Digital tools for sign language research: towards recognition and comparison of lexical signs

Fragkiadakis, M.

CHAPTER 6

Conclusions and Future Research

This dissertation has sought to contribute to the understanding of how digital tools using machine and deep learning methodologies can be applied in the processing, recognition, and comparative analysis of sign languages.

6.1 Conclusions

We began our journey by posing five research questions, which have guided the structure and the flow of our study.

Our first research question asked about the efficacy of machine learning and deep learning methodologies in the processing and recognition of sign languages. Through our work, we found that these methodologies provide us with the tools necessary to manage the multi-modal character of sign languages, but more research and refinements are needed to fully leverage their potential.

Secondly, we ventured into the realm of automation, asking how machine learning methodologies could be employed to develop a system that predicts and annotates sign and gestural sequences from video material automatically. Despite the presence of noise, poor video quality, and the challenge of multiple signers, our proposed pipelines proved robust in classifying manual activation, thereby significantly enhancing the efficiency of sign language annotation and comparison. Furthermore, we have shown that large training sets are not required for the categorization of low-level characteristics such as when someone is signing or gesturing in a video. While such necessity might be true for the training process of deep neural networks, it might not be required for tasks like the one described earlier. By using machine learning models and classifiers, researchers can inspect the features that contribute to the output of such models and further assess their validity expanding the information gained from their use.

The third research question sought to understand the intricacies of creating a sign-based query search tool, and how the user's proficiency level impacts the performance of the tool. We found a relationship between signing proficiency and the classification accuracy of our methods, which suggests the need for more nuanced design considerations that take into account the proficiency levels of the end users.

Our fourth research question examined the most effective machine learning and deep learning methodologies for recognizing signs from a finite lexicon. We leveraged the pre-trained pose estimation framework OpenPose and demonstrated the effectiveness of Dynamic Time Warping (DTW) for sign ranking and retrieval. Notably, this approach doesn't necessitate further training, like deep neural networks, thereby making the process more efficient. Additionally, the versatility of this method allows it to be used across diverse sign language lexicons, making it a broadly applicable solution for sign ranking and suggestion.

Lastly, our fifth research question prompted us to develop a tool to measure and visualize variations in dominant hand trajectories across different sign languages. Utilizing the techniques we experimented with for the previous research question, our tool showed promising results in quantifying lexical variance among sign languages. The technology proposed in this thesis may be used to automatically derive true and false friends across sign languages and allows for visualizations for the dominant hand's wrist location and trajectory. Such a process can alleviate part of the time-consuming requirement to manually annotate sign form parameters. We contend that the Dynamic Time Warping algorithm is an excellent tool for assessing variance in signs and sign languages. Several sign identification investigations have employed the movement and wrist placement of the dominant hand, as predicted by OpenPose, [18, 106]. However, to the best of my knowledge, this is the first time it has been utilized for measurable variance measurement.

Additionally, our technique has been designed to be user-friendly and easy to use, even for researchers who may not have a background in sign language processing or computer science. By providing clear instructions and a simple interface, we have made it possible for researchers to quickly and accurately compare sign language movement with minimal effort.

It is possible that our tool could be also used to conduct historical analysis of sign languages. This could be done by collecting and analyzing historical videos of sign language use. By comparing the movements of the dominant hand wrist in these records, researchers could potentially identify changes and evolution in the sign language over time. For example, the tool could be used to compare the movements of the dominant hand wrist in sign language records from different time periods to identify changes in the way that specific signs are produced. This could provide insights into the development and evolution of the sign language over time, and could help researchers to better understand the underlying mechanisms of language change.

Moreover, the tool could potentially be used to compare sign language movements across different regions or communities, in order to identify similarities and differences in the way that the language is used. This could provide insight into the cultural and social factors that shape the evolution of sign languages, and could help researchers to better understand the diversity and complexity of sign languages around the world.

6.2 Limitations

This research, while paving the way for advancements in the field, has a few important limitations that must be acknowledged. These restrictions not only provide a clear scope of our study but also highlight potential areas for further research and refinement.

Normalization Method and Lack of Consideration for Camera Angle or Signer's Rotation

Firstly, all our methods primarily consider the signer's location in the frame for normalization, disregarding any adjustments for camera angle or the signer's rotational variation. This might limit the applicability of our findings, as real-life applications often deal with varying recording conditions. The signer's location, orientation, and the angle of the recording device could introduce variability that is not accounted for in the current models. These factors might lead to differences in the visual perspective, which, in turn, can potentially impact the accuracy of our methodologies. Further work needs to be done in order to incorporate these factors into our models to make them more robust and adaptable in varying conditions.

Limited Improvement in Sign Retrieval with Finger Location Information

Secondly, one of our surprising observations was the marginal enhancement of our sign retrieval methods when additional finger location information was added. This seems counter-intuitive as one would expect that the inclusion of more detailed data, such as the positioning of individual fingers, would improve the accuracy of the system. However, our findings could hint at a broader problem prevalent in current pose estimation frameworks like OpenPose. Specifically, these frameworks might struggle with accurately capturing and utilizing intricate finger location data, especially when dealing with low-quality video material in poorly lit environments. Further compounding the problem is the absence of depth information, which can be particularly useful when interpreting nuanced finger movements and positions.

This outcome underscores a limitation of our study and points towards a potentially significant area of improvement in the field. Improving the accuracy of hand and finger pose estimation in the context of sign language processing can significantly boost the recognition accuracy, as handshape is a crucial phonological parameter in many sign languages. Future studies should consider this aspect and work on improving the accuracy and reliability of pose estimation frameworks in capturing intricate details of signs, especially handshapes.

Sample Size and Signer Proficiency

Lastly, our study has also highlighted the influence of signer proficiency on sign language classification. However, generalizing from these findings is constrained due to the small sample size. While our study was able to demonstrate a trend, the exploration of this relationship requires further studies with a more considerable participant pool, representing various proficiency levels.

Moreover, our current models do not account for the proficiency level of signers, which might lead to variable performance in real-world scenarios. Sign language proficiency could affect the consistency and accuracy of sign production, influencing the effectiveness of recognition models. Future research needs to account for this variation, and models should be made robust to handle the variability introduced by signer proficiency.

By acknowledging these limitations, we hope to motivate future research to address these challenges and continue the progress towards efficient and robust sign language processing and recognition tools.

6.3 Future research

The challenges and findings identified throughout our study indicate several paths for future research. One of the most critical needs is to develop more robust and adaptable methods to recognize signs under diverse real-life conditions.

The collection and annotation of sign language data in naturalistic conditions is another important area for future exploration. Such data will enable the construction of more comprehensive and diverse datasets, providing rich resources for training and testing sign language recognition systems.

Moreover, it is critical to underline the significance of interpretability in sign language research, which extends beyond the rigorous criteria for automated translation systems. While deep learning and machine learning models have made major contributions to advances in sign language detection and processing, their "black box" nature frequently leaves researchers with a lack of knowledge of their decision-making processes. This can be a significant impediment to improving and modifying these systems to better suit certain requirements and settings in sign language research.

Fostering such interpretability not only benefits in the continual development and refining of these tools, but also builds confidence and acceptability among researchers and end-users. Moving forward, tapping the full potential of machine and deep learning approaches in sign language research will require a balance of accuracy, efficiency, and interpretability.

Ultimately, the goal of our research and potential future studies in this domain is to enhance our understanding of sign languages and to develop more advanced, accurate, and accessible communication tools for the sign language community as well as sign language researchers and linguists. We hope that our study will stimulate further research, discussion, and development in this field.