Illinois Institute of Technology

College of Computing

Static Hand Gesture Recognition

Author:
Norah Khalaf Alotaibi

Supervisor:
Yan Yan

Problem description

Hand gestures form the main component of sign language vocabulary. Hand poses mainly constitute finger spelling of sign language vocabulary, which is used letter by letter to sign names, place names, ages, numbers, dates, years, and words that do not have predetermined signs. Hand gestures can be static or dynamic. Thus, automatic hand position recognition has been a hot research area, and there are many works on the same approach using vision-based and electronic signal-based methods. Vision-based techniques are more user-friendly and convenient than others when considering the complexity of the data acquisition process.

The difficulty of recognizing static hand gestures in the visual approach depends on many factors. First, the most important is the difficulty of detecting and segmenting hands from images taken with a complex background. Second, it is the difficulty of deriving good features that distinguish geometric differences in the appearance of the same hand position shown by different individuals and many challenges that Cause machine learning techniques to fail to recognize hand position from images / or videos.

Related research

Most existing works on vision-based hand gesture recognition mainly follow the steps of classical pattern analysis that go through image/video pre-processing, feature extraction, and classification. Cao et al.[1] proposed a hand posture recognition approach with heterogeneous feature fusion and obtained a recognition accuracy of 0.9916 on the publicly available Triesh data set. They trained a multiple kernel support vector machine classifier with feature vectors obtained by combining the shape context feature, pyramidal HOG feature, and SIFT-based bag of features for recognizing the various static hand gestures. Pugeault et al.[5] presented a multiclass random forest classifier to recognize 24 static signs in ASL (American Sign Language) alphabet with the features extracted through Gabor filters in four levels. They claimed a recognition rate of 0.49

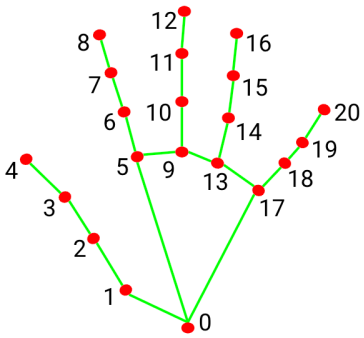
Despite promising results, the classic methods failed to recognize hand position in real-time scenarios. The main reason for this failure is the failure of traditional machine learning techniques to infer and recognize patterns from the input data accurately. The most critical challenges faced by the approach to identifying the position of the hand are the following:

- Difficulty detecting and segmentation hands from images taken with a complex background.
- Difficulty eliciting robust features that characterize geometric differences in the appearance of the same hand posture shown by different individuals.
- Having many categories of gestures with very little difference between the categories is another complex problem, especially in the automatic recognition of sign language.

Proposed work

This project aims to take advantage of the Mediapipe framework [3], provided by Google, which contains several ML solutions, including Hands Mediapipe [2], which is a high-resolution solution for real-time hand and finger tracking to infer (21) 3D landmarks of the hand, As in Figure (1).

This project proposes to build a hand descriptor that produces a set of features describing the hand, depending on the parameters of the hand 21 (joints). Such as, creating descriptors



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

Figure 1: Hand landmarks (1)

represents the relationship of a finger with other fingers (intersection, parallelism, etc.) and a relationship between a joint with a finger or with another joint, in addition to the direction of the hand and many hand descriptors. Hand descriptors will be represented through relationships and mathematical equations such as the angle between two vectors and the ratio of the intersection of two vectors, etc. What distinguishes the resulting descriptors from the proposed hand descriptor is that it is not affected by the hand's shape, size, or color because most of these descriptors are angles and proportions. Also, the proposed hand descriptor can be used in any project because the hand is not specific to sign language.

The project also aims to take advantage of Ensemble learning methods to increase the accuracy of the classification process.Training and testing will be conducted on many data sets, including American Sign Language (ASL) characters[4]. Figure (2) The phases of the proposed project.

