bring together researchers who focus on the various aspects of sign language understanding using tools from computer vision and linguistics. The workshop sought to promote a greater linguistic and historical understanding of sign languages within the computer vision community, to foster new collaborations and to identify the most pressing challenges for the field going forwards. The workshop was held in conjunction with the European Conference on Computer Vision (ECCV), 2020.

1 Introduction

In recent years, there has been considerable interest in tasks at the intersection of visual and linguistic modelling, motivated by progress on tasks such as visual dialogue, visual question answering and image captioning. As spatio-temporal linguistic constructs, sign languages represent unique challenges at the intersection of language and vision. For the last three decades, computer vision researchers have been studying sign languages in isolated recognition scenarios. However, large-scale continuous corpora are becoming increasingly available and the focus of the research community is transitioning towards continuous sign language recognition. Sign language translation and production in particular, present themselves as new frontiers that can be approached with modern techniques developed in the context of neural machine translation and generative modelling. In the SLRTP 2020 workshop, we aimed to bring together researchers to discuss the open challenges that lie at the intersection of sign language and computer vision. This report describes the themes covered by the event, statistics associated with workshop submissions and future directions.

2 Themes

The workshop covered several core themes through a series of invited keynote talks, which we describe next.

Processing Sign Languages: Linguistic, Technological, and Cultural Challenges. In this invited talk, Prof. Bencie Woll addressed three types of challenge: linguistic, technological and cultural – to researchers working on automated processing of sign languages. The talk offered a brief review of the typological properties of sign language structure, with emphasis on how they exploit the affordances provided by the use of articulators including the hands, upper body and face, and the properties of human visual perception. Technological challenges include the limited availability of tagged and annotated sign language corpora and researchers' lack of sign language awareness and skills. The most crucial challenge, however is cultural. There is little engagement with deaf communities, little attempt to find out whether proposed technology – often described as designed to help deaf people communicate – is what deaf people want and need. True commitment to accessibility involves consideration of all these factors, as well as long-term engagement with creating systemic change. To make progress, better partnerships between sign language linguists and software engineers is required. This includes support for the amount of work required to prepare comprehensive tagging and annotation of corpora, and inclusion within project teams of fluent signers (especially encouraging and supporting members of deaf communities and deaf scholars from diverse backgrounds to develop careers in technology). Most important of all is knowledge exchange with communities during the development of research.

Sign Language Recognition: From Dispersed to Comparable Research To identify the requirements for future research, a comprehensive survey of the existing state of the art in sign language recognition is necessary. In this invited talk, Oscar Koller gave an overview of the field, focusing on the move from disparate research to the current momentum it has gained. The talk looked into comparable research studies on the available benchmark data sets and analysed the statistics of popular sign language tasks to understand what is needed to continue on the field's accelerated journey to real accessibility [10]. Finally, it concluded with an investigation of how their work [11] helps to deal with the specific challenges present in sign language recognition.

Sign Language Technologies: What are We Hoping to Accomplish? This invited talk by Prof. Christian Vogler discussed elements of the negative perception of sign language recognition technologies in the deaf community and some of the history of how this perception has developed. The talk further provided an analysis of the challenges with the current state of the field [3], and what can be done to improve matters. It highlighted that collaboration with the deaf front and center is key, as is identifying realistic applications that people will want to use, based on inclusive principles that respect the community.

Creating Useful Applications with Imperfect, Sign-Language Technologies. Creating sign-language recognition and synthesis technologies is difficult, and state-of-the-art systems are still imperfect. This limitation presents a challenge for researchers in seeking resources to support dataset creation, user requirements gathering, and other critical infrastructure for the field. This invited talk by Prof. Matt Huenerfauth examined how it is possible to create useful applications in the near-term, to motivate research that would have long-term benefit to the field. Examples of funded projects that integrate imperfect sign-language technologies were discussed, including: providing automatic feedback for students learning American Sign Language (ASL) through analysis of videos of their signing, creating search-by-video interfaces for ASL dictionaries, generating understandable ASL animations to improve information access, and providing ASL content in reading-assistance software. The common thread is that the technologies at the core of each project (i.e. human animation synthesis or recognition of video of human motion) are all imperfect artificial-intelligence systems that occasionally fail in non-human-like ways. The talk discussed investigations of how to adapt these imperfect technologies for new domains, and using human-computer interaction research methods to evaluate alternative system designs. The goal is to enable users to cope with current limitations of these intelligent technologies so that they benefit from applications that employ them.

