Realtime Static Hand Gesture Recognition

Roy Amante A. Salvador

CS282 Mini Projects Presentation

roy.salvador@gmail.com

May 17, 2016

Outline

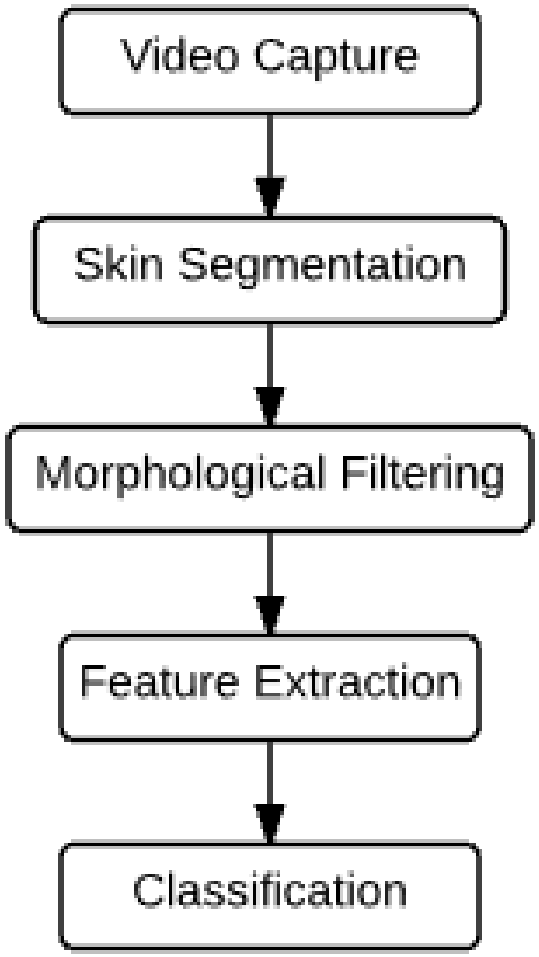
Hand Gesture Recognition

- Hand Gesture Recognition is a Computer Vision problem aimed in algorithmically interpretting and classifying hand gestures in video or images
- Gestures in general are part of human communication
- Can be categorized as static and dynamic
- Using it as interface in technological devices has been a developing trend

Hand Gesture Recognition Application

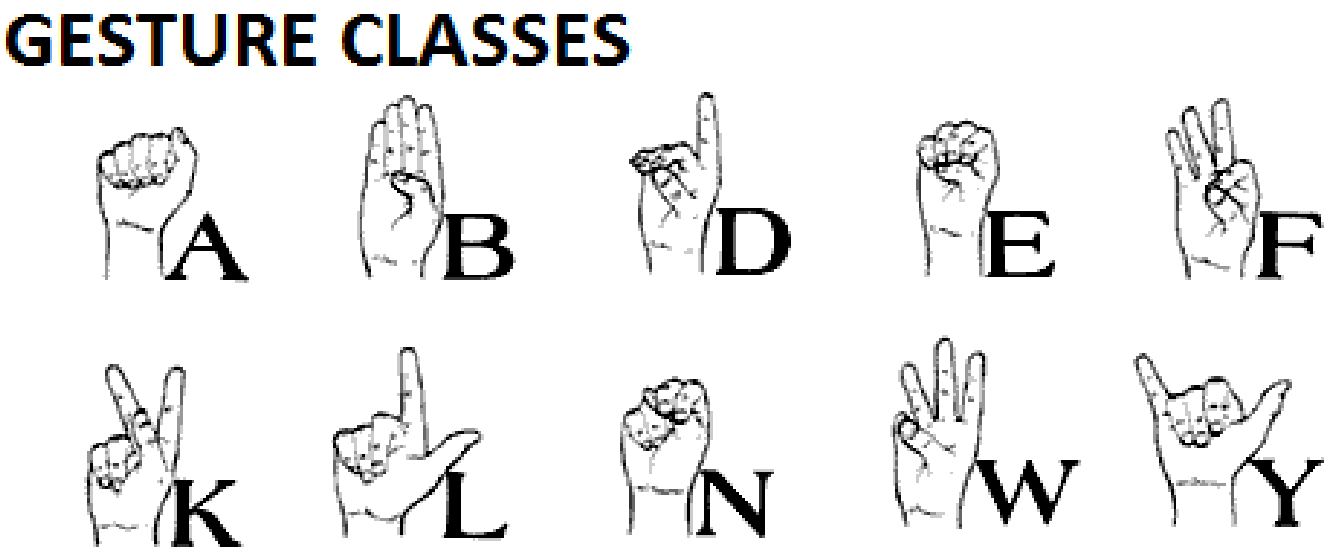
[1]Hand gesture images taken from The Gesture Interface: A Compelling Competitive Advantage in the Technology Race

