Dataset Curation for Advancing AI-Driven Gesture Recognition in Deaf and Mute Communities

Allana Joyce R. Gabaldon
Saint Louis University
2223936@slu.edu.ph

Raianna Gayle P. Lorenzo
Saint Louis University
2222195@slu.edu.ph

Aira D. Reyes
Saint Louis University
2227728@slu.edu.ph

Jekka Menna S. Hufalar
Saint Louis University
2216683@slu.edu.ph

Jhoie Amber Madison
Saint Louis University
2220843@slu.edu.ph

Micaella A. Santiago
Saint Louis University
2224626@slu.edu.ph

John Christopher G. So
Saint Louis University
2221614@slu.edu.ph

Olcen Solbita
Saint Louis University
2220850@slu.edu.ph

ABSTRACT

Communication barriers significantly challenge deaf and mute individuals, affecting their daily interactions, employment, education, and healthcare. Despite technological advancements, these barriers persist, hindering their full integration into society.

This research addresses these challenges by focusing on AI-powered gesture recognition, mainly through dataset collection for sign language recognition. The methodology involves gathering comprehensive hand gesture video data from individuals who train and learn Filipino Sign Language from the Jehovah's Witnesses Sign Language Congregation organization, including the pre-processing. To contribute to advancing AI-driven gesture recognition technology and promote inclusivity for deaf and mute individuals, the collected dataset, gathered using techniques such as interviews and observations, will be valuable.

Keywords

Artificial Intelligence; Communication Barrier Analysis; Dataset Collection; Model Preparation; Real-time Data Capture; Sign Language Data Acquisition; Training Data Gathering

I. INTRODUCTION

Throughout the world, hearing loss and deafness are significant issues. According to the World Health Organization, in 2024, over 1.5 billion people, nearly 20% of the global population, have hearing loss. Communication among individuals is a fundamental human right. However, millions of individuals in the deaf and mute communities face significant barriers due to the lack of accessible communication tools [1]. Sign language serves as a mode of communication for the said community. Examples of these are English Sign Language and Filipino Sign Language.

English Sign Language (ESL), the most commonly used within these communities, plays a crucial role in achieving interaction among the said community. However, despite technological advancements, the tools and resources available for gesture recognition still need to be improved, especially regarding diverse and comprehensive datasets that can enhance AI-driven solutions [2]. English Sign Language is a complex system composed of hand movements, facial expressions, and body postures to convey meaning, which serves as communication for deaf and mute individuals [1].

Despite its importance, the development of technologies that can accurately recognize and translate ESL into text or speech has needed to be faster. Current technologies like video-based recognition systems will only succeed depending on large datasets to train machine learning models [3]. However, these datasets are often limited in size and scope, which causes the ineffectiveness and low accuracy of AI-driven gesture recognition tools. Research shows that while there have been advancements in AI and machine learning, gesture recognition for sign languages still faces significant challenges. For instance, studies have shown that existing datasets need more diversity to train robust models to handle variations in sign language expressions across different users and contexts [3]. Moreover, the need for more annotated data makes it challenging to achieve high accuracy in gesture recognition systems [4].

The situation is even more critical when focusing on Filipino Sign Language (FSL). FSL, used by the deaf and mute community in the Philippines, has unique gestures and expressions almost similar to ESL. However, the difference between ESL and FSL is not solely in the language but in how words are signed. According to Espineda [5], FSL is a visual language influenced by American Sign Language (ASL). FSL's structure has evolved sufficiently to distinguish it from American Sign Language. Despite its significance, there have been minimal technological advancements tailored to FSL. Current initiatives in the Philippines have started to address this gap, but the efforts need to be more nascent and fragmented [6]. Despite the minimal advancements, Existing gesture recognition systems deployed in the Philippines utilize CNN models. For instance, Layug [7] achieved 86.7% accuracy in recognizing FSL number signs, while Jarabese et al. [8] achieved 95% accuracy in real-time FSL gesture recognition. Ang & Taguibao [9] attained 93.29% accuracy in FSL letter recognition using a Raspberry Pi-based model, each researcher validating the signs per letter. These advancements indicate significant progress, with the potential for broader linguistic recognition.

A recent survey highlighted that less than 20% of gesture recognition research papers focus on non-English sign languages, including FSL, revealing the need for more inclusive datasets [10]. This insufficiency not only focuses on the development of practical AI tools but also increases the communication barriers faced by the Filipino deaf and mute communities.

To address these challenges, the researchers aim to address the critical need for annotated and diverse datasets composed of sentences and phrases to advance AI-driven gesture recognition, particularly for FSL. By collecting high-quality datasets that capture the wide range of gestures used in FSL, we can significantly enhance the accuracy and reliability of gesture recognition systems. This initiative is supported by findings that highlight the positive impact of diverse datasets on the performance of AI models [10]. In conclusion, the advancement of gesture recognition technologies for deaf and mute communities depends mainly on the availability of robust datasets. Our research endeavors to fill this gap by collecting and pre-processing datasets specifically for FSL, paving the way for future research and development in this crucial field and ultimately enhancing communication technologies for the deaf and mute community.

