American Sign Language

Abhinuv Pitale
Pranavi Rambhakta
Varun Kumthekar
Omkar Dhande

Sign Language

Sign language is a system of communication using visual gestures and signs, as used by deaf people.

Goal

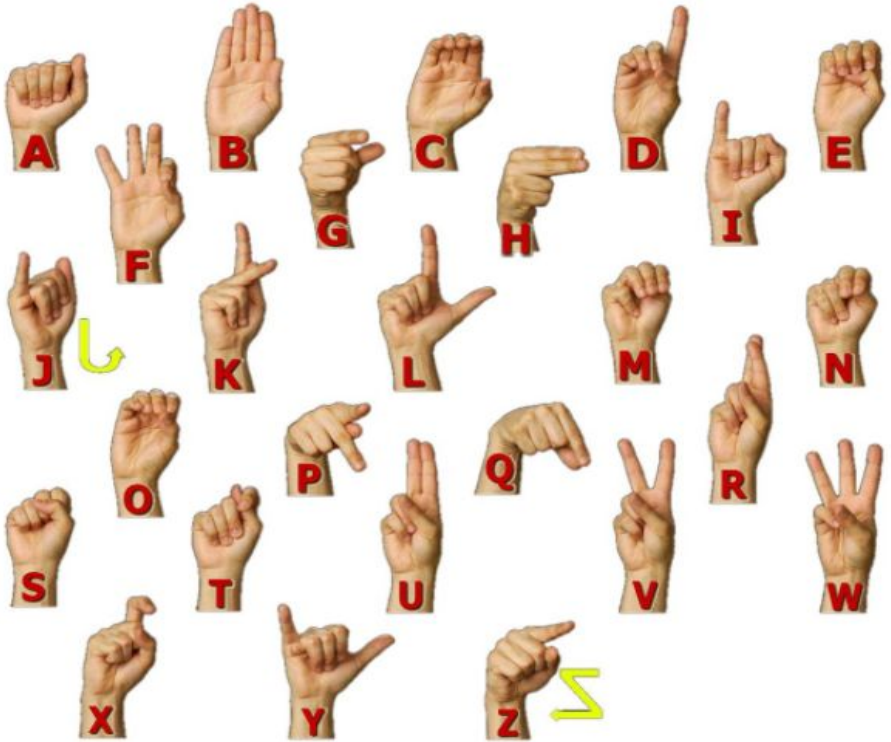
Construct words for detecting American Sign Language(ASL) Fingerspelling gestures from a video input.

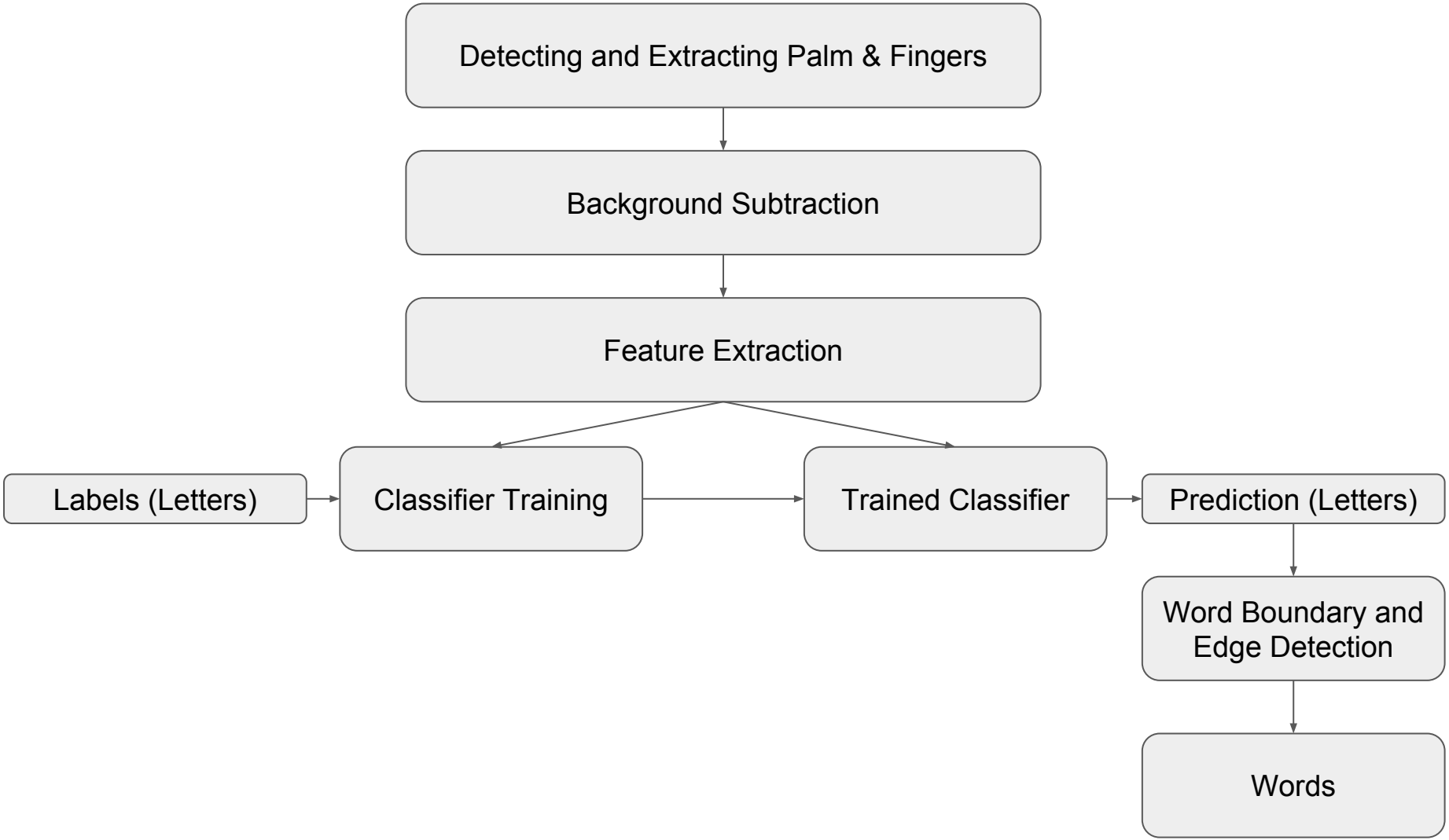
- For hearing impaired people sign language is the only way of communicating and expressing their emotions and thoughts. It is often difficult to understand this language and they have to keep translator along with them