System Architecture Overview



Dataset

- Hand gesture classes consist of ten static hand postures of the American Sign Language (ASL)



[1]Hand gesture images taken from http://www.linguistics.uconn.edu/asl/

Dataset

- Eight people are asked to hold static gestures using their left hand for more than ten seconds inside a controlled region within the entire frame
- Simple background specifically does not contain skin colored areas
- Asked to move around the frame and also towards and backwards the camera.
- Resulting video clips are sampled into images and tagged with the appropriate hand gesture class.
- Half of the clips for each class are used as training set while the other remaining half as test set.

Skin Segmentation

- Skin Segmentation is the process of locating the skin-like region of the image
- Image is converted to YC_bC_r color space

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Ignoring brightness Y information reduces the effect of uneven illumination
- Skin mask is formed using thresholding

$$77 \leq C_b \leq 127 \quad \textit{and} \quad 133 \leq Cr \leq 173$$

Morphological Filtering

- Morphological image processing pursues the goal of removing imperfections by accounting for the form and structure of the image
- Used an elliptic structuring element inscribed in a 5 x 5 rectangle
- Two iterations of Erosion followed by two iterations of Dilation operation
- Mask is smoothened by Gaussian Blurring before applying to the original image

Hand Region Extraction

- Contours of the masks are extracted
- The contour with the largest size is determined to be the biggest object in the frame and hence is assumed to be the hand
- Biggest contour is bounded by a box and becomes the hand region

Preprocessing Pipeline

(a) Original Hand Frame

(b) Conversion to $YCbCr$ Color Space

(c) Mask Creation Using Skin Pixel Thresholding

(d) Two iterations of Erosion

(e) Two iterations of Dilation and Gaussian Blurring

(f) Resulting segmented skin region bounded by a green box.

Histogram of Oriented Gradients (HOG)

- Histogram of Oriented Gradients (HOG) is a popular feature descriptor used in computer vision particularly in object detection and recognition
- First introduced and used in Pedestrian Detection in static images [Dalal, Triggs 2005]
- Counts occurrences of gradient orientation in localized portions of an image
- Differs from SIFT since HOG is computed on a dense grid of uniformly spaced cells

HOG Extraction [Scikit Image]

- Segmented hand region is turned into a grayscale image and resized to n by n pixels
- Image global normalization by applying Gamma compression ($\gamma < 1$) to reduce the effect of changes in illumination and shadowing.

$$V_{out} = AV_{in}^{\gamma}$$

- First order image gradients in x and y axes are which contour, silhouette and some texture information.

HOG Extraction [Scikit Image]

