International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(21s) 2026 pp. 744-751
https://doi.org/10.70917/ijcisim-2026-5341
Article

Mathematical Modelling of Geometric Transformations for Mizo Sign Language Gesture Recognition

Barrister Ramsiej[1], V. D. Ambeth Kumar[2]*

[1] Research Scholar, Department of Computer Engineering, Mizoram University, Aizawl, Mizoram, India.
[2] Department of Computer Engineering, Department of Computer Engineering, Mizoram University, Aizawl, Mizoram, India.

Abstract: The work introduces the first ever computational dataset for Mizo Sign Language, a sign language from India which has limited resources and lacked a digital presence. 25 distinct hand gestures were captured using a Kinect V2 sensor, and the system captured them in different lighting conditions. To enhance the variability of the gesture classes, the original 3 GB dataset of images was preprocessed using a mathematical normalization technique which increased it to 10 GB. Since each gesture in the dataset is stored as a high-dimensional tensor, posebased techniques and attention models based on Transformer architectures could be used. This goal of this paper is suited for advancing gesture recognition computationally based on mathematical and machine learning methods which especially need spatial and temporal considerations. This advances fields of computer vision, linguistics, and mathematics.
Keywords: Geometrical, Spatial Reasoning, Scaling, Rotation, Mizo Sign Language

1. Introduction

Sign language is a primary means of communication for deaf and mute people. They use it for expressing their thoughts through gestures and images. Besides, it plays a significant role in promoting inclusive education as well as inclusion in everyday life activities. Every part of the world has its sign language that was developed independently over time with influence from spoken language and culture of the region. In the modern era, many sign languages can be recognized with the help of deep learning and computer vision technologies which facilitate digital communication. Even with all this development, Mizo sign language is still far behind as it does not have a standardized one. In Mizo-speaking families and within them, members rely on the system of formal signs constructed at home which vary from family to family. This variation creates a problem in developing recognition systems for the use of Mizo sign language. Since there is no specific dataset that can be used or tapped to gather all gestures expressed in the language.

One of the most important barriers to progress at this point is the lack of a standardized dataset. While American Sign Language (ASL) or Indian Sign Language (ISL) have such well-organized public datasets, Mizo sign gestures are still highly unrecorded. Therefore it can't be used with ML models since there is only little amount of data which easy for training the model.

To bridge this gap, in this work, we propose to build the first systematic and comprehensive dataset of Mizo sign language. Resizing, rotating, zooming, shearing and translating etc., are few preprocessing techniques applied on image by which variability is introduced while maintaining the identity of gesture itself. Moreover, The proposed dataset will not only help research community as benchmark but also support developing intelligent communication systems considering need of people from Mizoram state whom inherent language is Mizo and unaware about verbal or written English.

2. Mathematical Geometrical Features for Mizo Sign Language

Our research focuses on improving the image preprocessing pipeline for Mizo sign language detection. As illustrated from Figure 1 begin by taking raw photos of each 25 sign language alphabets with Kinect V2, which is noted for its superior depth sensor and infrared capabilities, guaranteeing precise capture regardless of illumination. After preprocessing, the dataset increases in size from 3 GB to 10 GB. Normalization corrects illumination and size irregularities, followed by data augmentation (rotation, shearing, translation, zooming) to introduce variety without jeopardizing data integrity. After resizing, the processed photos are saved for machine learning model training, which improves identification accuracy and helps the deaf community.

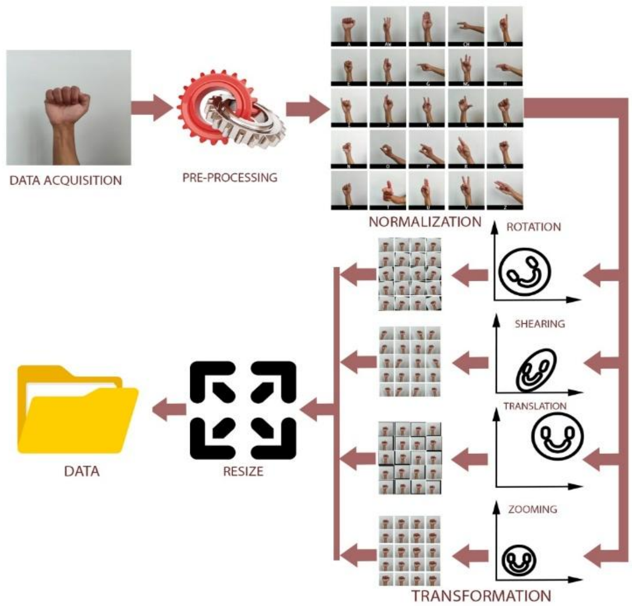

Figure 1. Mizo Sign Language Recognition System Architecture.