Turkish Sign Language Recognition at Boğaziçi University. In this invited talk, Prof. Lale Akarun describes work conducted at the Boğaziçi University Sign Language Group, which includes researchers from the domains of computer vision, speech and language processing, and linguistics. In the past, they have carried out projects on applications of sign language recognition, such as automated sign tutoring, and information kiosk for the Deaf in hospitals [5]. These applications involve sign verification and limited vocabulary isolated sign language recognition tasks. They have collected a dataset called BosphorusSign, which is open for researchers [17]. Their current work aims to find better visual embeddings that can generalize across different sign languages. They have shown that the embedding learnt with multitask learning, improves the performance of sign language translation [16]. In their latest work, they investigate unsupervised methodologies for finding hand shapes [23] and for sign unit discovery [19].

3 Programme and submissions

The SLRTP 2020 workshop received 25 high quality submissions comprising 13 full papers and 12 extended abstracts. Of these, 18 papers were accepted, of which 10 were full papers and 8 were extended abstracts.

The work of Bull *et al.* [4] introduced the problem of automatic segmentation of sign language into *Subtitle-Units* and provided a baseline for this task—such segmentations have direct application for translation and efficient subtitling of sign language content. The modelling of phonologically-meaningful subunits for sign language recognition was investigated by Borg *et al.* [2]: this provides not only a strong basis for recognition with deep learning-based approaches, but also improves the interpretability of the system. Motivated by the important role that facial expressions play in sign languages, da Silva *et al.* [22] develop a system that aims to perform FACS-based [7] action unit classification. Their approach is applied to a collected dataset of Brazilian Sign Language, Libras. An efficient system of sign language detection based on human pose estimation is presented by Moryossef *et al.* [15], who demonstrate its potential for video-conferencing applications. In [18], Parelli *et al.* investigate the use of 3D hand pose estimation, and show that it can be a valuable cue for sign language recognition. A plan for constructing an Auslan communication technologies pipeline encompassing sign recognition, production and natural sign language processing is proposed by Korte *et al.* [12]. Medical applications of automatic recognition are explored by Liang *et al.* [14], who develop a multi-modal toolkit to detect early stages of Dementia among British Sign Language users. The work of Gokce *et al.* [8] investigates multi-cue fusion and shows its effectiveness for improving sign language recognition. Polat *et al.* [19] consider instead the task of unsupervised sign discovery without labels using a k-nearest neighbours approach. The use of hand shape features for improving keyword search performance is investigated by Tamer *et al.* [24].

In addition to the full papers discussed above, the extended abstracts presented at the workshop explored a range of themes related to sign language recognition and production.

Belissen *et al.* [1] investigate the necessity and realizability of recognizing linguistic structures of sign languages, like classifiers, using natural corpora. Yin *et al.* [26] use popular transformer networks to improve the sign language translation performance. Duarte *et al.* [6] give a brief introduction of the newly curated How2Sign dataset, which is an extension of the large scale multi-modal How2 dataset [20]. Using the How2Sign dataset, Ventura *et al.* [25] explore continuous sign language video production conditioned on skeletal pose sequences. Kratimenos *et al.* [13] use state-of-the-art 3D pose estimation techniques to obtain parametric representations of signers and report improved multi-channel sign language recognition performance over using raw RGB images. Wizard-of-Oz experiments are conducted by Hassan *et al.* [9] to investigate the user satisfaction of sign language recognition systems. An iterative visual attention model is proposed by Shi *et al.* for fingerspelling sequence recognition in the wild [21]. Glasser *et al.* investigate sign language user interfaces, identify open questions and challenges including the Deaf and Hard of Hearing communities' interest in such technologies.