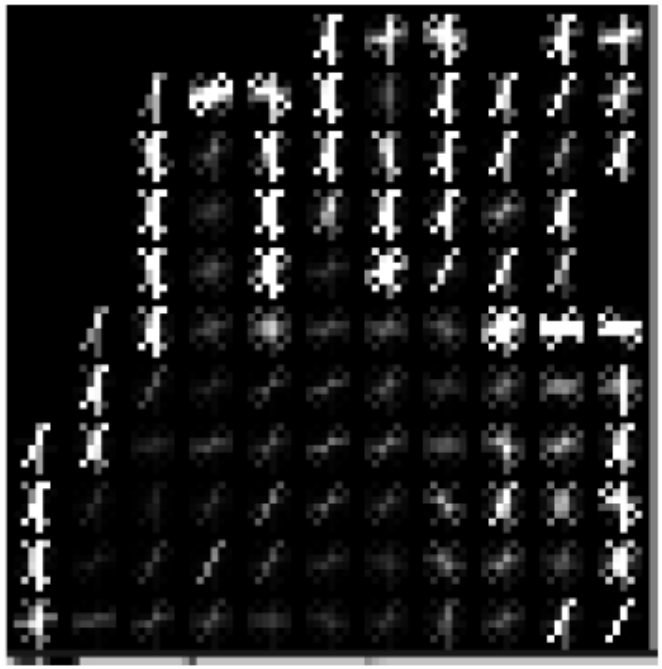
- The image window of size n by n is divided into blocks which are spatial regions within the image.
- Blocks are composed of 3 x 3 cells and each cell is made up of 8 x 8 pixels.
- Gradient magnitude and orientation is given by

$$||\nabla f|| = \sqrt{\left(\frac{\delta f}{\delta x}\right)^2 + \left(\frac{\delta f}{\delta y}\right)^2}$$

$$\theta = \tan^{-1}\left(\frac{\delta f}{\delta y} \div \frac{\delta f}{\delta x}\right)$$

HOG Extraction [Scikit Image]

- Gradient in cells are partitioned into 9 bins
- For each cell, a local 1-D histogram of gradient over all pixels in the image are computed



HOG Extraction [Scikit Image]

- An energy measure is accumulated over the cells in the block. This value is then used to normalize each cell in the block.
- Normalization further offers robustness to shadowing, illumination and edge contrast.
- The collection of HOG descriptors from all the blocks of a dense overlapping grid of blocks covering the window are flattened into a feature vector

Classification

- Support Vector Machine (SVM) is used as classifier
- A linear classifier which aims on maximizing the distance of near miss examples called Support Vectors from decision hyperplanes.
- Uses some kernel ϕ to map input to a richer dimensional space
- One vs All strategy is employed for Multiclass classification. A binary model with linear kernel is trained for each class and is fitted against all other classes.

Descriptor Size and Training Time

TABLE I: Descriptor Size and Training Time

Hand Region Frame Size (n x n)	Descriptor Size	Training Time (sec)
24	81	1
32	324	2
40	729	3
48	1296	5
56	2025	7
64	2916	11
72	3969	18
80	5184	23
88	6561	29
96	8100	36

