



# Midterm Presentation Attendance

## Report 1

### on Real-time Hand Gesture Recognition via Subspace Representation on Grassmann Manifold

**report submitter:** Mamanchuk Mykola, SID.202420671

**for speaker:** Santos Enoque S., SID.202320695

August 12, 2024

## 1 Introduction

During the presentation, a topic of discussion is about hand gesture recognition playing a crucial role in applications such as human-computer interaction and sign language interpretation. Traditional methods in this field often require extensive training data, significant computational resources, and struggle with variations in gesture representation.

Presenter introduces the Grassmann Subspace Method (GSM) as a novel approach to address these challenges. GSM requires no traditional training, is robust to affine transformations, and operates with low computational cost, all while achieving high accuracy in real-time scenarios. Building upon established pattern recognition techniques, GSM extends these methods to handle the complex task of real-time hand gesture recognition, specifically for the American Sign Language (ASL) alphabet.

**The key strengths of GSM** are stated to be **precise in recognition accuracy with minimal data requirements**, robustness in **real-time scenarios**, and potential for **generalization across different users and hand orientations**. This approach advances the practical applications of GSM in gesture recognition and opens new avenues for efficient, low-resource machine learning in human-computer interaction.

## 2 Background of Research and Theoretical Keypoints

Following concepts and metrics are introduced as a theoretical part of the research.

## 2.1 Subspace Method

The Subspace Method (SM) is a pattern recognition technique that compares an input vector  $\mathbf{p}$  to a set of  $k$  class subspaces in an  $f$ -dimensional space. The similarity  $S$  between  $\mathbf{p}$  and the  $i$ -th class subspace is calculated by:

$$S = \cos^2 \theta = \sum_{i=1}^{d_q} \frac{(\mathbf{p} \cdot \phi_i)^2}{\|\mathbf{p}\|^2},$$

where  $d_q$  is the dimension of the class subspace.

## 2.2 Mutual Subspace Method

The Mutual Subspace Method (MSM) extends SM by representing both the input and class as subspaces, capturing structural similarities with:

$$\text{sim}(S_i, S_k) = \frac{1}{t} \sum_{i=1}^t \cos^2 \theta_i.$$

## 2.3 Shape Subspace

A shape subspace for 3D objects is defined by the column vectors of a  $P \times 3$  matrix  $\mathbf{X}$ :

$$\mathbf{X} = (\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_P)^T,$$

where  $\mathbf{r}_p = (x_p, y_p, z_p)^T$  is the positional vector of the  $p$ -th point.

## 2.4 Grassmann Manifold

The Grassmann manifold  $\mathcal{G}(m, D)$  consists of  $m$ -dimensional subspaces in  $\mathbb{R}^D$ , represented by points on the manifold. Similarity between subspaces is calculated using the Grassmann kernel:

$$k(S) = k(S, S'_i) = \text{sim}(S_1, S_2).$$

This allows Principal Component Analysis (PCA) to be **applied on the manifold, improving recognition performance** by capturing geometric relationships.

# 3 Intro to Experiment Methodology

## 3.1 Data Collection and Preprocessing

The study utilized MediaPipe to extract 3D hand skeletons from images of ASL alphabet gestures, ranging from A to I. Each gesture was represented by a set of 3D coordinates that corresponded to key points on the hand, forming the basic dataset for subsequent processing.

### 3.2 Subspace Generation

To create a shape subspace for each hand gesture, Singular Value Decomposition (SVD) was employed. The process began by forming a shape matrix  $\mathbf{X}$  from the 3D hand skeleton points:

$$\mathbf{X} = (r_1, r_2, \dots, r_p)^T$$

The matrix was then centered by subtracting the mean position of all points. Following this, SVD was applied to decompose the centered matrix into its constituent components:

$$\mathbf{X}_c = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

The resulting shape subspace  $\mathbf{S}$  was defined by selecting the first  $d$  columns from the matrix  $\mathbf{U}$ .

### 3.3 Grassmann Kernel

The next step involved using a projection kernel on the Grassmann manifold to measure the similarity between different subspaces. This kernel, based on the Frobenius norm, was calculated as:

$$k(S_1, S_2) = \|S_1^T S_2\|_F^2$$

This allowed for a robust comparison of the subspace structures.

### 3.4 Projection onto Grassmann Manifold

The projection of subspaces onto the Grassmann manifold was carried out through a series of steps. Initially, the kernel matrix for reference subspaces was computed using the previously mentioned kernel function. This was followed by an eigendecomposition of the kernel matrix to derive the eigenvalues and eigenvectors. Each reference subspace was then projected onto the Grassmann manifold by combining these eigenvalues and eigenvectors in a specific manner. The same process was used to project input subspaces, ensuring that all gestures were represented in a comparable form on the manifold.

### 3.5 Grassmann Subspace Method

The GSM is stated to involve several key steps:

- Reference subspaces are projected onto the Grassmann manifold.
- Input gestures are transformed into subspaces and similarly projected.
- Similarities between input subspaces and reference subspaces are computed using the Grassmann kernel.
- Classification of the input gestures is performed using the Subspace Method on the Grassmann manifold, effectively distinguishing between different hand gestures.

## 4 Summary on Results

The research demonstrated strong performance in recognizing static hand gestures. The key strength is stated to be the ability to bypass the traditional training phase. Instead, **the method relies on projecting reference subspaces onto the Grassmann manifold** and computing similarities for classification, **making the process efficient**.

### 4.1 Performance Metrics

The system achieved an **average accuracy of 98.5%**, while **maintaining a real-time frame rate of 60** frames per second. These results highlight the potential of the GSM to provide accurate and efficient hand gesture recognition in real-time applications.

It is noted that the accuracy was determined through manual verification. The manual verification process involved human reviewers checking whether each gesture was correctly identified which introduces a degree of subjectivity (possibility of human error). To address these concerns, the researcher has **suggested the need for more automated and robust evaluation strategies** in future iterations.

### 4.2 Inferences

The high accuracy observed with no training phase highlights the effectiveness of employing Grassmann manifold representations for hand gesture recognition. By projecting hand shape subspaces onto the Grassmann manifold, the method effectively captures the intrinsic geometry of hand shapes, making it **robust to variations in scaling and rotation**.

However, the **reliance on manual verification for evaluation presents limitations**. This method can introduce biases and is not easily scalable for larger datasets or continuous operation.

## 5 Impact and Future Prospects

The research on hand gesture recognition using the GSM has significant implications for the field of computer vision, particularly in areas where real-time performance and minimal training data are crucial. The approach’s ability to maintain high accuracy without extensive training makes it a valuable tool for interactive applications in human-computer interaction, such as sign language translation and virtual/augmented reality interfaces.

Looking forward, the potential for this method to be adapted for dynamic gestures, integrated with temporal information, and enhanced with more sophisticated manifold learning techniques opens up avenues for more advanced and robust recognition systems. The flexibility of GSM to incorporate new gestures without retraining further highlights its potential to drive innovation in gesture-based interaction technologies, providing a foundation for future advancements in this field.

## References

1. **Mamanchuk N., University of Tsukuba**, Github, August 12, 2024. Available online: <https://github.com/RIFLE>
2. **Santos Enoque S., University of Tsukuba**, on Real-time Hand Gesture Recognition via Subspace Representation on Grassmann Manifold, 2024. Available online: [https://www.cs.tsukuba.ac.jp/lecture/midterm/local/24\\_data/mov/202320695.mp4](https://www.cs.tsukuba.ac.jp/lecture/midterm/local/24_data/mov/202320695.mp4) [Accessed: 2024-08-12]