Hendrik Matvejev, Priit Peterson, Karl Kristjan Puusepp

Estonian sign language fingerspelling recognition

Identifying Business Goals

Background

The Estonian Fingerspelling Sign Language Recognition project aims to be an interactive and easy-to-use resource for people learning Estonian sign language fingerspelling.

Business Goals

- Promote inclusivity and accessibility by bridging communication gaps through technology.
- Create a tool that aids in learning and practicing Estonian sign language.

Business Success Criteria

- High accuracy in recognizing Estonian sign language fingerspelling signs.
- User satisfaction with the usability and effectiveness of the recognition system.

Assessing the Situation

Inventory of Resources

- Access to the Google Cloud to train the model.
- Access to the Mediapipe library for hand tracking and gesture recognition.
- Team members.
- Volunteers for data gathering.

Requirements, Assumptions, and Constraints

Requirement: Resources about fingerspelling signs in the Estonian sign language.

Assumption: The Mediapipe library can accurately capture and analyze Estonian sign language signs.

Constraint: The data will probably have low variety, as it is based on a small set of individuals. Time. Mediapipe's lack of understanding of more complex 3-dimensional hand gestures.

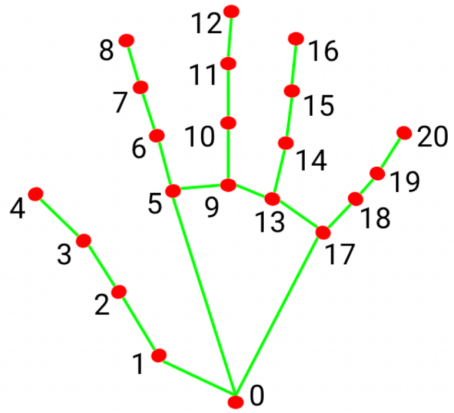
Risks and Contingencies

Risk: Insufficient diversity in the training data.

Contingency: Actively seek and include a diverse set of people in the training dataset.

Terminology

- Fingerspelling – The act of spelling out a word (usually a proper noun) letter by letter, using a set of fingerspelling signs from the Estonian sign language alphabet.
- Recognition – Creating a TensorFlow based model that outputs a label when given an image/images.
- Sign/gesture – A hand gesture, which in this case corresponds to a single letter in the Estonian alphabet.
- Estonian alphabet – The set of 32 symbols for letters in the Estonian language.
- Hand landmark – A datapoint (set of coordinates in 2-dimensional space of an image) representing a single point of a gesture. Each corresponds to a keypoint of a hand such that a set of 21 landmarks forms a hand gesture.



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Hand landmarks (Mediapipe).

Costs and Benefits

Costs: No significant costs
Benefits: The project is free, open-source and unmonetized, therefore there are no monetary benefits.

Defining Data-Mining Goals

Data-Mining Goals

- Creating a comprehensive and labeled dataset of all 32 fingerspelling signs in the Estonian sign language.
- Train a machine learning model using the Mediapipe library to accurately recognize Estonian sign language signs from images.
- Implement a user-friendly interface for capturing and processing images of signs.
- Develop a visualization component for users to see the recognized signs in real-time.

Data-Mining Success Criteria

- Model accuracy above 75%.
- Real-time recognition performance within acceptable latency.

- Positive user feedback on the user interface and overall system functionality.

Data Understanding

Gathering Data

Data Requirements

The primary data requirement is a diverse and comprehensive dataset of images containing different Estonian sign language signs. Each image should be labeled with the corresponding sign for supervised machine learning. The dataset needs to cover a wide range of hand shapes and orientations to ensure the model's robustness.

Verifying Data Availability

Upon conducting a preliminary search, we identified that no such dataset exists. Thus we will have to create the dataset ourselves.

Data gathering

We gathered data by photographing seven individuals (four men, three women – aged from 18 to 21 years old) in various locations, capturing over 200 images for each sign. The images were captured manually using a laptop webcam, which is in line with the final use case of the product, as the software is meant to be run through a live webcam feed. Participants were asked to slightly vary their signs to help the model learn both the essence of a sign and its spatial properties. This intentional diversity ensures the model's effectiveness in recognizing Estonian fingerspelling signs across different perspectives and real-world scenarios. The images were then pruned of any metadata and cropped to fit only the hands.

Define Selection Criteria

The selection criteria for creating the dataset includes:

1. Diversity: Ensuring representation of various signs, hand shapes, and orientations. Different individuals have different hand proportions which can affect the sign shape. To prevent this, we used individuals of various genders and hand sizes.

2. Quality: Taking images with clear and distinguishable signs to improve model accuracy.

3. Authenticity: Prioritize images captured in real-world scenarios to enhance the model's applicability in practical situations.

Describing Data

Exploring Data

This doesn’t apply since the dataset will be created by us.

Verifying Data Quality

1. **Motion blur** in images can compromise the accuracy of sign language recognition models. To mitigate this, images affected by motion blur were excluded, ensuring the dataset consists of clear and distinguishable hand gestures.

2. Real-world datasets may inadvertently include **incorrect gestures**. A manual review was conducted to correct mislabeled signs or gestures inconsistent with Estonian sign language, guaranteeing the dataset's quality.

3. Variations in lighting and hand orientation can lead to **false recognition of hand landmarks**. Through analysis, images with inaccurately detected landmarks were identified and excluded.

4. Recognizing **two-handed gestures** poses a challenge in sign language recognition. These had to be approached as a separate task.

5. The caveat of **only having adult hands in the dataset** is that the model may be less adept at recognizing signs when presented by children, potentially limiting its applicability in real-world scenarios.

Planning your project

1. **Data Collection:**
    - *Description:* Photograph seven individuals performing the signs in different locations.
    - *Duration:* 21 hours
    - *Team Contribution:* Each team member contributes 7 hours.
2. **Data Preprocessing:**
    - *Description:* Organize, label and clean the collected images, addressing issues like motion blur and incorrect gestures.
    - *Duration:* 15 hours
    - *Team Contribution:* Karl – 8h, Priit – 5h, Hendrik – 2h.
3. **Model Training:**
    - *Description:* Use Mediapipe library for training the sign language recognition model.
    - *Duration:* 21 hours
    - *Team Contribution:* Split tasks for training and parameter tuning, with each team member contributing 7 hours.
4. **Model Evaluation:**
    - *Description:* Assess the model's performance using a validation dataset and fine-tune as needed.

    - *Duration:* 12 hours
    - *Team Contribution:* Priit – 7h, Hendrik – 4h, Karl – 1h
5. **User Interface and Poster:**
    - *Description:* Build an interface in Jupyter for capturing and processing images of signs, including real-time visualization of recognized signs. Create a poster and visualizations for the poster session.
    - *Duration:* 24 hours
    - *Team Contribution:* Karl – 7h, Priit – 5h, Hendrik – 12h.

Methods and Tools:

- *Data Collection:* Laptop webcam.
- *Data Preprocessing:* Image processing libraries (OpenCV) and manual validation.
- *Model Training:* Mediapipe library, TensorFlow.
- *Model Evaluation:* Metrics like accuracy, precision, and recall.
- *User Interface Development and poster:* OpenCV, LaTeX, Mathplotlib.