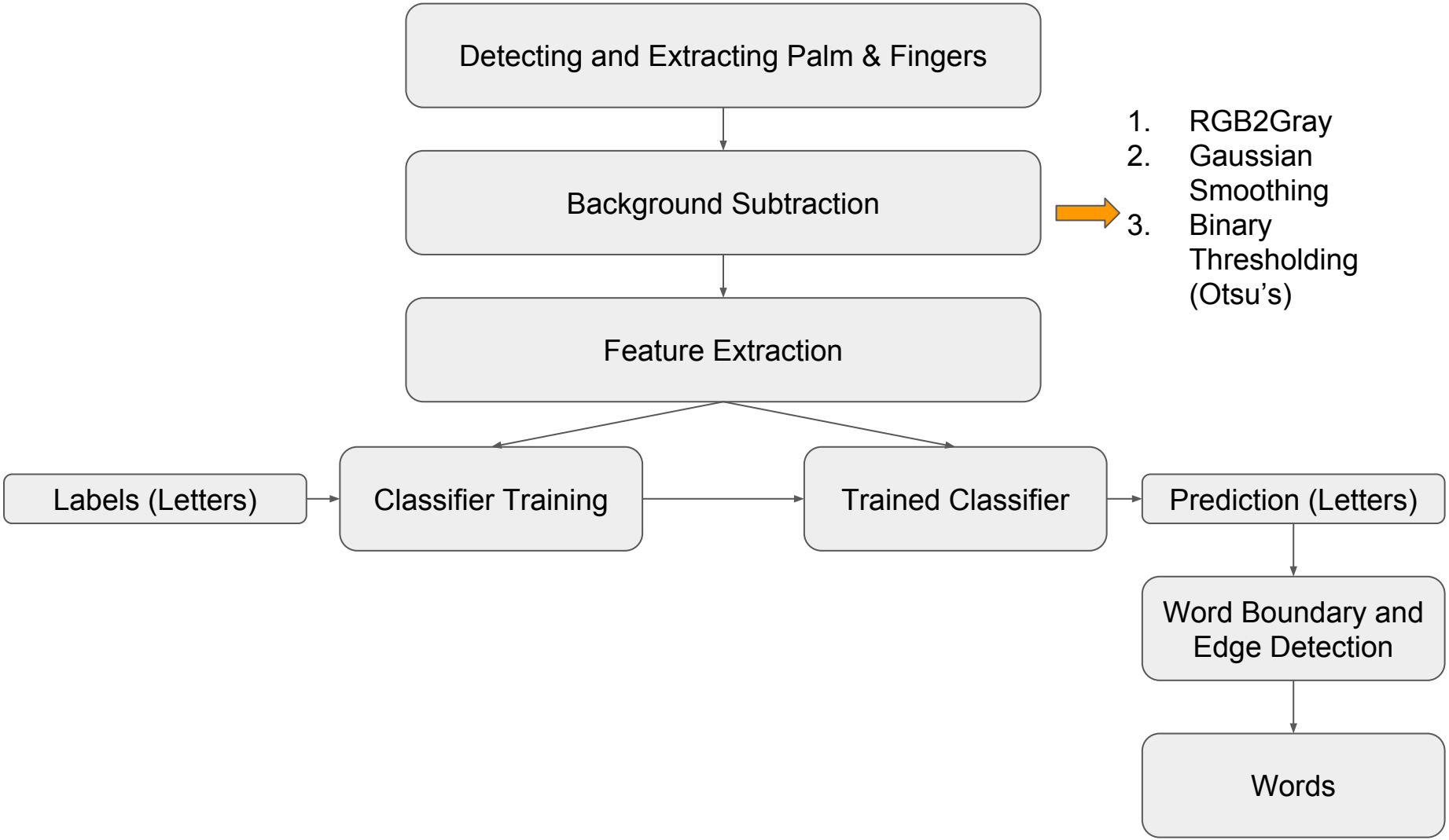
- Aim of developing hand gesture recognition system is to establish an interaction between human and computer and hence recognize the hand gesture automatically

