



Dynamic Hand Gesture Recognition

Problem Statement

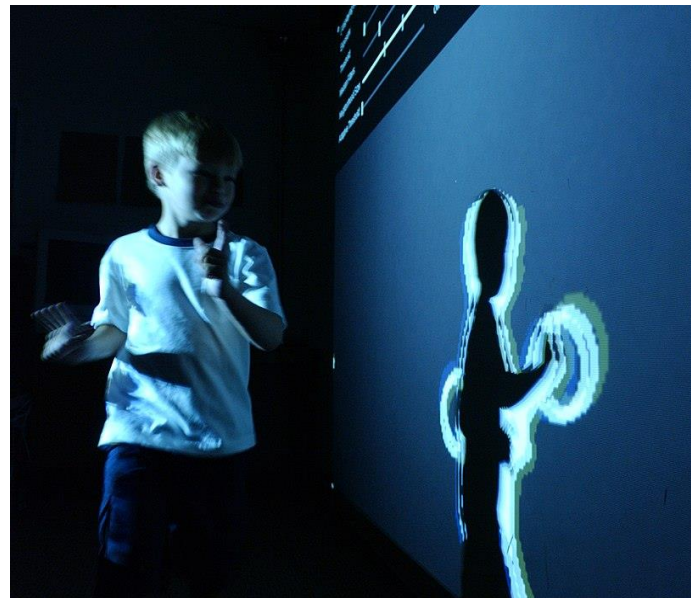
- ❖ To study the various techniques of recognition of hand gestures so as to explore a way of efficient computing other than traditional, physical input devices; more so through dynamic gesture recognition..

LET'S GET STARTED



Motivation

- Hand gestures are a natural way for human-computer interaction.
- Vision-based dynamic hand gesture recognition has become a hot research topic due to its various applications.
- With this project, we intend to explore the various techniques of hand gesture recognition by computer vision and how they work, through study of research in related concepts.
- We also intend to elaborate on a technique that conforms to our purposes and requirements, i.e. short-term sampling neural networks.



Literature Review

DYNAMIC HAND GESTURE RECOGNITION BASED ON 3D CONVOLUTIONAL NEURAL NETWORK MODELS – Zhang, Wang

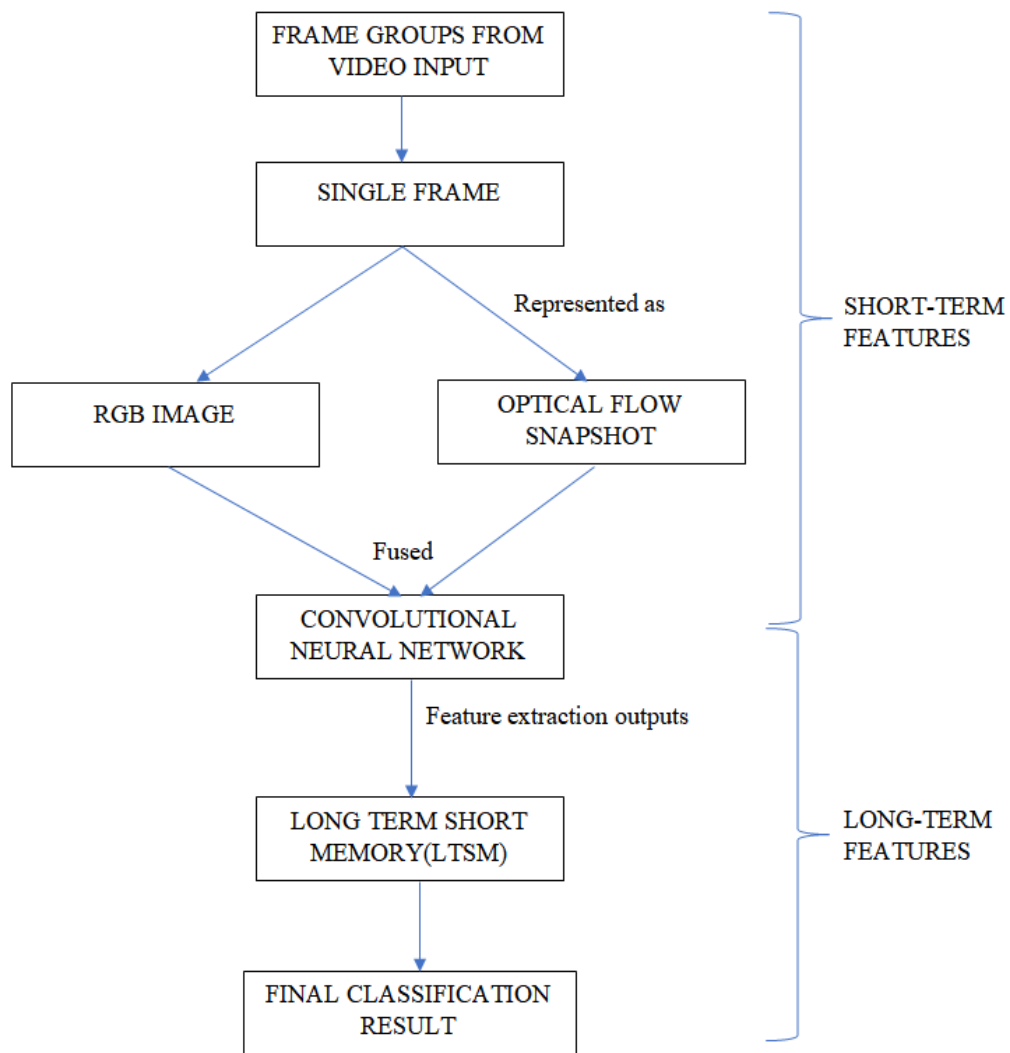
- This study applies a deep learning method to recognize hand gestures.
- Designed a 3-dimensional convolutional neural network model(CNN)
- 3D Convolution neural network can be seen as a variant of 2D convolution neural network extending 2dimension filter into 3 dimensions. This 3D filter shall slide in 3 directions to extract low-level features and its output's shape is a 3dimension space like a cuboid
- Uses the Jester V1.0 hand gesture dataset to train the model.
- According to the result of the training experiment, it got an average accuracy of 90%.