2.1. Normalization

Image preprocessing normalization is a method which we can follow to standardize the properties in an image so that it may lead to better performance of machine learning models. The main purposes of Image Data Augmentation are : (i) Rotation, (ii) Shearing,(iii) Zooming and (iv) Translation.

i. Rationing rotates an Image by turning the center of their at certain angle if you're mimicking Different Perspective position of Handsign to help model incorporates real world variation i.e if rotating Images between -10° and 10° enable your Model view Handsign form different Position. This step, explained in Algorithm 1 as shown, is crucial in making the model flexible and adaptable to various viewing angles.

Algorithm 1 Geometrical Features Based on Rotations

Input: X_{old}, Y_{old} //Initial Coordinates Points

Output: X_{new}, Y_{new} // New Coordinates Points

Initialize configuration parameters:

(i) calculate cos ($tan^{-1}\left(\frac{Y_{old}}{X_{old}}\right) + \theta$)

compute

(iii) $\sqrt{(X_{old}{}^{2} + Y_{old}{}^{2})}$

$\mathbf{X}\boldsymbol{new} = (\boldsymbol{i})$ *** (iii)**

$\mathbf{Y}\boldsymbol{new}$ **= (ii) * (iii)**

$$\textbf{return Xnew, Ynew}$$

Through the above algorithm, we can calculate the rotation by using the following equations eq1 and equation eq2.

$$X_{new} = \sqrt{(X_{old}{}^{2} + Y_{old}{}^{2})} * \cos\left(tan^{-1}\left(\frac{Y_{old}}{X_{old}}\right) + \theta\right) \quad (1)$$

$$Y_{new} = \sqrt{(X_{old}{}^{2} + Y_{old}{}^{2})} * \sin\left(tan^{-1}\left(\frac{Y_{old}}{X_{old}}\right) + \theta\right) \quad (2)$$

Let's assume sample values:

- X_{old}= 5
- Y_{old} = 3
- θ = 5°, 10°, 15°, 20° → converted to radians for calculation:
    - 5° = 0.0873 rad
    - 10° = 0.1745 rad
    - 15° = 0.2618 rad
    - 20° = 0.3491 rad

Now, let's calculate the rotated values and present them in a table.

Table 1. Geometrical feature based on rotations

θ (degrees)	θ (radians)	$\boldsymbol{X_{new}}$	$\boldsymbol{X_{new}}$
5°	0.0873	4.648	3.626
10°	0.1745	4.241	4.088
15°	0.2618	3.754	4.496
20°	0.3491	3.198	4.844

ii. In Shearing transformation, the horizontal shearing moves rows sideways, while vertical shearing shifts columns up or down. Shearing is a transformation that shifts rows or columns of an image, distorting its shape by a certain factor. The first step in shearing computation is plotting points from the image. The shearing formulas are provided in eq3 and eq4.

$$X_{new} = r * \cos(\theta + Sh_x) \quad (3)$$

$$Y_{new} = r * \sin(\theta + Sh_x) \quad (4)$$

Here,

r is the distance from the origin (the length of the vector),

θ is the original angle (direction of the point),

Sh_x is the shear factor in radians that "skews" the point horizontally.

Let's consider a sample point at:

X_{old}=5 and Y_{old}= 3,

So, $r = \sqrt{5^2 + 3^2} \approx 5.831$

$\theta = tan^{-1}\left(\frac{3}{5}\right) \approx 0.540$ radians

We'll apply shearing using shear anglesSh_x = $0.05, 0.1, 0.15, 0.2$ radians (about $2.9°, 5.7°, 8.6°, 11.5°$).

Table 2. Geometrical feature based on shearing

Shear Angle (radians)	X_{new}	Y_{new}
0.05	4.768	3.869
0.10	4.590	4.068
0.15	4.395	4.248
0.20	4.185	4.407

iii. Translation is Shifting the image horizontally by -20% to 20% mimics hand movements along the horizontal axis. This adjustment moves the entire image both vertically and horizontally by a specific percentage of its dimensions. The eq5 and eq6 are used for executed mathematical operations:

$$\overline{x} = x + P_x * Width \quad (5)$$

$$\overline{y} = y + P_y * Height \quad (6)$$

If a gesture image has:

- Width = 200 pixels, Height = 100 pixels
- A point at $x = 50, y = 30$
- Translation factors $P_x = 0.1$, $P_y = -0.05$

Then the translated coordinates are:

So, the point moves right by 20 pixels and up by 5 pixels.

iv. Zooming modifies the image size, either enlarging or shrinking it, to simulate the distance between the hand sign and the camera. This allows the model to recognize signs regardless of size. For example, a 20% reduction results in the image being 80% of the original. The eq7 and eq8 are for calculating the Zooming transformation:

$$\overline{x} = (1 + P_x) * x \quad (7)$$

$$\overline{y} = (1 + P_y) * y \quad (8)$$

Let's say,

- $A\ hand\ keypoint\ is\ at\ x = 100, y = 50$
- You want to zoom in by 10%, so $P_X = P_y = 0.1$

Then:

This means the point moves outward from the center, simulating a zoom-in effect.