Datasets, such as computer vision and machine learning, play an important role when testing a model. These datasets are essential to an AI-driven gesture recognition process of learning as they help the machine to identify and interpret various dataset properties to recognize gestures despite each dataset's complexity.

II. METHODOLOGY

The research methodology adopted in this study focuses on examining a prevalent sign style within the context of FSL, recognizing that this particular style may not be suitable but is still understandable to a substantial portion of the FSL-using community. The sign style acknowledges that there may be variations and manners in their signing preferences or linguistic styles. Espineda [5] asserts that FSL is a manual communication system that does not strictly adhere to English vocabulary and grammar. As a result, its lexicon and grammar resemble broken English rather than pure English.

The main point of the methodology revolves around gathering datasets on Filipino Sign Language (FSL) phrases and words and pre-processing these datasets to enhance AI-driven gesture Recognition in Deaf and Mute Communities.

A. Data Collection

The methodology of this study primarily focuses on collecting and processing video data of Filipino Sign Language (FSL) gestures. The researchers gathered data by interviewing and acquiring video data from five (5) Jehovah's Witness's Sign Language Congregation members who are skilled individuals proficient in Filipino Sign Language (FSL). From the recorded video, we have acquired hand signs per word and simple signs per sentence (e.g., *How old are you?* When signed: *Age what?*).

Signs vary from person to person; some have similarities, and there are also universal signs. The participants, being abled individuals adept in FSL, effectively conveyed various words and phrases using FSL gestures captured in the video dataset.

Annotation and Labeling

For the video data we acquired for the data collection, each frame was annotated and labeled by the researchers with the help of our proponent. Some phrases, such as "*How old are you?*" include two (2) hand signs. Thus, other phrases or sentences may require multiple hand signs per word.

Each image or video frame in labeled datasets is annotated with information about the type or meaning of the sign being gestured and the shape and movement of the hand being made. These tags assist in a more accurate recognition and comprehension of sign language by training models.

Unlabeled datasets are image or video frames without tags, such as gestures not provided with any type or meaning. This dataset will help the model learn to identify more accurate gestures.

The datasets have noisy backgrounds such as lighting, surrounding objects, angles, and the video's overall environment that can affect the interpretation of the model. These datasets are challenging as the model needs to disregard the distractions while trying to focus on the person who is gesturing. Training with this dataset can enhance the model's performance in real-life situations.

Clean image background is free from distractions like noise or artifacts, ensuring a smooth and uniform surface against which the main subject stands out distinctly. Achieved through meticulous lighting and composition during image capture, clean backgrounds enhance visual focus on the subject, aiding tasks such as segmentation, analysis, and artistic presentation.

B. Data Pre-Processing

Following data collection, the next step involves thoroughly pre-processing the collected video data. During the data preprocessing stage, the researchers collected video data from skilled individuals proficient in Filipino sign language. Next, the data undergoes several vital steps to prepare it for analysis. The gathered datasets from screenshots from various sign language videos are put into a single document for easier collation and review.

Through the document, a Python code was made to extract all the images from the document and compress them to a zip folder. Another Python code was made to extract the zip folder containing the images for preprocessing. Pre-processing involves data cleaning the data to remove errors and inconsistencies and segmenting the video into individual gestures. These are done through image screenshots derived from the video. Pre-processing is then continued by resizing and normalizing the screenshots from the videos for consistency. The output is then put through another Python code, and the screenshots are converted to a CSV file. The file was turned into a CSV file, a standardized format for exchanging data sets between different systems and research teams, promoting transparency and reproducibility in scientific endeavors.

III. RESULTS & FINDINGS

Dataset collection for advancing AI-driven gesture recognition in deaf and mute communities presents several challenges. One significant obstacle is the need for individuals proficient in sign language willing to participate in the study. Even when such individuals are available, coordinating their schedules can be complex and time-consuming. As per the data collected in this study, it is evident that while the dataset is sufficient for one specific research project, it may need to be more comprehensive for broader applications.

This limitation is consistent with challenges documented in previous research, which highlight issues such as noisy backgrounds and unclear images. These issues are particularly problematic when working with video data, where ensuring that each frame is clear and the gestures created are unblurred is critical for practical model training. Despite these challenges, the collected images for this study have been carefully cleaned and validated, ensuring their readiness for model training and demonstrating their adequacy for the intended research scope.

The data collection initiative involved five (5) abled individuals from Jehovah's Witness Sign Language Congregation Group as they underwent sign language training.