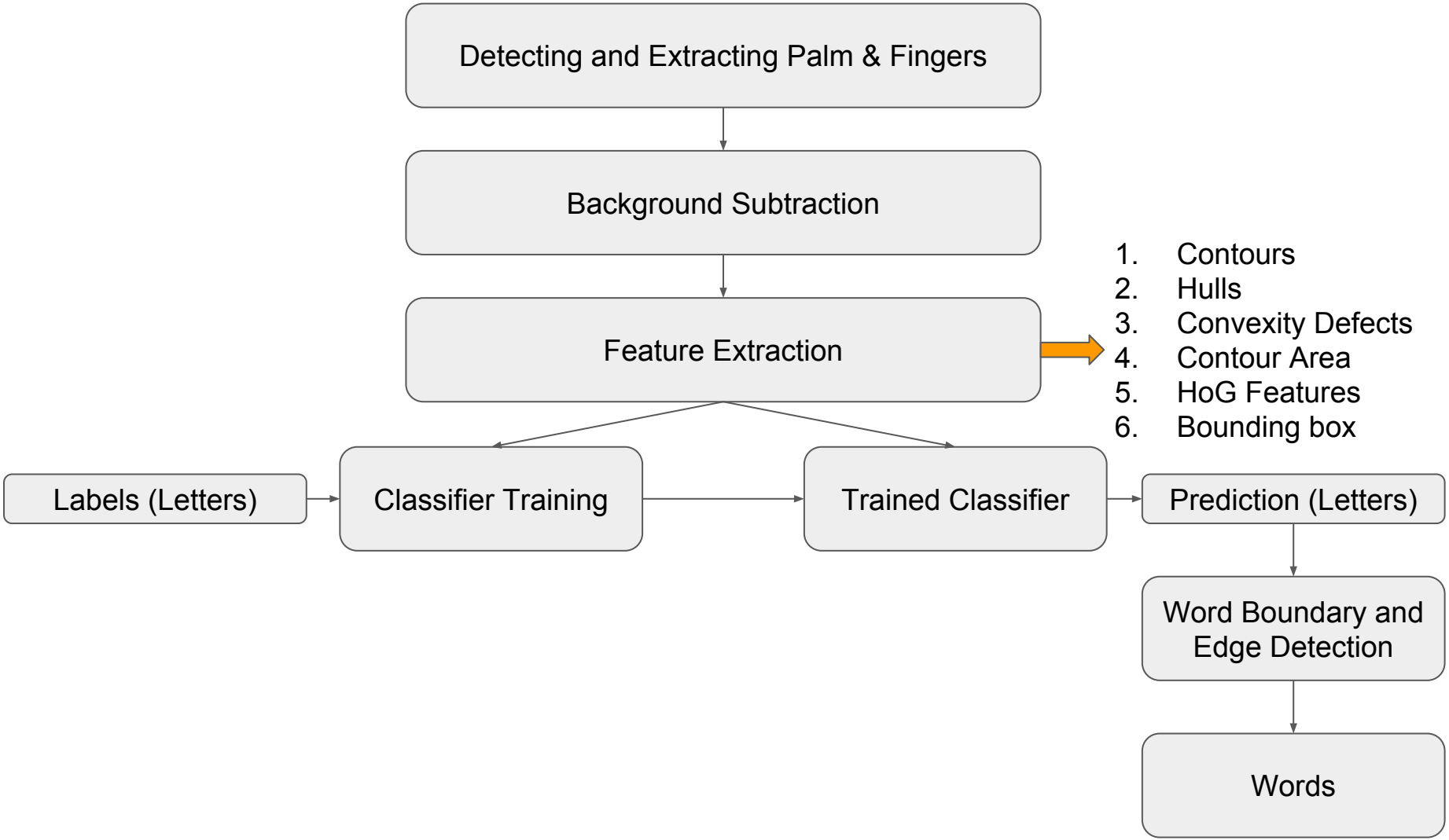
- By automatic hand gesture recognition we will be able to remove the boundaries between hearing and deaf people
- FingerSpelling -> This is a way of actually spelling out the letters of a particular words.

