Team ML_Lab

Anthony Allen, Parsh Patel and Caijun Qin

This report delineates an experimental design and project proposal of a machine learning algorithm capable of reading American Sign Language (ASL). The model presented at the end of this semester will use an image dataset created by the authors of this report combined with data supplied by a Special Topics Machine Learning class, EEL4930, at the University of Florida. This dataset provides 20 images from each group (with approximately 10 groups) of each ASL letter from A – I as defined by the American Sign Language University (ASLU) [1]. These images will be preprocessed to ensure they are all the same size and the signer's hand is front-facing. Ideally, this data will provide enough images to segment into training and validation sets. "Ground truth" can be generated by labelling each image with its representative letter.

Accuracy can be accessed via a simple classification approach. A confusion matrix can be generated to assess things like the positive and negative predicted values. Since the training data is "ground truth" this will be easily implementable. Additionally, since the dataset is limited, K-folds cross validation is the preferred approach to partition the images. Data will be partitioned into 5 folds; 3 training and 2 validation sets. Having multiple samples will allow for comparison of different models. A blind test set will be generated by the instructor of the machine learning course. However, a sixth fold can always be generated as a preliminary blind test step as well.

Some issues may ensue because of the nature of the data collection. The data consists of images taken of many different hands from different cameras. Variances in the orientation, scale, illumination, resolution, noise and background clutter can all be expected in this image set. To combat these, a generalized approach must be developed. Optimal correction values can be collected for each image and all of these can be averaged to find the most effective parameter values for all the images.

To extract meaningful features from the images, various image processing techniques can be used. Primarily, it will be important to segment the signer's hand from the rest of the image. This can be done with a simple masking approach or through the use of edge detection. Python offers a number of useful edge detection algorithms. However, for this specific application the Canny edge detection will be used. Unlike the Sobel approach, which identifies a gradient edge, the Canny method represents each shape with a line of single pixel width. Features can be extracted from the remaining shape. The machine learning algorithm will look at features like solidity, area, eccentricity, bounding box, centroid, circularity, equivalent diameter, orientation, extrema and much more. This approach may require heavy involvement of the designer. It will implement an auto correlation matrix to determine the importance of

each feature before a model can be selected. Alternatively, a neural network approach can be implemented. With this approach, various kernels are convolved on an image. Features of this kernel, such as diagonality and size can be collected to determine the best model similarly to the previous method.

References

[1] American Sign Language University, "ASLU," [Online]. Available: https://www.lifeprint.com/. [Accessed 10 February 2020].