Sign language recognition using Neural Networks

Motivation:

- Sign language recognition based on static images.
- Explore classification problem with CNN(Convolution Neural Networks), which works independent of engineered features.

Literature survey:

- CNN: Neocognitron- Self organizing Neural Network Models http://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf
- CNN for images: http://dl.acm.org/citation.cfm?id=303704
- Dropout: http://www.cs.toronto.edu/~nitish/msc_thesis.pdf

CNN Training:

1. Data Pre-processing:
    - Resolution:
      A fixed filter of size 5x5 has been used in the architecture. In order to find the resolution that goes well with the current filter, we have experimented with 128x128, 64x64 and 32x32 resolution images.

      While experimenting with architectural parameters, we used 32x32 images since it is computationally faster.

    - Augmentation:
      In order to make the model robust to variance in noise and translation, following operations have been performed on input training images.
        - Rotation
        - Translation
        - Horizontal flip
        - Noise

    - Feature representation:
      Experimentations have been done for the following feature representations
        - YUV: Y-channel provides gradient information which is discriminative feature for sign-language images
        - RGB: TBD
        - Canny-transformed Image: TBD

2. Training:
    - Architectures:
        - Depth of NN:
          In order to make sure that the receptive field of last layer of neurons to cover atleast 80% of the input, we tried 2- and 3-layered convolutions for 32x32 and 128x128 images respectively.
        - Number of feature maps:
          The combination of 16 and 256 feature maps has been used.
        - SGD and BGD:
          We experimented with Batch Gradient Descent on GPU using cuda.
        - Filter size: We used a standard filter size of 5x5

    - Dropouts: (TBD)

3. Results:

Architecture	Train Accuracy	Test Accuracy
yuv(3x128x128)->16x62x62->256x29x29->2-layered MLP(128->26)	100	56
yuv(3x128x128)->16x62x62->256x29x29->2-layered MLP(128->26) - Augmented data	99	55
yuv(3x128x128)->16x49x49->256x9x9->2-layered MLP(128->26)	87	46
yuv(3x64x64)->16x30x30->64x13x13->256x4x4->2-layered MLP(128->26)	100	63

Baseline: Bag of words model
For comparative study, we've trained a model based on Bag of words model. Model learns 600 visual words over the entire database and builds a histogram representation of each image with respect to learnt words. Using histogram as the feature, SVM was used as classifier.
We got 25.32% accuracy.

Regularization methods:
- Dropout: TBD.
- Momentum: TBD.

PS: Except for the things mentioned TBD(To be done), all have been done.