Representing Alphabets using Signs :

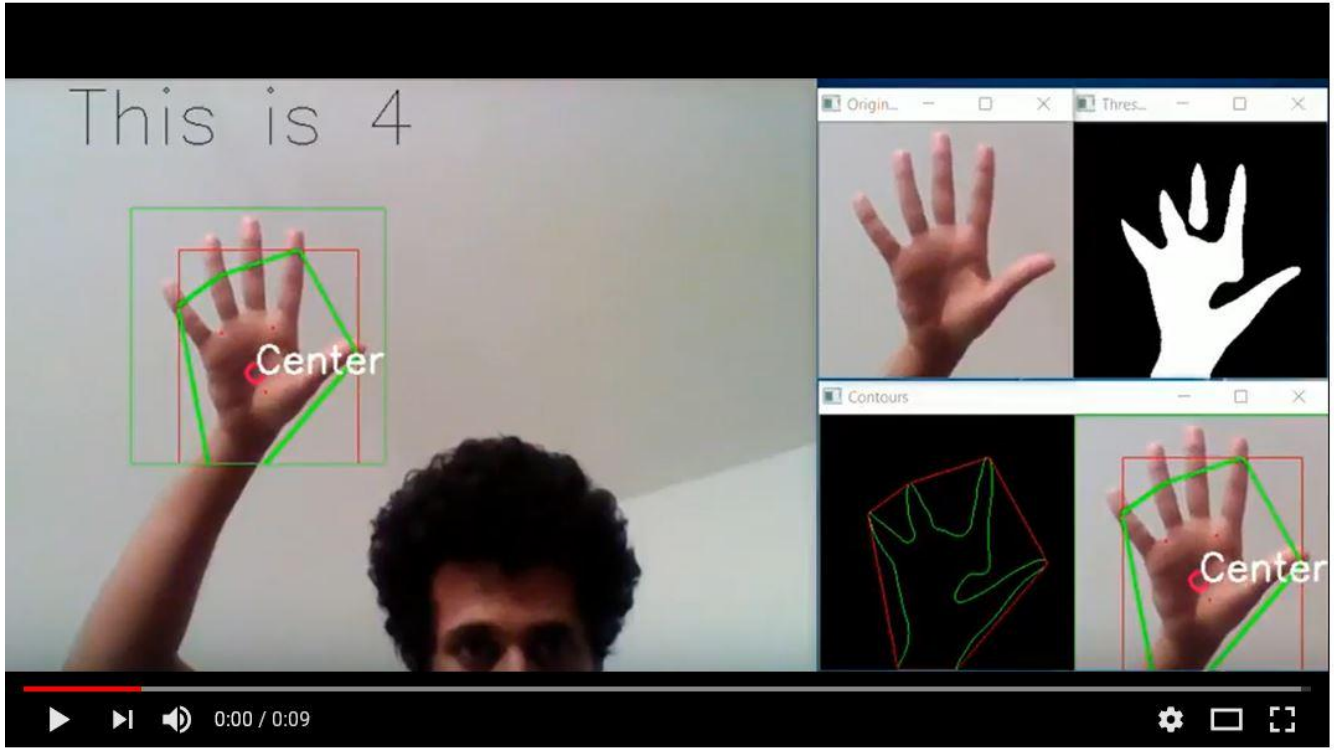



Detecting and Extracting Palm & Fingers
Background Subtraction
Feature Extraction
Labels (Letters)
Classifier Training
Trained Classifier
Prediction (Letters)
Word Boundary and Edge Detection
Words


Detecting and Extracting Palm & Fingers
Background Subtraction
1. RGB2Gray
2. Gaussian Smoothing
3. Binary Thresholding (Otsu's)
Feature Extraction
Labels (Letters)
Classifier Training
Trained Classifier
Prediction (Letters)
Word Boundary and Edge Detection
Words