4 Practical/logistical findings and recommendations

To meet the workshop objective of bringing sign language researchers from different communities together, we aimed to make the content and workshop discussion accessible to a broad audience. To this end, each submitted paper was accompanied by a short video describing the work, which was then captioned and translated into British Sign Language (BSL) and American Sign Language (ASL) by overlaid interpreters (or captioned with written English directly from sign language, where appropriate). Similarly, each invited talk was captioned and translated into ASL and BSL, or translated from ASL into written English. All discussions were translated live into both BSL and ASL.

Findings. Due to ongoing global health concerns, the workshop was held virtually via video conferencing software. This presented additional complexity in coordinating interactions, but also provided opportunities to improve accessibility by allowing recruitment of skilled interpreters from a global workforce without geographic constraints. This was particularly beneficial given the highly technical nature of the material covered in live interactions. Video conferencing software also had the additional benefit of allowing attendees to continue conversations with presenters through the chat functionality, even as other presentations continued (typically infeasible in a physical workshop).

Recommendations. Monolingual workshop organisation at computer vision conferences requires a considerable logistical effort: coordinating call-for-papers, submissions, reviews, paper decisions, sponsorship and the running of the workshop day itself. Provision of multi-lingual content requires additional planning: presentations must be sent to interpreters before the workshop to provide them with time to review the material and produce a translation. For live dialogue, interpreters must be sought who can attend the workshop and who feel comfortable with translating technical content. Finally, communication in multiple languages progresses more slowly than in one—the schedule of the workshop itself should be adjusted to reflect this. Our central recommendations are two-fold: (1) to start planning as early as possible—several months of work were required to coordinate SLRTP and there was still considerable time pressure at all stages of the process, (2) *dry-runs*—live interpretation through video conferencing adds complexity and the workshop chairs benefited from rehearsals of transitions between presentations, ensuring interpreters are visible at all times to ensure that content remains accessible.

5 Conclusion

The SLRTP 2020 workshop brought together researchers who work on various aspects of sign language understanding spanning techniques from linguistics to computer vision. In addition to providing a platform for a range of technical contributions, there were several key takeaways from the workshop. First, it is crucial that deaf communities and researchers are present at every stage of research projects and workshop/conference organisations about sign languages. Second, the focus of applied sign language research should be realistic applications that people will want to use, rather than those with no practical need. Third, significant further efforts are required in dataset collection if the research community is to benefit from recent advances in neural machine translation.

Acknowledgements

We would like to thank our sponsors Microsoft (AI for Accessibility Program), Google, and EPSRC project "ExTOL" (EP/R03298X/1). We would also like to thank Ben Saunders, Bencie Woll, Liliane Momeni, Oscar Koller, and Sajida Chaudhary for their help and advice, Robert Adam for ASL and BSL translations, Akbar Sikder and Esther Rose Bevan for BSL Interpretations, Anna Michaels and Brett Best from Arrow Interpreting for ASL Interpretations, and Katy Ryder and Tara Meyer from MyClearText for the live captioning.

References

1. Belissen, V., Braffort, A., Gouiffés, M.: Towards Continuous Recognition of Illustrative and Spatial Structures in Sign Language. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
2. Borg, M., Camilleri, K.P.: Phonologically-meaningful Subunits for Deep Learning-based Sign Language Recognition. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
3. Bragg, D., Koller, O., Bellard, M., Berke, L., Boudrealt, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., Morris, M.R.: Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In: Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS) (2019) 2
4. Bull, H., Gouiffès, M., Braffort, A.: Automatic Segmentation of Sign Language into Subtitle-Units. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 3
5. Camgöz, N.C., Kındıroğlu, A.A., Akarun, L.: Sign language recognition for assisting the deaf in hospitals. In: International Workshop on Human Behavior Understanding. pp. 89–101. Springer (2016) 3
6. Duarte, A., Palaskar, S., Ghadiyaram, D., De Haan, K., Metze, F., Torres, J., i Nieto, X.G.: How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
7. Ekman, P., Friesen, W.V.: Manual for the facial action coding system. Consulting Psychologists Press (1978) 4
8. Gökçe, c., Özdemir, O., Kındıroğlu, A.A., Akarun, L.: Score-level Multi Cue Fusion for Sign Language Recognition. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4