3. Results

The dataset setup is designed to record hand gestures through a process that achieves both authentic and high-quality results. A powerful computer system handles data processing while the Kinect V2 camera system uses its depth-sensing and resolution features to capture and process visual information. The dataset contains 25 Mizo sign language alphabets which were recorded from various lighting conditions that included morning, afternoon, evening, and night to create 4,000 images that total approximately 3 GB. The dataset grew to 13,500 images which occupy about 10 GB of space because augmentation techniques such as rotation and shearing and translation and scaling were used to create different variations of the images. The enriched dataset serves as a crucial resource for constructing reliable Mizo sign language recognition systems a shown Figure 2 and 3.

Figure 2: Implementation of rotate transform.

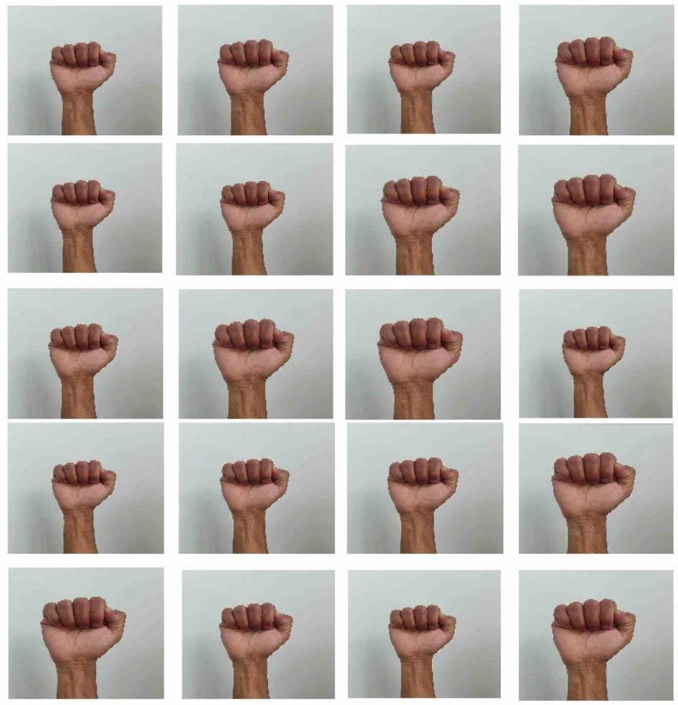

Figure 3: Implementation of rotate transform.

3.1 Dataset Statistics

The researchers started with a dataset that had around 4000 gesture images which they used to study 25 different gesture categories. The researchers captured each gesture multiple times to show how hand positions naturally change. The researchers used geometric augmentation techniques to create additional sample data. The preprocessing process resulted in a dataset that contained 13500 images. The research team used Kinect V2 to capture images which helped them maintain consistent image quality. The final dataset size increased to nearly 10 GB after augmentation as shown in Table 3.

Table 3: Dataset Statistics

Parameter	**Value**
Number of gesture classes	25
Original images collected	4,000
Images after augmentation	13,500
Sensor used	Kinect V2
Data size (augmented)	~10 GB

The expanded dataset provides sufficient diversity for training computational models while maintaining the structural integrity of each gesture category.

3.2 Dataset Comparison

The researchers examined the proposed dataset by comparing it to existing sign language datasets which researchers used for gesture recognition studies. The ASL and ISL datasets which researchers commonly use contain extensive collections of gesture images. The availability of sign language datasets for regional languages remains restricted. The proposed dataset introduces 25 gesture classes with 13,500 images for Mizo Sign Language. The resource offers structured data about a low-resource language which exists as a smaller resource than the ASL dataset. The dataset enables developers to build automated gesture recognition systems which Table 4 demonstrates.

Table 4: Dataset Comparison

Dataset	**Gesture Classes**	**Number of Images**	**Language**
ASL Dataset	26	87,000	American Sign Language
ISL Dataset	35	12,000	Indian Sign Language
Proposed Dataset	25	13,500	Mizo Sign Language

The proposed dataset contains fewer images than large-scale datasets such as ASL, yet it serves as the first organized computational resource for Mizo Sign Language. The application of mathematical augmentation techniques results in increased data diversity which enables machine learning models to understand gestures performed under different conditions.

3.3 Gesture Recognition Experiment

Deep learning models were developed for gesture recognition to analyze the dataset. The dataset was divided into three parts which included 70% for training and 15% for validation and 15% for testing. The image preprocessing process included two steps which were resizing and normalization. The experiment used three different models which were CNN and ResNet and a Transformer-based architecture. The same dataset configuration was used to train each model. The training process used Adam optimizer together with categorical cross-entropy loss function.