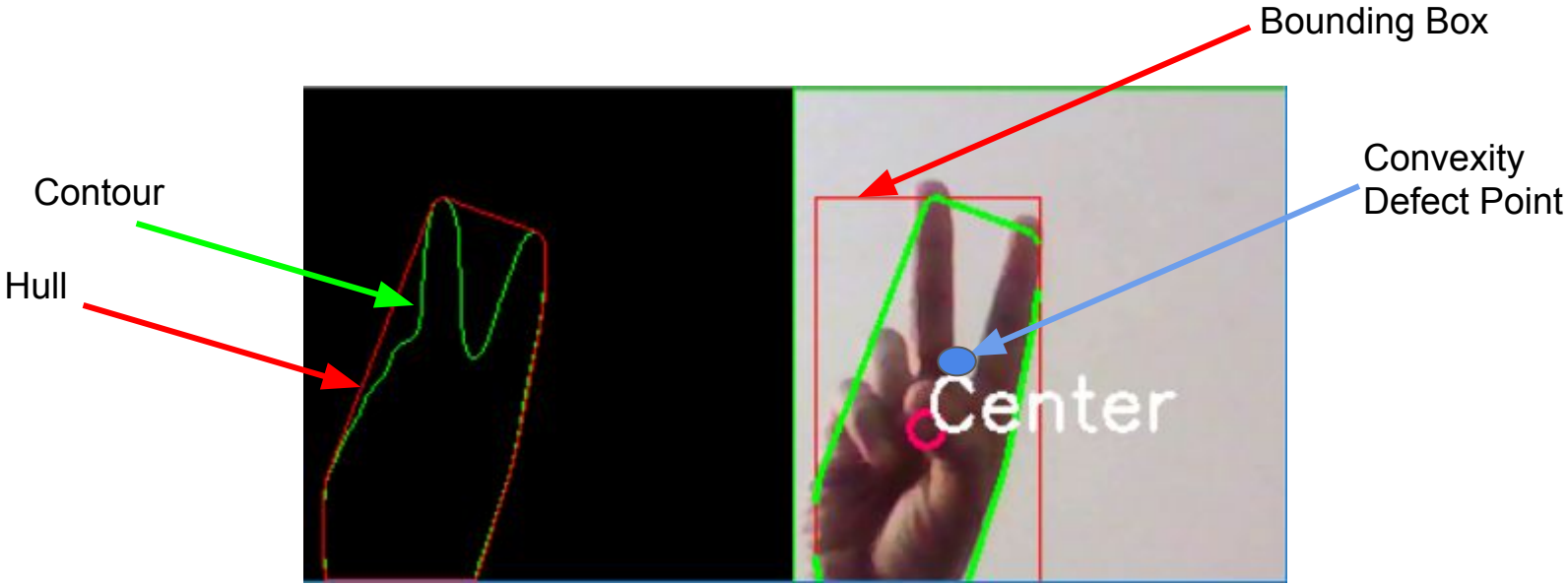

Detecting and Extracting Palm & Fingers
Background Subtraction
Feature Extraction
1. Contours
2. Hulls
3. Convexity Defects
4. Contour Area
5. HoG Features
6. Bounding box
Labels (Letters)
Classifier Training
Trained Classifier
Prediction (Letters)
Word Boundary and Edge Detection
Words

Features Video



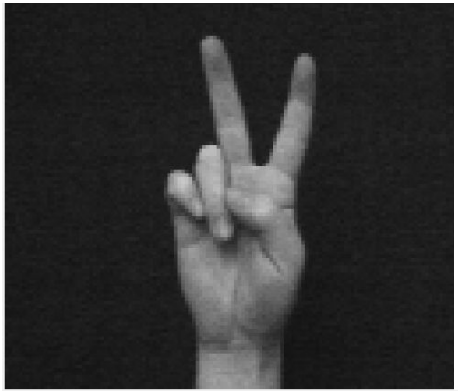
https://www.youtube.com/watch?v=AnV3OH_3D0E&feature=youtu.be

Feature Extraction



- **Contour** - Coordinates of the outline of the hand.
- **Hull -** Coordinates of the bulging points in the contour with maximum area (Here the extremities of the contour)
- **Convexity Defects -** List of coordinates which are part of the convex portion of the contour.
- **Bounding Box-** Coordinates of the 4 corners of the bounding rectangle.
- **HoG** of the thresholded image.

Existing Dataset Used - Triesch



Triesch dataset contains letters a,b,c and v against 3 different backgrounds as shown above.