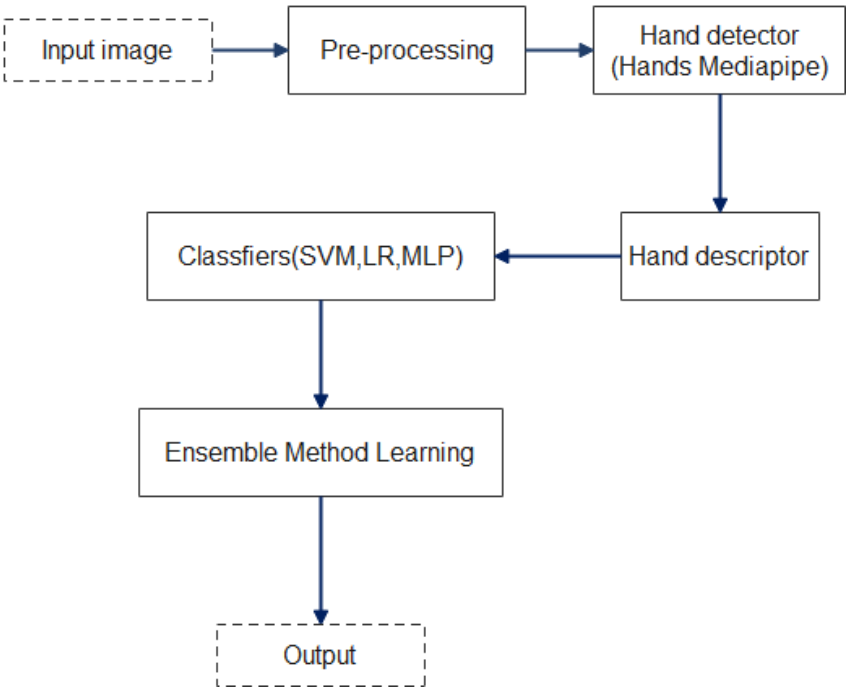


Figure 2: Phases of the proposed project (2)

Preliminary plan

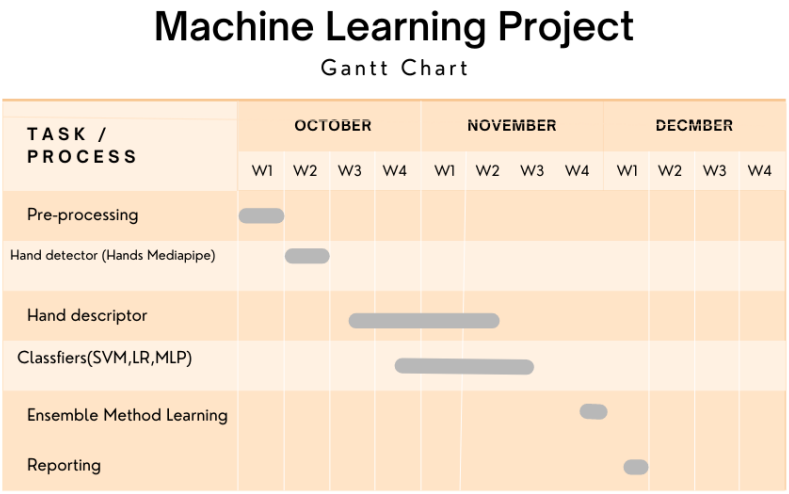


Figure 3: Preliminary plan (3)

References

[1] Jiangtao Cao et al. "Hand posture recognition based on heterogeneous features fusion of multiple kernels learning". In: *Multimedia Tools and Applications* 75.19 (2016), pp. 11909–11928.

[2] Google. *MediaPipe Hands solution*. URL: https://google.github.io/mediapipe/solutions/hands.html.

[3] Google. *MediaPipe Main*. URL: https://google.github.io/mediapipe/.

[4] kaggle. *ASL alphabets images for classification*. URL: https://www.kaggle.com/datasets/debashishsau/aslamerican-sign-language-aplhabet-dataset.

[5] Nicolas Pugeault and Richard Bowden. "Spelling it out: Real-time ASL fingerspelling recognition". In: *2011 IEEE International conference on computer vision workshops (ICCV workshops)*. IEEE. 2011, pp. 1114–1119.