Recognition Results

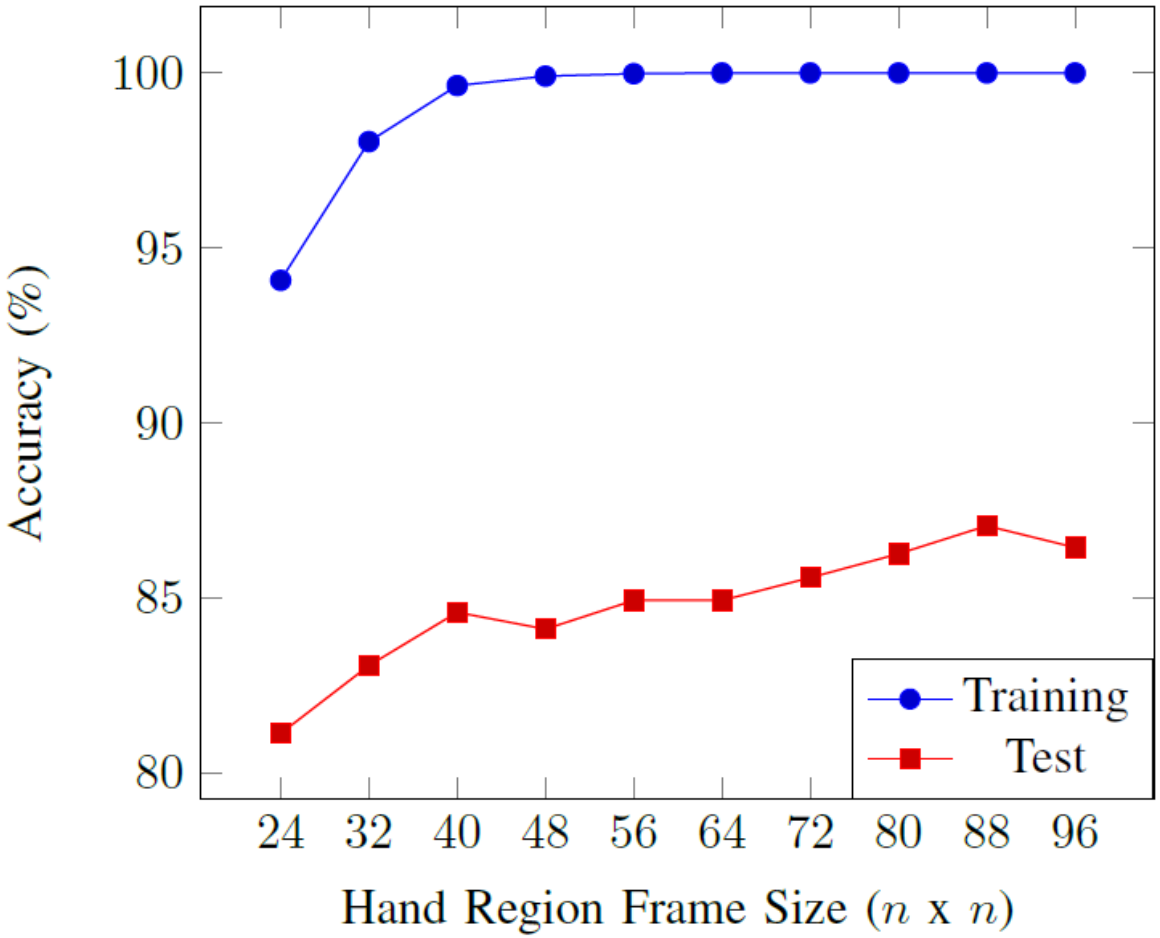


Fig. 5: Hand Region Frame Size vs Accuracy (%)

Recognition Results

TABLE III: Accuracy of the Hand Gesture Classes

Gesture	Accuracy (%)
A	82.14%
B	92.51%
D	90.65%
E	86.51%
F	93.62%
K	74.23%
L	94.71%
N	62.74%
W	96.46%
Y	93.97%
Overall	87.07%

Recognition Results

TABLE II: Hand Gesture Classification Confusion Matrix

		Ground Truth									
		A	B	D	E	F	K	L	N	W	Y
Classification	A	377	0	0	41	0	0	3	70	0	3
	B	0	494	17	0	16	11	1	6	0	4
	D	6	6	446	5	1	29	17	18	1	9
	E	32	21	16	468	13	0	0	34	3	3
	F	0	1	0	0	558	1	0	2	4	5
	K	2	1	0	0	3	409	2	0	8	5
	L	0	6	13	0	2	30	448	0	0	0
	N	42	0	0	17	3	0	0	261	0	0
	W	0	1	0	0	0	62	2	25	436	0
	Y	0	4	0	10	0	9	0	0	0	452

Conclusion

- Presented a successful implementation of a Hand Gesture Recognition system using the Histogram of Oriented Gradients (HOG) feature trained on a multi-class SVM with high recognition rate
- Possible improvement on this work is to apply techniques to lower dimensionality of the feature vector such as Principal Component Analysis (PCA)
- Can be used as baseline performance for comparing other approaches such as Deep Learning

References

Dalal, D., Triggs, B (2005)

Histograms of oriented gradients for human detection

IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2 (2005).

Prashan Premaratne (2014)

Human Computer Interaction Using Hand Gestures

School of Elec., Comp. and Telecom. Eng. The University of Wollongong

Springer Science+Business Media Singapore 2014 (2014).

Python Scikit Image

Histogram of Oriented Gradients

`http://scikit-image.org/docs/dev/auto_examples/plot_hog.html` (2016).

Demo