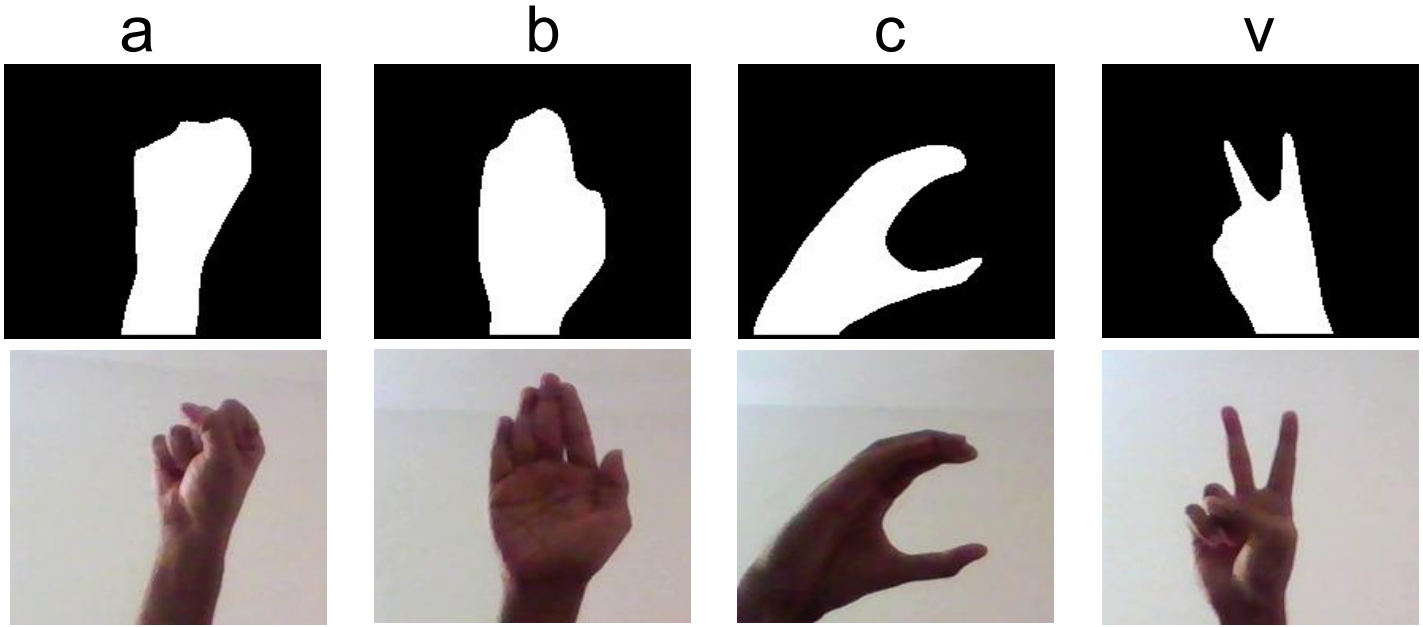
Why create another dataset?

- The created dataset images of letters recorded at different angles and at different distances wrt the camera.
    - This variation is introduced in order to introduce tolerances for orientation and scale.

Data Augmentation

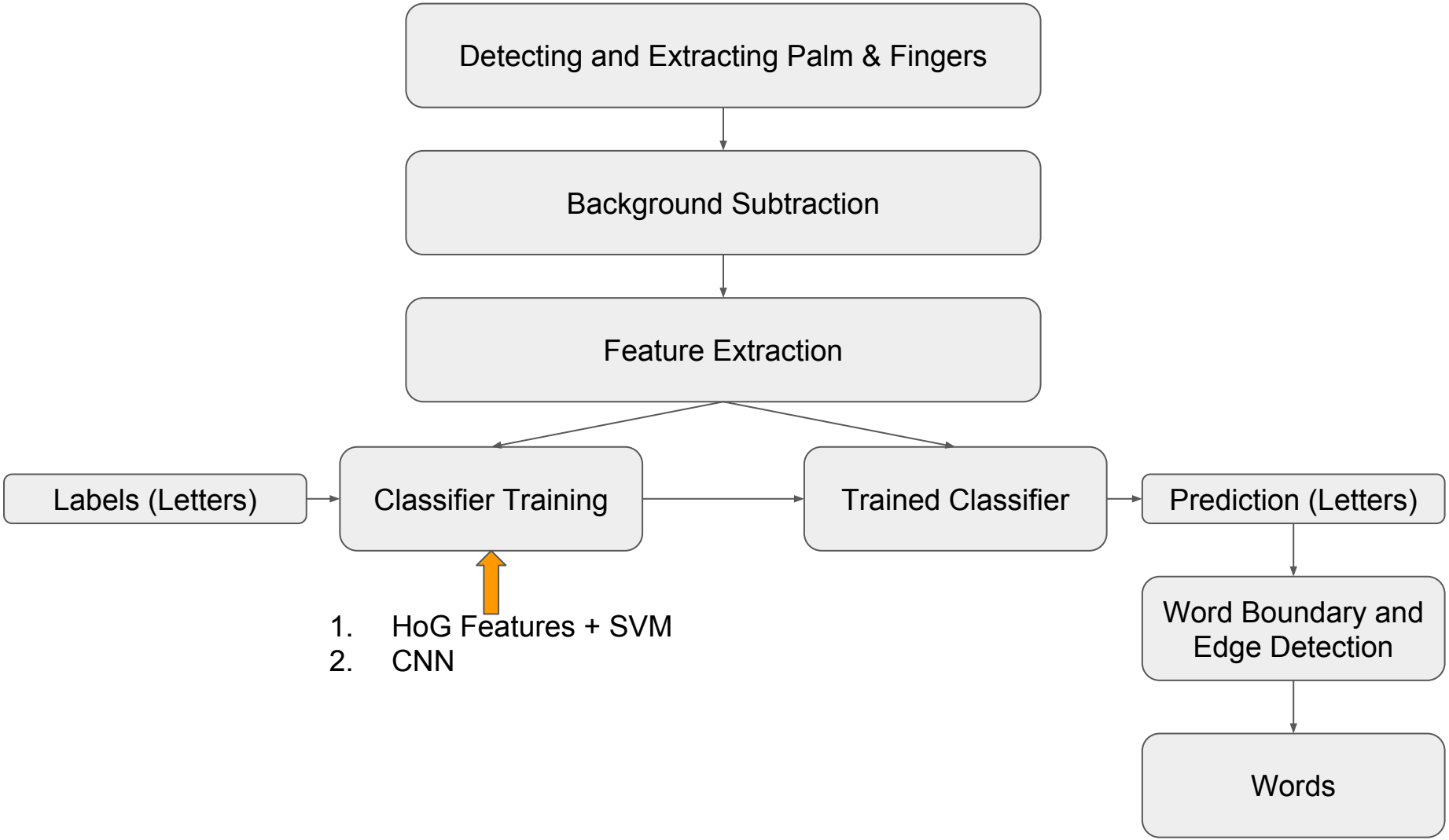
- Left-Right Hands -> We generate a set of rotated images from the existing images to help train our model to make it more robust.
- Tilt/Zoom -> Generated set of images with different levels of tilt and zoom to increase the variety of the training dataset.

