**Gesture Recognition Using Deep Learning**

Malhar Patwari (A20410420), Herick Jayesh Asmani (A20399752)

**Problem Description:**

You must have wondered what it would be like to control the TV, computer or any other device at your home with just a wave of your hand. These sci-fi dreams are becoming a reality with the advancement in the field of Deep Learning and their application in Computer Vision. Such vision-based hand gesture recognition also finds its application in advanced driver assistance systems (ADASs) which enables the driver to interact with the vehicle without affecting their concentration. Based on the above motivations, our aim is to build a system such that Given a real time video sequence of a hand gesture, the system should be able to interpret what that particular hand gesture signifies. However, the video-based gesture recognition problem is not trivial and is challenging due to: the intra and inter-persons variations in human hand gesture motion; inter-person variations in the shape and size of the human hand; illumination variations; and background noise.

**Method:**

We will address these issues by utilizing the Deep Learning framework such as the Long – term Recurrent Convolution Network (LRCN) which is based on LSTM and CNN. The Deep Learning framework will be used to extract the representative frames from the video sequence and classify the gesture. We will extract the representative frames from the video sequence using the deconvolution neural network (Deconvnet), a semantic segmentation framework. To extract the representative frames, we will utilize novel tiled image patterns and novel tiled binary patterns. The tiled image and binary patterns are represented in the form of a single image with multiple non-overlapping blocks of smaller images. The tiled image pattern will be used as the input to the Deconvnet, to encapsulate the video-based hand gesture motion within a single image. On the other hand, the tiled binary pattern will function as the output of the Deconvnet, and contain the information of the representative frames. The tiled image patterns are generated from the video sequence. While, the tiled binary pattern for the Deconvnet training will be generated from the training video sequence using the sparse modelling representative frame extraction algorithm (SMRF) as shown in the reference paper. The SMRF algorithm identifies the representative frames directly from the video sequence. Given the extracted representative frames, the LRCN will then be used to classify the gesture video sequence. The LRCN will be trained with the extracted representative frames and their corresponding gesture labels.

**Dataset:**

The dataset to be used for this project is CVRR VIVA dataset. This is a hand gesture dataset which was designed in order to study natural human activity under difﬁcult settings of cluttered background, volatile illumination, and frequent occlusion. The dataset was captured using a Kinect device under real-world driving settings and have a resolution of 115 x 250 pixels. It contains 885 intensity and depth video sequences of 19 different dynamic hand gestures performed by 8 subjects.

**Reference Papers:**

1. Deep Learning-Based Fast Hand Gesture Recognition Using Representative Frames. [Vijay John](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Vijay%22&searchWithin=%22Last%20Name%22:%22John%22&newsearch=true); [Ali Boyali](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Ali%22&searchWithin=%22Last%20Name%22:%22Boyali%22&newsearch=true); [Seiichi Mita](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Seiichi%22&searchWithin=%22Last%20Name%22:%22Mita%22&newsearch=true); [Masayuki Imanishi](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Masayuki%22&searchWithin=%22Last%20Name%22:%22Imanishi%22&newsearch=true); [Norio Sanma](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Norio%22&searchWithin=%22Last%20Name%22:%22Sanma%22&newsearch=true). [2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7794373)
2. Hand gesture recognition using deep learning. [Soeb Hussain](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Soeb%22&searchWithin=%22Last%20Name%22:%22Hussain%22&newsearch=true); [Rupal Saxena](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Rupal%22&searchWithin=%22Last%20Name%22:%22Saxena%22&newsearch=true); [Xie Han](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Xie%22&searchWithin=%22Last%20Name%22:%22Han%22&newsearch=true); [Jameel Ahmed Khan](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Jameel%20Ahmed%22&searchWithin=%22Last%20Name%22:%22Khan%22&newsearch=true); [Hyunchul Shin](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Hyunchul%22&searchWithin=%22Last%20Name%22:%22Shin%22&newsearch=true). [2017 International SoC Design Conference (ISOCC)](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8362894)