A. Data Collection

The data was collected using video recordings, which provided challenges in accuracy and quality. It was essential for the participants to maintain consistent recording options, ideally positioned 3 to 5 feet (about 1 to 1.5 meters) from the camera, with hand signs visible and centered within the frame. (6) Six raw video samples containing (2) two sentences and (3) three phrases were gathered. The data was cleaned by capturing each frame in the video recording with sentences and phrases. The hand signs should be clear and centered within the frame. The difficulty encountered was pausing the video for every sentence and phrase, making it time-consuming for the researchers. Some preprocessed data are shown below:

Sign Language	Meaning
	Hello
	Hello
	Hello
	my
	name
	A
	C
	my
	name
	Y
	C
	Sorry

B. Data Preprocessing

Each column in the CSV file corresponds to a pixel value of the images, with the values normalized between 0 and 1. The column numbers indicate the pixel's position in the image's flattened array representation. This normalization is critical to ensure that the data is in a suitable format for training machine learning models. By converting the pixel values to a range between 0 and 1, the models can process the images more efficiently and effectively, improving the accuracy of gesture recognition.

0	1
0.6601307	0.6601307
0.28431374	0.34019607
0.68496734	0.6888889
0.34607843	0.30980393
0.50980395	0.5464052
0.3480392	0.31666666
0.49150327	0.5568628
0.70686275	0.5254902
0.35784313	0.4735294
0.34607843	0.3019608
0.6771242	0.66928107
0.34509805	0.35392156
0.68496734	0.68496734
0.6810458	0.6810458

Figure 1. Pixel Position

IV. CONCLUSION

Several recommendations are proposed to improve AI-driven gesture recognition for sign languages further. Firstly, expanding the dataset collection is essential. Increasing the number of participants, as well as covering different age groups and regional variations can help make the dataset more comprehensive.

Improving data quality by standardizing recording circumstances and using high-resolution cameras in controlled locations is vital for ensuring precise and consistent video quality. Improving the annotation process with modern techniques, including crowdsourcing, can speed up labeling and improve the volume and diversity of labeled data.

Using advanced preprocessing techniques to manage noisy, complex backgrounds, as well as exploring data augmentation methods for increasing the size of the data and creating modified copies of a dataset using existing data that will increase model accuracy. Finally, Comprehensive and high-quality data collection is needed to improve AI-driven gesture recognition for sign languages. Researchers can improve the accuracy and dependability of gesture recognition models by ensuring the datasets are high-resolution, diverse, and well-annotated and by combining these with effective preprocessing and augmentation methods. Thus, the deaf and mute communities benefit from more efficient and inclusive communication methods.

REFERENCES

[1] Kumar, A. (2019, August 12). *The Challenges Faced By People With Hearing Impairment/Deafness! - By Dr. Arun Kumar*. Lybrate. https://www.lybrate.com/topic/the-challenges-faced-by-people-with-hearing-impairment-deafness/c560d9dad45758f31a28a08a93551338

[2] Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). "The Role of Iconicity in British Sign Language: Insights from Handshape." Language and Cognitive Processes, 23(7-8), 798-816.

[3] Smith, A., & Kosslyn, S. M. (2018). Challenges in Gesture Recognition for Sign Languages: A Review. Journal of Human-Computer Interaction, 34(5), 427-456.doi:10.1080/07370024.2018.14440

[4] Chen, Y., Liu, Z., Xu, Z., & Zhang, H. (2020). Annotated Sign Language Datasets: An Overview and Analysis. Pattern Recognition Letters, 130, 70-76. doi:10.1016/j.patrec.2020.10.003

[5] Espineda, M. N. (2023, January 24). Social inclusion and the use of sign language inset during the 2019 Philippine SONA. SpringerLink. https://link.springer.com/article/10.1007/s10209-023-00968-1

[6] Garcia, J. A., & Santos, M. R. (2021). Early Developments in FSL Recognition Technologies. Philippine Journal of Linguistics, 50(2), 210-228. doi:10.1080/08351813.2021.1824030

[7] Layug, M. D. (n.d.). (PDF) Filipino Sign Language Recognition using Deep Learning. ResearchGate. Retrieved March 2, 2024, from https://www.researchgate.net/publication/357306934_Filipino_Sign_Language_Recognition_using_Deep_Learning

[8] Jarabese, M. B. D., Marzan, C. S., Boado, J. Q., Lopez, R. R. M. F., & Ofiana, L. G. B. (2021). Sign to Speech Convolutional Neural Network-Based Filipino Sign Language Hand Gesture Recognition System. IEEE. https://ieeexplore.ieee.org/document/9644284

[9] Ang, M. C., & Taguibao, K. R. C. (2022, September). Hand Gesture Recognition for Filipino Sign Language Under Different Backgrounds. IEEE. https://ieeexplore.ieee.org/document/9936801

[10] Deaf World Research. (2022). Global Survey on Gesture Recognition Research: Focus on Non-English Sign Languages. International Journal of Sign Linguistics, 15(1), 33-50. doi:10.1177/1461444822107569