Dataset Created

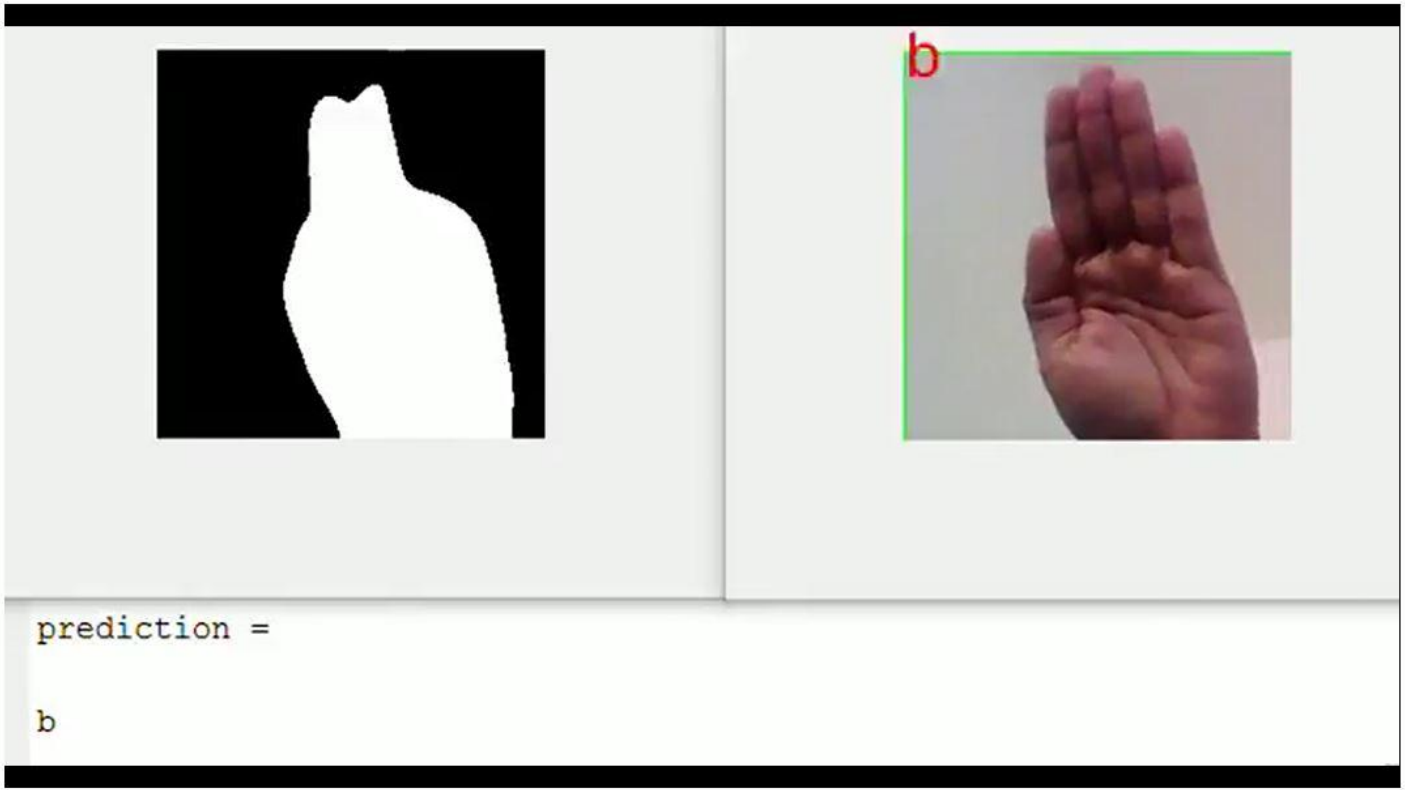


- This dataset contains 30 images for each of the above mentioned letters consisting of 10 images of 3 different persons.