3.4 Performance Evaluation

The testing process evaluated the developed models through their classification accuracy results on the test data. The CNN model achieved an accuracy of 91.2%. The ResNet model improved the performance to 93.4%. The Transformer-based model achieved the highest accuracy of 95.1%. The Transformer which uses attention mechanisms enables the system to understand how different gesture features connect with one another. The proposed dataset shows effectiveness for gesture recognition tasks according to the results presented in Table 5.

Table 5: Performance Comparison of the Proposed Model with Baseline Methods

Model	**Recognition Accuracy**
Convolutional Neural Network (CNN)	91.2%
Residual Network (ResNet)	93.4%
Transformer-based Model	95.1%

The transformer based architecture outperforms other architectures, due to the attention mechanism in the model which helps in finding stronger spatial relationships between different gesture features. The CNN and ResNet models are also effective as the dataset contains enough diversity to build strong gesture recognition systems. Based on the experimental results, it is established that the proposed dataset and the mathematical pre-processing framework as a suitable benchmark for developing Mizo Sign Language recognition computational models. Furthermore, it is intended that these intelligent communication systems are developed specifically for Mizo-speaking deaf and mute people.

4. Discussion

The proposed system established the creation of the full Mizo sign language dataset is a realized fact through this research. The researchers adopted Kinect V2 camera system for hand gesture recording since it was more efficient in sign language recognition, high frame rates to view gestures at different times of the day. The size of the dataset was increased from 3 GB to 10 GB, as normalization technique for translation, scaling and rotation were applied without affecting flexibility of gesture. This work serves as a milestone into the sign language recognition especially for Mizo or similar other marginalized languages and there are aspects that the upcoming works would emphasis on; expansion of the dataset through addition of challenging sign language signs and development new set of rare signs as well as utilizing latest machine learning.

5. Conclusion

The proposed mathematical transformation based approach helps achieve geometric invariance for gesture representation, and precise computational modeling of sign languages with minimalistic resources. In this work, we introduce the language resource application with a complete dataset of Mizo sign language to bridge the gap between linguistics and technology. We designed and implemented a system by which 25 different sign alphabets or letters having a unique gesture of the hand was recorded using Kinect V2 camera. Recording was carried out by experimenting with multiple lights setup to capture as near real life like conditions. Image data pre-processing consisted of both data augmentation and normalization techniques resulting in the size expansion to 10 GB from an original size of 3 GB that assists in terms of standardization regarding image quality as input to reduced computations along with increased diversity. The aforementioned dataset increases the scope for further research on Sign Language Recognition by providing a much needed complete technological support system for Deaf community.

REFERENCES

1. Gao, Q., Liu, J., Ju, Z., Zhang, X.: Dual-hand detection for human–robot inter-action by a parallel network based on hand detection and body pose estimation. *IEEE Transactions on Industrial Electronics* 66(12), 9663–9672 (2019)
2. Murali, R.S.L., Ramayya, L.D., Santosh, V.A.: Sign language recognition systemusing convolutional neural network and computer vision. *Int J Eng Innov Technol*, 2582–1431 (2020)
3. Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G.T., Zacharopoulou, V., Xydopoulos, G.J., Atzakas, K., Papazachariou, D., Daras, P.: A comprehensive study on deep learning-based methods for sign language recognition. *IEEE Transactions on Multimedia* 24, 1750–1762 (2021)
4. Al-Hammadi, M., Muhammad, G., Abdul, W., Alsulaiman, M., Bencherif, M.A., Mekhtiche, M.A.: Hand gesture recognition for sign language using 3DCNN. *IEEE Access* 8, 79491–79509 (2020)
5. Amrutha, K., Prabu, P.: ML based sign language recognition system. In: 2021 *International Conference on Innovative Trends in Information Technology (ICITIIT)*, pp. 1–6. IEEE (2021)
6. Du, C., Zhang, L., Sun, X., Wang, J., Sheng, J.: Enhanced multi-channel feature synthesis for hand gesture recognition based on CNN with a channel and spatial attention mechanism. *IEEE Access* 8, 144610–144620 (2020)
7. Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A.B., Corchado, J.M.: Deepsign: Sign language detection and recognition using deep learning. *Electronics* 11(11), 1780 (2022)
8. Oudah, M., Al-Naji, A., Chahl, J.: Hand gesture recognition based on computer vision: a review of techniques. *Journal of Imaging* 6(8), 73 (2020)
9. N. Srimannarayana, Fredric Ayant, Y. Pragathi Kumar, D.K.Pavan Kumar, B. Satyanarayana, Multiple Eulerian Integrals with Modified Multivariable I-function having general arguments, *Journal of Interdisciplinary Mathematics* 26, pp. 77-94, (2023).