9. Hassan, S., Alonzo, O., Glasser, A., Huenerfauth, M.: Effect of Ranking and Precision of Results on Users' Satisfaction with Search-by-Video Sign-Language Dictionaries. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
10. Koller, O.: Quantitative survey of the state of the art in sign language recognition. arXiv preprint arXiv:2008.09918 (2020) 2
11. Koller, O., Camgoz, C., Ney, H., Bowden, R.: Weakly supervised learning with multi-stream cnn-lstm-hmms to discover sequential parallelism in sign language videos. IEEE transactions on pattern analysis and machine intelligence (2019) 2
12. Korte, J., Bender, A., Gallasch, G., Wiles, J., Back, A.: A Plan for Developing an Auslan Communication Technologies Pipeline. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
13. Kratimenos, A., Pavlakos, G., Maragos, P.: 3D Hands, Face and Body Extraction for Sign Language Recognition. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
14. Liang, X., Angelopoulou, A., Kapetanios, E., Woll, B., Al-batat, R., Woolfe, T.: A Multi-modal Machine Learning Approach and Toolkit to Automate Recognition of Early Stages of Dementia among British Sign Language Users. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
15. Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S., Narayanan, S.: Real-Time Sign Language Detection using Human Pose Estimation. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
16. Orbay, A., Akarun, L.: Neural sign language translation by learning tokenization. arXiv preprint arXiv:2002.00479 (2020) 3
17. Özdemir, O., Kındıroğlu, A.A., Camgöz, N.C., Akarun, L.: Bosphorussign22k sign language recognition dataset. arXiv preprint arXiv:2004.01283 (2020) 3
18. Parelli, M., Papadimitriou, K., Potamianos, G., Pavlakos, G., Maragos, P.: Exploiting 3D Hand Pose Estimation in Deep Learning-Based Sign Language Recognition from RGB Videos. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
19. Polat, K., Saraçlar, M.: Unsupervised Discovery of Sign Terms by K-Nearest Neighbours Approach. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 3, 4
20. Sanabria, R., Caglayan, O., Palaskar, S., Elliott, D., Barrault, L., Specia, L., Metze, F.: How2: a large-scale dataset for multimodal language understanding. In: Proceedings of the Workshop on Visually Grounded Interaction and Language (ViGIL). NeurIPS (2018) 4
21. Shi, B., Del Rio, A.M., Keane, J., Brentari, D., Shakhnarovich, G., Livescu, K.: Fingerspelling Recognition in the wild with Iterative Visual Attention. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
22. da Silva, E.P., Costa, P.D.P., Kumada, K.M.O., De Martino, J.M., Florentino, G.A.: Recognition of affective and grammatical facial expressions: a study for Brazilian sign language. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4

23. Siyli, R.D., Gundogdu, B., Saraclar, M., Akarun, L.: Unsupervised key hand shape discovery of sign language videos with correspondence sparse autoencoders. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8179–8183. IEEE (2020) 3
24. Tamer, N.C., Saraçlar, M.: Improving Keyword Search Performance in Sign Language with Hand Shape Features. In: Proceedings of the European Conference on Computer Vision (ECCV), Sign Language Recognition, Translation and Production (SLRTP) Workshop (2020) 4
25. Ventura, L., Duarte, A., Giro-i Nieto, X.: Can Everybody Sign Now? Exploring Sign Language Video Generation from 2D Poses. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4
26. Yin, K., Read, J.: Attention is All You Sign: Sign Language Translation with Transformers. In: Sign Language Recognition, Translation and Production (SLRTP) Workshop - Extended Abstracts (2020) 4