- In addition to it there are
  - 30 images of Augmented Data with flips per letter
  - 30 images of rotated/zoomed Data per letter


Detecting and Extracting Palm & Fingers
Background Subtraction
Feature Extraction
Labels (Letters)
Classifier Training
Trained Classifier
Prediction (Letters)
1. HoG Features + SVM
2. CNN
Word Boundary and Edge Detection
Words

Demo



https://youtu.be/GhI2VwPll0k

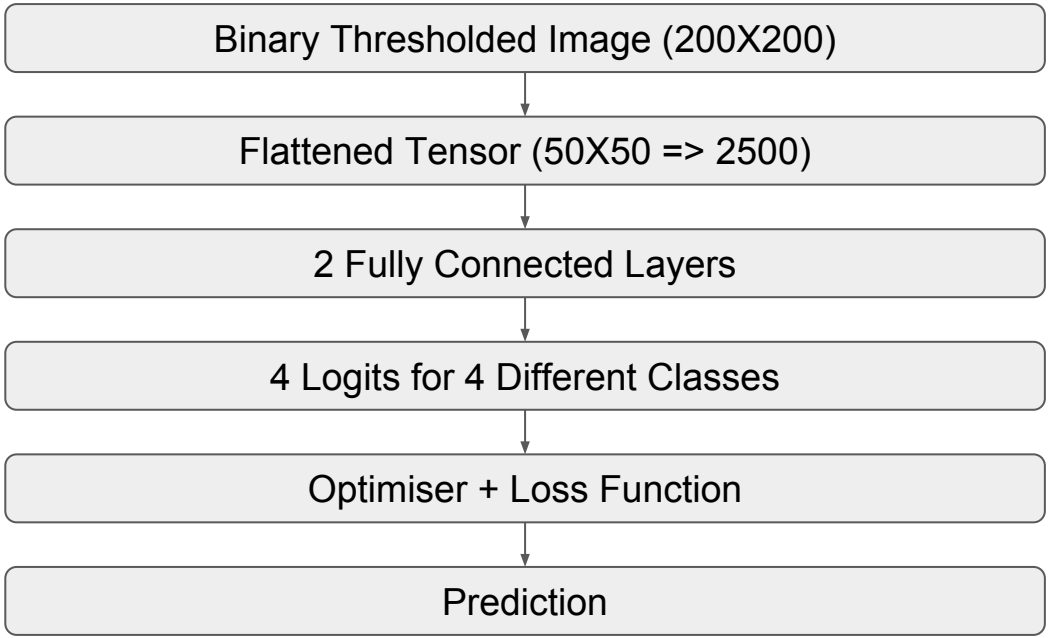
Preliminary Findings

1. HoG + SVM works really well on the Treisch dataset as it is nicely centered. But, when the model trained on this dataset is used in a live video, the predictions fail.
2. HoG + SVM has slightly worse accuracy for in-set when trained using our dataset. But the model trained using this dataset works a lot better with live video.

	In-Set Accuracy	Live Video Accuracy
Triesch	87%	~40%
Dataset Created	80%	~75%

Preliminary Findings

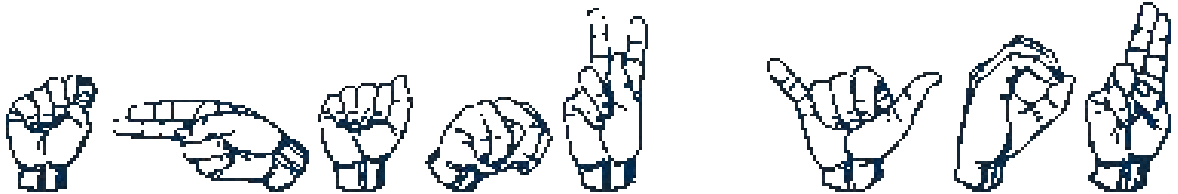
3. CNN has a better in-set accuracy(~90%) when trained with our Dataset.



Further work

- Till now we have built the dataset and implemented gesture recognition for words ‘a’, ‘b’, ‘c’ and ‘v’. We will extend this dataset to all the static characters

- We also plan to detect change in characters and form words from them. This will be done by identifying change in character as a gradient



T H A N K Y O U