TEMPORAL SEGMENT NETWORKS FOR ACTION RECOGNITION IN VIDEOS – Limin Wang, Zhe Wang, Xiong, Qiao, Lin, Tang, Gool

- Proposes a two-stream convolutional neural network
- In this architecture, spatial features and temporal features are learned by two separate networks and are then fused using support vector machines (SVM)
- However, the use of two separate learning networks means at least doubled training parameters and computational cost compared to using a single network.
- Moreover, fusing sampled frames with SVM does not catch long-range temporal information across consecutive sampled frames.

DYNAMIC HAND GESTURE RECOGNITION BASED ON SHORT-TERM SAMPLING NETWORKS – Zhang, Wang, Lan

- Presents a novel deep learning network for hand gesture recognition. The network integrates several well-proved modules together to learn both short-term and long-term features from video inputs and meanwhile avoid intensive computation.
- To learn short-term features, each video input is segmented into a fixed number of frame groups. A frame is randomly selected from each group and represented as an RGB image as well as an optical flow snapshot. These two entities are fused and fed into a convolutional neural network (ConvNet) for feature extraction.
- The ConvNets for all groups share parameters.
- To learn longterm features, outputs from all ConvNets are fed into a long short-term memory (LSTM) network, by which a final classification result is predicted. The new model has been tested with two popular hand gesture datasets, namely the Jester dataset and Nvidia dataset. The robustness of the model has also been proved with an augmented dataset with enhanced diversity of hand gestures.
- Uses the Jester dataset and Nvidia Dataset
- Achieved an accuracy if around 95%(Jester dataset) and 85%(Nvidia dataset)



Construct Dynamic Graphs for Hand Gesture Recognition via Spatial-Temporal Attention

- ❖ The key idea is to first construct a fully-connected graph from a hand skeleton, where the node features and edges are then automatically learned via a self-attention mechanism that performs in both spatial and temporal domains. In addition, a novel spatial-temporal mask is applied to significantly cut down the computational cost by 99%.

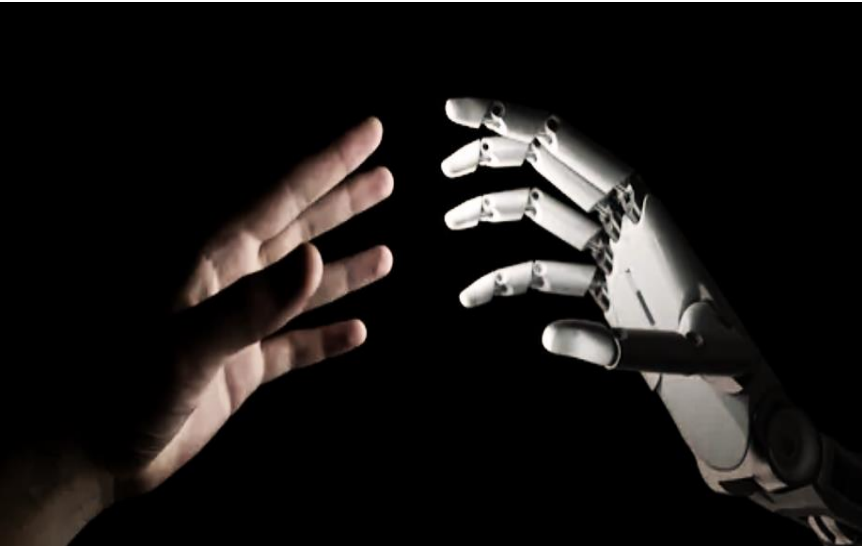
An Efficient PointLSTM for Point Clouds Based Gesture Recognition

- ❖ The key idea is to first construct a fully-connected graph from a hand skeleton, where the node features and edges are then automatically learned via a self-attention mechanism that performs in both spatial and temporal domains. In addition, a novel spatial-temporal mask is applied to significantly cut down the computational cost by 99%.

Hand Gesture Recognition on Cambridge

- ❖ [Fast and Robust Dynamic Hand Gesture Recognition via Key Frames Extraction and Feature Fusion]
- ❖ In this paper they combines image entropy and density clustering to exploit the key frames from hand gesture video for further feature extraction, which can improve the efficiency of recognition. Moreover, a feature fusion strategy is also proposed to further improve feature representation, which elevates the performance of recognition.

Feasibility and Scope



Traditional ways of human-computer interaction, namely keyboards, pointing devices, touchscreens and other physical input tools, are only as useful through the application of pressure through physical touch.

The recent pandemic has driven a major argument on the substantial and often, indispensable need for contactless operation of systems, irrespective of the scale of the industry.

Dynamic Hand Gesture Recognition can turn out to be an excellent solution to this problem, facilitating contact-less interaction with computing devices, at the same time providing dependable and respectable efficiency.



THANK YOU